12/3/2021 Final Class

Last time: least squares and SVD.

$\underline{A} \in \mathbb{R}^{m \times n}$, $m \geq n$.

goal: solve $\min_{\underline{x}} \| \underline{A}\, \underline{x} - \underline{b} \|_2^2$.

The minimum norm solution $\underline{z}^+ = \underline{V}^T \underline{x}^+$ has components:

$$z_i^+ = \begin{cases} \dfrac{u_i^T b}{\sigma_i} & i = 1, \dots, r \\ 0 & i = r+1, \dots, n. \end{cases}$$

$$\Rightarrow \boxed{\underline{x}^+ = \underline{V}\, \underline{z}^+ = \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} \underline{v}_i .}$$

A concern: small singular values.

We can use the SVD solution to the least squares problem to help us understand the impact of perturbations in $\underline{b}$, i.e. noise in $\underline{b}$.

Suppose: $\underset{\text{perturbed}}{\underline{\tilde{b}}} = \underset{\text{original}}{\underline{b}} + \underset{\text{noise}}{\underline{\varepsilon}}$

in solving $\min_{\underline{x}} \| \underline{A}\, \underline{x} - \underline{\tilde{b}} \|_2^2$, we get the minimum norm solution:

$$\underline{\tilde{x}}^+ = \sum_{i=1}^{r} \frac{\underline{u}_i^T \underline{\tilde{b}}}{\sigma_i} \underline{v}_i$$

$$= \underbrace{\sum_{i=1}^{r} \frac{\underline{u}_i^T \underline{b}}{\sigma_i} \underline{v}_i}_{\underline{x}^+} + \sum_{i=1}^{r} \frac{\underline{u}_i^T \underline{\varepsilon}}{\sigma_i} \underline{v}_i$$

$$\Rightarrow \underset{\text{perturbed solution}}{\underline{\tilde{x}}^+} - \underset{\text{original solution.}}{\underline{x}^+} = \sum_{i=1}^{r} \frac{\underline{u}_i^T \underline{\varepsilon}}{\sigma_i} \underline{v}_i$$

$$\| \underline{\tilde{x}}^+ - \underline{x}^+ \|_2^2 = \sum_{i=1}^{r} \frac{(\underline{u}_i^T \underline{\varepsilon})^2}{\sigma_i^2}.$$

$\leadsto \Delta$

Main point: If σ_i is small, even if the Noise $\|\underline{\varepsilon}\|_2$ is small, the difference between $\underline{\tilde{x}}^+$ and $\underline{x}^+$ might be very large. .

How to fix this: Regularized least squares, i.e. Tikhonov regularization.

$$\min_{\underline{x}} \left(\frac{1}{2} \| \underline{\underline{A}}\, \underline{x} - \underline{b} \|_2^2 + \underbrace{\frac{\lambda}{2} \| \underline{x} \|_2^2}_{\text{additional term.}} \right).$$

$\lambda > 0$

Notice: this can be expressed as a "normal" least squares problem as:

$$\min_{\underline{x}} \frac{1}{2} \left\| \begin{pmatrix} \underline{\underline{A}} \\ \sqrt{\lambda}\, \underline{\underline{I}} \end{pmatrix} \underline{x} - \begin{pmatrix} \underline{b} \\ \underline{0} \end{pmatrix} \right\|_2^2 .$$

For the regularized problem, the solution $\underline{x}_\lambda$ must solve the Normal equations:

$$\underbrace{\begin{pmatrix} \underline{\underline{A}} \\ \sqrt{\lambda}\,\underline{\underline{I}} \end{pmatrix}^T \begin{pmatrix} \underline{\underline{A}} \\ \sqrt{\lambda}\,\underline{\underline{I}} \end{pmatrix}}_{= (\underline{\underline{A}}^T \underline{\underline{A}} + \lambda \underline{\underline{I}})} \underline{x}_\lambda = \underbrace{\begin{pmatrix} \underline{\underline{A}} \\ \sqrt{\lambda}\,\underline{\underline{I}} \end{pmatrix}^T \begin{pmatrix} \underline{b} \\ \underline{0} \end{pmatrix}}_{= \underline{\underline{A}}^T \underline{b}.}$$

$$\Leftrightarrow \underbrace{(\underline{\underline{A}}^T \underline{\underline{A}} + \lambda \underline{\underline{I}})}\, \underline{x}_\lambda = \underline{\underline{A}}^T \underline{b}.$$

$\ast$ { Notice that we have shifted the eigenvalues of $\underline{\underline{A}}^T\underline{\underline{A}}$ "away from 0" by $\lambda > 0$.

Using again the SVD: $\underline{\underline{A}} = \underline{\underline{U}}\,\underline{\underline{\Sigma}}\,\underline{\underline{V}}^T$,

we can compute $\underline{x}_\lambda$:

$$\underline{x}_\lambda = \sum_{i=1}^{r} \frac{\sigma_i (\underline{u}_i^T \underline{b})}{\sigma_i^2 + \lambda} \underline{v}_i ,$$

shifted eigenvalues of $\underline{\underline{A}}^T \underline{\underline{A}}$!

For large singular values ($\sigma_i >> \lambda$)

$$\frac{\sigma_i (\underline{u}_i^T \underline{b})}{\sigma_i^2 + \lambda} \approx \frac{\underline{u}_i^T \underline{b}}{\sigma_i}$$

For very small singular values ($\sigma_i << \lambda$) and sufficiently large λ, we have

$$\sigma_i \frac{(\underline{u}_i^T \underline{b})}{\sigma_i^2 + \lambda} \approx 0.$$

Course recap:

→ vectors, operations, norms.
→ inner products, Cauchy-Schwarz.
→ electrical circuits, Householder reflectors
→ Factorizations
- LU ⟷ Gaussian elimination
- QR ⟷ Gram-Schmidt orthogonalization
- LL^T ⟷ Cholesky.

→ Subspaces, bases, dimension
→ Four Fundamental Subspaces, FTLA.

→ least squares $\left(\begin{array}{l}\text{easy to derive}\\ \text{Normal equations with}\\ \text{FTLA}\end{array}\right)$

↳ $\underline{b} = \underline{b}_R + \underline{b}_N$

$\underline{b}_R \in R(\underline{A})$, $\underline{b}_N \in N(\underline{A}^T)$

→ geometry of the Normal equations

→ projections $\underline{P}^2 = \underline{P}$, $\underbrace{\underline{P} = \underline{P}^T}_{\text{orthogonal.}}$

→ complex #'s.

→ eigenvalues / vectors + algebraic mult.

→ diagonalization. + geometric mult = $\dim N(\underline{A} - \lambda \underline{I})$.

diagonalizable $\Leftrightarrow$ algebraic mult. of each eigenvalue = geometric mult.

→ SVD !!