Solutions to Final Exam

Problem 1 *(9 points)* Explain why each of the following is NOT a valid autocorrelation function for a continuous-time, real-valued WSS random process:
(a) $R_X(\tau) = \begin{cases} e^{-\tau} & \tau \geq 0 \\ e^{2\tau} & \tau < 0 \end{cases}$ (b) $R_X(\tau) = e^{-|\tau|} I_{\{|\tau| \leq 1\}}$ (c) $R_X(\tau) = \frac{1+\tau^2}{1+\tau^4}$.

(a) Not symmetric: $R_X(\tau) \neq R_X(-\tau)$.
(b) Not continuous, even though it is continuous at zero. Violates equivalence of (i') and (iii') in Proposition 7.1.9.
(c) Violates Schwarz inequality, because, for example, $R_X(0) = 1 < R_X(\sqrt{0.5}) = 1.2$, or $R_X(\tau) = (1+\tau^2)(1-\tau^4+\tau^8+\cdots) = 1+\tau^2+O(\tau^4) > R(0)$ for all sufficiently small, nonzero values of τ. Alternatively, we could note that if R were the autocorrelation function of a random process X, then X would be mean square differentiable (because R_X is twice continuously differentiable) with $0 \leq R_{X'}(0) = -R_X''(0) = -2$, which is impossible. So, by argument by contradiction, R_X is not a valid autocorrelation function.

Problem 2 *(8 points)* (a) Suppose Z is a $N(\mu, \sigma^2)$ random variable. Express $E[Z^3]$ in terms of μ and σ^2.
(b) Suppose $\begin{pmatrix} X \\ Y \end{pmatrix}$ is a $N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right)$ random vector, where $|\rho| < 1$. Express $E[X^3|Y]$ in terms of ρ and Y.

(a) Z has the same distribution as $W + \mu$, where W is a $N(0, \sigma^2)$ random variable. Since $E[W] = 0$ and $E[W^3] = 0$, $E[Z^3] = E[W^3 + 3W^2 u + 3W\mu^2 + \mu^3] = 3\sigma^2\mu + \mu^3$. An alternative approach is to use the characteristic function of Z.
(b) The conditional distribution of X given $Y = y$ is $N(\rho y, 1-\rho^2)$. Therefore, the answer to this part is obtained by replacing μ by ρY and σ^2 by $1-\rho^2$ in the answer to part (a). That is, $E[X^3|Y] = 3\rho(1-\rho^2)Y + \rho^3 Y^3$.

Problem 3 *(12 points)* (This problem uses the notation $(f,g) = \int_{-\infty}^{\infty} f(t)g^*(t)dt$, and $||f|| = \sqrt{(f,f)}$. Recall that a real-valued Gaussian white noise process $Y = (Y_t : t \in \mathbb{R})$ with parameter $\sigma^2 = 1$ is a generalized random process with mean zero, $R_Y(\tau) = \delta(\tau)$, and $S_Y(\omega) = 1$ for all ω. The interpretation is that all random variables of the form (f, Y) are jointly Gaussian, mean zero, and $E[(f,Y)(g,Y)^*] = (f,g)$. The focus of this problem is an ordinary random process X which is approximately a white noise process.)
Let X be a real-valued stationary Gaussian process with mean zero and $R_X(\tau) = \frac{\alpha}{2}e^{-\alpha|\tau|}$. Note that for α large, R_X is an approximation to a delta function because it is narrow, supported near zero, and integrates to one.
(a) *(2 pts)* Give S_X and verify that for any fixed ω, $S_X(\omega) \to 1$ as $\alpha \to \infty$.
(b) *(5 pts)* For real-valued functions f and g, express $E[(X,f)(X,g)]$ in terms of S_X, $\widehat{f}$, and $\widehat{g}$.
(c) *(5 pts)* Suppose f and g are real-valued baseband functions, so that $\widehat{f}(\omega) = 0$ and $\widehat{g}(\omega) = 0$ for $|\omega| \geq \omega_o$. Give a sufficient condition involving α and ω_o to ensure that
$$|E[(f,X)(g,X)] - (f,g)| \leq (0.01)||f|| \cdot ||g||.$$

(a) $S_X(\omega) = \frac{\alpha^2}{\omega^2+\alpha^2} \to 1$ as $\alpha \to \infty$.
(b)

$$\begin{aligned} E[(f,X)(g,X)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)R_X(s-t)g(t)dsdt = (\widetilde{f} * R_X * g)(0) \\ &= \int_{-\infty}^{\infty} (\widehat{f}(\omega))^* S_X(\omega)\widehat{g}(\omega)\frac{d\omega}{2\pi} \quad \text{(by inverse transform)} \end{aligned}$$

Or, a slightly different approach is:

$$\begin{aligned} E[(f,X)(g,X)] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)R_X(s-t)g(t)dsdt = (f, R_X * g) \\ &= (\widehat{f}, \widehat{R_X * g})/2\pi \quad \text{(Parseval's relation)} \\ &= (\widehat{f}, S_X\widehat{g})/2\pi = \int_{-\infty}^{\infty} \widehat{f}(\omega)(\widehat{g}(\omega))^* S_X(\omega)\frac{d\omega}{2\pi}. \end{aligned}$$

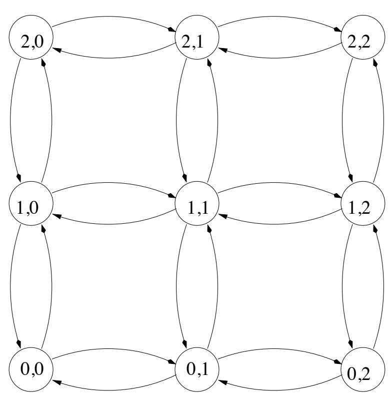
The above two expressions for $E[(f,X)(g,X)]$ are complex conjugates of each other, but since $E[(f,X)(g,X)]$ is real valued, the expressions are equal.
(c)

$$\begin{aligned} \Big| E[(f,X)(g,X)] - (f,g) \Big| &= \left| \int_{-\infty}^{\infty} \widehat{f}(\omega)(\widehat{g}(\omega))^*(S_X(\omega)-1)\frac{d\omega}{2\pi} \right| \\ &= \left| \int_{-\omega_o}^{\omega_o} \widehat{f}(\omega)(\widehat{g}(\omega))^*(S_X(\omega)-1)\frac{d\omega}{2\pi} \right| \\ &\le \int_{-\omega_o}^{\omega_o} |\widehat{f}(\omega)(\widehat{g}(\omega))^*||S_X(\omega)-1|\frac{d\omega}{2\pi} \end{aligned}$$

For $|\omega| < \omega_o$, $|S_X(\omega)-1| = \frac{\omega^2}{\alpha^2+\omega^2} \le \frac{\omega_o^2}{\alpha^2+\omega_o^2} \le 0.01$ if $\alpha^2 \ge 99\omega_o^2$, or $\alpha \ge \sqrt{99}\omega_o \approx 10\omega_o$. Under this condition,

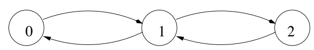
$$\begin{aligned} \Big| E[(f,X)(g,X)] - (f,g) \Big| &\le (0.01)\int_{-\omega_o}^{\omega_o} |\widehat{f}(\omega)(\widehat{g}(\omega))^*|\frac{d\omega}{2\pi} \\ &\le (0.01)\sqrt{\left(\int_{-\omega_o}^{\omega_o} |\widehat{f}(\omega)|^2\frac{d\omega}{2\pi}\right)\left(\int_{-\omega_o}^{\omega_o} |\widehat{g}(\omega)|^2\frac{d\omega}{2\pi}\right)} \qquad \text{(Schwarz inequality)} \\ &= (0.01)||f|| \cdot ||g|| \qquad \text{(Parseval's relation)} \end{aligned}$$

Problem 4 *(12 points)* Consider a continuous-time Markov process $X = (X_t : t \ge 0)$ with the state space $\mathcal{S} = \{0,1,2\} \times \{0,1,2\}$ and the transition rate diagram shown. The transition rates for all arrows in the diagram are equal to one.

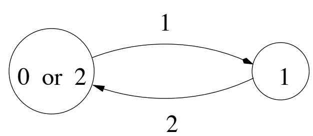


(a) Is the equilibrium distribution for this process equal to the uniform distribution, $(\frac{1}{9}, \ldots, \frac{1}{9})$? If so, justify. If not, find the equilibrium distribution.
(b) Let U_t and V_t denote the two coordinates of X_t for each t fixed. That is, $X_t = (U_t, V_t)$ for each t, where U_t and V_t both take values in $\{0,1,2\}$. Let $\pi(t) = (\pi_{0,0}(t), \pi_{0,1}(t), \ldots, \pi_{2,2}(t))$ denote the distribution of X at time t for $t \ge 0$. Under what conditions on the intital distribution, $\pi(0)$, are the processes $U = (U_t : t \ge 0)$ and $V = (V_t : t \ge 0)$ independent of each other? Explain.
(c) Find $\pi(t)$ for all $t \ge 0$ for the initial distribution $P\{X(0) = (1,1)\} = 1$. (Hint: Due to the particular structure of this problem, it can be solved with little computation.)

(a) Yes. It suffices to check that $\pi Q = 0$, which is equivalent to nine balance equations, one for each state. The balance equation for state (0,0) is $\pi_{0,0}(q_{00,01} + q_{00,10}) = \pi_{0,1}q_{01,00} + \pi_{1,0}q_{10,00}$. Since the rates are all one, this becomes $2\pi_{0,0} = \pi_{0,1} + \pi_{1,0}$, which holds for the uniform distribution. By symmetry, the balance equation for the other three corner states, $(0,2)$, $(2,0)$, and $(2,2)$, also hold. The balance equation for state $(0,1)$ is $3\pi_{0,1} = \pi_{0,0} + \pi_{1,1} + \pi_{0,2}$, which is also true for the uniform distribution. By symmetry, the balance equation for the other three side states, $(1,0)$, $(1,2)$, and $(2,1)$, also hold. Since the balance equations for eight of the states hold, the balance equation for the remaining state, $(1,1)$, must also hold (or it can be checked directly). Thus, the uniform distribution is the equilibrium distribution.
(b) For independence of the processes U and V, it is clearly necessary for U_0 to be independent of V_0. That is, it is necessary for $\pi(0)$ to have a product form: $\pi_{i,j}(0) = \pi_i^U(0)\pi_j^V(0)$ for two probability distributions $\pi^U(0)$ and $\pi^V(0)$ on $\{0,1,2\}$. Some thought shows, at least on an intuitive level, that this condition is also sufficient for U and V to be independent. The reason is that the dynamics of U don't depend on V, and vice versa. For example, for t and h fixed with h small, and $i,j \in \{0,1,2\}$, the conditional distribution of U_{t+h} given $(U_t, V_t) = (i,j)$ is the same, at least up to $o(h)$, for all values of j. Both U and V are Markov processes with the following transition rate diagram:



(To make this argument rigorous it is easier to begin in the reverse direction. If U and V are independent Markov processes on $\{0,1,2\}$ with initial distributions $\pi^U(0)$ and $\pi^V(0)$ respectively, then X defined by $X_t = (U_t, V_t)$ is a Markov process with inital state $\pi_{i,j}(0) = \pi_i^U(0)\pi_j^V(0)$ and the transition rate diagram given in the problem. Since a Markov jump process is uniquely determined by its initial distribution and transition rate diagram, the argument can be reversed.)
(c) For the process U, states 0 and 2 can be grouped into a super state, and the resulting process is still Markov, with the transition rate diagram shown:



By the solution for two-state Markov processes given in the notes, it follows that $\pi_1^U(t) = \frac{1+2e^{-t}}{3}$ and $\pi_0^U(t) + \pi_2^U(t) = \frac{2(1-e^{-t})}{3}$, for all $t \geq 0$. By symmetry, $\pi_0^U(t) = \pi_2^U(t)$, and $\pi^U(t) = \pi^V(t)$. Thus, $\pi^U(t) = \pi^V(t) = (\frac{1-e^{-t}}{3}, \frac{1+2e^{-t}}{3}, \frac{1-e^{-t}}{3})$. Therefore, for intial state $(1,1)$, the one-dimensional distributions of X are given by $\pi_{0,0}(t) = \frac{(1-e^{-t})^2}{9}$, $\pi_{0,1}(t) = \frac{(1-e^{-t})(1+2e^{-t})}{9}$, $\pi_{1,1}(t) = \frac{(1+2e^{-t})^2}{9}$, and the probabilites of other states are determined by symmetry.

Problem 5 *(12 points)* Suppose $Y = \theta s + W$, were s is a known fixed vector in $\mathbb{R}^n$, W is a mean-zero Gaussian random vector in $\mathbb{R}^n$ with a nonsingular covariance matrix K, and θ is a real-valued parameter to be estimated.
(a) Write the pdf of Y, which we write as $f_Y(y|\theta)$ to emphasize that it depends on θ.
(b) Derive the maximum likelihood (ML) estimate, $\widehat{\theta}(y)$, for an observed vector y.
(c) Derive the maximum a posteriori (MAP) estimate, $\widehat{\theta}_{MAP}(y)$, for an observed vector y, taking the prior distribution of θ to be $N(0,1)$: $f_\Theta(\theta) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{\theta^2}{2})$.

(a) Y has the $N(\theta s, K)$ distribution, so its pdf is

$$f_Y(y|\theta) = |K|^{-\frac{1}{2}}(2\pi)^{-\frac{n}{2}} \exp\left(-\frac{(y-\theta s)^T K^{-1}(y-\theta s)}{2}\right).$$

(b) Therefore, $\widehat{\theta}_{ML}(y)$ is the value of θ that minimizes the function $(y-\theta s)^T K^{-1}(y-\theta s)$ with respect to θ. This is a quadratic function of θ, so it is minimized at the unique point the derivative is zero. Calculating,

$$\begin{aligned}\frac{d(y-\theta s)^T K^{-1}(y-\theta s)}{d\theta} &= -s^T K^{-1}(y-\theta s)-(y-\theta s)^T K^{-1}s \\ &= -s^T K^{-1}y+2\theta s^T K s-y^T K^{-1}s \\ &= 2(\theta s^T K^{-1}s-y^T K^{-1}s),\end{aligned}$$

Setting the derivative equal to zero yields $\widehat{\theta}_{ML}(y)=\frac{y^T K^{-1}s}{s^T K^{-1}s}$.

(c) The MAP estimator $\widehat{\theta}_{MAP}(y)$ is the value of θ that maximizes $f_Y(y|\theta)f_\Theta(\theta)$ with respect to θ, or equivalently, taking negative logarithms, minimizes

$$\frac{(y-\theta s)^T K^{-1}(y-\theta s)}{2}+\frac{\theta^2}{2}$$

with respect to θ. This is again quadratic in θ, and by a modification of the calculation above, the derivative is given by $\theta s^T K^{-1}s-y^T K^{-1}s+\theta$. Setting the derivative equal to zero yields

$$\widehat{\theta}_{MAP}(y)=\frac{y^T K^{-1}s}{1+s^T K^{-1}s}.$$

An alternative derivation is to note that since Θ and Y are jointly Gaussian and have mean zero (under the Bayesian assumption) $\widehat{\theta}_{MAP}(y)=\widehat{\theta}_{ML}(y)=\widehat{E}[\Theta|Y=y]=\mathrm{Cov}(\Theta,Y)\mathrm{Cov(Y)}^{-1}y=s^T(ss^T+K)^{-1}y$. The two approaches yield equivalent answers, because $\frac{s^T K^{-1}}{1+s^T K^{-1}s}=s^T(ss^T+K)^{-1}$, as can be verified by multiplying both sides on the right by ss^T+K.

Problem 6 *(12 points)* Suppose $(X_n : n \geq 1)$ is a discrete-time Markov process with state space $\{0,1\}$, initial distribution $\pi(1)=(0.5,0.5)$, and one-step transition probability matrix $P=\begin{pmatrix}1-p & p\\ p & 1-p\end{pmatrix}$, where $0<p<1$. Let $S_n = X_1+\cdots+X_n$.

(a) *(3 pts)* Find $E[S_n]$ for all $n\geq 1$.

(b) *(3 pts)* Does $\lim_{n\to\infty}\frac{S_n}{n}$ exist in the a.s. sense? Justify your answer.

(c) *(6 pts)* Let $\theta>0$. The purpose of this problem is to compute $E[\exp(\theta S_n)]$ for all $n\geq 1$. (This could be used in the Chernoff inequality to provide a bound on large deviation probabilities of the form $P\{\frac{S_n}{n}\geq\alpha\}$.) Let $a_n = E[e^{\theta S_n}I_{\{X_n=0\}}]$ and $b_n=E[e^{\theta S_n}I_{\{X_n=1\}}]$. Note that $a_n+b_n=E[\exp(\theta S_n)]$. Identify a recursive way to compute (a_n,b_n) for all $n\geq 1$. Start by finding the initial condition (i.e. the value of (a_1,b_1).)

(a) The initial probability distribution $\pi(1)$ is the equilbrium distribution, that is, $\pi(1)P=\pi(1)$. Therefore, X_k has probability vector $\pi(1)$ for all $k\geq 1$. Hence,
$E[S_n]=E[X_1+\cdots+X_n]=E[X_1]+\cdots+E[X_n]=\frac{n}{2}$.

(b) Yes. Since X has a finite state space it is positive recurrent, and it is irreducible. Hence, the a.s. limit of the time averages of the X's exists and is equal to the statistical average, namely $E[X_n]=0.5$. (See Section 6.5.)

(c) Since $S_1=X_1$, the initial values are given by (a_1,b_1). Let $n\geq 1$ and use the fact $S_{n+1}=S_n+X_{n+1}$, the law of total probability, and the Markov property, to get

$$\begin{aligned}a_{n+1} &= E[e^{\theta S_{n+1}}I_{\{X_{n+1}=0\}}]=E[e^{\theta S_n}I_{\{X_{n+1}=0\}}]\\ &= E[e^{\theta S_n}I_{\{X_n=0,X_{n+1}=0\}}]+E[e^{\theta S_n}I_{\{X_n=1,X_{n+1}=0\}}]=a_n(1-p)+b_n p\end{aligned}$$

Similarly,

$$\begin{aligned}b_{n+1} &= E[e^{\theta S_{n+1}}I_{\{X_{n+1}=1\}}]=E[e^{\theta S_n}I_{\{X_{n+1}=1\}}]e^\theta\\ &= \left(E[e^{\theta S_n}I_{\{X_n=0,X_{n+1}=1\}}]+E[e^{\theta S_n}I_{\{X_n=1,X_{n+1}=1\}}]\right)e^\theta=a_n p e^\theta+b_n(1-p)e^\theta\end{aligned}$$

Summarizing, $(a_{n+1},b_{n+1})=(a_n,b_n)B$, where $B=\begin{pmatrix}1-p & pe^\theta\\ p & (1-p)e^\theta\end{pmatrix}$. Thus, $(a_n,b_n)=(0.5,(0.5)e^\theta)B^{n-1}$.

(Note: The recursive method is essentially the same as the forward part of the forward-backward algorithm for HMMs. The powers of B, and therefore $E[\exp(\theta S_n)]$, grow like λ_1^n, where λ_1 is the largest magnitude eigenvalue of B.)