Readings on Bayesian Estimation

Kay Chapters 10, 11, and 12. Moon & Stirling Sections 12.5–12.7. Anderson & Moore, Section 5.2 Pay particular attention to Kay Sections 10.5–10.6 and Appendix 10A (see the footnote below), as you will be assumed to know this material well. The Anderson & Moore discussion provides a brief, but intense, summary of the Bayes-optimal linear estimator, focussing on the geometric (i.e., Hilbert Space) interpretation of the linear MMSE estimation.

Note that chapter 14 of Kay provides a very useful summary of all (Bayesian and non-Bayesian) estimators discussed in the text.

Problems

1. Kay 10.1
2. Kay 10.3
3. Kay 10.4
4. Kay 11.1
5. Kay 11.2
6. Kay 11.10
7. Kay 11.18
8. Give conditions under which the Bayesian solution shown in Kay Equation (10.28) reduces to the Gauss–Markov solution shown in Kay (6.19). (Prove that your conditions result in (10.28) becoming equal to (6.19).)

9. Given the availability of a vector of real observations, y, consider the inverse problem,

\[ y = A\theta + w \]

where the known matrix A has full column-rank and the unobservable random vectors \( \theta \) and \( v \) are independent. Assume that \( \theta \sim N(\theta_0, \Sigma_0) \) and \( w \sim N(0, W) \), where \( \Sigma_0 \) and \( W \) are both full–rank. Notice that \( y \) must be a gaussian random variable (in fact \( y \) and \( \theta \) are jointly gaussian) and (more importantly for our purposes) \( \theta \) is conditionally gaussian when conditioned on the observation \( y \).\(^1\) This means that projected along any direction, \( d \), in \( \theta \)--parameter space the resulting univariate \( \theta \)--projection, \( d^T\theta \), is \( y \)--conditionally gaussian and hence conditionally unimodal at, and symmetric about, the \( d \)--projected conditional mean \( d^T\mathbb{E}\{\theta|y\} \). As a consequence of this fact the \( y \)--conditional mean, median (along any direction \( d \)), and mode of \( \theta \) coincide.

\(^1\)This is a general result: if \( p(y, \theta) \) is gaussian, then so is \( p(\theta|y) \). If you don’t know this important fact, read the very complete discussion and derivation given in Kay.
(a) Use the coincidence of the \( y \)-conditional mean and mode of \( \theta \) to show that the optimal minimum-mean-square estimator (mmse) of \( \theta \) can be found as a solution to a weighted least-squares problem. (Expressly derive and write out the quadratic loss function).

(b) Solve the resulting weighted least-squares problem. Show that for \( \Sigma_0 = \sigma_0^2 \cdot I \) and \( \sigma_0^2 \to \infty \) the solution reduces to the Gauss–Markov solution obtained for the classical statistical situation where the unknown \( \theta \) is assumed to be deterministic.

10. Kay 12.1

11. Kay 12.8

12. **Classical Oral Exam Questions.** Consider the completely nongaussian, but finite second–order moments, model

\[ y = x + n \]

where \( x \) and \( n \) are uncorrelated, zero–mean scalar random variables with variances \( \sigma_x^2 \) and \( \sigma_n^2 \) respectively. The scalar measurement, \( y \), is known, \( x \) and \( n \) are unknown.

(a) Fast! At the blackboard!! Under oral examination pressure!!! *Directly* derive the optimal linear mean-square estimators of \( x \) and \( n \) in terms of the quantities given above. For any one of the unknowns (\( x \), say) what is the relevant Wiener–Hopf equation?; How does it arise from an orthogonality principle?; What is the geometric interpretation of the orthogonality principle?

(b) What optimality properties hold for the estimators of \( x \) and \( n \), if \( x \) and \( y \) are now jointly gaussian?

(c) What are the various ways that one can estimate a random variable \( x \) given another random variable \( y \) (not necessarily linearly related) for all combinations of the cases \( x \) deterministic, \( x \) random, \( y \) deterministic, \( y \) random. In particular discuss Least–Squares, UMVUE, MLE, Gauss–Markov, MAP, and MMSE estimators under various modeling assumptions (e.g., Gaussian, Non–Gaussian, linearly related as in the first part of this problem, nonlinearly related, etc.)

13. Compute the optimal linear minimum mean-square error estimator of \( x \) given a measurement \( y \). Now compute the linear minimum mean-square error estimator of \( x \) given \( z = Ay + b \), where \( A \) is invertible. Finally, show that the two estimators are equivalent. Note that this problem justifies the claim made in class that linear regression on the innovations sequence is equivalent to linear regression on the original data sequence.

\[ ^2 \text{i.e., do not derive the general vector-matrix case first and then specialize to the scalar case. When students do this the committee members tend to interpret this as "solving from memory" rather than "solving from understanding." (I.e., a desperate clinging to known rote patterns.) Of course, you also want to be able to solve the general vector–matrix case too!} \]