Maximum Likelihood Estimator for the Linear Model with White Gaussian Noise

In many cases we are given a set of noisy observations \( x \in \mathbb{R}^m \) that can be described as a linear combination of some parameter \( \theta \in \mathbb{R}^n \). Our goal is to estimate \( \theta \):

\[
    x = H \theta + w, \quad x, w \in \mathbb{R}^m, \theta \in \mathbb{R}^n, H \in \mathbb{R}^{m \times n},
\]

\[
    w \sim N(0, I), m > n.
\]

where \( x \) is a known vector of observations, \( H \) is a known matrix, \( \theta \) is the parameter we want to estimate and \( w \) is unknown white Gaussian noise (\( w \) is white noise if \( \text{cov}(w_i, w_j) = 0 \)).

When the PDF (probability density function) is viewed as a function of the unknown parameter (with \( x \) fixed) it is termed the likelihood function. The maximum likelihood estimator \( \theta^* \) is defined as:

\[
    \theta^* = \arg \max_{\theta} p(x; \theta)
\]

For the linear model with white Gaussian noise \( w \sim N(0, I) \),

\[
    p(x; \theta) = \frac{1}{(2\pi)^{m/2}} e^{-(x-H\theta)^T(x-H\theta)/2}
\]

Hence

\[
    \theta^* = \arg \max_{\theta} p(x; \theta) = \arg \max_{\theta} \ln p(x; \theta) = \arg \max_{\theta} \frac{(x-H\theta)^T(x-H\theta)}{2} = \arg \min_{\theta} \|x-H\theta\|^2_2
\]

In other words, the ML (maximum likelihood) estimator for \( \theta \) minimizes the square error \( \|x-H\theta\|^2_2 \). We saw in class that \( \theta^* \) that minimizes the square error satisfies the normal equations:

\[
    H^T H \theta^* = H^T x
\]
Therefore the ML estimator for the linear model is:

$$\theta^* = (H^T H)^{-1} H^T x$$  \hspace{1cm} (6)

$\theta^*$ is a random variable (because it is a function of $x$ which is a random variable).

The expectation of $\theta^*$ is:

$$E(\theta^*) = E((H^T H)^{-1} H^T x) = (H^T H)^{-1} H^T E(x) =$$

$$= (H^T H)^{-1} H^T E(H \theta + w) = (H^T H)^{-1} H^T H \theta = \theta$$  \hspace{1cm} (7)

We say that $\theta^*$ is unbiased because $E(\theta^*) = \theta$.

The variance of $\theta^*$ is:

$$\text{var}(\theta^*) = \text{var}((H^T H)^{-1} H^T x) = (H^T H)^{-1} H^T \text{var}(x)((H^T H)^{-1} H^T) =$$

$$= (H^T H)^{-1} H^T \text{var}(H \theta + w) H (H^T H)^{-1} = (H^T H)^{-1} H^T I H (H^T H)^{-1} =$$

$$= (H^T H)^{-1}$$  \hspace{1cm} (8)