Proof: Asymptotic Covariance Reduction in Linear Regression

Objective

To prove that in standard linear regression with fixed observation noise variance σ^2, the covariance of the maximum likelihood estimator $\hat{\theta}_{\mathrm{ML}}$ decreases as the number of samples $D \to \infty$. Specifically,

$$\mathrm{Cov}(\hat{\theta}_{\mathrm{ML}}) \approx \frac{\sigma^2}{D} Q^{-1}, \quad \text{where } Q = \mathbb{E}[\phi(\zeta)\phi(\zeta)^\top]$$

1. Setup

We consider a standard linear regression model:

$$\gamma^{(d)} = \phi(\zeta^{(d)})^\top \theta^* + \varepsilon^{(d)}, \quad \varepsilon^{(d)} \sim \mathcal{N}(0, \sigma^2), \quad d = 1, \dots, D$$

Define the design matrix:

$$\Phi = \begin{bmatrix} \phi(\zeta^{(1)})^\top \\ \phi(\zeta^{(2)})^\top \\ \vdots \\ \phi(\zeta^{(D)})^\top \end{bmatrix} \in \mathbb{R}^{D \times B}, \quad \Gamma = \begin{bmatrix} \gamma^{(1)} \\ \gamma^{(2)} \\ \vdots \\ \gamma^{(D)} \end{bmatrix} = \Phi\theta^* + \varepsilon$$

where $\varepsilon \sim \mathcal{N}(0, \sigma^2 I_D)$.

2. Maximum Likelihood Estimator

The MLE (least-squares) solution is given by:

$$\hat{\theta}_{\mathrm{ML}} = (\Phi^\top\Phi)^{-1}\Phi^\top\Gamma$$

Substitute $\Gamma = \Phi\theta^* + \varepsilon$:

$$\hat{\theta}_{\mathrm{ML}} = \theta^* + (\Phi^\top\Phi)^{-1}\Phi^\top\varepsilon$$

3. Covariance of $\hat{\theta}_{\mathrm{ML}}$

Taking covariance on both sides:

$$\mathrm{Cov}(\hat{\theta}_{\mathrm{ML}}) = \mathrm{Cov}\left((\Phi^\top\Phi)^{-1}\Phi^\top\varepsilon\right)$$

Using the linear transformation rule:

$$\mathrm{Cov}(A\varepsilon) = A \cdot \mathrm{Cov}(\varepsilon) \cdot A^\top$$

Let $A = (\Phi^\top\Phi)^{-1}\Phi^\top$, and $\mathrm{Cov}(\varepsilon) = \sigma^2 I_D$, then:

$$\mathrm{Cov}(\hat{\theta}_{\mathrm{ML}}) = \sigma^2(\Phi^\top\Phi)^{-1}\Phi^\top\Phi(\Phi^\top\Phi)^{-1} = \sigma^2(\Phi^\top\Phi)^{-1}$$

4. Asymptotic Behavior as $D \to \infty$

Assume the samples $\zeta^{(d)}$ are i.i.d., and define:

$$Q := \mathbb{E}[\phi(\zeta)\phi(\zeta)^\top] \in \mathbb{R}^{B\times B}$$

Then by the **Law of Large Numbers**:

$$\frac{1}{D}\Phi^\top\Phi = \frac{1}{D}\sum_{d=1}^{D}\phi(\zeta^{(d)})\phi(\zeta^{(d)})^\top \xrightarrow{a.s.} Q$$

So for large D:

$$\Phi^\top\Phi \approx DQ \quad \Rightarrow \quad (\Phi^\top\Phi)^{-1} \approx \frac{1}{D}Q^{-1}$$

Hence:

$$\mathrm{Cov}(\hat{\theta}_{\mathrm{ML}}) = \sigma^2(\Phi^\top\Phi)^{-1} \approx \frac{\sigma^2}{D}Q^{-1}$$

5. Conclusion

As $D \to \infty$, the covariance of the parameter estimate $\hat{\theta}_{\mathrm{ML}}$ decreases proportionally to $1/D$, and we have:

$$\hat{\theta}_{\mathrm{ML}} \xrightarrow{p} \theta^* \quad \text{(consistency)}$$

Thus, with fixed noise variance σ^2, increasing the dataset size makes the estimate more concentrated and accurate.