1 General Least Square

Let $c_0, c_1, \ldots, c_n \in \mathbb{R}$ be unknown coefficients. We are given a sample of size s of measurements (z_i, y_i) where $z_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}$ for $i = 1, \ldots, s$. Furthermore let $f_1, \ldots, f_n$ be functions $f_k : \mathbb{R}^d \to \mathbb{R}$ for $k = 1, \ldots, n$. We assume that the relation between z and y is of the form

$$\sum_{k=1}^{n} c_0 + f_k(z) c_k \approx y.$$

Please set up the design matrix of the problem

$$\min_{c_0, c_1, \ldots, c_n} \sum_{i=1}^{s} \left(\sum_{k=1}^{n} c_0 + f_k(z_i) c_k - y_i \right)^2 .$$

Solution:

- Given: s measurements $(z_i, y_i),\ i = 1, \ldots, s$
- Model: $f(z) := \sum_{k=1}^{n} c_0 + f_k(z) c_k \approx y$
- Unknown coefficients: $x := (c_0, c_1, \ldots, c_n) \in \mathbb{R}^{n+1}$
- LQ-formulation:

$$\begin{aligned}
i = 1 : \quad & \sum_{k=1}^{n} c_0 + f_k(z_1) c_k = f(z_1) \overset{!}{=} y_1 \\
& \vdots \\
i = s : \quad & \sum_{k=1}^{n} c_0 + f_k(z_s) c_k = f(z_s) \overset{!}{=} y_s \\
\Leftrightarrow \quad & \underbrace{\begin{pmatrix} 1 & f_1(z_1) & \cdots & f_n(z_1) \\ \vdots & \vdots & \ddots & \vdots \\ 1 & f_1(z_s) & \cdots & f_n(z_s) \end{pmatrix}}_{=:A} \underbrace{\begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_n \end{pmatrix}}_{=:x} = \underbrace{\begin{pmatrix} y_1 \\ \vdots \\ y_s \end{pmatrix}}_{=:b} \\
\Leftrightarrow \quad & Ax = b
\end{aligned}$$

Thus $\hat{x}$ is LQ solution if it minimizes

$$\min_{x \in \mathbb{R}^{n+1}} \|Ax - b\|_2^2.$$

We find:

$$\begin{aligned}
\|Ax - b\|_2^2 &= \sum_{i=1}^{s} [Ax - b]_i^2 = \sum_{i=1}^{s} (f(z_i) - y_i)^2 \\
&= \sum_{i=1}^{s} \underbrace{\left(\sum_{k=1}^{n} c_0 + f_k(z_i) - y_i \right)^2}_{\in \mathbb{R} \ (\text{note: for } r \in \mathbb{R}:\ r^2 = \|r\|_2^2)}
\end{aligned}$$