9. Recap and Exercises

Gram Schmidt Proces:

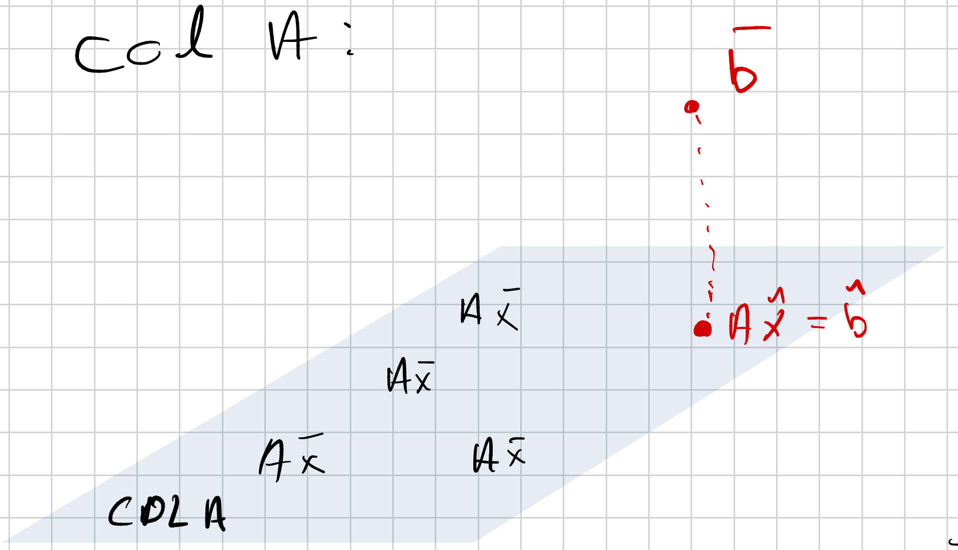
Given a basis, GSP can turn it into a orthogonal basis:

Let $\{\bar{x}_1 \ldots \bar{x}_p\}$ be a lin. ind. set:

$$\bar{v}_p = \bar{x}_p - \left(\sum_{i=1}^{p-1} \mathrm{proj}_{\bar{v}_i} x_p\right) = \bar{x}_p - \left(\sum_{i=1}^{p-1} \frac{\bar{x}_p \cdot \bar{v}_i}{\bar{v}_i \cdot \bar{v}_i}\, \bar{v}_i\right)$$

Least squares:

Given an inconsistent system $A\bar{x} = \bar{b}$ Least squares allows us to find the best approximate solution $\hat{x}$ such that $A\hat{x}$ has the smallest distance from $\bar{b}$ to col A:



Method:

last column

$$\text{Solve } \left[\, A^TA \quad A^T\bar{b} \,\right] \overset{RREF}{\sim} \left[\, I \quad \hat{x} \,\right]$$

$$\hat{x} = (A^TA)^{-1} \cdot (A^T b)$$ The normal Equation

Linear Models:

We use least squares to find the best fitted model for our data

Given n two tuples (x_i, y_i), let

$$\bar{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and let} \quad \bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix},$$

the design matrix X for $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

is $\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$ and the observation matrix is $\bar{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

$$\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} \quad X\hat{\beta} \quad \begin{matrix} \hat{\beta}_0 + \hat{\beta}_1 x_1 \\ \vdots \end{matrix} = \hat{y}$$

$$\bar{e} = (\bar{y} - \hat{y})$$

In order to estimate $\begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$ we solve

$$\begin{bmatrix} X^T X & X^T \bar{y} \end{bmatrix} \overset{RREF}{\sim} \begin{bmatrix} I & \begin{matrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{matrix} \end{bmatrix}$$

last column

We can modify the design Matrix to accommodate our hypothesised models

We can find $\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{bmatrix}$ by either

calculating each $\hat{y}_i = \beta_0 + \beta_1 x_i$ or all the $\hat{y}$'s

by doing $X\hat{\beta} = \hat{y}$

We obtain the error by finding

$$e = \| \bar{y} - X\hat{\beta} \|$$