Supplementary exercises

C1 *Monte Carlo study of the law of large numbers and the central limit theorem.*

Using, e.g. $M = 1000$, realizations, plot the sample mean of a sequence $\{v(t)\}_{t=1}^{N}$ of random variables as a function of the sample size N. The diagram should contain the mean (over the Monte Carlo simulations) of the sample means as well as the spread (over the Monte Carlo simulations) in the form of 3 times the standard deviation of the sample means of the M different realizations.

Use the following random variables $v(t)$:

a) Gaussian white noise with zero mean and variance 1

b) Uniformly distributed white noise and variance 1

c) Binary distributed white noise and variance 1

d) Filtered white noise according to

$$v(t) = e(t) + ae(t-1) \tag{1}$$

where $a = 0.95$ and $a = -0.95$, and where $e(t)$ has the distributions in a)–c).

e) Plot the histogram for the sample mean, normalized by $\sqrt{N}$, of $\{v(t)\}_{t=1}^{N}$, i.e.

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} v(t)$$

for the different Monte Carlo realizations and for different values of N. Use the different cases in a)–d). For which sample size does the distribution start to look normal? Does this depend on the distribution of the noise sequence? You can also use `kstest` to make a hypothesis test when the distribution start to look normal.

f) **(12 credit course only)** Derive an expression for the covariance of

$$\frac{1}{\sqrt{N}} \sum_{t=1}^{N} v(t)$$

as $N \to \infty$ when $\{v\}$ is a stationary sequence. Use this expression to explain why the results for (1) are so different for $a = 0.95$ and $a = -0.95$.

C2 *Optimal filtering*

Consider a system where, for $-\infty < t < \infty$,

$$y(t) = e(t) + ce(t-1)$$

where $\{e(t)\}$ is zero mean white noise with variance λ_e.

a) Suppose that $|c| < 1$. Determine the optimal one-step ahead predictor $\hat{y}(t+1|t)$ of $y(t+1)$ given $y(t), y(t-1), \ldots$.
What is the corresponding mean-square estimation error $\mathrm{E}[(y(t+1) - \hat{y}(t+1|t))^2]$?

Can the same predictor be used when $|c| = 1$ or $|c| > 1$?

b) Determine the optimal linear predictor of $y(t+1)$ given $y(t), y(t-1), \ldots, y(0)$ and the corresponding mean-square estimation error.

Hint: The system can be written on state-space form

$$\begin{aligned} x(t+1) &= e(t) \\ y(t) &= cx(t) + e(t) \end{aligned}$$

Observe that the optimal estimate of $y(t+1)$ is $c\hat{x}(t+1|t)$ where $\hat{x}(t+1|t)$ is the optimal state-estimate given $y(t), y(t-1), \ldots, y(0)$. Compute $\hat{x}(t+1|t)$ from the Kalman filter equations which, for the system

$$\begin{aligned} x(t+1) &= Fx(t) + Gw(t) \\ y(t) &= H^T x(t) + v(t) \end{aligned}$$

where $\mathrm{E}[x(0)] = x_o$, $\mathrm{E}[x(0)x^T(o)] = P_o$ and where $\mathrm{E}\left[\begin{bmatrix} w_k \\ v_k \end{bmatrix}\right] = 0$,

$$\mathrm{E}\left[\begin{bmatrix} w_k \\ v_k \end{bmatrix}\begin{bmatrix} w_k \\ v_k \end{bmatrix}^T\right] = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix}\delta_{k,l}, \text{ are given by}$$

$$\begin{aligned}
\hat{x}(0|-1) &= x_o \\
\hat{x}(t+1|t) &= F\hat{x}(t|t-1) + K(t)(y(t) - H^T\hat{x}(t|t-1)) \\
K(t) &= (F\Sigma(t|t-1)H + GS)(H^T\Sigma(t|t-1)H + R)^{-1} \\
\Sigma(t+1|t) &= (F - K(t)H^T)\Sigma(t|t-1)(F - K(t)H^T)^T + GQG^T \\
&\quad + K(t)RK^T(t) - GSK^T(t) - K(t)S^TG^T \\
\Sigma(0|-1) &= P_o
\end{aligned}$$

What is the Kalman filter if it is assumed that $y(0) = e(0)$ and

$$y(t) = e(t) + ce(t-1)$$

for $t = 1, 2, \ldots$. Also, derive the Kalman filter in this case by simple hand calculations.

c) Suppose that $\{e(t)\}$ is a Gaussian sequence. What is the optimal predictor of $y(t+1)$ given $y(t), y(t-1), \ldots, y(0)$?

d) Suppose that $|c| < 1$. Show that the Kalman filter converges to the time-invariant filter in a) as $t \to \infty$.

e) Suppose that $|c| > 1$. Show that the Kalman filter still converges. What is the corresponding time-invariant filter? What is the corresponding mean-square estimation error?

f) Suppose that $|c| = 1$. How does the Kalman filter behave then?

What is the conclusion from a) and f) regarding systems having zeros on the unit circle?

g) Suppose that $e(t)$ takes on the values ± 5 with equal probability and suppose that $c = 0.5$.

Determine the MMSE estimator of $y(1)$ given $y(0)$. Is it linear in the observations? Compare its performance with the Kalman filter. Also, derive the Kalman filter estimator for this problem directly.

C3 *Covariance matching*

Consider the system

$$y(t) = e(t) + ce(t-1)$$

where $e(t)$ is zero mean white noise with $\mathrm{E}[e^2(t)] = \lambda_e$ and $\mathrm{E}[e^4(t)] = \mu_e$, and where c is unknown.

a) What is $\mathrm{E}[y(t)y(t-1)]$?

b) Suppose that the noise variance λ_e is known. Use the result in a) to construct a simple unbiased estimator $\hat{c}_N$ of c.

c) Determine the variance of the estimator in b).

d) Implement your estimator in MATLAB and check your calculations in c) using Monte Carlo simulations.

e) Use the prediction error method (PEM) to estimate c (use the command `armax` in MATLAB. Compare the sample covariances of the estimator in b) and the PEM. Which one is better?

The result shows that there are more information regarding c in $\{y(t)\}_{t=1}^{N}$ than in the sample covariance for lag 1.

C4 *Sufficient statistics*

a) Consider an AR-1 system where, for $-\infty < t < \infty$,

$$y(t) + ay(t-1) = e(t)$$

where $e(t)$ is a zero mean Gaussian white noise with variance λ_e. Show that there exists a sufficient statistic $S(y^N) \in \mathbb{R}^3$ for $y^N := \{y(t)\}_{t=0}^N$, i.e. the $N+1$ measurements can be condensed into a three dimensional statistic without information loss.

Hint: Derive the probability density function of y^N by repeated use of Bayes' formula

$$p(y^t) = p(y_t|y^{t-1})p(y^{t-1})$$

b) Consider an MA-1 system

$$\begin{aligned} y(t) &= e(t) + ce(t-1), \quad t = 1, 2, \dots \\ y(0) &= e(0) \end{aligned}$$

where $e(t)$ is a zero mean Gaussian white noise with variance λ_e. Derive the probability density function of $y^N := \{y(t)\}_{t=0}^N$. Does it seem possible to find a low dimensional sufficient statistic as in a)?

C5 *Nonparametric identification*

Determine a non-parametric model for the data set `dataset1`.

How many resonance peaks does the system have?

Notice that the frequency response can be plotted with confidence intervals using `bode(model,'SD',3)`.

C6 *Parametric identification*

Determine a parametric model for the data set `dataset1` using prediction error identification.

How many resonance peaks does the system have?

Notice that the frequency response can be plotted with confidence intervals using `bode(model,'SD',3)`.

Compare your result with what you obtained in C5.