Solution to Quiz 1, ECE7252, February 27, 2008

1. Given the joint probability density function, $f(x,y)$:

$$f(x,y) = \frac{1}{\sqrt{(2\pi)^2(1-\rho^2)\sigma_x^2\sigma_y^2}} \exp\left[-\frac{1}{2(1-\rho^2)}\{(\frac{x-\mu_x}{\sigma_x})^2 + 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2\}\right], \quad (1)$$

(a) Because $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\tau^2}} \exp[-\frac{1}{\tau^2}(z-\zeta)^2]dz = 1$, and

$$(\frac{x-\mu_x}{\sigma_x})^2 + 2\rho(\frac{x-\mu_x}{\sigma_x})(\frac{y-\mu_y}{\sigma_y}) + (\frac{y-\mu_y}{\sigma_y})^2 = (1-\rho^2)(\frac{x-\mu_x}{\sigma_x})^2 + \frac{1}{\sigma_y^2}[(y-\mu_y)+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)]^2 \quad (2)$$

the marginal pdf, $f(x)$, which is Gaussian can be obtained as follows:

$$f(x) = \int_{-\infty}^{\infty} f(x,y)dy = \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp[-\frac{1}{\sigma_x^2}(x-\mu_x)^2], \quad (3)$$

Similarly, the marginal pdf of y is

$$f(y) = \int_{-\infty}^{\infty} f(x,y)dx = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp[-\frac{1}{\sigma_y^2}(y-\mu_y)^2]; \quad (4)$$

(b) The conditional pdf, $f(x|y)$, is also Gaussian with $f(x|y) = \frac{f(x,y)}{f(y)}$ by Bayes rule:

$$f(x|y) = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_x^2}} \exp\{-\frac{1}{(1-\rho^2)\sigma_x^2}[(x-\mu_x)^2 - 2\rho\frac{\sigma_x}{\sigma_y}(x-\mu_x)(y-\mu_y) + \frac{\sigma_x^2}{\sigma_y^2}(y-\mu_y)^2(1-(1-\rho^2))]\}, \quad (5)$$

or

$$f(x|y) = \frac{1}{\sqrt{2\pi(1-\rho^2)\sigma_x^2}} \exp\{-\frac{1}{(1-\rho^2)\sigma_x^2}[(x-\mu_x) - \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y)]^2\}; \quad (6)$$

(c) Clearly the conditional mean, $\mu_{x|y}$, and variance, $\sigma_{x|y}^2$, are:

$$\mu_{x|y} = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(y-\mu_y), \quad (7)$$

$$\sigma_{x|y} = \sigma_x\sqrt{1-\rho^2}; \quad (8)$$

(d) since $|\rho| < 1$, $\sigma_{x|y} \leq \sigma_x$, i.e. knowing y reduces the uncertainty of estimating x.

2. (a) Given an iid set of samples, $\mathbf{X} = \{x_i, i = 1, ..., n\}$, the log likelihood function can be expressed as: $\mathbf{L}(\mathbf{X}|\mu) = \log f(\mathbf{X}|\mu)$

$$\mathbf{L}(\mathbf{X}|\mu) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2, \tag{9}$$

similarly the log prior density is expressed as:

$$\log f(\mu) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\tau^2) - \frac{1}{2\tau^2}(\mu - \nu)^2. \tag{10}$$

The posterior density of μ given $\mathbf{X}$ is $f(\mu|\mathbf{X}) = \frac{f(\mathbf{X}|\mu)f(\mu)}{f(\mathbf{X})}$.

(b) Now the log posterior density is just $\log[f(\mu|\mathbf{X})] \propto \mathbf{L}(\mathbf{X}|\mu) + \log f(\mu)$. Taking the derivative over μ and set it to zero, we get

$$\frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) + \frac{1}{\tau^2}(\mu - \nu) = 0, \tag{11}$$

or with $w_1 = \frac{n}{\sigma^2}$, and $w_2 = \frac{1}{\tau^2}$

$$\hat{\mu}_{MAP} = \frac{w_1\bar{x} + w_2\nu}{w_1 + w_2}. \tag{12}$$

(c) When $n = 0$, it is clear that $\hat{\mu}_{MAP} = \nu$, i.e the prior mean. The same can be concluded if $\tau^2 << \frac{\sigma^2}{n}$, or we are more certain about μ a priori. On the other hand when $n = \infty$, $\hat{\mu}_{MAP} = \bar{x}$, i.e. the sample mean, or $\hat{\mu}_{MAP}$ is asymptotically equivalent to $\hat{\mu}_{ML}$.

3. Consider a linear regression model, $y = X_1\beta_1 + X_2\beta_2 + \epsilon$:

(a) The least square estimate,

$$\hat{b}_1 = (X_1^TX_1)^{-1}X_1^Tz = (X_1^TX_1)^{-1}X_1^T[y - (X_2^TX_2)^{-1}X_2^Ty]. \tag{13}$$

Since $E[y] = X_1\beta_1 + X_2\beta_2$, we have the expectation of $\hat{b}_1$ as:

$$E[\hat{b}_1] = (X_1^TX_1)^{-1}X_1^T(X_1\beta_1+X_2\beta_2)-(X_1^TX_1)^{-1}X_1^TX_2(X_2^TX_2)^{-1}X_2^T(X_1\beta_1+X_2\beta_2) \tag{14}$$

or

$$E[\hat{b}_1] = \beta_1 - (X_1^TX_1)^{-1}X_1^TX_2(X_2^TX_2)^{-1}X_2^TX_1\beta_1. \tag{15}$$

(b) Clearly $E[\hat{b}_1] = \beta_1$ if and only if $X_1^TX_2 = 0$, i.e X_1 and X_2 are decoupled so that the regression of y over X_1 and X_2 can be performed separately to produce unbiased estimates.

4. The third order LPC filter has the z-transform of the form:

$$A(z) = 1 - 0.875z^{-1} + 0.75z^{-2} - 0.25z^{-3} = 1 - \sum_{i=1}^{3} \alpha_i^{(3)} z^{-i}. \quad (16)$$

we can the use the backward iteration of the Durbin (Levinson) recursion to solve for the three PARCORs, k_3, k_2, and k_1, by using the following equations:

$$k_i = \alpha_i^{(i)}, i = 3, 2, 1, \quad (17)$$

and

$$\alpha_j^{(i-1)} = \frac{\alpha_j^{(i)} + \alpha_i^{(i)} \alpha_{i-j}^{(i)}}{(1 - k_i^2)}, 1 \leq j \leq i - 1. \quad (18)$$

Starting with $i = 3$, $k_3 = \alpha_3^{(3)} = 0.25$; so we can solve for

$$\alpha_1^{(2)} = \frac{\alpha_1^{(3)} + \alpha_3^{(3)} \alpha_2^{(3)}}{1 - k_3^2} = 0.733, \quad (19)$$

$$\alpha_2^{(2)} = \frac{\alpha_2^{(3)} + \alpha_3^{(3)} \alpha_1^{(3)}}{1 - k_3^2} = -0.567, \quad (20)$$

now for $i = 2$, we then have $k_2 = \alpha_2^{(2)} = -0.567$ and

$$\alpha_1^{(1)} = \frac{\alpha_1^{(2)} + \alpha_2^{(2)} \alpha_1^{(2)}}{1 - k_2^2} = 0.468, \quad (21)$$

so we finally continue with $i = 1$, and arrive at $k_1 = \alpha_1^{(1)} = 0.468$.

(b) since all PARCORs are less than one in magnitude, the all-pole filter, $H(z)$, is stable.

5. (a) To maximize the ratio take the derivative and set it to zero, we have:

$$\frac{\partial J_F(w)}{\partial w} = \frac{1}{w^T S_W w} [S_B w - J_F(w) S_W w] = 0. \quad (22)$$

Therefore the solution w^* satisfies $S_W^{-1} S_B w^* = J_F(w^*) w^*$, a generalized eigenvalue problem with w^* being an eigenvector of S_B, and $J_F(w^*)$ being the corresponding eigenvalue;

(b) With constrained optimization using a Lagrange formulation $J_{LL}(a) = a^T S_B a + \lambda(a^T S_B a - 1)$, we maximize $J_{LL}(a)$ with respect to a and λ by taking the derivatives and setting them to zero, we get

$$S_B a + \lambda S_W a = 0, \quad (23)$$

by multiplying a^T on both sides above we have $\lambda = -\frac{a^T S_W a}{a^T S_W a} = -J_L(a)$. The above equation also gives a generalized eigenvalue problem: $S_W^{-1} S_B a^* = J_L(a^*) a^*$, so $a^* = \frac{w^*}{||w^*||}$.