

Procesado Estadístico de Señal

Filtrado adaptativo



Departament de Teoria
del Senyal i Comunicacions

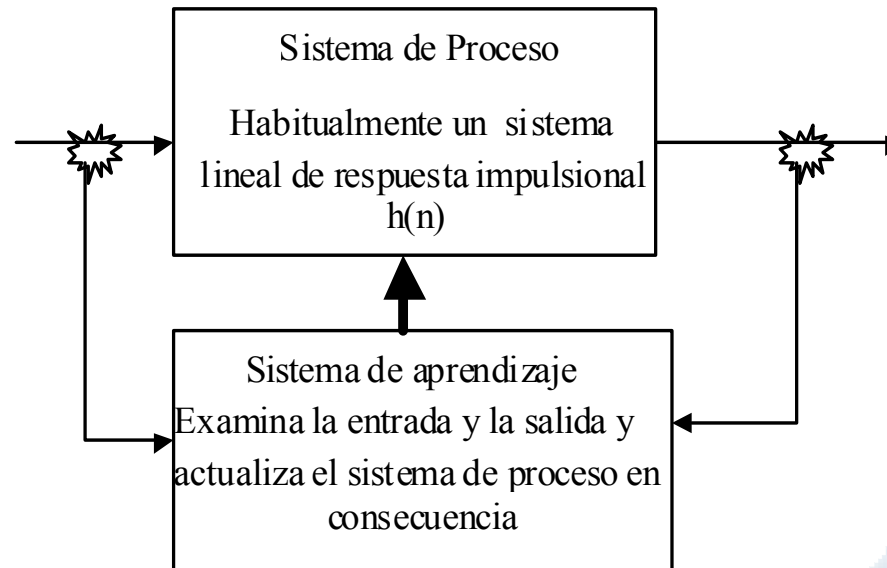


Índice

1. Señales y procesos estocásticos
2. Estimación espectral
3. Filtro de Wiener
4. Filtrado adaptativo
 - 4.1. Método de gradiente
 - 4.2. Diseño y convergencia del método de gradiente
 - 4.3. Algoritmo LMS y NLMS
5. Métodos transformados en codificación



- Adaptación automática al entorno de señal

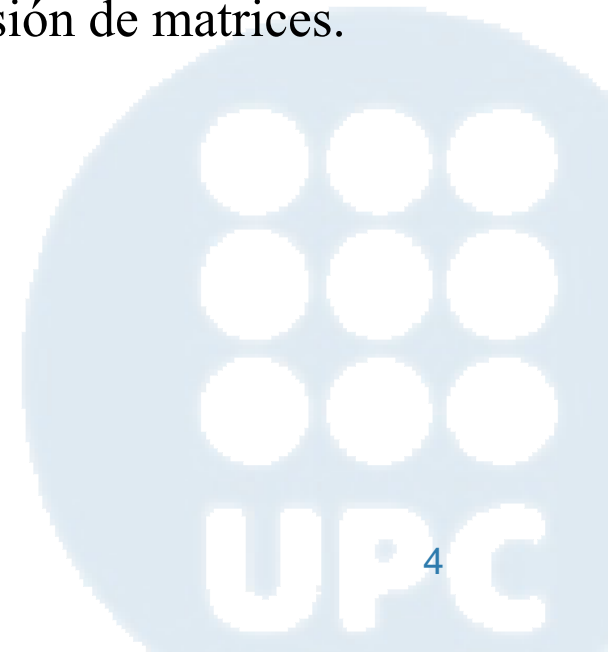


- Ejemplo: Control automático de ganancia de un amplificador; aprendizaje a partir de la potencia deseada y de la potencia a la salida del amplificador.

¿Por qué soluciones adaptativas?

Para aquellas aplicaciones del filtro de Wiener en las que

- Sistema variante: La estadística varía con el tiempo (por ejemplo, estimación/ecualización de canales variantes en el tiempo).
- Sistema invariante: convergencia secuencial a la solución de Wiener. Se desee implementar soluciones de bajo coste, sin inversión de matrices.



Método de gradiente

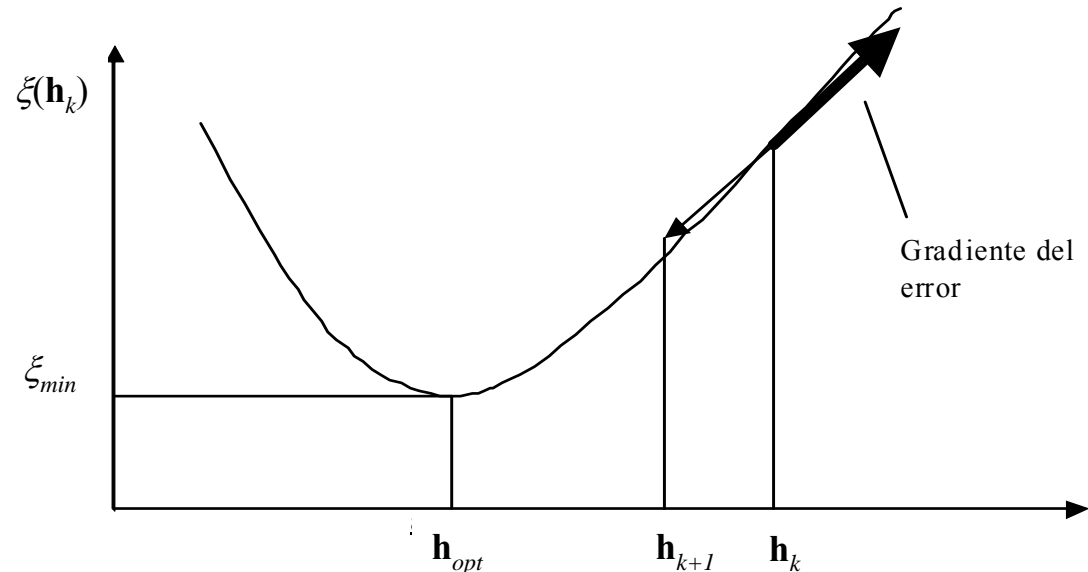
$$\xi(\mathbf{h}_k) = E\left\{\left|d(n) - \mathbf{h}_k^H \mathbf{x}(n)\right|^2\right\} = P_d + \mathbf{h}_k^H \mathbf{R} \mathbf{h}_k - \mathbf{h}_k^H \mathbf{r}_{dx} - \mathbf{r}_{dx}^H \mathbf{h}_k$$

$$\nabla \xi = \frac{\partial \xi(\mathbf{h}_k)}{\partial \mathbf{h}_k^*} = \mathbf{R} \mathbf{h}_{opt} - \mathbf{r}_{dx} = \mathbf{0} \quad \Rightarrow \quad \mathbf{h}_{opt} = \mathbf{R}^{-1} \mathbf{r}_{dx}$$

$$\xi(\mathbf{h}_k) = \xi_{min} + (\mathbf{h}_k - \mathbf{h}_{opt})^H \mathbf{R} (\mathbf{h}_k - \mathbf{h}_{opt}) \quad \xi_{min} = P_d - \mathbf{r}_{dx}^H \mathbf{R}^{-1} \mathbf{r}_{dx}$$

- Dependencia cuadrática del error con los coeficientes: existe un único mínimo.

Aprendizaje:
pasar desde un punto \mathbf{h}_k a otro mejor \mathbf{h}_{k+1} , tomando la dirección contraria al gradiente del MSE



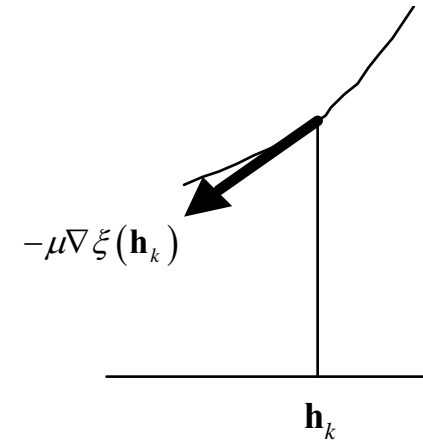
- Regla de aprendizaje:

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \mu \nabla \xi(\mathbf{h}_k)$$

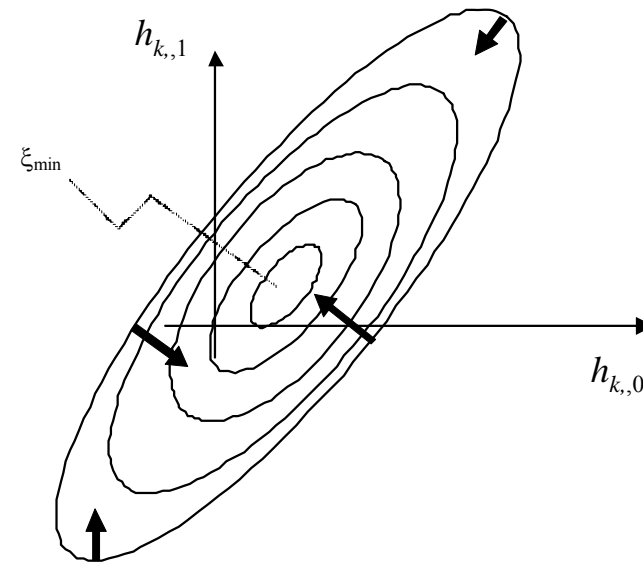


Paso de adaptación

determina la velocidad de aprendizaje



- Las curvas de nivel representan las líneas de igual error cuadrático medio
- El gradiente toma distintas direcciones dependiendo del punto



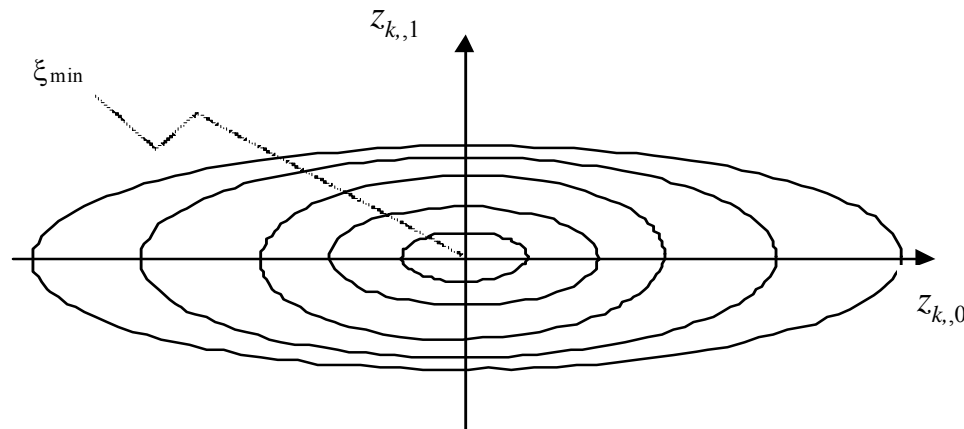
- Se realiza un cambio de variable

$$\xi(\mathbf{h}_k) = \xi_{\min} + (\mathbf{h}_k - \mathbf{h}_{opt})^H \mathbf{R} (\mathbf{h}_k - \mathbf{h}_{opt}) \quad \longrightarrow \quad \xi(\tilde{\mathbf{h}}_k) = \xi_{\min} + \tilde{\mathbf{h}}_k^H \mathbf{R} \tilde{\mathbf{h}}_k$$

- Las curvas de nivel quedan centradas en el origen
- Gradiente: $\nabla \xi(\tilde{\mathbf{h}}_k) = \mathbf{R} \tilde{\mathbf{h}}_k$
- El gradiente en los extremos de los ejes de las curvas de nivel pasa por el origen de coordenadas, es decir es de la forma: $m \tilde{\mathbf{h}}_k$
- Descomposición en autovectores y autovalores de la matriz de correlación

$$\xi = \xi_{\min} + \tilde{\mathbf{h}}_k^H \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H \tilde{\mathbf{h}}_k = \xi_{\min} + \mathbf{z}_k^H \mathbf{\Lambda} \mathbf{z}_k = \xi_{\min} + \sum_{i=1}^Q \lambda_i |z_{k,i}|^2$$

- Los ejes de las curvas de error en las nuevas coordenadas \mathbf{z}_k quedan alineados con los ejes de representación
- La curvatura buscada se obtiene de calcular la segunda derivada de la función de error respecto a $z_{k,i}$, siendo igual a $2\lambda_i$.
- La dirección del eje corto estará asociada al autovector de mayor autovalor, y viceversa.
- La excentricidad de las curvas de nivel de la función de error dependerá de lo distintos que sean los autovalores de la matriz de correlación.



Ejemplo 1. Función potencia del error para un filtro FIR de orden 2

$$\begin{aligned}\xi(h_0, h_1) &= r_{dd}(0) - 2 \operatorname{Re} \left(\sum_{k=0}^1 h_k^* r_{dx}(k) \right) + \sum_{k=0}^1 \sum_{i=0}^1 h_i h_k^* r_{xx}(k-i) = \\ &= r_{dd}(0) - 2 \operatorname{Re}(\mathbf{h}^H \mathbf{r}_{dx}) + \mathbf{h}^H \mathbf{R}_{xx} \mathbf{h}\end{aligned}$$

Caso 1: $\mathbf{R}_{xx} = \begin{bmatrix} 1.1 & 0.1 \\ 0.1 & 1.1 \end{bmatrix}$ $\mathbf{r}_{dx} = \begin{bmatrix} 0.5272 \\ -0.4458 \end{bmatrix}$ $r_{dd}(0) = 0.9486$

$$\mathbf{h} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{dx} = \begin{bmatrix} 0.5204 \\ -0.4526 \end{bmatrix}$$

Caso 2: $\mathbf{R}_{xx} = \begin{bmatrix} 40 & 39 \\ 39 & 40 \end{bmatrix}$ $\mathbf{r}_{dx} = \begin{bmatrix} 0.5272 \\ -0.4458 \end{bmatrix}$ $r_{dd}(0) = 0.9486$

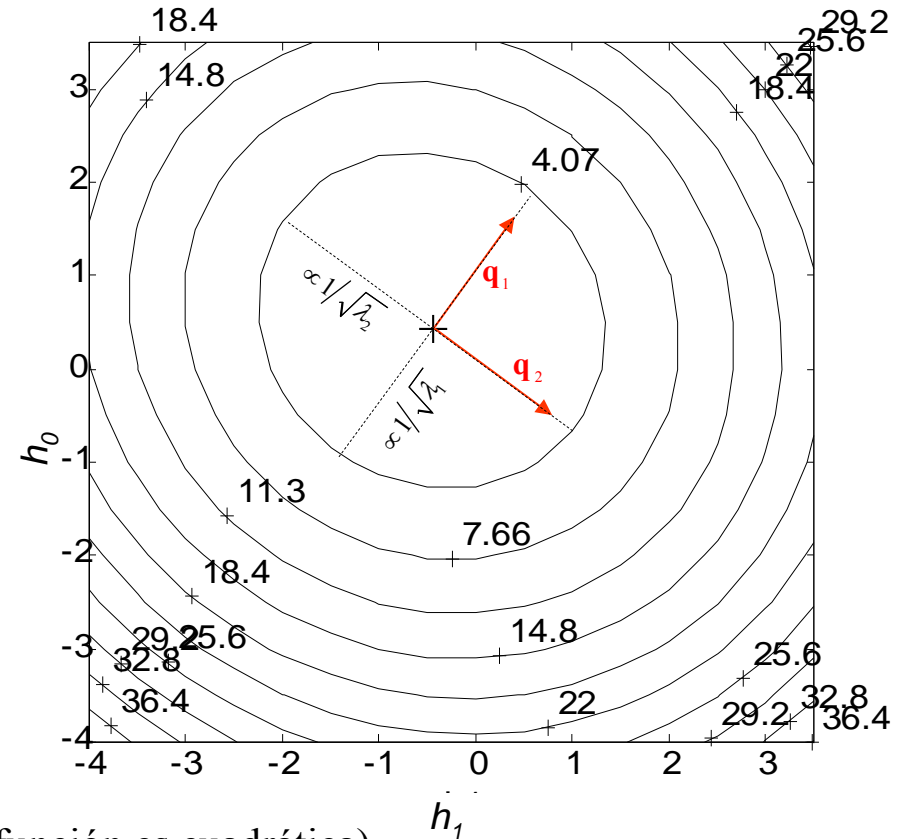
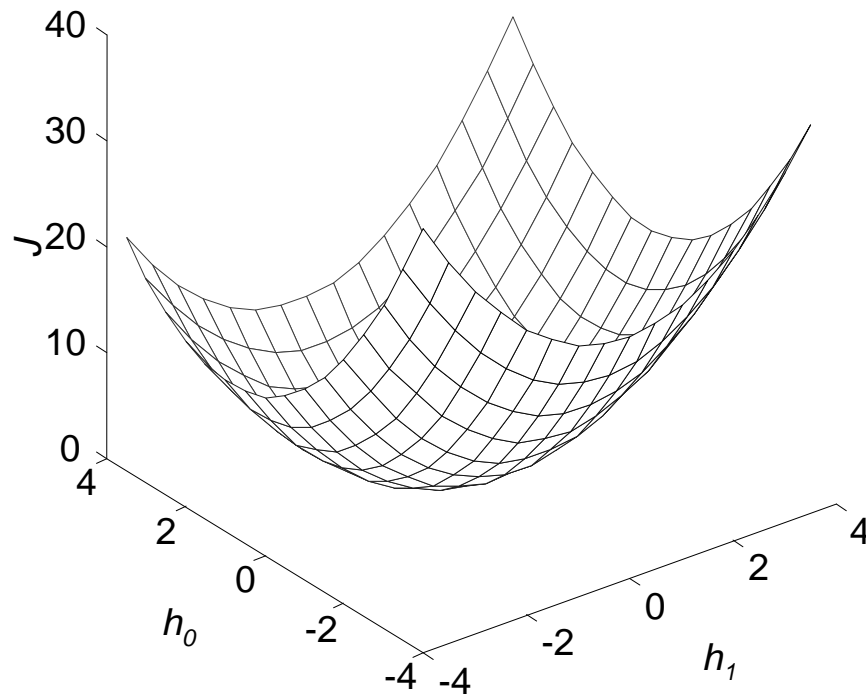
$$\mathbf{h} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{dx} = \begin{bmatrix} 0.487 \\ -0.486 \end{bmatrix}$$

Superficie de error (I)

Caso 1: Poca dispersión de autovalores

$$\lambda_1 = 1.2$$

$$\lambda_2 = 1.0$$



Existe un único punto singular y es un mínimo (la función es cuadrática).

Las secciones son elipses. El tamaño de los semiejes es proporcional al inverso de la raíz cuadrada de los autovalores.

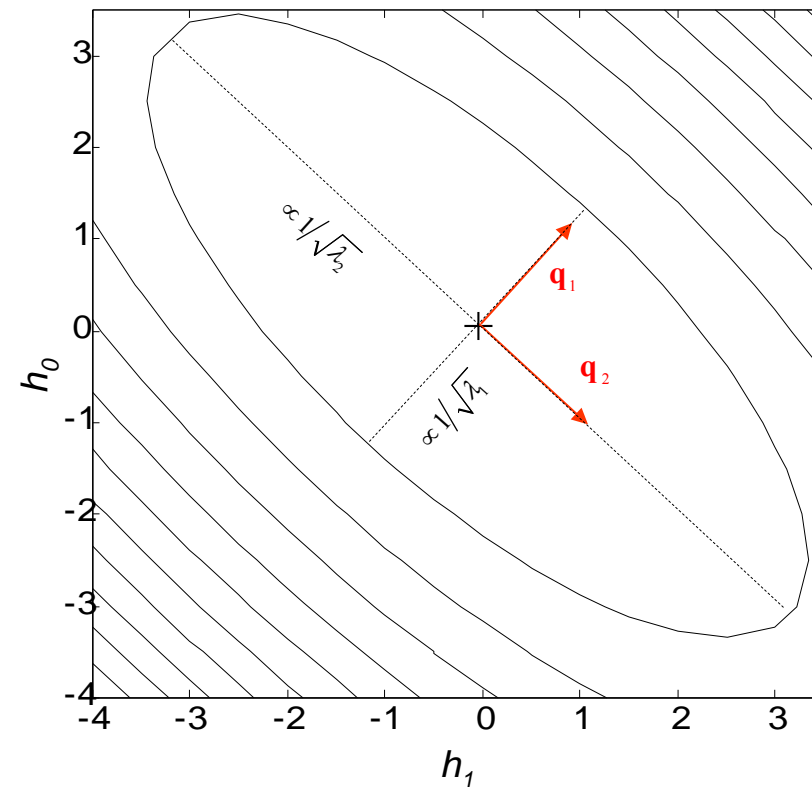
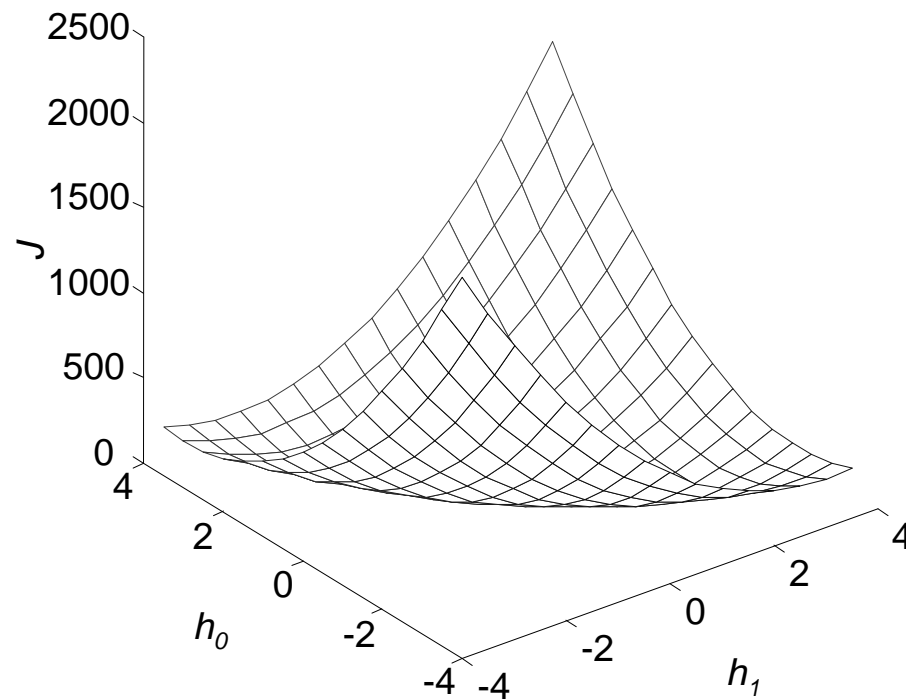
Los ejes de las elipses coinciden con las direcciones de los autovectores de la matriz de correlación.

Superficie de error (II)

Caso 2: Gran dispersión de autovalores

$$\lambda_1 = 79$$

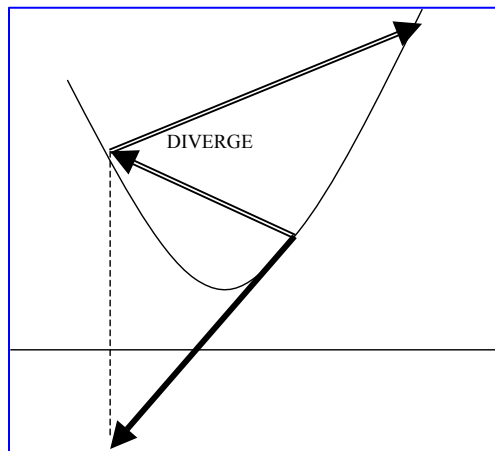
$$\lambda_2 = 1$$



- Los algoritmos denominados de gradiente emplean el gradiente del MSE para proporcionar la actualización de los coeficientes

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \mu \nabla \xi_k = \mathbf{h}_k - \mu (\mathbf{R}\mathbf{h}_k - \mathbf{p})$$

- Si el paso de adaptación es excesivamente grande el algoritmo diverge



$$\mathbf{h}_{k+1} = \mathbf{h}_k - \mu \nabla \xi(\mathbf{h}_k)$$

$$\xi(\mathbf{h}_k) = E \left\{ \left| d(n) - \mathbf{h}_k^H \mathbf{x}(n) \right|^2 \right\} = P_d + \mathbf{h}_k^H \mathbf{R} \mathbf{h}_k - \mathbf{h}_k^H \mathbf{r}_{dx} - \mathbf{r}_{dx}^H \mathbf{h}_k$$

$$\nabla \xi(\mathbf{h}_k) = \mathbf{R} \mathbf{h}_k - \mathbf{r}_{dx}$$

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \mu (\mathbf{R} \mathbf{h}_k - \mathbf{r}_{dx}) = (\mathbf{I} - \mu \mathbf{R}) \mathbf{h}_k + \mu \mathbf{r}_{dx}$$

- Restando el vector de coeficientes óptimo

$$\mathbf{h}_{k+1} - \mathbf{h}_{opt} = (\mathbf{I} - \mu \mathbf{R}) \mathbf{h}_k + \mu \mathbf{R} \mathbf{h}_{opt} - \mathbf{h}_{opt} = (\mathbf{I} - \mu \mathbf{R}) (\mathbf{h}_k - \mathbf{h}_{opt})$$

- Considerando la descomposición de la matriz de correlación en autovectores y autovalores y $\mathbf{Q} \mathbf{Q}^H = \mathbf{I}$

$$\mathbf{h}_{k+1} - \mathbf{h}_{opt} = (\mathbf{I} - \mu \mathbf{Q} \mathbf{\Lambda} \mathbf{Q}^H) (\mathbf{h}_k - \mathbf{h}_{opt}) = \mathbf{Q} (\mathbf{I} - \mu \mathbf{\Lambda}) \mathbf{Q}^H (\mathbf{h}_k - \mathbf{h}_{opt})$$

- Aplicando un cambio de variable $\mathbf{h}_k = \mathbf{Q}^H (\mathbf{h}_k - \mathbf{h}_{opt})$ y multiplicando por \mathbf{Q}^H

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

- **Convergencia:** cuando \mathbf{z}_k tiende al vector nulo.
- Para cada elemento del vector \mathbf{z}_k se tiene:

$$z_{k+1,i} = (1 - \mu\lambda_i) z_{k,i} = (1 - \mu\lambda_i)^2 z_{k-1,i} = \dots = (1 - \mu\lambda_i)^{k+1} z_{0,i}$$



$$|1 - \mu\lambda_i| < 1 \quad \forall i = 1, \dots, Q$$

- Matriz de correlación definida positiva, por tanto autovalores positivos

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

- Otra posible cota puede definirse a partir de la traza de \mathbf{R} , igual a la suma de los autovalores y por tanto mayor que el autovalor máximo

$$\mu < \frac{2}{\text{Traza}(\mathbf{R})} < \frac{2}{\lambda_{\max}}$$

- La constante de tiempo efectiva del sistema viene determinada por la respuesta transitoria que más tiempo tarda en atenuarse.
- La constante de tiempo vendrá determinada por el número de iteraciones necesarias para que el elemento de \mathbf{z} correspondiente al mínimo autovalor haya convergido

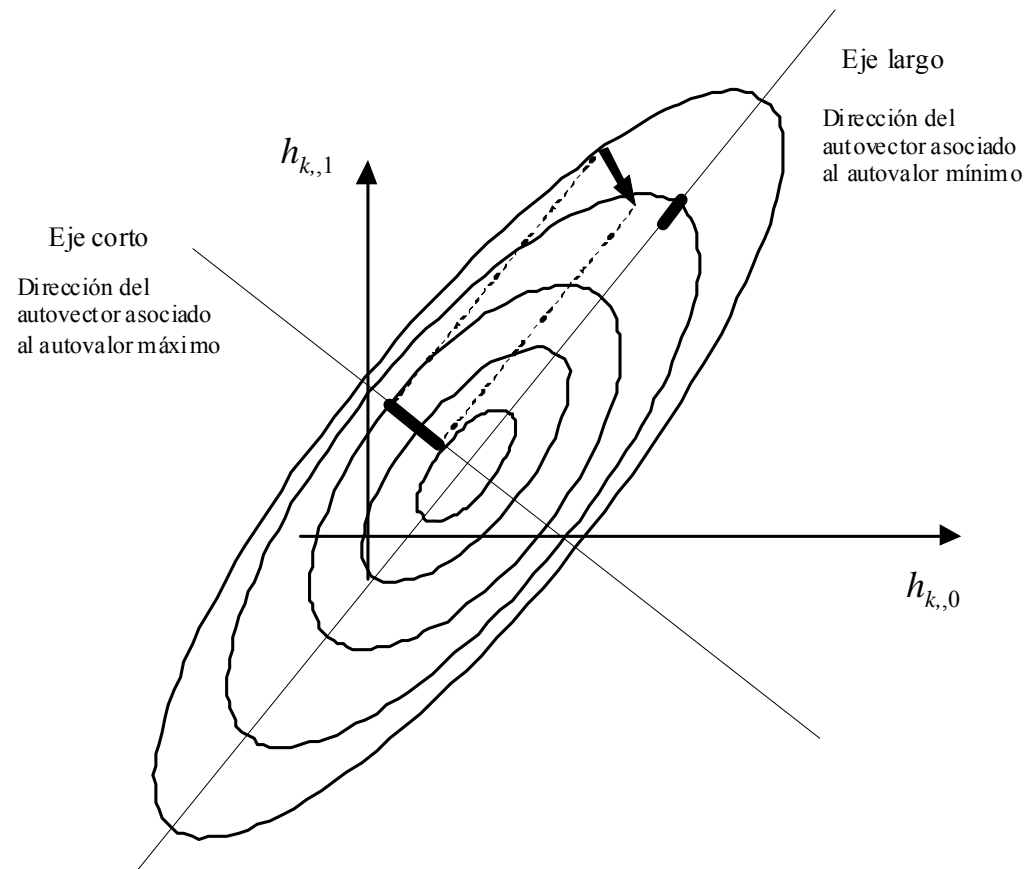
$$(1 - \mu\lambda_{\min})^{N_c} = 0,1 \Rightarrow N_c \ln(1 - \mu\lambda_{\min}) = -2.3$$
$$N_c = \frac{-2.3}{\ln(1 - \mu\lambda_{\min})} = \frac{-2.3}{\ln\left(1 - \frac{2\alpha}{\lambda_{\max}}\lambda_{\min}\right)}$$

- Si la dispersión de autovalores es mucho mayor que α

$$\ln(1+x) \approx x \quad x \ll 1 \qquad N_c = \frac{2.3}{2\alpha} \frac{\lambda_{\max}}{\lambda_{\min}}$$

- A mayor dispersión de autovalores y a menor α , menor velocidad de convergencia

- El autovalor mínimo limita la velocidad de convergencia, es decir, los modos débiles asociados a autovalores pequeños tardan más en converger que los fuertes
- El filtro aprende más rápido en los modos de alta energía, dejando para el final los de menor energía.

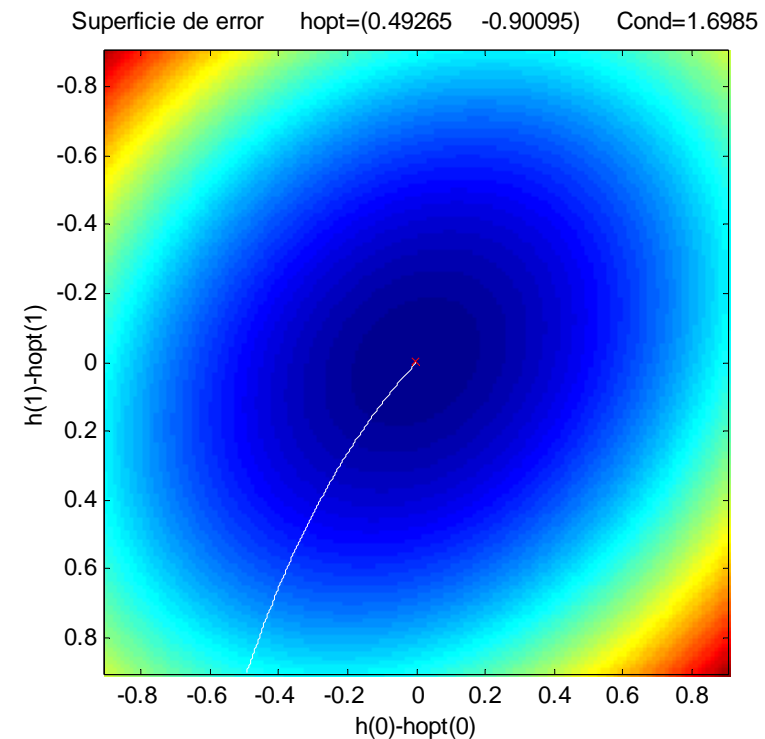
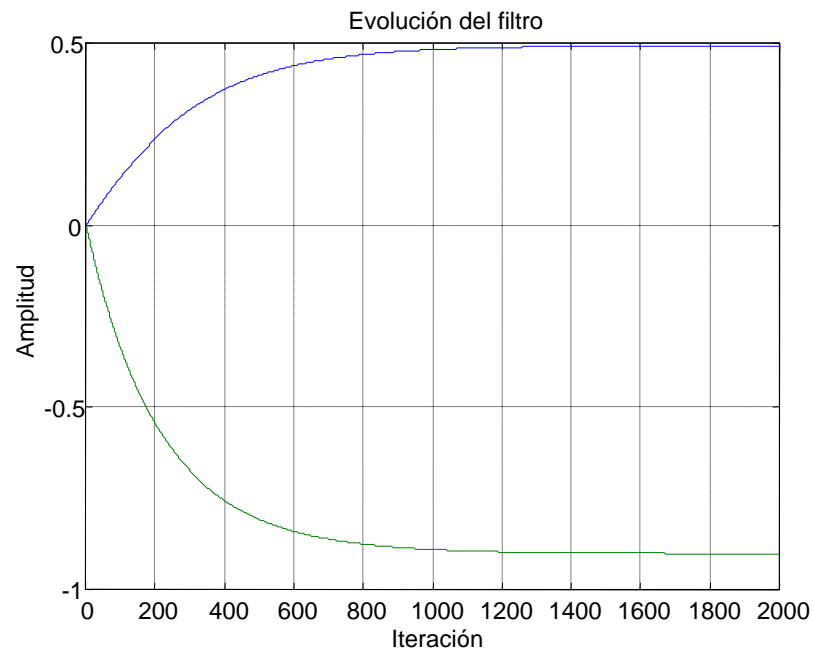


Ejemplo 2: predicción lineal de un proceso AR(2)

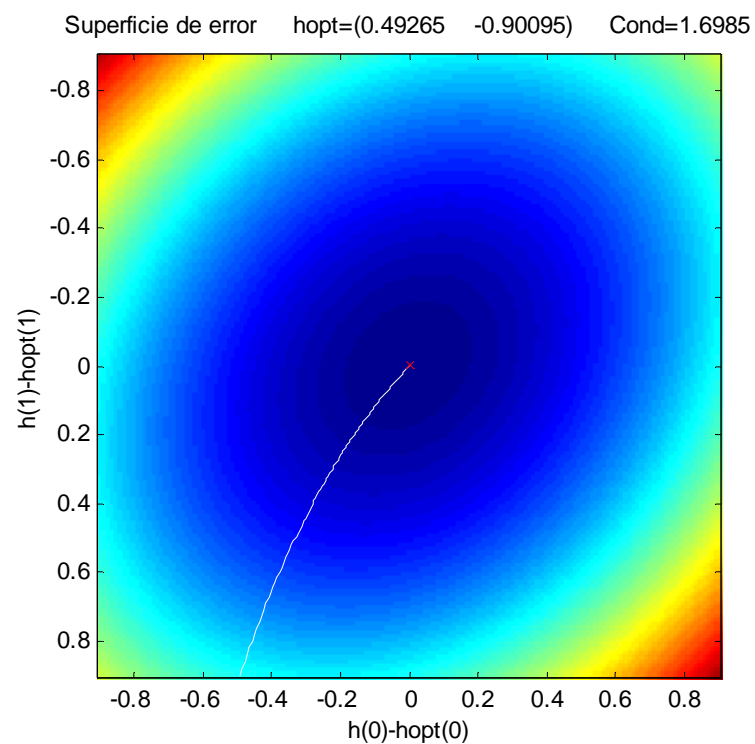
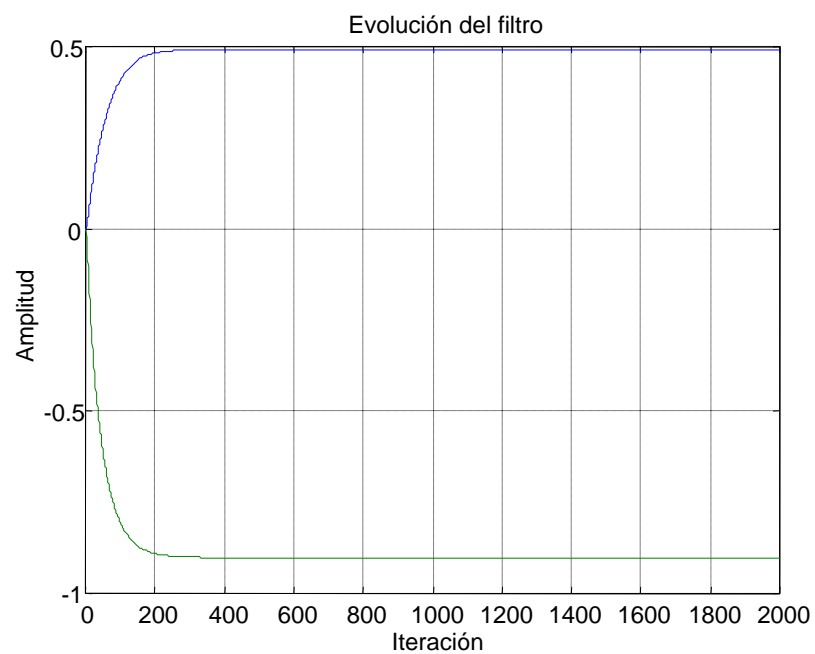
$$A(z) = 1 - 0.5z^{-1} + 0.9z^{-2} \quad \lambda_{\max} = 7.39$$

$$\lambda_{\max} / \lambda_{\min} = 1.71$$

$$\mathbf{h}_{opt} = [0.5 \quad -0.9]$$



$$\mu = 0,001$$



$\mu=0,005$

Filtrado adaptativo

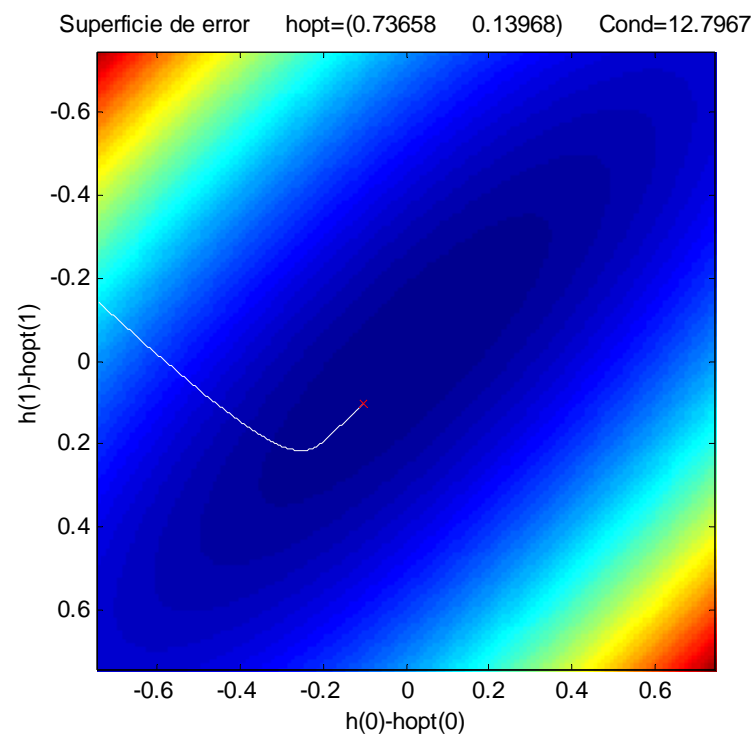
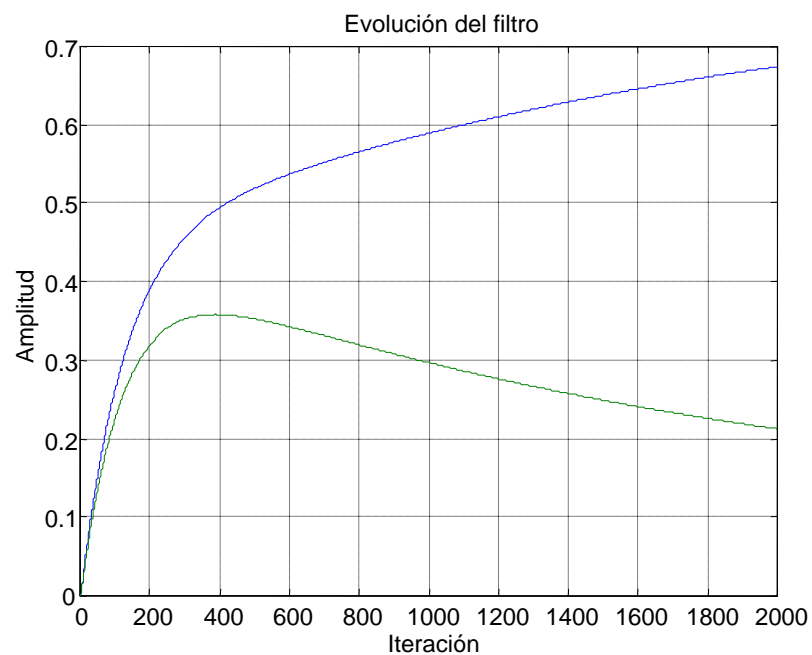
Ejemplo 3: predicción lineal de un proceso AR(2)

$$A(z)=1-0.75z^{-1}-0.125z^{-2}$$

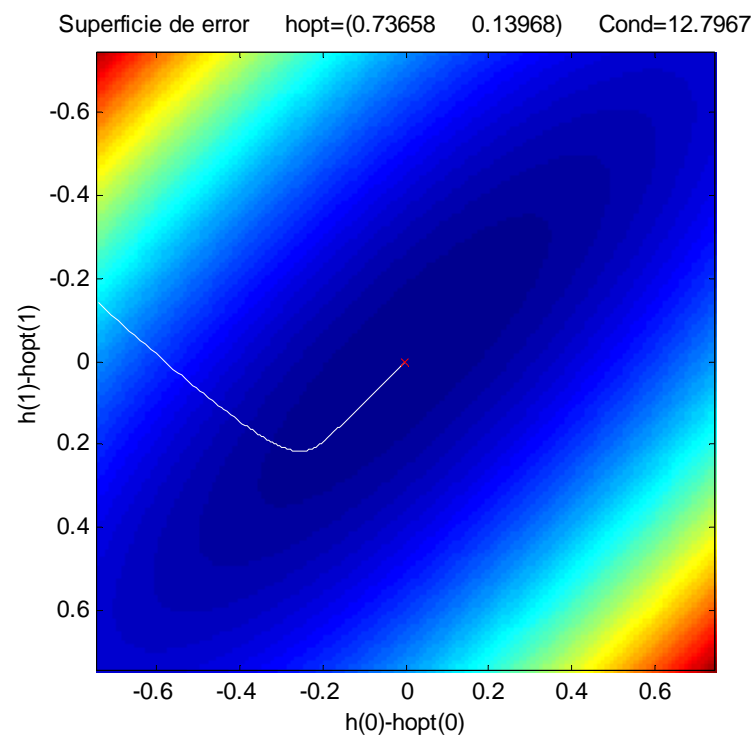
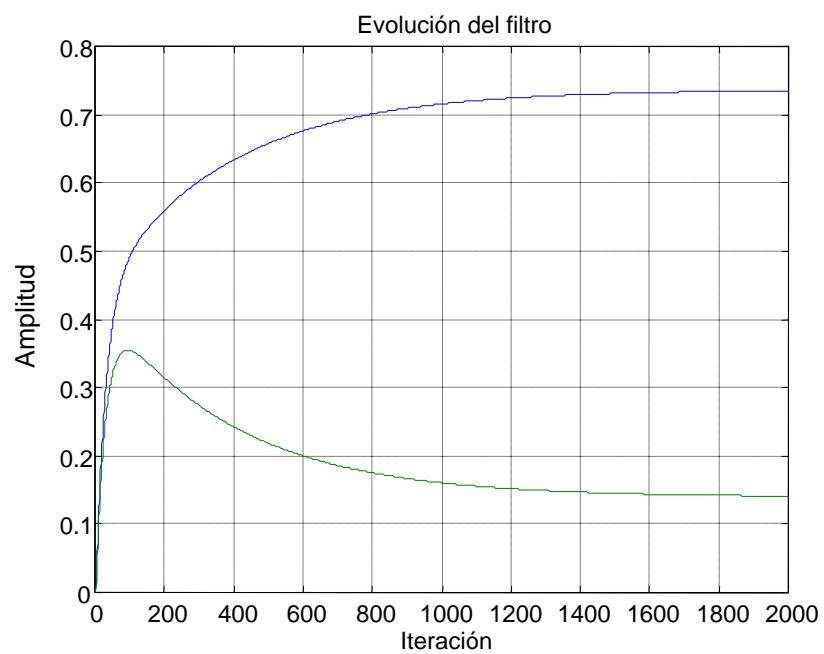
$$\lambda_{\max}=6.8$$

$$\lambda_{\max}/\lambda_{\min}=13$$

$$\mathbf{h}_{opt}=[0.75 \quad 0.125]$$



$$\mu=0,001$$

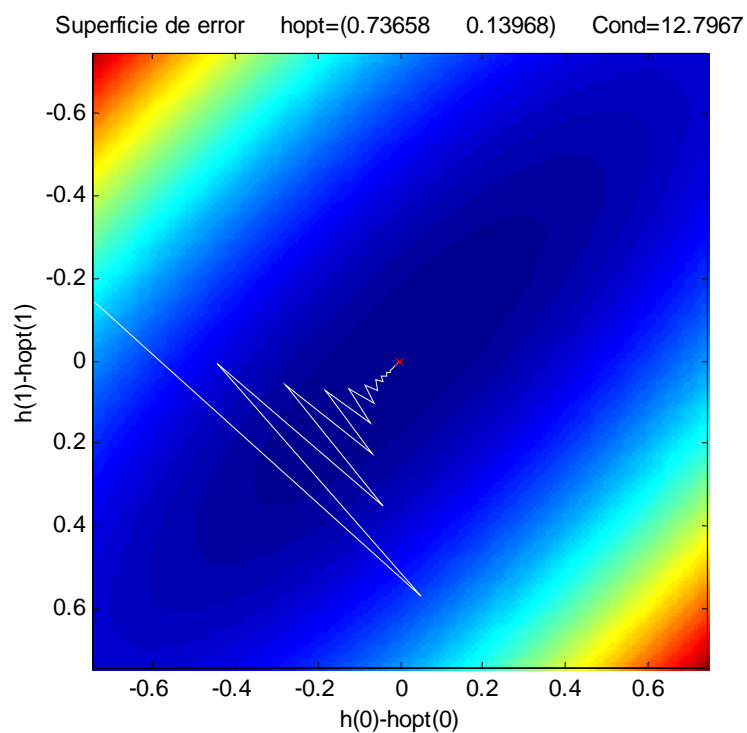


$\mu=0,005$

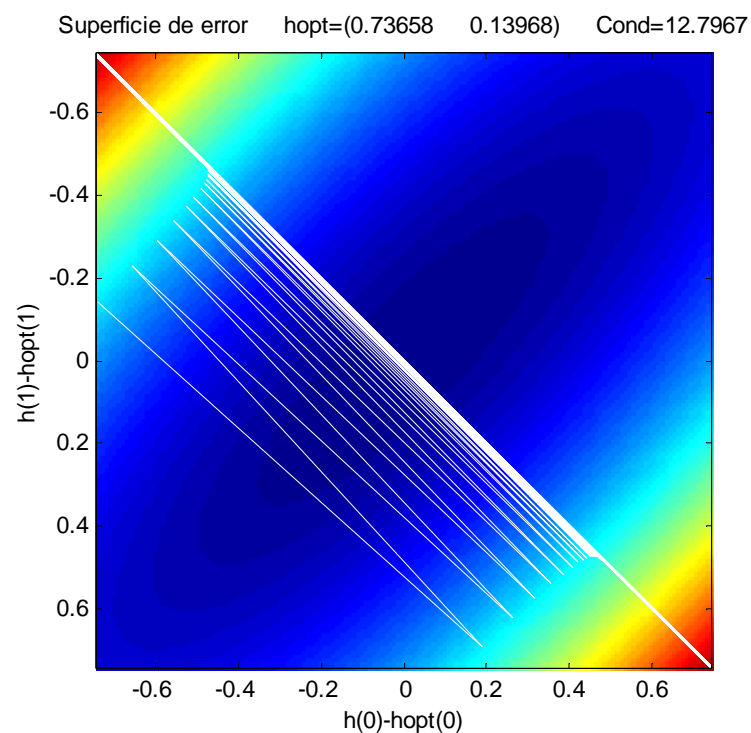
Ejemplo 4: Límite de la convergencia dependiendo de μ

Cotas: $2/(M \cdot r_x(0)) = 0.2714$

$2/\lambda_{\max} = 0.2930$



$\mu=0,25$



$\mu=0,2934$

Algoritmo de gradiente estocástico: LMS

- Estimación del gradiente:

$$\nabla \xi(n) = \mathbf{R}\mathbf{h}_k - \mathbf{p} = E\{\mathbf{x}(n)\mathbf{x}(n)^H\}\mathbf{h}_k - E\{d^*(n)\mathbf{x}(n)\}$$

- Se aproxima los valores esperados por sus valores instantáneos

$$\mathbf{h}_{k+1} = \mathbf{h}_k - \mu(\mathbf{x}(n)\mathbf{x}(n)^H \mathbf{h}_k - d^*(n)\mathbf{x}(n)) = \mathbf{h}_k - \mu\mathbf{x}(n)(y^*(n) - d^*(n)) = \mathbf{h}_k + \mu\mathbf{x}(n)e^*(n)$$

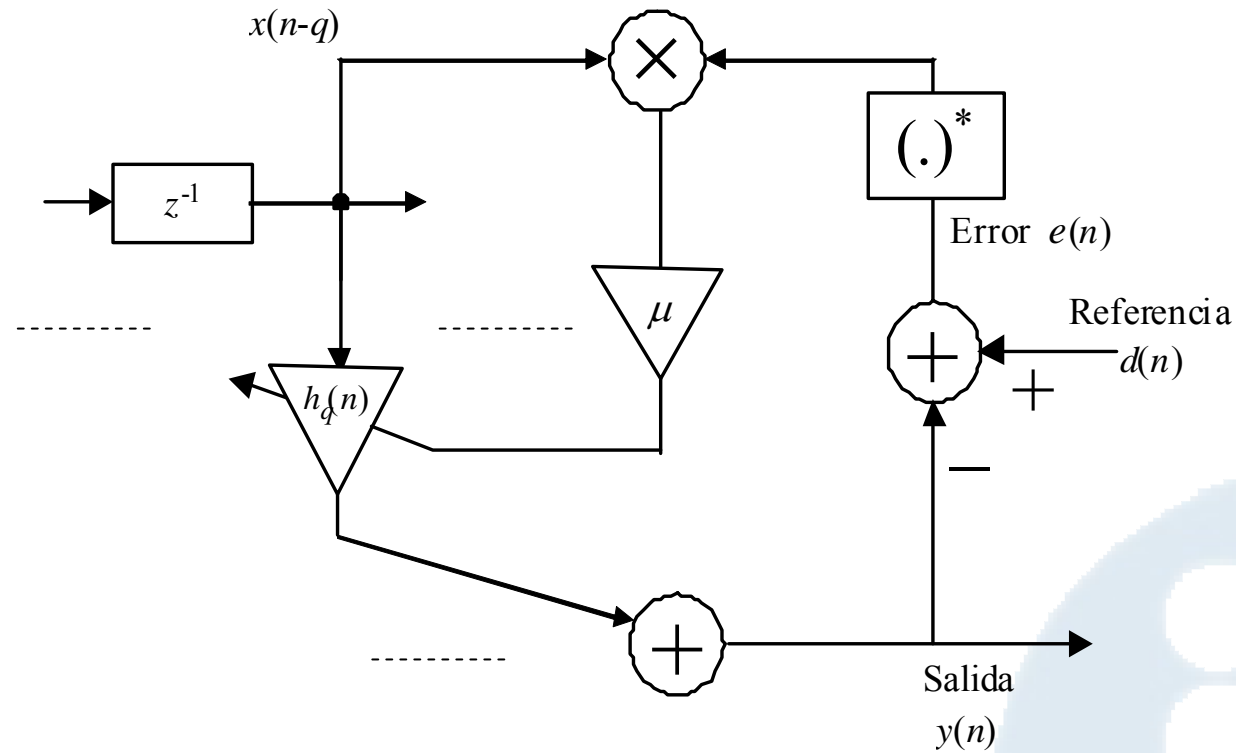
↑

$$y(n) = \mathbf{h}_k^H \mathbf{x}(n)$$

- Para cada valor de n podemos realizar iteraciones en k , aunque lo habitual es tomar $n = k$:

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu\mathbf{x}(n)e^*(n)$$

Implementación del LMS para el coeficiente q del filtro



Convergencia en media del LMS

$$\mathbf{h}(n+1) = \mathbf{h}(n) + \mu e^*(n) \mathbf{x}(n)$$

$$\mathbf{h}(n+1) - \mathbf{h}_{opt} = \mathbf{h}(n) - \mathbf{h}_{opt} + \mu e^*(n) \mathbf{x}(n)$$

Usando la expresión del error:

$$\begin{aligned} e(n) &= d(n) - \mathbf{h}^H(n) \mathbf{x}(n) = \\ &= e_{opt}(n) + \mathbf{h}_{opt}^H(n) \mathbf{x}(n) - \mathbf{h}^H(n) \mathbf{x}(n) = \\ &= e_{opt}(n) - \left(\mathbf{h}(n) - \mathbf{h}_{opt}(n) \right)^H \mathbf{x}(n) \end{aligned}$$

$$\mathbf{h}(n+1) - \mathbf{h}_{opt} = \left(\mathbf{I} - \mu \mathbf{x}(n) \mathbf{x}^H(n) \right) \left(\mathbf{h}(n) - \mathbf{h}_{opt} \right) + \mu e_{opt}^*(n) \mathbf{x}(n)$$

Obtengamos la esperanza matemática de esta expresión suponiendo que $\mathbf{x}(n)$ y $\mathbf{h}(n)$ son estadísticamente independientes:

$$E \left\{ \mathbf{h}(n+1) - \mathbf{h}_{opt} \right\} = \left(\mathbf{I} - \mu \mathbf{R}_x \right) E \left\{ \mathbf{h}(n) - \mathbf{h}_{opt} \right\} + \mu E \left\{ e_{opt}^*(n) \mathbf{x}(n) \right\}$$

La condición sobre el valor de μ que garantiza la convergencia del método de gradiente nos sirve para demostrar la convergencia en media de los coeficientes obtenidos mediante el LMS.

Es cero en el óptimo
por el principio de
ortogonalidad

La evolución de los coeficientes es “ruidosa” debido a que el gradiente está siendo estimado.

La condición de convergencia (esta vez en valor medio de \mathbf{h}) es la misma que para el algoritmo de gradiente:

$$0 < \mu \leq \frac{2}{M \cdot r_x(0)} \leq \frac{2}{\lambda_{\max}}$$

Convergencia en el caso no estacionario: utilizando un valor de μ fijo

- si se produce un **incremento de la potencia de señal**, el valor máximo del paso de adaptación requerido para la convergencia se reduce → posible divergencia
- si se produce un **decremento de la potencia de señal**, el valor máximo del paso de adaptación requerido para la convergencia aumenta → reducción en la velocidad de convergencia.

NLMS (Normalizado)

En condiciones no estacionarias es conveniente estimar la potencia de forma adaptativa a fin de garantizar la convergencia:

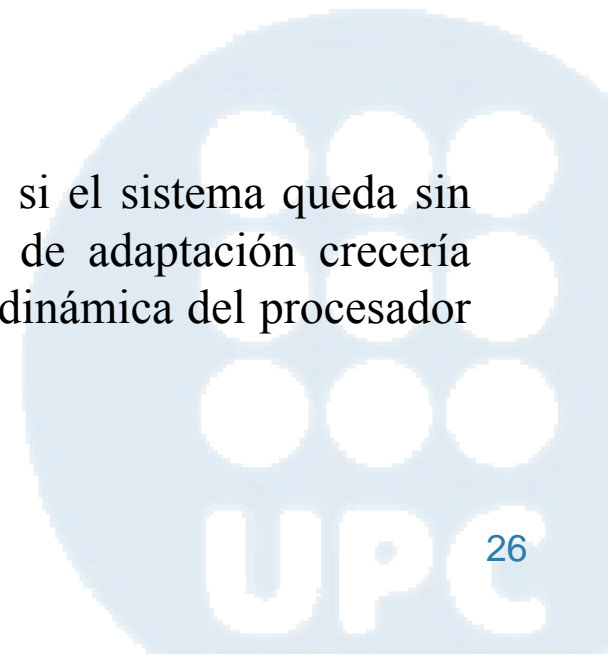
Método 1: $M \cdot \hat{r}_x(0; n) = \mathbf{x}^H(n) \mathbf{x}(n)$

Método 2: $\hat{r}_x(0; n) = \gamma \cdot \hat{r}_x(0; n-1) + (1 - \gamma) |x(n)|^2 \quad 0 < |\gamma| < 1$

De esta forma se define el **LMS normalizado** como aquel en el que:

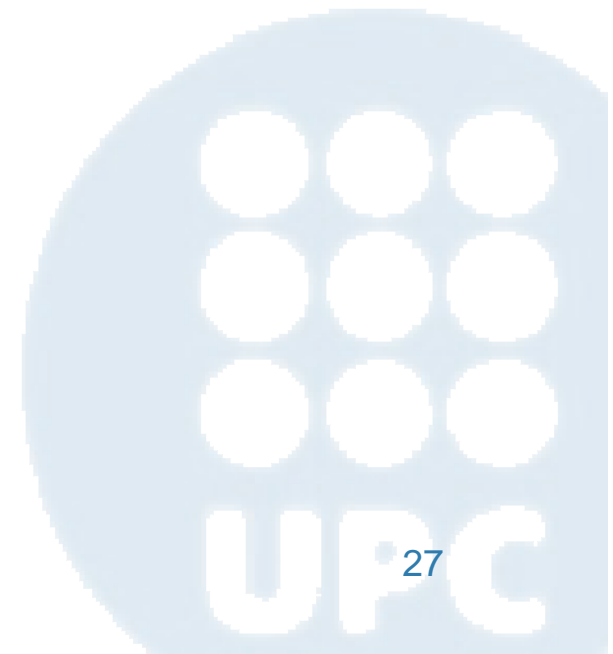
$$\mu(n) = \frac{\alpha}{P_{x,o} + M \cdot \hat{r}_x(0; n)} \quad 0 < \alpha < 2$$

$P_{x,o}$ es un umbral mínimo para el estimador de potencia: si el sistema queda sin señal de entrada o esta se hace muy pequeña, el paso de adaptación crecería mucho y los pesos también, produciendo saturación en la dinámica del procesador de los coeficientes.



Covarianza de los coeficientes del LMS

- El gradiente instantáneo es una variable aleatoria y no el gradiente determinista.
- El aprendizaje no finaliza, alcanza un mínimo nervioso alrededor del óptimo.
- Los coeficientes son una variable aleatoria cuya media es el filtro óptimo y cuya varianza debe determinarse



$$\begin{aligned}\mathbf{h}(n+1) - \mathbf{h}_{opt} &= \mathbf{h}(n) - \mathbf{h}_{opt} + \mu e^*(n) \mathbf{x}(n) = \\ &= (\mathbf{I} - \mu \mathbf{x}(n) \mathbf{x}^H(n)) (\mathbf{h}(n) - \mathbf{h}_{opt}) + \mu e_{opt}^*(n) \mathbf{x}(n)\end{aligned}$$

La matriz de covarianza de los coeficientes viene dada por:

$$\begin{aligned}\mathbf{S}(n+1) &= E \left\{ (\mathbf{h}(n+1) - \mathbf{h}_{opt}) (\mathbf{h}(n+1) - \mathbf{h}_{opt})^H \right\} = \\ &= E \left\{ (\mathbf{I} - \mu \mathbf{x}(n) \mathbf{x}^H(n)) (\mathbf{h}(n) - \mathbf{h}_{opt}) (\mathbf{h}(n) - \mathbf{h}_{opt})^H (\mathbf{I} - \mu \mathbf{x}(n) \mathbf{x}^H(n))^H \right\} + \mu^2 \xi_{\min} E \left\{ \mathbf{x}(n) \mathbf{x}^H(n) \right\}\end{aligned}$$

Suponiendo que $\mathbf{x}(n)$ y $\mathbf{h}(n)$ son estadísticamente independientes:

$$\mathbf{S}(n+1) = (\mathbf{I} - \mu \mathbf{R}_x) \mathbf{S}(n) (\mathbf{I} - \mu \mathbf{R}_x)^H + \mu^2 \xi_{\min} \mathbf{R}_x$$

En la convergencia \mathbf{S} no dependerá de n :

$$\mathbf{S} = (\mathbf{I} - \mu \mathbf{R}_x) \mathbf{S} (\mathbf{I} - \mu \mathbf{R}_x)^H + \mu^2 \xi_{\min} \mathbf{R}_x$$

$$0 = \mu \mathbf{R}_x \mathbf{S} - \mathbf{S} - \mathbf{R}_x \mathbf{S} \mathbf{R}_x^{-1} + \mu \xi_{\min} \mathbf{I}$$

Si $\mu \ll 1/\lambda_{\max}$

$$0 = \mu \xi_{\min} \mathbf{I} - \mathbf{S} - \mathbf{R}_x \mathbf{S} \mathbf{R}_x^{-1}$$

Esta ecuación tiene una solución única:

$$\mathbf{S} = \frac{\mu}{2} \xi_{\min} \mathbf{I}$$

En la convergencia los coeficientes varían de forma independiente

Convergencia y desajuste

El comportamiento errático de los coeficientes provoca un incremento en la potencia mínima del error:

$$\xi(\mathbf{h}(n)) = \xi_{\min} + \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right)^H \mathbf{R}_x \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right)$$

La potencia de ese incremento viene dada por:

$$\begin{aligned} \Delta \xi &= E \left\{ \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right)^H \mathbf{R}_x \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right) \right\} = E \left\{ \text{traza} \left(\left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right)^H \mathbf{R}_x \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right) \right) \right\} = \\ &= E \left\{ \text{traza} \left(\mathbf{R}_x \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right) \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right)^H \right) \right\} = \text{traza} \left(\mathbf{R}_x E \left\{ \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right) \left(\mathbf{h}(n+1) - \mathbf{h}_{opt} \right)^H \right\} \right) = \\ &= \text{traza} \left(\mathbf{R}_x \mathbf{S}(n+1) \right) \end{aligned}$$

En la convergencia:

$$\Delta \xi = \frac{\mu}{2} \xi_{\min} \cdot \text{traza}(\mathbf{R}_x) = \frac{\mu}{2} \xi_{\min} \cdot M \cdot r_x(0)$$

El incremento en la potencia mínima se mide con el desajuste:

$$\mathcal{M} = \frac{\Delta \xi}{\xi_{\min}} = \frac{\mu}{2} \cdot M \cdot r_x(0) = \frac{\alpha}{2}$$

Compromiso entre
velocidad de
convergencia y
desajuste

LMS: solución adaptativa

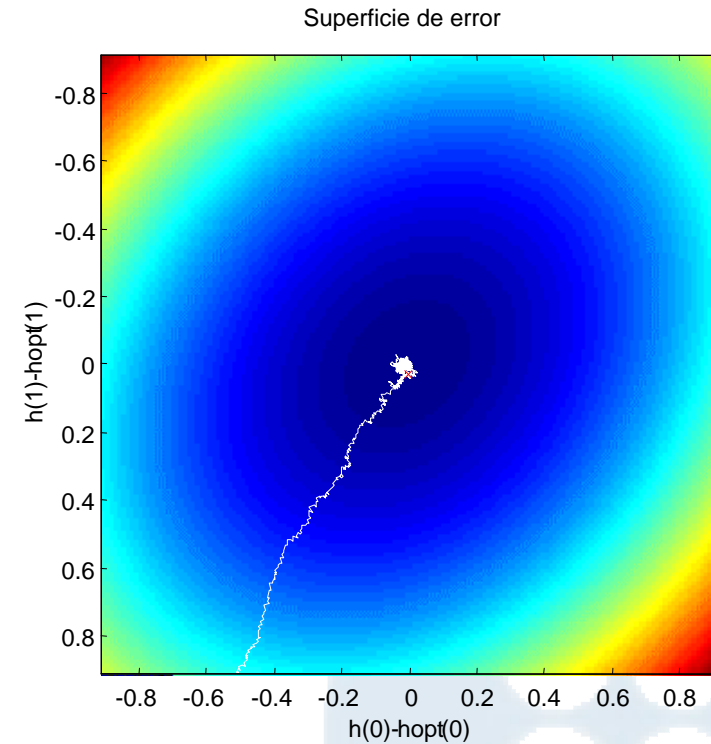
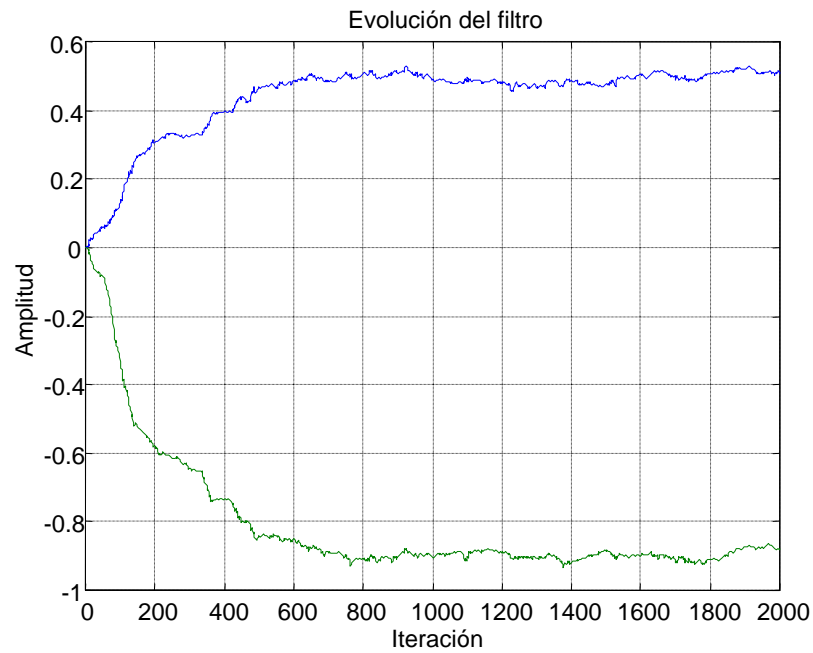
Ejemplo 2: predicción lineal LMS de un proceso AR(2)

$$A(z) = 1 - 0.5z^{-1} + 0.9z^{-2}$$

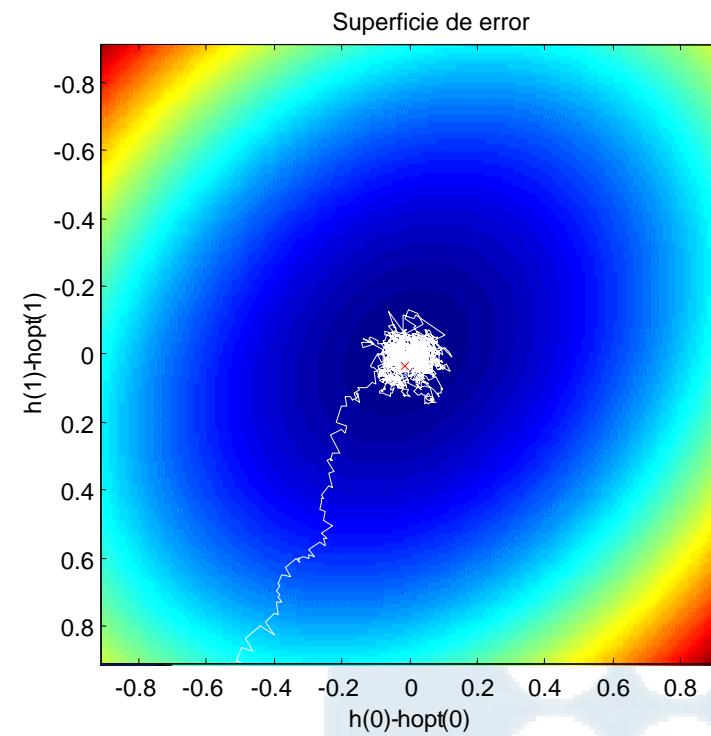
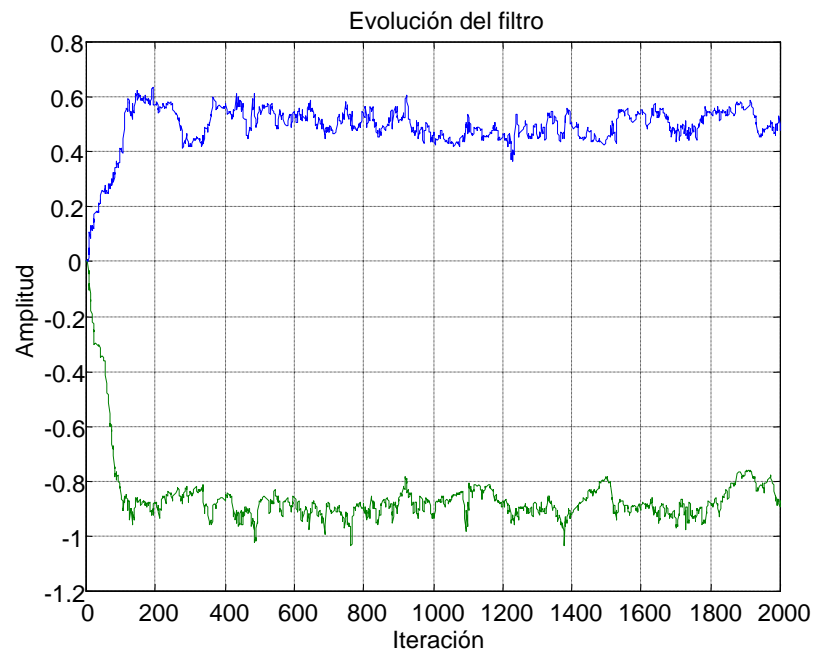
$$\lambda_{\max} = 7.39$$

$$\lambda_{\max} / \lambda_{\min} = 1.71$$

$$\mathbf{h}_{\text{opt}} = [0.5 \quad -0.9]$$



$$\mu = 0.001$$



$\mu=0,005$

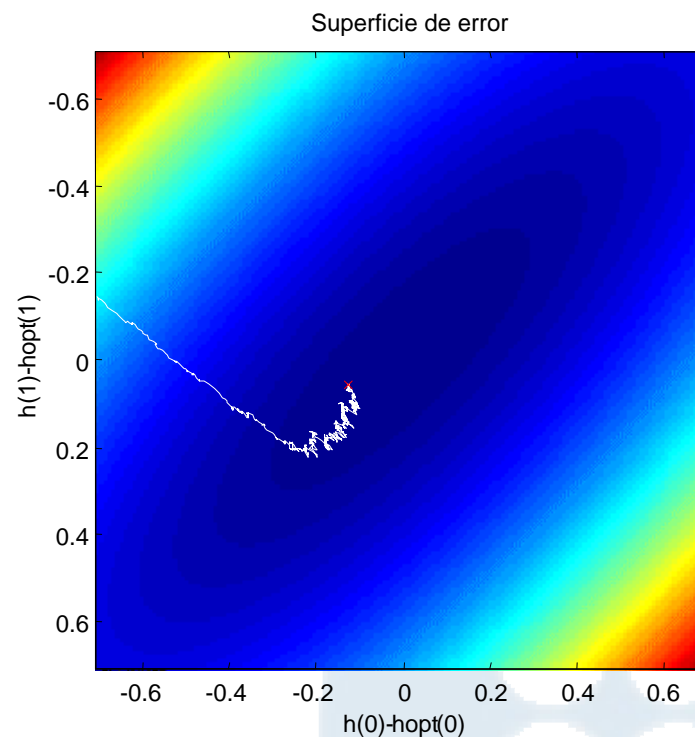
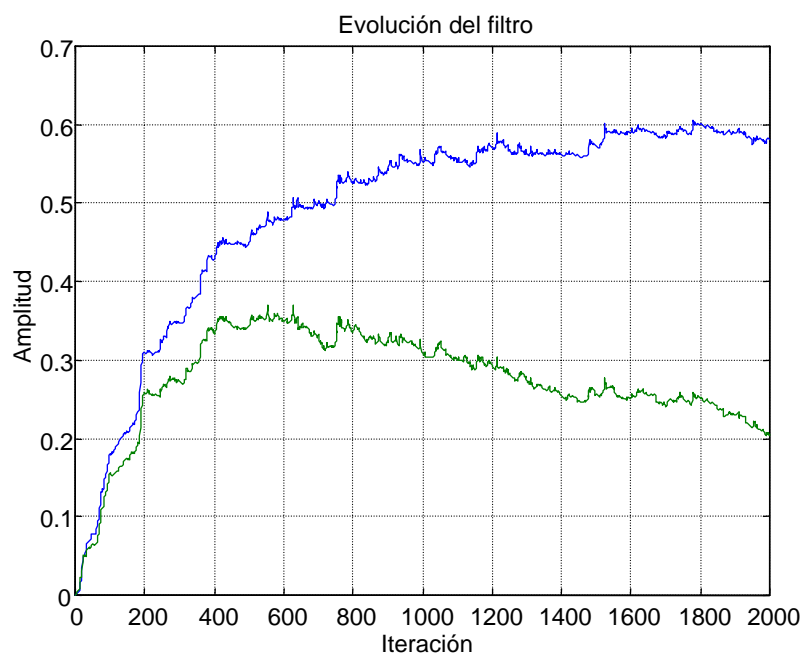
Ejemplo 3: predicción lineal de un proceso AR(2)

$$A(z)=1-0.75z^{-1}-0.125z^{-2}$$

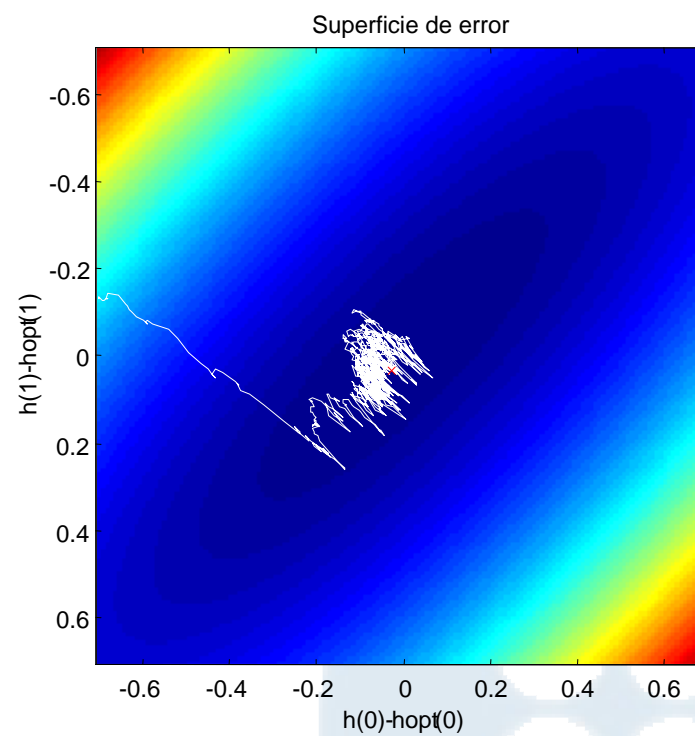
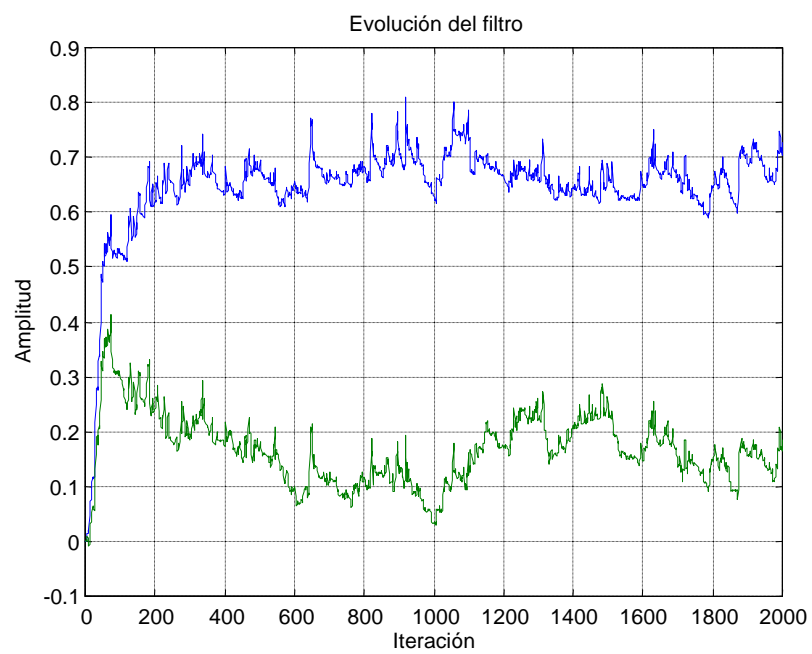
$$\lambda_{\max}=6.8$$

$$\lambda_{\max}/\lambda_{\min}=13$$

$$\mathbf{h}_{\text{opt}}=[0.75 \quad 0.125]$$



$$\mu=0,001$$



$\mu=0,005$