

Filtrado lineal óptimo

José Antonio Morán Moreno
Joan Claudi Socoró Carrié

PID_00175662



Los textos e imágenes publicados en esta obra están sujetos –excepto que se indique lo contrario– a una licencia de Reconocimiento-NoComercial-SinObraDerivada (BY-NC-ND) v.3.0 España de Creative Commons. Podéis copiarlos, distribuirlos y transmitirlos públicamente siempre que citéis el autor y la fuente (FUOC. Fundació per la Universitat Oberta de Catalunya), no hagáis de ellos un uso comercial y ni obra derivada. La licencia completa se puede consultar en <http://creativecommons.org/licenses/by-nc-nd/3.0/es/legalcode.es>

Índice

Introducción.....	5
1. El problema de la estimación lineal.....	9
2. La necesidad de medir magnitudes en optimización.....	11
3. Estimación lineal óptima.....	13
4. Revisión de conceptos básicos de estadística y señales aleatorias.....	15
4.1. Función densidad de probabilidad, media y varianza	16
4.2. Procesos estocásticos o señales aleatorias	17
4.3. Procesos estocásticos estacionarios	18
4.4. Procesos ergódicos	18
5. El filtro óptimo o filtro de Wiener.....	20
5.1. Filtro lineal óptimo de orden cero	20
5.2. Filtro lineal óptimo de orden N	23
5.3. El principio de ortogonalidad para un filtro de orden N	26
6. El predictor lineal.....	28
7. Conclusiones.....	29
8. Ejemplos prácticos de uso del filtrado lineal óptimo.....	30
8.1. Identificación de un sistema	30
Anexo.....	33

Introducción

En el módulo “Diseño de filtros discretos” hemos estudiado las características principales de los filtros y sus técnicas de diseño en el dominio frecuencial. Como bien conocéis, el dominio de la frecuencia resulta de gran utilidad en el ámbito del procesado de señal, permitiendo la representación de las señales en un dominio alternativo que nos permite extraer una información vital para muchas aplicaciones de telecomunicaciones.

Sabemos que si queremos realizar una transmisión de varias señales sin interferencias, sencillamente podemos poner la información en bandas frecuenciales separadas y una operación de filtrado sencilla nos ayudará a recuperar la información de cada una de las señales en el receptor. Los filtros y las técnicas de filtrado nos permiten el diseño de los dispositivos que permiten realizar esta separación en bandas frecuenciales y recuperar las señales originales.

No obstante, existen un buen número de aplicaciones y problemáticas reales donde no tenemos la suerte de tener las señales en bandas separadas, haciendo que el problema del filtrado frecuencial clásico deje de ser útil para resolver dichos problemas.

¿Podemos separar dos señales que se encuentran compartiendo la misma banda de frecuencias? ¿Existen técnicas de filtrado que nos permitan realizar esa operación?

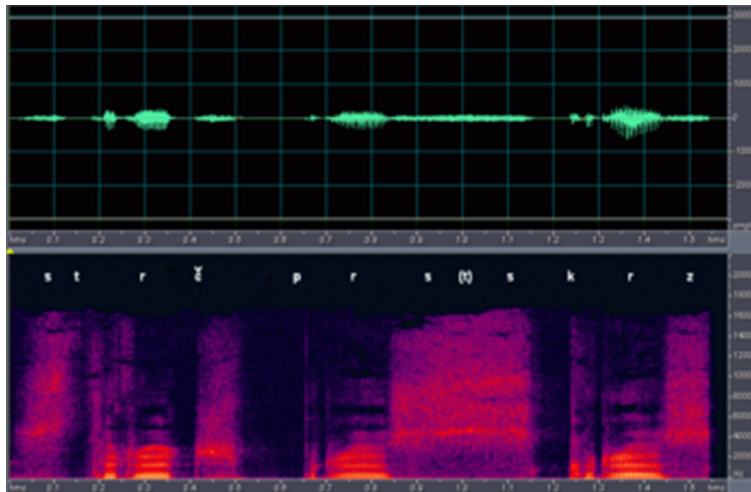
Pensemos en el ejemplo de una transmisión de audio desde un helicóptero. La señal de audio se recibe mezclada con el ruido del rotor del helicóptero que la contamina. Desde una perspectiva de filtrado clásico, lo básico que podríamos realizar es filtrar la señal recibida a la banda de frecuencias donde se encuentra la señal de voz. Si sabemos que la voz tiene un espectro entre 300 Hz y 3.400 Hz aproximadamente, podríamos filtrar la señal recibida a esa banda y mejorar la calidad por una reducción de la potencia de ruido.

La pregunta a realizar en este momento sería, **¿es eso lo máximo que podemos hacer o existe alguna posibilidad de mejorar las prestaciones de este filtrado básico?**

En este módulo analizaremos otras posibilidades de filtrado, las que llamaremos filtrado estadístico. Dentro de todas las posibilidades de filtrado estadístico, que son múltiples y variadas, nos centraremos en algunas específicas, que son las más utilizadas en el ámbito del procesado de señal y sus aplicaciones.

Aunque el estudio de la teoría de señal suele realizarse sobre señales determinísticas, la realidad es que la mayoría de aplicaciones, prácticamente en su totalidad, se realizan sobre señales de naturaleza aleatoria. La señal de voz humana, por mucho que se encuentre acotada en una banda frecuencial determinada, posee unas características frecuenciales que van variando con el tiempo.

Figura 1



Fuente: es.wikipedia.org

Tal y como se observa en el espectrograma, aunque el espectro está restringido frecuencialmente vemos cómo varía sustancialmente en función de los sonidos que se están emitiendo. Podemos imaginar que el filtrado frecuencial, a pesar de poderse utilizar en esta aplicación, tiene grandes limitaciones al no aprovechar todas las características del espacio de señal.

El filtrado óptimo pretende sentar las bases teóricas para poder abordar problemas de filtrado óptico basado en las características estadísticas de las señales implicadas.

Gran número de aplicaciones de procesamiento de señal requieren la utilización de técnicas estadísticas:

- Señales contaminadas por ruido externo (ruido rotor en conversación por helicóptero).
- Señal recibida por un sistema de comunicaciones distorsionada por el canal.
- Imagen capturada por una cámara distorsionada por los efectos del movimiento.

La mayoría de veces, los efectos distorsionantes en estos ejemplos de aplicaciones no son procesos estacionarios, sino que sus características estadísticas varían con el tiempo.

- El ruido del rotor del helicóptero o las revoluciones del motor del coche cambian durante la comunicación.
- Los canales de radiofrecuencia cambian sus características en función de las condiciones de propagación (hora del día, estación del año, lluvia, etc.)

Todo ello hace que las técnicas de filtrado deban trabajar en el dominio del espacio de señal estadístico, pero que a su vez dispongan de la capacidad de adaptarse a los cambios de entorno. Como podemos observar, las aplicaciones reales requieren de técnicas óptimas de procesado para dar solución a los problemas clásicos de comunicación y de procesado de señal.

1. El problema de la estimación lineal

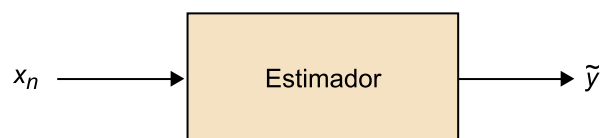
Cuando abordamos un problema de procesamiento estadístico y se pretende recuperar las características de una señal nos encontramos dentro de un problema de la teoría de la estimación.

La teoría de la estimación es una teoría matemática muy amplia que busca métodos para obtener un valor aproximado de un parámetro determinado a partir de los datos proporcionados por una muestra:

- La estimación de un valor de temperatura en función de los parámetros atmosféricos.
- La estimación de una señal de voz en una grabación policial distorsionada.
- La estimación de la señal transmitida por un sistema de telecomunicaciones en función de la señal recibida.
- La estimación de la velocidad de un objeto en un radar.

En todos los ejemplos anteriores, como podemos observar, resultan necesarias las muestras de unos parámetros y un parámetro objetivo para estimar. El sistema que se encargará de realizar la estimación será el estimador.

Figura 2



Existen diferentes criterios para diseñar el sistema de estimación siendo los siguientes algunos de los más conocidos:

- Estimador MAP (*maximum a posteriori*)
- Estimador ML (*maximum likelihood*)
- Estimador MS (*mean square*)
- Estimador LMS (*linear mean square*)

Los diferentes tipos de estimadores quedan fuera de los objetivos de esta asignatura. En cualquier caso, principalmente debéis quedaros con la idea de que los estimadores dependen del conocimiento de los parámetros estadísticos de las variables implicadas. Si tenemos un conocimiento muy completo de la función de probabilidad de las variables implicadas, se pueden desarrollar técnicas de estimación más fiables. Como contrapartida, cuanto más complejo

sea el conocimiento estadístico necesario, mayor será el coste computacional de las operaciones requeridas para hacer la estimación, y consecuentemente, será más compleja la implementación física del estimador, especialmente si buscamos aplicaciones en tiempo real.

En este escenario vemos que resulta importante el compromiso entre complejidad del estimador y el coste computacional del mismo, especialmente cuando en aplicaciones reales no podemos tener la certeza absoluta de la función densidad de probabilidad real de las variables implicadas.

Es por ello que en este módulo nos centraremos en los **estimadores lineales**, que corresponderán a estructuras de filtros lineales, que se diseñarán siguiendo un criterio estadístico para minimizar el error de estimación.

2. La necesidad de medir magnitudes en optimización

Cuando se abordan problemas de filtrado óptimo es necesario disponer de un criterio para definir magnitudes. En el caso de procesamiento de señal trabajaremos en espacios vectoriales, y consecuentemente en un espacio vectorial debemos tener una forma de medir las magnitudes de los vectores implicados.

La norma del vector estará definida según el tipo de vectores con los que se esté trabajando, siendo diferente si estamos en un espacio vectorial R^N o en un espacio de variables aleatorias. En cualquier caso, la norma vectorial debe cumplir una serie de propiedades que permitan operar entre vectores y disponer de criterios válidos de medición de magnitudes.

Las condiciones básicas que debe cumplir una **norma** son las siguientes.

Sea V un **espacio vectorial** sobre un **cuerpo** \mathbb{k} y \vec{x} un vector del espacio. Se dice que $\|\cdot\| : V \rightarrow \mathbb{R}$ es un operador que define la norma de \vec{x} , y escribimos $\|\vec{x}\|$, si cumple:

1) Para todo \vec{x} de V su norma ha de ser no negativa, y será cero si, y solo si, \vec{x} es el vector cero: $0 < \|\vec{x}\|$ si $\vec{x} \neq \vec{0}$, y $\|\vec{x}\| = 0 \Leftrightarrow \vec{x} = \vec{0}$.

2) Para todo \vec{x} de V y para todo k de \mathbb{k} , se satisface que $\|k\vec{x}\| = \|k\| \cdot \|\vec{x}\|$.

3) Para todo \vec{x} e \vec{y} de V se cumple que $\|\vec{x} + \vec{y}\| \leq \|\vec{x}\| + \|\vec{y}\|$ (desigualdad triangular).

Cualquier operador que cumpla estas tres condiciones, y en cualquier geometría, será un operador norma.

Página web

http://es.wikipedia.org/wiki/Norma_vectorial

Como vemos, los criterios de una norma son criterios sensatos que facilitan la operación entre vectores. La norma siempre ha de ser positiva, no tiene sentido tener vectores con norma negativa. Si multiplicamos un vector por un factor k , su norma se verá incrementada en el mismo factor, y por último, la tercera propiedad corresponde a la desigualdad triangular. La norma de la suma de vectores será siempre inferior o igual a la suma de normas.

En las aplicaciones de procesamiento estadístico, cuando se trabaje con variables aleatorias en problemas de optimización necesitamos un espacio vectorial con una norma bien definida que nos permita realizar las combinaciones lineales de los vectores implicados, así como las operaciones de medición de distancias necesarias para un proceso de optimización. Matemáticamente, esto se conoce como un espacio de Hilbert.

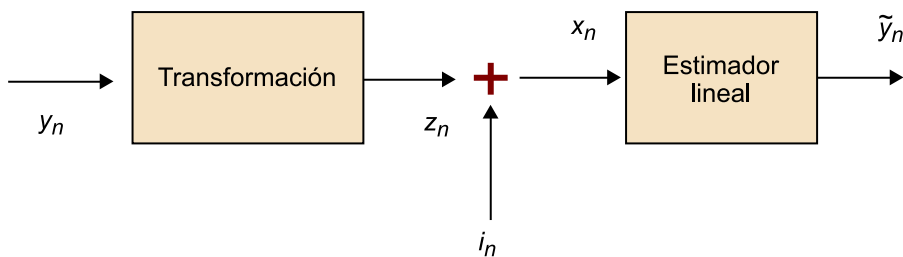
Ved también

No se realizará aquí una exposición detallada, pero podéis consultar las principales propiedades de un espacio de Hilbert en el anexo de este mismo módulo.

3. Estimación lineal óptima

Para abordar la teoría de la estimación lineal óptima, partiremos de un esquema general, a partir del cual se irán introduciendo los conceptos para desarrollar la teoría de la estimación lineal óptima. El modelo de partida general será el siguiente:

Figura 3



Ahora podremos identificar las diferentes señales implicadas en el proceso:

- y_n corresponde a la señal que queremos estimar. Esta señal no está disponible para realizar el procesamiento pero corresponde a la señal objetivo. En el caso de la aplicación de cancelación de ruido del helicóptero sería la voz real del piloto.
- z_n corresponde a la voz del piloto que mide el micrófono. En principio puede ser una señal muy similar pero estará distorsionada en mayor o menor medida debido a los efectos acústicos de la cabina, así como a las propias propiedades del micrófono.
- i_n corresponde a las señales de interferencia que se suman, en este caso sería el ruido del rotor así como otros ruidos que se puedan capturar en la cabina.
- x_n corresponde a la señal real con la que vamos a trabajar, son las muestras reales que se capturan a través del micrófono y que corresponden a la señal original, transformada con todas las posibles interferencias de ruido sumadas en la cabina.

El objetivo del estimador será utilizar las muestras de señal capturadas por el micrófono, x_n , para realizar la mejor estimación posible de la señal original, y_n , siguiendo el criterio de optimización definido.

Teniendo esto en cuenta, podemos ver que la bondad del estimador dependerá de la relación estadística que guarde la señal original con la señal transformada. Llevándolo al extremo, si la señal transformada pierde parte de información en la transformación, esta información no será recuperable.

Otro aspecto importante en la recuperación de la señal original es la independencia estadística de las señales de interferencia respecto de la original. Si la señal de interferencia se superpone en el campo de información de la señal original, esto perjudicará a la calidad de la estimación. En cambio, si las señales interferentes son estadísticamente independientes de la señal original, el procesador podrá aislar los dos espacios de señal y realizar una estimación de mejor calidad. En este ejemplo concreto, si el ruido del rotor es estadísticamente independiente de la voz, la estimación será de mejor calidad, pero si existe superposición en la información, se perderá calidad en la estimación.

A modo ilustrativo sabemos que los diferentes sonidos tienen características estadísticas diferentes, y la voz varía sus características en función de los sonidos que emite. Si tenemos un sonido que es muy diferente del rotor, por ejemplo una vocal, podemos esperar que la estimación sea de buena calidad, pero cuando estemos emitiendo sonidos fricativos, como por ejemplo la r que tiene una componente más similar al ruido de fondo, podemos esperar entonces una pérdida de calidad de la señal estimada.

En el caso concreto del estimador lineal óptimo, la variable estimada se generará a partir de la combinación lineal de las muestras de la señal de entrada, o lo que es lo mismo, como la salida de un filtro lineal de respuesta impulsional finita FIR:

$$\hat{y}[n] = \sum_{k=n_a}^{n_b} h^*[n-k]x[k]$$

Asimismo, al trabajar sobre espacios vectoriales, se considera que todas las operaciones entre las señales, ya sea de entrada como las interferencias, corresponden a combinaciones lineales de variables dentro del espacio vectorial.

Se entenderá por **filtro lineal óptimo** aquel que minimiza el valor esperado de la función de error al cuadrado, es decir:

$$E[(\hat{y}[n] - y[n])^2] \Big|_{MIN}$$

El problema de optimización se resolverá encontrando los coeficientes óptimos del filtro estimador que minimizan esta función de error.

4. Revisión de conceptos básicos de estadística y señales aleatorias

Como revisión de los conceptos básicos de probabilidad y señales aleatorias, en este apartado se presentarán las principales propiedades que conviene revisar para abordar con éxito el resto del módulo. Se considera que disponéis de conocimientos de probabilidad y estadística y los conceptos presentados en este punto no son más que una mera revisión de los conocimientos que ya tenéis.

En el ámbito de la estadística y la teoría de la probabilidad, se definen los **procesos estocásticos** como un concepto matemático que caracteriza una sucesión de variables aleatorias, que evolucionan en función de otra variable, que normalmente corresponde al tiempo.

En principio, en la definición completa de un proceso estocástico sería necesario conocer la función densidad de probabilidad de cada una de las variables para cada instante de tiempo; evidentemente esto resulta complejo de conocer en aplicaciones reales, así que se definirán un subconjunto de procesos estocásticos que se adaptan mejor a las aplicaciones reales.

Dentro del ámbito de las series temporales, existen diferentes ejemplos de procesos estocásticos:

- Señales del ámbito de la telecomunicación y la electrónica.
- Señales de aplicaciones biomédicas (encefalogramas, electrocardiogramas, etc.).
- Señales procedentes de procesos sísmicos o geológicos.
- Evolución de la población de una ciudad, municipio o país.
- Índices de bolsa.
- Manchas solares.

Y un largo etcétera, son ejemplos de procesos estocásticos que nos podemos encontrar en los ámbitos de aplicación.

La caracterización completa de un proceso estocástico es tan compleja que difícilmente podemos encontrar aplicaciones reales donde sea posible esa caracterización, de forma que tendremos que trabajar con algunos casos especiales de procesos estocásticos, con unas características particulares que facilitan su uso en las aplicaciones reales a partir de los datos conocidos.

Los procesos estocásticos se pueden definir en tiempo continuo o en tiempo discreto. En este caso nos centraremos en los procesos estocásticos en tiempo discreto que son los que utilizan en las aplicaciones de procesamiento de la señal digital.

4.1. Función densidad de probabilidad, media y varianza

Supóngase que x es una variable aleatoria con función densidad de probabilidad $p(x)$. Su media, varianza y su momento de segundo orden estarán definidos de la siguiente forma:

$$m = E[x] = \int_{-\infty}^{\infty} x p(x) dx \text{ como la media de la variable aleatoria.}$$

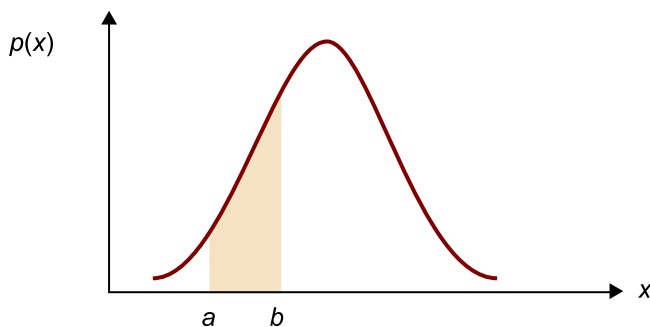
$$\sigma^2 = \text{var}(x) = E[(x - m)^2] = \int_{-\infty}^{\infty} (x - m)^2 p(x) dx \text{ como la varianza.}$$

$$E[x^2] = \int_{-\infty}^{\infty} x^2 p(x) dx \text{ como el momento de segundo orden.}$$

La probabilidad de que una variable aleatoria tenga un valor dentro de un

intervalo de valores $[a, b]$ viene definida por $\text{Prob}[a \leq x \leq b] = \int_a^b p(x) dx$.

Figura 4



La función densidad de probabilidad, como sabemos, está siempre normalizada a 1, dado que la probabilidad de que la variable adquiera cualquier valor en todo su conjunto de valores siempre será 1. Esto lo podemos representar de forma matemática de la siguiente forma:

$$\int_{-\infty}^{\infty} p(x) dx = 1$$

Esta misma propiedad nos permite determinar con un sencillo cálculo que:

$$\sigma^2 = \text{var}(x) = E[(x - m)^2] = E[x^2] - m^2$$

Las variables aleatorias pueden estar definidas sobre un dominio continuo o discreto. En este caso consideramos que la variable aleatoria está definida sobre un dominio continuo.

4.2. Procesos estocásticos o señales aleatorias

Un proceso aleatorio se define como una secuencia de variables aleatorias $\{x_0, x_1, x_2, \dots, x_n\}$ donde n habitualmente será la variable tiempo, o muestras temporales. En este caso, la descripción completa del proceso o señal aleatoria requeriría el conocimiento de todas las funciones densidad de probabilidad conjuntas:

$$p(x_0, x_1, x_2, \dots, x_n) \text{ para } n = 0, 1, 2, \dots$$

Podemos intuir que un conocimiento tan exhaustivo de la función densidad de probabilidad conjunta será prácticamente imposible de tener en aplicaciones reales, de forma que tendremos que conformarnos con un conocimiento más limitado de las mismas.

En las aplicaciones que se presentarán en esta asignatura normalmente se trabaja con señales de media cero. En cualquier caso, si la señal no presenta media cero, será fácil redefinir una señal equivalente restando el valor medio y simplificar así todo el proceso de análisis, que se complicaría en el caso de señales de media diferente a cero. Una vez finalizado todo el tratamiento, es tan fácil como volver a añadir la media al resultado final.

Señales de media cero

Si $E[x] = m$, podemos definir $x' = x - m$, donde x' es una variable de media cero.

Para trabajar con filtros lineales óptimos, más adelante veremos que tendremos que utilizar con frecuencia la función de autocorrelación. La función de autocorrelación proporciona un indicador de la influencia que tiene una variable aleatoria en un instante de tiempo sobre otra que se encuentra K muestras separada.

$$R_{xx}(n+k, n) = E[x_{n+k}x_n]$$

Vemos que la función de correlación determina la información estadística conjunta entre la variable en el instante n y aquella desplazada a $n+k$. Normalmente se conoce como *lag* a la separación relativa entre variables aleatorias.

4.3. Procesos estocásticos estacionarios

Como se ha comentado anteriormente, el conocimiento exhaustivo de la función densidad de probabilidad de los procesos estocásticos es algo muy complejo de conocer en aplicaciones reales. No obstante, los procesos estacionarios suponen un subconjunto que se adapta mucho mejor a las condiciones necesarias para desarrollar aplicaciones.

Se considera que un proceso estocástico es **estacionario en sentido estricto** cuando las funciones densidad de probabilidad de las variables aleatorias implicadas son invariantes ante un cambio de tiempo. Es decir, si realizamos un cambio en el índice temporal encontraremos la misma secuencia exacta de variables aleatorias. Como podemos imaginar, resulta complejo demostrar en aplicaciones reales que se cumple esta propiedad, de tal forma que aún tendremos que relajar un poco las condiciones a cumplir en caso de aplicaciones reales.

Un proceso estocástico será **estacionario en sentido amplio** cuando los dos momentos de primer orden, es decir, la media y la covarianza, se mantengan estables en el tiempo. Es decir, el valor de la media será constante para cualquier instante de tiempo y el valor de la correlación únicamente dependerá de la diferencia temporal entre variables, o sea, será igual la correlación entre la muestra 1 y 3 que entre la 5 y 7, o entre la 9 y la 11. Esto simplifica enormemente las condiciones que debe verificar un proceso en cuanto a caracterización de las variables aleatorias implicadas y hace mucho más sencillo realizar aplicaciones bajo estos supuestos.

$$R_{xx}(k) = E[x_{n+k}x_n] = E[x_{n'+k}x_{n'}]$$

Observamos que en este caso la función de autocorrelación no depende del valor temporal n sino que quedará completamente caracterizada por el valor del desplazamiento o *lag*.

Por otra parte, al diseñar filtros lineales, veremos que para el cálculo de los coeficientes óptimos tan solo nos será necesario conocer las medias y las correlaciones, de tal forma que los procesos estocásticos en sentido amplio son suficientes para poder realizar las aplicaciones.

4.4. Procesos ergódicos

Como hemos visto hasta el momento, hemos ido haciendo un proceso de simplificación de los procesos estocásticos y sus características estadísticas para poder acercarnos a aplicaciones reales. Existe otro paso que tendremos que realizar para poder utilizar todos estos conceptos en el campo de las aplicaciones reales.

Supongamos que tenemos una muestra de voz, y que esta la podemos caracterizar como un proceso estacionario en sentido amplio. El conocimiento de las estadísticas que nos permitiese estimar las medias y varianzas de las variables implicadas requeriría el conocimiento de múltiples realizaciones del mismo proceso. Esto sería en una aplicación disponer de múltiples realizaciones de la señal de voz, cosa que no siempre será posible en las aplicaciones reales.

Imaginemos el caso de una señal de voz que sirve de prueba policial y que requiere la cancelación de ruido de fondo para extraer claramente la conversación. En la prueba de voz, la única de la que dispondremos, probablemente tendremos la voz del sujeto objetivo camuflada con las interferencias u otras voces que puedan aparecen en la grabación. Suponiendo que el proceso es estacionario en sentido amplio para realizar la resolución del problema necesitaríamos la estimación de los momentos de primer orden y de segundo orden del proceso implicado. Al no disponer más que de una realización temporal del proceso, la única posibilidad que nos queda es estimar las medias estadísticas a partir de las medias temporales, ya que en cualquier otro caso sería imposible la resolución del problema al no disponer más que de una realización del proceso. En aquellos procesos donde el cálculo de las medias estadísticas se puede realizar a partir de medias temporales podemos hablar de procesos ergódicos. Tendremos que realizar esta suposición para poder trabajar con aplicaciones reales donde solo conozcamos una realización del proceso, que por otra parte son la mayoría de aplicaciones que nos encontraremos en los casos reales.

En conclusión, podemos decir que en las aplicaciones donde aplicaremos el filtrado lineal óptimo consideraremos que trabajamos con procesos estacionarios de segundo orden y ergódicos, y esto lo mantendremos así durante toda la asignatura a fin de resolver los problemas de filtrado lineal óptimo o de Wiener.

5. El filtro óptimo o filtro de Wiener

El filtrado lineal óptimo o filtro de Wiener corresponde al diseño de un filtro estadístico lineal que minimiza la función de coste:

$$E[(d[n])^2]_{MIN}$$

Siendo $e[n]$ la función de error de estimación, es decir, la diferencia entre el señal deseado y la salida del filtro.

A fin de adentrarnos progresivamente en las principales características del filtro lineal óptimo, abordaremos el problema por etapas, analizando en primer lugar las características de un filtro de orden cero, es decir, un filtro cuya respuesta impulsional está formada por un único coeficiente, para seguir profundizando con los conceptos básicos para filtros de orden mayor.

5.1. Filtro lineal óptimo de orden cero

En el caso particular de un filtro lineal óptimo de orden cero, la respuesta impulsional del filtro $h[n]$ sería un único coeficiente. La resolución del problema es determinar el valor de este coeficiente que hace que se minimice el error cuadrático medio. Para ello se debe realizar la minimización de la siguiente función de error:

$$E[(y[n] - \hat{y}[n])^2], \text{ donde } \hat{y}[n] = ax[n]$$

La minimización de esta función de error implicaría el cálculo de la derivada e igualar a cero. No sería necesario el cálculo de la segunda derivada ya que al ser una función cuadrática tan solo dispone de un mínimo, y los máximos los tendrá para valores del coeficiente tendiendo a infinito.

El desarrollo de la ecuación nos lleva a:

$$E[(y[n] - \hat{y}[n])^2] = E[(y[n] - ax[n])^2]$$

Y encontraremos la minimización derivando con respecto al coeficiente a e igualando a cero:

$$\frac{dE[(y[n] - ax[n])^2]}{da} = -2E[(y[n] - ax[n])x[n]] = 0$$

Y la solución de la ecuación sería:

$$E[y[n]x[n]] - aE[(x[n])^2] = 0$$

De tal forma que el valor óptimo del coeficiente es:

$$a = \frac{E[y[n]x[n]]}{E[(x[n])^2]}$$

El filtro lineal óptimo de orden cero nos aporta una resolución simple pero que ya permite interpretar algunos de los aspectos fundamentales de los estimadores lineales óptimos. Vemos que el coeficiente es directamente proporcional a la correlación cruzada que exista entre la señal objetivo $y[n]$ y la señal de entrada $x[n]$. Si la señal que pretendemos estimar y la señal origen no comparten ninguna información común (la correlación sería cero), el coeficiente del filtro óptimo sería cero. Esto quiere decir que si no hay información compartida entre una señal y la otra, la mejor estimación que podemos hacer es cero. A medida que exista una correlación cruzada superior entre la señal objetivo y la señal de entrada, mejor será la estimación que podremos realizar y menor será la señal de error.

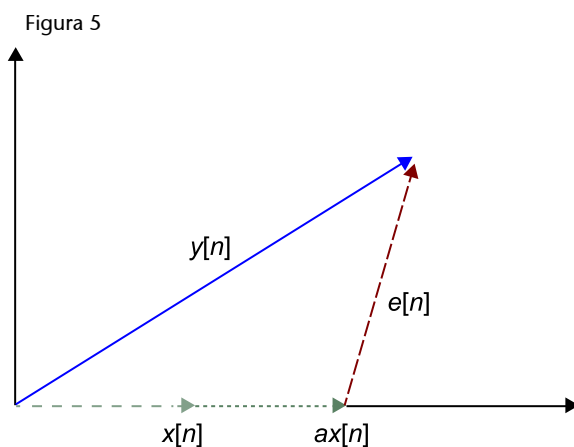
Nota

Es importante remarcar el hecho de que el filtro sea óptimo no quiere decir que el error sea muy pequeño, sino que el error es el menor posible con la información de la que se dispone. En el caso de señales incorreladas, la señal de error sería directamente la señal objetivo $y[n]$ dado que no se podría realizar ningún tipo de estimación mejor en ese caso.

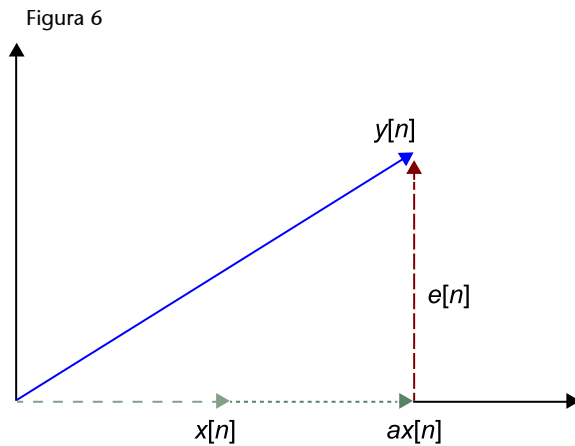
Para aclarar estos conceptos repasaremos algunas claves de geometría euclidiana que permita observar de forma visual estos conceptos. Teniendo en cuenta que las variables aleatorias con un producto escalar bien definido forman un espacio vectorial, podemos hacer uso de la geometría euclidiana para representar los conceptos de interés en el problema de optimización lineal óptima. Para ello se representarán las variables aleatorias como vectores y se utilizará como producto escalar la función de correlación entre variables.

Ejemplos

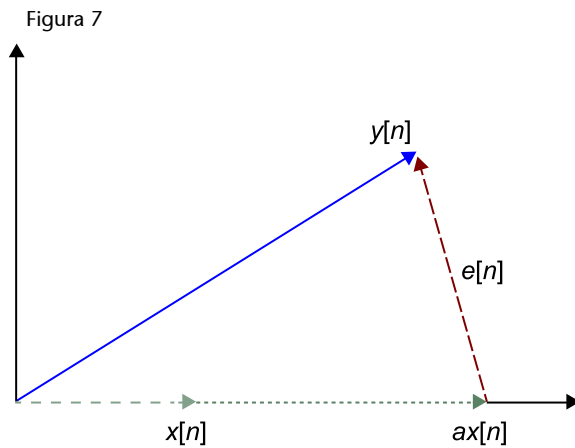
En la figura 5 podemos observar en azul oscuro la señal objetivo $y[n]$, en verde claro la señal de entrada $x[n]$ y en verde oscuro, la salida del filtro de un coeficiente $ax[n]$, siendo el vector de error el vector rojo.



En este primer ejemplo vemos que el coeficiente amplifica el vector pero todavía sigue quedando una parte del vector de error que proyecta sobre el eje x , de tal forma que el coeficiente puede aún variar actuando sobre el vector de error. Si a crece vemos que el vector de error todavía decrecerá.



En el segundo ejemplo (figura 6) se observa que el valor de a se ha incrementado y el vector de error tiene ahora mismo su valor mínimo, ya que cualquier modificación del valor a hará que el error crezca como se puede observar en la figura 7.



Haciendo uso de los conceptos expuestos previamente, y de forma puramente visual, determinamos que el filtro óptimo lineal, aquel que minimiza el vector de error lo obtendremos cuando el vector de error sea perfectamente ortogonal al espacio de información de entrada. Mientras el vector de error tenga proyección sobre el espacio de entrada, podremos seguir modificando el valor del coeficiente e ir reduciendo el error. En cambio cuando el vector de error sea perfectamente ortogonal, ya no será posible minimizar más el error y nos encontraremos en el caso óptimo.

Si recordamos la solución matemática para el caso óptimo:

$$\frac{dE[(y[n] - ax[n])^2]}{da} = -2E[(y[n] - ax[n])x[n]] = 0$$

Y a continuación reescribimos la ecuación de la siguiente manera:

$$E[(y[n] - ax[n])x[n]] = 0$$

$$E[e[n]x[n]] = 0$$

Encontramos la solución óptima cuando el vector de error es ortogonal al vector de entrada $x[n]$, es decir, la misma conclusión a la que se ha llegado mediante la interpretación geométrica de los vectores.

Este factor es el que se conoce como **principio de ortogonalidad**, que corresponde a la resolución óptima lineal cuando la función de coste es el error cuadrático medio. En lugar de plantear todas las ecuaciones y realizar las derivadas

pertinentes, se puede plantear el sistema de ecuaciones simplemente forzando que el error sea ortogonal a cada una de las variables de entrada de forma simultánea, cosa que simplifica en gran manera la operativa de resolución de este tipo de sistemas.

5.2. Filtro lineal óptimo de orden N

Una vez trabajados los conceptos del caso simple, se trata ahora de generalizar el problema a un filtro de orden N . Para que el desarrollo sea lo más completo posible lo realizaremos sobre los procesos estocásticos complejos, de tal forma que habremos realizado el caso más completo posible, siendo los procesos reales un caso particular de nuestro análisis.

En este ejemplo de análisis desarrollaremos primero el análisis matemático completo para obtener el sistema de ecuaciones que habrá que resolver, planteando finalmente las mismas ecuaciones a partir del principio de ortogonalidad, viendo la utilidad así del desarrollo realizado en el apartado anterior.

En el caso de un filtro de orden N de coeficientes complejos y en un espacio vectorial de procesos estacionarios y ergódicos complejos, tenemos la siguiente función de coste:

$$E[e^2]_{MIN} = E \left[\left(y - \vec{h}^H \vec{x} \right) \left(y - \vec{h}^H \vec{x} \right)^H \right]_{MIN}$$

Para el desarrollo de este cálculo se utiliza la notación matricial de la siguiente forma:

$$\begin{aligned} \hat{y} &= \vec{h}^H \vec{x} \\ e(y - \hat{y}) &= (y - \vec{h}^H \vec{x}) \\ \hat{y} &= \hat{y}[n] \quad y = y[n] \quad e = e[n] \\ \vec{x} &= \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-N] \end{bmatrix} \quad \vec{h} = \begin{bmatrix} h[0] \\ h[1] \\ \vdots \\ h[N] \end{bmatrix} \end{aligned}$$

Obsérvese que la salida del filtro corresponde al producto escalar de los coeficientes del filtro por un vector de muestras de entrada giradas en el tiempo y desplazadas. Este producto escalar no es más que la representación vectorial del sumatorio de convolución para la salida del filtro $\hat{y}[n] = \sum_{k=0}^N h^*[k]x[n-k]$, siendo $*$ el operador conjugado.

Ved también

Podéis revisar el anexo para hacer un repaso de los operadores vectoriales utilizados en el desarrollo.

La minimización de la función de coste J , definida como el valor esperado del error cuadrático medio, implica la minimización de la siguiente ecuación:

$$E[d^2] \big|_{MIN} = E[y y^H] - \vec{h}^H E[\vec{x} y^H] - E[\vec{y} \vec{x}^H] \vec{h} + \vec{h}^H E[\vec{x} \vec{x}^H] \vec{h} \big|_{MIN}$$

En este caso, al ser un problema N -dimensional, la minimización de la función de coste implicará la derivada vectorial igualada a cero. Al ser una función cuadrática, la resolución de la ecuación nos dará el mínimo de la función:

$$\frac{\partial E[d^2]}{\partial \vec{h}^*} = -E[\vec{x} y^H] + E[\vec{x} \vec{x}^H] \vec{h} = 0$$

Y la resolución de la misma nos lleva a:

$$\begin{aligned} -E[y^* \vec{x}] + E[\vec{x} \vec{x}^H] \vec{h} &= 0 \\ \vec{h}_{opt} &= E[\vec{x} \vec{x}^H]^{-1} E[y^* \vec{x}] = \vec{R}_{xx}^{-1} \vec{r}_{yx} \\ \vec{r}_{yx} = E[y^* \vec{x}] &= E \begin{bmatrix} y^*[n]x[n] \\ y^*[n]x[n-1] \\ \vdots \\ y^*[n]x[n-N] \end{bmatrix} = \begin{bmatrix} r_{yx}^*(0) \\ r_{yx}^*(1) \\ \vdots \\ r_{yx}^*(N) \end{bmatrix} \\ \vec{R}_{xx} = E[\vec{x} \vec{x}^H] &= E \begin{bmatrix} x[n]x^*[n] & x[n]x^*[n-1] & \dots & x[n]x^*[n-N] \\ x[n-1]x^*[n] & x[n-1]x^*[n-1] & \ddots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ x[n-N]x^*[n] & \dots & \dots & x[n-N]x^*[n-N] \end{bmatrix} = \\ &= \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \dots & r_{xx}[N] \\ r_{xx}[-1] & r_{xx}[0] & & \vdots \\ \vdots & & \ddots & \vdots \\ r_{xx}[-N] & \dots & \dots & r_{xx}[0] \end{bmatrix} \end{aligned}$$

Se observa que en el caso de un sistema N dimensional la resolución del problema implica la resolución de un sistema de ecuaciones de orden N , donde se implica el vector de correlación cruzada entre la señal objetivo y cada una de las entradas del sistema, así como la matriz de autocorrelación de todos los datos del sistema.

En el caso de trabajar con señales reales, la solución sería simplemente un caso particular del ejemplo anterior en el que los complementarios desaparecen al estar sobre el espacio de las señales reales.

$$\vec{r}_{yx} = E[y \vec{x}] = E \begin{bmatrix} y[n]x[n] \\ y[n]x[n-1] \\ \vdots \\ y[n]x[n-N] \end{bmatrix} = \begin{bmatrix} r_{yx}(0) \\ r_{yx}(1) \\ \vdots \\ r_{yx}(N) \end{bmatrix}$$

$$\begin{aligned}\tilde{\mathbf{R}}_{xx} = E[\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T] &= E\begin{bmatrix} x[n]x[n] & x[n]x[n-1] & \cdots & x[n]x[n-N] \\ x[n-1]x[n] & x[n-1]x[n-1] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ x[n-N]x[n] & \cdots & \cdots & x[n-N]x[n-N] \end{bmatrix} = \\ &= \begin{bmatrix} r_{xx}[0] & r_{xx}[1] & \cdots & r_{xx}[N] \\ r_{xx}[1] & r_{xx}[0] & & \vdots \\ \vdots & & \ddots & \\ r_{xx}[N] & \cdots & & r_{xx}[0] \end{bmatrix}\end{aligned}$$

Una vez desarrollado el caso N dimensional, plantearemos el caso de un filtro de orden 1 (dos coeficientes) para acabar de profundizar en los conceptos presentado hasta el momento. La interpretación sobre un espacio bidimensional permitirá completar los conceptos necesarios para comprender el proceso de optimización lineal.

En el caso concreto de dos coeficientes y para el caso real, el vector de correlación cruzada y la matriz de autocorrelación corresponderían a las siguientes ecuaciones:

$$\begin{bmatrix} r_{yx}[0] \\ r_{yx}[1] \end{bmatrix} = \begin{bmatrix} r_{xx}[0] & r_{xx}[1] \\ r_{xx}[-1] & r_{xx}[0] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

$$\begin{bmatrix} E[y[n]x[n]] \\ E[y[n]x[n-1]] \end{bmatrix} = \begin{bmatrix} E[x[n]x[n]] & E[x[n]x[n-1]] \\ E[x[n-1]x[n]] & E[x[n-1]x[n-1]] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

En este momento conviene utilizar esta matriz simple para plantear un par de escenarios que nos llevarán a comprender en mayor profundidad los conceptos expuestos.

En un primer escenario, imaginaremos un sistema de dos coeficientes del estilo $\hat{y}[n] = h[0]x[n] + h[1]x[n-1]$. En este caso se trata de un sistema con una respuesta impulsional de dos coeficientes donde consideraremos que $x[n]$ y $x[n-1]$ son señales incorreladas, o lo que es lo mismo en los conceptos de espacios vectoriales utilizados, dos señales ortogonales.

En este caso concreto, el sistema de ecuaciones a resolver sería:

$$\begin{bmatrix} E[y[n]x[n]] \\ E[y[n]x[n-1]] \end{bmatrix} = \begin{bmatrix} E[x[n]x[n]] & 0 \\ 0 & E[x[n-1]x[n-1]] \end{bmatrix} \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

El sistema de ecuaciones en este caso presentaría una solución trivial donde los coeficientes serían:

$$h[0] = \frac{E[y[n]x[n]]}{E[x[n]x[n]]} \text{ y } h[1] = \frac{E[y[n]x[n-1]]}{E[x[n-1]x[n-1]]}$$

Si prestamos atención a los coeficientes vemos que corresponden exactamente al coeficiente de proyección observado en el caso de una variable. Es decir, el filtro ideal resultará de proyectar la información deseada sobre cada una de las variables de entrada de forma independiente y realizar la combinación lineal de las aportaciones de todas las variables de entrada implicadas. En el caso de señales de entrada ortogonales, no será necesario resolver ningún sistema de ecuaciones puesto que las diferentes señales de entrada no se interfieren entre sí, de forma que la solución pasa por proyectar sobre cada uno de los vectores de entrada y sumar las aportaciones. La principal conclusión de este ejemplo es que si tenemos un espacio de señal de entrada formado por una base ortogonal, los coeficientes del filtro se pueden calcular de forma directa mediante las proyecciones sobre cada uno de los elementos de entrada.

En el escenario más genérico donde las dos variables de entrada están correladas, será necesario desacoplar la información común que presentan ambas mediante la resolución del sistema, que pasará por aplicar la inversa de la matriz de correlación a fin de eliminar la información cruzada que existe en los diferentes vectores de entrada.

La solución en este caso sería:

$$\begin{bmatrix} E[x[n]x[n]] & E[x[n]x[n-1]] \\ E[x[n-1]x[n]] & E[x[n-1]x[n-1]] \end{bmatrix}^{-1} \begin{bmatrix} E[y[n]x[n]] \\ E[y[n]x[n-1]] \end{bmatrix} = \begin{bmatrix} h[0] \\ h[1] \end{bmatrix}$$

Siendo necesaria la resolución de la matriz inversa para desacoplar la información conjunta de las señales de entrada.

5.3. El principio de ortogonalidad para un filtro de orden N

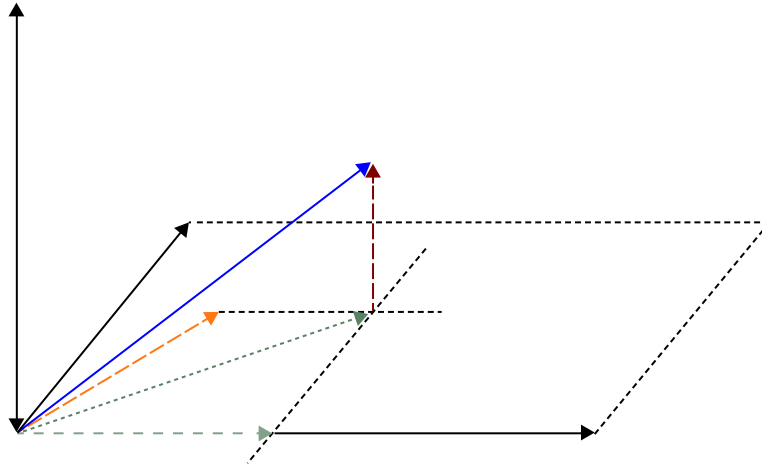
Tal y como hemos podido observar en los subapartados anteriores, el desarrollo matemático necesario para abordar la solución del filtro de Wiener requiere de la resolución de un problema de optimización aplicando derivadas vectoriales, que sin ser extremadamente complejas en su resolución, sí que entrañan una cierta complejidad y pueden inducir al error.

La interpretación geométrica presentada en el caso de dimensión 1 puede ser generalizada a dimensión N para llevarnos a un planteamiento más sencillo, directo e intuitivo del sistema de ecuaciones. En el caso de dimensión 2, nos encontramos con la necesidad de estimar un vector a partir de una base de dos componentes.

Tal y como se observa en la figura 8, imaginemos que pretendemos estimar el vector objetivo $y[n]$ a partir de una combinación lineal de $x[n]$ (verde claro) y $x[n-1]$ (amarillo). Sin necesidad de recurrir a una resolución compleja de un sistema de ecuaciones, y aplicando criterios puramente geométricos, vemos que el vector de error tendrá módulo mínimo cuando sea ortogonal al plano

formado por los vectores $x[n]$ y $x[n-1]$. Vemos igualmente que los dos vectores, al tener interferencia cruzada, se suman aplicando la regla del paralelogramo, que es una forma gráfica de eliminar la información cruzada que existe entre los dos vectores.

Figura 9



Ejemplo

Este ejemplo permite ampliar el principio de ortogonalidad a un espacio N -dimensional afirmando que la minimización del error cuadrático medio la obtendremos cuando el vector de error sea ortogonal a todos y cada uno de los vectores del espacio sobre el que se realiza la proyección, de tal forma que la solución la obtendremos cuando:

$$J = E[|y[n] - \hat{y}[n]|^2] = E\left[\left(y[n] - \sum_{k=0}^N x[n-k]h^*[k]\right)\left(y[n] - \sum_{k=0}^N x[n-k]h^*[k]\right)^*\right]$$

$$\frac{\partial J}{\partial h^*[i]} = -E\left[x[n-i]\left(y[n] - \sum_{k=0}^N x[n-k]h^*[k]\right)^*\right] = -E[x[n-i]e^*[n]] = 0$$

$$E[x[n-i]e^*[n]] = 0 \quad \forall i \in [0, 1, \dots, N]$$

En la última fila vemos que la solución del sistema la tendremos cuando el producto escalar entre el vector de error y los diferentes valores de la señal de entrada sea 0, o lo que es lo mismo, cuando el vector de error sea ortogonal a todas las componentes de la entrada del filtro ($x[n]$, $x[n-1]$, .., $x[n-K]$). La solución óptima la obtendremos cuando el vector de error sea ortogonal al vector de datos de entrada.

$$\left. \begin{aligned} \langle x[n]e[n] \rangle &= E[x[n]e^*[n]] = 0 \\ \langle x[n-1]e[n] \rangle &= E[x[n-1]e^*[n]] = 0 \\ &\vdots \\ \langle x[n-N]e[n] \rangle &= E[x[n-N]e^*[n]] = 0 \end{aligned} \right\}$$

Este criterio se podrá aplicar para cualquier tipo de filtrado lineal óptimo que se optimice según el criterio de los mínimos cuadrados, de forma que vemos que el principio de ortogonalidad representa una conceptualización extremadamente útil en aplicaciones de filtrado óptimo, y más adelante, en filtrado adaptativo.

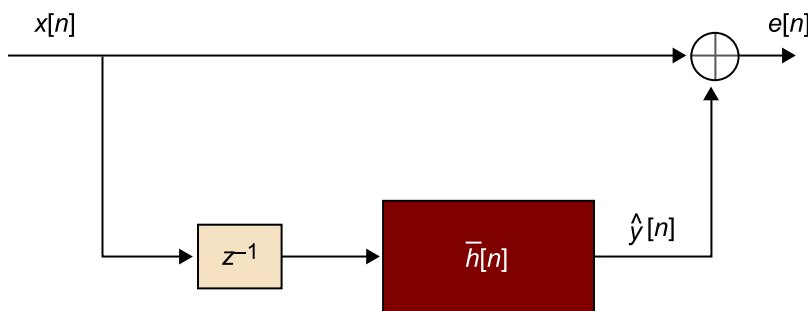
6. El predictor lineal

Dentro de las aplicaciones de procesamiento avanzado existe una de especial utilidad dentro del campo de las telecomunicaciones por su uso en técnicas de compresión de datos y aplicaciones de voz entre otras.

El predictor lineal es un caso particular de filtrado óptimo donde la señal objetivo es una muestra avanzada de la señal de entrada $y[n] = x[n+K]$, donde k será el orden del predictor. El predictor de orden 1 predice la muestra siguiente, el predictor de orden 2, $x[n+2]$ y así para cualquier orden de predicción. En la mayoría de aplicaciones prácticas, el uso del predictor suele centrarse en el predictor de orden 1, y, como se ha dicho, es muy útil en técnica de compresión de voz así como en técnicas de ecualización de canal.

En la siguiente figura observamos la representación de un predictor lineal.

Figura 8



A nivel de resolución de las ecuaciones de filtrado óptimo, la problemática del predictor lineal no aporta ninguna complejidad añadida respecto a las anteriores, más bien se trata de un caso particular del filtro de Wiener de gran utilidad en el ámbito del procesamiento.

La única diferencia es que el vector de correlación cruzada se transforma en un vector de autocorrelación desplazado de la siguiente forma:

$$E[y\vec{x}] = E \begin{bmatrix} y^*[n]x[n] \\ y^*[n]x[n-1] \\ \vdots \\ y^*[n]x[n-N] \end{bmatrix} = \begin{bmatrix} r_{yx}^*(0) \\ r_{yx}^*(1) \\ \vdots \\ r_{yx}^*(N) \end{bmatrix} = \begin{bmatrix} r_{xx}^*(k) \\ r_{xx}^*(k+1) \\ \vdots \\ r_{xx}^*(k+N) \end{bmatrix}$$

Con lo que el sistema de ecuaciones a resolver en el caso de un predictor sería el mismo sustituyendo el vector de correlación cruzada según la ecuación anterior.

7. Conclusiones

En este módulo hemos podido estudiar las principales características del filtrado lineal óptimo, sentando las bases para su resolución en aplicaciones prácticas, pero sobre todo, como introducción de los conceptos fundamentales del procesado adaptativo, dado que las características cambiantes de las aplicaciones reales nos llevarán en la mayoría de casos a implementaciones adaptativas.

La principal ventaja de los estimadores lineales respecto a otro tipo de estimadores más complejos es que para la resolución de los coeficientes óptimos del filtro tan solo es necesario el conocimiento de las funciones estadísticas de primer y segundo orden (medias y correlaciones). Dado que en muchas aplicaciones tan solo se dispone de una única realización del proceso, para poder hacer la estimación de las funciones estadísticas tendremos que recurrir al uso de las muestras temporales, es decir, calcular las medias y las correlaciones a partir de las muestras temporales del proceso. Recordemos que este aspecto es solo posible si el proceso es estacionario y ergódico. Será en este caso cuando resolveremos el problema utilizando un cálculo estadístico de la media y la correlación mediante las siguientes ecuaciones:

$$E[x] = \lim_{L \rightarrow \infty} \frac{1}{2L+1} \sum_{n=-L}^L x[n] \quad E[yx] = \lim_{L \rightarrow \infty} \frac{1}{2L+1} \sum_{n=-L}^L x[n]y[n]$$

En el caso de tener procesos limitados en el tiempo, como será habitual, restringiremos los límites del sumatorio a la ventana de datos de la que dispongamos, dado que esa será la mejor estimación posible que podremos realizar.

8. Ejemplos prácticos de uso del filtrado lineal óptimo

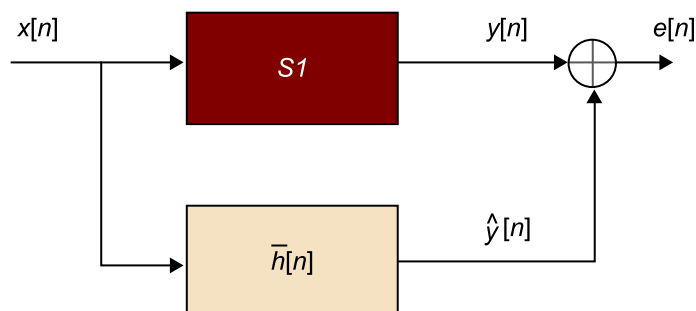
En este apartado, antes de finalizar el módulo, abordaremos algunos ejemplos típicos de aplicaciones lineales óptimas en el caso de filtros adaptativos a fin de ilustrar con estos ejemplos algunos de los conceptos más interesantes presentados en el módulo.

8.1. Identificación de un sistema

Supongamos que disponemos de un sistema $S1$ que es una caja negra de la que podemos conocer su respuesta. Es decir, si nosotros introducimos una señal de entrada, podemos medir cuál es la salida del sistema. El sistema en sí podría ser un equipo electromecánico, una sala acústica, un canal de comunicaciones o cualquier otro ejemplo que se nos pueda ocurrir.

El diagrama de bloques de diseño corresponderá al esquema que se muestra en la figura 10:

Figura 10



La idea es realizar una estimación de los coeficientes lineales que más se aproximan a la respuesta real del filtro. Si el filtro fuese un sistema lineal y el orden del estimador fuese lo suficientemente grande, la estimación sería prácticamente perfecta siempre y cuando se escogiesen apropiadamente las señales de test del sistema.

¿Qué tipo de señal de entrada sería conveniente utilizar para la implementación del filtro óptimo?

Si lo que nos interesa es realizar una estimación completa de la respuesta del filtro, necesitamos escoger como señal de entrada $x[n]$ una señal que sea rica en frecuencia, es decir, que sea capaz de estimular al sistema en toda la variedad posible de señales de entrada. Recordemos nuevamente que el filtro óptimo realizará la mejor estimación posible con la información de la que dispone, pero que eso no quiere decir que la estimación sea buena.

Ejemplo

A modo de ejemplo, si la señal excitación es una señal senoidal que tan solo excita una determinada frecuencia, el filtro de Wiener, por alto que sea el orden del filtro, tan solo podrá determinar la respuesta estimada a la frecuencia de la señal de entrada.

Teniendo en cuenta este aspecto, vemos que resulta importante en problemas de identificación de sistemas que la señal de entrada cubra las máximas dimensiones del espacio vectorial sobre el que se trabaja, y esto traducido al campo del procesado de señal sería que tenga componentes de todas las frecuencias posibles.

Ejemplo

Un ejemplo de señal de entrada sería la función delta discreta $\delta[n]$ que como bien sabemos tiene un espectro frecuencial plano, o un ruido blanco, que es capaz de cubrir todo el margen frecuencial. Si el sistema es un motor o una sala acústica, una delta no sería una función práctica dado que en el mundo analógico no la podríamos implementar, con lo que sería más conveniente trabajar con función tipo ruido blanco.

¿Cómo sabremos cuál es el orden apropiado del filtro?

Esta pregunta también resulta interesante, dado que *a priori* no tenemos ninguna información que nos indique cuál será el orden apropiado. Si por algún motivo conocemos este parámetro, procederemos a escoger el orden del filtro que se adapte al conocimiento que tengamos de la realidad.

Lo que ocurrirá si escogemos un orden menor será que el sistema no podrá adaptarse completamente y tendremos un error de estimación mayor y una identificación del sistema de peor calidad. Una buena solución es hacer diferentes evaluaciones incrementando el orden del filtro hasta que el error de estimación no decrezca.

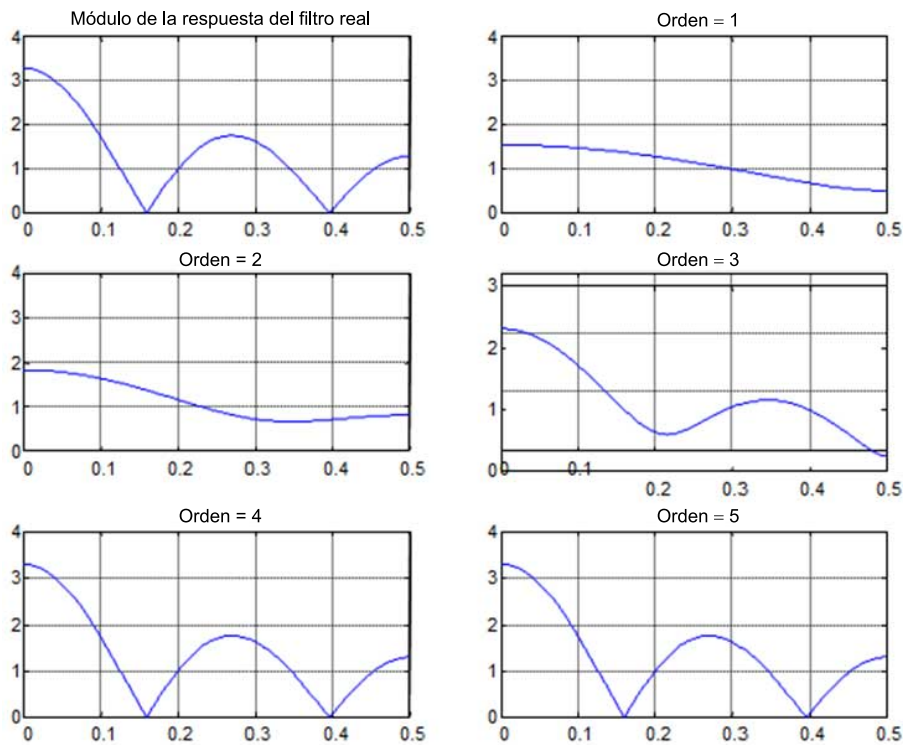
Si escogemos un orden superior al del sistema real, la estimación será buena pero los últimos coeficientes de la respuesta impulsional del filtro estimado serán cero. *A priori* esto no presenta un gran problema desde el punto de vista práctico, pero si desde el punto de vista computacional, dado que la resolución del filtro óptimo requiere el cálculo de la inversa de una matriz, una operación computacionalmente costosa y cuyo coste aumenta de forma exponencial. Si la implementación la estamos realizando en un DSP o una FPGA, esto puede suponer un problema para el hardware del sistema.

En la figura 11 se muestra la respuesta frecuencial estimada mediante un filtro de Wiener para un sistema cuya respuesta impulsional real es $h[n] = [1 \ 0,5 \ 0,3 \ 0,5 \ 1]$. En este caso se ha utilizado como señal de entrada ruido blanco gaussiano y se ha resuelto el filtro de Wiener mediante la estimación temporal de las funciones de correlación.

Se observa que en el caso del filtro de orden 1 (2 coeficientes) la estimación del sistema es bastante pobre, pero realiza una estimación de tipo paso-bajo a fin de aproximarse de la mejor forma posible a las bandas energéticas predominantes del sistema real. Este factor va ocurriendo a medida que aumentan

los coeficientes, donde vemos que la forma se va modulando hasta adaptarse completamente en el filtro de orden 4 (5 coeficientes). Si aumentamos todavía más el orden del filtro, vemos que ya no hay ninguna mejora en la estimación, puesto que el último coeficiente de la respuesta se hará cero.

Figura 11



Anexo

Espacios de Hilbert

Un espacio de Hilbert es una generalización de un espacio euclidiano que nos permite ampliar los conceptos de geometría euclidiana a otros tipos de vectores. El espacio de Hilbert está formado por un par $(V, \langle \cdot, \cdot \rangle)$ donde V es un espacio vectorial bien definido (finito o infinito dimensional) sobre el cuerpo de los números reales o complejos y $\langle \cdot, \cdot \rangle$ es un producto interior sobre V , con la condición adicional de que el espacio ha de ser métrico completo sobre la distancia inducida por el producto interno.

Que $\langle \cdot, \cdot \rangle$ sea un producto interno sobre V implica que $\langle \cdot, \cdot \rangle: V \times V \rightarrow K$ donde $K \in \{\mathbb{R}, \mathbb{C}\}$ que es el cuerpo subyacente sobre el que está definido el producto escalar. El operador $\langle \cdot, \cdot \rangle$ cumple asimismo las siguientes condiciones para todos los elementos $x, y, z \in V$ y para todos los escalares $a, b \in K$.

- $\langle x, y \rangle = \overline{\langle y, x \rangle}$ donde la barra implica conjugación compleja.
- $\langle ax + by, z \rangle = a \langle x, z \rangle + b \langle y, z \rangle$
- $\langle x, x \rangle \geq 0$, y la igualdad tan solo se cumple si, y solo si, $x = 0$.

Sobre este producto interno bien definido se induce una norma, que será el operador que servirá para poder medir distancias dentro del espacio vectorial, definida como $\|x\| = \sqrt{\langle x, x \rangle}$ que cumple las siguientes propiedades:

- $\|x\| = 0$ si, y solo si, $x = 0$.
- $\|ax\| = |a| \|x\|$
- $|\langle x, y \rangle| \leq \|x\| \|y\|$ (desigualdad de Cauchy-Schwarz).
- $\|x + y\| \leq \|x\| + \|y\|$ (desigualdad triangular).

Sin entrar en más profundidad en los detalles matemáticos expuestos, nos hemos de quedar con que los espacios de Hilbert permiten realizar mediciones y trabajar con distancias, según una norma bien definida que nos permitirá abordar problemas de optimización sobre espacios vectoriales de variables aleatorias.

Reglas de derivación con variable compleja y con funciones escalares de vectores complejos

Notación

Variable compleja: $x = x_R + jx_I \in \mathbb{C}$ donde $x_R = \operatorname{Re}[x] \in \mathfrak{R}$, $x_I = \operatorname{Im}[x] \in \mathfrak{I}$

Complejo conjugado: $x^* = x_R - jx_I \in \mathbb{C}$

Derivación de variable compleja	Ejemplos
$\frac{\partial}{\partial x} = \frac{1}{2} \left(\frac{\partial}{\partial x_R} - j \frac{\partial}{\partial x_I} \right)$	$\frac{\partial x}{\partial x} = \frac{1}{2} \left(\frac{\partial(x_R + jx_I)}{\partial x_R} - j \frac{\partial(x_R + jx_I)}{\partial x_I} \right) = \frac{1}{2} (1 - j \times j) = 1$ $\frac{\partial x^*}{\partial x} = \frac{1}{2} \left(\frac{\partial(x_R - jx_I)}{\partial x_R} - j \frac{\partial(x_R - jx_I)}{\partial x_I} \right) = \frac{1}{2} (1 - j \times (-j)) = 0$ $\frac{\partial x^* x}{\partial x} = \frac{dx^2}{dx} = \frac{\partial(x_R^2 + x_I^2)}{\partial x} = \frac{1}{2} \left(\frac{\partial(x_R^2 + x_I^2)}{\partial x_R} - j \frac{\partial(x_R^2 + x_I^2)}{\partial x_I} \right) =$ $= \frac{1}{2} (2x_R - j \times (2x_I)) = x^*$
$\frac{\partial}{\partial x^*} = \frac{1}{2} \left(\frac{\partial}{\partial x_R} + j \frac{\partial}{\partial x_I} \right)$	$\frac{\partial x}{\partial x^*} = \frac{1}{2} \left(\frac{\partial(x_R + jx_I)}{\partial x_R} + j \frac{\partial(x_R + jx_I)}{\partial x_I} \right) = \frac{1}{2} (1 + j \times j) = 0$ $\frac{\partial x^*}{\partial x^*} = \frac{1}{2} \left(\frac{\partial(x_R - jx_I)}{\partial x_R} + j \frac{\partial(x_R - jx_I)}{\partial x_I} \right) = \frac{1}{2} (1 + j \times (-j)) = 1$ $\frac{\partial x^2}{\partial x^*} = \frac{\partial(x_R^2 + j2x_I x_R - x_I^2)}{\partial x^*} = \frac{1}{2} (2x_R + j2x_I + j \times (2jx_R - 2x_I)) = 0$ $\frac{\partial x^* x}{\partial x^*} = \frac{dx^2}{dx^*} = \frac{\partial(x_R^2 + x_I^2)}{\partial x^*} = \frac{1}{2} \left(\frac{\partial(x_R^2 + x_I^2)}{\partial x_R} + j \frac{\partial(x_R^2 + x_I^2)}{\partial x_I} \right) =$ $= \frac{1}{2} (2x_R + j \times (2x_I)) = x$

Notación

Variable vector complejo: $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_N]^T \in \mathbb{C}^{N \times 1}$ vector columna, donde $x_k \in \mathbb{C}$, $k = 1, 2, \dots, N$

Constante vector complejo: $\mathbf{a} = [a_1 \ a_2 \ \dots \ a_N]^T \in \mathbb{C}^{N \times 1}$ vector columna, donde $a_k \in \mathbb{C}$, $k = 1, 2, \dots, N$

$\mathbf{b} = [b_1 \ b_2 \ \dots \ b_N]^T \in \mathbb{C}^{N \times 1}$ vector columna, donde $b_k \in \mathbb{C}$, $k = 1, 2, \dots, N$

Constante matriz compleja: $\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1N} \\ r_{21} & r_{22} & & r_{2N} \\ \vdots & & \ddots & \vdots \\ r_{N1} & r_{N2} & \dots & r_{NN} \end{bmatrix} \in \mathbb{C}^{N \times N}$ matriz cuadrada, donde $r_{ij} \in \mathbb{C}$, $i = 1, 2, \dots, N$; $j = 1, 2, \dots, N$

Operador hermítico (transposición+conjugación):

$$\mathbf{x}^H = (\mathbf{x}^T)^* = (\mathbf{x}^*)^T = [x_1^* \ x_2^* \ \dots \ x_N^*]^T \in \mathbb{C}^{N \times 1}$$

Derivación vectorial (gradiente)	Ejemplos
$\frac{\partial}{\partial \mathbf{x}} = \left[\frac{\partial}{\partial x_1} \ \frac{\partial}{\partial x_2} \ \dots \ \frac{\partial}{\partial x_N} \right]^T$	$\frac{\partial(2 + x_1 x_2 - x_1^2)}{\partial \mathbf{x}} = \begin{bmatrix} \partial(2 + x_1 x_2 - x_1^2) / \partial x_1 \\ \partial(2 + x_1 x_2 - x_1^2) / \partial x_2 \end{bmatrix} = \begin{bmatrix} x_2 - 2x_1 \\ x_1 \end{bmatrix}$

Derivación vectorial (gradiente)	Ejemplos
$\frac{\partial}{\partial \mathbf{x}^*} = \left[\frac{\partial}{\partial x_1^*} \quad \frac{\partial}{\partial x_2^*} \quad \dots \quad \frac{\partial}{\partial x_N^*} \right]^T$	$\frac{\partial(2 + x_1^* x_2 - x_1^2)}{\partial \mathbf{x}^*} = \begin{bmatrix} \partial(2 + x_1^* x_2 - x_1^2) / \partial x_1^* \\ \partial(2 + x_1^* x_2 - x_1^2) / \partial x_2^* \end{bmatrix} = \begin{bmatrix} x_2 \\ 0 \end{bmatrix}$
$\frac{\partial(a_1)}{\partial \mathbf{x}} = \mathbf{0}$	$\frac{\partial(1+j)}{\partial \mathbf{x}} = \begin{bmatrix} \partial(1+j) / \partial x_1 \\ \partial(1+j) / \partial x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}$
$\frac{\partial(a_1)}{\partial \mathbf{x}^*} = \mathbf{0}$	$\frac{\partial(1+j)}{\partial \mathbf{x}^*} = \begin{bmatrix} \partial(1+j) / \partial x_1^* \\ \partial(1+j) / \partial x_2^* \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}$
$\frac{\partial(\mathbf{a}^H \mathbf{x})}{\partial \mathbf{x}} = \mathbf{a}^*$	$\frac{\partial \begin{bmatrix} 1+j & 2-j \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}{\partial \mathbf{x}} = \begin{bmatrix} \partial((1+j)x_1 + (2-j)x_2) / \partial x_1 \\ \partial((1+j)x_1 + (2-j)x_2) / \partial x_2 \end{bmatrix} = \begin{bmatrix} 1+j \\ 2-j \end{bmatrix}$
$\frac{\partial(\mathbf{x}^H \mathbf{a})}{\partial \mathbf{x}^*} = \mathbf{a}$	$\frac{\partial \begin{bmatrix} x_1^* & x_2^* \end{bmatrix} \begin{bmatrix} 1+j \\ 2-j \end{bmatrix}}{\partial \mathbf{x}^*} = \begin{bmatrix} \partial((1+j)x_1^* + (2-j)x_2^*) / \partial x_1^* \\ \partial((1+j)x_1^* + (2-j)x_2^*) / \partial x_2^* \end{bmatrix} = \begin{bmatrix} 1+j \\ 2-j \end{bmatrix}$
$\frac{\partial(\mathbf{a}^H \mathbf{x})}{\partial \mathbf{x}^*} = \frac{\partial(\mathbf{x}^H \mathbf{a})}{\partial \mathbf{x}} = \mathbf{0}$	$\frac{\partial \begin{bmatrix} x_1^* & x_2^* \end{bmatrix} \begin{bmatrix} 1+j \\ 2-j \end{bmatrix}}{\partial \mathbf{x}} = \begin{bmatrix} \partial((1+j)x_1^* + (2-j)x_2^*) / \partial x_1 \\ \partial((1+j)x_1^* + (2-j)x_2^*) / \partial x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} = \mathbf{0}$
$\frac{\partial(\mathbf{x}^H \mathbf{R} \mathbf{x})}{\partial \mathbf{x}} = (\mathbf{x}^H \mathbf{R})^T$	$\frac{\partial \begin{bmatrix} x_1^* & x_2^* \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}{\partial \mathbf{x}} = \frac{\partial(r_{11}x_1^*x_1 + r_{12}x_1^*x_2 + r_{21}x_2^*x_1 + r_{22}x_2^*x_2)}{\partial \mathbf{x}} =$ $= \begin{bmatrix} \partial / \partial x_1 \\ \partial / \partial x_2 \end{bmatrix} \begin{bmatrix} r_{11}x_1^* + r_{21}x_2^* \\ r_{12}x_1^* + r_{22}x_2^* \end{bmatrix} = \begin{bmatrix} r_{11} & r_{21} \\ r_{12} & r_{22} \end{bmatrix} \begin{bmatrix} x_1^* \\ x_2^* \end{bmatrix} = \mathbf{R}^T \mathbf{x}^* = (\mathbf{x}^H \mathbf{R})^T$
$\frac{\partial(\mathbf{x}^H \mathbf{R} \mathbf{x})}{\partial \mathbf{x}^*} = \mathbf{R} \mathbf{x}$	$\frac{\partial \begin{bmatrix} x_1^* & x_2^* \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}}{\partial \mathbf{x}^*} = \frac{\partial(r_{11}x_1^*x_1 + r_{12}x_1^*x_2 + r_{21}x_2^*x_1 + r_{22}x_2^*x_2)}{\partial \mathbf{x}^*} =$ $= \begin{bmatrix} \partial / \partial x_1^* \\ \partial / \partial x_2^* \end{bmatrix} \begin{bmatrix} r_{11}x_1 + r_{12}x_2 \\ r_{21}x_1 + r_{22}x_2 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \mathbf{R} \mathbf{x}$

Demostración de la aplicación de las reglas de derivación en la obtención de la solución óptima del filtro de Wiener

En el apartado 5.2. “Filtro lineal óptimo de orden N” de este módulo partimos de la siguiente ecuación:

$$E[d^2]_{MIN} = E[y y^H] - \vec{h}^H E[\vec{x} y^H] - E[y \vec{x}^H] \vec{h} + \vec{h}^H E[\vec{x} \vec{x}^H] \vec{h} \Big|_{MIN}$$

Lo que se pretende es hallar el valor de \vec{h}^* que minimice esta función de coste, por tanto, lo que hay que hacer es:

$$\frac{\partial E[d^2]}{\partial \vec{h}^*} = \underbrace{\frac{E[y y^H]}{\partial \vec{h}^*}}_{(1)} - \underbrace{\frac{\vec{h}^H E[\vec{x} y^H]}{\partial \vec{h}^*}}_{(2)} - \underbrace{\frac{E[y \vec{x}^H] \vec{h}}{\partial \vec{h}^*}}_{(3)} + \underbrace{\frac{\vec{h}^H E[\vec{x} \vec{x}^H] \vec{h}}{\partial \vec{h}^*}}_{(4)} = 0$$

Vemos las reglas de derivación que hay que aplicar en cada caso:

$$\begin{aligned}
 (1) \quad & \rightarrow \left\{ \frac{\partial(a_1)}{\partial \mathbf{x}^*} = \mathbf{0} \right\} \rightarrow \frac{E[y y^H]}{\partial \vec{h}} = 0 \\
 (2) \quad & \rightarrow \left\{ \frac{\partial(\mathbf{x}^H \mathbf{a})}{\partial \mathbf{x}^*} = \mathbf{a} \right\} \rightarrow \frac{\vec{h}^H E[\vec{x} y^H]}{\partial \vec{h}} = E[\vec{x} y^H] \\
 (3) \quad & \rightarrow \left\{ \frac{\partial(\mathbf{a}^H \mathbf{x})}{\partial \mathbf{x}^*} = \frac{\partial(\mathbf{x}^H \mathbf{a})}{\partial \mathbf{x}} = \mathbf{0} \right\} \rightarrow \frac{E[y \vec{x}^H] \vec{h}}{\partial \vec{h}} = 0 \\
 (4) \quad & \rightarrow \left\{ \frac{\partial(\mathbf{x}^H \mathbf{R} \mathbf{x})}{\partial \mathbf{x}^*} = \mathbf{R} \mathbf{x} \right\} \rightarrow \frac{\vec{h}^H E[\vec{x} \vec{x}^H] \vec{h}}{\partial \vec{h}} = E[\vec{x} \vec{x}^H] \vec{h}
 \end{aligned}$$

Por tanto, se llega al resultado final que aparece en el mismo apartado:

$$\frac{\partial E[d^2]}{\partial \vec{h}} = -E[\vec{x} y^H] + E[\vec{x} \vec{x}^H] \vec{h} = 0$$