Filtrado adaptativo

Lorena Álvarez Pérez

Universidad Carlos III de Madrid

Curso académico 2023/2024

Índice

Hasta ahora...

- Habíamos asumido que la señal de entrada al filtro ($x[n]$) y la señal deseada ($d[n]$) eran **procesos estocásticos estacionarios en sentido amplio** (WSS).
    - El filtro se podía estudiar como un sistema LTI
- ¿Cómo se diseña el filtro cuando las señales no son estacionarias?
    - Si dividimos la señal en tramos que puedan "considerarse" estacionarios y calcular los coeficientes del filtro para cada trama de forma independiente...
        - Si los procesos estocásticos son de variación muy rápida, los intervalos donde asumimos estacionariedad serán demasiados cortos $\Rightarrow$ estimaciones poco eficientes de los parámetros de estudio.
        - Tratar de modo independiente los intervalos, dificulta el análisis de la señal de modo completo.
- **¡Solución! Filtrado adaptativo**

Punto de partida

- Filtro FIR óptimo cuando la señal de entrada ($x[n]$) y señal deseada ($d[n]$) son procesos estocásticos estacionarios WSS:

$$\mathbf{w} = \mathbf{R}_X^{-1}\mathbf{r}_{X,D}$$

- Si los procesos varían con el tiempo, la matriz de autocorrelación de la señal de entrada $\mathbf{R}_X$, y el vector de correlación cruzada entre la señal de entrada y la señal deseada $\mathbf{r}_{X,D}$, serán variantes en el tiempo.

$$\left\{\begin{array}{l}\mathbf{R}_X[n]\\ \mathbf{r}_{X,D}[n]\end{array}\right. \Rightarrow \mathbf{w}[n] \text{ (el filtro también es función del tiempo)}$$

¿Un **filtro de Wiener** para **cada instante** n?

¡Poca aplicación práctica! Si en escenarios estacionarios era difícil conseguir una caracterización completa de las funciones de autocorrelación, para cada instante se torna más inviable aún.

Obtención coeficientes - diseño incremental

- Partimos de que en el instante n, tendremos un vector de pesos $\mathbf{w}[n]$.
- El vector de pesos correspondiente al instante $n+1$, se calcula como:

$$\mathbf{w}[n+1] = \mathbf{w}[n] + \Delta\mathbf{w}[n]$$

- **Objetivo**: diseñar el **algoritmo para construir** $\Delta\mathbf{w}[n]$
- Propiedades "deseables" de dicho algoritmo:
  - En un problema estacionario, el filtro adaptativo debería producir una secuencia de actualizaciones de $\mathbf{w}[n]$ que converja al filtro óptimo:

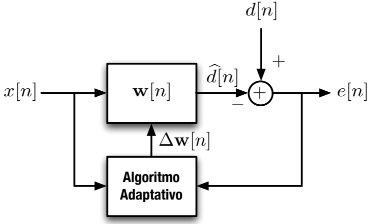
    $$\lim_{n\to\infty} \mathbf{w}[n] = \mathbf{R}_X^{-1}\mathbf{r}_{X,D}$$

    (se podría usar el filtro adaptativo en escenarios estacionarios, ahorrando la inversión de la matriz de autocorrelación).
  - No deberían tener que calcularse $\mathbf{R}_X[n]$ ni $\mathbf{r}_{X,D}[n]$
  - Con señales no estacionarias, el filtro debería ser capaz de adaptarse a los cambios en los estadísticos de las señales y realizar un seguimiento de la solución a medida que ésta evoluciona con el tiempo.

Esquema de un filtro adaptativo

El funcionamiento del filtro se divide en tres fases:

- Predicción: Cuando llega el vector $\mathbf{x}[n]$, el filtro realiza una predicción de la señal deseada $\widehat{d}[n] = \mathbf{w}[n]^\intercal \mathbf{x}[n]$
- Cálculo de error:
  Se le comunica al filtro el valor real de $d[n]$ y se calcula el error cometido por el filtro en la muestra: $e[n] = d[n] - \widehat{d}[n]$.
- Actualización:
  A partir de $e[n]$, se obtiene la actualización $\Delta\mathbf{w}[n]$ y, a continuación, los coeficientes $\mathbf{w}[n+1]$ que se usarán para procesar $x[n+1]$ que representa la observación que llega en el instante $n+1$.



Introducción

- Método iterativo de actualización del vector de pesos que considera la potencia del error de reconstrucción como una función de los pesos del filtro.
- Las actualizaciones del vector de pesos han de ir en la dirección de encontrar el mínimo de esta superficie de error.
- Sea $\xi[n] = \mathbb{E}\{e^2[n]\}$ (donde $e[n]^2 = f(\mathbf{w})$):
  - La dirección de máxima variación de $\xi[n]$ con $\mathbf{w}[n]$ viene dada por $\nabla_{\mathbf{w}}\xi[n]$.
  - Como nos interesa minimizar $\xi[n]$, la mejor actualización (para vector de pesos) es en la dirección opuesta al gradiente, $-\nabla_{\mathbf{w}}\xi[n]$.
  - Regla de actualización:

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \frac{\mu}{2}\nabla_{\mathbf{w}}\xi[n]$$

    (desde la posición actual, $\mathbf{w}[n]$, se da un paso de magnitud $\frac{\mu}{2}$ en la dirección dada por $-\nabla_{\mathbf{w}}\xi[n]$).
    A la constante μ se le conoce como paso de actualización.

Cálculo del gradiente

- El cálculo del gradiente $\nabla_{\mathbf{w}}\xi[n]$ se realiza a partir de la definición del error de reconstrucción: $e[n] = d[n] - \widehat{d}[n] = d[n] - \mathbf{w}^\intercal \mathbf{x}[n]$.

$$\begin{aligned}\nabla_{\mathbf{w}}\xi[n] &= \nabla_{\mathbf{w}}\mathbb{E}\{e[n]^2\} = \mathbb{E}\left\{\nabla_{\mathbf{w}} e[n]^2\right\} = \mathbb{E}\left\{\nabla_{\mathbf{w}}\left(d[n] - \mathbf{w}^\intercal \mathbf{x}[n]\right)^2\right\}\\ &= -2\mathbb{E}\left\{(d[n] - \mathbf{w}^\intercal \mathbf{x}[n])\mathbf{x}[n]\right\}\\ &= -2\mathbb{E}\left\{e[n]\mathbf{x}[n]\right\}\end{aligned}$$

  (Función de coste derivable $J(\mathbf{w})$, tal que para la solución óptima $\mathbf{w}_0$, se cumple que: $J(\mathbf{w}_0) < J(\mathbf{w})$).

- Por tanto, la regla de actualización de descenso por máxima pendiente será:

$$\boxed{\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\mathbb{E}\left\{e[n]\mathbf{x}[n]\right\}}$$

Aplicacion al filtro de Wiener

- La regla de descenso por máxima pendiente puede utilizarse en un caso de filtrado estacionario donde no convenga invertir las matrices de autocorrelación.
- Las actualizaciones vendrán dadas por:

$$
\begin{aligned}
\mathbb{E}\{e[n]\mathbf{x}[n]\} &= \mathbb{E}\{(d[n] - \mathbf{w}[n]^\intercal \mathbf{x}[n])\mathbf{x}[n]\} \\
&= \mathbb{E}\{(d[n]\mathbf{x}[n] - \mathbf{x}[n]^\intercal \mathbf{x}[n]\mathbf{w}[n])\} \\
&= \mathbb{E}\{d[n]\mathbf{x}[n]\} - \mathbb{E}\{\mathbf{x}[n]^\intercal \mathbf{x}[n]\}\mathbf{w}[n] \\
&= \mathbf{r}_{D,X} - \mathbf{R}_X \mathbf{w}[n]
\end{aligned}
$$

- El algoritmo de descenso por máxima pendiente será:

$$
\boxed{\mathbf{w}[n+1] = \mathbf{w}[n] + \mu\left(\mathbf{r}_{D,X} - \mathbf{R}_X \mathbf{w}[n]\right)}
$$

  Cuando el algoritmo encuentre un vector de pesos $\mathbf{w}[n]$ que coincida con el filtro de Wiener, el algoritmo se detiene, ya el que el filtro de Wiener verifica que:

$$
\mathbf{w}[n] = \mathbf{R}_X^{-1} \mathbf{r}_{D,X}
$$

  y el término de actualización sería nulo.

Paso de actualización

- Puede demostrarse que en escenarios estacionarios, el filtro adaptativo actualizado mediante el algoritmo de descenso por máxima pendiente converge al filtro óptimo si el paso de adaptación cumple:

$$0 < \mu < \frac{2}{\lambda_{max}}$$

donde λ_{max} es el autovalor más grande de $\mathbf{R}_X$.

Consideraciones

- El algoritmo de descenso por máxima pendiente es de **poca utilidad práctica**.
- Es necesario calcular $\mathbb{E}\{e[n]\mathbf{x}[n]\}$ y en la mayor parte de las aplicaciones, estos estadísticos se desconocen y deben estimarse a partir de los datos:
  - En casos estacionarios, las estimaciones de la autocorrelación de X_n y de la correlación cruzada entre X_n y D_n a partir de muestras, pueden ser más o menos, robustas.
  - En casos no estacionarios, es imposible conseguir estimaciones razonables de estos parámetros.

Introducción

- Una estrategia razonable puede ser sustituir dicha esperanza por una media muestral, calculada a partir de L observaciones previas de la señal de entrada, es decir, $x[n]$, $x[n-1]$, $x[n-2]$, …, $x[n-L+1]$, para obtener así estimaciones muestrales de dichos estadísticos:

$$\mathbb{E}\{e[n]\mathbf{x}[n]\} \approx \frac{1}{L}\sum_{l=0}^{L-1} e[n-l]\mathbf{x}[n-l]$$

- Nótese que hay una diferencia significativa entre L (número de instantes de tiempo que se utiliza para estimar los estadísticos) y P (número de coeficientes del filtro). En general, $L \gg P$.
  - P hace referencia al número de valores significativos de la función de autocorrelación para procesos estocásticos.
  - L denota el número de instantes de tiempo en los que se puede suponer que los estadísticos de estos procesos pueden considerarse estacionarios.

Introducción

- Caso particular: $L = 1$ (**algoritmo LMS**, del inglés *Least-Mean-Square algorithm*)[1].
  - Se utiliza el vector de entrada $\mathbf{x}[n]$ y error de reconstrucción actual $e[n]$ para realizar una estimación de rango uno de $\mathbb{E}\{E[n]X[n]\}$.
  - Ecuación de actualización:

$$\mathbf{w}[n+1] = \mathbf{w}[n] - \mu e[n]\mathbf{x}[n]$$

¡Observe que esta regla de actualización involucra vectores!

$$\begin{bmatrix} w_0[n+1] \\ w_1[n+1] \\ \vdots \\ w_{M-1}[n+1] \end{bmatrix} = \begin{bmatrix} w_0[n] \\ w_1[n] \\ \vdots \\ w_{M-1}[n] \end{bmatrix} + \begin{bmatrix} \mu e[n]x[n] \\ \mu e[n]x[n-1] \\ \vdots \\ \mu e[n]x[n-P+1] \end{bmatrix}$$

[1] Fue inventado en 1960 por Bernard Widrow (profesor de la Universidad de Standford) y su primer estudiante de doctorado (Ted Hoff).

Parámetro de actualización

- La convergencia del algoritmo LMS al filtro óptimo, en situaciones estacionarias, depende del parámetro μ (determina la magnitud de cada paso de actualización en la dirección opuesta al gradiente).
- El LMS converge *en la media al cuadrado* a la solución de Wiener si:
  - Se da independencia estadística entre el vector de pesos $\mathbf{w}[n]$ y los datos $\mathbf{x}[n]$
  - El paso de actualización (μ) verifica:

$$0 < \mu < \frac{2}{P\mathbb{E}\left\{x[n]^2\right\}}$$

donde P es el orden del filtro. En aplicaciones prácticas, $\mathbb{E}\left\{x[n]^2\right\}$ puede estimarse a partir de N valores de la señal $x[n]$ mediante:

$$\mathbb{E}\left\{x[n]^2\right\} \approx \frac{1}{N}\sum_{k=0}^{N-1} x[k]^2$$