DEPARTAMENT DE TEORIA DEL SENYAL I COMUNICACIONS	**Procesado de Señal** Fecha del examen: 10 de Junio 2008 Fecha de notas provisionales: 27 de Junio 2008 Periodo de alegaciones: 30 de Junio 2008 Fecha para notas revisadas: 2 de Julio 2008

Profesores: Albert Oliveras, Gregori Vázquez, Josep Vidal

Información adicional:

- Duración de la prueba: 3 h
- No se pueden usar libros, apuntes, calculadoras, teléfonos móviles, PDAs o dispositivos MP3
- Use hojas distintas para cada problema
- Justifique los resultados

Ejercicio 1

La estimación $\hat{\theta}_u$ de un parámetro escalar θ a partir de un conjunto de datos $\{x_i\}_{1\le i\le N}$ es no sesgada y de varianza $\mathrm{var}(\hat{\theta}_u)$ mínima (por tanto es el estimador denominado *MVU* o estimador "*Minimum-Variance and Unbiased*"). Así pues, el *error cuadrático medio* (*MSE*) es tal que:

$$MSE\left(\hat{\theta}_u\right)=E\left[\left|\theta-\hat{\theta}_u\right|^2\right]=\left|\theta-E\left[\hat{\theta}_u\right]\right|^2+\mathrm{var}\left(\hat{\theta}_u\right)=\mathrm{var}\left(\hat{\theta}_u\right)$$

En este ejercicio, comprobaremos que la solución anterior no es la que proporciona un *MSE* mínimo y que *existen estimaciones sesgadas que mejoran la varianza y el MSE del estimador*.

Considere que el estimador sesgado es de la forma siguiente:

$$\hat{\theta}_b=(1+m)\hat{\theta}_u$$

donde 'm' es una constante:

a. Obtenga el valor del sesgo cuadrático $bias^2\left(\hat{\theta}_b\right)=\left|\theta-E\left[\hat{\theta}_b\right]\right|^2$.

b. Obtenga el valor de la varianza $\mathrm{var}(\hat{\theta}_b)$ del nuevo estimador.

c. A partir de (a.) y (b.), indique el valor de 'm' que hace mínimo el $MSE(\hat{\theta}_b)$ en función del cociente $\rho=\theta^2/\mathrm{var}(\hat{\theta}_u)$. Compruebe que $-1\le m\le 0$.

d. Dibuje cualitativamente los términos $bias^2(\hat{\theta}_b)$, $\mathrm{var}(\hat{\theta}_b)$ y $MSE(\hat{\theta}_b)$ en función de la constante 'm' y justifique gráficamente la existencia de dicho mínimo.

Un ejemplo en el que el cociente $\rho=\theta^2/\mathrm{var}(\hat{\theta}_u)$ es constante, es el siguiente. Consideramos la siguiente función densidad de probabilidad exponencial:

$$f_x(x)=\begin{cases}(1/\theta)\exp(-x/\theta) & x\ge 0\\ 0 & x<0\end{cases}$$

La estimación *MVU* de θ y la varianza asociada para este caso vienen dadas por:

$$\hat{\theta}_u=(1/N)\sum_{i=1}^{N}x_i$$
$$\mathrm{var}(\hat{\theta}_u)=\theta^2/N$$

e. A partir del resultado obtenido en (c.), obtenga la expresión del estimador sesgado $\hat{\theta}_b$ de mínimo MSE para la distribución exponencial anterior y demuestre que $MSE(\hat{\theta}_b) < MSE(\hat{\theta}_u)$.

Lamentablemente, no siempre el cociente $\rho = \theta^2 / \mathrm{var}(\hat{\theta}_u)$ es constante, de modo que la constante 'm' pasaría a depender del parámetro θ y no sería posible aplicar la técnica anterior para reducir el *MSE*. Consideramos ahora el caso en que $\mathrm{var}(\hat{\theta}_u) = V$ es constante y nos planteamos hacer máxima la diferencia entre el $MSE(\hat{\theta}_b)$ y $MSE(\hat{\theta}_u)$ en un intervalo del parámetro a estimar, es decir, en el rango $|\theta| \le \theta_o$.

f. Obtenga el margen de valores de $(1+m)$ que verifican $MSE(\hat{\theta}_u) - MSE(\hat{\theta}_b) > 0$ en el rango $|\theta| \le \theta_o$.
g. Indique la solución sesgada $\hat{\theta}_b$ en función de la $\hat{\theta}_u$ que maximiza la diferencia $MSE(\hat{\theta}_u) - MSE(\hat{\theta}_b) > 0$ en el rango $|\theta| \le \theta_o$.
h. ¿Qué ocurre con la solución en (g.) si $\theta_o \to \infty$?

Ejercicio 2

En este problema definiremos un filtro predictor con componentes *forward* y *backward*: supondremos que disponemos de las muestras $\mathbf{x}_1(n) = \begin{bmatrix} x(n+1) \\ x(n+2) \end{bmatrix}$, $\mathbf{x}_2(n) = \begin{bmatrix} x(n-1) \\ x(n-2) \end{bmatrix}$ y queremos estimar $x(n)$ a partir de una combinación lineal de ellas:

$$\hat{x}(n) = \mathbf{h}_1^H \mathbf{x}_1 + \mathbf{h}_2^H \mathbf{x}_2$$

A partir de la minimización del error cuadrático medio:

$$\{\mathbf{h}_1, \mathbf{h}_2\} = \arg \min_{\mathbf{h}_1, \mathbf{h}_2} E\left\{ |x(n) - \hat{x}(n)|^2 \right\}$$

Se pide:

1) Encuentre las expresiones para los coeficientes del filtro, en función de las matrices de correlación y los vectores de correlación adecuados.
2) Demuestre que $\mathbf{h}_1 = \mathbf{h}_2^*$ a partir de las expresiones anteriores.
3) Determinar la potencia mínima del error.
4) ¿Cuál sería la mejor estimación de $x(n)$ si el proceso fuera blanco?
5) Escriba las expresiones de una solución adaptativa basada en el gradiente estocástico (LMS) para un vector **h** que contenga $\mathbf{h}_1$ y $\mathbf{h}_2$:

$$e(n) = x(n) - \hat{x}(n) = x(n) - \begin{bmatrix} \mathbf{h}_1^H & \mathbf{h}_2^H \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = x(n) - \mathbf{h}^H \mathbf{x}(n)$$

6) Razone cuales son las cotas del paso de adaptación que garantizan la convergencia del algoritmo.

Ejercicio 3

Va a evaluarse la ganancia obtenida en compresión de señales correladas gracias al uso de transformadas unitarias, respecto a la asignación de bits muestra a muestra (esquema PCM). Para el caso de codificación mediante métodos transformados, supondremos que se asignarán bits a cada uno de los coeficientes obtenidos mediante transformación de bloques de señal **x**. El vector de coeficientes transformados se calcula como $\mathbf{y} = \mathbf{A}^H \mathbf{x}$, donde

$$\mathbf{A}^H = \begin{bmatrix} \mathbf{a}_1^H \\ \mathbf{a}_2^H \\ \vdots \\ \mathbf{a}_N^H \end{bmatrix}$$

1. ¿Cuál deber ser la relación entre las filas de $\mathbf{A}^H$ para que la transformación sea unitaria?
2. Si el coeficiente transformado $y(k)$ se cuantifica con n_k bits, la potencia del error de cuantificación en ese coeficiente viene dada por

$$E\left\{\left|y_k - \hat{y}_k\right|^2\right\} = \frac{\sigma_{y_k}^2}{2^{2n_k}}, \text{ donde } \sigma_{y_k}^2 = E\left\{\left|y_k\right|^2\right\} \tag{0.1}$$

¿Cuál es la potencia media del error en la señal decodificada a partir de los coeficientes cuantificados, suponiendo que la transformación es unitaria?

Fijada una potencia del error en la señal decodificada, se desea minimizar la tasa promedio de bits por muestra de señal, que viene dada por $r = \frac{1}{N}\sum_{k=1}^{N} n_k$.

3. Determine cual es el número de bits a asignar por coeficiente transformado, minimizando r para un valor fijo D:

$$D = \frac{1}{N}\sum_{k=1}^{N} E\left\{\left|y_k - \hat{y}_k\right|^2\right\}$$

de la potencia media del error calculada en el apartado 2. Use el método de los multiplicadores de Lagrange. Al realizar la derivada, asuma que n_k puede tomar valores no enteros.

4. Asuma que el valor de D es suficientemente pequeño como para $n_k \geq 0 \;\; \forall k$. ¿Cuál es el valor óptimo de la tasa promedio de bits por muestra r, en función de la potencia media del error D? Si la transformada utilizada es la KLT, demuestre que $r_{KLT} = \frac{1}{2N}\log_2 \frac{\left|\mathbf{R}_x\right|}{D^N}$, donde el determinante de la matriz de correlación $\left|\mathbf{R}_x\right|$ puede indicarse como el producto de sus autovalores.

-------------***NOTA: LA RESPUESTA A LOS APARTADOS (5.), (6.) y (7.) ES OPCIONAL***---------------

Para comparar las prestaciones de este esquema con una cuantificación muestra a muestra (PCM), se asumirá que el proceso $x(n)$ es AR(1), con función de correlación $r_x(k) = \sigma_x^2 \rho^{|k|}$.

5. Determine el valor de r_{KLT} en función de D y de ρ , para la codificación de x(n) en bloques de N=2.
6. Usando la expresión (0.1), y asumiendo que el error es estacionario ¿cual es la potencia media del error de cuantificación cuando se cuantifica muestra a muestra de $x(n)$? Determine el valor de r_{PCM} necesario en este caso, en función de D.
7. ¿Qué esquema de codificación es más efectivo? Compare r_{KLT} y r_{PCM} en función de ρ , para ρ = 0 y $\rho \to 1$

NOTA: $\frac{d\left(2^x\right)}{dx} = 2^x \ln 2$

RESOLUCIÓN DE LOS EJERCICIOS

Ejercicio 1

(a.) *Vemos que* $\hat{\theta}_b = (1+m)\hat{\theta}_u$ *y* $E\left[\hat{\theta}_b\right] = (1+m)E\left[\hat{\theta}_u\right] = (1+m)\theta$. *El sesgo cuadrático es de la forma siguiente:*

$$bias^2\left(\hat{\theta}_b\right) = \left|\theta - E\left[\hat{\theta}_b\right]\right|^2 = \left|\theta - (1+m)\theta\right|^2 = m^2\theta^2$$

(b.) *Para el término de varianza:*

$$\operatorname{var}\left(\hat{\theta}_b\right) = E\left[\left|\hat{\theta}_b - E\left[\hat{\theta}_b\right]\right|^2\right] = E\left[\left|(1+m)\hat{\theta}_u - (1+m)\theta\right|^2\right] = (1+m)^2 E\left[\left|\hat{\theta}_u - E\left[\hat{\theta}_u\right]\right|^2\right] =$$

$$= (1+m)^2 \operatorname{var}\left(\hat{\theta}_b\right)$$

(c.) *El nuevo término de MSE es la suma de la varianza y el sesgo cuadrático:*

$$MSE\left(\hat{\theta}_b\right) = E\left[\left|\theta - \hat{\theta}_b\right|^2\right] = \left|\theta - E\left[\hat{\theta}_b\right]\right|^2 + \operatorname{var}\left(\hat{\theta}_u\right) = m^2\theta^2 + (1+m)^2 \operatorname{var}\left(\hat{\theta}_u\right)$$

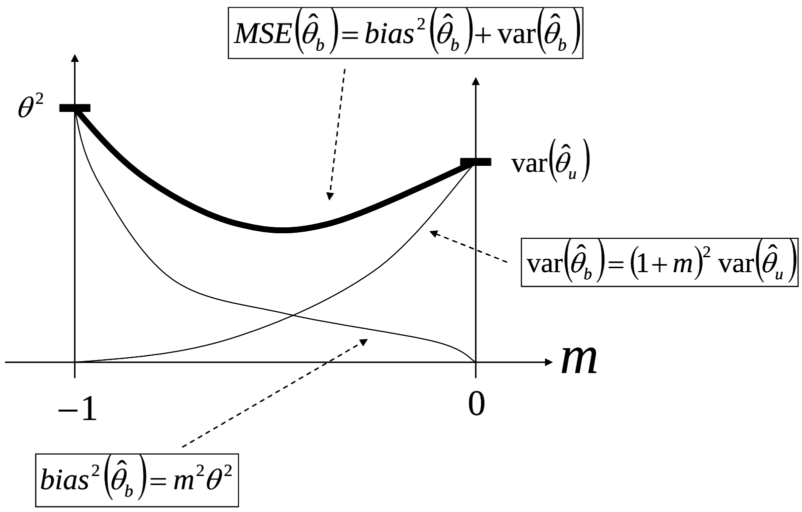
El mínimo se obtiene de la forma habitual:

$$\frac{\partial}{\partial m} MSE\left(\hat{\theta}_b\right) = \frac{\partial}{\partial m}\left(m^2\theta^2 + (1+m)^2 \operatorname{var}\left(\hat{\theta}_u\right)\right) = 2m\theta^2 + 2(1+m)\operatorname{var}\left(\hat{\theta}_u\right) = 0$$

El valor óptimo de 'm' se alcanza para:

$$m = -\frac{\operatorname{var}\left(\hat{\theta}_u\right)}{\theta^2 + \operatorname{var}\left(\hat{\theta}_u\right)} = -\frac{1}{1+\rho} \quad \textit{como } \rho = \frac{\theta^2}{\operatorname{var}\left(\hat{\theta}_u\right)} \geq 0 \Rightarrow -1 \leq m \leq 0$$

(d.)



(e.) *Tenemos que:*

$$\hat{\theta}_b = (1+m)\hat{\theta}_u = \left(1 - \frac{1}{1+\rho}\right)\hat{\theta}_u = \frac{\rho}{1+\rho}\hat{\theta}_u = \frac{\theta^2}{\theta^2 + \operatorname{var}\left(\hat{\theta}_u\right)}\hat{\theta}_u$$

$$MSE\left(\hat{\theta}_b\right) = m^2\theta^2 + (1+m)^2 \operatorname{var}\left(\hat{\theta}_u\right) = \frac{\rho}{1+\rho}\operatorname{var}\left(\hat{\theta}_u\right)$$

Para la distribución exponencial anterior: $\rho = N$ *y:*

$$MSE\left(\hat{\theta}_b\right) = \frac{\rho}{1+\rho}\operatorname{var}\left(\hat{\theta}_u\right) = \frac{N}{1+N}\operatorname{var}\left(\hat{\theta}_u\right) < \operatorname{var}\left(\hat{\theta}_u\right) = MSE\left(\hat{\theta}_u\right)$$

Especialmente de interés para tamaños de observación N *pequeños.*

(f.) La mejora en el MSE queda como la diferencia:

$$MSE(\hat{\theta}_b) - MSE(\hat{\theta}_u) = m^2\theta^2 + (1+m)^2 \operatorname{var}(\hat{\theta}_u) - \operatorname{var}(\hat{\theta}_u) =$$
$$= m^2\theta^2 + (1+m)^2 V - V < 0$$

Un punto importante es que la expresión anterior <u>siempre es máxima</u> para $|\theta| = \theta_o$ *en el intervalo* $|\theta| \le \theta_o$.

Luego:

$$m\left(2V + m(\theta^2 + V)\right) < 0 \quad \textit{como } m < 0 \Rightarrow 2V + m(\theta^2 + V) > 0$$

Por tanto:

$$m > -\frac{2V}{V+\theta^2} \Rightarrow (1+m) > \frac{\theta^2 - V}{\theta^2 + V}$$

(g.) La mejora en el MSE siempre es máxima para $|\theta| = \theta_o$ *en el intervalo* $|\theta| \le \theta_o$.

$$MSE(\hat{\theta}_b) - MSE(\hat{\theta}_u) = m^2\theta_o^2 + (1+m)^2 V - V < 0$$

Luego:

$$\frac{\partial}{\partial m}\left(MSE(\hat{\theta}_b) - MSE(\hat{\theta}_u)\right) = \frac{\partial}{\partial m}\left(m^2\theta_o^2 + (1+m)^2 V - V\right) = 0$$

Finalmente:

$$m = -\frac{V}{V+\theta_o^2} \Rightarrow \hat{\theta}_b = \frac{\theta_o^2}{V+\theta_o^2}\hat{\theta}_u$$

(h.) Vemos que para un rango de parámetros $|\theta| \le \theta_o$ *arbitrariamente grande* $\theta_o \to \infty$, *tenemos de la solución en (g.):*

$$\lim_{\theta_o \to \infty} \hat{\theta}_b = \lim_{\theta_o \to \infty} \frac{\theta_o^2}{V+\theta_o^2}\hat{\theta}_u \to \hat{\theta}_u$$

y no es posible aportar una solución al problema.

Ejercicio 2

1) $e(n) = x(n) - \hat{x}(n) = x(n) - \mathbf{h}_1^H\mathbf{x}_1 - \mathbf{h}_2^H\mathbf{x}_2$

$$J = E\left\{|e(n)|^2\right\} = E\left\{|x(n) - \hat{x}(n)|^2\right\} =$$
$$= r_x(0) - \mathbf{h}_1^H\mathbf{r}_{x_1} - \mathbf{r}_{x_1}^H\mathbf{h}_1 - \mathbf{h}_2^H\mathbf{r}_{x_2} - \mathbf{r}_{x_2}^H\mathbf{h}_2 - \mathbf{h}_1^H\mathbf{R}_{x_1x_2}\mathbf{h}_2 - \mathbf{h}_2^H\mathbf{R}_{x_2x_1}\mathbf{h}_1 + \mathbf{h}_1^H\mathbf{R}_{x_1}\mathbf{h}_1 + \mathbf{h}_2^H\mathbf{R}_{x_2}\mathbf{h}_2$$

Derivando e igualando a cero:

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{h}_1^*} E\left\{|e(n)|^2\right\} &= -E\left\{e(n)^*\mathbf{x}_1\right\} = -\mathbf{r}_{x_1} - \mathbf{R}_{x_1x_2}\mathbf{h}_2 + \mathbf{R}_{x_1}\mathbf{h}_1 = \mathbf{0} \\
\frac{\partial}{\partial \mathbf{h}_2^*} E\left\{|e(n)|^2\right\} &= -E\left\{e(n)^*\mathbf{x}_2\right\} = -\mathbf{r}_{x_2} - \mathbf{R}_{x_2x_1}\mathbf{h}_1 + \mathbf{R}_{x_2}\mathbf{h}_2 = \mathbf{0}
\end{aligned} \tag{0.2}$$

Resolvemos la primera y sustituimos en la segunda para obtener:

$$\mathbf{h}_1 = \left(\mathbf{R}_{x_1} - \mathbf{R}_{x_1x_2}\mathbf{R}_{x_2}^{-1}\mathbf{R}_{x_2x_1}\right)^{-1}\left(\mathbf{r}_{x_1} - \mathbf{R}_{x_1x_2}\mathbf{R}_{x_2}^{-1}\mathbf{r}_{x_2}\right)$$
$$\mathbf{h}_2 = \left(\mathbf{R}_{x_2} - \mathbf{R}_{x_2x_1}\mathbf{R}_{x_1}^{-1}\mathbf{R}_{x_1x_2}\right)^{-1}\left(\mathbf{r}_{x_2} - \mathbf{R}_{x_2x_1}\mathbf{R}_{x_1}^{-1}\mathbf{r}_{x_1}\right)$$

2) Para demostrar que las soluciones son conjugadas no hay más que desarrollar las matrices y vectores y comprobar que

$$\mathbf{R}_{x_1} = \mathbf{R}_{x_2}^* \qquad \mathbf{R}_{x_1x_2} = \mathbf{R}_{x_2x_1}^* \qquad \mathbf{r}_{x_1} = \mathbf{r}_{x_2}^*$$

3) A partir de las condiciones (0.2), la potencia mínima viene dada por
$J_{\min} = E\left\{e(n)^* x(n)\right\} = E\left\{\left(x(n) - \mathbf{h}_1^H \mathbf{x}_1 - \mathbf{h}_2^H \mathbf{x}_2\right)^* x(n)\right\} = r_x(0) - \mathbf{r}_{x_1}^H \mathbf{h}_1 - \mathbf{r}_{x_2}^H \mathbf{h}_2$

4) Si el proceso es blanco los coeficientes de los predictores son cero y la mejor estimación en el sentido del menor error cuadrático medio es $\hat{x}(n) = 0$.

5)
$$e(n) = x(n) - \hat{x}(n) = x(n) - \begin{bmatrix} \mathbf{h}_1^H & \mathbf{h}_2^H \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = x(n) - \mathbf{h}^H \mathbf{x}$$
$$\mathbf{h}(n) = \mathbf{h}(n-1) + \mu e^*(n)\mathbf{x}(n)$$

6) La cota en el valor de m que garantiza la convergencia viene dada por el máximo autovalor asociado a la matriz de correlación del vector **x**:

$$0 < \mu < \frac{2}{\lambda_{\max}\left(\mathbf{R}_{\mathbf{x}}\right)} \qquad \text{donde } \mathbf{R}_{\mathbf{x}} = \begin{bmatrix} \mathbf{R}_{\mathbf{x}_1} & \mathbf{R}_{\mathbf{x}_1\mathbf{x}_2} \\ \mathbf{R}_{\mathbf{x}_2\mathbf{x}_1} & \mathbf{R}_{\mathbf{x}_2} \end{bmatrix}$$

Ejercicio 3

1) Las filas deben cumplir $\mathbf{a}_i^H \mathbf{a}_j = \delta_{i-j}$
2) Si la transformación es unitaria la potencia del error cometido en los coeficientes transformados es la misma que la cometida sobre la señal:
$$J = \frac{1}{N} E\left\{(\mathbf{x} - \hat{\mathbf{x}})^H (\mathbf{x} - \hat{\mathbf{x}})\right\} = \frac{1}{N} \sum_{k=1}^{N} \frac{\sigma_{y_k}^2}{2^{2n_k}}$$
3) El multiplicador de Lagrange que hemos de minimizar vendrá dado por:
$$F(n_1, \ldots, n_N) = \frac{1}{N} \sum_{k=1}^{N} n_k + \lambda \left(D - \frac{1}{N} \sum_{k=1}^{N} \frac{\sigma_{y_k}^2}{2^{2n_k}} \right)$$
Derivando la expresión obtenemos un valor para el multiplicador de $\lambda = \frac{1}{D \ln 4}$ y para el número de bits por coeficiente: $n_k = \frac{1}{2} \log_2 \frac{\sigma_{y_k}^2}{D}$
4) Sumando el número de bits obtenido $r = \frac{1}{2N} \sum_{k=1}^{N} \log_2 \frac{\sigma_{y_k}^2}{D}$. Si la transformada utilizada es la KL, la potencia de cada coeficiente transformado es el autovalor de **x**, de forma que:
$$r = \frac{1}{2N} \sum_{k=1}^{N} \log_2 \frac{\sigma_{y_k}^2}{D} = \frac{1}{2N} \log_2 \prod_{k=1}^{N} \frac{\sigma_{y_k}^2}{D} = \frac{1}{2N} \log_2 \frac{\prod_{k=1}^{N} \lambda_k}{D^N} = \frac{1}{2N} \log_2 \frac{|\mathbf{R}_x|}{D^N}$$
5) La matriz de correlación para este proceso es $\mathbf{R}_x = \sigma_x^2 \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$ y su determinante es $|\mathbf{R}_x| = \sigma_x^4 \left(1 - \rho^2\right)$. De aquí se puede sustituir en la expresión anterior para determinar *r*.
6) El error de cuantificación por muestra en el caso PCM viene dado por la expresión de la potencia del error en la ecuación (0.1): $E\left\{|\varepsilon|^2\right\} = \frac{\sigma_x^2}{2^{2n}} = D$. El número de bits necesario por muestra será: $n = \frac{1}{2} \log_2 \frac{\sigma_x^2}{D}$

7) Comparando ambas expresiones obtenemos $r_{PCM} - r_{KLT} = \frac{1}{4}\log\left(1-\rho^2\right)^{-1/2}$. Si ρ tiende a cero, los dos esquemas son iguales. A medida que ρ se acerca a 1, la reducción en el número de bits usando la KLT se hace más importante.