



## INTRODUCCION

El concepto de filtro adaptativo, sugiere el de un dispositivo que intenta modelizar la relación entre señales en tiempo real de forma iterativa. Se diferencia de los filtros digitales en que éstos últimos tienen coeficientes invariantes en el tiempo, mientras que un filtro adaptativo puede cambiar su forma de comportarse, es decir, pueden cambiar sus coeficientes de acuerdo con un algoritmo adaptivo. De hecho, no se conocen los coeficientes del filtro cuando se diseña, debido a que estos coeficientes son calculados cuando el filtro se implementa y se reajustan automáticamente en cada iteración mientras dura su fase de aprendizaje.

El hecho de que estos filtros no sean invariantes temporales y que tampoco sean lineales, hace que su estudio sea más complejo que el de un filtro digital, ya que no se pueden aplicar, salvo en un par de excepciones, las transformaciones en frecuencia, dominio Z, etc.

El presente texto trata en forma introductoria el tema de filtros adaptivos. En principio se realiza una breve introducción sobre procesos aleatorios a tiempo continuo y estimación de parámetros. Seguidamente se define al filtro de Wiener, como método para la estimación estadística de la respuesta al impulso unitario de un sistema. Por último se aborda el tema de filtros adaptivos, haciendo referencia a los algoritmos adaptivos de máxima pendiente, el algoritmo LMS, el algoritmo RLS y filtro de Kalman.

## PROCESOS ALEATORIOS A TIEMPO CONTINUO

Cada vez que se realiza la observación o el sensado de algún evento temporal, se obtendrán señales eléctricas distintas cada vez que se quiera, debido a la naturaleza o características aleatorias implícitas que estas presentan. En el esquema de la Figura 1, se muestra el modelo típico de un sistema que puede ser lineal e invariante en el tiempo y de las diferentes señales temporales que interactúan. Por tratarse de señales aleatorias, estas son parte de las realizaciones de procesos aleatorios vinculados al sistema. De esta manera se cuenta con un proceso aleatorio de entrada al sistema  $\{X\}$ , que da lugar al proceso aleatorio de salida del sistema  $\{Y\}$ . Las realizaciones del proceso  $\{Y\}$ , estarán distorsionadas por las realizaciones (señal ruido) del proceso  $\{N\}$ , dando como resultado las realizaciones del proceso aleatorio  $\{Z\}$ .

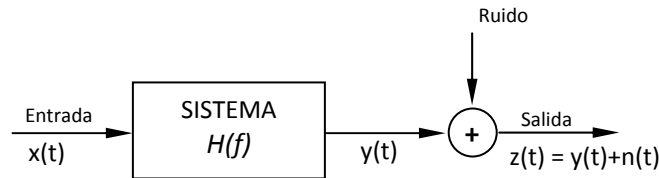


Figura 1: Esquema del sistema de análisis, donde la entrada del sistema corresponde a las realizaciones o señales eléctricas del transductor emisor del proceso  $\{X\}$ , la salida corresponde a las realizaciones o señales eléctricas del transductor receptor del proceso  $\{Z\}$  y  $H(f)$  la función de transferencia del sistema.

Al representar todas las realizaciones de un proceso, tal como indica la Figura 2, se observa a un conjunto de infinitas funciones del tiempo. Estas funciones están asignadas a un evento o punto muestral de un suceso o experimento aleatorio, de allí su designación como funciones muestrales. Si las mismas se estudian de manera separada, esto es, considerar solo un evento en particular  $\xi$ , se estaría ante la presencia de una onda de potencia finita, pudiéndose caracterizar a través del valor medio de la señal  $m(\xi)$ , del producto escalar de la función con su versión desplazada en diversos tiempos (lags) o función de autocorrelación  $R_{xx}(\tau, \xi)$  y con la potencia de la señal como autocorrelación en el origen  $P_x(\xi)$ , respectivamente:

$$\begin{aligned} m(\xi) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x(t, \xi) dt \\ R_{xx}(\tau, \xi) &= \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x^*(t, \xi) x(t + \tau, \xi) dt \\ P_x(\xi) &= R_{xx}(0, \xi) \end{aligned} \quad (1)$$

Al producto escalar de la función o función de autocorrelación se lo entiende como una medida del grado de similitud de la señal con su versión desplazada en el tiempo de  $\tau$ . El desplazamiento en el tiempo  $\tau$  es arbitrario y el significado de la función autocorrelación, es dar información de la variación en el tiempo de la señal en un sentido promediado (Hsu, 1997; Papoulis and Pillai, 2002; Manolakis, et al., 2005).

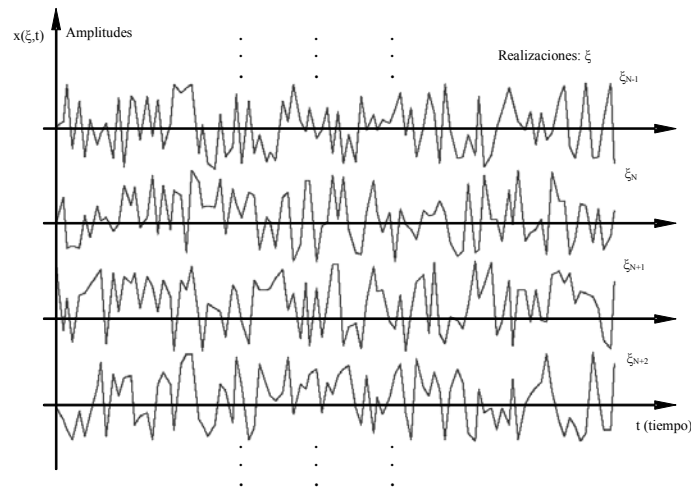


Figura 2: Proceso aleatorio representado por sus realizaciones (ensamble), resultado de distintos experimentos. Cada una de las realizaciones lleva asociada una forma de onda en el tiempo.

Cuando se considera del proceso aleatorio los distintos valores que toma cada evento para un tiempo fijo, se estará ante la presencia de una variable aleatoria (VA) caracterizada por su correspondiente función densidad de probabilidad o PDF (por sus siglas en ingles de Probability Density Function),  $f(x, t)$ , lo que permite determinar a lo parámetros estadísticos o momentos de orden  $n$  de la VA. En algunos casos conociendo estos parámetros estadísticos, se puede llegar a caracterizar la VA como por ejemplo las de distribución Gaussiana (Hsu, 1997; Papoulis and Pillai, 2002). Pero la caracterización del proceso es todavía escasa, ya que hasta ahora no se tiene una caracterización en donde se contemple la evolución conjunta tanto del tiempo como en cada uno de los eventos. Para ello es necesario disponer de una función densidad de probabilidad, teniendo en cuenta que un proceso aleatorio estará caracterizado con las PDF de las variables aleatorias que lo componen, por lo tanto, dado un conjunto de variables aleatorias obtenidas para instantes de tiempo  $t_1, t_2, \dots, t_n$  fijos; su caracterización probabilística puede representarse por una colección de PDF conjuntas:

$$\begin{aligned} &f(x_1, t_1) \\ &f(x_1, t_1; x_2, t_2) \\ &\vdots \\ &f(x_1, t_1; x_2, t_2; \dots; x_n, t_n) \end{aligned} \quad (2)$$

en donde la función PDF conjunta es la probabilidad de que el proceso tome valores en el intervalo  $(x_1, x_1+dx_1)$  en el tiempo  $t_1$ , en  $(x_2, x_2+dx_2)$  en el tiempo  $t_2$ , ..., en  $(x_n, x_n+dx_n)$  para en tiempo  $t_n$ . Como en el



caso de una VA, conociendo del proceso la PDF conjunta, es posible determinar sus parámetros estadísticos como por ejemplo la función media y los momentos de orden  $n$ , la función autocorrelación y la función covarianza, respectivamente:

$$\begin{aligned}\mu(t) &= E\{x(t)\} = \int_{-\infty}^{\infty} x \cdot f(x, t) dx \\ m_n(t) &= E\{x^n(t)\} = \int_{-\infty}^{\infty} x^n \cdot f(x, t) dx \\ R_X(t, \tau) &= E\{x^*(t)x(t+\tau)\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1^* x_2 \cdot f(x_1, t_1; x_2, t+\tau) dx_1 dx_2 \\ K_X(t, \tau) &= E\{[x(t) - \mu(t)][x(t+\tau) - \mu(t+\tau)]\} = R_X(t, \tau) - \mu(t)\mu(t+\tau)\end{aligned}\tag{3}$$

en donde  $E\{\}$ , representa el valor esperado. Del momento de orden 2 ( $n = 2$ ) se obtiene la función “error medio cuadrático” y de la función de covarianza, si  $t = \tau$  obtenemos la función de “varianza”.

## Estacionariedad y Ergodicidad

Existen algunos fenómenos que dan lugar a procesos aleatorios, cuyas propiedades o parámetros estadísticos se mantienen constantes en el tiempo, en ese caso, los procesos aleatorios son **estacionarios**. En este tipo de procesos, cuanto más abarque la independencia del tiempo sobre los parámetros estadísticos, mas estricto es el criterio de estacionariedad. De esta forma, se dice que un proceso es estacionario “en sentido amplio” de segundo orden, si su función media es independiente del tiempo y su función de autocorrelación depende solo del incremento temporal  $\tau$ . Además, si la PDF conjunta de segundo orden, es independiente de las traslaciones en el tiempo, se dice que el proceso aleatorio es estacionario “en sentido estricto” de segundo orden (Hsu, 1997; Papoulis and Pillai, 2002), por lo tanto se verifica lo siguiente:

$$\begin{aligned}\mu(t) &= E\{x(t)\} = \mu && \Rightarrow cte. \forall t \\ R_X(t, \tau) &= E\{x^*(t)x(t+\tau)\} = R_X(\tau) && \forall t \\ f(x_1, t; x_2, t+\tau) &= f(x_1; x_2, \tau) && \forall t\end{aligned}\tag{4}$$



Se entiende como la densidad espectral del proceso, a la potencia de todas las realizaciones, en promedio, respecto de un ancho de banda infinitesimal alrededor de cada frecuencia (Hsu, 1997). En el caso de un proceso estacionario, se verifica que su área es la potencia del proceso o autocorrelación en  $\tau$  igual a cero. Cuando el proceso es estacionario la densidad espectral de potencia será igual a la transformada de Fourier de la autocorrelación, esto es:

$$S_X(\omega) = \int_{-\infty}^{\infty} R_X(\tau) \cdot e^{-j\omega\tau} d\tau \quad (5)$$

Por lo que la función de auto correlación, se puede obtener a través de la transformada inversa de Fourier de la densidad espectral de potencia:

$$R_X(\tau) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_X \cdot e^{j\omega\tau} d\omega \quad (6)$$

Estas relaciones son posibles ya que se consideran a los procesos estacionarios como continuos, diferenciales e integrables en sentido medio cuadrático; a través del “Teorema de Wiener-Khinchin” (Papoulis and Pillai, 2002).

Para casos en que es necesario conocer la dependencia estadística entre dos procesos  $\{X\}$  e  $\{Y\}$ , o el grado de parecido de los mismos, se lo hace a través de la función correlación cruzada:

$$R_{XY}(t, \tau) = E\{x^*(t) \cdot y(t + \tau)\} \quad (7)$$

En el caso de que ambos procesos sean estacionarios, la función correlación cruzada será independiente del tiempo y tendrá las mismas consideraciones que la función autocorrelación, respecto de la continuidad, diferenciación e integración. En este caso su transformada de Fourier será la densidad espectral de potencia cruzada:



$$S_{XY} = \int_{-\infty}^{\infty} R_{XY}(\tau) \cdot e^{-j\omega\tau} d\tau \quad (8)$$

Al igual que en la función de autocorrelación, aplicando la transformada inversa de Fourier de la densidad espectral de potencia cruzada se obtiene la función de correlación cruzada.

A menudo cuando se trata de procesos estacionarios, los promedios en el ensamble pueden ser reemplazados con éxito por los promedios temporales de una muestra. A tales procesos se los llama **ergódicos**. Esta concepción artificial de los procesos estacionarios obedece a una limitación que se presenta en la práctica cuando no se dispone de más de una realización del proceso o de la PDF conjunta, en cuyo caso a partir de una realización, se pretende entender o determinar los parámetros estadísticos de un proceso aleatorio estacionario. En ese caso, se prueba que un proceso aleatorio estacionario de segundo orden ergódico presenta las siguientes igualdades:

$$\begin{aligned} \mu &= E\{x(t)\} = m = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x(t) dt \\ R_x(\tau) &= E\{x^*(t)x(t+\tau)\} = R_{xx}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \int_T x^*(t)x(t+\tau) dt \end{aligned} \quad (9)$$

## Procesos Aleatorios y Sistemas Lineales

Cuando las realizaciones de un proceso  $\{X\}$ , se aplican a un sistema como el de la Figura 1, con características de linealidad e invarianza en el tiempo, se obtiene a su salida otro proceso aleatorio denominado  $\{Y\}$ , cuyas realizaciones están relacionadas mediante la respuesta impulsional del sistema  $h(t)$  a través de la integral de convolución de la ecuación (10). En el caso de procesos estacionarios los parámetros estadísticos de los procesos están relacionados a través de la misma propiedad, por lo que la función media del proceso de salida del sistema estará dada por (11).

$$y(t) = \int_{-\infty}^{\infty} h(\zeta) \cdot x(t - \zeta) d\zeta \quad (10)$$

$$E\{y(t)\} = \int_{-\infty}^{\infty} h(\zeta) \cdot E\{x(t - \zeta)\} d\zeta = E\{x\} \cdot \int_{-\infty}^{\infty} h(\zeta) d\zeta = \mu_x \cdot H(0) = \mu_y \quad (11)$$



En la expresión (11) el termino  $H(0)$ , corresponde a la respuesta en frecuencia  $H(\omega)$  a frecuencia cero.

El parecido estadístico o correlación del proceso de salida con el proceso de entrada, será dado por la siguiente expresión:

$$R_{XY}(\tau) = \int_{-\infty}^{\infty} h(\zeta) \cdot R_X(\tau - \zeta) d\zeta = R_X(\tau) * h(\tau) \quad (12)$$

La expresión de la densidad espectral de potencia puede obtenerse de la siguiente manera:

$$S_{XY} = H(\omega) \cdot S_X(\omega) \quad (13)$$

De la misma forma, puede obtenerse la correlación cruzada entre los procesos  $\{Y\}$  y  $\{X\}$ , junto a la autocorrelación del proceso de salida de la siguiente manera:

$$\begin{aligned} R_{YX}(\tau) &= h^*(-\tau) * R(\tau) = R_{XY}(-\tau) \\ S_{YX}(\omega) &= H^*(\omega) \cdot S_X(\omega) = S_{XY}^*(\omega) \end{aligned} \quad (14)$$

Respecto del proceso de salida  $\{Y\}$ , la función de autocorrelación del proceso será función de la autocorrelación del proceso de entrada. Trasladando esta relación, a sus respectivas densidades espectrales de potencia, se tiene:

$$\begin{aligned} R_Y(\tau) &= h(\tau) * h^*(-\tau) * R_X(\tau) \\ S_Y(\omega) &= |H(\omega)|^2 S_X(\omega); \quad S_X(\omega) \in \Re \end{aligned} \quad (15)$$

Cuando las realizaciones del proceso de salida, se ven distorsionadas por una señal ruido como en la situación que se grafica en la Figura 1, las relaciones entre la entrada del sistema y la nueva salida que este tiene, van a ser afectadas por la señal ruido. Si esta señal es independiente de la señal de entrada, lo



importante a determinar es el grado de afectación de las realizaciones del proceso de salida, en términos de las densidades espectrales de los procesos. La manera de hacerlo, es observando aquellos contenidos de frecuencia que se generan a la salida del sistema y que no son producto de la relación entre la entrada y la salida en el sistema lineal. Esta idea es la que sugiere la definición de coherencia espectral (Manolakis et al., 2005), donde su modulo se expresa de la siguiente manera:

$$|\gamma(\omega)|^2 = \frac{S_{XY}(\omega) \cdot S_{YX}(\omega)}{S_X(\omega) S_Y(\omega)} \quad (16)$$

En el caso que la salida no se encuentre distorsionada, para cualquier  $H(\omega)$  lineal, la coherencia espectral es la unidad en todo el margen de frecuencia. Por otro lado, ante la presencia de ruido blanco no correlado con la entrada, la densidad espectral de potencia cruzada entre el ruido y la entrada  $S_{NX}(\omega)$ , la densidad espectral de potencia del ruido  $S_N(\omega)$  y la densidad espectral de potencia de la salida  $S_Y(\omega)$ , serán resueltos de la siguiente manera:

$$\begin{aligned} S_{NX}(\omega) &= 0 \\ S_N(\omega) &= F\{R_N(\tau)\} = F\{N_0 \cdot \delta(\tau)\} = N_0 \\ S_Y(\omega) &= S_X(\omega) |H(\omega)|^2 + S_N(\omega) \end{aligned} \quad (17)$$

en donde  $R_N(\tau)$  es la función de autocorrelación del ruido, igual a una delta de Dirac de peso  $N_0$ . Sustituyendo entonces las expresiones de (17) en la ecuación (16), se obtiene la siguiente expresión:

$$|\gamma(\omega)|^2 = \frac{1}{1 + \frac{S_N(\omega)}{|H(\omega)|^2 S_X(\omega)}} = \frac{1}{1 + \frac{1}{SNR(\omega)}} = \frac{SNR(\omega)}{SNR(\omega) + 1} \quad (18)$$

en donde  $SNR(\omega)$ , se conoce como relación señal-ruido del sistema en términos de la frecuencia  $\omega$ . Se puede observar en (18), como influye en el valor de la coherencia espectral la relación  $SNR$ , haciendo que esta decaiga por debajo de la unidad cuando esta relación es pequeña. En otras palabras, decir que





el contenido en una banda de frecuencia de una señal es coherente con el de otra señal en la misma banda, implica que una de las dos se puede obtener de la otra mediante una transformación lineal, es por ello que en la práctica se busca que la relación  $SNR$  sea lo mas grande posible.

## Estimación de Parámetros y Calidad de un Estimador

El concepto de ergodicidad es útil, cuando se pretende encontrar los parámetros estadísticos o caracterizar a un proceso estacionario con solo una realización del proceso. Pero la concepción de ergodicidad es limitada ya que las señales son de duración finita y las representaciones de los parámetros estadísticos a partir de los promedios temporales variarán entre las realizaciones o señales que se dispongan. Por otro lado, dado un proceso aleatorio como entrada de un sistema, puede ocurrir que el proceso no sea estacionario o que el sistema no sea lineal e invariante en el tiempo. Ante estas situaciones, utilizando una o pocas realizaciones del proceso aleatorio, solo se estará logrando la **estimación** de los parámetros estadísticos de un proceso aleatorio (Manolakis et al., 2005).

Dado un proceso aleatorio  $\{X\}$ , la estimación de un parámetro  $\vartheta$  a partir de una realización  $x(t, \xi)$  es equivalente entonces, a plantear cual es la probabilidad de que el valor correcto coincida con un valor dado de  $\vartheta$ . De esta manera el parámetro a estimar es una VA que este relacionada con los datos observados, caracterizada por su PDF condicional  $f(\theta / x(t, \xi))$ . Por lo tanto, el parámetro óptimo será aquel que maximiza la probabilidad condicional (Hsu, 1997; Papoulis and Pillai, 2002; Manolakis, et al., 2005):

$$\theta_{MAP} = \arg \max_{\theta_0} [f(\theta / x(t, \xi)) d\theta] = \arg \max_{\theta_0} [\Pr(\theta < \Theta < \theta + d\theta / x(t, \xi))] \quad (19)$$

a esta expresión se la denomina máximo a posteriori o MAP (por sus siglas en ingles de Maximum a Posteriori ). Este es un criterio óptimo para la estimación del parámetro  $\vartheta$ , pero acarrea la dificultad de necesitar de una función PDF condicional que muchas veces no se conoce o es difícil en determinar. Utilizando el teorema de Bayes para representar dicha probabilidad, se obtiene la expresión (20), en donde el denominador de esta expresión no participa del proceso de maximización de la probabilidad por lo que puede descartarse.

$$f(\theta / x(t, \xi)) = \frac{f(x(t, \xi) / \theta) \cdot f(\theta)}{f(x(t, \xi))} \propto f(x(t, \xi) / \theta) \cdot f(\theta) \quad (20)$$



El criterio MAP queda resuelto entonces, por la maximización de la PDF condicional de los datos condicionada al parámetro  $\theta$ , multiplicada por la PDF del parámetro  $\theta$ . La PDF condicional de los datos condicionada al parámetro  $\theta$ , se la conoce como verosimilitud (likelihood) y la PDF del parámetro  $\theta$ , como una información “a priori” del parámetro, respecto del conocimiento de su PDF. Cuando la PDF del parámetro  $\theta$  es uniforme o se tiene el máximo desconocimiento o la mínima información de su distribución, el criterio MAP se logra maximizando la verosimilitud o ML (por sus siglas en ingles de Maximum Likelihood), tal como se indica en la siguiente expresión:

$$\theta_{ML} = \arg \max_{\theta_0} [f(x(t, \xi) / \theta)] \quad (21)$$

Este criterio es más utilizado que el criterio MAP, ya que es menos compleja la determinación de su PDF, sin embargo a diferencia del MAP, este criterio es subóptimo debido a la poca fiabilidad que se tiene del conocimiento de la PDF del parámetro  $\theta$ .

Otra manera de realizar la estimación consiste en estimar el parámetro que menos se desvía del valor correcto, en términos del error medio cuadrático. El criterio se conoce como MSE (por sus siglas en ingles de Mean Square Error) dado por la siguiente expresión:

$$\theta_{MSE} = \arg \min_{\theta_0} [E\{(\theta - \theta_0)^2 / x(t, \xi)\}] = \arg \min_{\theta_0} \left[ \int_{-\infty}^{\infty} (\theta - \theta_0)^2 f(\theta / x(t, \xi)) d\theta \right] \quad (22)$$

Se demuestra de (22) que el criterio MSE es la función media condicional del parámetro  $\theta$  condicionada a los datos, ya que esta función, es la que anula la derivada del error medio cuadrático respecto del parámetro correcto, esto es:

$$\frac{\partial E\{(\theta - \theta_0)^2 / x(t, \xi)\}}{\partial \theta_0} = \theta_0 - \int_{-\infty}^{\infty} \theta \cdot f(\theta / x(t, \xi)) d\theta = 0 \quad (23)$$

$$\theta_{MSE} = \int_{-\infty}^{\infty} \theta \cdot f(\theta / x(t, \xi)) d\theta$$



Por lo tanto, cuando el valor de la media de la PDF condicional coincide con el valor máximo de su distribución, el criterio MSE coincidirá con el MAP, siendo en ese caso un criterio óptimo. Existen muchas funciones de distribución de probabilidad, incluida la Gaussiana, donde su función media coincide con el valor máximo de su distribución.

Existe una manera de evaluar la calidad de los estimadores (MAP, ML, MSE o la de cualquier otro) a partir de las siguientes propiedades de **ausencia de sesgo**, **eficiencia**, **consistencia** y **suficiencia** (Papoulis and Pillai, 2002).

Si al considerar el caso de que el parámetro a estimar existe y se tiene la seguridad de que vale  $\vartheta$ , se puede interpretar la tarea de estimar a partir de las observaciones, como una VA  $\hat{\theta}$  cuyo valor medio de las estimaciones realizadas sea igual al parámetro correcto. En el caso que esto no sea posible se dice que el estimador es **sesgado** o que tiene sesgo  $b$ , definido por:

$$b^2 \equiv [E\{\hat{\theta}\} - \theta]^2 \quad (24)$$

Cuando la duración de las realizaciones aumenta tendiendo a infinito, el sesgo tiende a cero, en ese caso se dice que el estimador es asintóticamente insesgado y se define de la siguiente manera:

$$\lim_{n \rightarrow \infty} E\{\hat{\theta}_n\} = \theta \quad (25)$$

Pero la ausencia de sesgo no es condición suficiente para definir un buen estimador, se busca que la VA  $\hat{\theta}$  tenga una desviación respecto de la media, pequeña o mínima varianza, y esta característica entre estimadores insesgados dirá cuán **eficiente** es uno del otro. Si el estimador es asintóticamente insesgado y además su varianza tiende a cero, se dice que el estimador tiene la propiedad de **consistencia**, esto es:

$$\begin{aligned} E\{\hat{\theta}_n\} &\rightarrow \theta \\ \sigma_{\theta}^2 &\equiv E\left[(\hat{\theta} - E\{\hat{\theta}\})^2\right] \rightarrow 0 \end{aligned} \quad (26)$$

Para cada estimador insesgado se obtendrá entonces, una varianza mayor que cero, si las observaciones son de duración finita. El más eficiente de estos estimadores, tendrá una varianza acotada por el límite



de Cramer-Rao, como se indica en (27). Si la varianza del estimador insesgado, coincide con el límite de Cramer-Rao, se dice que tiene **eficiencia absoluta** o que el estimador tiene la propiedad de **suficiencia**.

$$\sigma_{\theta}^2 \geq \left[ E \left\{ \left( \frac{\partial}{\partial \theta} \log [f(x(t, \xi) / \theta)] \right)^2 \right\} \right]^{-1} \quad (27)$$

## FILTRO WIENER

El filtro de Wiener es uno de los filtros lineales óptimos más importantes, que permite determinar o estimar la respuesta o el comportamiento de un sistema, a la vez que mejora la apariencia de señales digitales en lo que respecta a la relación señal – ruido SNR.

Se define entonces a  $x[n]$  una señal discreta de  $N$  elementos, cuya expresión vectorial es igual a:

$$\underline{x}^T = [x(0), x(1), \dots, x(n), \dots, x(N-2), x(N-1)] \quad (28)$$

De la misma manera, se define a  $h[n]$  como la respuesta impulsional de un filtro FIR a estimar, de  $Q$  coeficientes de longitud, cuyas características den lugar a la señal discreta  $\hat{y}[n]$  lo más parecida posible a una señal discreta de referencia denominada  $d[n]$  de  $N+Q-1$  muestras de longitud. El parecido entre las señales, se lo evalúa en términos del mínimo error medio cuadrático o MSE (Hsu, 1997; Papoulis and Pillai, 2002; Manolakis, et al., 2005), como se indica en (29). En la Figura 3 se resume tal situación.

$$\xi = E \{ |\varepsilon[n]|^2 \} = E \{ |d[n] - \hat{y}[n]|^2 \} \quad (29)$$

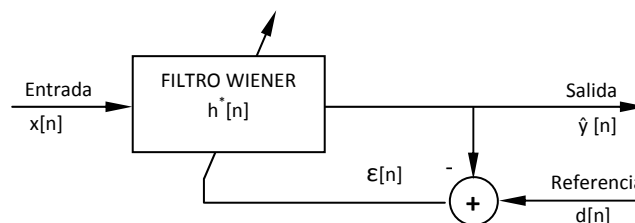




Figura 3: Esquema del filtro de Wiener, con indicación de datos, salida, referencia y el error.

La señal de salida del sistema puede encontrarse a través de la convolución de la señal de entrada  $x[n]$  con la respuesta impulsional conjugada  $h^*[n]$  del sistema a estimar, cuya representación matricial es la siguiente:

$$\hat{y}[n] = h^*[n] * x[n] = \underline{h}^H \cdot \underline{X}_n = \underline{h}^H \cdot \begin{bmatrix} x[n] \\ x[n-1] \\ \vdots \\ x[n-Q+1] \end{bmatrix} \quad (30)$$

el superíndice  $H$  del vector de respuesta impulsional indica que el vector es transpuesto conjugado, el vector  $\underline{X}_n$  esta conformado por los desplazamientos (hasta  $Q$ ) de la señal  $x[n]$  (run-time vector), obteniendo de esta manera un vector de salida de  $N+Q-1$  elementos.

Al desarrollar el valor esperado de (29), teniendo en cuenta la ecuación (30), se obtiene la expresión del error a minimizar:

$$\xi = E \left\{ |d[n]|^2 \right\} + \underline{h}^H \cdot \underline{R} \cdot \underline{h} - \underline{h}^H \cdot \underline{P} - \underline{P}^H \cdot \underline{h} \quad (31)$$

siendo  $\underline{R} = E \left\{ \underline{X}_n \cdot \underline{X}_n^H \right\}$  la matriz de correlación de la señal de entrada  $x[n]$  con los  $Q$  primeros valores de la función de autocorrelación, matriz del tipo Toeplitz de  $Q \times Q$  elementos dado por (32);  $\underline{P} = E \left\{ \underline{X}_n \cdot d^*[n] \right\}$  es el vector de  $Q$  elementos, cuyas entradas corresponden a los valores de la correlación cruzada entre la señal de entrada  $x[n]$  y la señal de referencia  $z[n]$  hasta el orden  $Q$ . Derivando la expresión del error respecto del vector  $\underline{h}^H$  e igualando a cero, se obtiene el vector óptimo para la respuesta impulsiva del sistema o filtro Wiener, cuya expresión esta dada por (33). La señal de salida  $\hat{y}[n]$  de este sistema será entonces la de ecuación (34).



$$\underline{\underline{R}} = E\{\underline{X}_n \cdot \underline{X}_n^H\} = \begin{bmatrix} r(0) & r(1) & \cdots & r(Q-1) \\ r(-1) & r(0) & \cdots & r(Q-2) \\ \vdots & \vdots & \ddots & \vdots \\ r(-Q+1) & r(-Q+2) & \cdots & r(0) \end{bmatrix} \quad (32)$$

$$\underline{h}_{opt} = \underline{\underline{R}}^{-1} \cdot \underline{P} \quad \text{filtro Wiener} \quad (33)$$

$$\hat{y}[n] = \underline{h}_{opt}^H \cdot \underline{X}_n = \underline{P}^H \underline{\underline{R}}^{-1} \cdot \underline{X}_n \quad (34)$$

El error mínimo MSE que se obtiene a partir del vector óptimo de la respuesta impulsiva se obtiene de reemplazar (33) en (31), obteniendo:

$$\xi_{min} = E\{|d[n]|^2\} - \underline{P}^H \cdot \underline{\underline{R}}^{-1} \cdot \underline{P} = E\{|d[n]|^2\} - \underline{h}_{opt}^H \cdot \underline{\underline{R}} \cdot \underline{h}_{opt} \quad (35)$$

Por lo expuesto, si minimizar el error es minimizar la norma del vector  $\varepsilon[n]$ , utilizando la definición de producto escalar, el error será ortogonal al plano de los datos de la señal de entrada, por lo que el principio de ortogonalidad es el que se expresa en (36). Por lo tanto, cuanto menos se parezca la señal de entrada del error o dicho de otra manera no correladas, los coeficientes de la respuesta impulsiva estarán más cerca del valor óptimo.

$$\underline{\varepsilon} \perp \underline{X}_n \Rightarrow E\{\varepsilon^*[n]x[n-q]\} = 0 \quad \text{con} \quad q = 0, \dots, Q-1 \quad (36)$$

Para analizar la forma de trabajar del filtro de Wiener en el dominio de la frecuencia, es necesario extender las ecuaciones de diseño para un filtro de respuesta impulsional infinita o filtro IIR, en lugar del caso del filtro FIR causal. Considerando entonces para el diseño, la correlación de la señal de referencia con el de la señal de entrada al filtro de  $Q$  coeficientes, se tiene:



$$R_{ZX}(l) = \sum_{q=0}^{Q-1} h(q) \cdot R_X(l-q) \quad (37)$$

Si en la expresión (37) los límites de la sumatoria van desde  $-\infty$  a  $\infty$ , el comportamiento del filtro Wiener puede representarse en el dominio de la frecuencia, de la siguiente manera:

$$H(\omega)S_X(\omega) = S_{ZX}(\omega) \quad \Leftrightarrow \quad H^*(-\omega)S_X(\omega) = S_{ZX}(\omega) \quad (38)$$

siendo  $H(\omega)$  la respuesta en frecuencia del filtro Wiener. Ya que se trata de un sistema discreto se debería escribir  $H(z)|_{\exp(j\omega T)}$ , con  $H(z)$  como la transformada Z de los coeficientes del filtro y con  $T$  como el periodo de muestreo. Aplicando el teorema de Parseval a la expresión (35), se puede obtener la expresión del mínimo error MSE en términos de las densidades espectrales de potencia como:

$$\begin{aligned} \xi_{\min} &= E\{|d[n]|^2\} - \underline{h}_{opt}^H \cdot \underline{P} = \frac{1}{2\pi} \int_{-\infty}^{\infty} [S_Z(\omega) - H^*(\omega) \cdot S_{ZX}(\omega)] \cdot d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} S_Z(\omega) \left[ 1 - \frac{S_{ZX}(\omega) \cdot S_{ZX}(\omega)}{S_X(\omega) \cdot S_Z(\omega)} \right] \cdot d\omega = \frac{1}{2\pi} \int_{-\infty}^{\infty} S_Z(\omega) [1 - |\gamma(\omega)|^2] \cdot d\omega \end{aligned} \quad (39)$$

Se observa en (39), la dependencia del mínimo error MSE con la coherencia e implícitamente con la relación SNR, entre los datos de la señal de entrada y de la señal de referencia. De esta manera, la coherencia espectral permite anticipar cuan efectivo será el filtro de Wiener, mostrando en que frecuencias va a existir un mínimo error MSE (Manolakis et al., 2005).

Por la limitada cantidad de realizaciones que se dispone por lo general de los procesos aleatorios implicados y por el desconocimiento de sus PDF conjuntas, la matriz de autocorrelación de la señal de entrada  $x[n]$ , como el vector de correlación cruzada entre la señal de entrada  $x[n]$  y la señal de referencia  $z[n]$ , con sus respectivas densidades espectrales, van a ser estimadas. Manolakis et al. (2005), demuestra que un buen estimador de la matriz de autocorrelación y de la función de autocorrelación, es el que se expresa en (40). Particularmente, el estimador de la función de autocorrelación, se lo denomina “estimador sesgado normalizado” de la autocorrelación.



$$\hat{\underline{R}} = \frac{1}{N} \cdot \text{Toeplitz}(\underline{X}_n \cdot \underline{X}_n^H) \quad (40)$$

$$\hat{R}_x(m) = \frac{1}{N} \sum_{n=0}^{N-|m|-1} x^*[n] \cdot x[n+m] \quad m = 0, \dots, N-1$$

Mientras que para la densidad espectral de potencia de la señal  $x[n]$ , un estimador bastante utilizado es el de la siguiente expresión:

$$S_x(k) = \frac{1}{\underline{S}^H \underline{S}} \underline{S}^H \cdot \underline{R} \cdot \underline{S} = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n] \cdot e^{j2\pi kn/N} \right|^2 = \frac{1}{N} |X(k)|^2 \quad (41)$$

$$\underline{S}^H = \begin{bmatrix} 1 & e^{j2\pi/N} & \dots & e^{j2\pi/N(Q-1)} \end{bmatrix}$$

donde  $X(k)$ , es la transformada discreta de Fourier o DFT.

## FILTROS ADAPTIVOS

Debido a que en la naturaleza es posible registrar señales, con características de no estacionariedad, el filtro de Wiener no resulta apropiado para su utilización, ya que este solo brinda la solución óptima para el caso de sistemas estacionarios.

Considerando el supuesto de no estacionariedad, para el caso del filtro Wiener FIR de la expresión (33), la no estacionariedad implica el considerar los procesos aleatorios como no estacionarios y que los coeficientes del filtro que minimizan el error de la expresión (31) dependen de  $n$ .

De esta manera el concepto del filtro adaptativo (variante en el tiempo) consiste en encontrar para cada instante de tiempo  $n$  de iteración del algoritmo, un conjunto de coeficientes óptimos, a través de considerar las actualizaciones de los pesos del filtro:

$$\underline{h}_{n+1} = \underline{h}_n + \mu f(\underline{h}_n) \quad (42)$$



donde  $\mu f(\underline{h}_n)$  es un factor de corrección que se aplica a los coeficientes  $\underline{h}_n$  en el instante  $n$  para obtener el nuevo conjunto de coeficientes  $\underline{h}_{n+1}$  en el instante  $n + 1$ . Esta ecuación de actualización es la base de los algoritmos adaptivos y el diseño de cada filtro adaptivo requiere definir el factor  $\mu f(\underline{h}_n)$ . El algoritmo utilizado para aplicar la corrección del factor  $\mu f(\underline{h}_n)$  debe ser tal, que en un escenario estacionario, el conjunto de coeficientes del filtro  $\underline{h}_n$  converja a la solución óptima de Wiener de la expresión (33). Mientras que para señales no estacionarias, el filtro debe ser capaz de adaptarse a los cambios estadísticos y alcanzar la solución al evolucionar en el tiempo. En la Figura 4, se muestra el esquema general de los filtros adaptivos. En los párrafos siguientes se presenta una síntesis de los algoritmos típicos de filtros adaptivos. Puede ampliar el tema de filtros adaptivos en Manolakis et al. (2005).

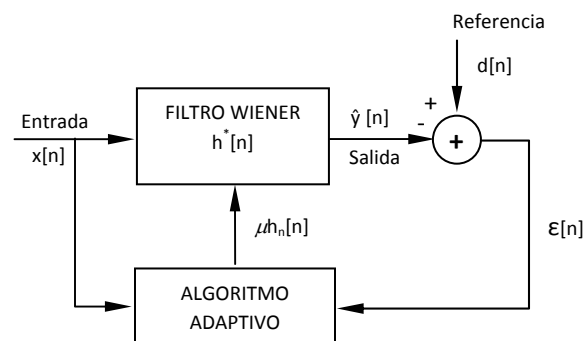


Figura 4: Esquema general de los filtros adaptivos, con indicación de datos, salida, referencia y el error.

## Método de Gradiente de Máxima Pendiente

A partir de los coeficientes óptimos del filtro Wiener y del valor de mínimo error de la expresión (35), es posible contar con una expresión del error MSE para cualquier respuesta impulsional en función de la óptima:

$$\xi(\underline{h}_n) = \xi_{\min} + (\underline{h}_n - \underline{h}_{opt})^H \underline{\underline{R}} (\underline{h}_n - \underline{h}_{opt}) \quad (43)$$

siendo  $n$  un número que expresa las iteraciones en el proceso de aprendizaje.



Como puede observarse en la expresión (43), existe una dependencia cuadrática del valor del error con los coeficientes y cuyo mínimo error se logra cuando se esta ante los coeficientes óptimos del filtro Wiener. La superficie lograda de esta expresión, tiene el aspecto de un paraboloide, cuyas curvas de nivel corresponden a elipses.

En la Figura 5 puede observarse el mecanismo de aprendizaje a realizar, desde un estado  $\underline{h}_n$  a otro estado mejor  $\underline{h}_{n+1}$ , el cual consiste en tomar la dirección contraria al gradiente del MSE y trasladarse una determinada cantidad en dicha dirección. De lo anterior, se deduce la siguiente regla de aprendizaje:

$$\underline{h}_{n+1} = \underline{h}_n - \mu \nabla \xi(\underline{h}_n) = \underline{h}_n - \mu (\underline{R} \underline{h}_n - \underline{P}) \quad (44)$$

en donde la cantidad  $\mu$  corresponde al paso de adaptación, el cual determina la velocidad de aprendizaje que se desea imprimir al sistema.

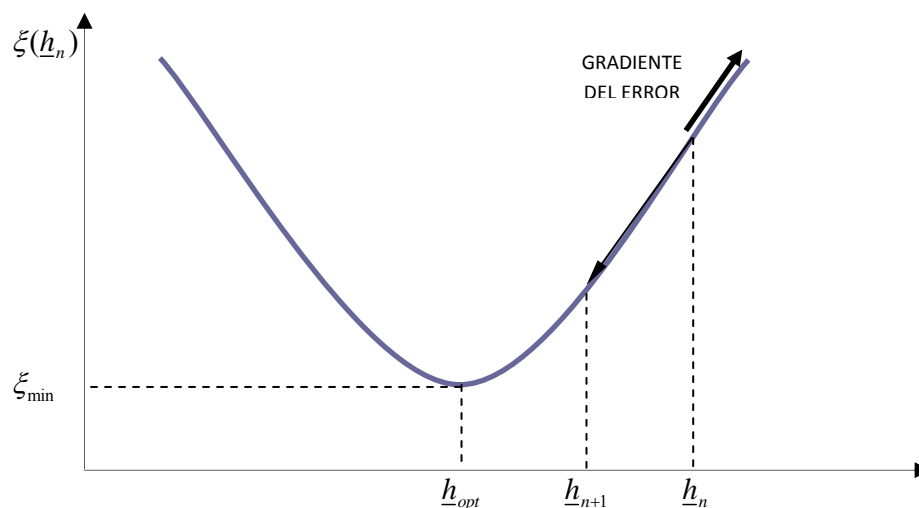


Figura 5: Curva del MSE en función de los coeficientes. Su comportamiento cuadrático conlleva que el gradiente cambiado de signo, en cualquier posición, marca la dirección a seguir para alcanzar el mínimo.

Si bien la representación de la Figura 5 es adecuada para la explicación, una representación en dos dimensiones revela que el método del gradiente tan solo denota una dirección de menor error. Sin embargo, dependiendo de su posición, no necesariamente esta dirección corresponde a la dirección del mínimo error (ver Figura 6). En dicha figura, las curvas de nivel representan las líneas de igual error medio cuadrático. Sólo en el caso de que estas líneas de igual error sean circunferencias, el gradiente en cualquier punto apuntará siempre al mínimo error.

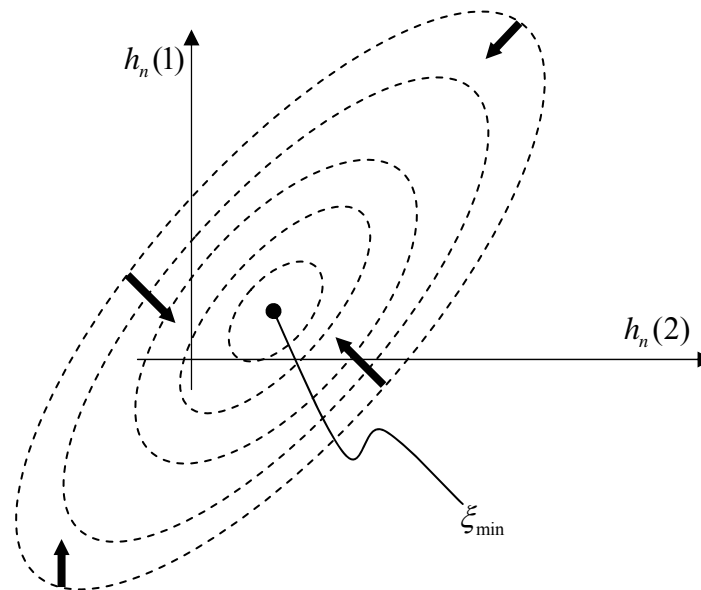


Figura 6: Curvas de nivel del error para el caso de un filtro de dos coeficientes, en donde se observan diversos vectores de gradiente. Las magnitudes de los vectores gradientes son es inversamente proporcionales a la separación entre las curvas de nivel.

Realizando un cambio de variables en la expresión (43) de manera de comprobar la relación entre la forma de las curvas de nivel y la matriz de correlación, se tiene:

$$\xi(\tilde{\underline{h}}_n) = \xi_{\min} + \tilde{\underline{h}}_n^H \underline{\underline{R}} \tilde{\underline{h}}_n \quad (45)$$



Este cambio de variables producirá el centrado de las curvas de nivel en el valor óptimo  $\underline{h}_{opt}$ . De este modo, el gradiente del error de la expresión (43), en cualquier punto de la superficie, puede expresarse en términos de la nueva variable como:

$$\nabla \xi(\tilde{\underline{h}}_n) = \underline{\underline{R}} \tilde{\underline{h}}_n \quad (46)$$

En particular, el gradiente en los puntos extremos de los ejes de las curvas de nivel es un vector que pasa por el origen de coordenadas y por lo tanto es de la forma  $k\tilde{\underline{h}}_n$ , por lo que las direcciones de los ejes principales vienen dadas por los autovectores de la matriz  $\underline{\underline{R}}$ . Se demuestra entonces, la siguiente expresión entre los autovalores de  $\underline{\underline{R}}$  y la superficie de error:

$$\xi(\underline{z}_n) = \xi_{\min} + \tilde{\underline{h}}_n^H \left( \underline{\underline{E}} \underline{\underline{\Lambda}} \underline{\underline{E}}^H \right) \tilde{\underline{h}}_n = \xi_{\min} + \underline{z}_n^H \underline{\underline{\Lambda}} \underline{z}_n = \xi_{\min} + \sum_{i=1}^Q \lambda_i |z_n(i)|^2 \quad (47)$$

en donde  $\underline{\underline{E}}$  corresponde a la matriz de vectores propios y  $\underline{\underline{\Lambda}}$  a la matriz diagonal de autovalores de la matriz  $\underline{\underline{R}}$ . De esta manera se logra una rotación del sistema coordenado, quedando los ejes de las curvas de error alineados con las nuevas coordenadas  $\underline{z}_n = \underline{\underline{E}}^H (\underline{h}_n - \underline{h}_{opt})$  (ver Figura 7).

La curvatura en este nuevo sistema coordenado, se obtiene de calcular la segunda derivada de la expresión del error de (47) respecto de  $z_n(i)$ , igual a  $2\lambda_i$ . Por lo que la dirección del eje de menor longitud del grafico de la Figura 7, estará asociada al vector propio de mayor autovalor, y viceversa. Manolakis et al. (2005) enuncia que la excentricidad de las curvas de nivel de la función de error depende de cuan distintos son los autovalores de la matriz de correlación. Esta observación esta íntimamente relacionada con la velocidad de convergencia.

Considerando la regla de aprendizaje de la expresión (44) y teniendo en cuenta el cambio de variable realizado para obtener el gradiente de la expresión (46), se deduce la regla de aprendizaje en torno a la variable  $\tilde{\underline{h}}_n$ :

$$\underline{\tilde{h}}_{n+1} = \underline{\tilde{h}}_n - \mu \underline{R} \underline{\tilde{h}}_n \quad (48)$$

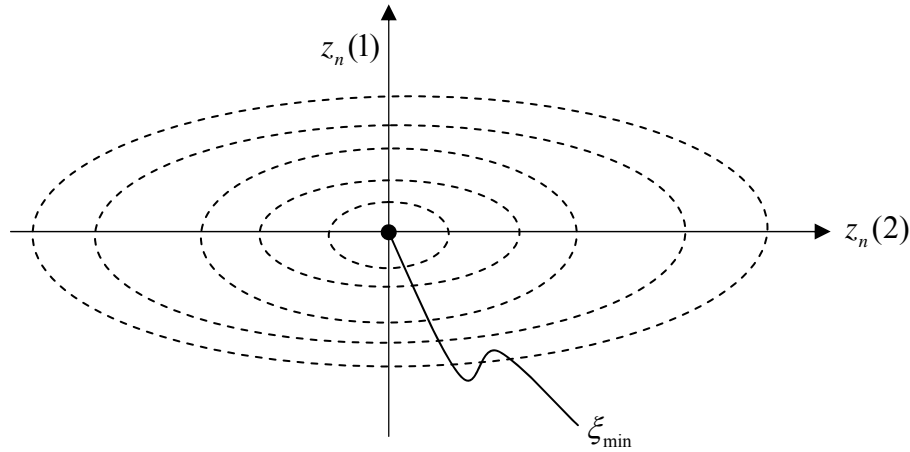


Figura 7: Curvas de error de la superficie de la figura V.4, representadas sobre las nuevas variables  $\underline{z}_n$ .

Debido a que la matriz de vectores propios  $\underline{E}$  es ortogonal, se cumple que:

$$\underline{E}^H \underline{E} = \underline{E} \underline{E}^H = \underline{I} \quad (49)$$

siendo  $\underline{I}$  la matriz identidad de de  $Q \times Q$  elementos. Por lo tanto de multiplicar en ambos miembros de la expresión (48) por la matriz  $\underline{E}^H$ , se obtiene la regla de aprendizaje desacoplada en torno a la variable  $\underline{z}_n$ :

$$\underline{z}_{n+1} = (\underline{I} - \mu \underline{\Lambda}) \underline{z}_n \quad (50)$$

$$\begin{bmatrix} z_{n+1,0} \\ z_{n+1,1} \\ \vdots \\ z_{n+1,Q-1} \end{bmatrix} = \begin{bmatrix} 1 - \mu \lambda_0 & 0 & \cdots & 0 \\ 0 & 1 - \mu \lambda_1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 - \mu \lambda_{Q-1} \end{bmatrix} \begin{bmatrix} z_{n,0} \\ z_{n,1} \\ \vdots \\ z_{n,Q-1} \end{bmatrix}$$



De esta manera la adaptación del  $i$ -ésimo coeficiente o modo es igual a:

$$z_{n+1,i} = (1 - \mu\lambda_i)z_{n,i} \quad (51)$$

siendo la serie geométrica de razón  $(1 - \mu\lambda_i)$  igual:

$$z_{n,i} = (1 - \mu\lambda_i)^n z_{0,i} \quad (52)$$

en donde  $z_{0,i}$  corresponde a los  $i$ -ésimos coeficientes iniciales del proceso de aprendizaje. Por lo, lo que la condición de convergencia es igual:

$$|1 - \mu\lambda_i| < 1, \quad \forall i \quad (53)$$

Esta condición habría de verificarse para cada uno de los autovalores pero, dado que la matriz de autocorrelación es definida positiva, sus autovalores son positivos y ordenados, en consecuencia si la condición (53) verifica para el autovalor máximo se verificara para todos los demás. Conjuntamente, como el paso de adaptación es positivo, para no cambiar la dirección del gradiente, se llega fácilmente a la condición de convergencia:

$$0 < \mu < \frac{2}{\lambda_{\max}} \quad (54)$$

es decir, el paso de adaptación viene limitado por el doble de la inversa del autovalor máximo de la matriz de autocorrelación.



## El Algoritmo LMS

En el método de máxima pendiente es necesario conocer los parámetros estadísticos de de segundo orden para el calculo del error MSE. En ese caso es necesario conocer la matriz de autocorrelación  $\underline{\underline{R}} = E\{\underline{X}_n \cdot \underline{X}_n^H\}$  y el vector  $\underline{P} = E\{\underline{X}_n \cdot d^*[n]\}$ , que en la práctica son dificultosos de encontrar. El algoritmo LMS (por sus siglas en ingles de Least Mean Square), realiza una estimación instantánea de dichos parámetros estadísticos, siendo iguales a:

$$\begin{aligned}\underline{\underline{\hat{R}}} &= \underline{X}_n \cdot \underline{X}_n^H \\ \underline{\hat{P}} &= \underline{X}_n \cdot d^*[n]\end{aligned}\tag{55}$$

Por lo que a partir de dicha estimación, se obtiene la siguiente regla de de aprendizaje:

$$\underline{h}_{n+1} = \underline{h}_n - \mu(\underline{\hat{R}}\underline{h}_n - \underline{\hat{P}}) = \underline{h}_n + \mu\underline{X}_n(d^*[n] - \hat{y}^*[n]) = \underline{h}_n + \mu\underline{X}_n \varepsilon^*[n]\tag{56}$$

Resulta evidente que una cualidad deseable de un algoritmo adaptivo es que converja a la solución óptima en el menor número de iteraciones posible. Para medir la tasa de convergencia o la constante de tiempo asociada, es de utilidad optar el tipo de convergencia a tener en cuenta. Se definen entonces dos tipos de convergencia a considerar en el algoritmo LMS:

$$\begin{aligned}\lim_{n \rightarrow \infty} E\{\underline{h}_n\} &= \underline{h}_{opt} && \text{Convergencia en media} \\ \lim_{n \rightarrow \infty} E\{\xi(n)\} &= \xi(\infty) = cte && \text{Convergencia en MSE}\end{aligned}\tag{57}$$

En el caso de la convergencia media, el concepto surge de considerar a los coeficientes del filtro como una variable aleatoria cuya media es el filtro óptimo. Para el caso de la convergencia MSE, la media del error MSE es un valor  $\xi(\infty)$  constante. Si este valor es mayor que  $\xi_{\min}$ , queda de manifiesto el grado de desajuste del algoritmo, el cual se define de la siguiente manera:



$$D = \frac{E\{\xi(n)\} - \xi_{\min}}{\xi_{\min}} 100\% \quad (58)$$

En la Figura 8 se muestra la curva de aprendizaje del algoritmo adaptivo, en donde se ilustra el concepto de desajuste. Esta curva de aprendizaje no es otra cosa que la representación gráfica del error MSE en función de  $n$ .

Determinando la convergencia en media de los pesos del filtro en el algoritmo LMS a partir de la ecuación (56), se obtiene:

$$\begin{aligned} E\{\underline{h}_{n+1}\} &= E\{\underline{h}_n\} + \mu E\{\underline{X}_n(d^*[n] - \underline{X}_n^H \underline{h}_n)\} \\ &= E\{\underline{h}_n\} + \mu [\underline{P} - \underline{R} E\{\underline{h}_n\}] \\ &= [\underline{I} - \mu \underline{R}] E\{\underline{h}_n\} + \mu \underline{P} \end{aligned} \quad (59)$$

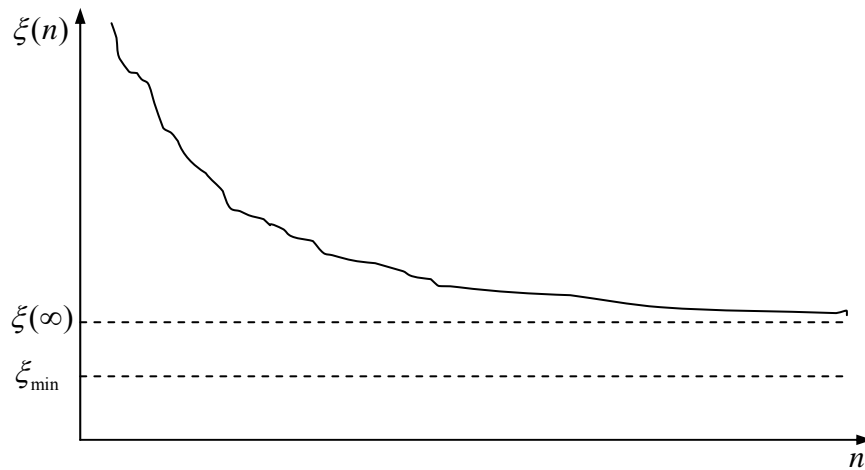


Figura 8: Curva de aprendizaje del algoritmo adaptivo

Recordando el cambio de variables  $\underline{z}_n$  y la diagonalización de la matriz  $\underline{R}$  se obtiene se obtiene la regla de aprendizaje desacoplada:

$$E\{\underline{z}_{n+1}\} = [\underline{I} - \mu \underline{\Lambda}] E\{\underline{z}_n\} \quad (60)$$





Así, el algoritmo LMS converge en media desde  $\underline{h}_0$  hasta  $\underline{h}_{opt}$ , si y sólo si se cumple la condición mostrada en la ecuación (54). Sin embargo, este criterio de estabilidad no es aplicable en la práctica debido a la dificultad de calcular el valor  $\lambda_{\max}$  cuando el tamaño del filtro es grande. De allí que para aplicaciones prácticas, se realiza una estimación de este valor de la siguiente forma:

$$\lambda_{\max} \leq tr(\underline{R}) = \sum_{i=0}^{Q-1} \lambda_i = Q R_X(0) = Q P_X \quad (61)$$

donde  $tr(\underline{R})$  es la traza de la matriz  $\underline{R}$  y donde  $P_X = R_X(0) = E\{x^2[n]\}$  denota la potencia de la señal  $x[n]$ . Por lo tanto el parámetro  $\mu$  queda acotado del siguiente modo:

$$0 < \mu < \frac{2}{Q P_X} \quad (62)$$

El análisis anterior sólo asegura la convergencia en media de los coeficientes del filtro, sin embargo, asegurar la convergencia en media no garantiza la estabilidad en varianza. Por lo tanto, para que los coeficientes también converjan en varianza hay que exigir una condición más restrictiva para el factor de convergencia  $\mu$ :

$$0 < \mu < \frac{2}{3 \lambda_{\max}} \quad (63)$$

que se traduce en términos de la potencia de la señal de entrada en la siguiente condición:

$$0 < \mu < \frac{2}{3 Q P_X} \quad (64)$$



En cuanto al desajuste del algoritmo LMS, el cálculo se realiza basado en la descomposición de la estimación del gradiente en dos componentes: el gradiente real y un ruido de gradiente. Calculando la covarianza de este ruido una vez que los coeficientes han convergido, se prueba la siguiente expresión:

$$D = \frac{\sum_{i=0}^{Q-1} \frac{\mu \lambda_i}{2 - \mu \lambda_i}}{1 - \sum_{i=0}^{Q-1} \frac{\mu \lambda_i}{2 - \mu \lambda_i}} \quad (65)$$

Si se cumple que  $\mu \lambda_i \ll 1$  se tiene:

$$D \approx \frac{\mu \text{tr}(\underline{\underline{R}})}{2 - \mu \text{tr}(\underline{\underline{R}})} \quad (66)$$

si además se cumple que  $\mu \text{tr}(\underline{\underline{R}}) \ll 1$  entonces:

$$D \approx \frac{\mu}{2} \text{tr}(\underline{\underline{R}}) = \frac{\mu}{2} Q P_x \quad (67)$$

En resumen, el LMS es un algoritmo que no requiere a priori un conocimiento de la estadística de las señales, requiere pocas operaciones por muestra, ofrece un buen compromiso entre prestaciones y simplicidad, es sensible a la dispersión de valores propios de la matriz de autocorrelación lo cual define la velocidad de convergencia del algoritmo y es robusto ante problemas de ruido, errores de cuantificación y empleo de aritmética finita.

## El Algoritmo RLS

Una posibilidad que se aparta de los métodos de gradiente es recurrir a la estimación, muestra a muestra, de las componentes de la respuesta al impulso optima de la solución de Wiener a partir de una solución recursiva. Este es el objeto del algoritmo de mínimos cuadrados recursivo o RLS (por sus siglas



en inglés de Recursive Least Squares). Dichas estimaciones se realizan a través del promedio de un número  $M$  de muestras más recientes del vector de datos  $\underline{X}_n$  (run-time vector) y de la señal referencia  $d[n]$ .

Planteando la minimización del error  $\varepsilon[n] = d[n] - \hat{y}[n]$  desde un punto de vista determinístico y no estocástico como el planteado en la expresión (29), se tiene el error cuadrático o función de costo a minimizar:

$$\begin{aligned}\xi(n) &= \sum_{m=0}^n \alpha^{n-m} |\varepsilon[m]|^2 = \sum_{m=0}^n \alpha^{n-m} |d[m] - \underline{h}_n^H \cdot \underline{X}_m|^2 \\ &= \sum_{m=0}^{n-1} \alpha^{n-m} |d[m] - \underline{h}_n^H \cdot \underline{X}_m|^2 + \alpha |d[n] - \underline{h}_n^H \cdot \underline{X}_n|^2 \\ &= \alpha \xi(n-1) + |d[n] - \underline{h}_n^H \cdot \underline{X}_n|^2\end{aligned}\quad (68)$$

en donde  $\varepsilon[m]$  es el error instantáneo y la constante  $\alpha$ ,  $0 < \alpha \leq 1$ , es el factor de olvido exponencial. De esta manera, los coeficientes del filtro que reducen el error cuadrático se especifican a través de las actualizaciones de las siguientes ecuaciones:

$$\begin{aligned}\lim_{n \rightarrow \infty} \hat{\underline{R}}_n &= \lim_{n \rightarrow \infty} \{ \alpha \hat{\underline{R}}_{n-1} + \underline{X}_n \underline{X}_n^H \} = \underline{\underline{R}} \\ \lim_{n \rightarrow \infty} \hat{\underline{P}}_n &= \lim_{n \rightarrow \infty} \{ \alpha \hat{\underline{P}}_{n-1} + \underline{X}_n d^*[n] \} = \underline{\underline{P}}\end{aligned}\quad (69)$$

Una vez realizada la actualización partir de la expresión (69), se obtiene el vector de coeficientes de la siguiente manera:

$$\underline{h}_{n+1} = \hat{\underline{R}}_{n+1}^{-1} \hat{\underline{P}}_{n+1}\quad (70)$$

El procedimiento expuesto para actualizar los pesos del filtro adolece de dos problemas. El primero es que no se dispone, por el momento de una dependencia explícita entre los coeficientes en el instante  $n$  y el instante  $n-1$ . El segundo problema, es que el procedimiento, requiere de la inversión de la matriz



de autocorrelación. La solución a ambos problemas radica en la solución del segundo de problema, a través de la utilización del denominado lema de la inversa (Manolakis et al., 2005). Este mecanismo, permite escribir la ecuación de recursión para la matriz de autocorrelación de la expresión (69) en términos de la matriz inversa, de la siguiente manera:

$$\hat{\underline{R}}_{n+1}^{-1} = \alpha^{-1} \hat{\underline{R}}_n^{-1} - \left( \alpha^{-1} \hat{\underline{R}}_n^{-1} \underline{X}_{n+1} \right) \left( \underline{I} + \alpha^{-1} \underline{X}_{n+1}^H \hat{\underline{R}}_n^{-1} \underline{X}_{n+1} \right)^{-1} \left( \alpha^{-1} \underline{X}_{n+1}^H \hat{\underline{R}}_n^{-1} \right) \quad (71)$$

Rescribiendo la ecuación (70) a partir de la ecuación (69), se tiene:

$$\underline{h}_{n+1} = \hat{\underline{R}}_{n+1}^{-1} \left( \alpha \hat{\underline{P}}_n + \underline{X}_{n+1} d^*[n+1] \right) \quad (72)$$

expresión que da lugar, en conjunto con la ecuación (71), a la actualización de los coeficientes del filtro en forma recursiva. Reagrupando los términos, se obtiene la siguiente expresión general de actualización de los coeficientes del filtro:

$$\underline{h}_{n+1} = \underline{h}_n + \underline{k}_{n+1} \varepsilon^*[n+1] \quad (73)$$

en donde  $\underline{k}_{n+1}$  es el vector de ganancias y cuya expresión es:

$$\underline{k}_{n+1} = \left( \alpha^{-1} \hat{\underline{R}}_n^{-1} \underline{X}_{n+1} \right) \left( \underline{I} + \alpha^{-1} \underline{X}_{n+1}^H \hat{\underline{R}}_n^{-1} \underline{X}_{n+1} \right)^{-1} \quad (74)$$

Simplificando la ecuación (71) a través de utilizar la igualdad de la ecuación (74), se tiene:



$$\hat{\underline{R}}_{n+1}^{-1} = \alpha^{-1} \left( \hat{\underline{R}}_n^{-1} - \underline{k}_{n+1} \underline{X}_{n+1}^H \hat{\underline{R}}_n^{-1} \right) \quad (75)$$

El algoritmo RLS es sin duda el mejor algoritmo adaptativo para la minimización del error, ya que sus prestaciones no dependen de la dispersión de autovalores como en el caso de los métodos de gradiente. La convergencia es del orden de la longitud del filtro, es decir, para un filtro de  $Q$  coeficientes el algoritmo demanda  $Q$  iteraciones o vectores de datos para llegar a la convergencia. Por otro lado, el grado de desajuste que se obtiene se minimiza con valores del factor de olvido  $\alpha$  cercanos a la unidad.

## Filtro de Kalman

El objetivo de esta sección es introducir a las nociones básicas relacionadas con el filtrado de Kalman como uno de los algoritmos de filtros adaptivos. El filtro tiene su origen en el documento de Kalman (1960) donde describe una solución recursiva para el problema del filtrado lineal de datos discretos. Otras introducciones al filtro de Kalman pueden ser halladas en Prasad and Mahalanabis (1980), Balakrishnan (1984), Schutz (1994), Papoulis and Pillai (2002), Manolakis et al. (2005).

La solución es óptima por cuanto el filtro combina toda la información observada y el conocimiento previo acerca del comportamiento del sistema para producir una estimación del estado de tal manera que el error es minimizado estadísticamente. El término recursivo significa que el filtro recalcula la solución cada vez que una nueva observación o medida ruidosa es incorporada en el sistema.

El filtro de Kalman es el principal algoritmo para estimar sistemas dinámicos representados en la forma de **espacio-estado**. En esta representación, el sistema es descrito por un conjunto de variables denominadas de estado. El estado contiene toda la información relativa al sistema a un cierto punto en el tiempo. Esta información debe permitir la inferencia del comportamiento pasado del sistema, presente o futuro, dependiendo si la problemática a encarar por parte del filtro de Kalman es el alisado, el filtrado o la predicción respectivamente.

Planteando el concepto del filtro de Kalman a partir de la estimación de la respuesta al impulso de un sistema  $\underline{h}_n$ , el cual se representa a través de la siguiente ecuación lineal de estado:

$$\underline{h}_{n+1} = \underline{F} \underline{h}_n + \underline{v}_n \quad (76)$$



con una ecuación de medida u observación igual a:

$$d^*[n] = \underline{H}_n^H \underline{h}_n + \underline{w}_n = \underline{X}_n^H \underline{h}_n + \underline{w}_n \quad (77)$$

La matriz cuadrada  $\underline{F}_n$ , relaciona a los estados en los periodos de tiempo discreto  $n$  y  $n+1$ , y que en este caso de estimación corresponde a la matriz identidad (Prasad and Mahalanabis, 1980). La matriz rectangular  $\underline{H}_n^H$ , relaciona el vector de estado  $\underline{h}_n$  con el vector de medida u observaciones  $\underline{d}_n^*$ . Si se cuenta con los datos de entrada  $x[n]$  del sistema a estimar, la matriz  $\underline{H}_n^H = \underline{X}_n^H$  (run-time vector). Por otro lado, si solo se cuenta con la señal de referencia o medida  $d[n]$ , la matriz  $\underline{H}_n^H$  es igual a:

$$\underline{H}_n^H = \underline{D}_{n-1}^H = \begin{bmatrix} d^H[n-1] & d^H[n-2] & \cdots & d^H[n-Q] \end{bmatrix} \quad (78)$$

siendo en ese caso un problema de predicción lineal. Los procesos aleatorios correspondientes a los vectores  $\underline{v}_n$  y  $\underline{w}_n$  representan el ruido del estado y de la medida respectivamente. Estos ruidos son del tipo ruido gaussiano cumpliéndose lo siguiente:

$$\begin{aligned} E\{\underline{h}_n \underline{v}_n^H\} &= \underline{0} ; E\{\underline{v}_n\} = \underline{0} ; E\{\underline{v}_n \underline{v}_n^H\} = \underline{V}_n \\ E\{\underline{w}_n \underline{d}_n^H\} &= \underline{0} ; E\{\underline{w}_n \underline{h}_n^H\} = \underline{0} ; E\{\underline{w}_n\} = \underline{0} ; E\{\underline{w}_n \underline{w}_n^H\} = \underline{W}_n \end{aligned} \quad (79)$$

En definitiva, el modelo de filtro establece el vector  $d^*[n]$  como señal disponible y a utilizar para conseguir la estimación más precisa posible del vector de estado.

Partiendo de la base de que se dispone de una estimación del vector de medida, ya que solo se conoce la matriz  $\underline{H}_n^H$ , se tiene:

$$\hat{d}^*[n] = \underline{H}_n^H \hat{\underline{h}}_n \quad (80)$$



Por lo que existirá un error, el cual servirá para mejorar la estimación del vector de estado y que se define de la siguiente manera:

$$\varepsilon[n] = d^*[n] - \hat{d}^*[n] = \underline{H}_n^H (\underline{h}_n - \hat{\underline{h}}_n) + \underline{w}_n = \underline{H}_n^H \tilde{\underline{h}}_n + \underline{w}_n \quad (81)$$

siendo  $\tilde{\underline{h}}_n = \underline{h}_n - \hat{\underline{h}}_n$  el error de estado. Teniendo en cuenta que el ruido de medida está incorrelado con el vector de estado, se puede calcular la matriz de covarianza del vector error de medida en función de la matriz de covarianza del vector de error de estado de la siguiente manera:

$$\underline{\xi}_n = E\{\varepsilon[n]\varepsilon^H[n]\} = \underline{H}_n^H E\{\tilde{\underline{h}}_n \tilde{\underline{h}}_n^H\} \underline{H}_n + \underline{W}_n = \underline{H}_n^H \underline{K}_n \underline{H}_n + \underline{W}_n \quad (82)$$

donde  $\underline{K}_n$  es la matriz de covarianza del vector de error de estado. Esta ecuación revela, que el error de medida viene dado por dos términos: la potencia del ruido de medida y el error de estado a través de su matriz de covarianza.

Conocido el error de medida, se pretende encontrar la manera de utilizar este error para mejorar la estimación del estado. Lo inmediato sugiere que la nueva estimación del vector de estado se modifique de acuerdo al error de medida. Por lo que se establecen las siguientes ecuaciones de adaptación en términos del estado estimado y del error de estado respectivamente:

$$\begin{aligned} \hat{\underline{h}}_{n+1} &= \underline{F}_n \hat{\underline{h}}_n + \underline{k}_n \varepsilon[n] \\ \tilde{\underline{h}}_{n+1} &= \underline{F}_n \tilde{\underline{h}}_n + \underline{v}_n - \underline{k}_n \varepsilon[n] \end{aligned} \quad (83)$$

donde  $\underline{k}_n$  es la matriz de ganancia, la cual se determina a partir de considerar la ortogonalidad entre el error de estado a estimar y el error de medida, esto es:



$$E\{\tilde{\underline{h}}_{n+1} \underline{\varepsilon}^H[n]\} = \underline{0} \quad (84)$$

resolviendo este principio en la ecuación (83), se obtiene:

$$\underline{0} = \underline{F}_n E\{\tilde{\underline{h}}_n \underline{\varepsilon}^H[n]\} - \underline{k}_n \underline{\xi}_n = \underline{F}_n \underline{K}_n \underline{H}_n - \underline{k}_n \underline{\xi}_n \quad (85)$$

teniendo en cuenta, las características de independencia del ruido de medida y despejando la matriz de ganancia de la expresión (85), se obtiene:

$$\underline{k}_n = \underline{\xi}_n^{-1} \underline{F}_n \underline{K}_n \underline{H}_n \quad (86)$$

En este momento, disponiendo de una estimación del vector de estado y de su matriz de covarianza, al disponer del error de medida, se calcula la matriz de ganancia y se actualiza la estimación del vector de estado. Para completar la iteración es necesario dejar disponible la estimación de la matriz de covarianza del nuevo estado. Esta ecuación, que permite actualizar también la matriz de covarianza del error de estado, se obtiene de la ecuación (83), teniendo presente la expresión ya obtenida para la matriz de ganancia:

$$\underline{K}_{n+1} = \underline{F}_n \underline{K}_n \underline{F}_n^H - \underline{k}_n \underline{\xi}_n \underline{k}_n^H + \underline{V}_n \quad (87)$$

En resumen, las ecuaciones que se utilizan para derivar el filtro de Kalman se pueden dividir en dos grupos: las que actualizan el tiempo o ecuaciones de predicción y las que actualizan los datos observados o ecuaciones de actualización. Las del primer grupo son responsables de la proyección del estado al momento  $n+1$  tomando como referencia el estado en el momento  $n$  y de la actualización intermedia de la matriz de covarianza del estado. El segundo grupo de ecuaciones son responsables de la retroalimentación, es decir, incorporan nueva información dentro de la estimación anterior con lo cual se llega a una estimación mejorada del estado.





Las ecuaciones que actualizan el tiempo pueden también ser pensadas como ecuaciones de pronóstico, mientras que las ecuaciones que incorporan nueva información pueden considerarse como ecuaciones de corrección. Efectivamente, el algoritmo de estimación final puede definirse como un algoritmo de pronóstico-corrección para resolver numerosos problemas. Así el filtro de Kalman funciona por medio de un mecanismo de proyección y corrección al pronosticar el nuevo estado y su incertidumbre y corregir la proyección con la nueva medida. Este ciclo se muestra en la Figura 9 .

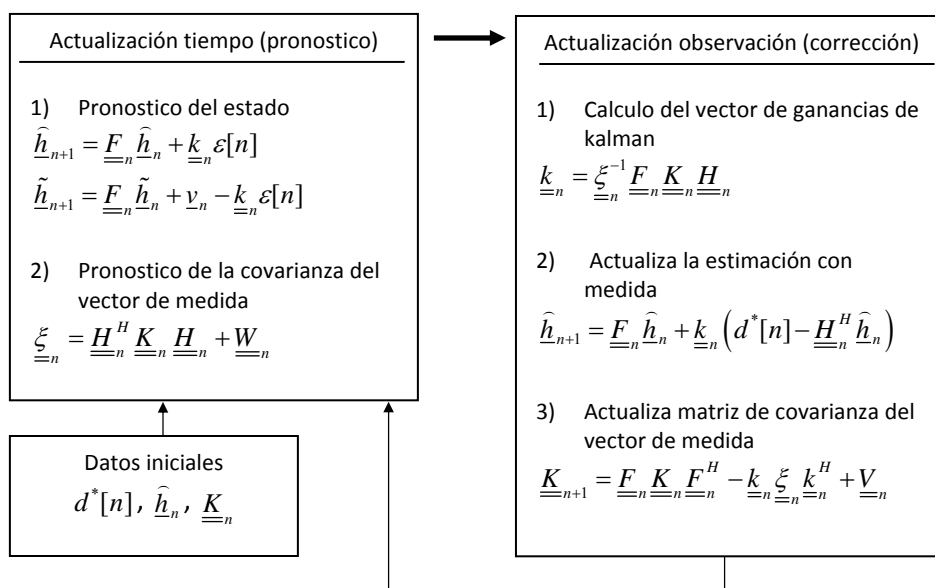


Figura 9: Esquema completo del concepto general del filtro de Kalman

## REFERENCIAS

- Balakrishnan, A. V., 1984. Kalman Filtering Theory. Optimization Software, Inc. Publication Division, New York.
- Hsu, H., 1997. Theory and Problems of Probability, Random Variables, and Random Processes. Mcgraw-Hill, New York.
- Kalman, R. E., 1960. A New Approach to Linear Filtering and Prediction Problems. Transactions of the ASME–Journal of Basic Engineering, 82 (Series D): 35-45.
- Manolakis, D. G., Ingle, V. K. and Kogon, S. M., 2005. Statistical and Adaptive Signal Processing: Spectral Estimation, Signal Modeling, Adaptive Filtering, and Array Processing. Artech House Inc.



Oppenheim, A. and Willsky, A. S., 1983. Signals and Systems. Prentice Hall Inc.

Oppenheim, A. V. and Schaffer, R. W., 1999. Discrete-Time Signal Processing. 2nd Ed. Prentice-Hall Signal Processing Series.

Papoulis, A. and Pillai, S.U., 2002. Probability, Random Variables and Stochastic Processes. 4 Edition , Mc. Graw Hill, New York 10020.

Prasad, S. and Mahalanabis, A. K., 1980. Adaptive Filter Structures for Deconvolution of Seismic Signals. IEEE Transactions on Geoscience and Remote Sensing, Vol. GE-18, No. 3.

Proakis, J. G. and Manolakis, D. G., 1996. Digital Signal Processing: Principles, Algorithms and Applications. Third Edition. Prentice Hall Inc.