ME491 Lecture 6.

1.4. Pseudo-Inverse, Least Squares, and Regression

A lot of the times in engineering, we want to solve the following equation :

$$Ax = b$$

Here, this equation represents a linear system of equations.

In ME360, we learned how to use $PA = LU$ to solve the equation exactly, and we know there exists one unique solution.
However, in data-intensive scenarios, often, we can't find an exact or unique solution.

- Under-determined system

$$A \in \mathbb{C}^{n \times m}, \quad n \ll m \qquad (\text{short-fat matrix})$$

- Over-determined system

$$A \in \mathbb{C}^{n \times m}, \quad n \gg m \qquad (\text{tall-skinny matrix})$$

(For more discussion of the solution space of A, check out P20-21)

For the over-determined case, we can find a solution x that minimizes the sum-squared error $\|Ax - b\|_2^2$, the so-called least-square solution, note: $\|Ax - b\|_2$ is also minimized.

We can use SVD to find solution x.

1. Substitute an exact truncated SVD for A.

$$A = \tilde{U} \tilde{\Sigma} \tilde{V}^*$$

2. Invert $\tilde{U}$, $\tilde{\Sigma}$, and $\tilde{V}^*$, resulting.

$$A^{\dagger} \triangleq \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^* \Rightarrow A^{\dagger}A = \tilde{V}\tilde{V}^*$$

Note: $A^{\dagger}A$ will only equal the identity $I_{m \times m}$ if the truncated SVD captures all non-zero singular values. $A^{\dagger}A$ only approximates the identity.

3. Find the least-square solution

$$A\tilde{x} = b \Rightarrow A^{\dagger}A\tilde{x} = A^{\dagger}b \Rightarrow \tilde{x} = \tilde{V}\tilde{\Sigma}^{-1}\tilde{U}^* b$$

(Note, the derivation to get to the solution for $\tilde{x}$ is not the best, and the youtube videos made by book author can also be confusing. If you are interested in this derivation, search "pseudo-inverse" + "SVD")

* Condition Number

In ME 360, we computed condition number of a matrix using the infinite norm. Here, we define the condition number using singular values.

$$K(A) = \frac{\sigma_{max}(A)}{\sigma_{min}(A)}$$

- Condition number is a measure of how sensitive matrix multiplication and inversion are to errors in the input.

Let's remind ourselves the effect of a large condition number.

For linear system of equations $Ax = b$, the vector x is not specified perfectly with some error ϵ_x, then we have

$$A(x + \epsilon_x) = b + \epsilon_b.$$

ϵ_b: corresponding error in b.

In the worst case scenario, we have

$$A(x+\epsilon_x) = \underbrace{\sigma_{min} x}_{b} + \underbrace{\sigma_{max} \epsilon_x}_{\epsilon_b}$$

$$\frac{\|b\|}{\|\epsilon_b\|} = \frac{\|x\|}{\|\epsilon_x\|} \times \frac{\sigma_{min}}{\sigma_{max}} = \frac{\|x\|}{K(A)\|\epsilon_x\|}$$

If $K(A)$ is big, then the signal-to-noise ratio is small, which is undesirable.

Similarly, if b has an error ϵ_b, the estimated solution for $x+\epsilon_x$ is

$$x + \epsilon_x \approx A^{+}(b+\epsilon_b) = \frac{1}{\sigma_{max}} b + \frac{1}{\sigma_{min}} \epsilon_b$$

One-Dimensional Linear Regression

Regression is an important tool to find the general trend in data, and there are many ways to go about it. Here, we show how to use SVD to do 1D Linear Regression, and we will show different methods in the future.

$$\begin{bmatrix} | \\ b \\ | \end{bmatrix} = \begin{bmatrix} | \\ a \\ | \end{bmatrix} x = \hat{U}\tilde{\Sigma}\tilde{V}^* x \Rightarrow x = \tilde{V}\tilde{\Sigma}^{-1}\hat{U}^* b$$

Here: $\tilde{\Sigma} = \|a\|_2$, $\tilde{V} = 1$, $\hat{U} = \frac{a}{\|a\|_2} \Rightarrow x = \frac{a^* b}{\|a\|_2^2}$