B3 - Apuntes de filtrado lineal

Nombre del curso: Teoría Moderna de la Detección y Estimación

Autores: Miguel Lázaro Gredilla

Universidad Carlos III de Madrid

Contents

5

Filtrado Lineal

5.1 Introduccion

Un problema común en estimación es el de querer determinar los coeficientes de un filtro lineal con M coeficientes a partir de la sola observación de las entradas y salidas de este. A esta tarea, así como a otras relacionadas, se la conoce con el nombre genérico de "filtrado lineal". En este bloque mostraremos como las técnicas descritas en el bloque B1 pueden ser usadas para diseñar estimadores ML, MAP, MAD y MMSE de los coeficientes de dicho filtro, así como de futuras salidas del filtro si se conocen las correspondientes entradas.

5.2 El problema de filtrado

Suponga que se utiliza un filtro desconocido $\mathbf{s} = [s_1, s_2, \ldots, s_M]^\top$ para filtrar una señal $u[n]$. Al resultado se le suma cierto ruido gaussiano $\varepsilon[n]$ iid de media nula y varianza σ_ε^2, dando lugar a una observación $x[n]$. Es decir,

$$x[n] = u[n]s_1 + u[n-1]s_2 + \ldots + u[n-M+1]s_M + \varepsilon[n],$$

o, lo que es lo mismo, compactando $\mathbf{u}[n] = [u[n], u[n-1], \ldots, u[n-M+1]]$, podemos decir que

$$x[n] = \mathbf{u}[n]\mathbf{s} + \varepsilon[n].$$

El problema de filtrado consiste entonces en estimar los coeficientes s de un filtro a partir de un conjunto de entradas y salidas observadas, así como estimar la salida x_* correspondiente a una nueva entrada $\mathbf{u}_*$.

Si disponemos de las señales $u[n]$ y $x[n]$ en el intervalo $1 \leq n \leq N$, dispondremos de un total de $N - M + 1$ parejas entrada-salida, $\{\mathbf{u}[n], x[n]\}_{n=M}^N$. Estos datos observables pueden expresarse de forma más compacta como

$$\mathbf{U} = \begin{bmatrix} \mathbf{u}[M] \\ \mathbf{u}[M+1] \\ \vdots \\ \mathbf{u}[N] \end{bmatrix} = \begin{bmatrix} u[M] & u[M-1] & \ldots & u[1] \\ u[M+1] & u[M] & \ldots & u[2] \\ \vdots & \vdots & \ddots & \vdots \\ u[N] & u[N-1] & \ldots & u[N-M+1] \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} x[M] \\ x[M+1] \\ \vdots \\ x[N] \end{bmatrix}.$$

5.3 Solución ML

El propio planteamiento del problema nos proporciona la verosimilitud de los coeficientes del filtro s dada la observación n-ésima:

$$p(x[n]|\mathbf{s}) = \mathcal{N}(x[n]|\mathbf{u}[n]\mathbf{s}, \sigma_\varepsilon^2)$$

Dado un conjunto de observaciones, simplemente tomamos el producto de las anteriores verosimilitudes, ya que los términos de ruido son independientes

$$p(\mathbf{x}|\mathbf{s}) = \prod_{n=M}^{N} \mathcal{N}(x[n]|\mathbf{u}[n]\mathbf{s}, \sigma_\varepsilon^2) = \mathcal{N}(\mathbf{x}|\mathbf{U}\mathbf{s}, \sigma_\varepsilon^2\mathbf{I}).$$

El valor de s que maximiza $p(\mathbf{x}|\mathbf{s})$ es

$$\hat{\mathbf{s}}_{\text{ML}} = (\mathbf{U}^\top\mathbf{U})^{-1}\mathbf{U}^\top\mathbf{x}$$

5.4 Filtro de Wiener

El filtro de Wiener $\mathbf{s}_{\text{Wiener}}$ es el filtro que minimiza el error cuadrático entre la salida observada $x[n]$ y la salida producida al ser utilizado para filtrar la entrada $u[n]$. Se puede demostrar fácilmente que dicho filtro puede expresarse como:

$$\mathbf{s}_{\text{Wiener}} = R_{uu}^{-1} r_{ux},$$

donde R_{uu} es la matriz de correlación de la señal de entrada $u[n]$ y r_{ux} es el vector de correlación cruzada entre $u[n]$ y $x[n]$. Desafortunadamente, estas dos cantidades son desconocidas en general, por lo que en la mayoría de las ocasiones, el filtro de Wiener no puede calcularse. Sin embargo, es frecuente usar la expresión anterior usando estimaciones muestrales para la matriz de correlación $\hat{R}_{uu} = \frac{1}{N-M+1}\mathbf{U}^\top\mathbf{U}$ y el vector de correlación cruzada $\hat{r}_{ux} = \frac{1}{N-M+1}\mathbf{U}^\top\mathbf{x}$. El resultado es una aproximación al filtro de Wiener $\hat{\mathbf{s}}_{\text{Wiener}} = \hat{R}_{uu}^{-1}\hat{r}_{ux}$ que minimiza el error cuadrático muestral (a menudo llamada "estimación least-squares") y que coincide con la solución ML, es decir $\hat{\mathbf{s}}_{\text{Wiener}} = \hat{\mathbf{s}}_{\text{ML}}$.

A medida que el número de muestras disponibles para la estimación de los estadísticos R_{uu} y r_{ux} aumenta, dichas estimaciones se vuelven más precisas, de manera que $\hat{\mathbf{s}}_{\text{Wiener}}$ y por tanto $\hat{\mathbf{s}}_{\text{ML}}$ coinciden asintóticamente con el verdadero filtro de Wiener.

5.5 Solución Bayesiana

Para obtener un estimador Bayesiano de s necesitamos conocer su probabilidad a priori $p(\mathbf{s})$. Aunque en general ésta es desconocida, es sensato utilizar

$$p(\mathbf{s}) = \mathcal{N}(\mathbf{s}|\mathbf{0}, \sigma_s^2\mathbf{I}),$$

ya que considera aceptable cualquier conjunto de coeficientes reales, y supone que estos tienen media nula y una dispersión fijada por σ_s^2. También es posible fijar $\sigma_s^2 \to \infty$ para conseguir una distribución uniforme. En cualquier caso, el uso de esta distribución a priori permite obtener de manera analítica la distribución a posteriori.

Conocidas la verosimilitud y la distribución a priori, podemos obtener la distribución a posteriori

$$p(\mathbf{s}|\mathbf{x}) = \mathcal{N}(\mathbf{s} \mid (\mathbf{U}^\top\mathbf{U} + \tfrac{\sigma_\varepsilon^2}{\sigma_s^2}\mathbf{I})^{-1}\mathbf{U}^\top\mathbf{x},\ \ \sigma_\varepsilon^2(\mathbf{U}^\top\mathbf{U} + \tfrac{\sigma_\varepsilon^2}{\sigma_s^2}\mathbf{I})^{-1}),$$

que inmediatamente nos proporciona los siguientes estimadores para $\mathbf{s}$:

$$\hat{\mathbf{s}}_{\text{MMSE}} = \hat{\mathbf{s}}_{\text{MAP}} = \hat{\mathbf{s}}_{\text{MAD}} = (\mathbf{U}^\top\mathbf{U} + \tfrac{\sigma_\varepsilon^2}{\sigma_s^2}\mathbf{I})^{-1}\mathbf{U}^\top\mathbf{x}$$

Nótese que suponer una distribución a priori uniforme (usando $\sigma_s^2 \to \infty$ en la expresión anterior) convierte la solución MAP en la solución ML obtenida anteriormente.

También podemos plantearnos el problema de estimar la salida x_* correspondiente a la entrada $\mathbf{u}_*$, a la vista de los mismos datos. Para ello, calculamos

$$\begin{aligned} p(x_*|\mathbf{x}) &= \int_{\mathbb{R}^M} p(x_*|\mathbf{s})p(\mathbf{s}|\mathbf{x})d\mathbf{s} \\ &= \mathcal{N}(x_* \mid \mathbf{u}_*(\mathbf{U}^\top\mathbf{U} + \tfrac{\sigma_\varepsilon^2}{\sigma_s^2}\mathbf{I})^{-1}\mathbf{U}^\top\mathbf{x},\ \ \sigma_\varepsilon^2 + \sigma_\varepsilon^2\mathbf{u}_*(\mathbf{U}^\top\mathbf{U} + \tfrac{\sigma_\varepsilon^2}{\sigma_s^2}\mathbf{I})^{-1}\mathbf{u}_*^\top), \end{aligned}$$

de manera que

$$\hat{x}_{*\text{MMSE}} = x_{*\text{MAP}} = x_{*\text{MAD}} = \mathbf{u}_*(\mathbf{U}^\top\mathbf{U} + \tfrac{\sigma_\varepsilon^2}{\sigma_s^2}\mathbf{I})^{-1}\mathbf{U}^\top\mathbf{x} = \mathbf{u}_*\hat{\mathbf{s}}_{\text{MMSE}}$$

5.6 Cálculo online

Es posible obtener las soluciones anteriores de manera online.

5.6.1 Solución ML

Se puede obtener una aproximación a $\hat{\mathbf{s}}_{\text{ML}}$ online con coste computacional $\mathcal{O}(M)$ sin más que notar que

$$\hat{\mathbf{s}}_{\text{ML}} = \operatorname*{argmax}_{\mathbf{s}} p(\mathbf{x}|\mathbf{s}) = \operatorname*{argmin}_{\mathbf{s}} ||\mathbf{x} - \mathbf{U}\mathbf{s}||^2$$

y a continuación usar gradiente estocástico para minimizar $||\mathbf{x}-\mathbf{U}\mathbf{s}||^2$. La actualización de coeficientes que debe iterarse para efectuar la minimización es en este caso

$$\mathbf{s} \leftarrow \mathbf{s} - \mu(x[n] - \mathbf{u}[n]\mathbf{s})\mathbf{u}[n],$$

donde μ es un paso de adaptación "suficientemente pequeño". A este algoritmo se le llama least mean squares (LMS).

5.6.2 Solución Bayesiana

Se puede obtener de manera exacta $\hat{\mathbf{s}}_{\text{MMSE}}$ a medida que se dispone de más muestras (N aumenta) sin necesidad de rehacer todos los calculos, reusando la solución anterior. Para ello, se define $\mathbf{P}^{(N)} = (\mathbf{U}^\top\mathbf{U} + \frac{\sigma_\varepsilon^2}{\sigma_s^2}\mathbf{I})^{-1}$, $\mathbf{r}^{(N)} = \mathbf{U}^\top\mathbf{x}$ y se usa el siguiente cálculo recursivo

$$\begin{aligned}
\mathbf{P}^{(N+1)} &= \mathbf{P}^{(N)} - \frac{\mathbf{P}^{(N)}\mathbf{u}[N+1]^\top\mathbf{u}[N+1]\mathbf{P}^{(N)}}{1+\mathbf{u}[N+1]\mathbf{P}^{(N)}\mathbf{u}[N+1]^\top} \\
\mathbf{r}^{(N+1)} &= \mathbf{r}^{(N)} + \mathbf{u}[N+1]x[N+1] \\
\mathbf{s}^{(N+1)} &= \mathbf{P}^{(N+1)}\mathbf{r}^{(N+1)},
\end{aligned}$$

que sólo tiene un coste $\mathcal{O}(M^2)$ por paso (a diferencia de aplicar la ecuación original completa en cada paso, que tendría coste $\mathcal{O}(M^3)$). A este algoritmo se le llama recursive least squares (RLS).

5.7 Problemas

5.1. Considere la siguiente secuencia

$$u[1]\ldots u[7] \equiv 0.7,\ -0.1,\ 0.7,\ -0.2,\ -0.1,\ 1.5,\ -1.1$$

que se alimenta como entrada a un filtro lineal de tres coeficientes $\mathbf{s} = [s_1, s_2, s_3]^\top$. Se conocen los siguientes elementos de la secuencia de salida, (corrompidos con ruido Gaussiano de varianza 0.25):

$$x[1]\ldots x[6] \equiv -0.60,\ 1.13,\ 0.57,\ 0.42,\ 1.25,\ -2.58$$

a) ¿Cuál es la estimación ML de s? (filtro de Wiener basado en estadísticos aproximados).

b) Utilice el filtro obtenido para predecir $x[7]$, $\hat{x}_{\text{ML}}$.

c) Calcule las estimación MMSE, MAP y MAD de s asumiendo que la pdf a priori de sus componentes es $s_i \sim \mathcal{N}(0,1)$.

d) Obtenga la estimación MMSE de $x[7]$, $\hat{x}_{\text{MMSE}}$.

e) Calcule el error cuadrático esperado en la predicción b). (Es decir, la esperanza de $(\hat{x}_{\text{ML}} - x[7])^2$ a la vista de los datos disponibles).

f) Calcule el error cuadrático esperado en la predicción d). (Es decir, la esperanza de $(\hat{x}_{\text{MMSE}} - x[6])^2$ a la vista de los datos disponibles).

References

1. Hayes M H (1996) Statistical Digital Signal Processing and Modeling. John Wiley and Sons, New York, EE.UU.
2. Oppenheim A, Schaffer R (1999) Discrete-Time Signal Processing 2nd Ed. Prentice Hall, New York, EE.UU. Thesis, Columbia University, New York