ECE531 Screencast 6.5: Sequential LMMSE Estimation

D. Richard Brown III

Worcester Polytechnic Institute

Introduction

Recall the LMMSE estimator

$$\hat{\theta}_{\mathsf{LMMSE}}(y) = \mathrm{E}[\Theta] + \mathrm{cov}(\Theta, Y)\left[\mathrm{cov}(Y,Y)\right]^{-1}(Y - \mathrm{E}[Y]).$$

Suppose we wish to compute this each time we get a new sample. Denoting $\hat{\theta}_{\mathsf{LMMSE}}[n]$ as the LMMSE estimator based on observations $Y[n] = [Y_0, \dots, Y_n]$, can can write

$$\begin{aligned}
\hat{\theta}_{\mathsf{LMMSE}}[-1] &= \mathrm{E}[\Theta] \text{ (no observations)} \\
\hat{\theta}_{\mathsf{LMMSE}}[0] &= \mathrm{E}[\Theta] + \mathrm{cov}(\Theta, Y[0])\left[\mathrm{cov}(Y[0], Y[0])\right]^{-1}(Y[0] - \mathrm{E}[Y[0]]) \\
\hat{\theta}_{\mathsf{LMMSE}}[1] &= \mathrm{E}[\Theta] + \mathrm{cov}(\Theta, Y[1])\left[\mathrm{cov}(Y[1], Y[1])\right]^{-1}(Y[1] - \mathrm{E}[Y[1]]) \\
&\vdots
\end{aligned}$$

The problem here is that each time we get a new observation, we have to recompute the covariances and the matrix inverse. This is going to become computationally prohibitive when n gets large.

Sequential LMMSE Estimation: Main Idea

Suppose you already have computed $\hat{\theta}_{\mathsf{LMMSE}}[n-1]$. The next observation Y_n arrives. We would like to have a computationally efficient way of computing $\hat{\theta}_{\mathsf{LMMSE}}[n]$ as a function of $\hat{\theta}_{\mathsf{LMMSE}}[n-1]$ and Y_n, i.e.

$$\hat{\theta}_{\mathsf{LMMSE}}[n] = f(\hat{\theta}_{\mathsf{LMMSE}}[n-1], Y_n)$$

To proceed with this idea, we need to assume we have a linear model

$$\underbrace{\begin{bmatrix} Y_0 \\ \vdots \\ Y_n \end{bmatrix}}_{Y[n]} = \underbrace{H[n]}_{(n+1)\times p} \underbrace{\Theta}_{p\times 1} + \underbrace{\begin{bmatrix} W_0 \\ \vdots \\ W_n \end{bmatrix}}_{W[n]}$$

with $W[n] \sim \mathcal{N}(0, \sigma^2 I)$ and independent of the parameter Θ and $H[n]$ is a known mixing matrix. Note that each of these matrices/vectors (except Θ) gets a new row each time a new observation arrives.

Sequential LMMSE Estimation: Main Result

Partition $H[n]$ as

$$H[n] = \begin{bmatrix} H[n-1] \\ h^\top[n] \end{bmatrix}$$

where $H[n-1] \in \mathbb{R}^{n \times p}$ and $h^\top[n] \in \mathbb{R}^{1 \times p}$. Then

$$\hat{\theta}_{\text{LMMSE}}[n] = \hat{\theta}_{\text{LMMSE}}[n-1] + K[n]\left(Y_n - h^\top[n]\hat{\theta}_{\text{LMMSE}}[n-1]\right)$$

with

$$K[n] = \frac{\Sigma[n-1]h[n]}{\sigma^2 + h^\top[n]\Sigma[n-1]h[n]} \in \mathbb{R}^{p \times 1}$$

representing the "Kalman gain" and with

$$\Sigma[n] = (I - K[n]h^\top[n])\Sigma[n-1] \in \mathbb{R}^{p \times p}$$

representing the estimation error covariance of our LMMSE estimator $\hat{\theta}_{\text{LMMSE}}[n]$.

Sequential LMMSE Estimation: Intuition

We have

$$\hat{\theta}_{\mathsf{LMMSE}}[n] = \hat{\theta}_{\mathsf{LMMSE}}[n-1] + K[n]\left(Y_n - h^\top[n]\hat{\theta}_{\mathsf{LMMSE}}[n-1]\right)$$

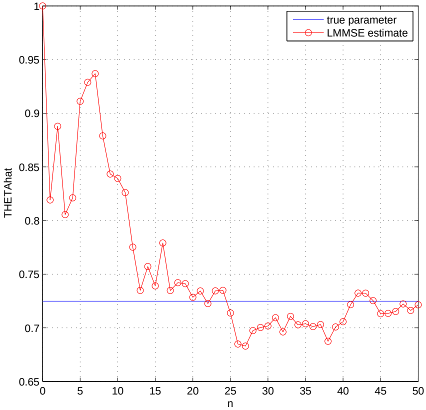
Note that $h^\top[n]\hat{\theta}_{\mathsf{LMMSE}}[n-1]$ can be thought of as a one-step prediction of Y_n given observations $Y_0, \dots, Y_{n-1}$.

The term $Y_n - h^\top[n]\hat{\theta}_{\mathsf{LMMSE}}[n-1]$ is the prediction error and is often called the "innovation".

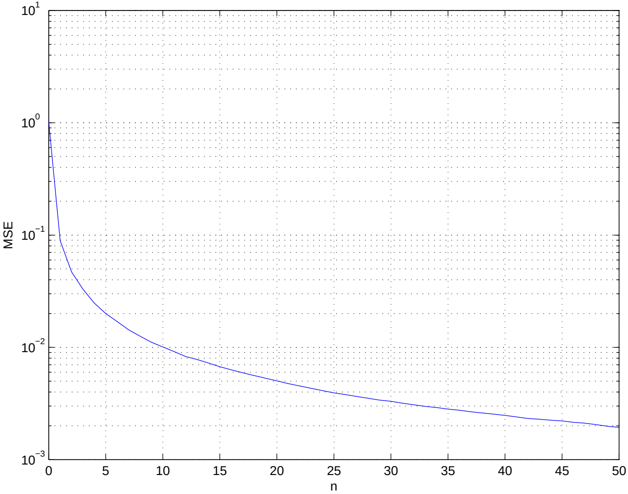
- If we predicted Y_n perfectly, then $\hat{\theta}_{\mathsf{LMMSE}}[n] = \hat{\theta}_{\mathsf{LMMSE}}[n-1]$.
- If our prediction is imperfect, you can think of $K[n]$ as providing the weighting/gain in how much the innovation updates our sequential LMMSE estimate.
- $K[n]$ becomes small when the estimation error covariance $\Sigma[n-1]$ becomes small.

Example: Sequential LMMSE Estimation

Random scalar parameter $\Theta \sim \mathcal{N}(1,1)$. Observation model: $Y_k = \Theta + W_k$ where $\Theta \in \mathbb{R}$ and $W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 0.1)$ independent of parameter.



Example: Sequential LMMSE Estimation



Averaged over 5000 runs.