

TESIS DOCTORAL

**Criterios de Teoría de la Información en
Procesado Adaptativo de la Señal
y Aplicaciones**

**Dpt. Teoria del Senyal i Comunicacions
Universitat Politècnica de Catalunya**

*Autor: Josep Sala i Alvarez
Director: Gregori Vázquez i Grau*

4.3 Modelo de Varias Fuentes

El vector de coeficientes solución del criterio estadístico definido en secciones anteriores no es simplemente función de la no-linealidad escogida sino también del tipo de estadística de la señal. En esta sección consideraremos un modelo de señal más elaborado, incluyendo varias fuentes a diferencia de la sección 4.2. Como resultado más importante de esta sección cabe destacar la relación del vector de coeficientes con la función de regeneración de la referencia. Definiremos la señal recibida de la siguiente forma,

$$\mathbf{r}_k = \sum_{i=1}^{i=N_u} a_i^k \mathbf{s}_k + \mathbf{n}_k = \mathbf{S}_k \mathbf{a}_k + \mathbf{n}_k \quad (4.3.1)$$

de forma análoga a como se introdujo en la sección 3.3.2, donde hemos agrupado todas las fuentes en el vector \mathbf{a}_k y los vectores de steering correspondientes en la matriz \mathbf{S}_k . Supondremos que todos los vectores de steering son linealmente independientes. Si éste no fuese el caso, podríamos reducir el orden del modelo a un número menor de fuentes equivalentes igual al rango de la matriz de steering. Supondremos que todas las señales son estadísticamente independientes y que el número de fuentes N_u es menor que la dimensión del vector de steering, es decir, que la dimensión del espacio, $N_u < N$. De este modo podemos hablar libremente de subespacio de señal y subespacio de ruido. Al igual que se expuso en la sección 4.2, estudiaremos la aplicación de la no-linealidad propuesta a la recuperación de una fuente en presencia de interferencias (el resto de las fuentes en el modelo de \mathbf{r}_k).

Recurriendo al Teorema Fundamental de la Esperanza, la función de coste se expresa de la siguiente forma,

$$J = -E_{y_k} \psi_\phi(y_k) = -E_{\mathbf{a}_k, \mathbf{n}_k} \psi_\phi(y_k(\mathbf{a}_k, \mathbf{n}_k)) = -E_{\mathbf{a}_k, \mathbf{n}_k} \psi_\phi(y_k) \quad (4.3.2)$$

donde el operador esperanza implica ahora al conjunto de todas las fuentes y al vector de ruido. Siguiendo el mismo desarrollo expuesto en la sección 4.2 y definiendo las siguientes variables aleatorias,

$$y_k = \mathbf{w}_k^H \mathbf{r}_k = \mathbf{w}_k^H \mathbf{S}_k \mathbf{a}_k + \mathbf{w}_k^H \mathbf{n}_k = y_k^a + \eta_k \quad (4.3.3)$$

siendo y^a la señal sin ruido, el gradiente de la ecuación de coste nos lleva a,

$$\begin{aligned} \nabla_{\mathbf{w}_k^H} J &= \nabla_{\mathbf{w}_k^H} - E_{\mathbf{a}_k} (\psi_\phi * p_{\eta_k})(y_k^a) \\ &= -E_{\mathbf{a}_k} (\dot{\psi}_\phi * p_{\eta_k}) \mathbf{S}_k \mathbf{a}_k - E_{\mathbf{a}_k} \left(\psi_\phi * \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right) \mathbf{R}_k \mathbf{w}_k \end{aligned} \tag{4.3.4}$$

con lo cual tras igualación a cero del gradiente obtenemos la expresión para el vector de coeficientes,

$$\mathbf{w}_k = -\mathbf{R}_k^{-1} \mathbf{S}_k \frac{E_{\mathbf{a}_k, \eta_k} \dot{\psi}_\phi(y_k) \mathbf{a}_k}{E_{\mathbf{a}_k} \left(\psi_\phi * \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right)} = \mathbf{R}_k^{-1} \mathbf{S}_k \alpha_k \tag{4.3.5}$$

lo cual significa que cualquier vector que minimice el criterio se puede generar a partir de como máximo N_u vectores linealmente independientes pertenecientes al subespacio generado por la matriz $\mathbf{R}_k^{-1} \mathbf{S}_k$. Los coeficientes de la combinación lineal se hallan agrupados en el vector α_k . Definimos la base transformada de los vectores de steering a partir de la inversa de la matriz de correlación del ruido como,

$$\bar{\mathbf{S}}_k = \mathbf{R}_k^{-1} \mathbf{S}_k \Rightarrow \mathbf{w}_k = \bar{\mathbf{S}}_k \alpha_k, \quad \bar{\mathbf{y}}_k = \bar{\mathbf{S}}_k^H \mathbf{r}_k \tag{4.3.6}$$

Puede demostrarse que la multiplicación del vector de datos por la matriz de steering produce un conjunto de estadísticos suficientes. Como se procedió en el caso anterior, debemos determinar si la función de densidad que minimiza el criterio se encuentra también polarizada. Dicha información se encuentra contenida en el vector de coeficientes α_k . Deberíamos poder demostrar también si en el mínimo de la función de coste, la fuente recuperada presenta interferencia de las otras fuentes. La solución de mínima varianza para la recuperación de todas las fuentes viene dada por la siguiente expresión,

$$\mathbf{W}_{MV}^H = (\mathbf{S}_k^H \mathbf{R}_k^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_k^{-1} = (\bar{\mathbf{S}}_k^H \mathbf{S}_k)^{-1} \bar{\mathbf{S}}_k^H = \Phi_k^{-1} \bar{\mathbf{S}}_k^H \tag{4.3.7}$$

donde Φ^{-1} denota a la matriz de decorrelación de señales. Evaluaremos ahora el gradiente de la función de coste para cada columna de \mathbf{W}_{MV} y comprobaremos si produce o no un mínimo. Se habrá demostrado así que la distribución que minimiza el criterio se encuentra polarizada con respecto a la de mínima varianza. Llegamos al punto donde debemos particularizar la no-linealidad de la función de coste. Supongámosla igual a la función de log-likelihood de la fuente deseada i .

Calculamos el gradiente con respecto al vector de coeficientes α_k ,

$$\nabla_{\alpha_k^H} J = -\nabla_{\alpha_k^H} E_{\bar{y}_k} \ln p_{a_k^i+n_k^i}(\alpha_k^H \bar{y}_k) = -E_{a_k, n_k} \frac{\dot{p}_{a_k^i+n_k^i}(\alpha_k^H \bar{y}_k)}{p_{a_k^i+n_k^i}(\alpha_k^H \bar{y}_k)} \bar{y}_k, \quad \begin{cases} p_{a_k^i+n_k^i} = p_{\alpha_k^H \bar{y}_k} \\ \bar{y}_k = S_k^H R_k^{-1} r_k \end{cases} \quad (4.3.8)$$

Demostramos a continuación que este gradiente no es cero cuando el vector α_k es tal que recupera la fuente i -ésima según el criterio de mínima varianza y que por tanto la distribución minimizadora se encuentra polarizada. Si premultiplicamos y dividimos con respecto a la matriz de decorrelación Φ^{-1} , tenemos,

$$\nabla_{\alpha_k^H} J = -\Phi_k E_{a_k, n_k} \frac{\dot{p}_{a_k^i+n_k^i}(\alpha_k^H \bar{y}_k)}{p_{a_k^i+n_k^i}(\alpha_k^H \bar{y}_k)} \Phi_k^{-1} \bar{y}_k = -\Phi_k v \quad (4.3.9)$$

Puesto que la inversa de la matriz de decorrelación (el producto escalar de los vectores de steering en la norma R^{-1}) es de rango completo por definición, su Kernel o subespacio nulo tiene dimensión cero. Por lo tanto, si como mínimo alguno de los elementos del vector v es diferente de cero, el gradiente no puede ser igual al vector nulo y la distribución se encuentra polarizada. Examinemos entonces la componente i -ésima del vector v . Dado que el vector de coeficientes α_k es tal que recupera la i -ésima fuente tenemos que,

$$[\Phi_k^{-1} \bar{y}_k]_i = z_k^i = a_k^i + n_k^i = \alpha_k^H \bar{y}_k \quad (4.3.10)$$

Por consiguiente,

$$[v]_i = E_{a_k, n_k} \frac{\dot{p}_{a_k^i+n_k^i}(\alpha_k^H \bar{y}_k)}{p_{a_k^i+n_k^i}(\alpha_k^H \bar{y}_k)} [\Phi_k^{-1} \bar{y}_k]_i = E_{z_k^i} \frac{\dot{p}_{z_k^i}(z_k^i)}{p_{z_k^i}(z_k^i)} z_k^i = \int z_k^i \dot{p}_{z_k^i}(z_k^i) dz_k^i = -1 \quad (4.3.11)$$

con lo cual, al ser la componente i -ésima de n diferente de cero, tenemos que el gradiente de la función de coste no tiene un mínimo en la solución de mínima varianza,

$$[v]_i = -1 \Rightarrow \nabla_{\alpha_k^H} J = -\Phi_k v \neq 0 \quad (4.3.12)$$

y por tanto la distribución que minimiza el criterio se encuentra polarizada.

A continuación examinaremos el significado del vector de coeficientes α_k . El cálculo del gradiente de la función de coste nos conduce a la conocida expresión,

$$\nabla_{\mathbf{w}_k^H} J = -E_{\mathbf{a}_k, \mathbf{n}_k} \left(\frac{\partial}{\partial y_k} \psi_\phi \right) \nabla_{\mathbf{w}_k^H} y_k = -E_{\mathbf{a}_k, \mathbf{n}_k} \dot{\psi}_\phi(y_k) \mathbf{r}_k = \mathbf{0} \quad (4.3.13)$$

Desglosando el vector \mathbf{r}_k en sus componentes de señal y de ruido obtenemos,

$$S_k E_{\mathbf{a}_k, \mathbf{n}_k} \dot{\psi}_\phi(y_k) \mathbf{a}_k = -E_{\mathbf{a}_k, \mathbf{n}_k} \dot{\psi}_\phi(y_k) \mathbf{n}_k \quad (4.3.14)$$

lo cual es equivalente a,

$$E_{\mathbf{a}_k, \mathbf{n}_k} \dot{\psi}_\phi(y_k) \mathbf{a}_k = -\Phi_k^{-1} E_{\mathbf{a}_k, \mathbf{n}_k} \dot{\psi}_\phi(y_k) S_k^H R_k^{-1} \mathbf{n}_k \quad (4.3.15)$$

La expresión del vector de coeficientes óptimo queda expresado entonces según la ecuación 4.3.5 como,

$$\begin{aligned} \mathbf{w}_k &= R_k^{-1} S_k \frac{1}{\lambda} \Phi_k^{-1} \left(-E_{\mathbf{a}_k, \mathbf{n}_k} \dot{\psi}_\phi(y_k) S_k^H R_k^{-1} \mathbf{n}_k \right) = \mathbf{W}_{MV} \xi_k \\ \xi_k &= -\frac{1}{\lambda} E_{\mathbf{a}_k, \mathbf{n}_k} \dot{\psi}_\phi(y_k) S_k^H R_k^{-1} \mathbf{n}_k \quad , \quad \lambda = E_{\mathbf{a}_k} \left(\psi_\phi^* \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right) \end{aligned} \quad (4.3.16)$$

donde los coeficientes de la combinación lineal de los vectores de mínima varianza agrupados en la matriz \mathbf{W}_{MV} resultan ser iguales a la correlación del error generalizado (representado por la derivada de la función de coste) con el ruido que afecta al vector de estadísticos suficientes, $S_k^H R_k^{-1} \mathbf{n}_k$.

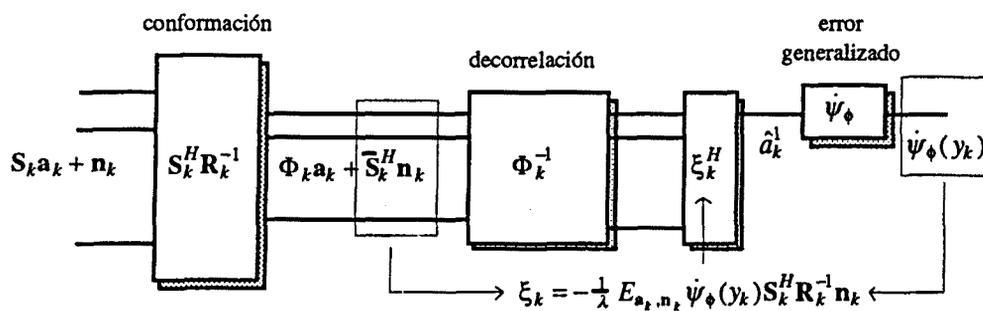


Figura 4.3.1: representación esquemática del funcionamiento de la función de coste de error generalizado propuesto con respecto al filtro de mínima varianza de la fuente 1.

Para la recuperación de la fuente 1 utilizaremos la no-linealidad log-likelihood. En el caso óptimo de mínimos cuadrados, cuando conocemos los vectores de steering correspondientes a cada fuente, tenemos que el vector ξ_k^i que recupera cada fuente es proporcional al vector

de pinning π_i asociado (un uno en la posición 'i' y ceros en el resto), lo cual traducido a términos de correlación según la ecuación 4.3.16 significa que el ruido a la salida de la fuente 'i' se encuentra incorrelado con los procesos de ruido a la salida de la etapa de conformación correspondientes a las fuentes $j \neq i$. Es decir, la etapa de decorrelación realiza una predicción del ruido a la salida i-ésima del conformador en base a los procesos de ruido del resto de las salidas a fin de minimizar la potencia de ruido que afecta a la i-ésima fuente. Estableciendo un paralelo entre el caso óptimo MMSE y la aplicación de la función de coste propuesta, el vector ξ_k^i se obtiene en base al error regenerado y no al error real.

En lo que sigue, deduciremos una expresión de ligadura análoga a la desarrollada en la sección 4.2 entre el parámetro varianza óptimo y el equivalente en este caso al escalado (ξ_k). Veremos a lo largo del desarrollo algunas relaciones interesantes sobre el comportamiento del algoritmo. Procediendo según el método usual, reformularemos el gradiente en términos de las funciones $q(\cdot)$ y del vector de datos. Denominamos el vector de salida de la solución de mínima varianza a una entrada r_k como u_k , de forma que la salida del filtro pueda expresarse en función del vector de coeficientes ξ_k en lugar de α_k ,

$$u_k = \Phi_k^{-1} \bar{y}_k, \quad \xi_k = \Phi_k \alpha_k \Rightarrow z_k = z_k^i = \xi_k^H u_k = \alpha_k^H \bar{y}_k \tag{4.3.17}$$

donde sin pérdida de generalidad suponemos que tratamos de recuperar la fuente '1'. Véase la figura 4.3.1.1. El cálculo del gradiente de la función de coste respecto a ξ_k nos conduce a,

$$\nabla_{\xi_k^H} J = -E_{a_k, n_k} E_{a_k'} (z_k - a_k')^* q(z_k, a_k') u_k = 0 \tag{4.3.18}$$

Dado que ahora todo quedará expresado en función del vector solución de mínima varianza u_k , podemos indexar el operador esperanza en función de éste en base al Teorema Fundamental de la Esperanza. Desglosando la ecuación anterior llegamos a,

$$E_{u_k} E_{a_k'} z_k^* u_k q(z_k, a_k') = E_{u_k} E_{a_k'} a_k'^* u_k q(z_k, a_k') \Rightarrow E_{u_k} z_k^* u_k = E_{u_k} \hat{a}_k^* u_k \tag{4.3.19}$$

donde observamos que la correlación cruzada de la salida del filtro, z_k con la solución de mínima varianza es igual a la correlación de la señal estimada con dicha solución. Operando de forma distinta con la misma ecuación surge una expresión explícita para el vector (alternativa a la de la figura 4.3.1). Consideremos la expresión de z_k en la ecuación 4.3.17 y sustituyamos en la ecuación 4.3.19,

$$\tag{4.3.20}$$

$$E_{u_t} \mathbf{u}_k \mathbf{u}_k^H \xi_k = E_{u_t} \hat{\mathbf{a}}_k^* \mathbf{u}_k \Rightarrow \mathbf{R}_{uu} \xi_k = E_{u_t} \hat{\mathbf{a}}_k^* \mathbf{u}_k \Rightarrow \xi_k = \mathbf{R}_{uu}^{-1} E_{u_t} \hat{\mathbf{a}}_k^* \mathbf{u}_k$$

obteniendo una expresión de ξ_k no basada en términos de correlación cruzada de procesos de ruido como en la figura 4.3.1, sino en términos de la matriz de correlación de los datos a la salida del decorrelador de mínima varianza. En el caso óptimo de mínima varianza, la correlación cruzada $E \hat{\mathbf{a}}_k^* \mathbf{u}_k$ es precisamente una de las columnas de la matriz de correlación \mathbf{R}_{uu} , con lo cual el vector ξ_k bloquea todas las salidas del decorrelador, excepto una con amplitud unidad.

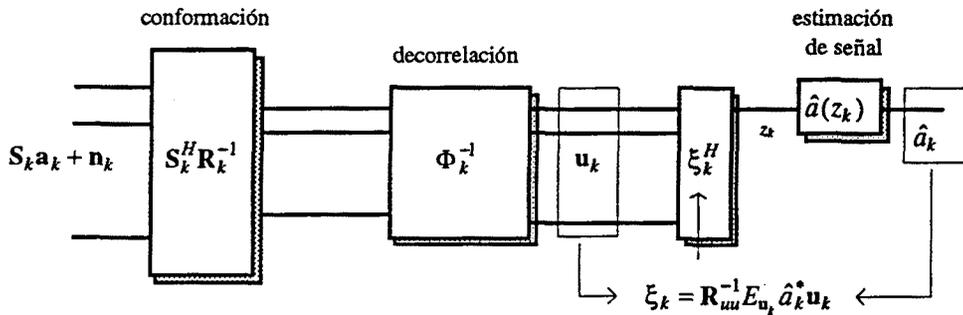


Figura 4.3.2: segundo método de obtención de los coeficientes del filtro de recuperación de la fuente 1 a partir de las correlaciones cruzadas entre los datos.

Podemos hallar también una expresión para el vector de coeficientes α_k que obtiene z_k a partir de la salida de la etapa de conformación. La relación entre los vectores de coeficientes α_k y ξ_k viene dada a partir de la matriz de decorrelación. Por consiguiente,

$$z_k = \xi_k^H \mathbf{u}_k = \xi_k^H \Phi_k^{-1} \bar{\mathbf{y}}_k = \alpha_k^H \bar{\mathbf{y}}_k \Rightarrow \alpha_k = \Phi_k^{-1} \xi_k \tag{4.3.21}$$

Multiplicando ambos lados de la ecuación 4.3.21 por la matriz de decorrelación llegamos a una expresión alternativa para el vector de coeficientes α_k ,

$$\alpha_k = \Phi_k^{-1} \mathbf{R}_{uu}^{-1} E_{u_t} \hat{\mathbf{a}}_k^* \mathbf{u}_k = \Phi_k^{-1} \mathbf{R}_{uu}^{-1} \Phi_k^{-1} E_{u_t} \hat{\mathbf{a}}_k^* \Phi_k \mathbf{u}_k \Rightarrow \alpha_k = \mathbf{R}_{\bar{\mathbf{y}}\bar{\mathbf{y}}}^{-1} E_{\bar{\mathbf{y}}_t} \hat{\mathbf{a}}_k^* \bar{\mathbf{y}}_k \tag{4.3.22}$$

El resultado anterior se representa esquemáticamente en la figura 4.3.3.

Multiplicando ambos lados de la ecuación 4.3.20 por el vector ξ_k hermítico, obtenemos una relación que utilizaremos más adelante, y que llamamos ecuación de conservación de la fase,

$$E_{z_t} |z_k|^2 = E_{z_t} z_k \hat{\mathbf{a}}_k^* \Rightarrow \text{Im}\{E_{z_t} z_k \hat{\mathbf{a}}_k^*\} = 0 \tag{4.3.23}$$

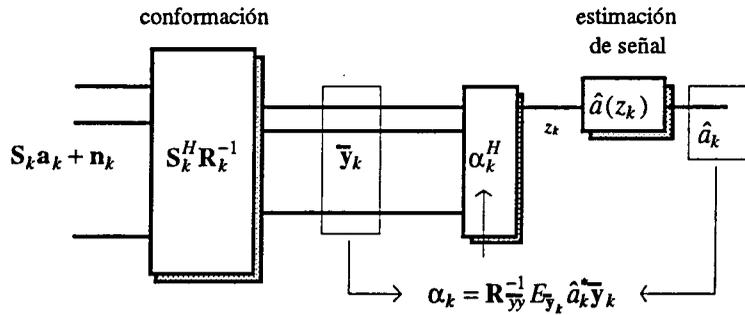


Figura 4.3.3: obtención del vector de coeficientes a_k a partir de la correlación de la referencia con los datos de salida del conformador.

Es decir, en esperanza, el error de fase entre z_k y \hat{a}_k es nulo. A continuación, reproducimos y expandimos la expresión del parámetro varianza ya obtenido en la ecuación 3.5.4.4 utilizando las propiedades de las funciones $q(\cdot)$,

$$\begin{aligned} \sigma_i^2 &= E_{z_k} E_{a'_k} |z_k - a'_k|^2 q(z_k, a'_k) \\ &= E_{z_k} |z_k|^2 - E_{z_k} z_k^* \hat{a}_k - E_{z_k} z_k \hat{a}_k^* + E_{z_k} E_{a'_k} |a'_k|^2 q(z_k, a'_k) \end{aligned} \tag{4.3.24}$$

Recurriendo a la propiedad 4.3.22 tenemos,

$$\sigma_i^2 + E_{z_k} |z_k|^2 = E_{z_k} E_{a'_k} |a'_k|^2 q(z_k, a'_k) \tag{4.3.25}$$

De aquí podemos deducir fácilmente una expresión de ligadura entre el parámetro varianza y el vector ξ_k por simple sustitución de z_k ,

$$E_{z_k} |z_k|^2 = E_{z_k} \xi_k^H \mathbf{u}_k \mathbf{u}_k^H \xi_k = \xi_k^H \mathbf{R}_{uu} \xi_k \Rightarrow \sigma_i^2 + \xi_k^H \mathbf{R}_{uu} \xi_k = E_{z_k} E_{a'_k} |a'_k|^2 q(z_k, a'_k) \tag{4.3.26}$$

En el caso particular de una modulación de amplitud constante e igual a uno se reduce a la siguiente expresión,

$$\sigma_i^2 + \xi_k^H \mathbf{R}_{uu} \xi_k = 1 \tag{4.3.27}$$

Podemos observar en presencia de interferencias también una contracción de la constelación. Para ello consideremos una expansión de la matriz de correlación \mathbf{R}_{uu} en la correlación de los datos y del ruido a la salida. suponemos en este caso que todas las fuentes son de amplitud unidad (independientes). Entonces,

$$\tag{4.3.28}$$

$$\mathbf{R}_{uu} = \mathbf{I} + [\mathbf{S}_k^H \mathbf{R}_k^{-1} \mathbf{S}_k]^{-1} = \mathbf{I} + \Phi_k^{-1}$$

Sustituyendo en la ecuación 4.3.27 obtenemos,

$$\sigma_t^2 + \xi_k^H \xi_k \left(1 + \frac{\xi_k^H \Phi_k^{-1} \xi_k}{\xi_k^H \xi_k} \right) = 1 \quad (4.3.29)$$

Utilizando las propiedad del cociente de Rayleigh, podemos llegar a una cota superior para el vector en función de la potencia del ruido a la entrada,

$$0 \leq \lambda_{\min}(\Phi_k^{-1}) = \frac{1}{\lambda_{\max}(\Phi_k)} \leq \frac{\xi_k^H \Phi_k^{-1} \xi_k}{\xi_k^H \xi_k} \Rightarrow \xi_k^H \xi_k \leq \frac{1}{1 + \lambda_{\max}^{-1}(\Phi_k)} \quad (4.3.30)$$

En el caso de una única fuente podemos ver como este resultado se reduce a la desigualdad ya deducida en la ecuación 4.2.4.12.

4.3.1 Equivalencia con una Referencia Conocida

En este apartado estableceremos un paralelo entre las relaciones derivadas anteriormente para los vectores \mathbf{a}_k y $\boldsymbol{\psi}_k$ con regeneración de referencia y con referencia externa, es decir, de los datos reales sin ruido. Como función de coste paralela respecto al segundo caso, consideraremos la minimización de la potencia del error entre la salida del filtro y la referencia disponible a la que llamamos d_k . Es decir,

$$J_{mse} = E_{z_k} |z_k - d_k|^2 = E_{\mathbf{u}_k} |\xi_k^H \mathbf{u}_k - d_k|^2 \quad (4.3.1.1)$$

El cálculo del gradiente de la función de coste nos lleva a,

$$\nabla_{\xi_k^H} J_{mse} = E_{\mathbf{u}_k} (\xi_k^H \mathbf{u}_k - d_k)^* \mathbf{u}_k = (E_{\mathbf{u}_k} \mathbf{u}_k \mathbf{u}_k^H) \xi_k - E_{\mathbf{u}_k} d_k^* \mathbf{u}_k = \mathbf{0} \quad (4.3.1.2)$$

Reorganizando términos llegamos a,

$$\xi_k = \mathbf{R}_{uu}^{-1} E_{\mathbf{u}_k} d_k^* \mathbf{u}_k \quad \propto \quad \xi_k = \mathbf{R}_{uu}^{-1} E_{\mathbf{u}_k} \hat{a}_k^* \mathbf{u}_k \quad (4.3.1.3)$$

donde podemos apreciar la equivalencia entre la solución de mínimos cuadrados y la solución expuesta en la ecuación 4.3.1.17. No obstante, aunque la forma es idéntica, ambas soluciones difieren en el término de la referencia. Establecemos así una analogía entre ambos criterios. Podemos observar en la ecuación 4.3.1.2 que también se cumple el principio de ortogonalidad entre el error y los datos.

Al igual que hemos llegado a una expresión para el vector reducido ξ_k en la ecuación 4.3.1.3, es posible también llegar a su análoga en el vector w_k . En efecto, retomemos la expresión del gradiente de la función de coste con respecto a w_k e igualémoslo a cero,

$$E_{r_k} E_{a'_k} q(z_k, a'_k) (z_k - a'_k)^* r_k = 0 \Rightarrow E_{r_k} E_{a'_k} q(z_k, a'_k) z_k^* r_k = E_{r_k} E_{a'_k} q(z_k, a'_k) a_k^* r_k \quad (4.3.1.4)$$

Sustituyendo z_k por $w_k^H r_k$ en el lado izquierdo de la anterior igualdad y expresando el lado derecho en términos de la referencia regenerada. llegamos a la siguiente igualdad,

$$E_{r_k} E_{a'_k} q(z_k, a'_k) r_k r_k^H w_k = E_{r_k} r_k r_k^H w_k = E_{r_k} \hat{a}_k^* r_k \quad (4.3.1.5)$$

donde hemos utilizado la propiedad de que la esperanza de la función de calidad con respecto a la variable a'_k es la unidad. Finalmente, llegamos a la expresión deseada para el vector de coeficientes w_k ,

$$w_k = R_{rr}^{-1} E_{r_k} \hat{a}_k^* r_k \quad (4.3.1.6)$$

Hemos visto hasta ahora dos formas diferentes de expresar el vector de coeficientes. La primera, mediante los vectores ξ_k o bien α_k , y la segunda expresada en función de la matriz de autocorrelación de los datos r_k ,

$$\begin{aligned} (1) \quad w_k &= R_{nn}^{-1} S_k \alpha_k = R_{nn}^{-1} S_k \Phi_k^{-1} \xi_k, \quad \xi_k = R_{uu}^{-1} E_{u_k} \hat{a}_k^* u_k \\ (2) \quad w_k &= R_{rr}^{-1} E_{r_k} \hat{a}_k^* r_k \end{aligned} \quad (4.3.1.7)$$

Con lo cual ambas expresiones deben coincidir. A continuación justificaremos este hecho y extraeremos algunas conclusiones importantes de la coincidencia de ambas formas de representación, (1) y (2). Recuérdese como se vió en la ecuación 4.3.16, que la forma 1 es equivalente a una combinación lineal de las columnas de la matriz W_{MV} , la cual es solución del criterio de mínima varianza o de referencia espacial para la recuperación de todas las fuentes con el escalado correcto y potencia de ruido mínima. Partiremos de la forma 2 y

mediante transformaciones matriciales llegaremos a la forma 1, obteniendo como sub-producto una expresión alternativa para el vector de coeficientes α_k .

Consideremos en primer lugar la expresión de la matriz de correlación de los datos según el modelo de señal de varias fuentes,

$$\mathbf{R}_{rr} = E_{r_k} \mathbf{r}_k \mathbf{r}_k^H = \mathbf{S}_k (E \mathbf{a}_k \mathbf{a}_k^H) \mathbf{S}_k^H + \mathbf{R}_{nn} = \mathbf{S}_k \mathbf{P}_{aa} \mathbf{S}_k^H + \mathbf{R}_{nn} \quad (4.3.1.7)$$

donde introducimos la matriz de correlación de la secuencia \mathbf{a}_k , \mathbf{P}_{aa} , diagonal y positiva, puesto que actuamos bajo el supuesto de fuentes incorreladas e independientes. Dado que la expresión de la forma 2 requiere de la inversa de la matriz de correlación, recurrimos al Lema de Inversión de Matrices,

$$\mathbf{R}_{rr}^{-1} = \mathbf{R}_{nn}^{-1} - \mathbf{R}_{nn}^{-1} \mathbf{S}_k (\mathbf{P}_{aa}^{-1} + \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \quad (4.3.1.8)$$

Expandiendo la ecuación 4.3.1.7(2) según el modelo de señal, llegamos a,

$$\mathbf{w}_k = \mathbf{R}_{rr}^{-1} E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{r}_k = \mathbf{R}_{rr}^{-1} \mathbf{S}_k E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{a}_k + \mathbf{R}_{rr}^{-1} E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{n}_k = \mathbf{w}_1 + \mathbf{w}_2 \quad (4.3.1.9)$$

donde observamos que el cálculo del vector de coeficientes puede desglosarse en dos términos, \mathbf{w}_1 y \mathbf{w}_2 . Consideremos la expresión del primero a partir de las ecuaciones 4.3.1.8-9,

$$\begin{aligned} \mathbf{w}_1 &= \mathbf{R}_{rr}^{-1} \mathbf{S}_k E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{a}_k = \left(\mathbf{R}_{nn}^{-1} - \mathbf{R}_{nn}^{-1} \mathbf{S}_k (\mathbf{P}_{aa}^{-1} + \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \right) \mathbf{S}_k E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{a}_k \\ &= \mathbf{R}_{nn}^{-1} \mathbf{S}_k \left(\mathbf{I} - (\mathbf{P}_{aa}^{-1} + \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \right) E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{a}_k \\ &= \mathbf{R}_{nn}^{-1} \mathbf{S}_k \mathbf{A} E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{a}_k \end{aligned} \quad (4.3.1.10)$$

en función de la matriz cuadrada \mathbf{A} , de dimensión igual al número de fuentes, definida como,

$$\mathbf{A} \stackrel{def}{=} \mathbf{I} - (\mathbf{P}_{aa}^{-1} + \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k = (\mathbf{P}_{aa}^{-1} + \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{P}_{aa}^{-1} \quad (4.3.1.11)$$

Consideremos ahora la expresión de \mathbf{w}_2 a partir de las ecuaciones 4.3.1.8-9,

$$\mathbf{w}_2 = \mathbf{R}_{rr}^{-1} E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{n}_k = \left(\mathbf{R}_{nn}^{-1} - \mathbf{R}_{nn}^{-1} \mathbf{S}_k (\mathbf{P}_{aa}^{-1} + \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \right) E_{r_k} \hat{\mathbf{a}}_k^* \mathbf{n}_k \quad (4.3.1.12)$$

Viendo la estructura que presenta w_1 en la ecuación 4.3.1.10 y la estructura que presenta w_k en la ecuación 4.3.1.7(1), y siendo $w_2 = w_k - w_1$, w_2 debe presentar la misma estructura. Por lo tanto, existe un vector β_k tal que,

$$w_2 = R_{nn}^{-1} S_k \beta_k = \left(R_{nn}^{-1} - R_{nn}^{-1} S_k (P_{aa}^{-1} + S_k^H R_{nn}^{-1} S_k)^{-1} S_k^H R_{nn}^{-1} \right) E_{r_t} \hat{a}_k^* n_k \quad (4.3.1.12)$$

aislamos el vector β_k mediante premultiplicación por la matriz adecuada,

$$\begin{aligned} \beta_k &= \left(S_k^H R_{nn}^{-1} S_k \right)^{-1} S_k^H \left(R_{nn}^{-1} - R_{nn}^{-1} S_k (P_{aa}^{-1} + S_k^H R_{nn}^{-1} S_k)^{-1} S_k^H R_{nn}^{-1} \right) E_{r_t} \hat{a}_k^* n_k \\ &= \left(I - (P_{aa}^{-1} + S_k^H R_{nn}^{-1} S_k)^{-1} S_k^H R_{nn}^{-1} S_k \right) \left(S_k^H R_{nn}^{-1} S_k \right)^{-1} S_k^H R_{nn}^{-1} E_{r_t} \hat{a}_k^* n_k \end{aligned} \quad (4.3.1.13)$$

Podemos simplificar la última expresión, reconociendo la matriz W_{MV} ya citada anteriormente en el extremo derecho de la igualdad. Operando obtenemos,

$$\beta_k = \left(I - (P_{aa}^{-1} + S_k^H R_{nn}^{-1} S_k)^{-1} S_k^H R_{nn}^{-1} S_k \right) E_{r_t} \hat{a}_k^* W_{MV}^H n_k = A E_{r_t} \hat{a}_k^* W_{MV}^H n_k \quad (4.3.1.14)$$

Quedando expresado el vector β_k como el producto de la matriz A definida en 4.3.1.11 por la correlación cruzada entre la referencia regenerada \hat{a}_k conjugada con el vector de ruido obtenido a la salida de la matriz W_{MV} correspondiente al criterio de mínima varianza. Podemos entonces expresar finalmente el vector de coeficientes como,

$$w_k = R_{nn}^{-1} S_k \alpha_k \quad , \quad \alpha_k = A \left(E_{r_t} \hat{a}_k^* a_k + E_{r_t} \hat{a}_k^* W_{MV}^H n_k \right) \quad (4.3.1.15)$$

o bien en términos de la matriz de mínima varianza,

$$w_k = W_{MV} \Phi_k \left(P_{aa}^{-1} + \Phi_k \right)^{-1} P_{aa}^{-1} \left(E_{r_t} \hat{a}_k^* a_k + E_{r_t} \hat{a}_k^* W_{MV}^H n_k \right) \quad (4.3.1.16)$$

De las anteriores ecuaciones, 4.3.1.15-16, podemos ver claramente el efecto de interferencia residual reflejado por la correlación cruzada $E \hat{a}_k^* a_k$, donde cada componente refleja la correlación existente entre la referencia regenerada y cada una de las fuentes del modelo. Observamos también la influencia del ruido en el vector de coeficientes mediante el segundo sumando de la ecuación 4.3.1.15, el cual resulta ser nulo en un esquema de referencia temporal donde no existe incertidumbre en la secuencia transmitida, puesto que entonces, al no contener la referencia términos de ruido, queda incorrelada con el término n_k .

4.4 Hessiano de la función de coste

En este apartado estudiaremos el Hessiano de la función de coste a partir de las expresiones del gradiente deducidas en el apartado anterior. Suponemos el caso sencillo de una arquitectura lineal en los datos y en los coeficientes. El análisis del Hessiano nos facilitará ver en qué condiciones obtenemos mínimos de la función de coste y por otra parte nos conducirá a la definición de parámetros de interés de la función de coste y a relaciones interesantes entre ellos. Expresamos el Hessiano como el gradiente de cada uno de los componentes del gradiente,

$$H_{w_k} J = \nabla_{w_k} \nabla_{w_k^H} J = \nabla_{w_k} E_{a_k, n_k} E_{a_k'} (y_k - a_k')^* q(y_k, a_k') r_k \tag{4.4.1}$$

Intercambiando gradiente y esperanza y reorganizando términos llegamos al gradiente de la expresión entre llaves,

$$H_{w_k} J = E_{a_k, n_k} r_k E_{a_k'} \nabla_{w_k} \left\{ \zeta_k = (y_k - a_k')^* q(y_k, a_k') \right\} \tag{4.4.2}$$

Aplicando la regla de la cadena y utilizando el hecho de que el gradiente de y_k respecto los coeficientes del filtro es nulo, el gradiente de ζ_k queda expresado como,

$$\nabla_{w_k} y_k = 0 \Rightarrow \nabla_{w_k} \zeta_k = \left(\frac{\partial}{\partial y_k} \zeta_k \right) \nabla_{w_k} y_k^* = \left(q(y_k, a_k') + (y_k - a_k')^* \frac{\partial}{\partial y_k} q(y_k, a_k') \right) r_k^H \tag{4.4.3}$$

Donde sólo nos queda hallar la derivada parcial de la función $q(\cdot)$ respecto a y_k^* . Recurriendo a la expresión de $q(\cdot)$, tenemos,

$$\begin{aligned} \frac{\partial}{\partial y_k} q(y_k, a_k') &= \frac{\partial}{\partial y_k} \frac{e^{-|y_k - a_k'|^2 / \sigma_i^2}}{E_{a_k'} e^{-|y_k - a_k'|^2 / \sigma_i^2}} = \frac{\frac{\partial}{\partial y_k} e^{-|y_k - a_k'|^2 / \sigma_i^2}}{E_{a_k'} e^{-|y_k - a_k'|^2 / \sigma_i^2}} - \frac{e^{-|y_k - a_k'|^2 / \sigma_i^2} \frac{\partial}{\partial y_k} E_{a_k'} e^{-|y_k - a_k'|^2 / \sigma_i^2}}{\left(E_{a_k'} e^{-|y_k - a_k'|^2 / \sigma_i^2} \right)^2} \\ &= - \frac{e^{-|y_k - a_k'|^2 / \sigma_i^2}}{E_{a_k'} e^{-|y_k - a_k'|^2 / \sigma_i^2}} \frac{1}{\sigma_i^2} (y_k - a_k') + \frac{e^{-|y_k - a_k'|^2 / \sigma_i^2} E_{a_k'} e^{-|y_k - a_k'|^2 / \sigma_i^2} \frac{1}{\sigma_i^2} (y_k - a_k')}{\left(E_{a_k'} e^{-|y_k - a_k'|^2 / \sigma_i^2} \right)^2} \end{aligned} \tag{4.4.4}$$

Restituyendo la expresión de $q(\cdot)$ en el primer y segundo términos de la ecuación anterior obtenemos,

$$\tag{4.4.5}$$

$$\begin{aligned}
\frac{\partial}{\partial y_k} q(y_k, a'_k) &= -\frac{1}{\sigma_i^2} (y_k - a'_k) q(y_k, a'_k) + q(y_k, a'_k) E_{a'_k} \frac{1}{\sigma_i^2} (y_k - a''_k) q(y_k, a''_k) \\
&= -\frac{1}{\sigma_i^2} (y_k - a'_k) q(y_k, a'_k) + q(y_k, a'_k) \frac{1}{\sigma_i^2} (y_k - E_{a'_k} a''_k q(y_k, a''_k)) \\
&= \frac{1}{\sigma_i^2} (a'_k - E_{a'_k} a''_k q(y_k, a''_k)) q(y_k, a'_k) = \frac{1}{\sigma_i^2} (a'_k - \hat{a}_k) q(y_k, a'_k)
\end{aligned}$$

Volviendo a la ecuación 4.4.3 el gradiente de ζ_k queda expresado como,

$$\nabla_{\mathbf{w}_k} \zeta_k = \left(1 + \frac{1}{\sigma_i^2} (y_k - a'_k)^* (a'_k - \hat{a}_k) \right) q(y_k, a'_k) \mathbf{r}_k^H \quad (4.4.6)$$

Volviendo otro paso atrás llegamos a la expresión final para el Hessiano,

$$\begin{aligned}
\mathbf{H}_{\mathbf{w}_k} J &= E_{a_i, n_i} \mathbf{r}_k E_{a'_i} \left(1 + \frac{1}{\sigma_i^2} (y_k - a'_k)^* (a'_k - \hat{a}_k) \right) q(y_k, a'_k) \mathbf{r}_k^H \\
&= E_{a_i, n_i} \mathbf{r}_k \mathbf{r}_k^H + E_{a_i, n_i} \mathbf{r}_k \mathbf{r}_k^H \frac{1}{\sigma_i^2} E_{a'_i} (y_k - a'_k)^* (a'_k - \hat{a}_k) q(y_k, a'_k)
\end{aligned} \quad (4.4.7)$$

con lo cual el Hessiano queda descompuesto en la suma de dos términos. El primero coincide con la matriz de correlación de los datos de entrada, mientras que el segundo constituye también una matriz de correlación ponderada por un término que depende de forma compleja del error estimado de detección. Examinemos con más detalle el factor escalar del segundo término al que denominamos $\gamma(y_k)$,

$$\begin{aligned}
\gamma(y_k) &= \frac{1}{\sigma_i^2} E_{a'_i} (y_k - a'_k)^* (a'_k - \hat{a}_k) q(y_k, a'_k) = \frac{1}{\sigma_i^2} E_{a'_i} y_k^* (a'_k - \hat{a}_k) q(y_k, a'_k) - \frac{1}{\sigma_i^2} E_{a'_i} a_k'^* (a'_k - \hat{a}_k) q(y_k, a'_k) \\
&= \frac{1}{\sigma_i^2} y_k^* \hat{a}_k - \frac{1}{\sigma_i^2} y_k^* \hat{a}_k - \frac{1}{\sigma_i^2} E_{a'_i} a_k'^* (a'_k - \hat{a}_k) q(y_k, a'_k) = -\frac{1}{\sigma_i^2} E_{a'_i} a_k'^* (a'_k - \hat{a}_k) q(y_k, a'_k)
\end{aligned} \quad (4.4.8)$$

Donde por una parte hemos demostrado que la salida del filtro se encuentra incorrelada con respecto al error de detección promediado a través de las funciones $q(\cdot)$,

$$E_{a'_i} y_k^* (a'_k - \hat{a}_k) q(y_k, a'_k) = 0 \quad (4.4.9)$$

y por otra obtenemos una expresión reducida del Hessiano de la función de coste,

$$\begin{aligned}
\mathbf{H}_{\mathbf{w}_k} J &= E_{a_i, n_i} \mathbf{r}_k \mathbf{r}_k^H - E_{a_i, n_i} \mathbf{r}_k \mathbf{r}_k^H \frac{1}{\sigma_i^2} E_{a'_i} a_k'^* (a'_k - \hat{a}_k) q(y_k, a'_k) \\
&= E_{a_i, n_i} \mathbf{r}_k \mathbf{r}_k^H + E_{a_i, n_i} \gamma(y_k) \mathbf{r}_k \mathbf{r}_k^H
\end{aligned} \quad (4.4.10)$$

Vemos pues que el factor γ representa la correlación promediada a través de $q(\cdot)$ de la señal sin ruido con el error de estimación. Podemos demostrar que en el caso de relación señal a

ruido elevada, el factor γ tiende a cero, por lo cual el Hessiano de la función de coste es igual a la matriz de correlación de los datos. Una expresión alternativa del factor γ viene dada por,

$$\gamma(y_k) = \frac{1}{\sigma_i^2} \left(E_{a_k'} |a_k'|^2 q(y_k, a_k') - |\hat{a}_k|^2 \right) \quad (4.4.11)$$

como resulta inmediato demostrar a partir de la ecuación 4.4.8. La expresión anterior del factor γ sugiere una nueva interpretación de su significado. Es posible relacionarlo con alguna definición del error de estimación de la muestra de señal recibida?. Podemos ver que sí si tomamos la definición de dicho error como el promedio sobre las funciones $q(\cdot)$ de la norma cuadrática del error entre la señal estimada \hat{a}_k y cada a_k' de señal, es decir,

$$\sigma_d^2 \stackrel{def}{=} E_{a_k'} (a_k' - \hat{a}_k)^* (a_k' - \hat{a}_k) q(y_k, a_k') \quad (4.4.12)$$

El desarrollo de la ecuación anterior nos relaciona rápidamente σ_d^2 con el factor γ ,

$$\begin{aligned} \sigma_d^2 &= E_{a_k'} a_k'^* (a_k' - \hat{a}_k) q(y_k, a_k') - \hat{a}_k^* E_{a_k'} (a_k' - \hat{a}_k) q(y_k, a_k') = E_{a_k'} a_k'^* (a_k' - \hat{a}_k) q(y_k, a_k') - \hat{a}_k^* (\hat{a}_k - \hat{a}_k) \\ &E_{a_k'} a_k'^* (a_k' - \hat{a}_k) q(y_k, a_k') = -\sigma_i^2 \gamma(y_k) \end{aligned} \quad (4.4.13)$$

con lo cual deducimos que el factor γ es menos el error de estimación definido anteriormente y que por lo tanto es siempre menor que cero. La expresión final del Hessiano se reduce entonces a,

$$\mathbf{H}_{w_k} J = E_{a_k, n_k} \left(1 - \frac{\sigma_d^2(y_k)}{\sigma_i^2} \right) \mathbf{r}_k \mathbf{r}_k^H \quad (4.4.14)$$

Como resultado secundario obtenemos que el Hessiano es siempre menor que la matriz de correlación de los datos en el sentido de definición positiva,

$$\mathbf{H}_{w_k} J < E_{a_k, n_k} \mathbf{r}_k \mathbf{r}_k^H \quad (4.4.15)$$

En los mínimos de la función de coste, el Hessiano debe ser definido positivo, de donde encontramos una relación final que el algoritmo ciego debe cumplir,

$$\mathbf{H}_{w_k} J > 0 \Rightarrow E_{a_k, n_k} \mathbf{r}_k \mathbf{r}_k^H > E_{a_k, n_k} \frac{\sigma_d^2(y_k)}{\sigma_i^2} \mathbf{r}_k \mathbf{r}_k^H \quad (4.4.16)$$

Dicha relación establece un límite fundamental para la relación señal a ruido mínima bajo la cual el algoritmo ciego basado en la función de coste propuesta puede trabajar. En efecto, supongamos el caso particular de una señal BPSK recibida en un array y a la que se suma ruido Gaussiano de varianza σ_n . Según se demuestra en el apéndice 8.1, la función de coste evaluada a la salida del ecualizador puede representarse como la esperanza respecto a la distribución BPSK de la no linealidad convolucionada con la función de densidad del ruido. A medida que vamos aumentando la varianza del ruido, llegamos a un punto en que la convolución con la densidad del ruido enmascara los dos mínimos de la no-linealidad de forma que la minimización del criterio no puede asegurar la recuperación de la señal BPSK.

$$J = -E_{z_k = \lambda a_k + \eta_k(\sigma_n)} \ln p_{a_k + n_k}(z_k) = -E_{z_k^a = \lambda a_k} \left(p_{-\eta_k(\sigma_n)} * \ln p_{a_k + n_k} \right) (z_k^a) = -E_{z_k^a = \lambda a_k} \chi(z_k^a, \sigma_n) \tag{4.4.17}$$

Este efecto se debe al hecho de que la función de densidad de la variable z_k ya tampoco presenta dos máximos distinguibles y por consiguiente no se adapta a la forma de la no-linealidad log-likelihood. En la siguiente figura presentamos un ejemplo de la evolución de la no-linealidad χ definida en la ecuación anterior,

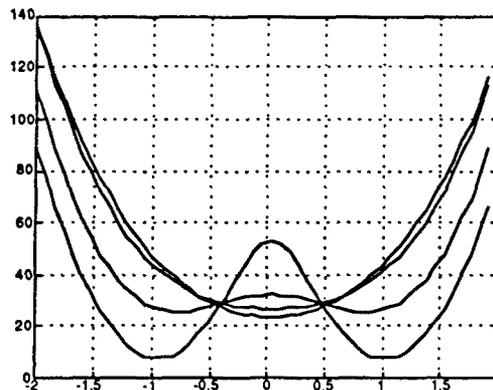


Figura 4.4.1: Representación de la no-linealidad $\chi(z_k, \sigma_n)$ en la ecuación anterior en función de la varianza σ_n del ruido Gaussiano aditivo. Obsérvese como para varianzas crecientes los dos valles de la no-linealidad se suavizan hasta desaparecer. Véase también el efecto de contracción de la constelación: los mínimos suben de nivel y se desplazan hacia el centro para valores crecientes de σ_n .

Podemos establecer a continuación una relación entre el parámetro varianza óptimo, σ_1^2 , y la varianza de detección definida anteriormente, σ_d^2 . Consideremos la expresión de la primera definida en la ecuación 4.3.24 y expandámosla,

$$\tag{4.4.18}$$

$$\begin{aligned} \sigma_r^2 &= E_{y_k} E_{a_k'} |y_k - a_k'|^2 q(y_k, a_k') \\ &= E_{y_k} \left(y_k^* E_{a_k'} (y_k - a_k') q(y_k, a_k') - E_{a_k'} a_k'^* (y_k - a_k') q(y_k, a_k') \right) \\ &= E_{y_k} \left(y_k^* (y_k - \hat{a}_k) - y_k \hat{a}_k^* + E_{a_k'} |a_k'|^2 q(y_k, a_k') \right) \end{aligned}$$

Recurriendo entonces a la ecuación 4.4.11, introducimos el factor γ en la ecuación,

$$\sigma_r^2 = E_{y_k} \left(y_k^* (y_k - \hat{a}_k) - y_k \hat{a}_k^* + |\hat{a}_k|^2 + \sigma_r^2 \gamma(y_k) \right) = E_{y_k} \left(|y_k - \hat{a}_k|^2 + \sigma_r^2 \gamma(y_k) \right) \tag{4.4.19}$$

con lo cual, pasando el factor γ a la izquierda de la ecuación obtenemos la siguiente relación fundamental entre ambas varianzas,

$$\sigma_r^2 + E_{y_k} \sigma_d^2(y_k) = E_{y_k} |y_k - \hat{a}_k|^2 \tag{4.4.20}$$

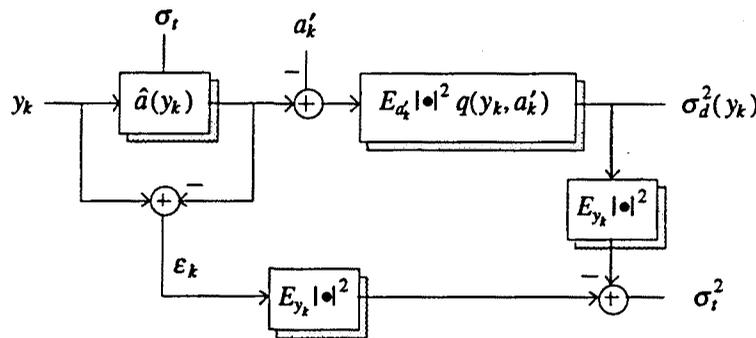


Figura 4.4.2: relación entre las varianzas involucradas en la definición de la función de coste.

Podemos obtener algunas conclusiones de esta relación. Entre ellas una cota superior del parámetro varianza es la potencia del ruido a la salida del filtro en relación a la referencia regenerada o potencia del error generalizado,

$$\sigma_r^2 \leq E_{y_k} |y_k - \hat{a}_k|^2 \tag{4.4.21}$$

La ecuación 4.4.18 nos permite hallar una relación final entre las potencias de la señal filtrada y_k y de la señal regenerada \hat{a}_k en términos de las diferentes varianzas de ruido representadas en la figura 4.4.1. Expandamos el término de error generalizado de la ecuación 4.4.8 incluyendo también la relación de conservación de la fase deducida en la ecuación 4.3.23,

$$\tag{4.4.22}$$

$$\begin{aligned}\sigma_t^2 + E_{y_k} \sigma_d^2(y_k) &= E_{y_k} \left(|y_k|^2 + |\hat{a}_k|^2 - 2 \operatorname{Re}\{\hat{a}_k^* y_k\} \right) , \quad E_{y_k} |y_k|^2 = E_{y_k} \hat{a}_k^* y_k \\ &= -E_{y_k} |y_k|^2 + E_{y_k} |\hat{a}_k|^2\end{aligned}$$

con lo cual llegamos finalmente a la relación de potencias deseada,

$$\sigma_t^2 + E_{y_k} \sigma_d^2(y_k) + E_{y_k} |y_k|^2 = E_{y_k} |\hat{a}_k|^2 \quad (4.4.23)$$

4.4.1 Otras Arquitecturas

En el apartado anterior consideramos el caso particular de una arquitectura lineal en los coeficientes y en los datos. Podemos obtener una expresión más general del gradiente y el Hessiano (sólo la correspondiente al gradiente se presenta en este apartado) para arquitecturas no lineales en los coeficientes. Aunque en este trabajo suponemos siempre linealidad en los datos, veremos que la expresión que deducimos a continuación no requiere de esta condición. Incluso ejemplos sencillos de aplicación (conformación de haz más eualización) pueden conducir a arquitecturas cúbicas en los coeficientes (algunos de ellos restringidos en la recta real: véase el apartado de aplicaciones). Si observamos con detenimiento el desarrollo de las demostraciones del apartado anterior, notamos que se ha recurrido a la premisa de arquitectura lineal sólo en aquellos puntos donde era necesario calcular el gradiente de la arquitectura. Utilizando ámpliamente la hipótesis general de no-analiticidad de la función de coste y de la arquitectura llegamos a la siguiente expresión general para el gradiente de la función de coste (donde suponemos el caso general de vectores de coeficientes complejos, las ecuaciones son fácilmente particularizables al caso de vectores reales por simple sustitución),

$$\nabla_{\mathbf{w}_k^H} J = -E_{a_t, n_t} \left(\frac{\partial}{\partial y_k} \ln p_{a_t + \eta_t}(y_k) \right) \nabla_{\mathbf{w}_k^H} y_k - E_{a_t, n_t} \left(\frac{\partial}{\partial y_k^*} \ln p_{a_t + \eta_t}(y_k) \right) \nabla_{\mathbf{w}_k^H} y_k^* \quad (4.4.2.1)$$

No obstante, las funciones de densidad derivadas a partir de la convolución de una función de densidad de una variable aleatoria determinada más ruido Gaussiano cumplen la siguiente propiedad,

$$(4.4.2.2)$$

$$\begin{aligned} \frac{\partial}{\partial y_k} p_{a_k + \eta_k}(y_k) &= \frac{\partial}{\partial y_k} E_{a'_k} \frac{1}{\pi\sigma^2} e^{-|y_k - a'_k|^2 / \sigma^2} = E_{a'_k} \frac{1}{\pi\sigma^2} e^{-|y_k - a'_k|^2 / \sigma^2} \frac{1}{\sigma^2} (y_k - a'_k)^* \\ &= \left(\frac{\partial}{\partial y_k^*} p_{a_k + \eta_k}(y_k) \right)^* \end{aligned}$$

puesto que los términos esperanza y conjugado son intercambiables. El gradiente de la función de coste de expresa en función del error generalizado y el gradiente con respecto a la arquitectura de la siguiente forma (no es posible una mayor descomposición del gradiente de la arquitectura en términos generales),

$$\nabla_{\mathbf{w}_k^H} J = -E_{a_k, n_k} \varepsilon^*(y_k) \nabla_{\mathbf{w}_k^H} y_k - E_{a_k, n_k} \varepsilon(y_k) \nabla_{\mathbf{w}_k^H} y_k^* \tag{4.4.2.3}$$

Vemos pues que para el caso de un vector de coeficientes complejo donde la arquitectura de filtrado se define en términos del producto de su hermítico con el vector de datos, el segundo término de la ecuación anterior (gradiente de la arquitectura conjugada) se anula, obteniendo la expresión usual del gradiente. El paso al Hessiano se deriva fácilmente.

4.5 Expansión de la función de coste mediante cumulantes

En este apartado deduciremos una relación entre la función de coste no condicionada y los cumulantes de una variable aleatoria aún por definir. El objetivo es demostrar que con la función de coste definida estamos manejando implícitamente estadísticos suficientes de todos los ordenes. Consideremos la expresión de la potencia del error $|\varepsilon_J|^2$ definida en la ecuación 3.4.1.1.6 y comparémosla con la expresión de la función generadora de cumulantes de una variable aleatoria u cualquiera,

$$\frac{1}{\sigma_i^2} |\varepsilon_J|^2 = -\ln E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} \Leftrightarrow \Psi_u(s) = \ln E_u e^{su} \tag{4.5.1}$$

Donde se considera en general que la variable s puede tomar cualquier valor en el plano complejo, es decir, se define en el sentido de Laplace. La particularización $s = -j\omega$ implica una definición en el sentido de Fourier. A la vista de la semejanza formal entre ambas expresiones nos preguntamos si existe una particularización de la función generadora de cumulantes de la variable aleatoria u tal que resulte la potencia del error $|\varepsilon_J|^2$ normalizada mediante el parámetro varianza. En efecto, consideremos la siguiente particularización de la variable aleatoria u función de a'_k ,

$$u(a'_k) = |z_k - a'_k|^2 \Rightarrow E_u e^{su} = E_{a'_k} e^{s|z_k - a'_k|^2} \tag{4.5.2}$$

donde obtenemos una relación entre los argumentos de los logaritmos de las expresiones 4.5.1 mediante aplicación del Teorema Fundamental de la Esperanza. Destacamos que la variable z_k se considera igual a una constante en la antedicha expresión. Vemos finalmente que tras sustitución del parámetro frecuencia por un valor adecuado llegamos a la relación entre la función de coste o potencia del error $|\varepsilon_J|^2$ y una expansión de cumulantes,

$$\begin{aligned} s = \frac{-1}{\sigma_i^2} &\Rightarrow \frac{1}{\sigma_i^2} |\varepsilon_J|^2 = -\Psi_{|z_k - a'_k|^2} \left(\frac{-1}{\sigma_i^2} \right) = -\ln \sum_{v=0}^{+\infty} \frac{(-1)^v}{v! \sigma_i^{2v}} E_{a'_k} |z_k - a'_k|^{2v} \\ &= -\sum_{v=0}^{+\infty} \frac{1}{v!} cum_v \left[|z_k - a'_k|^2 \right] \left(\frac{-1}{\sigma_i^2} \right)^v \geq 0 \end{aligned} \tag{4.5.3}$$

donde la serie de cumulantes obtenida es definida positiva por construcción de $|\varepsilon_J|^2$. La función de coste se puede expresar finalmente como,

$$J = \ln \pi \sigma_i^2 - \sum_{v=0}^{+\infty} \left(\frac{-1}{\sigma_i^2} \right)^v \frac{1}{v!} E_{z_k} cum_v \left[|z_k - a'_k|^2 \right] \tag{4.5.4}$$

Hemos visto pues que la función de coste propuesta puede interpretarse como un criterio implícito basado en estadísticas de orden superior utilizando inteligentemente información lateral de la función de densidad deseada mediante el operador esperanza E_{a_k} presente en la ecuación 4.5.1. Podemos interpretar a la vista de la ecuación 4.5.3 que la función de coste tiende a minimizar una serie de los cumulantes del error, es decir, tiende a forzar un error Gaussiano.

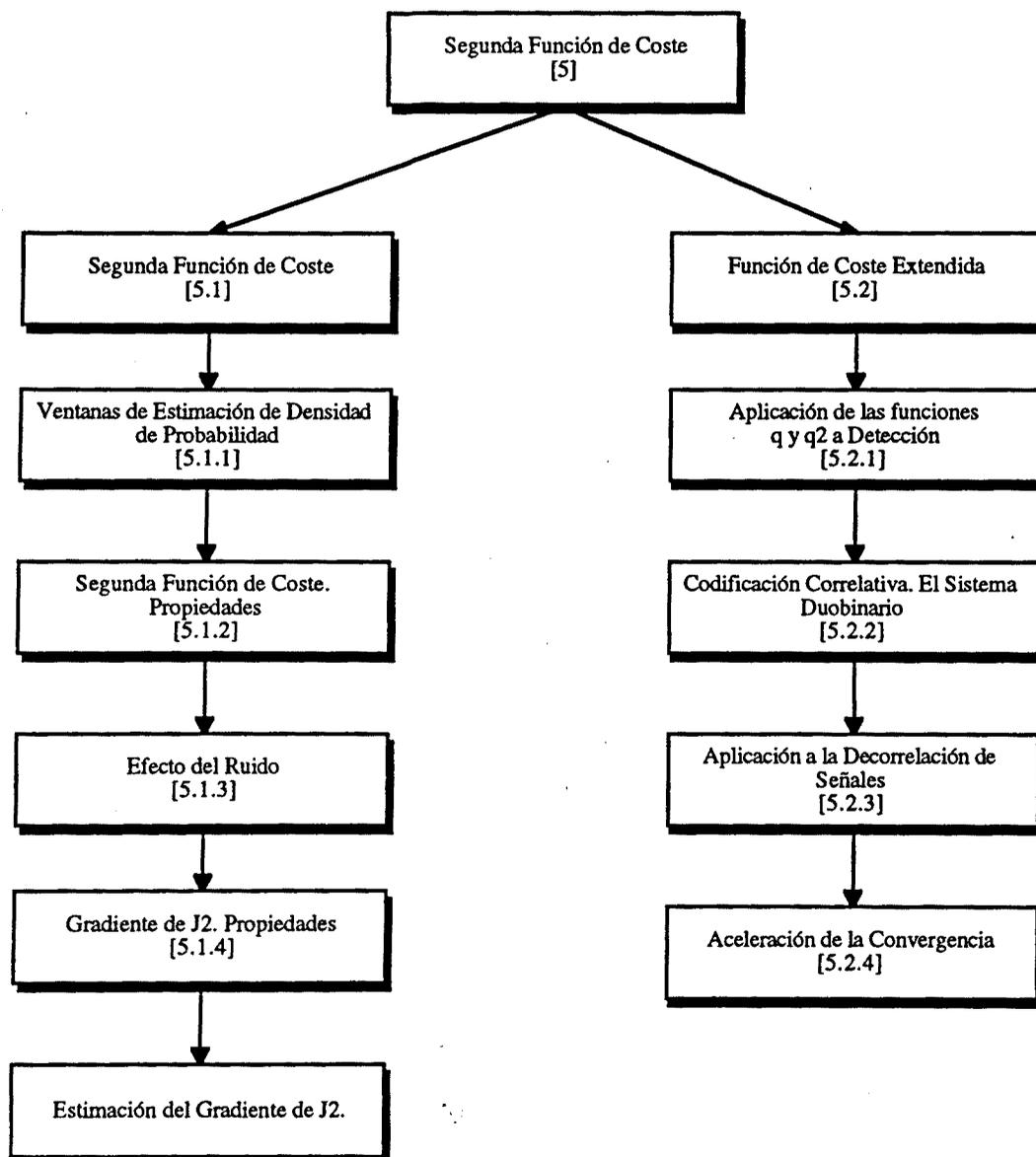
Basándonos en la ecuación 4.5.1 podemos hallar una nueva interpretación de la función característica. Consideremos la función de densidad de probabilidad de una variable aleatoria a_k+n_k donde n_k es ruido Gaussiano y a_k cualquier distribución, tenemos entonces que,

$$p_{a_k+n_k}(z_k) = \frac{1}{\pi\sigma_k^2} \Phi_{|z_k-a_k|^2} \left(\frac{-1}{\sigma_k^2} \right) = \frac{1}{\pi\sigma_k^2} \sum_{v=0}^{+\infty} \frac{(-1)^v}{v!\sigma_k^{2v}} E_{a_k} |z_k - a_k|^{2v} \quad (5.5.3.4)$$

Al considerar la función característica de la norma del error, obtenemos la función de densidad de la suma a_k+n_k .

Capítulo V

Segunda Función de Coste



5.1 Segunda Función de Coste

Hasta ahora hemos estudiado la función de coste J proveniente de uno de los términos de la discriminación de Kullback-Leibler. Tal como se razonó en la sección 3.2, el hecho de quedarse con un único término provenía del desconocimiento de la función de densidad real a la salida del sistema adaptativo. Debido a esto, según sean las condiciones del problema, pueden aparecer efectos no deseados. En esta sección introduciremos una nueva función de coste que solventa estas deficiencias, buscando un criterio más cercano a la discriminación KL. Examinemos primero problemas presentados por la primera función de coste,

$$J = -E_{z_k} \ln p_{a_k+n_k}(z_k) = \ln \pi \sigma_i^2 - E_{z_k} \ln E_{a_k} e^{-|z_k - a_k|^2 / \sigma_i^2} \quad (5.1.1)$$

1.- Convergencia a la solución trivial: consideremos el caso simple de una función de densidad de probabilidad con un máximo en cero, por ejemplo una modulación duobinaria. Puede ocurrir según se inicialicen los coeficientes del filtro adaptativo que el vector de coeficientes tienda a la solución nula. Entonces la función de coste no reacciona puesto que al estar la salida permanentemente a cero, su gradiente o error generalizado es también cero.

2.- Efecto de captura a señales con función de densidad diferente a la deseada: este efecto ocurre por ejemplo cuando nos hallamos en presencia de dos variables aleatorias que presentan máximos comunes en sus funciones de densidad respectivas (p.ej. BPSK y QPSK). Debido a que la primera función de coste no es sensible a la frecuencia con que se da cada uno de los máximos a la salida del filtro adaptativo, puede suceder que utilizando una función de coste basada en una BPSK capturemos la señal QPSK, y viceversa. Dependiendo de la potencia de cada una de las señales, el mínimo asociado a la señal QPSK cuando utilizamos la función de coste BPSK puede ser incluso menor que cuando capturamos la señal BPSK deseada. No obstante, aunque la primera función de coste presenta este problema en adquisición, suponiendo que converja al mínimo deseado, su comportamiento en seguimiento es correcto.

3.- Insensibilidad a la independencia estadística: en el caso de que deseemos aplicar la función de coste definida como hasta ahora a la separación de señales de distribución conocida (utilizando su función de densidad conjunta) mediante una matriz de separación o de decorrelación, puede suceder dependiendo de las condiciones iniciales de la matriz y de la potencia de las señales, que las señales recuperadas a la salida de la etapa separadora no sean independientes dos a dos, sino que, por ejemplo, la señal más potente aparezca en

todas las salidas. Esta es otra consecuencia de la insensibilidad de la función de coste a la frecuencia con que se da cada máximo de la función de densidad conjunta. Es decir, al recuperar la misma señal en todas las salidas, solo se está dando un subconjunto de todos los centroides posibles de la constelación conjunta del vector de señales de salida.

Para situaciones como las anteriores en que puedan aparecer señales interferentes y funciones de densidad deseadas con máximos en cero se hace necesario introducir ciertas modificaciones en la función de coste.

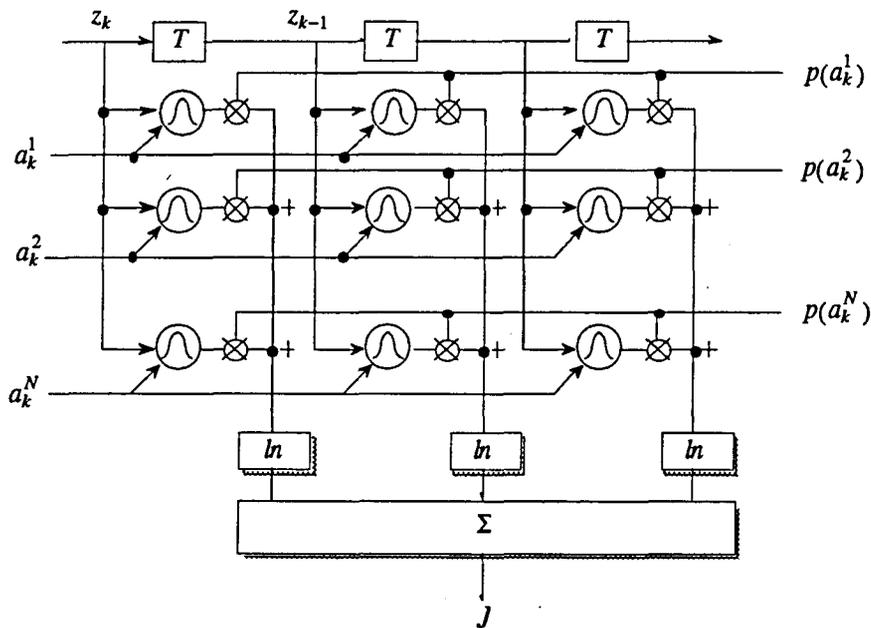


Figura 5.1.1: evaluación de la función de coste J

Hemos visto hasta ahora que la función de coste propuesta adolece de algunos problemas como el efecto de captura a distribuciones de densidad de probabilidad erróneas debido al hecho de que no es sensible a las probabilidades de cada centroide de la constelación, sino sólo a su carácter de máximo. Proponemos en esta sección una función de coste que al mismo tiempo que sea sensible a la varianza del ruido, sea más robusta frente al efecto de captura. No obstante, la nueva función seguirá siendo en principio multimodal aunque habremos eliminado modos problemáticos de la función de coste antigua. Básicamente, se trata de introducir sensibilidad a la dependencia estadística. Partiendo de la función de coste anterior, realizamos un simple intercambio de operadores esperanza para obtener la nueva función,

$$J = -E_{z_t} \ln E_{a_i} \frac{1}{\pi\sigma_i^2} e^{-|z_t - a_i|^2 / \sigma_i^2} \Leftrightarrow J_2 = -E_{a_i} \ln E_{z_t} \frac{1}{\pi\sigma_i^2} e^{-|z_t - a_i|^2 / \sigma_i^2} \tag{5.1.2}$$

En las figuras 5.1.1 y 5.1.2 comparamos gráficamente como se evalúan ambas funciones de coste,

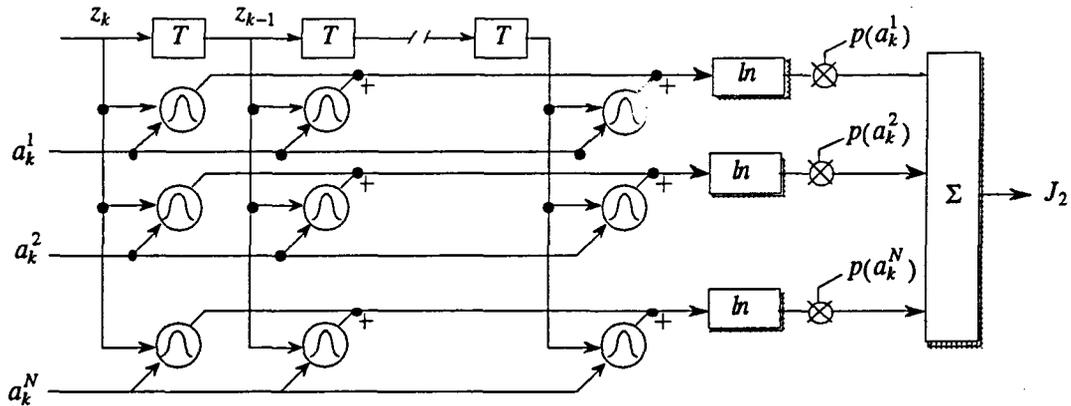


Figura 5.1.2: evaluación de la función de coste J_2 .

5.1.1 Ventanas de Estimación de Densidad de Probabilidad.

Introduciremos aquí el concepto de estimación de una densidad de probabilidad para el caso general de variables aleatorias continuas y reales, suponiendo procesos ergódicos. La generalización a variables aleatorias complejas es trivial. Una función de densidad se define a partir de la función de distribución como,

$$p_X(x_o) = \lim_{\Delta x \rightarrow 0} \frac{P(x_o + \Delta x) - P(x_o)}{\Delta x} = \lim_{\Delta x \rightarrow 0} \frac{\Pr(x_o \leq x < x_o + \Delta x)}{\Delta x} \tag{5.1.1.1}$$

Ahora bien, la probabilidad de que la variable aleatoria X caiga en un determinado intervalo puede expresarse mediante la esperanza de la función indicatriz, I , del intervalo, como,

$$\Pr(x_o \leq x < x_o + \Delta x) = E_X I(x; [x_o, x_o + \Delta x]) = E_X \Pi \left(\frac{x - (x_o + \frac{1}{2} \Delta x)}{\Delta x} \right), \quad \Pi(a) \stackrel{def}{=} \begin{cases} 1 & |a| < \frac{1}{2} \\ 0 & |a| > \frac{1}{2} \end{cases} \tag{5.1.1.2}$$

Debido a que trataremos con funciones de densidad continuas, no es necesario detenerse en la definición de Π en la frontera $|a|=1/2$. Entonces, la función de densidad puede estimarse como,

$$\tag{5.1.1.3}$$

$$p_X(x_0) = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} E_X I(x; [x_0, x_0 + \Delta x])$$

En la práctica, la estimación perfecta de la función de densidad requiere un registro de datos infinito para poder realizar el límite $\Delta x \rightarrow 0$. Por lo cual, para registros de datos finitos se debe emplear una función indicatriz o ventana de estimación de anchura no nula, $\Delta x > 0$. O bien si se dispone de un registro de datos infinito y se utiliza una ventana $\Delta x > 0$, la estimación de la función de densidad es polarizada.

Veremos que en el planteamiento de la nueva función de coste J_2 aparece de forma natural el concepto de función indicatriz y de estimación de densidad mediante una ventana, en este caso no rectangular sino Gausiana. Es decir, pueden considerarse una infinidad de ventanas de estimación $\phi(x; [x_0, x_0 + \Delta x])$ con la propiedad asintótica siguiente, sin que éstas sean necesariamente rectangulares,

(5.1.1.4)

$$p_X(x_0) = \lim_{\Delta x \rightarrow 0} \frac{1}{\Delta x} E_X \phi(x; [x_0, x_0 + \Delta x])$$

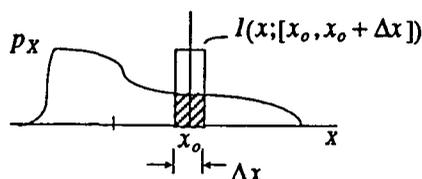


Figura 5.1.1.1: Estimación mediante ventanas de una función de densidad.

Tendremos ocasión de comprobar esta afirmación más adelante. Puesto que la función indicatriz es de hecho una no-linealidad sobre los datos, su esperanza equivale a filtrar la salida de una no-linealidad, lo cual es fácilmente implementable disponiendo de una realización de un proceso estocástico ergódico.

Podemos establecer un paralelo entre la estimación de una densidad de probabilidad y los algoritmos de estimación espectral. Al igual que para estimar la contribución de energía de un proceso a una determinada frecuencia, utilizamos un filtro de ancho de banda estrecho sintonizado a dicha frecuencia y estimamos la potencia de salida, en la estimación de una densidad de probabilidad utilizamos una no-linealidad estrecha 'sintonizada' a un determinado valor y filtramos mediante el operador esperanza. Los efectos de interferencia de otras frecuencias cuando la banda del filtro no es suficientemente estrecha son totalmente análogos a los de densidad de probabilidad de valores de la variable aleatoria contiguos al estimado.

5.1.2 Segunda Función de Coste. Propiedades

Podemos ver a partir de esta definición que el argumento del logaritmo neperiano puede considerarse como una estimación de densidad de probabilidad donde la exponencial Gaussiana juega el papel de función indicatriz. Examinemos las figuras 5.1.1.1-2. Podemos ver que ambas están basadas en una red de no-linealidades o funciones indicatrices Gaussianas. Mientras que en la primera función de coste, el filtrado se realiza en base a columnas de nodos, en la segunda se realiza en base a filas. En ambos casos, antes de realizar la acumulación final, se aplica una no-linealidad (logarítmica) a cada una de las acumulaciones parciales. Básicamente, los problemas anteriormente enumerados de la primera función de coste son debidos a que al realizar una integración por columnas, perdemos información estadística puesto que utilizamos una única muestra temporal. Es decir, no sabemos con qué frecuencia ocurre cada uno de los centroides. En cambio, al realizar un filtrado por filas, conservamos esta información.

En la implementación real de la función de coste, se puede utilizar el gradiente instantáneo de la no-linealidad, con lo cual se pierde información estadística. Esto no ocurre con la nueva función de coste puesto que conserva historia previa de las señales.

Consideremos ahora la siguiente norma del error definida análogamente a la norma del error de la ecuación 3.4.1.1.6 como,

$$|\varepsilon_{2,J}|^2 \stackrel{def}{=} -\sigma_i^2 E_{z_k} \ln E_{z_k} e^{-|z_k - a_k'|^2 / \sigma_i^2} \quad (5.1.2.1)$$

podemos introducir la función \hat{p} como una estimación de la densidad de probabilidad. Entonces,

$$\hat{p}(z_k = a_k') \stackrel{def}{=} \frac{1}{\pi\sigma_i^2} E_{z_k} e^{-|z_k - a_k'|^2 / \sigma_i^2} \quad (5.1.2.2)$$

Es importante tener en cuenta que \hat{p} es de hecho una función de densidad de probabilidad puesto que su integral sobre toda su región de soporte es igual a la unidad. En efecto, consideremos,

$$\int_C \hat{p}(z_k = a_k') da_k' = \int_C \frac{1}{\pi\sigma_i^2} E_{z_k} e^{-|z_k - a_k'|^2 / \sigma_i^2} da_k' = E_{z_k} \int_C \frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k'|^2 / \sigma_i^2} da_k' = E_{z_k} 1 = 1 \quad (5.1.2.3)$$

Podemos además ver qué densidad de probabilidad resulta al utilizar el estimador anterior teniendo en cuenta la expresión para la densidad de probabilidad de la suma de dos variables aleatorias independientes presentada en la Sección 5.2.4,

$$\hat{p}(z_k = a'_k) = \frac{1}{\pi\sigma_t^2} E_{z_k} e^{-|z_k - a'_k|^2 / \sigma_t^2} = p_{z_k + n_t}(a'_k) \quad \text{con} \quad p_{n_t}(n_t) = \frac{1}{\pi\sigma_t^2} e^{-|n_t|^2 / \sigma_t^2} \quad (5.1.2.4)$$

Vemos pues que con tal estimador obtenemos una densidad de probabilidad equivalente a la de la densidad que deseamos estimar más un término de ruido gaussiano aditivo de varianza igual a la apertura σ_t de la ventana Gaussiana utilizada. Este efecto es debido a un 'leakage' de densidad de probabilidad proveniente de los valores cercanos a $z_k = a'_k$, que caen dentro del lóbulo de la ventana. Dicho efecto es análogo al 'leakage' de frecuencia presente en algoritmos de estimación espectral debido al tipo de ventana temporal que se utilice. En el límite asintótico cuando la varianza tiende a cero obtenemos una estimación exacta de la función de densidad,

$$\begin{aligned} \lim_{\sigma_t \rightarrow 0} \hat{p}(z_k = a'_k) &= \lim_{\sigma_t \rightarrow 0} p_{z_k + n_t}(a'_k) = \lim_{\sigma_t \rightarrow 0} p_{z_k}(a'_k) * p_{n_t}(a'_k) = p_{z_k}(a'_k) * \lim_{\sigma_t \rightarrow 0} p_{n_t}(a'_k) = \\ &= p_{z_k}(a'_k) * \delta(a'_k) = p_{z_k}(a'_k) \end{aligned} \quad (5.1.2.5)$$

Tenemos entonces que,

$$J_2 = \ln \pi \sigma_t^2 + \frac{1}{\sigma_t^2} |E_{2,J}|^2 = -E_{a'_k} \ln \hat{p}(a'_k) = E_{a'_k} \ln \frac{p(a'_k)}{\hat{p}(a'_k)} - E_{a'_k} \ln p(a'_k) \quad (5.1.2.6)$$

De donde podemos observar que se puede relacionar la función de error anterior con una función similar a la discriminación (a la que podemos llamar análogamente pseudo-discriminación),

$$J_2 = \hat{D}(a'_k || z_k) + H(a'_k) \quad (5.1.2.7)$$

Este desarrollo de momento es válido para variables aleatorias de densidad continua. Examinemos qué sucede para variables aleatorias discretas. Para variables discretas debe tenerse en cuenta que ambos sumandos de la ecuación anterior divergen independientemente pero su suma converge. Debido a que la anterior norma del error puede descomponerse en dos sumandos, la pseudo-discriminación y la entropía diferencial de la variable aleatoria deseada sin ruido, tenemos que la minimización de la función de coste es análoga a la minimización de la pseudo-discriminación puesto que el segundo término, la entropía, es independiente de cual sea la distribución de la variable aleatoria z_k . Evidentemente, es

deseable que cuanto menor sea la varianza de ruido que afecte a z_k , tanto menor sea la pseudo-discriminación y que la función de densidad estimada se acerque tanto más a la deseada

5.1.3 Efecto del Ruido

Hemos visto hasta ahora que el criterio anterior es equivalente a la minimización de una función de discriminación respecto a la variable aleatoria objetivo (en el caso de variables aleatorias continuas). Intuitivamente, dada una variable aleatoria objetivo a_k , y por otra parte una variable $z_k = \mathbf{w}^H(a_k \mathbf{s}_k + \mathbf{n}_k)$ la minimización de la discriminación con respecto al vector de coeficientes induce a pensar que $z_k \rightarrow a_k + n_k$, con n_k ruido Gaussiano de mínima varianza. Pongamos un caso ejemplo en que a_k es una variable aleatoria BPSK y supongamos una restricción del tipo $\mathbf{w}^H \mathbf{s}_k = 1$. La función de coste puede expresarse según la ecuación 5.1.2 como,

$$J_2 = -E_{a'_k} \ln p_{z_k + n_k}(a'_k) = -E_{a'_k} \ln p_{a_k + n_k + n_k}(a'_k) = -E_{a'_k} \ln E_{a_k} \frac{1}{\pi \sigma^2} e^{-|a'_k - a_k|^2 / \sigma^2} \Bigg|_{\sigma^2 = \sigma_n^2 + \sigma_s^2} \tag{5.1.3.1}$$

siendo,

$$\sigma_n^2 = E|n_k|^2 = \mathbf{w}_k^H \mathbf{R}_{nn} \mathbf{w}_k \tag{5.1.3.2}$$

Obtenemos finalmente para el caso de una variable aleatoria discreta (para el caso de una variable continua es necesario evaluar una integral) que la función de coste se expresa como,

$$J_2 = -\sum_{i'=1}^M p_{i'} \ln \left(\frac{1}{\pi(\sigma_n^2 + \sigma_s^2)} \sum_{i=1}^M p_i e^{-|a^{i'} - a^i|^2 / (\sigma_n^2 + \sigma_s^2)} \right) \tag{5.1.3.3}$$

Donde M es la dimensión del alfabeto y el conjunto $\{p_i\}$ constituye la probabilidad de los centroides de la constelación. Representamos en la siguiente figura la función de coste en términos de σ^2 . Obsérvese su valor decreciente para varianzas de ruido sucesivamente menores.

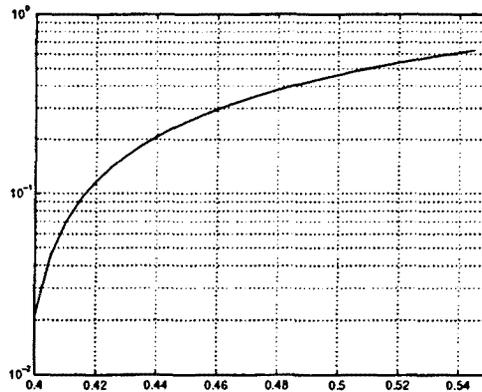


Figura 5.1.3.1: Comportamiento de la función de coste de la ecuación 5.1.3.3 en función de la potencia de ruido. Obsérvese como un decrecimiento de la función de coste conlleva una disminución de la potencia de ruido.

5.1.4 Gradiente de J_2 . Propiedades

La optimización de la función de coste propuesta se realizará igualmente con técnicas de gradiente. Veremos también en este apartado algunas relaciones importantes con respecto a la función de regeneración bajo esta función de coste, como por ejemplo cómo se expresa el vector óptimo de coeficientes en analogía a la solución de Wiener. Calculamos el gradiente con respecto al vector de coeficientes hermítico,

$$\nabla_{\mathbf{w}_k^H} J_2 = \frac{1}{\sigma_i^2} E_{a_k'} \frac{E_{r_k} e^{-|z_k - a_k'|^2 / \sigma_i^2} (z_k - a_k')^* r_k}{E_{r_k} e^{-|z_k - a_k'|^2 / \sigma_i^2}} = \frac{1}{\sigma_i^2} E_{a_k'} E_{r_k} q_2(z_k, a_k') (z_k - a_k')^* r_k \tag{5.1.4.1}$$

donde definimos el siguiente conjunto análogo de funciones q_2 ,

$$q(z_k, a_k) = \frac{e^{-|z_k - a_k|^2 / \sigma_i^2}}{E_{a_k'} e^{-|z_k - a_k'|^2 / \sigma_i^2}} \Leftrightarrow q_2(z_k, a_k) = \frac{e^{-|z_k - a_k|^2 / \sigma_i^2}}{E_{r_k} e^{-|z_k - a_k|^2 / \sigma_i^2}} \tag{5.1.4.2}$$

con las siguiente propiedad análoga,

$$E_{a_k'} q(z_k, a_k') = 1 \Leftrightarrow E_{r_k} q_2(z_k, a_k) = 1 \tag{5.1.4.3}$$

Tenemos ahora que la evaluación de las funciones q_2 , no se hace en el eje estadístico sino en el eje temporal y son por tanto sensibles a la distribución de probabilidad real a la salida del filtro. Es decir, las funciones $q_2(\cdot)$ se encuentran ahora normalizadas no respecto a la esperanza de la variable aleatoria \hat{a}_k , sino con respecto a la esperanza de la variable aleatoria real que tenemos a la salida del sistema. Esta definición presenta mayor robustez que en la primera función de coste puesto que el denominador de $q_2(\cdot)$ es directamente expresable en términos de la densidad de probabilidad estimada para la variable z_k , como destacamos en la siguiente ecuación,

$$q_2(z_k, a_k) = \frac{\frac{1}{\pi\sigma_t^2} e^{-|z_k - a_k|^2 / \sigma_t^2}}{\hat{p}(z_k = a_k)} \tag{5.1.4.4}$$

A diferencia de lo que ocurría en la primera función de coste, donde opcionalmente se permitía la actualización del parámetro σ_t , tal procedimiento carece de sentido en este contexto puesto que el significado de σ_t no es el de una varianza de ruido, sino el de un parámetro de resolución en la estimación de una densidad de probabilidad. En su lugar, deben actualizarse las funciones $q_2(\cdot)$, las cuales dependen de la estadística de la salida z_k . Las funciones de calidad para la primera función de coste eran en cambio fijas y precalculables. También se define ahora un error generalizado en función del cual se expresa el gradiente de la función de coste de la ecuación 5.1.4.1,

$$\varepsilon_{2,k} \stackrel{def}{=} E_{a'_k} (z_k - a'_k) q_2(z_k, a'_k) \Rightarrow \nabla_{\mathbf{w}_k^*} J_2 = \frac{1}{\sigma_t^2} E_{\mathbf{r}_k} \varepsilon_{2,k}^* \mathbf{r}_k \tag{5.1.4.5}$$

Observemos la similitud formal con la primera función de coste, donde ahora el error generalizado sí que es función de las propiedades estimadas de cada centroide. Podemos definir igualmente una función de regeneración de referencia como,

$$\hat{a}_{2,k} \stackrel{def}{=} E_{a'_k} a'_k q_2(z_k, a'_k) \tag{5.1.4.6}$$

Debido a que con la segunda función de coste nos ha sido posible definir una función $q_2(\cdot)$, propiedades análogas a las deducidas en la sección 4.3 en relación a la función de regeneración siguen siendo aplicables. Su demostración sigue idénticas líneas. Tomemos la expresión del gradiente de la ecuación 5.1.4.1 e igualémoslo a cero para que se cumpla la condición de mínimo,

$$E_{a'_k} E_{\mathbf{r}_k} q_2(z_k, a'_k) (z_k - a'_k)^* \mathbf{r}_k = \mathbf{0} \Rightarrow E_{a'_k} E_{\mathbf{r}_k} q_2(z_k, a'_k) z_k^* \mathbf{r}_k = E_{a'_k} E_{\mathbf{r}_k} q_2(z_k, a'_k) a_k^* \mathbf{r}_k \tag{5.1.4.7}$$

Sustituyamos ahora z_k por $w_k^H r_k$ en el lado izquierdo de la ecuación, al mismo tiempo que expresamos el lado derecho en términos de la referencia regenerada $\hat{a}_{2,k}$,

$$E_{a'_k} E_{r_k} q_2(z_k, a'_k) r_k r_k^H w_k = E_{r_k} \hat{a}_{2,k}^* r_k \quad (5.1.4.8)$$

Finalmente, aislamos el vector de coeficientes w_k mediante inversión de la matriz del lado izquierdo de la ecuación anterior,

$$w_k = [E_{a'_k} E_{r_k} q_2(z_k, a'_k) r_k r_k^H]^{-1} E_{r_k} \hat{a}_{2,k}^* r_k = \tilde{R}_{rr}^{-1} E_{r_k} \hat{a}_{2,k}^* r_k \quad (5.1.4.9)$$

Nótese la gran similitud que presenta dicha ecuación con la solución de Wiener. Hemos definido una matriz de "pseudo-correlación" de los datos, \tilde{R}_{rr} con tilde, donde cada producto diádico del vector de datos r_k , $r_k r_k^H$, aparece afectado por un factor de peso (definido positivo ya que se expresa como una esperanza de la función $q_2(\cdot)$ también definida positiva) función de la salida del filtro z_k y de la distribución que deseamos recuperar, a'_k ,

$$\tilde{R}_{rr} \stackrel{def}{=} E_{r_k} \zeta(z_k) r_k r_k^H, \quad \zeta(z_k) \stackrel{def}{=} E_{a'_k} q_2(z_k, a'_k) \neq 1 \\ \zeta(z_k) \geq 0 \quad (5.1.4.10)$$

Análogamente se cumple la propiedad de conservación de la fase de la función de regeneración, la cual ya deducimos para la primera función de coste en la ecuación 4.3.23. Tomemos la ecuación 5.1.4.7 y multipliquemos a ambos lados por el hermítico del vector de coeficientes,

$$E_{a'_k} E_{r_k} q_2(z_k, a'_k) z_k^* w_k^H r_k = E_{a'_k} E_{r_k} q_2(z_k, a'_k) a_k^* w_k^H r_k \quad (5.1.4.11)$$

sustituyendo en ambos lados el factor $w_k^H r_k$ por z_k y expresando el lado derecho en términos de la referencia regenerada, llegamos a la ecuación de conservación de la fase para la segunda función de coste,

$$E_{r_k} \zeta(z_k) |z_k|^2 = E_{r_k} \hat{a}_{2,k}^* z_k \Rightarrow \text{Im}\{E_{r_k} \hat{a}_{2,k}^* z_k\} = 0 \quad (5.1.4.12)$$

donde a diferencia de la primera función de coste aparece ahora el factor de peso (definido positivo) presente en la función de pseudo-correlación. Hemos visto en la ecuación 5.1.4.9 que el vector de coeficientes puede expresarse en términos de una matriz de pseudo-

correlación de los datos. Como vimos en el estudio de la primera función de coste en el capítulo 4, existe una forma alternativa para expresar el vector de coeficientes: en función de la matriz de correlación del ruido de entrada. En el siguiente apartado llegaremos a una solución de este tipo.

5.1.4.1 Versión alternativa del vector de coeficientes

Consideremos ahora el gradiente de la función de coste desde otro punto de vista a semejanza de como se realizó la primera deducción del vector de coeficientes para la primera función de coste en la sección 4.3. Previamente al cálculo del gradiente, expresemos la segunda función de coste en términos de la función de densidad de probabilidad del ruido Gaussiano a la salida del sistema como,

$$J_2 = -E_{a_k'} \ln E_{z_k} \frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k'|^2 / \sigma_i^2} = -E_{a_k'} \ln E_{\mathbf{a}_k, \mathbf{n}_k} \frac{1}{\pi\sigma_i^2} e^{-|\mathbf{w}_k^H \mathbf{S}_k \mathbf{a}_k + \mathbf{w}_k^H \mathbf{n}_k - a_k'|^2 / \sigma_i^2} \quad (5.1.4.1.1)$$

donde hemos sustituido la expresión de z_k por su valor en función del modelo de varias fuentes. Utilizando la propiedad del intercambio de convoluciones descrita en el apéndice 8.1, podemos eliminar la esperanza sobre el vector \mathbf{n}_k mediante modificación de la no-linealidad,

$$J_2 = -E_{a_k'} \ln E_{\mathbf{a}_k} \left(\frac{1}{\pi\sigma_i^2} e^{-|x - a_k'|^2 / \sigma_i^2} * p_{\eta_k}(x) \right) \Bigg|_{x = \mathbf{w}_k^H \mathbf{S}_k \mathbf{a}_k}, \quad p_{\eta_k}(x) = \frac{1}{\pi\sigma_w^2} e^{-|x|^2 / \sigma_w^2} \Bigg|_{\sigma_w^2 = \mathbf{w}_k^H \mathbf{R}_{nn} \mathbf{w}_k} \quad (5.1.4.1.2)$$

donde la no-linealidad se expresa ahora como la convolución de la Gaussiana con la función de densidad del ruido a la salida del sistema y evaluada en la variable aleatoria z_k a la que se le ha sustraído el término de ruido. Llamémosla $z_k^a = \mathbf{w}_k^H \mathbf{S}_k \mathbf{a}_k$. Mediante \mathbf{R}_{nn} denotamos la matriz de correlación del ruido de entrada. Por analogía con la suma de dos variables aleatorias sabemos que la convolución de dos Gaussianas debe dar otra Gaussiana, cuyo momento de segundo orden es igual a la suma de momentos (o potencias) de las dos Gaussianas individuales. Por lo tanto podemos expresar la función de coste como,

$$J_2 = -E_{a_k'} \ln E_{\mathbf{a}_k} \frac{1}{\pi(\sigma_i^2 + \sigma_w^2)} e^{-|z_k^a - a_k'|^2 / (\sigma_i^2 + \sigma_w^2)} \Bigg|_{\substack{z_k^a = \mathbf{w}_k^H \mathbf{S}_k \mathbf{a}_k \\ \sigma_w^2 = \mathbf{w}_k^H \mathbf{R}_{nn} \mathbf{w}_k}} = -E_{a_k'} \ln \psi(a_k') \quad (5.1.4.1.3)$$

donde definimos la función $\psi(\cdot)$ para simplificar la notación. A partir de la ecuación anterior y teniendo en cuenta las reglas de derivación ya expuestas en capítulos anteriores (sobre la derivación de funciones no-analíticas), el gradiente anterior se puede formular de la siguiente forma,

$$\nabla_{\mathbf{w}_k^H} J_2 = -E_{a_k'} \psi^{-1} \cdot E_{\mathbf{a}_k} \left\{ -\frac{1}{\pi(\sigma_r^2 + \sigma_w^2)^2} \mathbf{R}_{nn} \mathbf{w}_k + \frac{|z_k^e - a_k'|^2}{\pi(\sigma_r^2 + \sigma_w^2)^3} \mathbf{R}_{nn} \mathbf{w}_k - \frac{(z_k^e - a_k')^*}{\pi(\sigma_r^2 + \sigma_w^2)^2} \mathbf{S}_k \mathbf{a}_k \right\} e^{-|z_k^e - a_k'|^2 / (\sigma_r^2 + \sigma_w^2)} \quad (5.1.4.1.4)$$

Agrupando términos llegamos a,

$$\nabla_{\mathbf{w}_k^H} J_2 = E_{a_k'} \psi^{-1} \cdot E_{\mathbf{a}_k} \frac{1}{\pi(\sigma_r^2 + \sigma_w^2)^2} e^{-|z_k^e - a_k'|^2 / (\sigma_r^2 + \sigma_w^2)} \left\{ \left(1 - \frac{|z_k^e - a_k'|^2}{\sigma_r^2 + \sigma_w^2} \right) \mathbf{R}_{nn} \mathbf{w}_k + (z_k^e - a_k')^* \mathbf{S}_k \mathbf{a}_k \right\} \quad (5.1.4.1.5)$$

Igualando la expresión anterior del gradiente a cero llegamos finalmente a una expresión para el vector de coeficientes \mathbf{w}_k que minimiza el criterio,

$$\begin{aligned} \mathbf{w}_k &= \mathbf{R}_{nn}^{-1} \mathbf{S}_k \frac{E_{a_k'} \psi^{-1} \cdot E_{\mathbf{a}_k} e^{-|z_k^e - a_k'|^2 / (\sigma_r^2 + \sigma_w^2)} (z_k^e - a_k')^* \mathbf{a}_k}{E_{a_k'} \psi^{-1} \cdot E_{\mathbf{a}_k} e^{-|z_k^e - a_k'|^2 / (\sigma_r^2 + \sigma_w^2)} \left(1 - \frac{|z_k^e - a_k'|^2}{\sigma_r^2 + \sigma_w^2} \right)} \\ &= \mathbf{R}_{nn}^{-1} \mathbf{S}_k \alpha_k \end{aligned} \quad (5.1.4.1.6)$$

Obteniendo una expresión para el vector de coeficientes con la misma estructura a la que llegabamos para la primera función de coste.

5.1.5 Estimación del gradiente de J_2 .

En la primera función de coste se realizaba una estimación del gradiente en base a tomar el valor instantáneo de la no-linealidad correspondiente. En cambio, la evaluación del gradiente de J_2 requiere un mayor esfuerzo por el hecho de tratarse inherentemente de una linealidad con memoria, es decir, si reproducimos a continuación la ecuación 5.1.4.1,

$$\nabla_{\mathbf{w}_k^H} J_2 = \frac{1}{\sigma_r^2} E_{a_k'} \frac{E_{\mathbf{r}_k} e^{-|z_k - a_k'|^2 / \sigma_r^2} (z_k - a_k')^* \mathbf{r}_k}{E_{\mathbf{r}_k} e^{-|z_k - a_k'|^2 / \sigma_r^2}} = \frac{1}{\sigma_r^2} E_{a_k'} E_{\mathbf{r}_k} q_2(z_k, a_k') (z_k - a_k')^* \mathbf{r}_k \quad (5.1.5.1)$$

observamos como la esperanza con respecto a los datos es absolutamente necesaria para evaluar no tan sólo la función $q_2(z_k, a_k)$ sino también la esperanza del triple producto de

$q_2(z_k, a_k)$ con el error $z_k - a_k$ conjugado y el vector de datos \mathbf{r}_k . El procedimiento que utilizaremos será estimar por separado los términos del numerador y del denominador del cociente de la ecuación anterior. La división del término vectorial del numerador y del término escalar del denominador nos proporcionará una estimación del error generalizado. Realizaremos la estimación mediante un factor de olvido o alisado de una no-linealidad en los datos tal como sigue. Sean v_k y δ_k las estimaciones del numerador y del denominador, respectivamente. Entonces,

$$\left. \begin{aligned} v_{k+1}(a'_k) &= (1-\beta)v_k(a'_k) + \beta e^{-|z_k - a'_k|^2 / \sigma_i^2} (z_k - a'_k)^* \mathbf{r}_k \\ \delta_{k+1}(a'_k) &= (1-\beta)\delta_k(a'_k) + \beta e^{-|z_k - a'_k|^2 / \sigma_i^2} \end{aligned} \right\} \Rightarrow \hat{\nabla}_{\mathbf{w}_k^H} J_2 = \frac{1}{\sigma_i^2} E_{a'_k} \frac{v_k(a'_k)}{\delta_k(a'_k)} \quad (5.1.5.2)$$

El valor $1/\beta$ nos determinará la longitud de la ventana temporal equivalente utilizada para implementar la estimación de la esperanza estadística con respecto a los datos \mathbf{r}_k . Aunque existen sin duda mejores estimadores de los términos v_k y δ_k , estos presentan un coste computacional menor. Podría haberse procedido mediante una suma de L términos de la no-linealidad para cada iteración, no obstante, el coste computacional para obtener una estimación fiable es desorbitante. Téngase en cuenta que en el cálculo de estos términos se encuentra implícita la evaluación de densidades de probabilidad, por lo cual la longitud de la ventana temporal equivalente utilizada para la estimación dependerá del número de centroides de la constelación. Es decir, para una constelación de M símbolos, necesitaremos una ventana de longitud aproximadamente $100M$ para que cada uno de los símbolos se de por promedio 100 veces (supuestos éstos equiprobables). Esto nos fija el factor β a $1/100M$. Tómese el ejemplo de eculización de una QPSK con $L=400$, o de decorrelación de dos señales QPSK con $L=100 \cdot 4 \cdot 4 = 1600$. Evidentemente, un cálculo directo es inviable. En el caso de que los símbolos no sean equiprobables, postulamos un factor β igual a,

$$\beta = \frac{\min\{p(a_k), 1 \leq k \leq M\}}{100} \quad (5.1.5.3)$$

Es decir, forzamos que el símbolo menos probable se dé 100 veces en promedio. Un efecto lateral de la estimación propuesta del gradiente es que para una iteración no utilizamos el valor actual del vector de coeficientes para todos los términos de la ventana temporal equivalente, sino que se utilizan valores pasados debido al promediado exponencial. Para ser absolutamente rigurosos, deberían recalcularse todas las salidas z_k dentro de la ventana temporal mediante el vector de coeficientes actual, \mathbf{w}_k y re-evaluar el cociente. Si el paso de actualización de los coeficientes es suficientemente pequeño, podemos considerar el promediado exponencial anterior como una aproximación suficientemente buena. Este efecto

lateral es mayor durante el período de adquisición, en que el vector de coeficientes va variando. En convergencia, y suponiendo condiciones estacionarias, este efecto no es tan grave puesto que la solución del criterio es un vector de coeficientes invariante en el tiempo. Este algoritmo de estimación del gradiente se ha validado en las simulaciones incluso para valores relativamente elevados del paso de actualización de los coeficientes, obteniendo buenos resultados. Con lo cual, podemos afirmar que éste será un efecto de segundo orden.

La elección del parámetro varianza σ_t también reviste cierta importancia debido a su efecto sobre los mínimos de la función de coste. En efecto, supongamos que deseamos ecualizar una señal duobinaria con función de densidad de probabilidad $p(a_k) = \delta(a_k+1)/4 + \delta(a_k)/2 + \delta(a_k-1)/4$. Para una distribución discreta, el parámetro σ_t nos controla la anchura de la ventana utilizada para estimar las probabilidades de cada centroide. Conviene establecer un límite máximo de esta anchura, puesto que pueden darse casos en que la distribución real "engañe" a la función de coste.

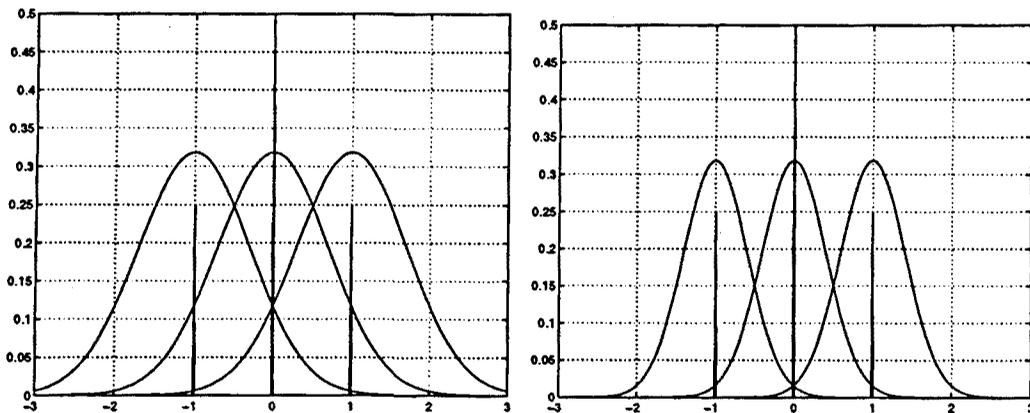


Figura 5.1.5.1: Representación de las ventanas utilizadas en la segunda función de coste para la estimación de las probabilidades de los centroides. Obsérvese en la de la izquierda (parámetro $\sigma_t^2=1$) como para un vector de coeficientes nulo, las probabilidades estimadas para cada centroide son aproximadamente igual a las del sistema duobinario, puesto que para una salida nula (función de densidad de masa unitaria en el origen), la delta correspondiente coincide aproximadamente con la mitad de la amplitud máxima de las Gaussianas centradas en +1 y -1. Parece así como si efectivamente se diesen dichos centroides. En cambio, aumentando la resolución de las ventanas (derecha, con parámetro $\sigma_t^2=1/3$), eliminamos este efecto al situar las deltas en las colas del resto de las gaussianas excepto la de interés. Para un funcionamiento correcto del algoritmo conviene utilizar ventanas lo más disjuntas posible.

En efecto supongamos un caso como el representado en la figura 5.1.5.1.a donde el vector de coeficientes del ecualizador se mueve alrededor de la solución trivial cero. Debido a la anchura de las ventanas, se estima una probabilidad no despreciable para los centroides +1 y -1, siendo la máxima probabilidad estimada la correspondiente al centroide cero, puesto que

una fracción no despreciable de las muestras cae en los límites de resolución de la ventana. Por consiguiente, la distribución que "ve" la función de coste es bastante parecida a la distribución objetivo de una señal duobinaria y aparece un mínimo local al que se engancha el ecualizador. Resulta conveniente entonces reducir el parámetro σ_t para evitar este efecto. En algunas pruebas realizadas hemos observado que para $\sigma_t=1$, el ecualizador se engancha a la solución $w_k=0$, mientras que si ponemos $\sigma_t^2=1/3$ conseguimos ecualizar la señal, eliminando el carácter de mínimo que tenía la solución $w_k=0$.

5.2 Función de Coste Extendida

En las funciones de coste que se han propuesto hasta el momento, el objetivo primordial ha sido forzar la distribución de una variable aleatoria escalar a la salida de un filtro adaptativo. En un caso más general conviene estudiar cómo se puede forzar la función de distribución conjunta de varias variables aleatorias. Existen ciertas aplicaciones donde la disponibilidad de funciones de coste basadas en este criterio pueden resultar de utilidad:

- Ecualización ciega de señales utilizando información sobre la codificación de los datos: En este caso, y suponiendo que no se haya realizado un 'scrambling' de los símbolos, la codificación se manifiesta como una dependencia estadística precisa entre símbolos cercanos. Resulta especialmente adecuado para datos codificados convolucionalmente o en esquemas de codificación correlativa (caso duobinario).
- Decorrelación de señales en sistemas de comunicación multiusuario : En este caso, y conociendo la estadística de las señales, nos interesa recuperar los símbolos transmitidos por cada usuario, eliminando la interferencia de múltiple acceso (Capítulo III) provocada por la presencia de usuario simultáneos en el canal. El criterio que se aplica es el de forzar la distribución conjunta de todos los usuarios a la salida de una matriz de decorrelación.

Podemos aplicar la filosofía de la distribución conjunta a cualquiera de las dos funciones de coste hasta ahora propuestas. En lo sucesivo consideraremos solamente la segunda, J_2 . Los resultados que se obtienen considerando la primera son totalmente análogos. Supondremos el modelo usual de datos y un filtro adaptativo para recuperar la señal distribución conjunta deseada. La obtención de la salida del filtro en instantes diferentes se calcula mediante la organización de los datos de entrada en una matriz de datos X_k definida como sigue,

(5.2.1)

$$\mathbf{z}_k^T = \mathbf{w}_k^H [\mathbf{r}_k, \dots, \mathbf{r}_{k-N+1}] = \mathbf{w}_k^H \mathbf{X}_k \Rightarrow \mathbf{z}_k = \mathbf{X}_k^T \mathbf{w}_k^*$$

Supongamos la segunda función de coste donde la evaluación se realiza sobre una función de distribución conjunta.

$$J_2 = -E_{\mathbf{a}_k} \ln E_{\mathbf{z}_k} \frac{1}{\pi^N \det \mathbf{R}_t} e^{-(\mathbf{z}_k - \mathbf{a}_k)^H \mathbf{R}_t^{-1} (\mathbf{z}_k - \mathbf{a}_k)} = -E_{\mathbf{a}_k} \ln E_{\mathbf{z}_k} \frac{1}{\pi^N \det \mathbf{R}_t} e^{-\|\mathbf{z}_k - \mathbf{a}_k\|_{\mathbf{R}_t}^2} \quad (5.2.2)$$

donde utilizamos una parametrización del ruido según la matriz de correlación \mathbf{R}_t . Bajo la hipótesis de ruido incorrelado la función de coste se reduce a,

$$\mathbf{R}_t = \sigma_t^2 \mathbf{I} \Rightarrow J_2 = -E_{\mathbf{a}_k} \ln E_{\mathbf{z}_k} \frac{1}{\pi^N \sigma_t^{2N}} e^{-\|\mathbf{z}_k - \mathbf{a}_k\|_2^2 / \sigma_t^2} = -E_{\mathbf{a}_k} \ln \hat{p}_{\mathbf{z}_k}(\mathbf{a}_k') \quad (5.2.3)$$

Las prestaciones dependerán de cuán cercana sea la parametrización de la matriz del ruido a la real. En cualquier caso, cuando la norma de dicha matriz tienda a cero obtenemos una estimación no polarizada de la función de densidad. La función Gaussiana en este contexto realiza una estimación de la densidad de probabilidad de cada uno de los centroides de la constelación. A fin de hallar una expresión para el error generalizado en el caso extendido, calculamos el gradiente de la función de coste respecto al hermítico del vector de coeficientes,

$$\nabla_{\mathbf{w}_k^H} J_2 = -\frac{1}{\sigma_t^2} E_{\mathbf{a}_k} \left(\sum_{i=1}^N \frac{\partial}{\partial z_k^i} \ln \hat{p}_{\mathbf{z}_k}(\mathbf{a}_k') \nabla_{\mathbf{w}_k^H} z_k^i + \sum_{i=1}^N \frac{\partial}{\partial z_k^{i*}} \ln \hat{p}_{\mathbf{z}_k}(\mathbf{a}_k') \nabla_{\mathbf{w}_k^H} z_k^{i*} \right) \quad (5.2.4)$$

Ahora bien, el segundo sumando es cero puesto que la salida conjugada no depende del hermítico del vector de coeficientes. Por consiguiente,

$$\nabla_{\mathbf{w}_k^H} J_2 = -\frac{1}{\sigma_t^2} E_{\mathbf{a}_k} \sum_{i=1}^N \frac{E_{\mathbf{X}_k} e^{-\|\mathbf{z}_k - \mathbf{a}_k'\|_2^2 / \sigma_t^2} (z_k^i - a_k'^i)^* \mathbf{r}_{k-i+1}}{\hat{p}_{\mathbf{z}_k}(\mathbf{a}_k')} = -\frac{1}{\sigma_t^2} E_{\mathbf{a}_k} \sum_{i=1}^N E_{\mathbf{X}_k} q_2(\mathbf{z}_k, \mathbf{a}_k') (z_k^i - a_k'^i)^* \mathbf{r}_{k-i+1} \quad (5.2.5)$$

Finalmente, intercambiando esperanza y sumatorio, llegamos a la siguiente expresión para el gradiente de la función de coste. Definiendo el vector error $\mathbf{e}'_k = \mathbf{z}_k - \mathbf{a}_k'$,

$$\nabla_{\mathbf{w}_k^H} J_2 = -\frac{1}{\sigma_t^2} E_{\mathbf{a}_k} E_{\mathbf{X}_k} q_2(\mathbf{z}_k, \mathbf{a}_k') \sum_{i=1}^N (z_k^i - a_k'^i)^* \mathbf{r}_{k-i+1} = -\frac{1}{\sigma_t^2} E_{\mathbf{a}_k} E_{\mathbf{X}_k} q_2(\mathbf{z}_k, \mathbf{a}_k') \mathbf{X}_k \mathbf{e}'_k \quad (5.2.6)$$

Intercambiando esperanzas, llegamos a la expresión del vector de error generalizado $\boldsymbol{\varepsilon}_{2,k}$,

$$(5.2.7)$$

$$\varepsilon_{2,k} \stackrel{def}{=} E_{\mathbf{a}'_k} q_2(\mathbf{z}_k, \mathbf{a}'_k) \mathbf{e}'_k \Rightarrow \nabla_{\mathbf{w}'_k} J_2 = -\frac{1}{\sigma_r^2} E_{\mathbf{x}_k} \mathbf{X}_k \varepsilon_{2,k}^*$$

Igualmente podemos postular una regeneración de referencia obtenida como,

$$\hat{\mathbf{a}}_k \stackrel{def}{=} E_{\mathbf{a}'_k} q_2(\mathbf{z}_k, \mathbf{a}'_k) \mathbf{a}'_k \tag{5.2.8}$$

Observamos pues que al introducir la información disponible sobre la codificación de los símbolos, llegamos simplemente a versiones vectoriales de las ecuaciones ya deducidas. El hecho de considerar la nueva función de coste no ha variado nada, podemos definir igualmente un error generalizado y una función de regeneración de referencia y expresar el gradiente de la función de coste como la correlación entre este error y la matriz de datos de entrada. Al considerar la función de distribución conjunta simplemente añadimos robustez al algoritmo. La codificación de los símbolos reduce el número de secuencias posibles que se transmiten, es decir, existe un conjunto de 'palabras código' o vectores de símbolos permitidos y un conjunto complementario de 'palabras error' o no permitidas. Si por efecto del ruido, la palabra más cercana es una palabra error, al utilizar una función de coste marginal (sobre una sola muestra), realizaríamos la actualización del vector de coeficientes de forma errónea. En cambio, utilizando la distribución conjunta, detectaríamos la poca fiabilidad de la muestra y la actualización se realizaría de una forma más conservadora. Se regeneraría una referencia en función de las palabras código más cercanas en norma Euclídea según la ecuación 5.2.3. De disponer de información sobre la correlación real del ruido, deberíamos sustituir la norma Euclídea por la distancia de Mahalanobis, o en la norma de la inversa de la matriz de correlación del ruido \mathbf{R}_t .

5.2.1 Aplicación de las funciones q y q_2 a detección

El hecho de utilizar las funciones q y q_2 en la regeneración de la referencia nos permite definir un algoritmo de detección (Bayesiano) mediante maximización de $q(\mathbf{z}_k, \mathbf{a}'_k)$ sobre la distribución \mathbf{a}'_k de,

$$\bar{\mathbf{a}}_k = \arg \max_{\mathbf{a}'_k} \{q_2(\mathbf{z}_k, \mathbf{a}'_k)\} = \arg \min_{\mathbf{a}'_k} \{-\ln q_2(\mathbf{z}_k, \mathbf{a}'_k)\} \tag{5.2.1.1}$$

Dada la secuencia \mathbf{z}_k , ésta nos induce un conjunto de, llamémosle métricas, a cada valor posible de la distribución deseada. Decidimos un símbolo como aquel que lleva asociada la

menor métrica, o menos logaritmo de la función q_2 . Podemos ver que éste representa un algoritmo de detección totalmente 'soft', puesto que no se realiza cuantificación de los datos de entrada. Si las señales han sido codificadas, el mismo algoritmo realizará directamente la decodificación. Es decir, si suponemos que una secuencia original de información, \mathbf{b}_k , se ha mapeado sobre la secuencia de canal \mathbf{a}_k , el proceso de decodificación consiste en el mapeado inverso,

$$\bar{\mathbf{a}}_k \rightarrow \bar{\mathbf{b}}_k \Leftrightarrow \bar{\mathbf{b}}_k = \arg \max_{\mathbf{b}_k} \{q_2(\mathbf{z}_k, \mathbf{a}'_k(\mathbf{b}_k))\} \quad (5.2.1.2)$$

Una forma menos robusta que en 5.2.8 de regenerar la referencia pero computacionalmente más eficiente consistiría en utilizar aquella secuencia de canal asociada al \mathbf{b}_k decodificado.

5.2.2 Codificación Correlativa. El Sistema Duobinario.

Un ejemplo clásico de correlación correlativa es el sistema duobinario, ampliamente utilizado en comunicaciones espaciales. En este sistema se introduce interferencia intersimbólica controlada en el modulador. Este método posibilita la reducción del ancho de banda necesario del canal a transmitir (intuitivamente resulta fácil de ver puesto que la señal ya no es tan "blanca"). Es decir, dada una secuencia de símbolos $a(n)$, se genera una secuencia codificada $c(n)$ mediante un filtrado lineal previo a la conformación de pulso de la siguiente forma,

$$c(n) = \sum_{k=0}^{N_{cc}-1} a(k)h_{cc}(n-k) = a(n)*h_{cc}(n) \rightarrow s(t) = \sum_{k=-\infty}^{+\infty} c(k)p(t-kT_s) \quad (5.2.2.1)$$

donde $h_{cc}(n)$ constituye la respuesta impulsional del codificador correlativo y $s(t)$ la señal en tiempo continuo emitida por el modulador. $p(t)$ y T_s constituyen el pulso conformador y el tiempo de símbolo, respectivamente.

Si consideramos el espectro de la secuencia $a(n)$, que suponemos blanco bajo la hipótesis de símbolos incorrelados, el espectro de la secuencia $c(n)$ aparece coloreado por el efecto de filtrado. Es decir, habremos reducido el nivel de potencia de la señal $s(t)$ en la banda alta de frecuencias, o equivalentemente, evitamos transiciones bruscas o variaciones de fase rápidas de la señal temporal $s(t)$, contribuyendo a una reducción del ancho de banda efectivo. La reducción se justifica intuitivamente por la densidad de transiciones por cero de la señal

codificada. Un caso particular viene ilustrado por el sistema duobinario donde se utiliza un filtro sumador $H_{cc}(z^{-1}) = (1+z^{-1})/2$ La constelación resultante aparece en la figura 5.2.2.1.

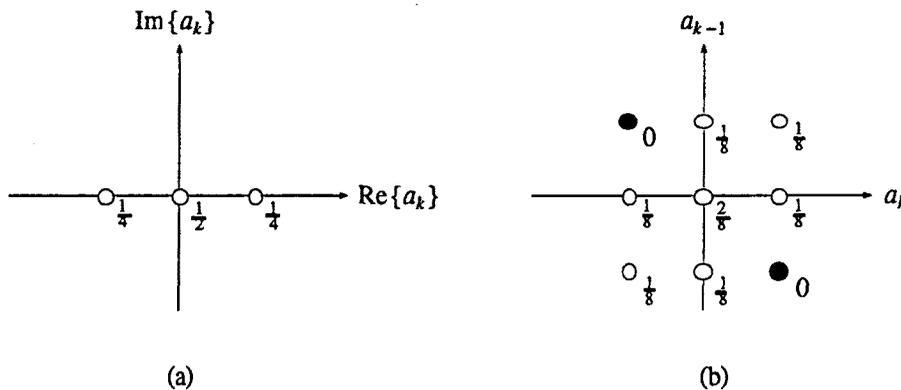


Figura 5.2.2.1: Representación de las constelaciones del sistema duobinario con las probabilidades de cada centroide. (a) constelación de un símbolo (b) constelación de dos símbolos, con un total de siete centroides. Podemos observar en esta constelación como aparecen dos centroides prohibidos debido a que no se permiten transiciones de símbolos de alta frecuencia (las parejas [1,-1] y [-1,1]), reduciendo así el ancho de banda.

5.2.3 Aplicación a la Decorrelación de Señales

En este tipo de aplicación forzamos también la distribución conjunta de una variable aleatoria vectorial. A diferencia de la aplicación en codificación, cada una de las señales resultantes, se obtiene a partir de un vector de coeficientes diferente. En el caso anterior, se obtenía mediante un vector de datos diferente (desplazado), o análogamente mediante vectores de coeficientes desplazados entre sí. Por tanto, aquí deberemos realizar el gradiente con respecto a una matriz de coeficientes. [Caso particular de separación de señales en fase cuadratura o por diversidad de polarización en antenas, equivalente a ecualización de una señal QPSK]. Veremos que las ecuaciones no varían sustancialmente respecto al caso de la codificación. En esta aplicación obtendremos un vector de errores generalizados, correspondiente a cada una de las componentes del vector de señal. Comparemos a continuación como se obtiene el vector de salidas en ambos casos, codificación (izquierda) y decorrelación o separación (derecha),

$$\mathbf{z}_k^T = \mathbf{w}_k^H \mathbf{X}_k = \mathbf{W}(\mathbf{w}_k)^H \mathbf{r}_k \Leftrightarrow \mathbf{z}_k = \mathbf{W}_k^H \mathbf{r}_k \tag{5.2.3.1}$$

donde las columnas de la matriz \mathbf{W}_k constituyen los filtros con que se recupera cada uno de los usuarios. Vemos pues, que el caso de la codificación puede plantearse como un caso

particular de la decorrelación/separación, donde la matriz \mathbf{W}_k aparece parametrizada linealmente mediante un vector de coeficientes \mathbf{w}_k . Naturalmente, es necesario conocer la estadística de las señales. En la función de coste podemos ver como aparecen productos cruzados entre las señales. Si se llega a la condición de independencia vemos que podemos linealizar la función de coste, es decir forzamos la anulación de los coeficientes (cumulantes) cruzados en todos los ordenes del desarrollo respecto el parámetro $-1/(\sigma_i)^2$, según la descomposición presentada

$$\begin{aligned}
 J_2 &= -E_{\mathbf{a}'_k} \ln E_{\mathbf{z}_k} \frac{1}{\pi^N \sigma_i^{2N}} e^{-|\mathbf{z}_k - \mathbf{a}'_k|^2 / \sigma_i^2} = -E_{\mathbf{a}'_k} \ln E_{\mathbf{z}_k} \prod_{i=1}^N \frac{1}{\pi \sigma_i} e^{-|z_k^i - a_i'^2|^2 / \sigma_i^2} \\
 &\rightarrow -E_{\mathbf{a}'_k} \ln \prod_{i=1}^N E_{z_k^i} \frac{1}{\pi \sigma_i} e^{-|z_k^i - a_i'^2|^2 / \sigma_i^2} = -\sum_{i=1}^N E_{\mathbf{a}'_k} \ln E_{z_k^i} \frac{1}{\pi \sigma_i} e^{-|z_k^i - a_i'^2|^2 / \sigma_i^2}
 \end{aligned}
 \tag{5.2.3.2}$$

Es decir,

$$J_2 = -E_{\mathbf{a}'_k} \ln \sum_{i_1, \dots, i_N} \frac{1}{i_1! \dots i_N!} m_{i_1, \dots, i_N}(\mathbf{a}'_k) \left(\frac{-1}{\sigma_i^2}\right)^{i_1 + \dots + i_N} = -\sum_{i_1, \dots, i_N} \frac{1}{i_1! \dots i_N!} \left(\frac{-1}{\sigma_i^2}\right)^{i_1 + \dots + i_N} E_{\mathbf{a}'_k} cum_{i_1, \dots, i_N} \|\mathbf{z}_k - \mathbf{a}'_k\|_2^2
 \tag{5.2.3.3}$$

En el caso de cumplirse la independencia tendremos que en función de la delta de Kronecker, los coeficientes de la expansión anterior cumplen la siguiente igualdad,

$$E_{\mathbf{z}_k} cum_{i_1, \dots, i_N} \|\mathbf{z}_k - \mathbf{a}'_k\|_2^2 = \delta(i_1, \dots, i_N) C_i(\mathbf{a}'_k)
 \tag{5.2.3.4}$$

Evidentemente, en un caso general no podremos asegurar que pueda llegarse a la independencia mutua entre todas las salidas, debido a que el ruido presenta correlación no nula entre ellas. De hecho, tal condición no queda garantizado en el caso de decorrelación de señales CDMA. Por consiguiente, aparecerán términos cruzados no nulos, aunque su contribución se verá minimizada.

Calculando el gradiente con respecto a la matriz de coeficientes hermítica llegamos a,

$$\varepsilon_{2,k} \stackrel{def}{=} E_{\mathbf{a}'_k} q_2(\mathbf{z}_k, \mathbf{a}'_k) \mathbf{e}'_k \Rightarrow \nabla_{\mathbf{W}_k^H} J_2 = -\frac{1}{\sigma_i^2} E_{\mathbf{r}_k} \mathbf{r}_k \varepsilon_{2,k}^H
 \tag{5.2.3.5}$$

Calculemos a continuación el gradiente de la función de coste,

$$\nabla_{\mathbf{W}_k^H} J_2 = -\frac{1}{\sigma_i^2} E_{\mathbf{a}'_k} \left(\sum_{i=1}^N \frac{\partial}{\partial z_k^i} \ln \hat{p}_{\mathbf{z}_k}(\mathbf{a}'_k) \nabla_{\mathbf{W}_k^H} z_k^i \right)
 \tag{5.2.3.6}$$

donde el término conjugado es nulo por definición de la arquitectura. Podemos expresar entonces el gradiente de la arquitectura según,

$$z_k^i = \pi_i^T \mathbf{W}_k^H \mathbf{r}_k = \text{tr}(\mathbf{W}_k^H \mathbf{r}_k \pi_i^T) \Rightarrow \nabla_{\mathbf{W}_k^H} z_k^i = \mathbf{r}_k \pi_i^T \quad (5.2.3.7)$$

Sustituyendo en la ecuación 5.2.3.6, llegamos a,

$$\nabla_{\mathbf{W}_k^H} J_2 = -\frac{1}{\sigma_r^2} E_{\mathbf{a}_k'} E_{\mathbf{r}_k} q_2(\mathbf{z}_k, \mathbf{a}_k') \sum_{i=1}^N (z_k^i - a_k^{i'})^* \mathbf{r}_k \pi_i^T = -\frac{1}{\sigma_r^2} E_{\mathbf{a}_k'} E_{\mathbf{r}_k} q_2(\mathbf{z}_k, \mathbf{a}_k') \mathbf{r}_k \mathbf{e}_k'^H \quad (5.2.3.8)$$

donde hemos definido el siguiente error,

$$\mathbf{e}_k' \stackrel{\text{def}}{=} \sum_{i=1}^N (z_k^i - a_k^{i'})^* \pi_i \quad (5.2.3.9)$$

Definiendo el error generalizado asociado, llegamos a la expresión deseada del gradiente,

$$\boldsymbol{\varepsilon}_{2,k} \stackrel{\text{def}}{=} E_{\mathbf{a}_k'} q_2(\mathbf{z}_k, \mathbf{a}_k') \mathbf{e}_k' \Rightarrow \nabla_{\mathbf{W}_k^H} J_2 = -\frac{1}{\sigma_r^2} E_{\mathbf{r}_k} \mathbf{r}_k \boldsymbol{\varepsilon}_k^H \quad (5.2.3.10)$$

Vemos entonces que en la situación de equilibrio, el vector de errores generalizados debe estar completamente incorrelado con los datos.

5.2.4 Aceleración de la Convergencia

Hemos observado en las simulaciones que la velocidad de convergencia de la segunda función de coste depende en gran medida de la potencia relativa de las señales interferentes con respecto a la deseada. La utilización de pasos de actualización elevados para acelerar la convergencia provoca muchas veces la captura del vector de coeficientes a soluciones no deseadas, parecidas no obstante a la distribución perseguida. El margen de rechazo de las señales interferentes se ve así reducido. Se recomienda pues utilizar pasos de actualización más pequeños, puesto que así garantizamos una implementación temporal más exacta de los operadores esperanza (recuérdese que por razones de implementación se ha cambiado ligeramente la formulación de la función de coste). El margen de rechazo de las interferentes se ve así mejorado.

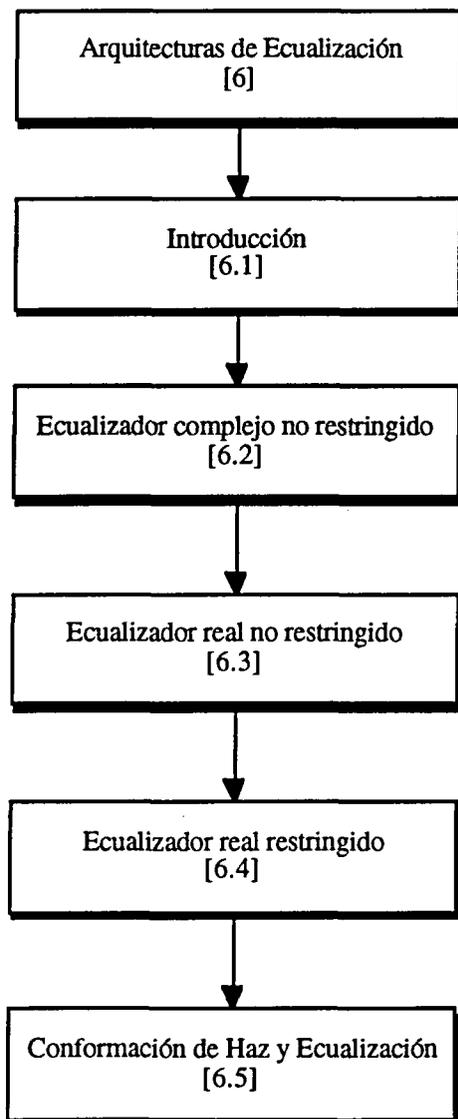
Una técnica harto conocida de aceleración consiste en premultiplicar el vector de datos por una matriz de filtrado dependiente de los datos. Podemos visualizar cada una de las señales concurrentes en el vector de datos, como modos o resonancias, de tal forma que los modos

menos potentes se hallan asociados a velocidades de convergencia más lentas. Si se construye la matriz de autocorrelación del vector de datos, su espectro de autovalores y por consiguiente su dispersión determinará la velocidad de convergencia del algoritmo. El procedimiento utilizado normalmente consiste en ecualizar el vector de datos mediante la raíz cuadrada de la inversa de la matriz de autocorrelación. De este modo, normalizamos la potencia de todas las señales presentes en el registro de datos. Este filtrado previo es equivalente a variar la topología del mínimo de la función de coste. Una gran dispersión de autovalores se asocia a curvas de nivel de elevada excentricidad (aunque no podemos en principio asegurar que sean hiperelipsoides debido al acusado carácter no-lineal de la función de coste), mientras que una dispersión nula se manifiesta como curvas de nivel hiperesféricas. En este segundo caso, el gradiente apunta siempre en la dirección del mínimo de la función de coste y la convergencia mejora apreciablemente.

El procedimiento mediante el cual se calcula la matriz de pre-filtrado o de decorrelación de los datos, se basa en la descomposición de la matriz de autocorrelación en factores de Cholesky. El cálculo asociado a las técnicas de aceleración de convergencia requiere normalmente un volumen de operaciones mayor que en algoritmos del tipo LMS. Un factor adicional a tener en cuenta en la evaluación de un algoritmo no consiste tan solo en la calidad de la señal conseguida sino también en el número total de operaciones necesario para obtener dicha calidad.

Capítulo VI

Arquitecturas de Ecuación



6.1 Introducción

En este capítulo nos ocuparemos de la definición de diversas arquitecturas de ecualización y de la deducción de los algoritmos de actualización correspondientes. Para ello utilizaremos básicamente la primera función de coste. El algoritmo de actualización correspondiente a una de las arquitecturas aquí presentadas (ecualizador complejo no restringido) ya se ha deducido en la sección 5.1.4 para la segunda función de coste. Seguiremos idéntico procedimiento para el resto de arquitecturas.

En este apartado derivaremos las ecuaciones para la actualización de los coeficientes de un ecualizador no restringido. Supondremos por tanto un modelo de señal banda base complejo y un modelo de canal también complejo. Suponemos que una secuencia de símbolos se modula sobre un pulso de Nyquist $p(t)$ y se transmite a través de un canal de respuesta impulsional $h_c(t)$. La señal recibida se expresa en tiempo continuo según la siguiente igualdad,

$$r(t) = \sum_{k=-\infty}^{k=+\infty} a_k (p * h_c)(t - kT) + n_r(t) \quad (6.1.1)$$

donde $n_r(t)$ representa un proceso de ruido Gaussiano aditivo. La sincronización con la secuencia de símbolos así como el error en la fase de la portadora se hallan implícitos en la respuesta impulsional del canal, ya que pueden modelarse linealmente. La arquitectura de la demodulación se define en términos de un filtro lineal $h_a(t)$ que produce salidas diezmadas a la velocidad de símbolo, tal que los estimandos de la secuencia se expresan según la siguiente ecuación,

$$\hat{a}_k = (p * h_a)(t - kT) \quad (6.1.2)$$

Nótese que una operación lineal sobre los datos de llegada es suficiente para resolver efectos estacionarios como el error de sincronismo o la rotación en la fase de la portadora. Esta arquitectura también puede compensar errores de frecuencia sin necesidad de un NCO (Number Controlled Oscillator) siempre que la incertidumbre del Doppler sea mucho más pequeña que el ancho de banda de la señal. El error de sincronización puede corregirse con un simple desplazamiento temporal del filtro $h_a(t)$. En una implementación digital suponemos que tomamos las muestras a la velocidad $1/T_s$, siendo la señal disponible discreta con $r_k = r(kT_s)$. El filtro h_a se toma como un filtro FIR de N_c coeficientes. Por lo

tanto, si agrupamos muestras sucesivas de la señal de llegada en un vector: $\mathbf{r}_k = [r(k), r(k-1), \dots, r(k-N_c+1)]^T$, la señal de llegada puede expresarse como la convolución de la secuencia de símbolos transmitidos a_k con una secuencia vectorial \mathbf{h}_r que contiene las muestras de la respuesta impulsional en el formato anteriormente definido para r ,

$$\mathbf{r}_k = a_k \mathbf{h}_r + \mathbf{n}_r(k) = \sum_{i=-\infty}^{i=+\infty} a_i \mathbf{h}_r(k-i) + \mathbf{n}_r(k) \quad (6.1.3)$$

Los símbolos estimados se obtienen entonces mediante producto escalar con el vector de coeficientes \mathbf{w} (de \mathbf{h}_a),

$$\hat{a}_k = \mathbf{w}^H \mathbf{r}_k = a_k \mathbf{w}^H \mathbf{h}_r(0) + \sum_{i \neq 0} a_i \mathbf{w}^H \mathbf{h}_r(k-i) + n_a(k) \quad (6.1.4)$$

El objetivo del filtro \mathbf{w} es minimizar la potencia de los términos interferentes, ruido Gaussiano y convolutivo, para minimizar la probabilidad de error de bit de a_k . Se desearía que \mathbf{w} fuese ortogonal a todos los vectores $\{\mathbf{h}_r(i), i \neq 0\}$, aunque esto raramente sucede puesto que \mathbf{h}_r tiene longitud infinita. Los coeficientes de \mathbf{w} se actualizan para minimizar la función de coste.

A continuación pasaremos a estudiar tres tipos de ecualizadores: el ecualizador complejo no restringido, el ecualizador real no restringido y el ecualizador real restringido. Las diferencias entre estos ecualizadores se encuentran en la arquitectura de sus coeficientes. El algoritmo adaptativo que apliquemos será el mismo en cualquier caso.

6.2 Ecualizador complejo no restringido

En este ecualizador permitimos que todos sus coeficientes sean complejos sin restricción. Aunque el coste computacional de filtrado es mayor que en las otras dos opciones, presenta alguna ventaja en ciertos escenarios. Un posible problema que puede presentar un demodulador digital basado en una estrategia de muestreo 'I-Q sampling' (es decir, los canales de fase y cuadratura se generan analógicamente, realizando a continuación una conversión analógica a digital en base a dos A/D) es la descompensación de canales debido a diferencias en los filtros analógicos. Tal efecto puede corregirse ecualizando cada canal con filtros diferentes. Podemos absorber también la corrección de fase de la portadora en el proceso de la ecualización, con lo cual un ecualizador complejo cumple ambas funciones. Esto es posible puesto que la función de densidad de probabilidad deseada a la salida del

filtro fuerza la posición de los centroides de la constelación en unos puntos determinados. Definimos la arquitectura del ecualizador w como,

$$y_k = \mathbf{w}^H \mathbf{r}_k \tag{6.2.1}$$

siendo w un vector de coeficientes complejos. Utilizamos la función de coste anteriormente descrita. Debido a consideraciones prácticas, prescindimos del operador esperanza y utilizamos la no-linealidad instantánea para una actualización de coeficientes en base al gradiente. Podemos desglosar entonces el gradiente de la función de coste como el producto de su derivada evaluada en la salida del filtro o error generalizado, por el gradiente de la arquitectura, que en este caso es igual al vector de datos \mathbf{r}_k ,

$$\nabla_{\mathbf{w}^H} -\ln p_{z_k}(y_k) = -\left(\frac{\partial}{\partial y_k} \ln p_{z_k}(y_k)\right) \nabla_{\mathbf{w}^H} y_k = -\varepsilon^*(y_k) \mathbf{r}_k \tag{6.2.2}$$

Puesto que suponemos una modulación lineal con una constelación discreta, el error generalizado se expresa como,

$$\varepsilon(y_k) = y_k - \sum_{i \in \mathcal{C}} a_i p_{a_i} q(y_k, a_i) \tag{6.2.3}$$

Pasemos a considerar la ecuación de actualización de la no-linealidad. Según el modelo, estamos considerando ruido Gaussiano aditivo. Por tanto, el único parámetro determinante de la no-linealidad es la varianza de dicho ruido. El gradiente con respecto a la varianza al cuadrado viene dado por,

$$\nabla_{\sigma_i^2} -\ln p_{z_k}(y_k) = -p_{z_k}^{-1}(y_k) \frac{\partial}{\partial \sigma_i^2} p_{z_k}(y_k) \quad , \quad p_{z_k}(y_k) = E_{a_i} \frac{1}{\pi \sigma_i^2} e^{-|z_k - a_i|^2 / \sigma_i^2} \tag{6.2.4}$$

Expresándolo en términos de las funciones $q(\cdot)$ llegamos a la siguiente igualdad,

$$\nabla_{\sigma_i^2} = \frac{1}{\sigma_i^2} \left(1 - \frac{1}{\sigma_i^2} \sum_{i \in \mathcal{C}} |z_k - a_i|^2 p_{a_i} q(z_k, a_i) \right) = \frac{1}{\sigma_i^2} \left(1 - \frac{\hat{\sigma}^2}{\sigma_i^2} \right) \tag{6.2.5}$$

Siendo pues equivalente a un error normalizado entre la varianza estimada y el parámetro varianza. Vemos que en el óptimo se cumplirá la igualdad entre varianzas. Como conclusión importante cabe destacar que la ecuación de actualización de la varianza es independiente de la arquitectura del sistema adaptativo y depende simplemente de las funciones $q(\cdot)$. Al

adaptar la función de coste misma en el algoritmo adaptativo debemos tener cuidado puesto que puede cambiar la importancia relativa del paso de actualización que estemos utilizando. En este caso, la actualización del parámetro varianza conduce a un cambio de la curvatura de la función de coste en el mínimo. En el caso de relaciones señal a ruido altas, el mínimo es muy picudo y una actualización de la varianza puede conducir a posibles divergencias del algoritmo adaptativo. Resulta aconsejable normalizar el paso de actualización a la curvatura de la función de coste, ya sea premultiplicando por la inversa del Hessiano o por el método más sencillo de normalizarlo por la inversa de la varianza, $1/\sigma_i^2$.

Existen otros métodos alternativos al del gradiente para realizar la actualización del parámetro varianza. Consideremos la expresión deducida en la Sección 4.2.4 para el cálculo del parámetro varianza óptimo mediante igualación a cero del gradiente de la función de coste,

$$\sigma_i^2 = E_{a_i'} |z_k - a_k'|^2 q(z_k, a_k') \quad (6.2.6)$$

Realizamos entonces la actualización mediante un factor de olvido μ_σ de la siguiente forma,

$$\hat{\sigma}_{i,k+1}^2 = (1 - \mu_\sigma) \hat{\sigma}_{i,k}^2 + \mu_\sigma E_{a_i'} |z_k - a_k'|^2 q(z_k, a_k') \quad (6.2.7)$$

En el capítulo de simulaciones realizaremos algunas comparaciones con ambos métodos de actualización.

6.3 Ecuador real no restringido

En este tipo de ecualizador suponemos que no existen efectos de descompensación de canales fase-cuadratura y que por tanto los efectos del canal pueden corregirse con un ecualizador real. No obstante, la corrección de la fase de la portadora no puede llevarse a cabo a partir de un ecualizador real y deberemos introducir un fador de rotación. La arquitectura del vector de coeficientes viene entonces dada por,

$$\mathbf{w} = e^{-j\theta} \mathbf{w}_r, \quad \mathbf{w}_r \in \mathfrak{R}^N \quad (6.3.1)$$

donde \mathbf{w}_r es un vector de coeficientes reales. Procediendo como siempre, el gradiente de la función de coste se expresa como el error generalizado por el gradiente con respecto a la

arquitectura. En el caso actual, la arquitectura viene definida por el vector de parámetros fase de la portadora más coeficientes del ecualizador: $[\theta, \mathbf{w}_r^T]$. Debemos tener en cuenta al calcular el gradiente, que ahora tenemos que calcular la derivada de la función de coste con respecto a y_k y a su conjugado puesto que ahora el gradiente de las arquitecturas y_k e y_k^* son ambos no nulos. La aplicación de la regla de la cadena nos conduce a,

$$\nabla_{\mathbf{w}_r^T} - \ln p_{z_k}(y_k) = - \left(\frac{\partial}{\partial y_k} \ln p_{z_k}(y_k) \right) \nabla_{\mathbf{w}_r^T} y_k - \left(\frac{\partial}{\partial y_k^*} \ln p_{z_k}(y_k) \right) \nabla_{\mathbf{w}_r^T} y_k^* \quad (6.3.2)$$

Recurriendo a una propiedad de las funciones de densidad derivadas a partir de la suma de una variable aleatoria más ruido Gaussiano (aunque esta propiedad no es exclusiva de este tipo de distribuciones) tenemos que,

$$\frac{\partial}{\partial y_k^*} \ln p_{z_k}(y_k) = \left(\frac{\partial}{\partial y_k} \ln p_{z_k}(y_k) \right)^* \quad (6.3.3)$$

cumpléndose idéntica relación en el gradiente respecto a la arquitectura,

$$\nabla_{\mathbf{w}_r^T} y_k^* = \left(\nabla_{\mathbf{w}_r^T} y_k \right)^* \quad (6.3.4)$$

llegamos a una relación coherente para la actualización de los pesos del ecualizador,

$$\nabla_{\mathbf{w}_r^T} - \ln p_{z_k}(y_k) = -2 \operatorname{Re} \left\{ \left(\frac{\partial}{\partial y_k} \ln p_{z_k}(y_k) \right) \nabla_{\mathbf{w}_r^T} y_k \right\} \quad (6.3.5)$$

Idéntico proceso se sigue con relación a la fase de la portadora, con lo cual las ecuaciones de actualización se basan entonces en el siguiente gradiente,

$$\nabla = -2 \operatorname{Re} \left\{ \varepsilon^*(y_k) \begin{bmatrix} -j y_k \\ e^{-j\theta} \mathbf{r}_k \end{bmatrix} \right\} \quad (6.3.3)$$

lo cual es coherente puesto que las actualizaciones de la fase de la portadora y del ecualizador deben ser ambos números reales.

6.4 Ecuador real restringido.

En este tipo de ecualizador realizamos un mapeado de la estructura del ecualizador óptimo. Es decir, suponemos el ecualizador formado por la concatenación de dos: un ecualizador de muestras y un ecualizador o decorrelador de símbolos. De este modo podemos implementar ecualizadores de longitud igual a los anteriores pero reduciendo el número de parámetros a adaptar. Aunque los coeficientes del ecualizador de símbolos dependen de los del ecualizador de muestras, al forzar un inventariado finito de los datos al implementar filtros FIR, los supondremos independientes en el proceso de adaptación. El cálculo de los coeficientes del ecualizador de símbolos es complejo ya que es necesario calcularlos como el respuesta impulsional inversa de un filtro conocido. Suponiendo ecualización real, llegamos a la siguiente definición de la arquitectura, incluyendo la rotación de la fase de la portadora,

$$y_k = e^{-j\theta} \mathbf{w}_2^T \mathbf{X} \mathbf{w}_1, \quad \mathbf{X}^T = [\mathbf{r}_{k-N'}, \dots, \mathbf{r}_{k+N'}] \quad (6.4.1)$$

donde se ha definido la matriz de datos \mathbf{X} como compuesta de bloques de datos de entrada desplazados un símbolo entre sí. La matriz abarca un total de $N=2N'+1$ símbolos. El vector \mathbf{w}_1 corresponde al ecualizador de muestras mientras que \mathbf{w}_2 corresponde al ecualizador de símbolos. Nótese que la arquitectura pasa a ser no-lineal en los coeficientes de ecualización. Podemos llegar aún más lejos en la parametrización del ecualizador global imponiendo la condición de que \mathbf{w}_2 sea un vector centro-simétrico. Es fácil introducir dicha parametrización a partir de la redefinición de \mathbf{w}_2 en base a una matriz de restricciones Π ,

$$\mathbf{w}_2 = \Pi \mathbf{v}_2, \quad \Pi^T = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \quad (6.4.2)$$

que hemos representado esquemáticamente para un caso particular. La definición final de la arquitectura se convierte entonces en,

$$y_k = e^{-j\theta} \mathbf{v}_2^T \Pi^T \mathbf{X} \mathbf{w}_1 \quad (6.4.4)$$

En la misma línea que en la Sección 6.3 y definiendo el vector de parámetros como $[\theta, \mathbf{w}_1^T, \mathbf{v}_2^T]$ obtenemos la siguiente ecuación de actualización,

$$(6.4.5)$$

$$\nabla = -2 \operatorname{Re} \left\{ \varepsilon^*(y_k) \begin{bmatrix} -jy_k \\ e^{-j\theta} \mathbf{X}^T \Pi \mathbf{v}_2 \\ e^{-j\theta} \Pi^T \mathbf{X} \mathbf{w}_1 \end{bmatrix} \right\}$$

en el caso que decidiésemos prescindir de la matriz de restricción, simplemente es necesario hacerla igual a la matriz unidad \mathbf{I} .

6.5 Conformación de haz y ecuación

En esta aplicación se supone que realizamos conjuntamente la conformación de haz en un array seguida de ecuación. Implícitamente esto es equivalente a ecualizar cada uno de los sensores con el mismo filtro. Tal estructura de ecuación resulta óptima para un array de banda estrecha donde las interferencias sólo presentan correlación espacial pero no temporal. Supondremos que la arquitectura de ecuación es cualquiera de las anteriormente expuestas. Definamos primeramente la nueva notación. Suponemos que el snapshot de datos en el instante k viene dado por el vector \mathbf{r}_k , que el vector de conformación viene dado por el vector complejo (sin restricciones) \mathbf{w}_s y que la salida del conformador se denota por c_k , el ecualizador temporal se denota por \mathbf{w}_t y la salida final por y_k ,

$$c_k = \mathbf{w}_s^H \mathbf{r}_k, \quad y_k = \mathbf{w}_t^H c_k \quad \rightarrow \quad y_k = \mathbf{w}_s^H \mathbf{X} \mathbf{w}_t^*, \quad \mathbf{X}^T = [\mathbf{r}_{k-N'}, \dots, \mathbf{r}_{k+N'}] \tag{6.5.1}$$

donde la matriz de datos \mathbf{X} contiene los vectores de $2N'+1$ snapshots consecutivos. En la figura 6.5.1 representamos un esquema de la arquitectura.

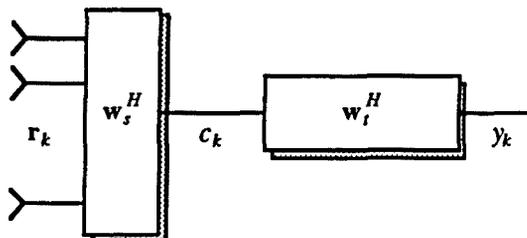


Figura 6.5.1: Estructura conjunta de conformación de haz y ecuación de banda estrecha. Todos los sensores se filtran por el mismo ecualizador.

A diferencia de lo que ocurría en ecuación de sensor único, no es necesario introducir un parámetro adicional para compensar la rotación de fase de la portadora puesto que tal corrección se engloba en los coeficientes complejos no restringidos del conformador de haz.

No se admiten restricciones en los coeficientes de \mathbf{w}_s a fin de tener absoluta libertad en la cancelación de interferencias. A continuación calcularemos las ecuaciones de actualización de los coeficientes para una arquitectura de ecuación general y particularizaremos para las tres estructuras de ecuación temporal presentadas en apartados anteriores.

El gradiente respecto a los coeficientes espaciales se calcula sin mayores problemas a partir de la ecuación 6.5.1,

$$\nabla_{\mathbf{w}_i^H} - \ln p_{z_t}(y_k) = - \left(\frac{\partial}{\partial y_k} \ln p_{z_t}(y_k) \right) \nabla_{\mathbf{w}_i^H} y_k = - \varepsilon^*(y_k) \mathbf{X} \mathbf{w}_i^* \quad (6.5.2)$$

Es decir, la función de coste impone en el óptimo incorrelación entre el error generalizado y el vector resultante de filtrar la señal de cada sensor por el ecuador temporal. El cálculo del gradiente en cuanto a los coeficientes del ecuador presenta ya mayor dificultad al tener en cuenta una estructura generalizada. Supongamos que los coeficientes temporales \mathbf{w}_t dependen de un conjunto alternativo de parámetros \mathbf{u}_t según la relación $\mathbf{w}_t = \mathbf{w}_t(\mathbf{u}_t)$. Supondremos que ambos vectores \mathbf{w}_t e \mathbf{u}_t son vectores de números reales puesto que el caso complejo restringido no presenta mayor dificultad. El gradiente debe pues calcularse respecto a este nuevo conjunto de parámetros,

$$\begin{aligned} \nabla_{\mathbf{u}_i^T} - \ln p_{z_t}(y_k) &= - \left(\frac{\partial}{\partial y_k} \ln p_{z_t}(y_k) \right) \nabla_{\mathbf{u}_i^T} y_k - \left(\frac{\partial}{\partial y_k^*} \ln p_{z_t}(y_k) \right) \nabla_{\mathbf{u}_i^T} y_k^* \\ &= -2 \operatorname{Re} \left\{ \varepsilon^*(y_k) \nabla_{\mathbf{u}_i^T} y_k \right\} \end{aligned} \quad (6.5.3)$$

Podemos escribir una ecuación de carácter general que implementa un ecuador cuadrático en los coeficientes a semejanza de la ecuación 6.4.1. Consideremos las muestras de la señal a la salida del conformador de haz, c_k , agrupadas en una matriz \mathbf{C} , donde las filas contiguas se encuentran desplazadas un símbolo entre sí. Supongamos también dos vectores de parámetros \mathbf{u}_1 y \mathbf{u}_2 que implementan en ecuador de muestras y un ecuador de símbolos, respectivamente. Podemos entonces representar la salida y_k de la siguiente forma,

$$y_k = \mathbf{u}_2^T \Pi^T \mathbf{C} \mathbf{u}_1 \quad (6.5.4)$$

Donde Π representa una posible matriz de restricciones en el caso que se desee imponer centro-simetría en el ecuador de símbolos (podríamos proceder análogamente para el vector \mathbf{u}_1 considerando una segunda matriz de restricción). Vemos fácilmente que el caso

general de un ecualizador de sólo muestras se consigue particularizando el vector \mathbf{u}_2 a un escalar, procediendo igualmente con el vector \mathbf{u}_1 para un ecualizador de sólo símbolos. El gradiente respecto a los coeficientes temporales queda finalmente expresado de forma general como,

$$\nabla_{\mathbf{u}_1^T} = -2 \operatorname{Re} \left\{ \varepsilon^*(y_k) \begin{bmatrix} \mathbf{C}^T \Pi \mathbf{u}_2 \\ \Pi^T \mathbf{C} \mathbf{u}_1 \end{bmatrix} \right\} \quad (6.5.5)$$

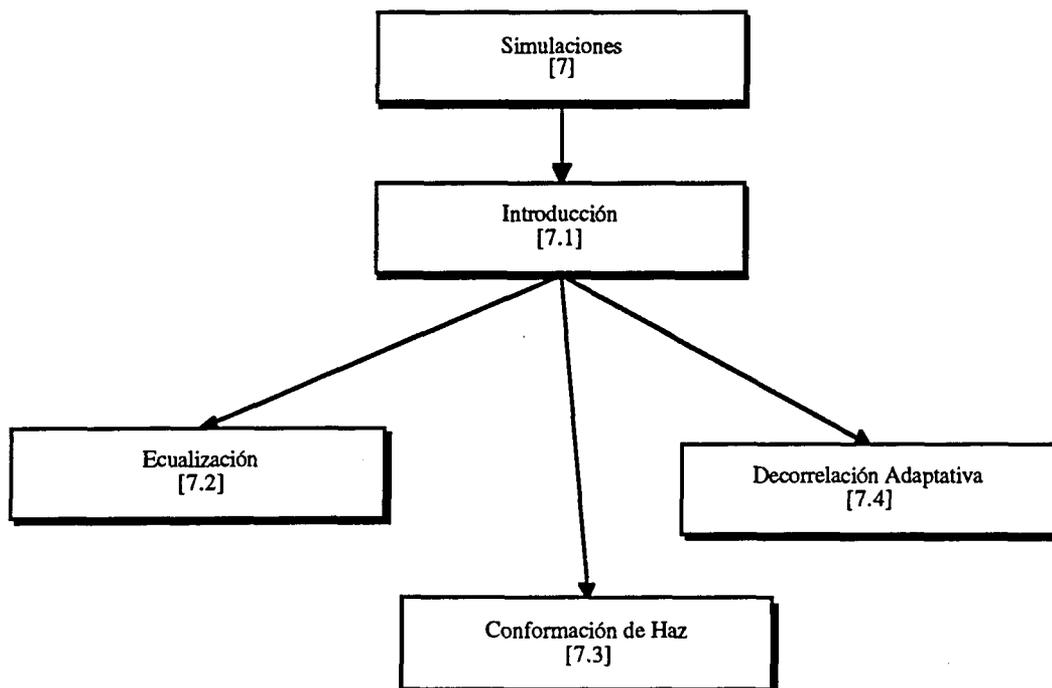
La interpretación de las ecuaciones de actualización puede realizarse con mayor claridad si consideramos el caso del ecualizador complejo no restringido. Las ecuaciones de actualización vienen representadas por,

$$\begin{aligned} \mathbf{w}_s(k+1) &= \mathbf{w}_s(k) + \mu_s \varepsilon^*(y_k) \mathbf{X} \mathbf{w}_r^* \\ \mathbf{w}_r(k+1) &= \mathbf{w}_r(k) + \mu_r \varepsilon^*(y_k) \mathbf{X}^T \mathbf{w}_s^* \end{aligned} \quad (6.5.6)$$

Vemos pues que para la actualización de los pesos del conformador utilizamos los snapshots filtrados por el ecualizador con lo cual aumentamos la calidad de los valores incrementales al reducir la potencia de ruido en la ecuación de actualización. Lo mismo sucede en la actualización de los pesos del ecualizador temporal, se utilizan los datos filtrados espacialmente, con lo cual el valor incremental presenta una mayor calidad. Esta interdependencia funciona bien en seguimiento, aunque en adquisición ocurren efectos de enmascaramiento. El ecualizador no puede empezar a converger a menos que el conformador no enfoque aproximadamente a la fuente deseada. No obstante se ha podido demostrar en las simulaciones que el algoritmo acaba convergiendo. Puesto que la función de coste es sensible a la estadística de las señales y menos a su potencia, el demodulador rechazará señales de estadística diferente. Como por ejemplo fuentes espacialmente localizadas de ruido Gausiano.

Capítulo VII

Simulaciones



7.1 Introducción

En los capítulos anteriores se han propuesto dos funciones de coste para sistemas adaptativos basadas en el concepto de Referencia Estadística. En este capítulo procederemos a analizar las prestaciones de ambas en un conjunto de diferentes escenarios a fin de validar los algoritmos propuestos. Consideraremos tres aplicaciones básicas en comunicaciones digitales,

- **Ecualización Ciega:** se analizará el comportamiento de J y J_2 para diversas condiciones de SNR y características del canal de transmisión.
- **Conformación de Haz:** se analizará el comportamiento de J y J_2 para ecualización y conformación de haz conjunta.
- **Decorrelación adaptativa:** se analizará el comportamiento de la función de coste extendida para J_2 en la decorrelación (o separación) de dos y tres usuarios que comporten el mismo canal.

El comportamiento de los algoritmos de ecualización ciega depende de las características del canal por el que se ha transmitido la señal recibida $r(t)$ y de la estadística de la señal de entrada al canal. Definiremos primero algunos canales típicos utilizados en las simulaciones, así como parámetros de interés a tener en cuenta en la elección de un canal determinado. Como se vió anteriormente en el capítulo sobre decorrelación de señales CDMA, la calidad del sistema se mide en base a cuán correlados se encuentren los usuarios, o bien símbolos cercanos en el caso dual de ecualización. La correlación entre símbolos no determina solamente el nivel de ruido que se obtendrá en la etapa de detección sino que también repercute en la velocidad de convergencia de los algoritmos adaptativos (dispersión de autovalores de la matriz de correlación en el caso del LMS). Definiremos el pulso de canal equivalente como aquel obtenido mediante convolución de la respuesta impulsional del canal con el pulso de conformación del modulador, $p_{eq}(n)$, así como su secuencia de autocorrelación, $r_{eq}(n)$ (el equivalente escalar de la secuencia matricial $\mathbf{R}(n)$ definida en la ecuación 2.6.3.8). Dicha secuencia, muestreada a múltiplos del tiempo de símbolo nos cuantifica la extensión de la interferencia intersimbólica y por ende nos estima la dificultad en ecualizar el canal asociado. Se considerarán únicamente modelos de canal de tipo FIR, aunque con una amplia casuística en cuanto a sus diagramas polos-zeros. Consideraremos los siguientes canales:

(7.1.1)

$$h_1(z^{-1}) = 1 - \frac{1}{2}z^{-4} + \frac{1}{4}z^{-8}$$

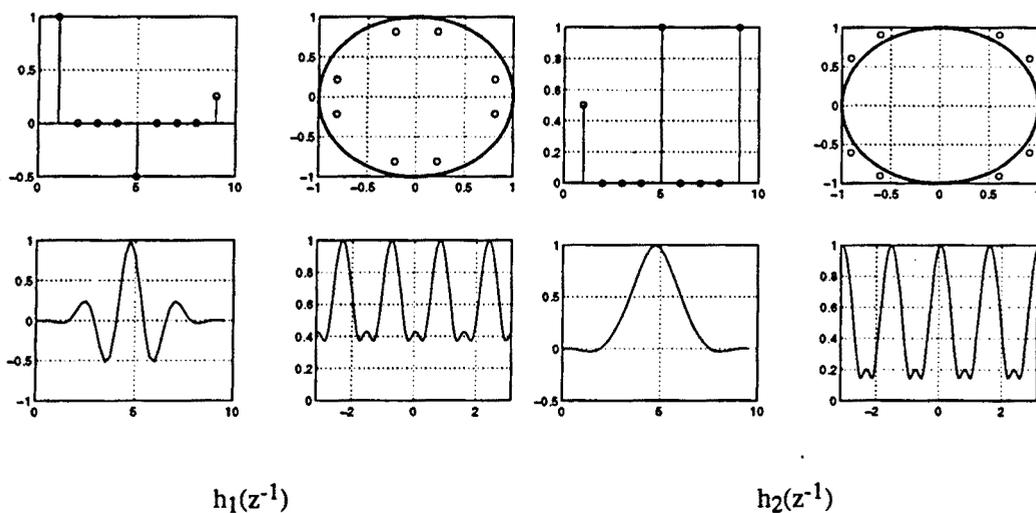
$$h_2(z^{-1}) = \frac{1}{2} + z^{-4} + z^{-8}$$

$$h_3(z^{-1}) = 1 - z^{-2} - z^{-3} + 3z^{-4} + 7z^{-5} + 2z^{-6} - 2z^{-7} + 4z^{-8} + 4z^{-9}$$

$$h_4(z^{-1}) = -0.1274 + 0.5542z^{-1} - 1.0973z^{-2} - 0.7313z^{-3} + 1.4047z^{-4} - 0.6202z^{-5} + 0.2371z^{-6} - 1.5868z^{-7} - 0.4015z^{-8} - 0.7707z^{-9}$$

los cuales representamos en la figura 7.1.1. Como puede observarse en las gráficas, h_1 y h_2 se han escogido como filtros de fase mínima (todos los ceros en el interior del círculo unidad) y de fase no-mínima (todos los ceros en el exterior del círculo unidad) respectivamente. h_3 y h_4 son ambos de fase mixta aunque realzan zonas distintas del espectro. La interferencia intérsimbólica abarca un margen de aproximadamente cinco símbolos en la totalidad de los casos.

En todas la simulaciones hemos supuesto que en adquisición no se disponía de información sobre el sincronismo de las señales, realizando ésta de forma totalmente ciega. Todos los experimentos se han basado en ecualizadores “fractionally-spaced” o fraccionarios. Es decir, no se realiza una combinación lineal de muestras a la salida del filtro adaptado, diezmadadas a una única muestra por símbolo, sino que se toman todas las muestras disponibles para estimar el símbolo recibido. La recuperación de sincronismo deviene así un problema de ecualización. La disponibilidad de información lateral sobre el sincronismo permitiría la implementación de ecualizadores más largos a un menor coste computacional como ya se ha expuesto en el Capítulo VI.



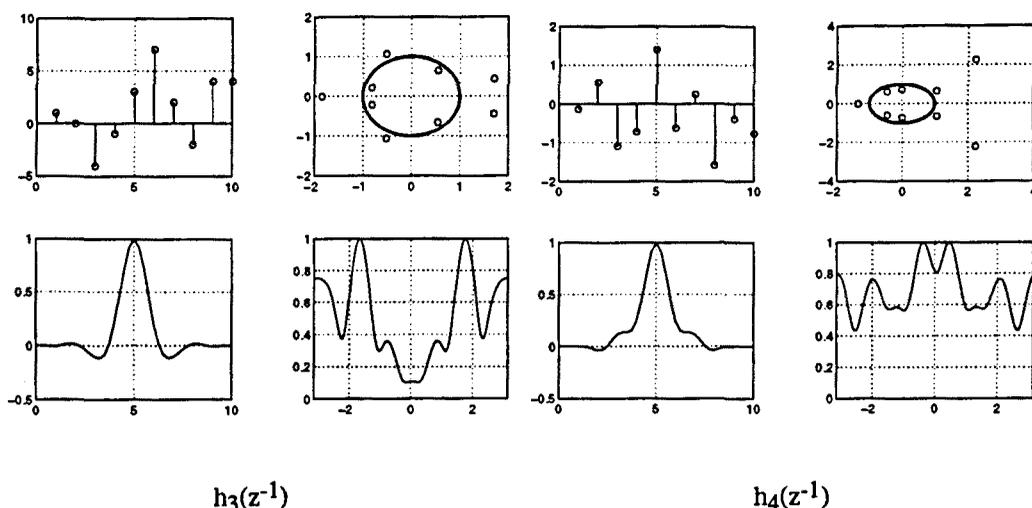


Figura 7.1.1: En esta figura representamos los cuatro canales utilizados en las simulaciones. Cada una de las representaciones está compuesta por cuatro gráficas: en la primera aparecen las muestras de la respuesta impulsional del canal, en la segunda representamos el diagrama de ceros (hemos supuesto modelos MA para los canales), en la tercera representamos la función de autocorrelación de la respuesta impulsional y finalmente en la cuarta la respuesta en frecuencia del canal.

En las simulaciones de decorrelación adaptativa supondremos un modelo de señal construido a partir de varios usuarios, cada uno de una estadística conocida, y modulados digitalmente sobre pulsos conformadores desconocidos. Es decir,

$$r(k) = \sum_{n=1}^{N_u} \sum_{m=-\infty}^{+\infty} a_n(m) p_n(k - mN_{ss}) + n(n) \quad , \quad z_k = W_k^H r_k$$

la señal recibida se corresponde con una mezcla convolutiva de las secuencias de usuario $a_n(k)$. El decorrelador se construye en base a una matriz rectangular W_k , la cual se adapta mediante la segunda función de coste según el algoritmo expuesto en la sección 5.2.3. La señal a la salida del decorrelador se obtiene mediante el producto de la hermítica de W_k con el vector de datos r_k . Suponemos la señal $r(k)$ perturbada mediante ruido Gaussiano aditivo $n(n)$. Imponemos la restricción de que todas las N_u señales presenten el mismo periodo de símbolo de N_{ss} muestras.

7.2 Ecuación

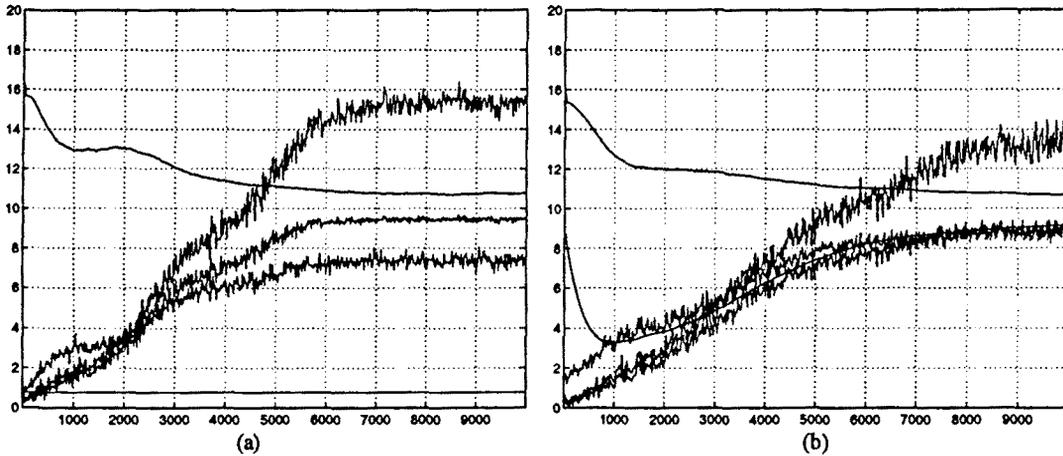


Figura 7.2.1-a: En esta figura representamos varias curvas características de la primera función de coste aplicada a la ecuación de una señal BPSK. En la curva A representamos la evolución de la relación señal útil a ruido Gaussiano a la salida del ecualizador en dB, mientras que la curva B representa la relación señal útil a interferencia intersimbólica. La curva C muestra la relación señal útil a potencia total, ruido más interferencia. La curva D representa la evolución de la relación señal a ruido más interferencia evaluada como el inverso de la varianza estimada según la ecuación $\sigma_t^2 = E_{a_k} z_k - a_k |z_k|^2 q(z_k, a_k)$. La curva E representa la evolución del parámetro α real. Debido a que no lo actualizamos en esta simulación, podemos ver como mantiene su valor constante. Obsérvese que la estimación del parámetro varianza es polarizada al no coincidir con la real. Se ha elegido un pulso square-root raised cosine como pulso conformador. 1) Número de símbolos: 10000. 2) Muestras por símbolo: 4. 3) pulso truncado a 3 símbolos. 4) factor de roll-off: 0.4. 5) SNR en el canal: 5 dB. 6) rotación de fase: $\pi/4$. 5) canal utilizado: $1-0.5z^{-4}+0.25z^{-8}$. 6) paso de actualización de los coeficientes, μ_w : 0.002. 7) paso de actualización de $\alpha=1/\sigma_t^2$, μ_α : 0. 8) Desalineamiento del vector de datos con el sincronismo real de la señal: 1 muestra.

Figura 7.1.1-b: según la simulación correspondiente a la figura anterior, representamos una comparación entre la SNR real y la SNR estimada al realizar una actualización del parámetro α , igual al inverso de σ_t^2 . Podemos ver como ahora existe convergencia de las tres curvas: C, D y E. Nótese que esta coincidencia no ocurre en adquisición sino en convergencia.

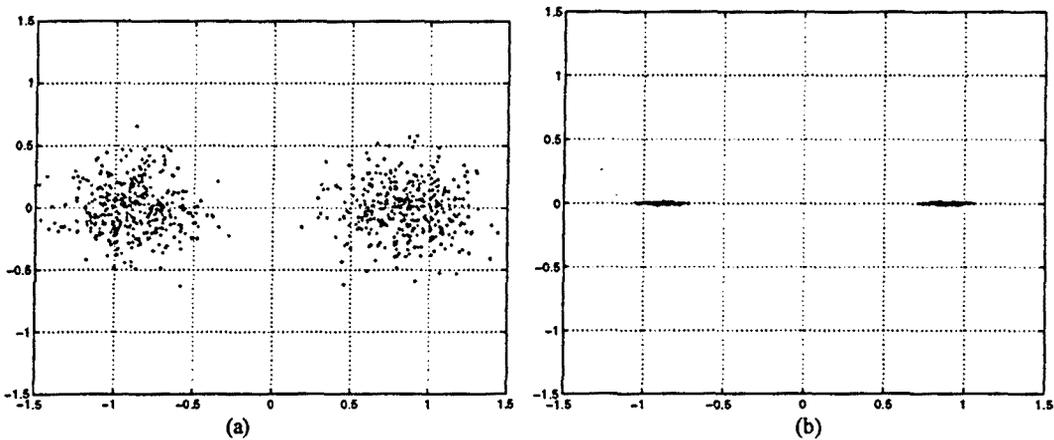


Figura 7.2.2: en esta figura representamos una realización particular en la que puede observarse la constelación obtenida en convergencia. A la izquierda podemos ver la constelación de salida, mientras que a la derecha hemos sustraído el ruido Gaussiano, apareciendo solo la interferencia intersimbólica residual. Vemos como se intenta forzar una disminución de la interferencia. En la segunda figura podemos observar claramente el efecto de contracción de la constelación. Puesto que forzamos una distribución BPSK, la componente imaginaria de la constelación es nula. Obsérvese como la varianza de la interferencia intersimbólica del canal de cuadratura (derecha) es mucho menor que la del canal en fase.

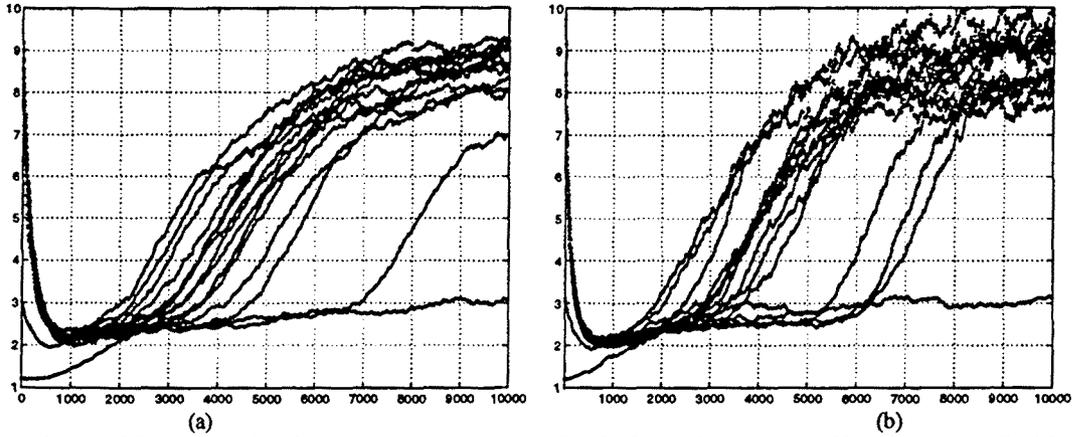


Figura 7.2.3: En estas dos figuras representamos la evolución del parámetro α para 15 realizaciones. En (a) se actualiza α según el gradiente de la función de coste, mientras que en (b) utilizamos el procedimiento descrito en la ecuación 6.2.7. Obsérvese que existe una dispersión de los tiempos de adquisición del algoritmo. Los pasos de normalización se han escogido como $\mu_w=0.002$ para los coeficientes y $\mu_\omega=0.002$ para la actualización de α .

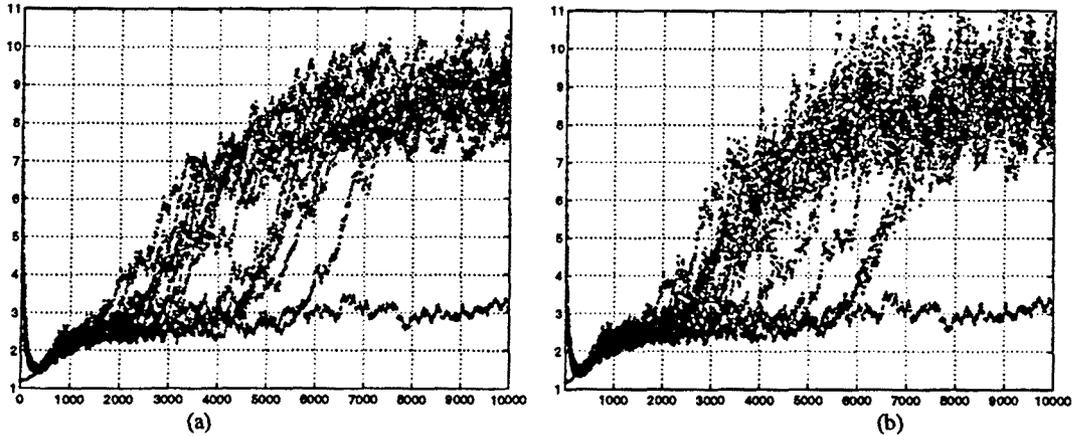


Figura 7.2.4: En estas dos figuras representamos la evolución del parámetro α para 15 realizaciones. En (a) se actualiza α según el gradiente de la función de coste, mientras que en (b) utilizamos el procedimiento descrito en la ecuación 6.2.7. Los pasos de normalización se han escogido como $\mu_w=0.002$ para los coeficientes y $\mu_\omega=0.01$ para la actualización de α . Los tiempos de adquisición disminuyen ligeramente, puesto que el factor dominante es el paso de actualización de los coeficientes.

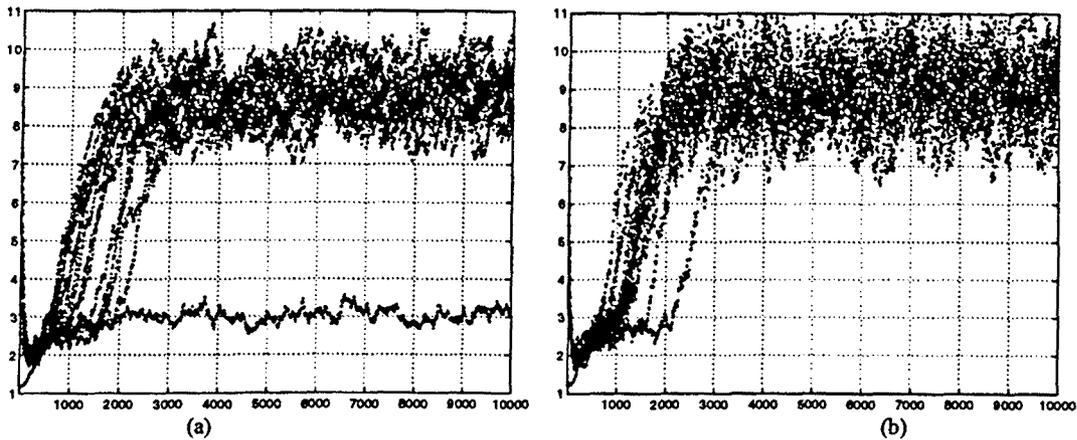


Figura 7.2.5: En estas dos figuras representamos la evolución del parámetro α para 15 realizaciones. En (a) se actualiza α según el gradiente de la función de coste, mientras que en (b) utilizamos el procedimiento descrito en la ecuación 6.2.7. Los pasos de normalización se han escogido como $\mu_w=0.01$ para los coeficientes y $\mu_\omega=0.01$ para la actualización de α . Los tiempos de adquisición disminuyen ostensiblemente al aumentar el paso de actualización. No obstante, la constelación se ve afectada, resultando aconsejable utilizar pasos de actualización diferentes en adquisición y seguimiento.

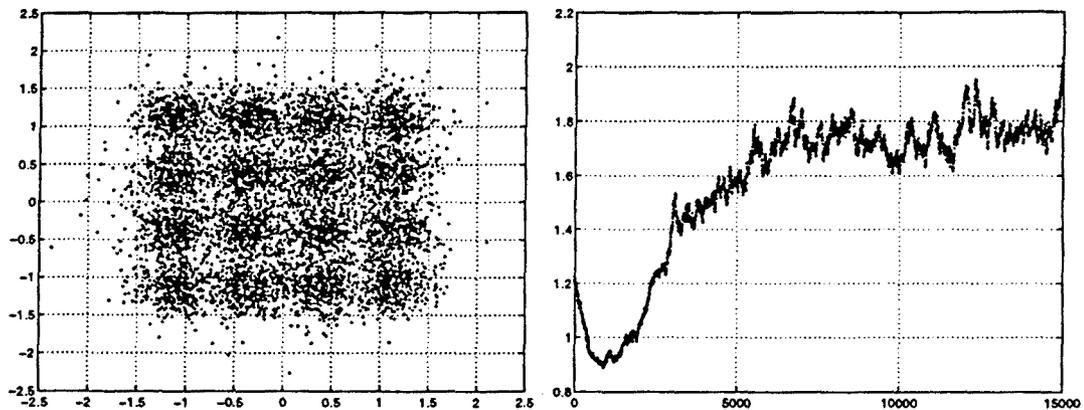


Figura 7.2.6: En estas dos gráficas ilustramos el problema de los mínimos locales de la función de coste. Dicho problema es especialmente grave para estadísticas de entrada con un número elevado de centroides, encontrándose relacionada la población de mínimos con la dimensión del alfabeto. En la primera figura observamos como la función de coste sólo ha conseguido proporcionar una versión escalada de la constelación 16-QAM deseada. Este efecto de produce debido a que para el valor de escala conseguido, los centroides de las esquinas de la constelación coinciden (erróneamente) con los centroides interiores de la pdf utilizada para la función de coste. Ya dependiendo del canal pueden existir determinadas combinaciones que den lugar a constelaciones de salida no presentes a la entrada, simplemente debido a que la ponderación con unos coeficientes determinados del ecualizador de los datos a la entrada “caen” en algunos centroides de la función de coste. Recordemos que en la sección 4.3 habíamos determinado que todos los mínimos locales se hallan en el subespacio de señal. En la segunda gráfica aparece la evolución temporal del parámetro α , aunque con poco sentido físico en este caso puesto que no forzamos exactamente la estadística deseada. Este experimento se ha realizado con los valores, 1) Número de símbolos: 15000. 2) relación SNR a la entrada: 22 dB. 3) Pulso square-root raised cosine muestreado a 4 veces la frecuencia de señalización y truncado a 3 símbolos. 4) Los coeficientes del canal corresponden a: $-0.5 + z^{-4} - 0.5z^{-8} + 0.25z^{-12}$. 5) $\mu_{\alpha} = \mu_w = 0.005$. 6) Número de coeficientes del ecualizador: 45.

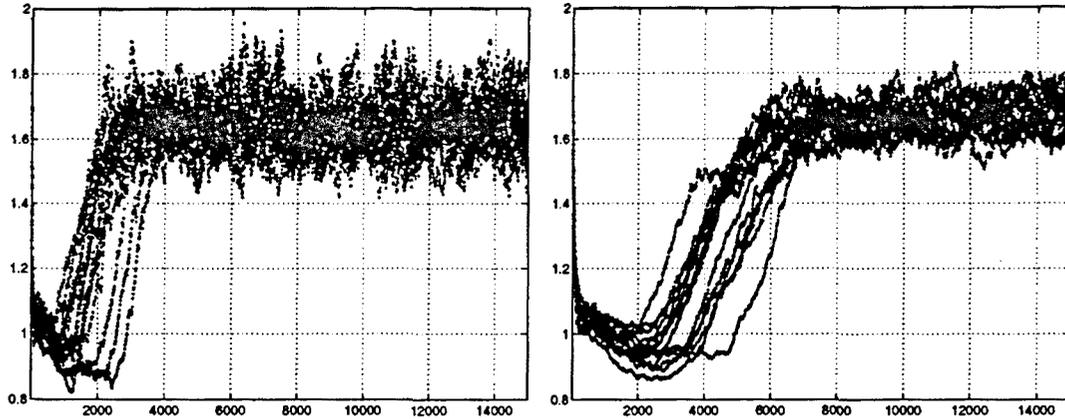


Figura 7.2.7: En estas dos figuras representamos hemos aplicado una función de coste basada en la estadística de una señal QPSK a la eculización de una señal 64-QAM. Para la primera hemos elegido $\mu_w = 0.005$, y $\mu_w=0.002$ para la segunda. El resto de los parámetros se mantiene invariable en ambos casos. Los problemas inherentes a la superpoblación de mínimos locales se resuelven así fácilmente. Observamos que los tiempos de adquisición se sitúan en unos márgenes razonables. Nótese que en este caso, el significado del parámetro a no es el del inverso de una potencia de ruido σ^2 debido a la no coincidencia de la estadística presente a la entrada y la forzada por la función de coste. No obstante, estas gráficas reflejan fielmente el tiempo de adquisición del algoritmo y su desajuste ya en la etapa de convergencia (compárese con llos gráficos correspondientes en la figura 2). Podemos apreciar la poca dispersión de los tiempos de adquisición para las 15 realizaciones representadas. Ambas simulaciones se han realizado para los siguientes valores:

- 1) Número de símbolos: 15000. 2) Muestras por símbolo: 4. 3) pulso Sqaure-root raised cosine truncado a 3 símbolos. 4) factor de roll-off: 0.4. 5) SNR en el canal: 28 dB. 6) rotación de fase: $\pi/4$. 5) Canal utilizado: $1-0.5z^{-4}+0.25z^{-8}$. 6) paso de actualización $\mu_\alpha=\mu_w$.

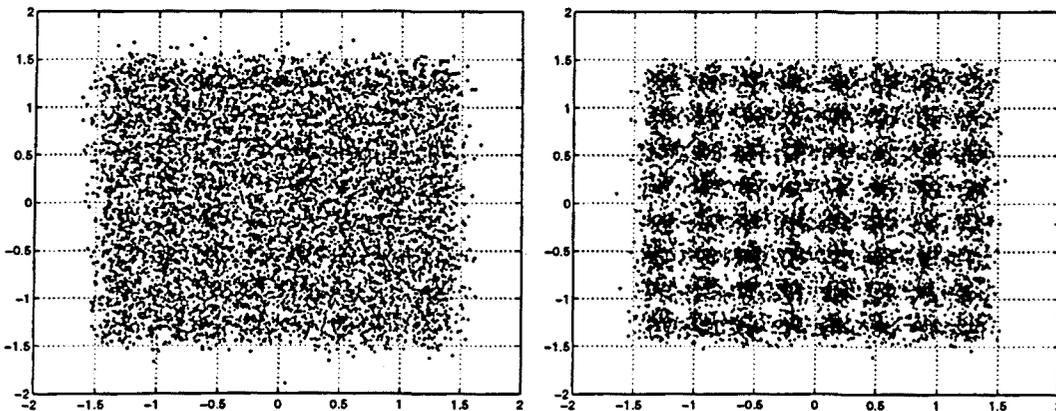


Figura 7.2.8: En estas dos figuras representamos las constelaciones obtenidas correspondientes a los dos casos de la figura 7.2.7. Aunque la primera aparece muy ruidosa, nótese en las figuras anteriores que su tiempo de adquisición es bastante rápido (paso de actualización grande). En el segundo caso obtenemos ya una constelación aceptable pero con un tiempo de adquisición mayor. Vemos pues que para constelaciones cuadradas de un gran número de centroides, la utilización de una función de coste QPSK presenta notables ventajas.

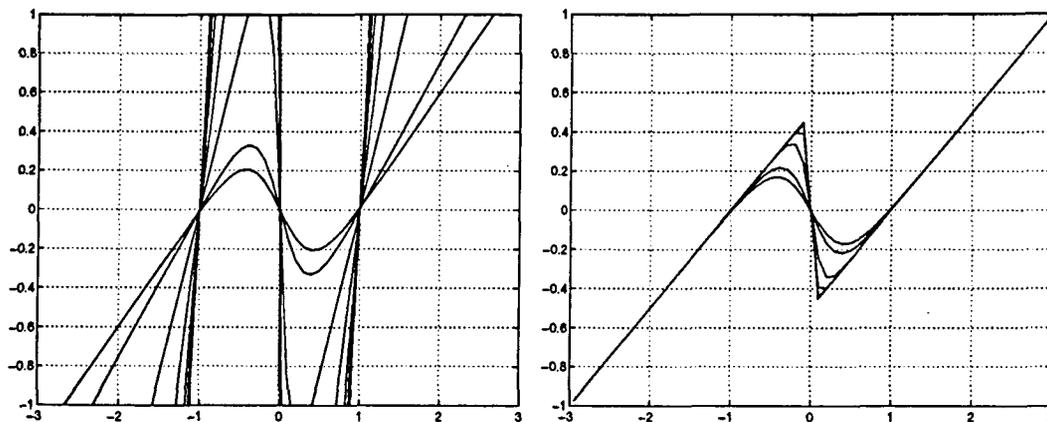


Figura 7.2.9: En esta figura hemos representado en primer lugar la evolución temporal de la curva de error generalizado, o derivada de la función de coste, para la ecualización de una señal BPSK. Podemos observar como a medida que pasamos de la etapa de adquisición a la de convergencia, el margen lineal de la curva de error va creciendo a medida que aumenta su pendiente. El aumento de pendiente es debido al hecho de que estamos cambiando la curvatura de la función de coste en línea, en función de la estadística de la señal de salida. A medida que convergemos, la relación señal a ruido más interferencia va en aumento, paralelamente a una mayor definición de las campanas Gaussianas que conforman la función de densidad de probabilidad. Su curvatura en los picos, o la pendiente del error generalizado en los cruces por cero de pendiente positiva por consiguiente aumenta. Este efecto de aumento de la pendiente es contraproducente puesto que equivale a un mayor desajuste de los coeficientes o a un paso de actualización equivalente mayor. Solucionamos este problema introducido por el algoritmo adaptativo mediante normalización del error generalizado por el parámetro α , tal como se representa en la siguiente figura. Vemos entonces que se mantiene la pendiente del error normalizado en los cruces por cero para todos los valores de α , en cambio, se reduce el margen de enganche, puesto que la varianza del ruido más las interferencias se reduce paralelamente. Podemos ver como la curva final converge a una curva fija determinada por la máxima SINR obtenible. Para entornos con relación señal a ruido alta, la curva de error generalizado va tendiendo a la misma que en el algoritmo DD (Decision-Directed), en el que implícitamente se supone probabilidad de error baja en una decisión hard del símbolo. Los valores utilizados para esta realización son:

- 1) Número de símbolos: 10000.
- 2) $\mu_{\alpha} = \mu_w = 0.006$.
- 3) SNR a la entrada: 15 dB.
- 4) canal utilizado de fase mixta: $-0.5 + z^{-4} - 0.5z^{-8} + 0.25z^{-12}$.
- 5) Número de coeficientes del ecualizador: 50.
- 6) Número de muestras por símbolo: 4.

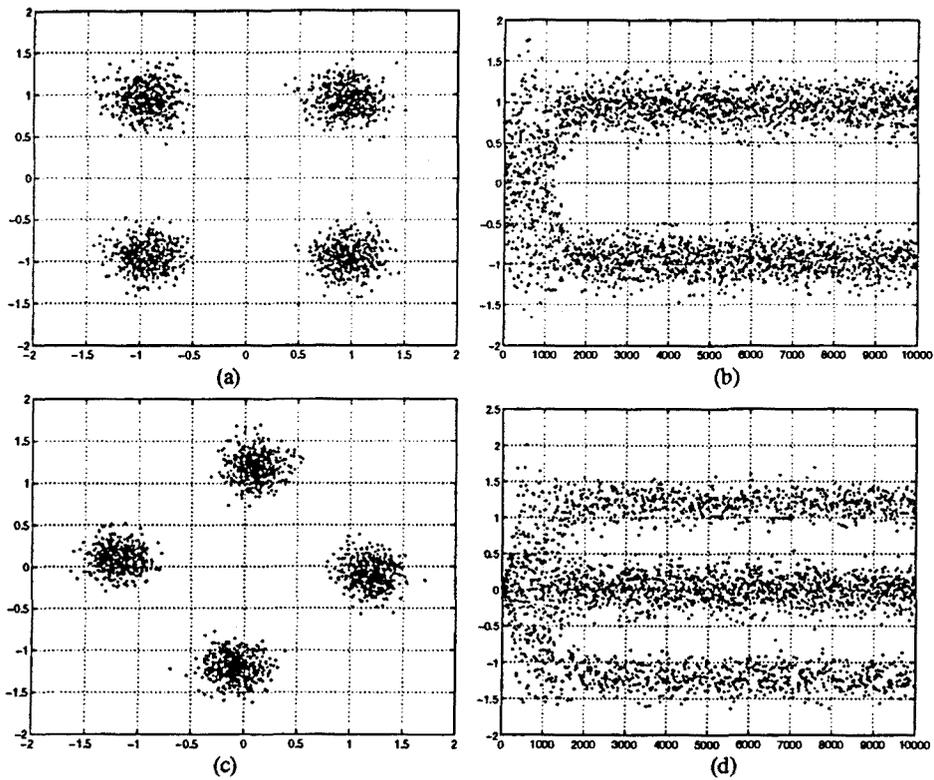


Figure 7.2.10: Ecuación de una señal QPSK forzando una distribución QPSK (gráficas a y b) y forzando una distribución de módulo constante (gráficas c y d).según el algoritmo presentado en la sección 3.6.1. Nótese que el tiempo de adquisición para el segundo algoritmo es ligeramente mayor. La ambigüedad de fase se debe a que el algoritmo sólo utiliza información sobre el módulo de la correlación. Los valores escogidos para la simulación son: 1) Número de símbolos: 10000. 2) Pulso square-root raised cosine truncado a 3 símbolos y factor de roll-off 0.4. 3) canal utilizado: h_4 .

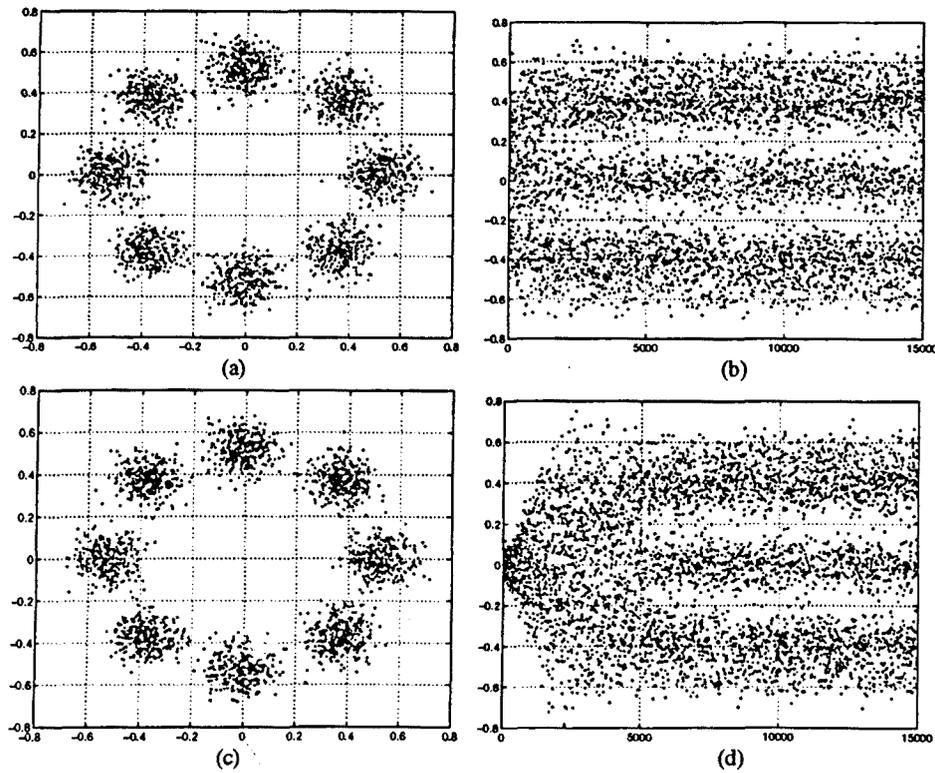


Figura 7.2.11: Experimento idéntico a la figura anterior utilizando una constelación 8-PSK.

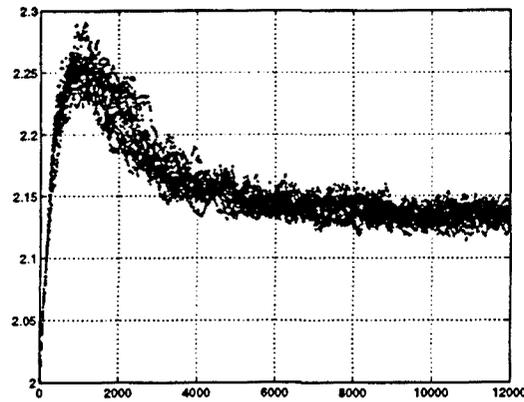


Figura 7.2.11: Evolución temporal de la segunda función de coste para la ecualización de una señal duobinaria, evaluada para 15 realizaciones. El crecimiento inicial observado en las gráficas es debido al periodo de tiempo que tarda en llenarse la memoria de la función no-lineal. Los valores escogidos para la simulación son: 1) Número de símbolos: 12000. 2) Muestras por símbolo: 4. 3) Pulso square-root-raised cosine truncado a 3 símbolos. 4) Factor de roll-off: 0.4. 5) relación SNR en el canal: 10 dB. 6) coeficientes del canal: $0.5+z^{-4}+z^{-8}$. 7) Pasos de actualización: $\mu_{\alpha}=\mu_{\omega}=0.003$ 8) $a=1.2$.

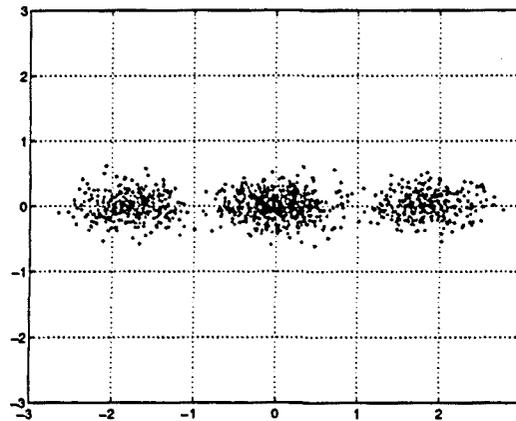


Figura 7.2.12: Constelación de la señal duobinaria en convergencia. A diferencia de la primera función de coste, esta vez se consigue que el vector de coeficientes no tienda a la solución trivial.

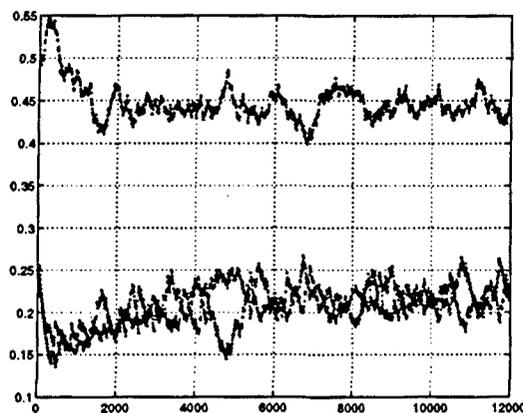


Figura 7.2.13: Evolución temporal de las probabilidades estimadas mediante la función de coste. La gráfica superior corresponde a la densidad de probabilidad del símbolo $a=0$, mientras que el resto corresponde a los dos símbolos polares $a=\{+2,-2\}$. Se han tomado como estimaciones iniciales de estos valores en la iteración cero los valores nominales de las probabilidades: $p(0)=1/2$, $p(+2)=1/4$ y $p(-2)=1/4$. Obsérvese que las probabilidades estimadas se encuentran la mayor parte de las veces por debajo de sus valores nominales, debido al paso de las muestras de salida por una función no-lineal (campana de Gauss), como se había justificado anteriormente.

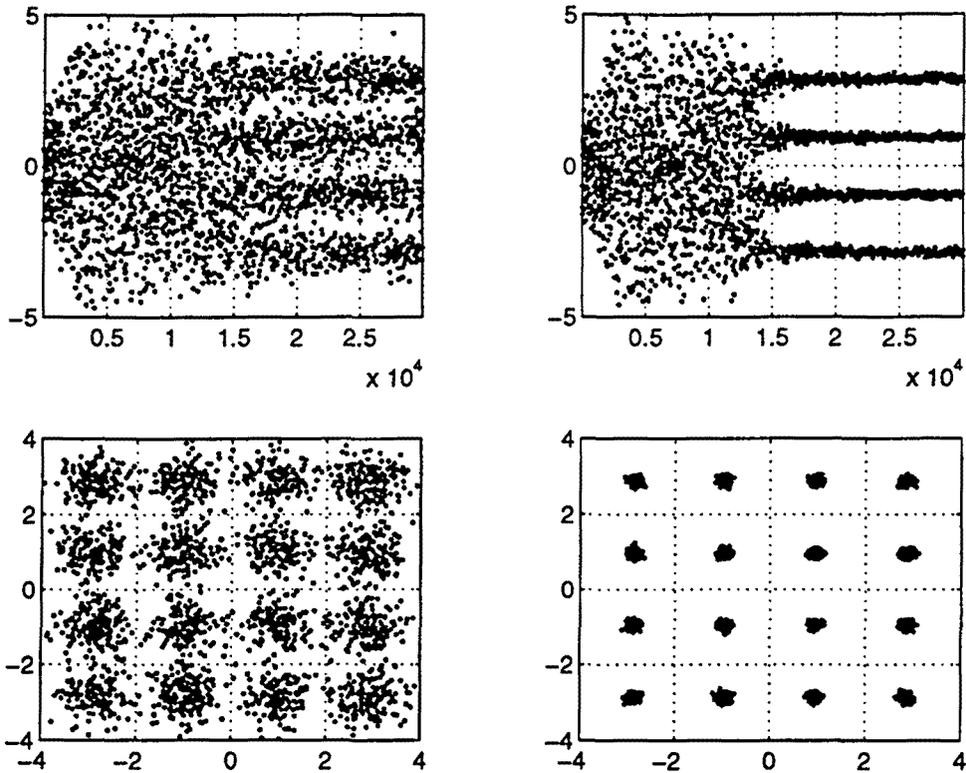


Figura 7.2.14: Representación de la evolución de la constelación en función del tiempo para la equalización de una señal 16-QAM utilizando la segunda función de coste. En la fila superior representamos la evolución del canal en fase para la misma realización con ruido (izquierda) y sustrayendo el ruido (derecha). En la gráfica superior derecha podemos ver como la función va abriendo el ojo de la constelación hasta eliminar casi completamente la interferencia intersimbólica, efecto no muy apreciable en presencia del ruido. En las gráficas inferiores representamos la constelación ya en convergencia. Aunque la relación SNR a la salida del equalizador es baja, podemos apreciar que la constelación aparece perfectamente centrada en la escala correcta, a diferencia de lo que ocurría en la simulación con la primera función de coste. Esta función de coste resulta ser más robusta en cuanto a errores de escalado, puesto que un mínimo local puede volverse inestable al transcurrir el tiempo si algunos de los centroides de la constelación no se dan con la densidad de probabilidad requerida. Esta simulación se ha realizado para los siguientes valores: 1) Número de símbolos: 30000. 2) Muestras por símbolo: 4. 3) Pulso square-root raised cosine truncado a 3 símbolos. 4) Factor de roll-off: 0.4. 5) Relación SNR en el canal: 10 dB. 6) Rotación de fase: $\pi/4$. 7) $\mu_w = 0.0001$. 8) El paso de actualización utilizado en la estimación de las probabilidades, μ_p , se ha escogido igual a 1/100 de la probabilidad de los centroides: 6.25×10^{-4} , supuestos éstos equiprobables. 9) Número de coeficientes del equalizador: 40. 10) Factor α de estimación de las probabilidades: 1.2. 11) Coeficientes del canal: [-0.1274, 0.5542, -1.0973, -0.7313, 1.4047, -0.6202, 0.2371, -1.5868, -0.4015, -0.7707].

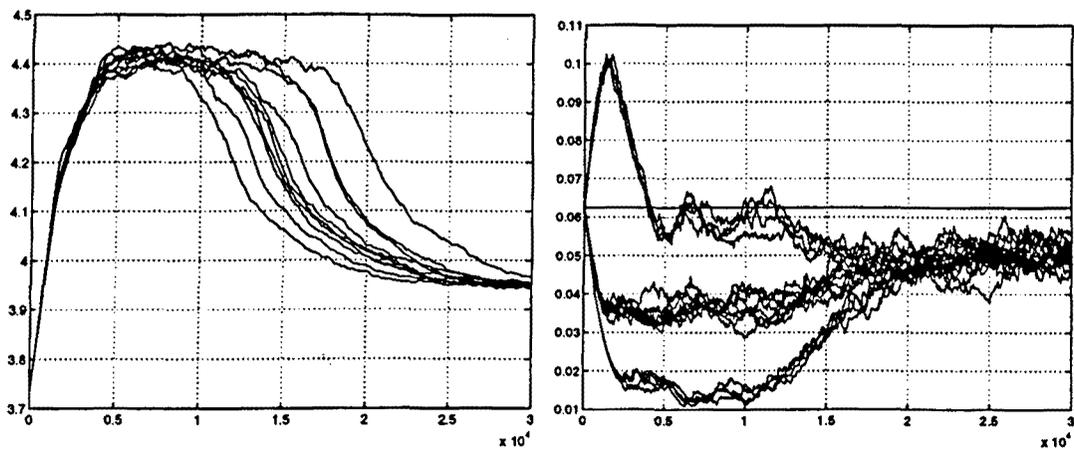


Figura 7.2.15: En la primera gráfica representamos la evolución temporal de la función de coste para 10 realizaciones distintas. Obsérvese que el tiempo de adquisición puede presentar una dispersión acusada. En la segunda gráfica representamos la evolución temporal de las probabilidades estimadas para una realización particular. Ya en convergencia, dichas probabilidades tienden a un valor inferior a 1/16, la probabilidades de un centroide cualquiera.

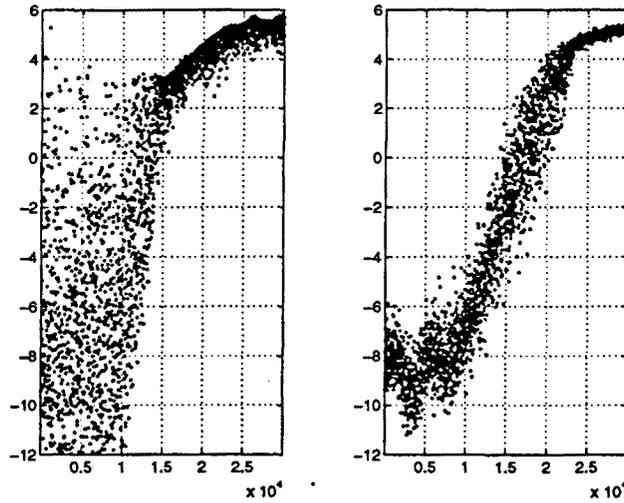


Figura 7.2.16: Evolución temporal de la relación señal a ruido más interferencia intersimbólica a la salida del ecualizador para una realización particular (izquierda) y un promedio sobre varias realizaciones (derecha). Obsérvese que en la segunda gráfica el codo de adquisición a seguimiento es más suave debido a la dispersión ya observada en los tiempos de adquisición. Podemos considerar el margen entre el límite inferior y superior de esta curva como el intervalo en el que se llega a la convergencia.

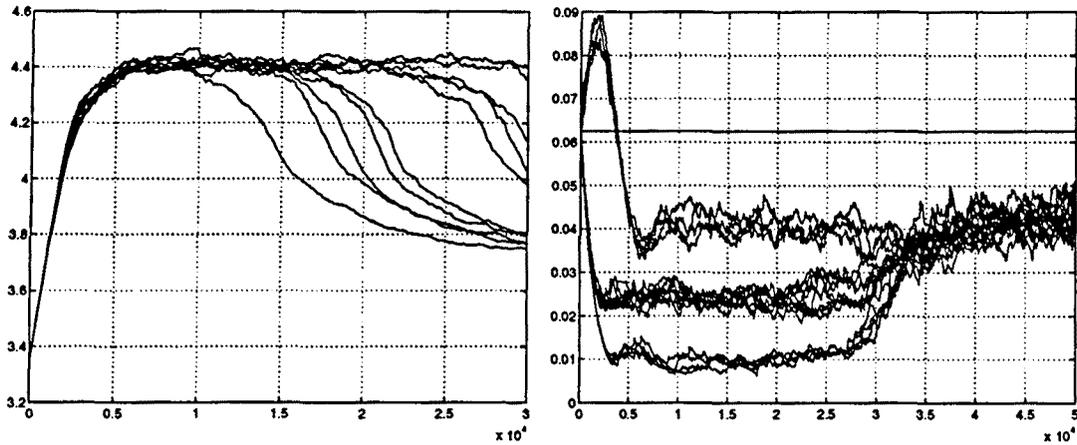


Figura 7.2.17: a) Evolución temporal de la función de coste para un factor $\alpha=2.2$ para un conjunto de 10 realizaciones. Podemos ver como el aumento de este parámetro repercute en un aumento del tiempo de adquisición. b) Evolución de las probabilidades estimadas para un registro de 50000 símbolos. Vemos como se llega a la convergencia en 30000 iteraciones. El aumento de α conduce a una disminución de las probabilidades estimadas con respecto al caso anterior, debido a que ahora menos muestras caen dentro de la ventana de resolución de la función ino-lineal.

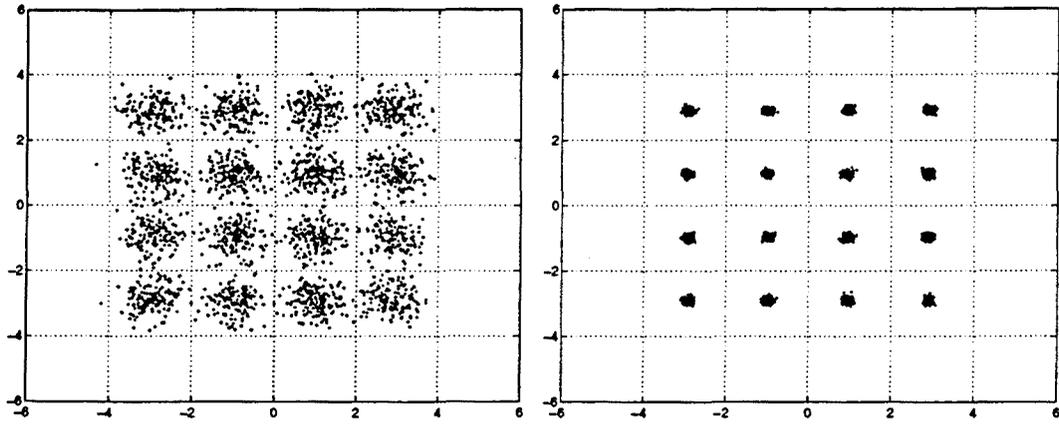


Figura 7.2.18: Representación de la constelación obtenida en convergencia para el caso anterior con ruido (izquierda) y sustrayendo el ruido (derecha).

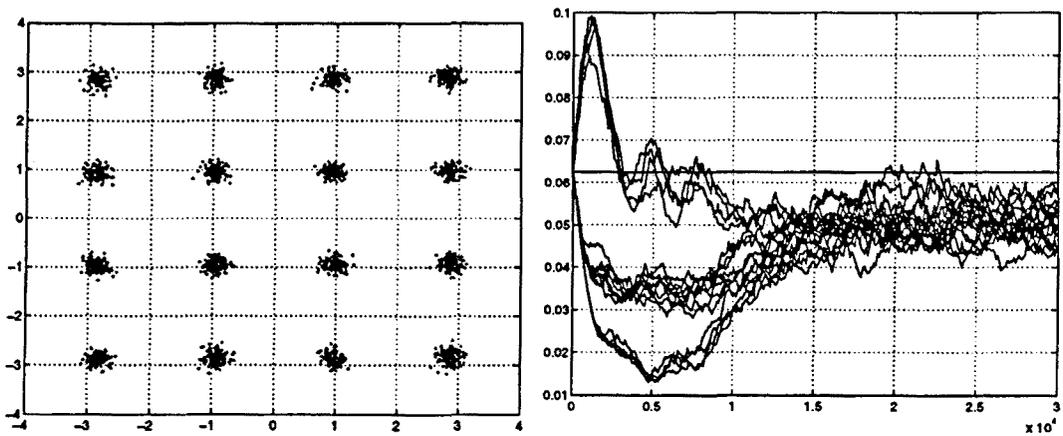


Figura 7.2.19: Manteniendo constantes los demás valores de las simulaciones anteriores, en este caso hemos utilizado un paso de actualización de los coeficientes mayor, $\mu_w = 0.008$. Representamos en la primera figura la constelación obtenida en convergencia a la que se ha sustraído el ruido. La evolución temporal de las probabilidades estimadas se representa en la segunda figura, obtenidas con un factor $\alpha=1.2$.

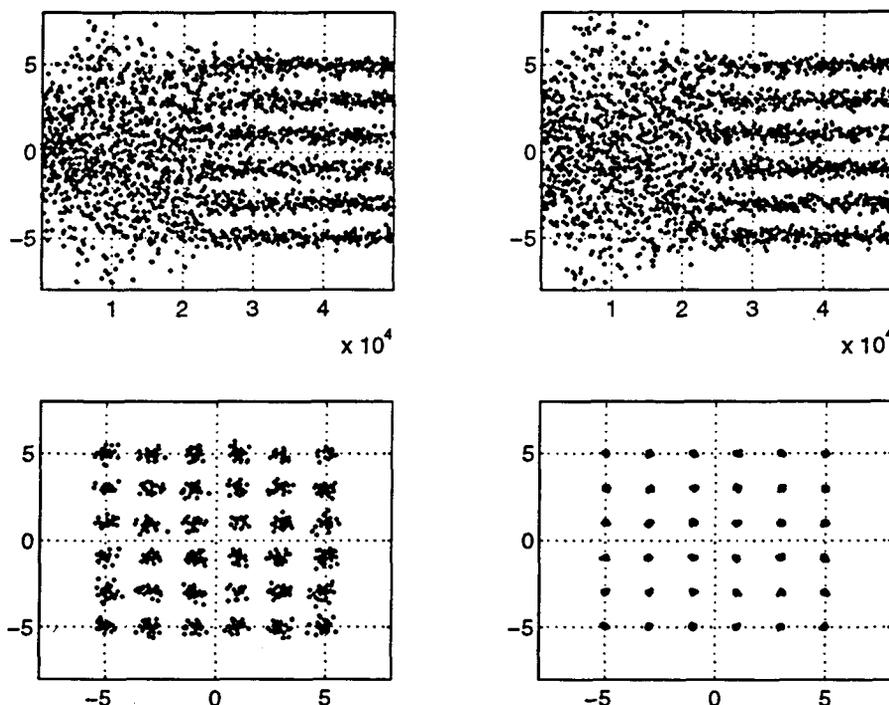


Figura 7.2.20: Ecualización de una señal 36-QAM utilizando la segunda función de coste. En la primera fila representamos la apertura del diagrama de ojo en función del tiempo para los canales de fase (izquierda) y cuadratura (derecha). En la segunda fila aparece la constelación obtenida en convergencia con ruido (izquierda) y sustrayendo el ruido (derecha). Los valores escogidos para la simulación son: 1) Número de símbolos: 50000. 2) Muestras por símbolo: 4. 3) Pulso square-root raised cosine truncado a tres símbolos. 4) Factor de roll-off: 0.4. 5) rotación de fase: $\pi/4$. 6) factor $\alpha=1.25$. 7) Paso de actualización de los coeficientes, $\mu_w=0.005$. 8) Paso de actualización de las probabilidades estimadas: $3e-4$. 9) Número de coeficientes del ecualizador: 30. 10) Coeficientes del canal: $[1,0,-4,-1,3,7,2,-2,4,4]$.

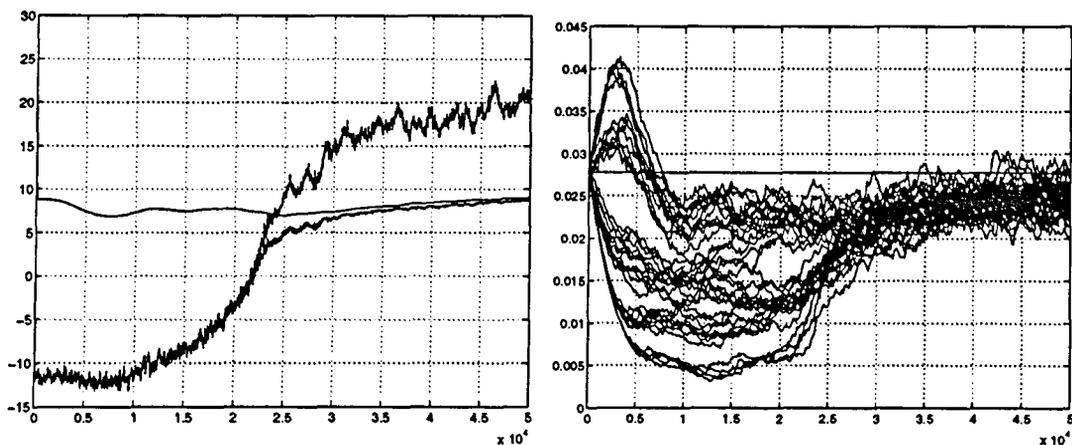


Figura 7.2.21: Izquierda: Evolución de la relación SNR, la relación señal a interferencia intersimbólica y la relación señal a ruido más interferencia a la salida del ecualizador en dB para el caso anterior. Derecha: Evolución de las probabilidades estimadas para el caso anterior.

7.3 Decorrelación Adaptativa

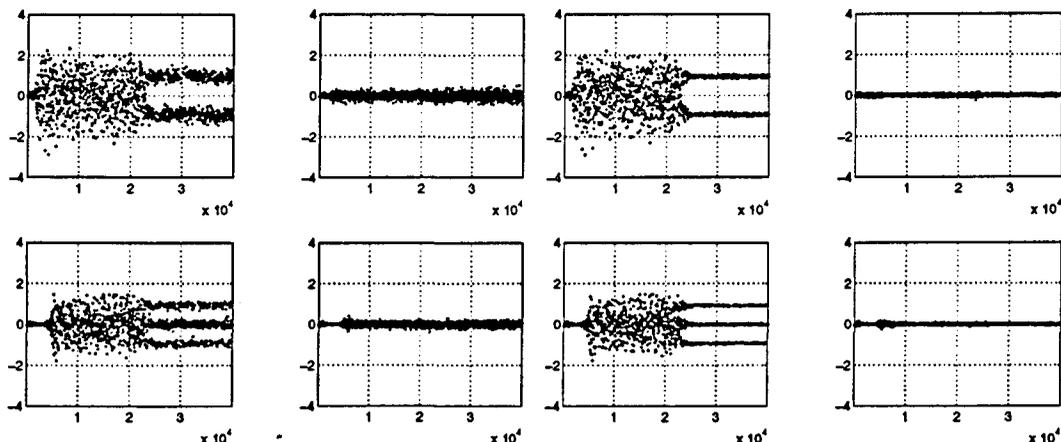


Figura 7.31: En este experimento se aplica la segunda función de coste a la decorrelación de dos usuarios que comparten el mismo canal. El primer y segundo usuarios utilizan una modulación BPSK y duobinaria respectivamente. 1) Representación de los canales fase-quadratura con ruido. 2) Representación de los canales fase-quadratura sin ruido. Se han escogido los siguientes valores para la simulación (idénticos para ambos usuarios): 1) Número de símbolos: 40000. 2) Muestras por símbolo:4. 3) Pulso-square-root raised cosine con factor de roll-off 0.4 truncado a tres símbolos. 4) Relación SNR en el canal de la señal compuesta por ambos usuarios: 10 dB. 5) Paso de actualización de los coeficientes: 0.008. 6) Paso de actualización de las probabilidades estimadas: 0.001. 7) Número de coeficientes del ecualizador: 40. 8) Factor α de la función de coste: 4. En general, y según demuestran las simulaciones, el tiempo de convergencia de la solución a la decorrelación de usuarios suele ser sustancialmente mayor que para la ecualización de un único usuario.

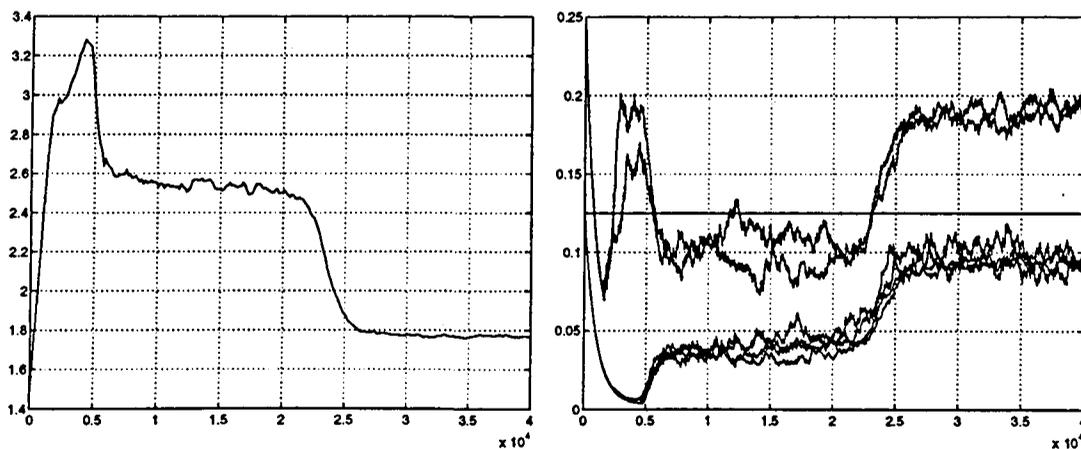


Figura 7.3.2: 1) Representación de la evolución temporal de la función de coste. 2) Representación de la evolución temporal de las probabilidades estimadas. El carácter no lineal del criterio puede observarse a través de la evolución irregular de la curva de la función de coste.

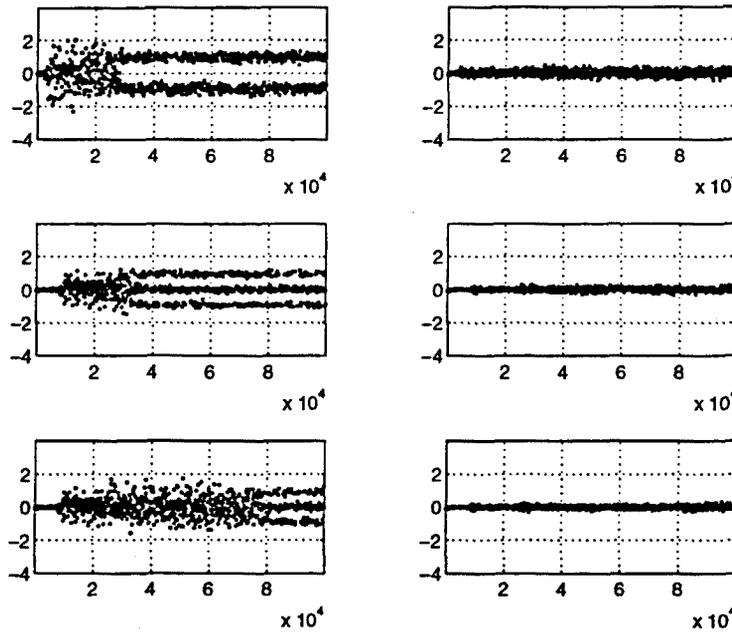


Figura 7.3.3: En este experimento se aplica la segunda función de coste a la decorrelación de tres usuarios que comparten el mismo canal. El primer usuario utiliza una modulación BPSK y el segundo y tercer usuarios utilizan una modulación duobinaria. En la gráfica se ha representado la evolución de los canales fase y cuadratura para cada uno de los usuarios con ruido. Se ha generado un pulso conformador diferente para cada usuario de tal modo que sus correlaciones cruzadas sean no nulas. Se han escogido los siguientes valores para la simulación (idénticos para ambos usuarios): 1) Número de símbolos: 100000. 2) Muestras por símbolo:4. 3) Relación SNR en el canal de la señal compuesta por ambos usuarios: 10 dB. 4) Paso de actualización de los coeficientes: 0.008. 5) Paso de actualización de las probabilidades estimadas: 0.001. 6) Número de coeficientes del ecualizador: 40. 7) Factor α de la función de coste: 4. En general, y según demuestran las simulaciones, el tiempo de convergencia de la solución a la decorrelación de usuarios suele ser sustancialmente mayor que para la ecualización de un único usuario.

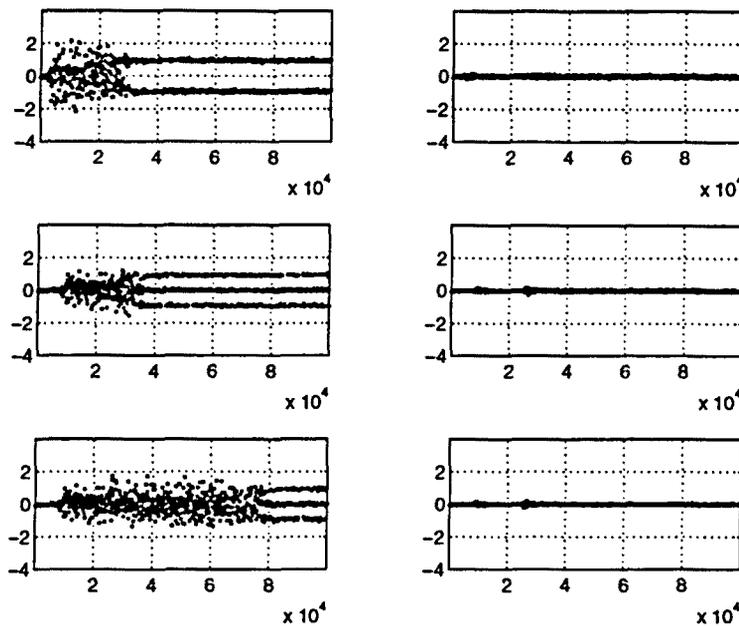


Figura 7.3.4: Representación de los canales fase-cuadratura del ejemplo anterior sustrayendo los procesos de ruido.

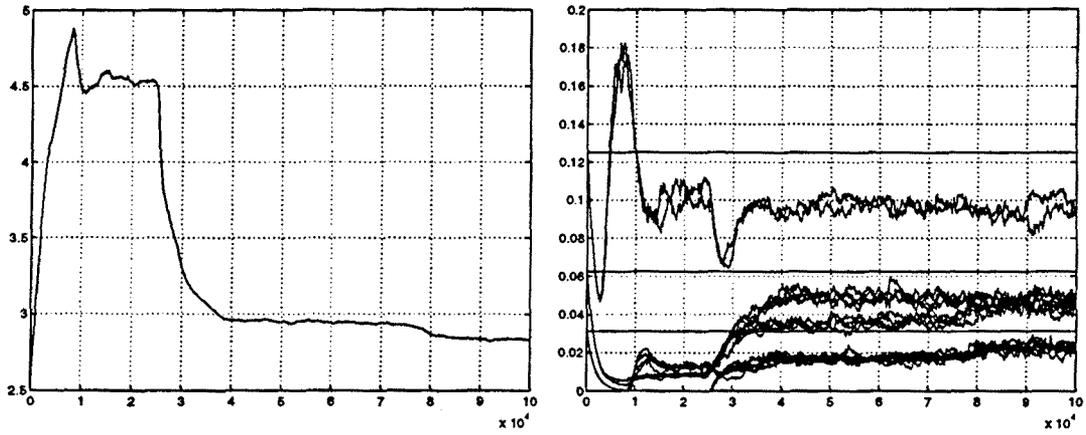


Figura 7.3.5: 1) Representación de la evolución temporal de la función de coste. 2) Representación de la evolución temporal de las probabilidades estimadas. Con respecto al caso de decorrelación de dos usuarios observamos un patente aumento del tiempo de convergencia del algoritmo.

7.4 Conformación de Haz

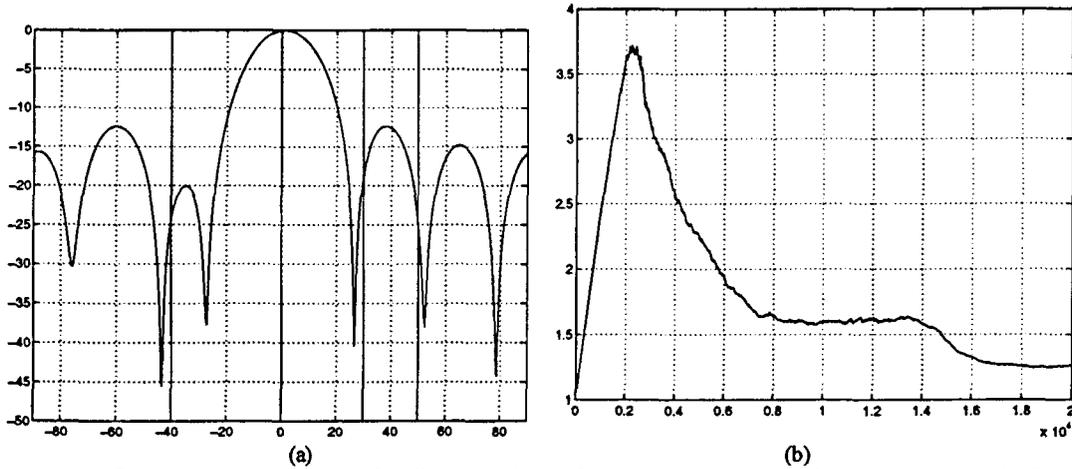


Figura 7.4.1: (a) representación del diagrama de conformación de haz obtenido en convergencia (b) Evolución de la función de coste. En esta simulación hemos considerado un array de 7 sensores sobre el que inciden 4 señales: dos de ellas interferentes Gaussianas situadas a 50 y -40 grados con un nivel de potencia de unos 16 dB por encima del nivel de ruido presente en los sensores. Las dos restantes, una señal duobinaria (señal deseada) situada a 0 grados, y una interferente QPSK situada a 30 grados, con un nivel de potencia 9.96 dB superior al de la señal duobinaria. El número de coeficientes del ecualizador situado a continuación del beamformer es de 30. Se ha utilizado la segunda función de coste para permitir discriminar la señal duobinaria de la interferente QPSK. Se ha utilizado un paso de actualización de los coeficientes del ecualizador y del beamformer de 0.0025, mientras que el paso de actualización de las probabilidades estimadas es de 0.0015. El número total de símbolos utilizados es de 20000. Obsérvese en la figura 7.4.1(a) como el conformador de haz rechaza las señales interferentes situadas en las direcciones -40, 30 y 50 grados respectivamente.

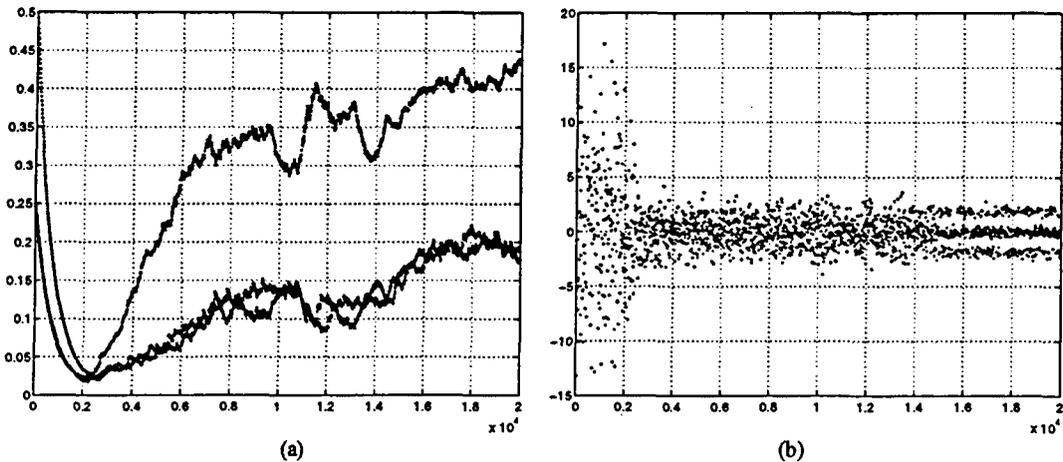


Figura 7.4.2: (a) representación de las probabilidades estimadas para la señal duobinaria en la simulación descrita en la figura 7.4.1. Obsérvese como las curvas convergen a valores cercanos a las probabilidades de la señal duobinaria sin ruido: 0.25 y 0.5. Debido a efectos conjuntos de la apertura de la ventana gaussiana de estimación de las probabilidades y del nivel de ruido subyacente, las probabilidades estimadas permanecen por debajo de los valores máximos de 0.5 y 0.25. (b) Evolución del canal de fase.

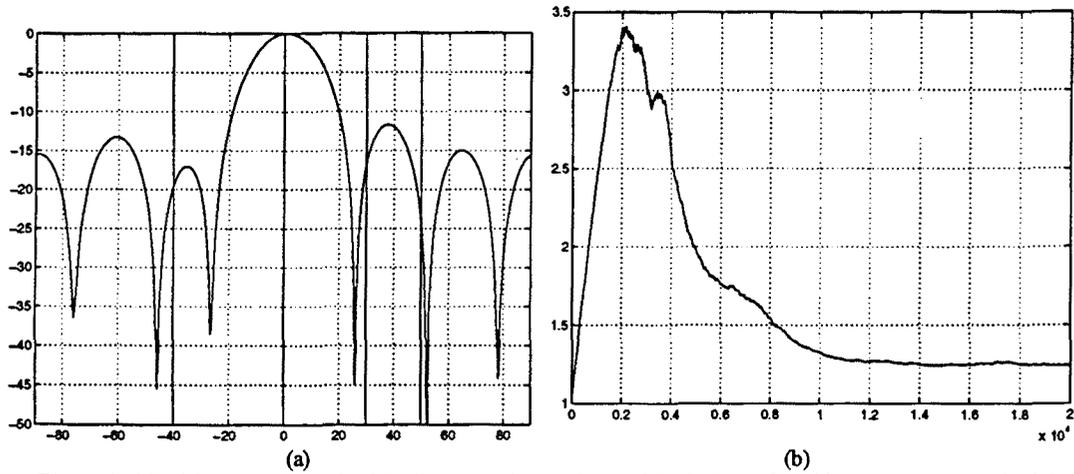


Figura 7.4.3: (a) representación del diagrama de conformación de haz obtenido en convergencia (b) Evolución de la función de coste. En esta simulación hemos considerado un array de 7 sensores sobre el que inciden 4 señales: dos de ellas interferentes Gaussianas situadas a 50 y -40 grados con un nivel de potencia de unos 16 dB por encima del nivel de ruido presente en los sensores. Las dos restantes, una señal duobinaria (señal deseada) situada a 0 grados, y una interferente 16-QAM situada a 30 grados, con un nivel de potencia 10.14 dB superior al de la señal duobinaria. El número de coeficientes del ecualizador situado a continuación del beamformer es de 30. Se ha utilizado la segunda función de coste para permitir discriminar la señal duobinaria de la interferente QPSK. Se ha utilizado un paso de actualización de los coeficientes del ecualizador y del beamformer de 0.0025, mientras que el paso de actualización de las probabilidades estimadas es de 0.0015. El número total de símbolos utilizados es de 20000. Obsérvese en la figura 7.4.3(a) como el conformador de haz rechaza las señales interferentes situadas en las direcciones -40, 30 y 50 grados respectivamente.

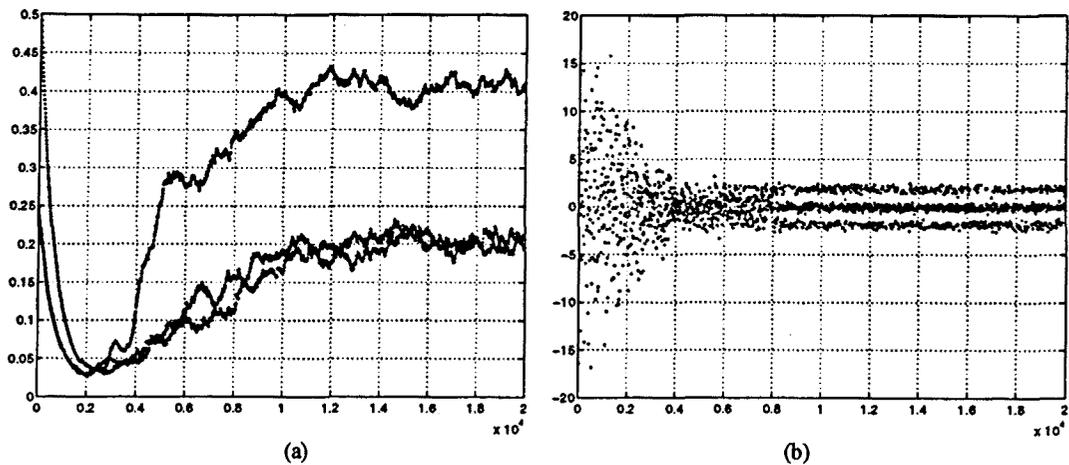


Figura 7.4.4: (a) representación de las probabilidades estimadas para la señal duobinaria en la simulación descrita en la figura 7.4.3. Obsérvese como las curvas convergen a valores cercanos a las probabilidades de la señal duobinaria sin ruido: 0.25 y 0.5. Debido a efectos conjuntos de la apertura de la ventana gaussiana de estimación de las probabilidades y del nivel de ruido subyacente, las probabilidades estimadas permanecen por debajo de los valores máximos de 0.5 y 0.25. (b) Evolución del canal de fase.

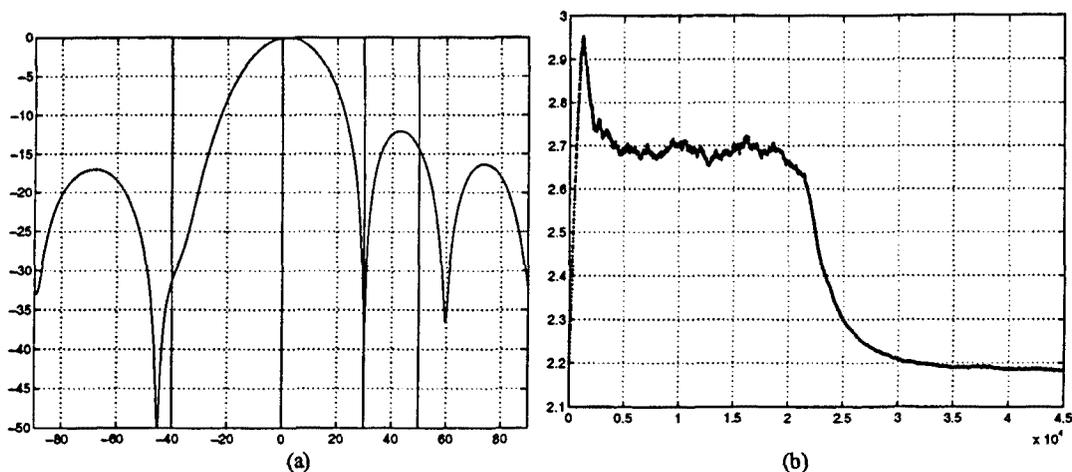


Figura 7.4.5: (a) representación del diagrama de conformación de haz obtenido en convergencia (b) Evolución de la función de coste. En esta simulación hemos considerado un array de 7 sensores sobre el que inciden 4 señales: dos de ellas interferentes Gaussianas situadas a 50 y -40 grados con un nivel de potencia de unos 23 dB por encima del nivel de ruido presente en los sensores. Las dos restantes, una señal 9-QAM (señal deseada) situada a 0 grados, y una interferente QPSK situada a 30 grados, con un nivel de potencia 0.52 dB superior al de la señal deseada. El número de coeficientes del ecualizador situado a continuación del beamformer es de 30. Se ha utilizado la segunda función de coste para permitir discriminar la señal 9-QAM de la interferente QPSK. Se ha utilizado un paso de actualización de los coeficientes del ecualizador y del beamformer de 0.004, mientras que el paso de actualización de las probabilidades estimadas es de 0.001. El número total de símbolos utilizados es de 45000. Obsérvese en la figura 7.4.5(a) como el conformador de haz rechaza las señales interferentes situadas en las direcciones -40, 30 y 50 grados respectivamente.

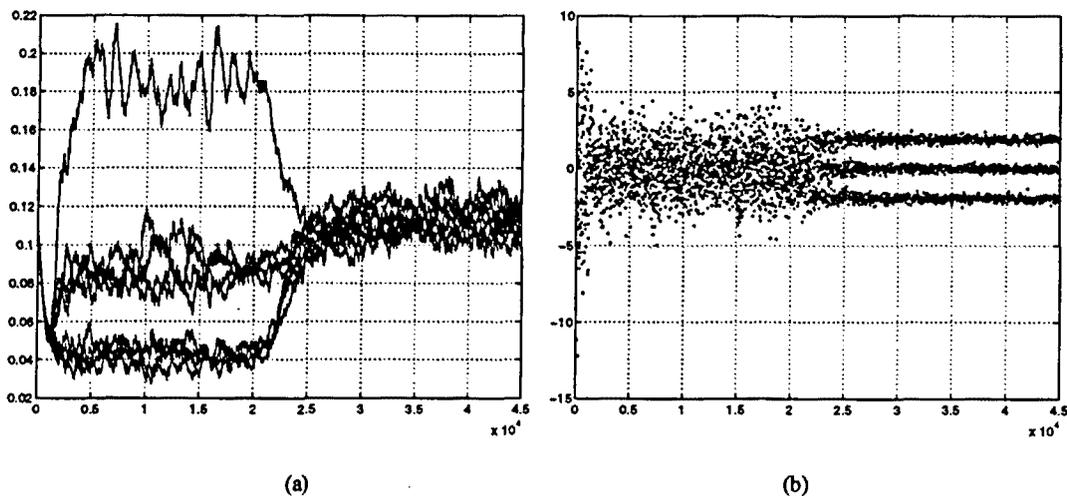


Figura 7.4.6: (a) representación de las probabilidades estimadas para la señal 9-QAM en la simulación descrita en la figura 7.4.5. Obsérvese como las curvas convergen a valores cercanos a las probabilidades de la señal 9-QAM sin ruido: 1/9. Nótese como las estimaciones de las probabilidades se han vuelto más nerviosas en comparación con las simulaciones de la señal duobinaria. Esto es debido a que hemos triplicado el número de centroides de la constelación y a que no hemos variado relativamente el paso de actualización de las probabilidades. El filtrado que se realiza es entonces menor.

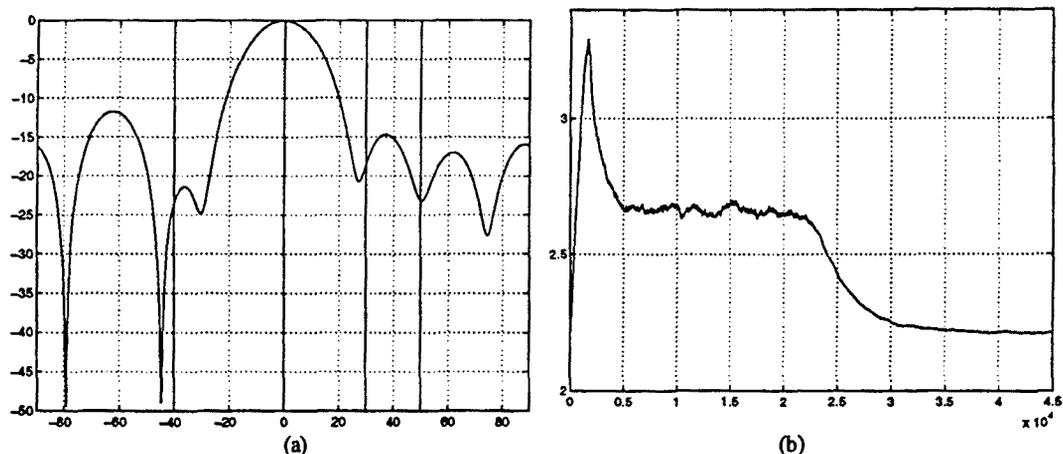


Figura 7.4.7: (a) representación del diagrama de conformación de haz obtenido en convergencia (b) Evolución de la función de coste. En esta simulación hemos considerado un array de 7 sensores sobre el que inciden 4 señales: dos de ellas interferentes Gaussianas situadas a 50 y -40 grados con un nivel de potencia de unos 23 dB por encima del nivel de ruido presente en los sensores. Las dos restantes, una señal 9-QAM (señal deseada) situada a 0 grados, y una interferente QPSK situada a 30 grados, con un nivel de potencia 6.51 dB superior al de la señal deseada. El número de coeficientes del ecualizador situado a continuación del beamformer es de 30. Se ha utilizado la segunda función de coste para permitir discriminar la señal deseada de la interferente QPSK. Se ha utilizado un paso de actualización de los coeficientes del ecualizador y del beamformer de 0.004, mientras que el paso de actualización de las probabilidades estimadas es de 0.001. El número total de símbolos utilizados es de 45000. Obsérvese en la figura 7.4.7(a) como el conformador de haz rechaza las señales interferentes situadas en las direcciones -40, 30 y 50 grados respectivamente.

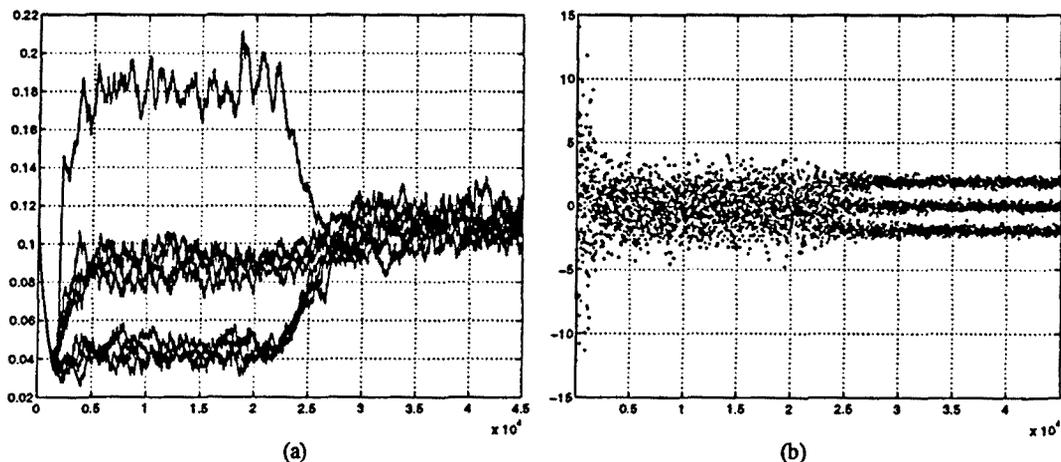


Figura 7.4.8: (a) representación de las probabilidades estimadas para la señal 9-QAM en la simulación descrita en la figura 7.4.7. Obsérvese como las curvas convergen a valores cercanos a las probabilidades de la señal 9-QAM sin ruido: 1/9. Nótese como las estimaciones de las probabilidades se han vuelto más nerviosas en comparación con las simulaciones de la señal duobinaria. Esto es debido a que hemos triplicado el número de centroides de la constelación y a que no hemos variado relativamente el paso de actualización de las probabilidades. El filtrado que se realiza es entonces menor.

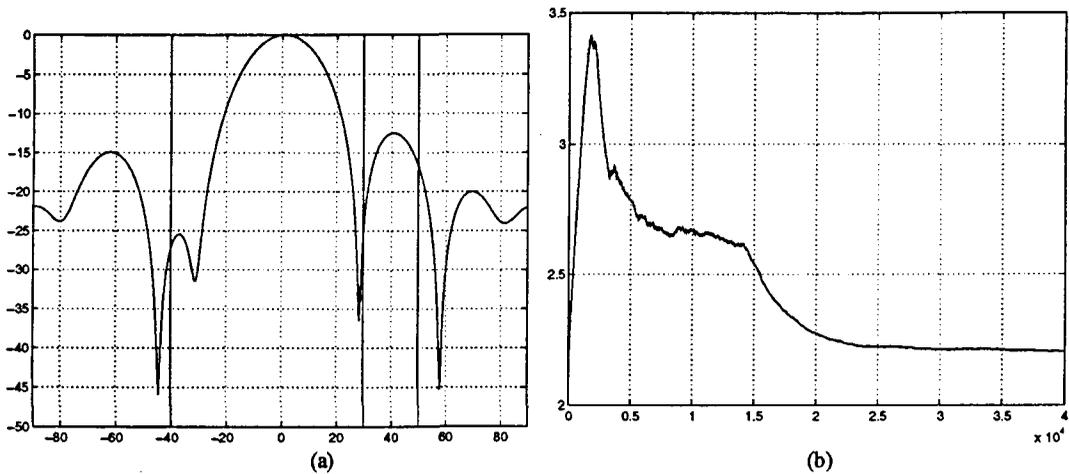


Figura 7.4.9: (a) representación del diagrama de conformación de haz obtenido en convergencia (b) Evolución de la función de coste. En esta simulación hemos considerado un array de 7 sensores sobre el que inciden 4 señales: una de ellas interferente Gaussiana situada a 50 grados con un nivel de potencia de unos 20 dB por encima del nivel de ruido presente en los sensores. Las dos restantes, una señal 9-QAM (señal deseada) situada a 0 grados, y dos interferentes, una QPSK situada a 30 grados, con un nivel de potencia 4 dB superior al de la señal 9-QAM, y otra interferente duobinaria situada a -40 grados y con un nivel de potencia 2 dB superior a la señal deseada. El número de coeficientes del ecualizador situado a continuación del beamformer es de 30. Se ha utilizado la segunda función de coste para permitir discriminar la señal 9-QAM de las interferentes. Se ha utilizado un paso de actualización de los coeficientes del ecualizador y del beamformer de 0.0035, mientras que el paso de actualización de las probabilidades estimadas es de 0.001. El número total de símbolos utilizados es de 40000. Obsérvese en la figura 7.4.9(a) como el conformador de haz rechaza las señales interferentes situadas en las direcciones -40, 30 y 50 grados respectivamente.

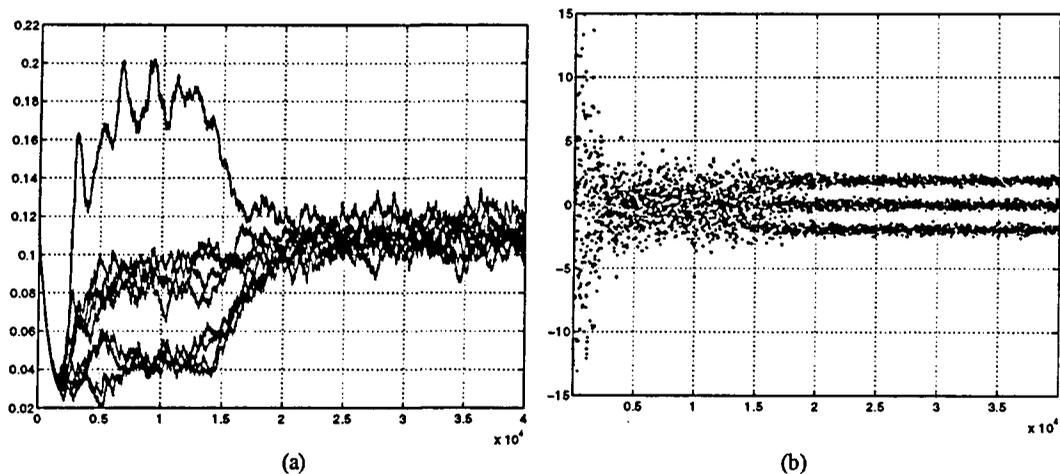


Figura 7.4.10: (a) representación de las probabilidades estimadas para la señal 9-QAM en la simulación descrita en la figura 7.4.9. Obsérvese como las curvas convergen a valores cercanos a las probabilidades de la señal 9-QAM sin ruido: $1/9$. Nótese como las estimaciones de las probabilidades se han vuelto más nerviosas en comparación con las simulaciones de la señal duobinaria. Esto es debido a que hemos triplicado el número de centroides de la constelación y a que no hemos variado relativamente el paso de actualización de las probabilidades. El filtrado que se realiza es entonces menor.

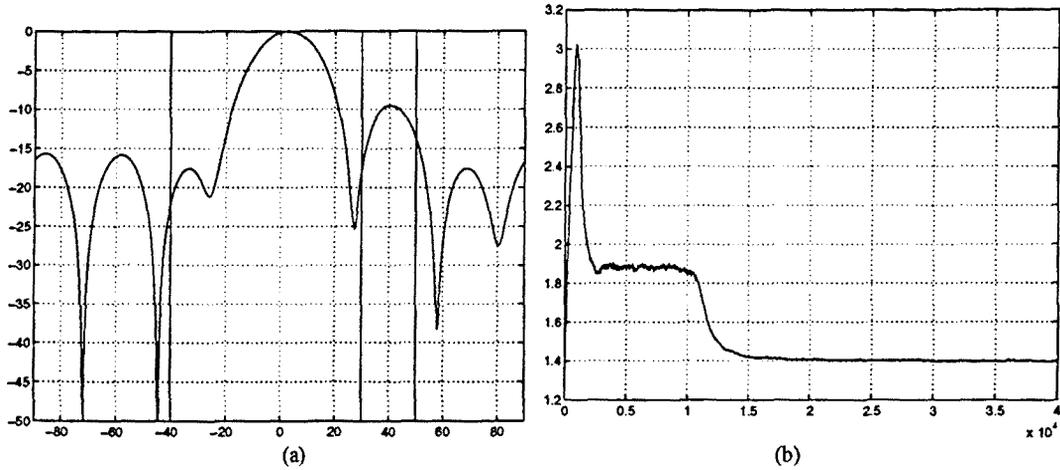


Figura 7.4.11: (a) representación del diagrama de conformación de haz obtenido en convergencia (b) Evolución de la función de coste. En esta simulación hemos considerado un array de 7 sensores sobre el que inciden 4 señales: una de ellas interferente Gaussiana situada a 50 grados con un nivel de potencia de unos 20 dB por encima del nivel de ruido presente en los sensores. Las dos restantes, una señal QPSK (señal deseada) situada a 0 grados, y dos interferentes, una 9-QAM situada a 30 grados, con un nivel de potencia 6.5 dB superior al de la señal QPSK, y otra interferente 9-QAM situada a -40 grados y con un nivel de potencia 6.5 dB superior a la señal deseada El número de coeficientes del ecualizador situado a continuación del beamformer es de 30. Se ha utilizado la segunda función de coste para permitir discriminar la señal QPSK de las interferentes. Se ha utilizado un paso de actualización de los coeficientes del ecualizador y del beamformer de 0.003, mientras que el paso de actualización de las probabilidades estimadas es de 0.0025. El número total de símbolos utilizados es de 40000. Obsérvese en la figura 7.4.11(a) como el conformador de haz rechaza las señales interferentes situadas en las direcciones -40, 30 y 50 grados respectivamente.

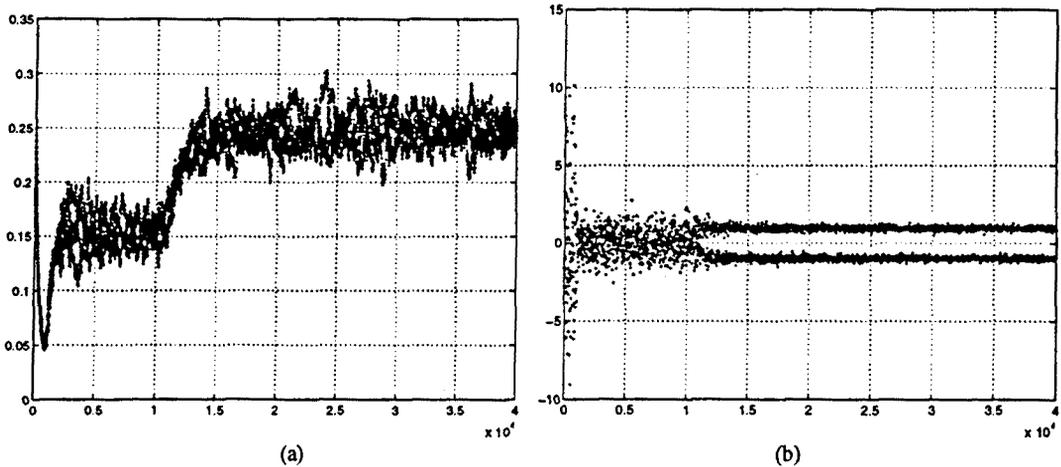
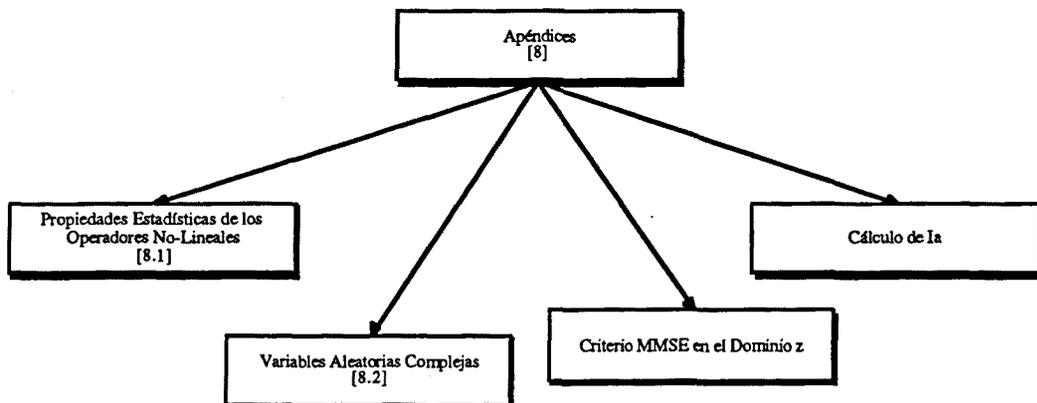


Figura 7.4.12: (a) representación de las probabilidades estimadas para la señal QPSK en la simulación descrita en la figura 7.4.11. Obsérvese como las curvas convergen a valores cercanos a las probabilidades de la señal QPSK sin ruido: 1/4.

Capítulo VIII

Apéndices



8.1 Propiedades Estadísticas de los Operadores No-Lineales

En este apéndice estudiaremos ciertas propiedades estadísticas de operadores no lineales utilizadas en algunos capítulos de este trabajo. Nos concentraremos específicamente en el estudio de la esperanza de un operador no lineal que actúa sobre varios procesos estocásticos. Definiremos también la notación utilizada para el tratamiento de expresiones complejas en términos de esperanzas estadísticas con respecto a variables aleatorias distintas.

8.1.1 El operador esperanza

Consideremos una función no lineal de varias variables aleatorias definimos su esperanza de la siguiente forma,

$$E\psi(a_1, \dots, a_N) = E_{a_1, \dots, a_N} \psi(a_1, \dots, a_N) \quad (8.1.1)$$

donde en los subíndices de E explicitamos sobre qué variables se calcula la esperanza. Entendemos pues que el subíndice del operador indica cual es la función de densidad que debe utilizarse cuando recurrimos a la notación integral, es decir,

$$E_{a_1, \dots, a_N} \psi(a_1, \dots, a_N) = \int_{C^N} p_{a_1, \dots, a_N}(a_1, \dots, a_N) \psi(a_1, \dots, a_N) da_1, \dots, da_N \quad (8.1.2)$$

En el caso particular de que todas las variables aleatorias sean independientes dos a dos, podemos factorizar el operador esperanza en base a los operadores de cada variable individual,

$$E_{a_1, \dots, a_N} \psi(a_1, \dots, a_N) = E_{a_1} E_{a_2} \dots E_{a_N} \psi(a_1, \dots, a_N) \quad (8.1.2)$$

en todas sus permutaciones posibles. Hecho que tiene su traducción directa en la factorización de las densidades de probabilidad marginales. El caso mas general de factorización cuando existe dependencia estadística entre variables se formula en términos de las variables condicionadas de la siguiente forma,

$$(8.1.3)$$

$$E_{a_1, \dots, a_N} \psi(a_1, \dots, a_N) = E_{a_1} E_{a_2 | a_1} \dots E_{a_N | a_{N-1}, \dots, a_1} \psi(a_1, \dots, a_N)$$

donde tal descomposición es análoga a la factorización de la función de densidad conjunta en funciones de densidad condicionadas, tal como aparecen en los subíndices.

8.2 Esperanza de una no-linealidad

Consideremos dos procesos estocásticos definidos por las variables aleatorias a_k y b_k independientes entre sí y consideremos también la esperanza de una no-linealidad ψ aplicada sobre un tercer proceso z_k suma de los anteriores. Supondremos que ambos procesos son escalares y complejos. Definimos la magnitud estadística J o función de coste como,

$$J = E_{z_k} \psi(z_k) \tag{8.1.1}$$

Según el teorema fundamental de la esperanza podemos expresar la función de coste como,

$$J = E_{a_k, b_k} \psi(a_k + b_k) \tag{8.1.2}$$

donde observamos que podemos factorizar el operador esperanza en base a la suposición de independencia entre ambas variables. Para la deducción pasaremos la ecuación anterior a forma integral en base al teorema fundamental de la esperanza,

$$E_{a_k} E_{b_k} \psi(a_k + b_k) = E_{a_k} \int_C p_{b_k}(b_k) \psi(a_k + b_k) db_k \tag{8.1.3}$$

Vemos que el término integral puede interpretarse como la convolución de la no-linealidad con la función de densidad del proceso b_k . Efectuemos entonces el siguiente cambio de variable : $b_k = -c_k$,

$$E_{a_k} \int_C p_{b_k}(-c_k) \psi(a_k - c_k) d(-c_k) = E_{a_k} - \int_C p_{-b_k}(c_k) \psi(a_k - c_k) dc_k \tag{8.1.4}$$

Por lo tanto,

$$J = E_{a_k, b_k} \psi(a_k + b_k) = E_{a_k} (\psi * p_{-b_k})(a_k) \tag{8.1.5}$$

La igualdad anterior es igualmente aplicable al caso de procesos estocásticos vectoriales y operadores no-lineales de varias variables, donde la convolución pasa a ser multi-dimensional. La interpretación de la ecuación anterior es que traducimos la incertidumbre añadida a la variable a_k por b_k (es decir, el ensanchamiento de su función de densidad) por el ensanchamiento de la no-linealidad. Observemos la figura 8.1.1 donde hemos calculado la esperanza de forma directa y según 8.1.1. Vemos fácilmente que ambos términos coinciden.

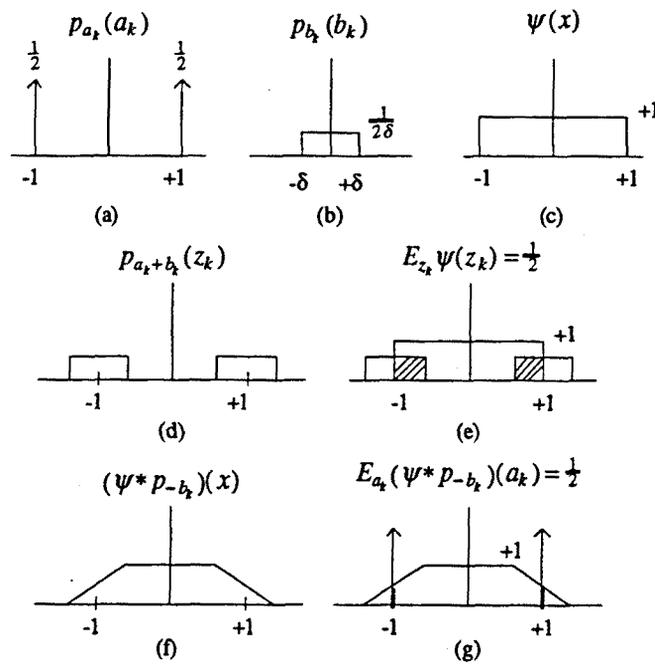


Figura 8.1.1: caso particular del intercambio de convoluciones

8.2 Variables Aleatorias Complejas

Trabajaremos con cierta frecuencia con funciones de densidad de probabilidad de variables aleatorias complejas. Se define este tipo de funciones de densidad como la función de densidad conjunta de sus partes reales e imaginarias. Algunas propiedades de estas funciones nos serán de utilidad. Cabe destacar que las partes real e imaginaria de una variable aleatoria compleja no son necesariamente independientes. Denotémoslas respectivamente como z_r y z_i , tenemos entonces que,

$$p_Z(z) \stackrel{def}{=} p_{Z_r}(z_r) p_{Z_i|Z_r}(z_i|z_r) = p_{Z_i}(z_i) p_{Z_r|Z_i}(z_r|z_i) \quad (8.2.1)$$

Podemos considerar una variable aleatoria compleja como una variable aleatoria vectorial de segundo orden. Si existe independencia entre las partes real e imaginaria tendremos,

$$p_Z(z) = p_{Z_r}(z_r) p_{Z_i}(z_i) \quad (8.2.2)$$

De las dos propiedades anteriores podemos deducir que la función de densidad de una variable compleja es una función no analítica por el hecho de que se puede expresar de forma irreducible en función de sus partes real e imaginaria. Es decir,

$$p_Z(z) \stackrel{def}{=} p_{Z_r}(z_r) p_{Z_i|Z_r}(z_i|z_r) = p_{Z_i} \left(\frac{z+z^*}{2} \right) p_{Z_r|Z_i} \left(\frac{z+z^*}{2} \middle| \frac{z-z^*}{2j} \right) \quad (8.2.3)$$

Podemos considerar a modo de ejemplo el caso de una variable compleja Gaussiana de componentes incorreladas que simule el ruido complejo fase-cuadratura. Su función de distribución depende del módulo al cuadrado de la variable ruido, siendo claramente una función no-analítica. Por consiguiente, siempre que debamos derivar una función de densidad compleja deberemos aplicar la regla de la cadena respecto a la variable aleatoria y a su conjugada.

I Suma de una constante

Consideremos la variable $z' = z + \beta$, su función de densidad viene dada por,

$$\begin{aligned} p_{Z'=z+\beta}(z') &= p_{Z'_r=z_r+\beta_r}(z'_r) p_{Z'_i=z_i+\beta_i|Z'_r}(z'_i|z'_r) = p_{Z_r}(z'_r - \beta_r) p_{Z_i|Z_r}(z'_i - \beta_i | z'_r - \beta_r) \\ &= p_Z(z' - \beta) \end{aligned} \quad (8.2.3)$$

II Multiplicación por una constante

Consideremos la variable aleatoria $z'=\alpha z$. La multiplicación por la constante compleja α puede considerarse equivalente a un cambio de coordenadas del vector z en el espacio \mathfrak{R}^2 .

$$\begin{bmatrix} z'_r \\ z'_i \end{bmatrix} = \begin{bmatrix} \alpha_r & -\alpha_i \\ \alpha_i & \alpha_r \end{bmatrix} \begin{bmatrix} z_r \\ z_i \end{bmatrix} = \mathbf{A} \begin{bmatrix} z_r \\ z_i \end{bmatrix} \quad (8.2.4)$$

La función de densidad de z' viene dada por,

$$p_{Z'=\alpha Z}(z') = \frac{1}{|\mathbf{A}|} p_Z(\mathbf{A}^{-1}z') \quad (8.2.5)$$

donde $|\mathbf{A}|$ denota el determinante del Jacobiano de la transformación. En este caso, $\det(\mathbf{A})=\alpha^*\alpha=|\alpha|^2$. Nótese que la pre-multiplicación por la inversa de \mathbf{A} es análoga a la división compleja por α , por lo tanto,

$$p_{Z'=\alpha Z}(z') = \frac{1}{|\alpha|^2} p_Z \left(\text{Im} \frac{z'}{\alpha} \right) p_{Z_1|Z_r} \left(\text{Re} \frac{z'}{\alpha} \mid \text{Im} \frac{z'}{\alpha} \right) = \frac{1}{|\alpha|^2} p_Z \left(\frac{z'}{\alpha} \right) \quad (8.2.6)$$

III Función de densidad Gaussiana

Supongamos una variable aleatoria compleja cuyas componentes real e imaginaria son variables Gaussianas e incorreladas de la misma varianza σ_r . La función de densidad conjunta viene entonces dada por el producto de las densidades marginales,

$$p_n(n) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{-n_r^2/2\sigma_r^2} \frac{1}{\sqrt{2\pi}\sigma_r} e^{-n_i^2/2\sigma_r^2} = \frac{1}{2\pi\sigma_r^2} e^{-(n_r^2+n_i^2)/2\sigma_r^2} \quad (8.2.7)$$

Pero la varianza del ruido complejo es el doble que la del ruido real, por lo tanto,

$$p_n(n) = \frac{1}{\pi\sigma^2} e^{-|n|^2/\sigma^2} \quad , \quad \sigma^2 = 2\sigma_r^2 \quad (8.2.8)$$

8.3 Criterio MMSE en el Dominio z

El criterio de mínimos cuadrados constituye el algoritmo de estimación óptimo bajo condiciones Gaussianas. Su formulación precisa de la matriz de correlación del ruido. No obstante, cuando consideramos ruido temporalmente correlado y condiciones estacionarias, el planteamiento del criterio MMSE asintótico en el dominio temporal conduce al manejo de matrices de correlación (y de sus inversas) de dimensión infinita. Resulta más práctico en estos casos buscar criterios equivalentes que faciliten el desarrollo de los algoritmos óptimos. Formularemos tales criterios para el caso vectorial general.

8.3.1 Planteamiento

Definamos la realización de un proceso estocástico vectorial Gaussiano como la secuencia de vectores \mathbf{e}_k cuya función de correlación viene dada por la siguiente expresión,

$$\mathbf{R}_{ee}(k) = E\mathbf{e}_m\mathbf{e}_{m+k}^T \tag{8.3.1.1}$$

Definiendo el vector bloque \mathbf{e}_N y la matriz bloque \mathbf{R}_N o matriz de correlación del vector bloque \mathbf{e}_N como sigue,

$$\mathbf{e}_N^T = [\mathbf{e}^T(-N'), \dots, \mathbf{e}^T(+N')] \quad , \quad \mathbf{R}_N = \begin{bmatrix} \mathbf{R}_{ee}(0) & \mathbf{R}_{ee}^H(1) & \mathbf{R}_{ee}^H(2) & \dots & & \\ \mathbf{R}_{ee}(1) & \mathbf{R}_{ee}(0) & \mathbf{R}_{ee}^H(1) & & & \\ \mathbf{R}_{ee}(2) & \mathbf{R}_{ee}(1) & \mathbf{R}_{ee}(0) & & & \\ \vdots & & & \ddots & & \\ & & & & \mathbf{R}_{ee}^H(1) & \\ & & & & \mathbf{R}_{ee}(1) & \mathbf{R}_{ee}(0) \end{bmatrix} \tag{8.3.1.2}$$

El criterio MMSE temporal para un bloque finito de longitud N de los datos se define a partir de la función de densidad conjunta del vector \mathbf{e}_N como,

$$\sigma_N^2 = \frac{1}{N} \mathbf{e}_N^H \mathbf{R}_N^{-1} \mathbf{e}_N = \frac{1}{N} \sum_{i=-N'}^{+N'} \sum_{i'=-N'}^{+N'} \mathbf{e}^H(i) [\mathbf{R}_N^{-1}]_{ii'} \mathbf{e}(i') \tag{8.3.1.3}$$

donde $[\cdot]_{ii'}$ denota el subbloque (i,i') de la matriz considerada. El factor de escala 1/N se ha introducido para evitar la divergencia cuando N tiende a infinito. Evidentemente, para longitudes cada vez mayores, la dimensión de la matriz de correlación tiende a infinito.

8.3.2 Desarrollo

Procedamos de forma equivalente con la transformada discreta de Fourier de la secuencia de error. Definimos para ello la transformada de una secuencia vectorial de la siguiente forma,

$$\mathbf{e}(w_N^{-k}) = \frac{1}{N} \sum_{n=-N'}^{n=+N'} \mathbf{e}(n) w_N^{-kn} \quad , \quad w_N^{-1} = e^{-j\frac{2\pi}{N}} \quad (8.3.2.1)$$

Denominando Γ_N a la matriz de correlación bloque entre bins de frecuencia distintos, obtenemos la siguiente expresión del error en el dominio de la frecuencia,

$$\phi_N^2 = \frac{1}{N^2} \sum_{k=-N'}^{+N'} \sum_{k'=-N'}^{+N'} \mathbf{e}(w_N^{-k})^H [\Gamma_N^{-1}]_{kk'} \mathbf{e}(w_N^{-k'}) \quad (8.3.2.2)$$

A continuación calcularemos la expresión de cada una de las submatrices de la ecuación 8.3.2.2. La expresión para el bloque (k,k') vendrá dada por definición como,

$$\begin{aligned} [\Gamma_N]_{kk'} &= E \mathbf{e}(w_N^{-k}) \mathbf{e}(w_N^{-k'})^H = \frac{1}{N^2} E \sum_{n=-N'}^{+N'} \mathbf{e}(n) w_N^{-kn} \sum_{n'=-N'}^{+N'} \mathbf{e}^H(n') w_N^{+k'n'} \\ &= \frac{1}{N^2} E \sum_{n,n'} \mathbf{e}(n) \mathbf{e}^H(n') w_N^{-(kn-k'n')} = \frac{1}{N^2} \sum_{n,n'} \mathbf{R}_{ee}(n-n') w_N^{-(kn-k'n')} \end{aligned} \quad (8.3.2.3)$$

A continuación consideramos dos casos disjuntos de la ecuación 8.3.2.3: (a) aquel en que $k=k'$ y (b) con $k \neq k'$,

$$\begin{aligned} [\Gamma_N]_{kk} &= \frac{1}{N^2} \sum_{n,n'} \mathbf{R}_{ee}(n-n') w_N^{-k(n-n')} = \frac{1}{N^2} \sum_{\delta} \mathbf{R}_{ee}(\delta) w_N^{-k\delta} \sum_{n-n'=\delta} 1 \\ &= \frac{1}{N} \sum_{\delta} \left(1 - \frac{|\delta|}{N}\right) \mathbf{R}_{ee}(\delta) w_N^{-k\delta} \end{aligned} \quad (8.3.2.4)$$

donde podemos observar que para N finito, la matriz de correlación en frecuencia corresponde a la transformada discreta de la secuencia de correlación envantada con una ponderación triangular. A medida que N aumente hasta infinito, el enventado tenderá a una ponderación uniforme. Pasemos ahora a evaluar el caso (b) con $k \neq k'$,

$$(8.3.2.5)$$

$$\begin{aligned}
 [\Gamma_N]_{kk'} &= \frac{1}{N^2} \sum_{n,n'} \text{Ree}(n-n') w_N^{-(kn-k'n')} = \frac{1}{N^2} \sum_{\delta} \text{Ree}(\delta) \sum_{n-n'=\delta} w_N^{-(kn-k'n')} \\
 &= \frac{1}{N^2} \sum_{\delta} \text{Ree}(\delta) \sum_{n,\delta} w_N^{-(kn-k'n+k'\delta)} = \left(\frac{1}{N} \sum_{\delta} \text{Ree}(\delta) w_N^{-k'\delta} \right) \left(\frac{1}{N} \sum_n w_N^{-(k-k')n} \right) = [\Gamma_N]_{kk'} \delta_{kk'}
 \end{aligned}$$

donde el segundo término es igual a cero para $k \neq k'$ por tratarse del sumatorio sobre un número entero de períodos de una exponencial compleja de frecuencia entera, siendo igual por consiguiente a la delta de Kronecker o DFT de la función uniforme. Hemos demostrado que bins de frecuencias distintas están incorrelados para las frecuencias enteras de la DFT. La matriz bloque de correlación en frecuencia es por consiguiente diagonal para todo N , siendo los términos de la diagonal igual al espectro de potencia del proceso en un bin de frecuencia determinado. No obstante, si interpolamos la DFT entre bins enteros, existe correlación entre dichos términos. Tal efecto es debido a que el envanado rectangular en tiempo se traduce en una convolución del espectro con una función sinc que introduce correlación entre frecuencias que se encuentren entre sí a una distancia inferior a la anchura efectiva de su lóbulo principal. Sustituyendo la expresión 8.3.2.5 en la ecuación 8.3.2.2 y teniendo en cuenta que las matrices correspondientes a los términos cruzados son cero, llegamos a la siguiente expresión de la función error, donde el doble sumatorio se ha reducido a un sumatorio simple,

$$\phi_N^2 = \frac{1}{N} \sum_{k=-N'}^{+N'} e^{(w_N^{-k})^H} [\Gamma_N^{-1}]_{kk} e^{(w_N^{-k'})} \tag{8.3.2.6}$$

8.3.3 Paso al Dominio z

El traducción al dominio z del criterio deducido en el apartado anterior es inmediato mediante paso al límite, haciendo tender a infinito la dimensión de la ventana temporal. Definamos la secuencia de correlación transformada junto con la transformada del proceso e_k de la siguiente forma,

$$\begin{aligned}
 \text{Ree}(z^{-1}) &= \sum_{k=-\infty}^{k=+\infty} \text{Ree}(k) z^{-k} \\
 e(z^{-1}) &= \sum_{k=-\infty}^{k=+\infty} e(k) z^{-k}
 \end{aligned} \tag{8.3.3.1}$$

El conjunto de frecuencias $\{2\pi k/N, 0 \leq k \leq N-1\}$ de la DFT de orden N tiende a llenar el intervalo $[0, 2\pi)$ o bien $[-\pi, +\pi)$ a medida que N tiende a infinito, por lo cual la k -ésima

potencia del factor de twiddle, w_N^{-k} pasa a ser sustituido por el fasor $e^{-j\omega}$. Entonces, efectuando el proceso al límite de la ecuación 8.3.2.6 llegamos a la siguiente expresión para el error asintótico,

$$\begin{aligned} \phi^2 &= \lim_{N \rightarrow \infty} \phi_N^2 = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=-N'}^{+N'} e^{(w_N^{-k})^H} [\Gamma_N^{-1}]_{kk} e^{(w_N^{-k'})} \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{(e^{-j\omega})^H} \Gamma^{-1}(e^{-j\omega}) e^{(e^{-j\omega})} d\omega \end{aligned} \tag{8.3.3.2}$$

teniendo en cuenta que el diferencial de frecuencia angular es un infinitésimo equivalente de $2\pi/N$. En la anterior ecuación se ha utilizado la propiedad asintótica equivalente a la deducida en 8.3.2.4,

$$\Gamma(z^{-1}) = \sum_{\delta} R_{ee}(\delta) z^{-\delta} = R_{ee}(z^{-1}) \tag{8.3.3.3}$$

Podemos definir finalmente el criterio MMSE en el dominio z de las dos formas equivalentes,

$$\begin{aligned} \phi^2 &= \frac{1}{2\pi j} \oint_{C_1} e^{(z^{-1})^H} R_{ee}^{-1}(z^{-1}) e^{(z^{-1})} z^{-1} dz \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{(e^{-j\omega})^H} R_{ee}^{-1}(e^{-j\omega}) e^{(e^{-j\omega})} d\omega \end{aligned} \tag{8.3.2.4}$$

Cada uno de los bins de frecuencia está ponderada por el inverso del espectro de potencia en ese punto.

8.4 Cálculo de I_A

En este apéndice calcularemos una expresión cerrada de la integral que define la función $I_A(\rho)$ en términos de funciones conocidas. La demostración se basa en la transformación de la integral $I_A(\rho)$ en una ecuación diferencial de solución previamente conocida. Prescindiendo del término $2A/\sigma^2$, definimos la integral equivalente $A(x)$ y calculamos su primera y segunda derivadas,

$$A(x) = \int_{-\pi}^{+\pi} e^{x \cos \theta} d\theta \quad , \quad \frac{d}{dx} A = \int_{-\pi}^{+\pi} \cos \theta e^{x \cos \theta} d\theta \quad , \quad \frac{d^2}{dx^2} A = \int_{-\pi}^{+\pi} \cos^2 \theta e^{x \cos \theta} d\theta \tag{8.4.1}$$

Centrémonos ahora en la primera derivada y efectuemos una integración por partes,

$$\frac{d}{dx} A = e^{x \cos \theta} \sin \theta \Big|_{-\pi}^{+\pi} + x \int_{-\pi}^{+\pi} \sin^2 \theta e^{x \cos \theta} d\theta = x \int_{-\pi}^{+\pi} \sin^2 \theta e^{x \cos \theta} d\theta \tag{8.4.2}$$

pasando el argumento x dividiendo al lado izquierdo de la ecuación, podemos relacionar la primera y segunda derivadas de $A(x)$ con la función misma,

$$\frac{d^2}{dx^2} A + \frac{1}{x} \frac{d}{dx} A = \int_{-\pi}^{+\pi} (\sin^2 \theta + \cos^2 \theta) e^{x \cos \theta} d\theta = A(x) \tag{8.4.3}$$

Consideremos ahora la ecuación diferencial modificada de Bessel de orden n y su conjunto de soluciones,

$$\left(x^2 \frac{d^2}{dx^2} + x \frac{d}{dx} - (x^2 + n^2) \right) y = 0 \quad \Leftrightarrow \quad y = c_1 I_n(x) + c_2 K_n(x) \tag{8.4.4}$$

Donde las funciones $I_n(x)$ y $K_n(x)$ son las funciones modificadas de Bessel de primera y segunda especie respectivamente. Vemos pues que si particularizamos para $n=0$ y dividimos la ecuación por x^2 , obtenemos la misma igualdad que en la ecuación (8.4.3). $A(x)$ puede entonces obtenerse como solución d la ecuación modificada de Bessel con las condiciones de contorno adecuadas para determinar los valores de la combinación lineal, c_1 y c_2 . Podemos obtener fácilmente una condición de contorno para el caso particular $x=0$ en que la integral $A(x)$ es calculable e igual a 2π ,

$$A(0) = 2\pi = c_1 I_0(0) + c_2 K_0(0) = c_1 1 + c_2 \infty \quad \Rightarrow \quad c_1 = 2\pi, c_2 = 0 \tag{8.4.5}$$

de donde deducimos la expresión final de $I_A(\rho)$,

$$I_A(\rho) = A\left(\frac{2A}{\sigma^2}\rho\right) = 2\pi I_0\left(\frac{2A}{\sigma^2}\rho\right) \quad (8.4.6)$$

Capítulo IX
Conclusiones

9 Conclusiones

El objetivo último de esta tesis lo ha constituido el estudio de nuevos criterios de diseño de funciones de coste para algoritmos adaptativos. Hemos estudiado dos criterios ya conocidos bajo una nueva perspectiva: los sistemas de Referencia Temporal y de Referencia Espacial. Siendo nuestro principal interés la adaptación ciega, hemos propuesto el criterio de Referencia Estadística cuya idea fundamental consiste en forzar una distribución de probabilidad deseada a la salida de un filtro adaptativo. Este es un criterio que surge de forma natural en algoritmos ciegos, puesto que no disponemos de información espacial ni temporal de la señal o señales que deseamos recuperar. Sí en cambio, de su estadística.

El criterio de Referencia Estadística constituye únicamente una filosofía bajo la que basar el diseño de funciones de coste. En este trabajo hemos presentado dos posibles funciones de coste que implementan el criterio. La primera, y la más sencilla, hemos demostrado que permite la ecualización ciega bajo ciertas condiciones en términos del tipo de estadística de la señal de entrada. Problemas que surgían en cierto tipo de distribuciones se han solventado mediante la introducción de una segunda función de coste.

Un concepto importante que ha surgido del criterio de Referencia Estadística es el de Regeneración de la Referencia. en el criterio de Referencia Temporal se dispone de información total o parcial de la secuencia transmitida, hecho que permite una adaptación rápida del sistema. La ausencia de esta información en los dos métodos de Referencia Estadística que hemos estudiado en esta tesis se resuelve de forma natural mediante la aparición de un operador no-lineal, función de la estadística, que regenera los datos transmitidos y permite la construcción de una función de error de forma análoga a los criterios de Referencia Temporal. La regeneración no-lineal de los datos parte del conocimiento de las funciones de calidad del sistema. Hemos visto que para ambas funciones de coste, las funciones de calidad se actualizan conjuntamente con los coeficientes, a fin de garantizar un comportamiento óptimo. Dicha actualización es sin embargo opcional en el caso de la primera función de coste.

En general, debido al fuerte carácter no-lineal de los criterios de Referencia Estadística, resulta difícil llegar a expresiones cerradas que cuantifiquen el comportamiento de los algoritmos. Hemos visto en cambio, que debido a la estructura especial de las funciones de coste que hemos propuesto, es posible llegar a expresiones para los vectores de coeficientes que minimizan el criterio. En particular, un resultado importante lo encontramos en la

analogía con el criterio de mínima varianza para el caso de una única fuente. Para ambas funciones de coste hemos visto que el vector de coeficientes se encuentra contenido en el espacio de señal (modificado por la matriz inversa de correlación del ruido), lo cual garantiza un comportamiento coherente de la función de coste.

Hemos visto que el criterio de Referencia Estadística resulta muy versátil en el sentido de que no hace restricciones en cuanto a la estadística de la señales o señales de entrada y salida del sistema. Es decir, puede utilizarse ya sea para recuperar señales escalares como por ejemplo en ecualización y conformación de haz, o bien para señales vectoriales como en decorrelación de señales CDMA, conformación de haz múltiple y separación de fuentes. Tampoco se introducen restricciones en cuanto a la arquitectura del sistema. A este respecto hemos visto ejemplo de arquitecturas lineales, cuadráticas y de órdenes incluso superiores en los coeficientes.

9.1 Futuras Líneas de Investigación.

Al finalizar este trabajo creemos que aún queda mucho por hacer en este campo, sobretodo en el diseño de nuevas funciones de coste, basadas quizá en filosofías alternativas a la discriminación de Kullback, la cual ha constituido el punto de partida de este trabajo.

Los algoritmos de actualización desarrollados en esta tesis se han basado únicamente en técnicas de gradiente. El estudio de algoritmos alternativos, apropiados para funciones multimodales, como por ejemplo "Simulated Annealing", pueden abrir nuevos campos de trabajo en la optimización de criterios de Referencia Estadística. En particular, aunque hasta ahora no se ha considerado, creemos que la aplicación de Redes Neuronales puede resultar interesante en este contexto.

Capítulo X
Bibliografía

10 Bibliografía

- [AND79]. Brian D. O. Anderson, John B. Moore. "Optimal Filtering". Prentice-Hall. Electrical Engineering Series. 1979.
- [ASC89]. Gerd Ascheid, Martin Oerder, Johannes Stahl, Heinrich Meyr. "An All Digital Receiver Architecture for Bandwidth Efficient Transmission at High Data Rates". IEEE Transactions on Communications VOL-37 NO-8 August 1989.
- [BEL70]. Richard Bellman. "Introduction to Matrix Analysis". 1970.
- [BLA87] Richard E. Blahut. "Principles and Practice of Information Theory". Addison Wesley 1987.
- [BRO83]. Robert Grover Brown. "Introduction to Random Signal Analysis and Kalman Filtering". Wiley. 1983.
- [BRO91]. Dana H. Brooks, Chrisostomos L. Nikias, "Cross-Bicepstrum and Cross-Tricepstrum Approaches to Multichannel Deconvolution", International Signal Processing Workshop on Higher Order Statistics. Chamrousse-France, July, 10/12, 1992.
- [CAS93] Luís Castedo, Aníbal Figueiras-Vidal, "Blind Equalization and Interference Suppression" based on Cyclostationary Signal Properties", Conf. on Adaptive Methods and Emergent Techniques for Signal Processing and Communications, Universidade de Vigo, 1993.
- [Che91] Yuanjie Chen, Chrysostomos I. Nikias, John G. Proakis, "CRIMNO: Criterion with Memory non-Linearity for Blind Equalization", International Signal Processing Workshop on Higher Order Statistics, Chamrousee, France, July, 10/12, 1992.
- [Cid93] "Recurrent Radial Basis Functions Networks for Optimal Symbol by Symbol Equalization", Conf. on Adaptive Methods and Emergent Techniques for Signal Processing and Communications, Universidade de Vigo, 1993.
- [Com91] Pierre Common, "Independent Component Analysis", International Signal Processing Workshop on Higher Order Statistics". Chamrousse, France, July, 10/12, 1992.
- [Com93] Pierre Comon, Jean-Louis Lacoume, "Statistiques d'Ordres Superieurs pour le Traitement du Signal". Fascicule III, Les Houches, école predoctorale de physique. August 28th-September 10th, 1993.
- [FRA80]. L. E. Franks. "Carrier and Bit Synchronization in Data Communications- A Tutorial Review". IEEE Transactions on Communications VOL-28 NO-8 pp. 1107-1121. August 1980.
- [GAG78]. Robert M. Gagliardi. "Introduction to Communications Engineering". Wiley Interscience 1978.
- [GARF86]. Floyd M. Gardner. "A BPSK/QPSK Timing-Error Detector for Sampled Receivers". IEEE Transactions on Communications VOL-34 NO-5 May 1986.
- [GARF88] Floyd M. Gardner. "Demodulator Reference Recovery Techniques Suited For Digital Implementation". European Space Agency. Final Report. ESTEC Contract No. 6847/86/NL/DG. 1988.

- [GARW85]. William A. Gardner. "Introduction to Random Processes with Applications to Signals and Systems". New York: Macmillan. 1985.
- [GARW86]. William A. Gardner. "The Role of Spectral Correlation in Design and Performance Analysis of Synchronizers". IEEE Transactions on Communications VOL-34 NO-8 November 1986.
- [GARW91]. William A. Gardner. "Exploitation of Spectral Redundancy in Cyclostationary Signals". IEEE SP Magazine. April 1991.
- [GIT92]. Richard D. Gitlin, Jeremiah F. Hayes, Stephen B. Weinstein. "Data Communications Principles". Plenum Publishing Corporation. 1992.
- [HAY91]. Simon Haykin. "Adaptive Filter Theory". Prentice-Hall. 1991.
- [HON93] M.Honig, U.Mazhow, "MMSE Linear Interference Suppression for Multi-User Communications", Conf. on Adaptive Methods and Emergent Techniques for Signal Processing and Communications, Universidade de Vigo, 1993.
- [IZZ89]. Luciano Izzo, Luigi Paura, Giovanni Poggi. "Multiple Source Localization,: a new method exploiting the cyclostationarity property". Douzieme Colloque GRETSI - Juan les Pins. 12-16 June 1989.
- [IZZ92]. Luciano Izzo, Luigi Paura, Giovanni Poggi. "An Interference-Tolerant Algorithm for Localization of Cyclostationary Signal Sources". IEEE Transactions on Signal Processing Vol-40 NO-7 pp. 1682-1686. July 1992.
- [JER92]. Michel C. Jeruchim, Philip Balaban, K. Sam Shanmugan. "Simulation of Communication Systems. Applications of Communications Theory". 1992.
- [KAJ93] A. Kajiwarra, "Mobile Satellite CDMA System Robust to Doppler Shift", IEEE International Conference on Communications, ICC'93, Geneva May, 23-26,1993.
- [KAM91] K.D. Kammeyer, B.Jelonek, "A Cumulant Zero-Matching Method for the Blind System Identification", International Signal Processing Workshop on Higher Order Statistics, Chamrousse, France, July, 10/12, 1992.
- [KIR83] S. Kirkpatrick, C.D. Jr. Gelatt, M. P. Vecchi. "Optimization by Simulated Annealing". Science, 220. pp. 671-680. 1983.
- [KOH90] R. Kohno, H.Imai, M. Hatori, S. Pasupathy, "Combinations of an Adaptive Array Antenna and a Canceller of Interference for Direct Sequence Spread-Spectrum Multiple-Access Systems". IEEE Journal on Selected Areas in Communications, VOL.8,No.4,1990.
- [LAG91] M.A. Lagunas, G. Vázquez, "Array Processing from Third Order Functions", International Signal Processing workshop on Higher Order Statistics. Chamrousse-France, July, 10/12, 1992.
- [LAG93] M.A.Lagunas, G. Vázquez, "Adaptive Arrays in Communications", Conf. on Adaptive Methods and Emergent Techniques for Signal Processing and Communications, Universidade de Vigo, 1993.
- [LUP90] Ruxandra Lupas, Sergio Verdú, "Near-Far Resistance of Multi-User Detectors in

- Asynchronous Channels", IEEE Transactions on Communications, VOL. 38, pp-496-508, April 1990.
- [MEY90]. Heinrich Meyr, Gerd Ascheid. "Synchronization in Digital Communication". Volume 1. John Wiley & Sons. 1990.
- [NIK91] Chrysostomos L. Nikias, "Blind Deconvolution Using Higher-Order Statistics", International Signal Processing Workshop on Higher Order Statistics, Chamrousse-France, July, 10/12, 1992.
- [NOW93] Steven J. Nowlan, Geoffrey Hinton, "A Soft-Decision Directed LMS Algorithm for Blind Equalization". IEEE Transactions on Communications, VOL. 41, No.2, February 1993.
- [OER88]. Martin Oerder, Heinrich Meyr. "Digital Filter and Square Timing Recovery". IEEE Transactions on Communications VOL-36 NO-5 May 1988.
- [PAP84] Athanasios Papoulis. "Probability, Random Variables and Stochastic Processes". Second Edition. McGraw-Hill. 1984.
- [RIB92a]. Jaume Riba, Gregori Vázquez. "Descripciones Tiempo-Frecuencia en la Recuperación Conjunta de Parámetros en Modulaciones Digitales". URSI 1992.
- [RIB92b]. Jaume Riba, Gregori Vázquez, et al. "Parallel Processing for Joint Maximum Likelihood Synchronisers at High Data Rates". Third International Workshop on Digital Signal Processing Techniques Applied to Space Communications. September 1992.
- [RIB93]. Jaume Riba, Gregori Vázquez, "Adaptive Joint Frequency & Timing Estimation Exploiting The Cyclostationarity Property." Adaptive Methods and Emergent Techniques For Signal Processing and Communications. June 1993., Bayona., Spain. Cost #229WG.1+2 Second Vigo Workshop. Special issue of the SIGNAL PROCESSING Journal.
- [RIB94] "Recursive Bayes Risk Parameters Estimation From The Cyclic Autocorrelation Matrix" Jaume Riba, Gregori Vázquez. ICASSP 94.
- [SAL94a] "Separation of Digital Communication Signals through Joint Space-Time Decorrelation". Josep Sala, Gregori Vázquez. Proceedings of ICASSP'94 (International Conference on Acoustics, Speech and Signal Processing), Oral Session. Adelaide, Austràlia. 19-22 Abril 1994.
- [SAL94b] "Adaptive Blind Equalization and Demodulation Without Channel and Signal Parameter Extraction". Josep Sala, Gregori Vázquez. Proceedings of EUSIPCO'94 (European Signal Processing Conference), Oral Session. Edimburg, Escòcia. 13-16 Setembre 1994.
- [SAL94c] "Mobile User Tracking Algorithms for Adaptive Arrays using Joint Beamforming and Equalization Techniques". Josep Sala, Gregori Vázquez. Proceedings of the RACE Mobile Telecommunications Workshop. Amsterdam, 17-19 Maig 1994.
- [SAL94d] "Global Beamforming and Demodulation for Linearly Modulated Signals". Josep Sala, Gregori Vázquez. Proceedings of SEM'94, Seventh SP Workshop on Statistical Signal and Array Processing. Quebec, 26-29 Juny, 1994.
- [SAL94e] "Adaptive Suppression of Co-Channel Interference in Digital Multi-User

Schemes". Josep Sala, Gregori Vázquez, Miguel Angel Rey. Proceedings of ESA's Fourth Workshop on Digital Signal Processing Techniques Applied to Space Communications. King's College, Londres, Sept. 1994.

[SAL94f] "A Blind Equalization Algorithm Based On A Statistical Reference of the Signal Constellation". Josep Sala, Gregori Vázquez. Proceedings of COST #229 Final Joint Vigo Workshop. 19-21 Octubre 1994

[SAL94g] "A Cost Function for Blind Signal Recovery Based on an Implicit Cumulant Expansion". Josep Sala, Gregori Vázquez. Joint IEEE Workshop on Higher Order Statistics. Begur, Girona, Juny 1995.

[SAL94h] "Arrayed Decorrelation of DS-CDMA Signals" submitted to "IEEE Transactions on Information Theory" Josep Sala, Gregori Vázquez .

[SCH90]. Scharf. "Statistical Signal Processing; Detection, Estimation and Time Series Analysis". 1990.

[VAZ91]. Gregori Vázquez, Josep Sala, Jaume Riba, Miguel A. Lagunas, Juan Fernandez. "Advanced Demodulator Using Parallel Signal Processing Techniques". Theoretical Report. Work erformed for ALCATEL ESPACIO under Contract No.: ESA/ESTEC-9246/90/NL/US(SC). December 1991.

[VAZ92] "Zero-Error Detection Algorithms and Parallel Processing for Frequency and Timing Estimation in Digital Communications" Gregori Vázquez, Jaume Riba, Meritxell Lamarca, Josep Sala. Proceedings of the Third International Workshop on Digital Signal Processing Techniques Applied to Space Communications. ESTEC, Noordwijk, The Netherlands. 23-25 Setembre 1992.

[VER92] Sergio Verdú, "Multi-User Demodulation", Proceedings of the Third International Workshop on Digital Signal Processing Techniques Applied to Space Communications. ESTEC, Noordwijk, the Netherlands, Session 3, pp. 18-36, 23-25 September 1992.

[XU92] Guanghai Xu, Thomas Kailath, "Direction of Arrival Estimation via Exploitation of Cyclostationarity - A combination of Temporal and Spatial Processing", IEEE Transactions on Signal Processing, VOL. 40, No. 7, July 1992.

[Zhe 91] Fu-Chung Zheng, Stephen McLaughlin, Bernard Mulgrew, "2nd and 4th Order Cumulant based Blind Equalization Algorithm for Non-Minimum Phase Channels", International Signal Processing Workshop on HOS. Chamrousse-France, July, 10/12, 1992.