Filtrado óptimo y adaptativo

El filtro de Wiener

La figura 1 muestra un problema general de filtrado óptimo. Las señales $x(n)$, y $d(n)$ son realizaciones de dos procesos estocásticos $X(n)$ y $D(n)$, respectivamente. $X(n)$ se conoce como proceso de entrada y $D(n)$ se conoce como proceso de referencia. Estos procesos estocásticos se suponen conjuntamente estacionarios en sentido amplio (WSS). $W(z)$ representa la función de transferencia de un filtro FIR con coeficientes ajustables. Las señales $y(n)$ son realizaciones del proceso estocástico a la salida del filtro, $Y(n)$.

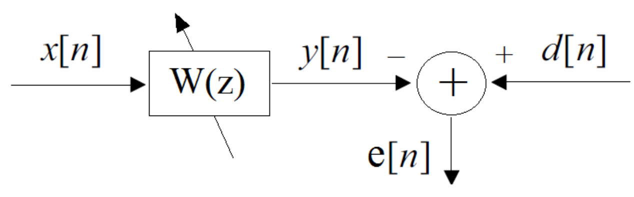

Figura 1: Filtrado óptimo.

El objetivo es diseñar el filtro $W(z)$, de M coeficientes $\mathbf{w} = (w_0, w_1, \cdots, w_{M-1})^T$, que minimice el error cuadrático medio (MSE) entre las señales $y(n)$ y $d(n)$.

$$MSE = E[e(n)^2] = E[(d(n) - y(n))^2]. \tag{1}$$

La respuesta impulsional del filtro es

$$w(n) = w_0\, \delta(n) + w_1\, \delta(n-1) + w_2\, \delta(n-1) + \ldots + w_{M-1}\, \delta(n-(M-1)).$$

Entonces, la salida del filtro

$$y(n) = x(n) * w(n) = \sum_{k=0}^{M-1} w_k\, x(n-k) = \mathbf{w}^T\, \mathbf{x}(n), \tag{2}$$

Jesús Pérez Arriaga

donde

$$\mathbf{w} = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{M-1} \end{bmatrix}, \quad \mathbf{x}(n) = \begin{bmatrix} x(n) \\ x(n-1) \\ \vdots \\ x(n-M+1) \end{bmatrix}.$$

Combinando (1) y (2), el MSE se puede expresar como función de los coeficientes del filtro

$$\begin{aligned} MSE = J(\mathbf{w}) &= E[(d(n) - \mathbf{w}^T\,\mathbf{x}(n))^2] \\ &= E[d(n)^2 + (\mathbf{w}^T\,\mathbf{x}(n))^2 - 2\mathbf{w}^T\,\mathbf{x}(n)d(n)] \\ &= r_d(0) + \mathbf{w}^T\mathbf{R}_x\mathbf{w} - 2\,\mathbf{w}^T\mathbf{p}, \end{aligned} \tag{3}$$

donde

$$\mathbf{R}_x = \begin{bmatrix} r_x(0) & r_x(1) & \dots & r_x(M-1) \\ r_x(1) & r_x(0) & \dots & r_x(M-2) \\ \vdots & \vdots & \ddots & \vdots \\ r_x(M-1) & r_x(M-2) & \dots & r_x(0) \end{bmatrix}, \quad \mathbf{p} = \begin{bmatrix} r_{x,d}(0) \\ r_{x,d}(-1) \\ \vdots \\ r_{x,d}(1-M) \end{bmatrix}.$$

$\mathbf{R}_x$ es la matriz de covarianza del vector $\mathbf{x}[n]$, cuyos elementos son muestras de la función de autocorrelación del proceso $x[n]$. Es una matriz Toeplitz simétrica. $\mathbf{p}$ es un vector formado por muestras de la función de correlación entre los procesos $x(n)$ y $d(n)$. Finalmente, $r_d[n]$ es la función de autocorrelación del proceso $d(n)$.

Entonces, el filtro que minimiza el MSE (filtro óptimo) será

$$\mathbf{w}_{opt} = \arg\min_{\mathbf{w}}\ J(\mathbf{w})$$

Igualando a cero el gradiente de (3) se llega a las llamadas ecuaciones normales,

$$\nabla J(\mathbf{w}) = 2\mathbf{R}_x\mathbf{w} - 2\mathbf{p} = 0 \quad \Leftrightarrow \quad \mathbf{R}_x\mathbf{w}_{opt} = \mathbf{p}. \tag{4}$$

Entonces los coeficientes del filtro óptimo serán

$$\mathbf{w}_{opt} = \mathbf{R}_x^{-1}\mathbf{p}. \tag{5}$$

Por tanto, para obtener el filtro óptimo solo se necesita conocer la función de autocorrelación del proceso a la entrada, $r_x(m)$ y la función de correlación cruzada entre el proceso a la entrada y el proceso de referencia, $r_{x,d}(m)$.

Sustituyendo (5) en (3), se obtiene el MSE para el filtro óptimo

$$\begin{aligned} J(\mathbf{w}_{opt}) &= r_d(0) + \mathbf{w}_{opt}^T \mathbf{R}_x \mathbf{w}_{opt} - 2\,\mathbf{w}_{opt}^T \mathbf{p} \\ &= r_d(0) - \mathbf{w}_{opt}^T \mathbf{p}. \end{aligned}$$

Además, es fácil demostrar que el MSE para un filtro $\mathbf{w}$ se puede escribir en función de $\mathbf{w}_{opt}$ como sigue

$$J(\mathbf{w}) = J(\mathbf{w}_{opt}) + (\mathbf{w} - \mathbf{w}_{opt})^T \mathbf{R}_x (\mathbf{w} - \mathbf{w}_{opt}).$$

Se puede demostrar que la función $J(\mathbf{w})$ solo tiene un mínimo, y este es el filtro de Wiener (5).

Aplicación práctica del filtro de Wiener

En la aplicación del filtro de Wiener a muchos problemas prácticos se distinguen dos fases: la fase de entrenamiento y la fase operacional.

La fase de entrenamiento comprende todas las operaciones/cálculos necesarios para calcular los coeficientes del filtro de Wiener. En muchos casos prácticos las características de los procesos $x(n)$ y $d(n)$ son desconocidas. Entonces, para calcular el filtro de Wiener, es necesario estimar $r_x(m)$ y $r_{x,d}(m)$ previamente. Si son procesos conjuntamente ergódicos, $r_x(m)$ y $r_{x,d}(m)$ se pueden estimar a partir de una sola realización de $x(n)$ y $d(n)$.

Una vez obtenido el filtro óptimo se pasa a la parte operacional, donde las realizaciones del proceso $x(n)$ se filtran óptimamente con el filtro de Wiener obtenido en la fase de entrenamiento.

En la práctica no es estrictamente necesario que los procesos $x(n)$ y $d(n)$ sean conjuntamente WSS. Si $r_x(m)$ y $r_{x,d}(m)$ cambian lentamente con el tiempo, se pueden realizar fases de entrenamiento periódicamente para disponer del filtro óptimo en cada momento. El periodo entre fases de entrenamiento consecutivas debe ser menor que el periodo en que $r_x(m)$ y $r_{x,d}(m)$ cambian significativamente.

El algoritmo LMS

El filtro de Wiener se puede obtener, sin necesidad de invertir la matriz de autocorrelación, mediante un procedimiento iterativo, que se denomina “Algoritmo de Máximo Descenso” o “Steepest Descent”. Esta técnica está en la base del algoritmo LMS (Least Mean Squares) y, por ello, la describimos aquí en primer lugar.

La expresión (3) representa una función de coste cuadrática en los coeficientes del filtro y cuyo mínimo (el único que tiene) es el filtro de Wiener que viene dado por (5). El Algoritmo de Máximo Descenso obtiene la solución de Wiener moviéndose por la función de coste en la dirección contraria al gradiente. Es decir, empezando en un filtro cualquiera $\mathbf{w}_0$, el algoritmo actualiza los coeficientes del filtro en cada iteración de acuerdo a

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \nabla J(\mathbf{w}) \tag{6}$$

donde μ es un paso de adaptación (también denominada constante de aprendizaje) que controla cuanto nos movemos en la dirección contraria al gradiente.

Considerando la expresión del gradiente de la función de coste (4), el Algoritmo de Máximo Descenso se reduce a

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \left(\mathbf{p} - \mathbf{R}_x \mathbf{w}_n \right). \tag{7}$$

Se puede demostrar que el algoritmo de máximo descenso converge a la solución de Wiener si $0 < \mu \leq 2/\lambda_{max}$, donde λ_{max} es el mayor autovalor de la matriz $\mathbf{R}_x$.

Tanto en el Algoritmo de Máximo Descenso, como en la derivación teórica del filtro de Wiener, es necesario conocer (o estimar) la matriz de autocorrelación y el vector de correlación cruzada. El algoritmo LMS, propuesto por Widrow y sus colaboradores en el año 1960, supone un salto conceptual respecto al Algoritmo de Máximo Descenso que consiste en substituir el gradiente exacto de la función de coste por una estima instantánea del mismo. Teniendo en cuenta que la función de coste es el error cuadrático medio (ecuación (3)), (6) se puede escribir así

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \nabla \left(E[e(n)^2] \right). \tag{8}$$

Substituyendo el gradiente exacto por una estima instantánea del mismo

$$\mathbf{w}_{n+1} = \mathbf{w}_n - \mu \nabla \left(e(n)^2 \right). \tag{9}$$

Finalmente, teniendo en cuenta que $e(n) = d(n) - \mathbf{w}_n^T \mathbf{x}(n)$, el algoritmo LMS queda

$$\mathbf{w}_{n+1} = \mathbf{w}_n + \mu \, e(n) \, \mathbf{x}(n). \tag{10}$$

El algoritmo se inicializa en algún valor adecuado, $\mathbf{w}_0$, dependiendo de la aplicación (aleatoriamente, todo ceros, una delta en el centro del filtro, etc...). Se puede demostrar que el algoritmo converge si se cumple que $0 < \mu < 2/(M\sigma_x^2)$.

Es de destacar la notable simplicidad del algoritmo, que lo convierten en el algoritmo adaptativo más ampliamente utilizado.