

**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*



*El que no mata, engreixa*  
*Aforisme Popular*



## INDICE DE CONTENIDO

- 1**    **Introducción**
- 2**    **Criterios de Referencia Espacial y Temporal**
  - 2.1    El Criterio de Máxima Verosimilitud (ML).
  - 2.2    Solución de Wiener y Solución ML.
  - 2.3    Características de la Solución de Wiener.
  - 2.4    Eliminación de Interferencias.
  - 2.5    Espacio Estadístico Normado.
  - 2.6    Decorrelación en Array.
    - 2.6.1    Introducción.
    - 2.6.2    Modelo de Señal.
    - 2.6.3    El ADD.
    - 2.6.4    Análisis de Prestaciones
    - 2.6.5    El Caso Límite. Probabilidad de Error Asintótica.
    - 2.6.6    Apéndice A. Criterios de Detección.
    - 2.6.7    Apéndice B. Ruido Espacialmente Coloreado.
    - 2.6.8    Apéndice C. Ruido Temporal y Espacialmente Coloreado.
  - 2.7    Decorrelación síncrona.
    - 2.7.1    Decorrelación Asistida (I).
    - 2.7.2    Decorrelación Asistida (II).
  - 2.8    Ecuación.
    - 2.8.1    Modelo de la Señal.
    - 2.8.2    Ecuación óptimo.
    - 2.8.3    Detección óptima.
    - 2.8.4    Ecuación óptimo en array.
  - 2.9    Modelado y Eliminación de Interferencias.
- 3**    **Criterios de Referencia Estadística. Primera Función de Coste.**
  - 3.1    Introducción.
  - 3.2    Una Función de Coste basada en Referencia Estadística
  - 3.3.    Modelo de Señal.
    - 3.3.1    Modelo de una Fuente.
    - 3.3.2    Modelo de Varias Fuentes.
  - 3.4    Función de Coste.
    - 3.4.1    Función de Coste Condicionada.
      - 3.4.1.1    Relación entre la Función de Coste Condicionada y no Condicionada.

### 3.5 Algoritmo de Gradiente basado en la Referencia Estadística.

3.5.1 Obtención del Gradiente de la Función de Coste.

3.5.2 Regeneración de la Referencia.

3.5.3 Función de Regeneración.

3.5.3.1 Contracción de la constelación.

3.5.4 Estimación de la Varianza del Ruido en el Modelo.

### 3.6 Ejemplos de Aplicación.

3.6.1 Señal de Envolvente Constante.

3.6.2 Señales Multi-amplitud.

3.6.3 Señal QPSK.

### 3.7 Resumen del Algoritmo de Referencia Estadística. Primera Función de Coste.

### 3.8 Conclusiones

## 4 Estudio de la Función de Coste.

### 4.1 Introducción.

4.1.1 Modelo de una Fuente.

4.1.2 Modelo de Varias Fuentes.

4.1.3 Hessiano de la Función de Coste.

4.1.4 Expansión de la Función de Coste Mediante Cumulantes.

### 4.2 Modelo de una Fuente. Introducción.

4.2.1 Propuesta de una Función de Coste.

4.2.2 Mínimos Locales de la Función de Coste.

4.2.3 Mínimo Absoluto de la Función de Coste.

4.2.4 Obtención del Parámetro Varianza.

### 4.3 Modelo de Varias Fuentes.

4.3.1 Equivalencia con una Referencia Conocida.

### 4.4 Hessiano de la Función de Coste.

4.4.1 Otras arquitecturas.

### 4.5 Expansión de la Función de Coste Mediante Cumulantes.

## 5 Segunda Función de Coste

### 5.1 Segunda Función de Coste.

5.1.1 Ventanas de Estimación de Densidad de Probabilidad.

5.1.2 Segunda Función de Coste. Propiedades.

5.1.3 Efecto del Ruido.

5.1.4 Gradiente de  $J_2$ . Propiedades.

- 5.1.5 Estimación del Gradiente de  $J_2$ .
- 5.2 Función de Coste Extendida.
  - 5.2.1 Aplicación de las funciones  $q$  y  $q_2$  a Detección.
  - 5.2.2 Codificación Correlativa. El sistema Duobinario.
  - 5.2.3 Aplicación a la Decorrelación de Señales.
  - 5.2.4 Aceleración de la Convergencia.
- 5.3 Resumen del Algoritmo de Referencia Estadística. Segunda Función de Coste.
- 6 Arquitecturas de Ecuación.**
  - 6.1 Introducción.
  - 6.2 Ecuación complejo no restringido.
  - 6.3 Ecuación real no restringido.
  - 6.4 Ecuación real restringido.
  - 6.5 Conformación de Haz y Ecuación.
- 7 Simulaciones**
  - 7.1 Introducción.
  - 7.2 Ecuación.
  - 7.3 Decorrelación Adaptativa.
  - 7.4 Conformación de Haz.
- 8 Apéndices.**
  - 8.1 Propiedades Estadísticas de los Operadores No-Lineales.
    - 8.1.1 El Operador Esperanza.
    - 8.1.2 Esperanza de una No-Linealidad.
  - 8.2 Variables Aleatorias Complejas.
  - 8.3 Criterio MMSE en el dominio  $z$ .
  - 8.4 Cálculo de  $I_A$ .
- 9 Conclusiones**
  - 9.1 Futuras Líneas de Investigación.
- 10 Bibliografía.**



**Capítulo I**  
**Introducción**

---



## 1.1 Introducción

En esta tesis hemos querido considerar básicamente el diseño de *funciones de coste* para *adaptación ciega* en un amplio margen de aplicaciones en el ámbito de las comunicaciones digitales. Qué pretendemos significar mediante los términos "adaptación" y "ciega", y mediante "funciones de coste" ?. Lo ilustraremos brevemente con unos ejemplos, los cuales constituirán el campo de aplicación en el que validaremos los resultados teóricos de este trabajo. Consideremos los siguientes problemas en comunicaciones digitales,

El problema de ecualización: la máxima capacidad a la que podemos transmitir información a través de un canal viene limitada por las características de éste, en términos de dos parámetros básicos como demostró Shannon: el ancho de banda del canal y la relación señal a ruido en recepción. La calidad o probabilidad de error con la que podemos recuperar o demodular los datos transmitidos no dependen únicamente de los dos parámetros anteriormente mencionados, sino que vienen determinados por el conocimiento que tengamos de la naturaleza del canal. Es decir, en multitud de situaciones reales, dicho conocimiento es, con suerte, parcial, excepto en canales prácticamente ideales y de gran ancho de banda, como los sistemas de transmisión fija por satélite. En el resto de los casos, la respuesta impulsional del canal introduce una distorsión desconocida por el receptor en la señal que envía el transmisor. En el caso de modulaciones lineales, donde enviamos secuencias de un mismo pulso, modulado por los símbolos de la información, este pulso, llamado de conformación o de "shaping", se convoluciona con la respuesta impulsional del canal, variando su forma. El filtro que utiliza normalmente un demodulador, adaptado a la forma conocida del pulso de shaping, filtra el ruido de forma óptima permitiendo que la detección se realice con mínima probabilidad de error. El efecto del canal distorsiona la salida del filtro adaptado en los instantes óptimos de muestreo, introduciendo interferencia de símbolos adyacentes, o interferencia intersimbólica.

La velocidad a la que puede eliminar el efecto de las interferencias es sumamente importante para garantizar una pérdida de información mínima y optimizar el flujo neto de transmisión. Por ello, un método sencillo de adaptación consiste en introducir en el canal de transmisión secuencias de símbolos, llamadas secuencias de entrenamiento, conocidas previamente por el demodulador. El algoritmo de ecualización puede así utilizar ese conocimiento para identificar el sistema y eliminar la interferencia intersimbólica. No obstante, interesa reducir el porcentaje de tiempo que se toma para transmitir estas secuencias, puesto que la capacidad de información neta del enlace disminuye. Es entonces

donde aparece el concepto de ecualización ciega como una forma de garantizar la eliminación de la interferencia intersimbólica en ausencia de referencias sobre los datos transmitidos o sobre las características del canal.

Generalmente, los métodos de ecualización ciega adolecen de una velocidad de convergencia considerablemente más lenta que los de ecualización asistida. No obstante, una vez en seguimiento, permiten ir actualizando los coeficientes del filtro de ecualización en función de las características del canal. Estas técnicas permiten aumentar el flujo de información al no ser ya necesario el envío periódico de las secuencias de entrenamiento (no obstante, estas siempre aparecen en sistemas de comunicación reales por la necesidad de "formatear" u organizar la información que se transmite por el canal). El tema de ecualización ciega es ya un tema maduro, con un gran número de publicaciones. Prosigue sin embargo la investigación en este campo debido a la increíble variedad de soluciones que ofrece. Algunas de estas soluciones constituirán resultados de esta tesis.

El problema de decorrelación. Sistemas DS-CDMA: como habíamos citado anteriormente, la capacidad de un canal viene limitada por el ancho de banda y la relación señal a ruido. Otro factor a tener en cuenta en sistemas de comunicaciones más sofisticados es el número de usuarios simultáneamente presentes en el mismo canal. Este hecho, debido básicamente a la interferencia de unos usuarios sobre otros, o a la necesidad de reducir tal interferencia, al máximo, limita la capacidad del canal referida a un usuario particular. Los sistemas CDMA o "Code Division Multiple Access" constituyen una forma de compartición simultánea del canal entre varios usuarios. Los datos provenientes de un usuario particular pueden recuperarse mediante conocimiento de la secuencia pseudo-aleatoria, o de "spread", de donde proviene el término "Direct Sequence", que se ha utilizado para ensanchar el ancho de banda de la información y poder así compartir el canal con interferencia mínima del resto de usuarios activos en ese momento.

En un entorno real de comunicaciones por espectro ensanchado y acceso múltiple, no siempre puede garantizarse una mínima interferencia del resto de los usuarios del canal. De hecho, el grado en que un usuario afecta a los restantes dependerá de dos factores: (a) de la potencia en recepción del usuario considerado y (b) de los valores de correlación cruzada de la secuencia PN del usuario con los interferidos. En recepción, cada usuario se recupera mediante un filtro adaptado a su secuencia. La presencia de otros usuarios introducirá interferencia a la salida de este filtro en el instante óptimo de muestreo, degradando el proceso de detección y aumentando la probabilidad de error.

Entendemos por decorrelación de usuarios en un sistema CDMA el proceso mediante el cual, tomando las señales a la salida de los filtros adaptados de cada usuario, podemos eliminar las interferencias de unos usuarios sobre otros y proporcionar una señal limpia a los algoritmos de detección. Decimos entonces que se ha decorrelado la interferencia de múltiple acceso. El problema que surge en este instante es qué algoritmo utilizar para garantizar el proceso de decorrelación. Veremos que las funciones de coste que proponemos en este trabajo son aplicables también a la decorrelación ciega de señales CDMA. El problema de decorrelación es en cierta forma análogo al de ecualización, donde intentamos eliminar no la interferencia de otros usuarios, sino la de otros símbolos.

El problema de conformación de haz: hemos visto en los dos problemas anteriores como la capacidad de un canal está limitada por el ancho de banda, la relación señal a ruido y la presencia de usuarios simultáneos. En el problema de conformación de haz tenemos también usuarios simultáneos, aunque la separación de éstos no se realiza en base a una secuencia de spread, sino en base a una dimensión adicional a la del tiempo: la dimensión de espacio. En este caso entendemos que el sistema receptor está formado por una agrupación de antenas. Es decir, nos vemos obligados no a procesar una señal escalar, sino una señal vectorial. Cada uno de los usuarios del canal realiza el acceso en base a una técnica conocida como SDMA o "Space Division Multiple Access". En esta, cada usuario incide sobre el array desde una dirección espacial diferente, de forma que el array receptor puede separar en ángulo a cada uno de los usuarios. No siempre este concepto de separación angular resulta apropiado para todas las situaciones. En otros casos sustituimos el término conformación de haz o "beamforming" por el término combinación óptima de señales u "Optimal Combining".

El proceso de separación de usuarios y de filtrado del ruido espacial o eliminación de interferencias se realiza en base al proceso de beamforming, en que se realiza un producto escalar entre el "snapshot", es decir, el vector que contiene muestras de señal de cada sensor en un determinado instante de tiempo, y el vector de coeficientes correspondiente a cada usuario. El vector de coeficientes o la dirección de llegada de un usuario puede conocerse a priori en algunos casos. Hablamos entonces de Referencia Espacial, en contraste al concepto de Referencia Temporal, en el que conocíamos la secuencia de llegada.

El problema de conformación de haz mediante algoritmos ciegos constituye un problema muy interesante de procesado adaptativo. Consideremos un ejemplo sencillo de un escenario: un receptor formado por cierto número de sensores que recibe señales provenientes de varios usuarios. No disponemos de ningún tipo de referencia espacial de

ninguno de los usuarios y carecemos también de las secuencias de entrenamiento o referencia temporal que nos posibilitarían su recuperación. Como debemos proceder para recuperar cada uno de ellos ?, y si nos interesa únicamente una señal de entre las posibles, por ejemplo, una señal con modulación QPSK, cómo deberemos actuar?. Aunque existen técnicas de estimación espectral que podrían aplicarse para solucionar este problema, como MUSIC, su implementación resulta costosa y no se consideran como técnicas de beamforming. Por ejemplo, no nos permitiría distinguir si una posible señal presente en un pico del espectro es QPSK o bien CPM. Es decir, pretendemos diseñar un conformador de haz que adaptativamente sepa discernir entre señales de estadísticas diferentes.

En los tres ejemplos anteriores hemos ido introduciendo conceptos básicos de procesado adaptativo. Aquellos que nos interesa destacar especialmente son los criterios de Referencia Espacial y Referencia Temporal sobre los que se basa en diseño de funciones de coste, normalmente de tipo error cuadrático medio, para algoritmos de adaptación. La idea motriz de esta tesis consiste en añadir un tercer criterio a los dos anteriormente citados: el criterio de **Referencia Estadística**. Es decir, en la recuperación de señales no utilizamos información de una realización particular de la señal (referencia temporal) o de la estructura del modelo de señal (referencia espacial) sino que tomamos como referencia la distribución de probabilidad de la señal que deseamos recuperar, que suponemos, naturalmente, conocida. Este criterio es especialmente robusto en el ejemplo anterior donde deseábamos recuperar una señal QPSK de entre todas las incidentes en el array. Incluso, la recuperación es posible en presencia de interferencias más potentes que la misma señal deseada. Otro ejemplo como el de decorrelación de señales CDMA, presenta un cariz algo distinto. Es decir, en este caso las señales de todos los usuarios tienen la misma estadística. Como se recupera entonces cada uno de ellos y se evita que el usuario más potente. con la distribución de probabilidad deseada, p.ej. BPSK, aparezca en todas las salidas del decorrelador?. En este caso debemos utilizar un criterio de Referencia Estadística que utilice información sobre la función de densidad de probabilidad conjunta de todos los usuarios, es decir, que implícitamente incorpore la mayor independencia estadística entre las salidas del decorrelador, dos a dos. Incluso, criterios de este tipo aplicados a algoritmos de equalización, pueden permitirnos utilizar información del código con que se modulan los datos.

El objetivo de esta tesis es pues enfocar el problema del filtrado y demodulación adaptativas desde el punto de vista de teoría de la información. Se demostrará que la disponibilidad de información lateral sobre las funciones de densidad de probabilidad de las señales deseadas puede mejorar las prestaciones de los sistemas adaptativos. La principal

aplicación que se ha tenido en mente es la demodulación de señales de comunicación digitales (tanto para demoduladores de sensor único como para arrays). entendemos que no sólo el filtrado de por sí (ecualización espacial y temporal), sino también la estimación adaptativa de todos los parámetros (fase de la portadora y sincronismo) relevantes al proceso de la demodulación se engloban en el concepto de filtrado adaptativo. Mostraremos que ambos conceptos se interrelacionan de una forma natural al considerarlos desde el enfoque aquí expuesto. Los resultados teóricos se validarán para unos cuantos escenarios predefinidos.

En el contexto de comunicaciones, el diseño de un demodulador que cumpla unos criterios de optimalidad necesita de información a priori sobre la estadística de la señal recibida,  $r(t)$ . Cuando quiera que la estadística de dicha señal se desvíe de las premisas básicas, el demodulador deja de ser óptimo en el sentido descrito. Una forma de atacar este problema consiste en estimar el conjunto de parámetros (estadísticos) que caracterizan el proceso  $r(t)$  para controlar el vector de estado del demodulador. Podemos afirmar que el comportamiento de tal demodulador adaptativo será dependiente de los datos. Es por lo tanto necesario diseñar algoritmos que a partir de los datos conduzcan el demodulador a su configuración óptima ( la que maximiza el criterio de optimalidad cuando todos los parámetros relevantes sobre la estadística de la señal y del ruido son perfectamente conocidos). La elección de un criterio particular y/o de una arquitectura de demodulación es un aspecto determinante de las prestaciones del sistema adaptativo final.

## 1.2 Perspectiva

En este apartado exponemos una perspectiva de la literatura existente en este campo. En los últimos veinte años han aparecido en la literatura una serie de algoritmos adaptativos y de estimación de parámetros. Los campos de aplicación se extienden desde RADAR y Comunicaciones a Bioingeniería y Procesado de Datos Sísmicos. Como regla general en el diseño de un sistema adaptativo, uno intenta beneficiarse del conocimiento a priori de alguna característica o propiedad estadística de las señales. El uso de este conocimiento nos permite deducir un criterio de optimalidad o función de coste. El sistema adaptativo evoluciona con el fin de minimizar esta función de coste. Multitud de criterios se han propuesto hasta la actualidad. Podemos citar los basados en Estadísticos de Orden Superior (HOS), Procesado de Señales Cicloestacionarias (CSSP), Redes Neuronales (NN), Métodos de Bussgan etc.

La utilización de HOS es de fundamental importancia por sus propiedades de incrementar la relación señal a ruido cuando la señal de interés esta inmersa en ruido Gaussiano, y por el hecho de que conserva la información de fase. Una propiedad básica es que todos los cumulantes de orden superior ( $n > 2$ ) de un proceso Gaussiano son nulos. Este hecho ha encontrado amplia aplicación en ecualización de canales de fase mixta y no-mínima ya que la respuesta impulsional del canal queda definida de forma no ambigua a partir de los cumulantes de orden superior de la señal recibida (suponiendo entrada no-Gausiana) [Zhe91][Sha91][Kam91]. Es de aplicación generalizada en algoritmos ciegos el uso de transformadas no lineales de los datos. Según [Nik91], los algoritmos de deconvolución ciega se pueden clasificar como del tipo Bussgang o Poliespectrales [Bro91] dependiendo de si la no-linealidad se sitúa a la salida o a la entrada del ecualizador. Se verá más adelante que el algoritmo aquí desarrollado se enmarca en la primera categoría. Los HOS también se han explotado [Lag91] en la estimación de direcciones de llegada (DOA) y conformación de haz a partir de snapshots de tercer orden. La ventaja de estas técnicas es que nos aseguran que sólo obtendremos las direcciones de llegada de aquellas fuentes correspondientes a procesos no-Gaussianos. La separación ciega de señales independientes constituye también un campo interesante de aplicación de los HOS [Com91][Com93]. El propósito de estos algoritmos es el de separar varias señales que han experimentado una mezcla instantánea o convolutiva sólo a partir del conocimiento de su independencia mutua. Es en este caso donde se utiliza el mínimo de información a priori sobre las estadísticas de las señales. En general, la aplicación de HOS no está libre de inconvenientes. Cabe destacar su elevado coste computacional y su lenta velocidad de convergencia.

El procesado de señales cicloestacionarias explota otra característica estadística de los procesos estocásticos, la periodicidad de los momentos y cumulantes de una señal. En estos casos, la función del sistema adaptativo es recuperar sólo aquellas señales que presenten una determinada ciclofrecuencia. Estos métodos han encontrado amplia aplicación en ecualización ciega puesto que la mayoría de las señales de comunicación presentan un carácter cicloestacionario. Un procedimiento posible es evaluar la correlación cíclica de la señal recibida de tal forma que señales de ciclofrecuencia diferente no contribuyan significativamente [Vaz92]. Una ventaja proporcionada por estos métodos es su mayor robustez frente a interferencias. El efecto de captura (recuperación de señales no deseadas) que presenta el algoritmo CMA, entre otros, se resuelve parcialmente [Cas93]. Las capacidades de rechazo de las interferencias que proporcionan estos algoritmos han encontrado rápida aplicación en procesado de arrays para anular interferencias espacialmente

localizadas en la etapa de conformación de haz [Izz89][Izz92], también vía procesado temporal [Xu92].

Los mejores resultados se obtienen cuando el sistema adaptativo se entrena con señales de referencia. El algoritmo intenta modificar el vector de estado del sistema para minimizar la función de error dada. Los métodos de conformación de haz con referencia temporal o espacial (TRB y SRB) constituyen técnicas aplicadas en terminales de comunicación avanzados [Lag93]. TRB es más robusto frente a problemas de calibración. En las técnicas de referencia temporal, una secuencia de datos previamente conocida por el receptor se envía por el canal. Las estimaciones de los parámetros del canal y de la señal son de mayor calidad que en otros métodos puesto que la incertidumbre de cuál es la secuencia de símbolos recibida es nula (adaptación asistida por los datos). Una vez que el demodulador se ha inicializado y empiezan a recibirse los símbolos de la información, diferentes técnicas pueden utilizarse para el seguimiento (aunque también es posible mantener la transmisión de una señal de referencia en un canal paralelo, ya sea en tiempo o en frecuencia). Los algoritmos pueden clasificarse en tres grandes grupos: DA (Data-Aided), DD (Decision Directed) o NDA (Non-Data-Aided).[Gar86]. Las técnicas DD son similares a TR en el sentido en que la señal de referencia está basada en decisiones de los símbolos recibidos y no en un conocimiento a priori. Presentan por tanto peor comportamiento, aunque hemos observado prestaciones excelentes de técnicas DD a bajos niveles de  $E_b/N_0$  en el diseño de demoduladores avanzados que utilizan información de los códigos convolucionales [Vaz91]. En cambio, en los algoritmos NDA no se realizan decisiones reales de los símbolos recibidos.

Para finalizar podemos decir que las prestaciones de un sistema adaptativo se encuentran en relación directa con la cantidad de información disponible de las señales recibidas. En la práctica habitual de comunicaciones, tenemos un buen conocimiento de la estadística de las señales recibidas (lo cual no ocurre en todas las aplicaciones) y utilizaremos este conocimiento para la recuperación de la señal deseada.

### 1.3 Resumen

El tema central de este trabajo es el diseño de funciones de coste para algoritmos adaptativos basados en criterios de Teoría de la Información. Dada una señal recibida, nos interesa recuperar la información transmitida en base a conocimientos apriorísticos del modelo de señal. Podemos aplicar diversos criterios, es decir, plantear la minimización equivalente de una función de coste determinada en función de los parámetros de un sistema que nos asegura una recuperación más o menos fiable de la información. El tipo de información previa de que dispongamos nos determinará qué enfoque debemos tomar. Básicamente este trabajo se ha estructurado en tres grandes partes. En la primera, compuesta por el capítulo II, consideramos varios ejemplos generales en los que disponemos de una información muy precisa del modelo de señal recibida y estudiamos cuales son los algoritmos (ML) de recuperación de la información. Como caso general del que se deduce por ejemplo el algoritmo de ecualización hemos considerado la decorrelación de señales CDMA en arrays. Aunque de un importante interés teórico en cuanto a los límites fundamentales de prestaciones, este tipo de algoritmos no están justificados para una implementación práctica donde desconocemos gran parte de la información supuestamente conocida en ellos. Esto nos conduce al diseño de algoritmos adaptativos en los que mediante una función de coste sensatamente escogida podemos estimar en línea los parámetros del sistema necesarios para la demodulación de los datos (implícitamente eliminación de interferencias, filtrado óptimo de ruido, ecualización,...). En la segunda parte, capítulos III, IV y V, proponemos dos funciones de coste y algunas variantes de ellas deducidas que posibilitan tales objetivos, realizando un estudio teórico exhaustivo de sus propiedades estadísticas. Mediante estas funciones de coste intentamos forzar una estadística determinada, la de los datos deseados, a la salida del sistema. Tal algoritmo se clasifica como ciego puesto que no utilizamos información ninguna de la secuencia real de datos transmitidos. Aunque como resultado importante cabe destacar el paralelismo existente con la solución de Wiener para sistemas con secuencias de entrenamiento conocidas. De hecho, a lo largo del trabajo se destacan las propiedades de las soluciones que minimizan el criterio propuesto. Dada la filosofía bajo la cual hemos derivado el criterio, hemos adoptado el nombre de Algoritmo de Referencia Estadística (SRA), en contraste con el concepto de Referencia Temporal o Espacial en procesamiento de arrays. En la tercera parte, capítulos VI y VII, aplicamos las funciones de coste deducidas en el capítulo anterior a un número de aplicaciones de interés como ecualización, conformación adaptativa de haz y decorrelación de señales, las cuales podemos ver que constituyen aspectos diferentes de un problema general. Vemos que el algoritmo presenta prestaciones dignas de tenerse en cuenta tanto en

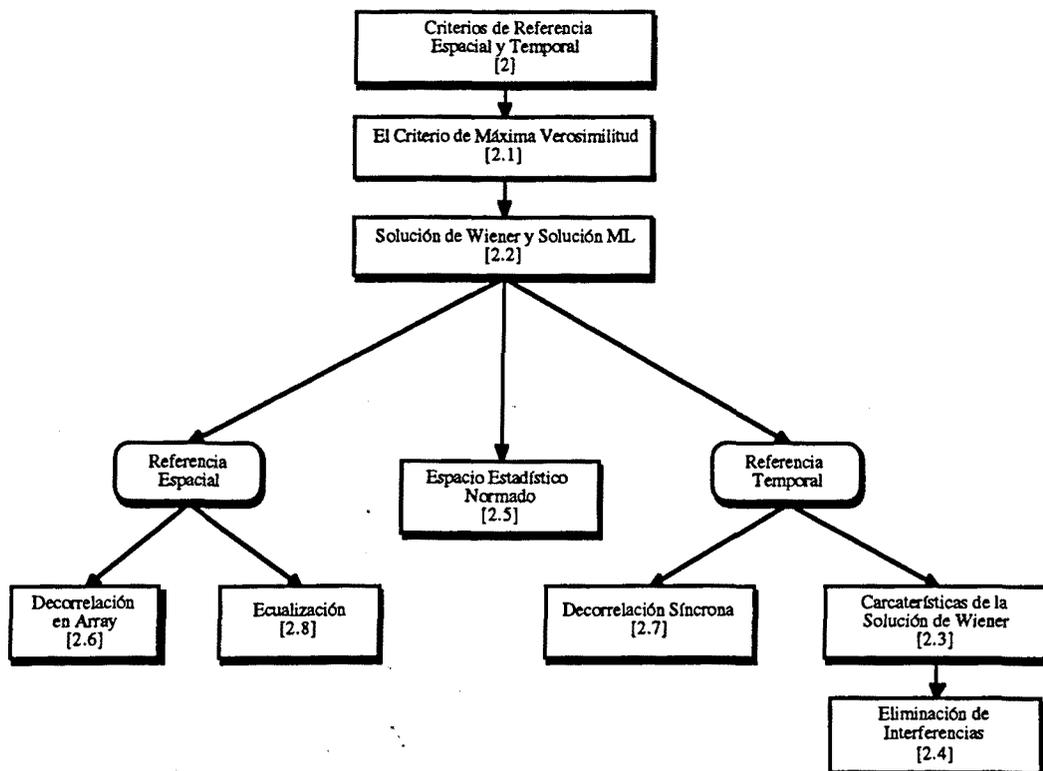
adquisición como en seguimiento. El tema de ecualización ciega ha suscitado multitud de trabajos durante muchos años y creemos que la perspectiva que se le ha dado en este trabajo contribuye aunque sea minimamente al conocimiento de estos problemas. Por tanto concluimos presentado algunas posibles vías de continuación de este trabajo, destacando la inclusión de la información sobre la codificación de los datos en los algoritmos de ecualización o conformación de haz.



## Capítulo II

### Criterios de Referencia Espacial y Temporal

---





## 2 Criterios de Referencia Espacial y Temporal

Dado un sistema de comunicaciones, estamos interesados en recuperar la información transmitida en el terminal receptor. No obstante siempre existe cierto desconocimiento de las características del canal que el receptor debe compensar o de las interferencias que debe eliminar. En secciones posteriores estudiaremos cómo puede solucionarse este inconveniente cuando desconocemos los datos transmitidos. Existen dos métodos ampliamente utilizados para el entrenamiento de un receptor adaptativo, basados en el tipo de conocimiento que se tenga de la señal recibida. En el primer método se supone que no existe incertidumbre en cuanto a la secuencia transmitida por la fuente deseada y que existe un conocimiento perfecto "a priori". Hablamos en este caso de métodos de **referencia temporal** o basados en secuencias de entrenamiento. En el segundo método se conoce la estructura de los datos, esto es, en un contexto de arrays, implicaría conocer la dirección de llegada de la señal deseada. Hablamos en este caso de métodos de **referencia espacial**. Supondremos siempre que la señal deseada se obtiene mediante una combinación lineal de los datos en base a un vector de coeficientes. En este estudio introductorio supondremos un modelo de varias fuentes y veremos cuál es el vector de coeficientes que minimiza el criterio deseado. En general, dados criterios diferentes, obtendremos diferentes vectores de coeficientes que minimizan la función de coste asociada, aunque con buenas propiedades si el criterio se ha elegido de forma correcta. En esta sección estudiamos dos criterios: veremos qué relación existe entre la solución de Wiener, la cual asociamos a métodos de referencia temporal, y la solución ML, asociada a métodos de referencia espacial. Podemos ver como ambos criterios (ver Capítulo 3) pueden reducirse mediante utilización de la función de coste desarrollada en este capítulo al criterio conocido de mínimos cuadrados, al estar condicionada la función de densidad por la secuencia de datos conocida.

Supondremos el modelo de señal definido por,

$$\mathbf{r}_k = \mathbf{S}_k \mathbf{a}_k + \mathbf{n}_k \stackrel{def}{=} \mathbf{d}_k + \mathbf{n}_k, \quad E_{\mathbf{n}_k} \mathbf{n}_k \mathbf{n}_k^H = \sigma_n^2 \mathbf{R}_{nn} \quad (2.1)$$

donde  $\mathbf{r}_k$  representa la señal recibida,  $\mathbf{S}_k$  la matriz de "steering" (supondremos que sus columnas forman un conjunto de vectores linealmente independientes),  $\mathbf{a}_k$  la secuencia vectorial de señales (consideraremos a todas las señales incorreladas dos a dos),  $\sigma_n^2 \mathbf{R}_{nn}$  la matriz de covarianza del ruido  $\mathbf{n}_k$ ,  $\mathbf{d}_k$  la componente de señal de  $\mathbf{r}_k$  o señal de referencia, obtenida como el producto  $\mathbf{S}_k \mathbf{a}_k$ . Consideraremos que  $\mathbf{R}_{nn}$  es una matriz fija, variando la potencia global del ruido en base al escalado de la varianza  $\sigma_n$ . Igualmente, supondremos

fija la matriz de steering y variaremos la potencia de las fuentes según las amplitudes de las componentes de  $\mathbf{a}_k$ .

## 2.1 El Criterio de Máxima Verosimilitud (ML)

Frecuentemente en Comunicaciones y en Procesado de Señal se necesita inferir el valor de determinados parámetros a partir de los datos recibidos (observaciones). El criterio de Máxima Verosimilitud utiliza una parametrización de la función de densidad de los parámetros dados los datos recibidos. El criterio consiste en escoger aquel parametro que maximice dicha función de densidad. Supongamos que un sistema produce un determinado vector de observaciones  $\mathbf{x}$  a partir de un vector de parámetros de entrada  $\mathbf{a}$ . Se denota el estimador ML como una función de los datos:  $\hat{\mathbf{a}}_{ML}(\mathbf{x})$ , donde,

$$\hat{\mathbf{a}}_{ML}(\mathbf{x}) = \arg \max_{\mathbf{a}} p_{\mathbf{a}|\mathbf{x}}(\mathbf{a}|\mathbf{x}) \quad (2.1.1)$$

En este trabajo hemos tenido en cuenta únicamente el modelo lineal en el cual la relación entre ambos vectores viene dada por  $\mathbf{x}=\mathbf{S}\mathbf{a}+\mathbf{n}$ , siendo  $\mathbf{S}$  una matriz de transformación y  $\mathbf{n}$  un término de ruido aditivo. Generalmente, la aplicación rigurosa de este principio conduce a estimadores con una gran carga computacional. Existen soluciones de compromiso subóptimas respecto al criterio ML pero ricas en interpretaciones y de complejidad abordable. La vía que hemos seguido es imponer restricciones en cuanto a la arquitectura del estimador, en particular, forzamos que éste sea lineal. El criterio ML restringido deviene entonces la siguiente maximización,

$$\hat{\mathbf{a}}_{LCML}(\mathbf{x}) = \arg \max_{\mathbf{w}} p_{\mathbf{a}|\mathbf{x}}(\mathbf{w}^H \mathbf{x}|\mathbf{x}) \quad (2.1.2)$$

Dada una estadística discreta para el vector de parámetros  $\mathbf{a}$  (usual en sistemas de comunicaciones donde éste representa los símbolos transmitidos), el principio ML equivale a una maximización exhaustiva de la función 2.1.1 sobre una rejilla multidimensional formada por todos los valores posibles de  $\mathbf{a}$  con una complejidad de crecimiento exponencial (medida en número necesario de operaciones) según el tamaño del alfabeto de  $\mathbf{a}$ . Si en cambio imponemos restricciones de tipo lineal en la arquitectura del estimador, su complejidad pasa a ser lineal en la dimensión de  $\mathbf{a}$ . Nótese que según la ecuación anterior, el término estimado  $\hat{\mathbf{a}}_{LCML}(\mathbf{x})$  puede ser una variable aleatoria de distribución continua, aún

siendo la distribución de  $\mathbf{a}$  discreta. Esto implica simplemente que debemos realizar una etapa posterior de detección a la ya realizada de filtrado mediante los coeficientes  $\mathbf{w}$ .

Un concepto importante en estimación lo constituye la suficiencia estadística. Intuitivamente, hablamos de estadísticos suficientes como de aquellos valores obtenidos a partir de una transformación de los datos sin pérdida de información respecto a los parámetros  $\mathbf{a}$  subyacentes, tales que dicha transformación equivale a una compresión de los datos. Es decir, eliminamos toda la redundancia necesaria para la estimación de los parámetros  $\mathbf{a}$ . Denominando  $\mathbf{b}$  al conjunto de estadísticos suficientes de  $\mathbf{a}$ , tenemos,  $\mathbf{b} = T(\mathbf{x})$ . Para el modelo lineal citado anteriormente tenemos que  $T(\cdot)$  es, también, una transformación lineal.

Veremos en el siguiente apartado como la aplicación de los conceptos aquí expuestos conduce a la especificación de dos criterios diferentes para la estimación de los parámetros  $\mathbf{a}$ . Introduciremos los conceptos de referencia espacial y temporal.

## 2.2 Solución de Wiener y Solución ML

En este apartado estudiaremos dos funciones de coste posibles para el filtrado óptimo de los datos del modelo de señal de la sección anterior: (a) en el criterio ML forzaremos un ajuste sobre los datos de entrada al sistema,  $\mathbf{r}_k$ , mientras que en (b), aplicaremos el criterio de mínimos cuadrados a la salida del sistema sobre una referencia conocida, y que conducirá a la conocida solución de Wiener. Ambos criterios se ilustran en la figura 2.2.1.

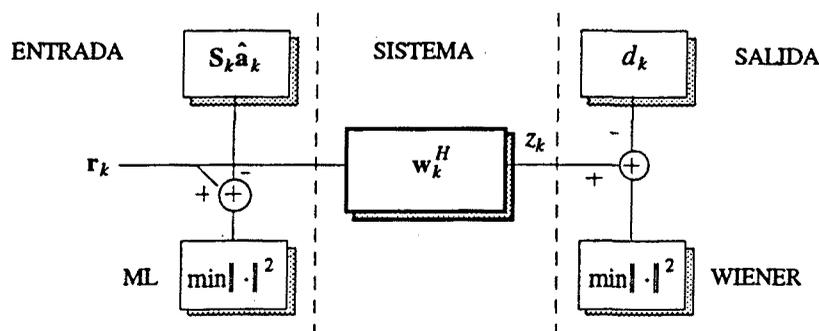


Figura 2.2.1: Representación de los criterios de diseño de un sistema para la recuperación de una señal deseada. En la parte superior de la figura aparece el modelo apriorístico de la señal: la parametrización a partir de los vectores de steering en el criterio ML y el conocimiento de la señal de referencia en el criterio de mínimo error cuadrático medio (Wiener).

En el criterio ML restringido planteamos la maximización de la densidad de probabilidad de los datos a la entrada condicionada por un conocimiento previo del número de fuentes y de las matrices de steering y de correlación del ruido, es decir, utilizamos una referencia espacial o información lateral sobre la estructura de los datos. Los argumentos de la maximización son las secuencias  $\mathbf{z}_k$  que mejor se ajustan al modelo de señal. Bajo condiciones de ruido Gaussiano, el criterio a maximizar se expresa como,

$$\max_{\mathbf{z}_k = \mathbf{W}_k^H \mathbf{r}_k} p_{\mathbf{r}_k | \mathbf{z}_k}(\mathbf{z}_k | \mathbf{r}_k) = \max_{\mathbf{z}_k = \mathbf{W}_k^H \mathbf{r}_k} \frac{1}{\pi^N \det \mathbf{R}_m} e^{-(\mathbf{r}_k - \mathbf{S}_k \mathbf{z}_k)^H \mathbf{R}_m^{-1} (\mathbf{r}_k - \mathbf{S}_k \mathbf{z}_k)} \quad (2.2.1)$$

donde la restricción aparece implícita como combinación lineal del vector de datos  $\mathbf{r}_k$ :  $\mathbf{z}_k = \mathbf{W}_k^H \mathbf{r}_k$ . Para la obtención de la secuencia  $\mathbf{z}_k$  en función de los datos de llegada, minimizamos la función de coste asociada, es decir, el argumento de la exponencial anterior, con respecto a la secuencia,

$$\nabla_{\mathbf{z}_k^H} J_{ML} = \mathbf{0} \Rightarrow \mathbf{z}_k = (\mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_m^{-1} \mathbf{r}_k \quad (2.2.2)$$

en donde aparece la expresión correcta de la matriz de coeficientes  $\mathbf{W}_k$  multiplicando al vector de datos. Puede demostrarse (Sección 2.5) que ésta equivale a una minimización de la potencia de ruido a la salida del sistema con restricciones introducidas por el conocimiento de los vectores de steering, puesto que estamos utilizando el término  $\mathbf{S}_k$ . Es decir, barreos el espacio de los datos  $\mathbf{r}_k$  en base a los componentes  $\mathbf{z}_k$  como  $\mathbf{S}_k \mathbf{z}_k$ .

En cambio, en un criterio de ajuste por mínimos cuadrados a la salida del sistema o de referencia temporal, se suponen conocidas las secuencias de cada fuente pero no los vectores de steering ni la correlación del ruido de entrada. Veremos que bajo estas condiciones, los vectores de coeficientes difieren bajo ambos criterios. Ahora, la matriz  $\mathbf{W}_k$  será aquella que mejor combine los datos de entrada para minimizar la siguiente función de coste, siendo  $\mathbf{d}_k$  la referencia conocida. De 2.1.1,

$$J_W = E(\mathbf{d}_k - \mathbf{z}_k)^H \mathbf{R}_{\eta\eta}^{-1} (\mathbf{d}_k - \mathbf{z}_k) + \ln(\pi^{N_u} \det \mathbf{R}_{\eta\eta}) \quad , \quad \mathbf{z}_k = \mathbf{W}_k^H \mathbf{r}_k \quad (2.2.3)$$

donde  $N_u$  denota al número de usuarios o dimensión del vector  $\mathbf{z}_k$  y  $\mathbf{R}_{\eta\eta}$  la matriz de peso del criterio. Para la obtención de la secuencia  $\mathbf{z}_k$  óptima, igualamos el gradiente a cero, llegando a la solución de Wiener, de la cual destacamos la no dependencia de la matriz de peso  $\mathbf{R}_{\eta\eta}$ ,

(2.2.4)

$$\nabla_{\mathbf{w}_k^H} J_W = 0 \Rightarrow \mathbf{W}_k = \mathbf{R}_r^{-1} E \mathbf{d}_k^H \mathbf{r}_k$$

Añadimos que el ajuste sobre una referencia a la salida no necesita implicar a todas las componentes de  $\mathbf{a}_k$ . Puede plantearse simplemente con respecto a un usuario como veremos en lo sucesivo. Demostraremos en los apartados siguientes que, en general, las soluciones proporcionadas por ambos criterios de referencia no coinciden excepto en el límite cuando la varianza del ruido tiende asintóticamente a cero. La solución ML es un estimador no polarizado de mínima varianza de cada una de las fuentes, lo cual no ocurre en el criterio de Wiener. Ésto es debido a que en este segundo criterio, al no disponer de información de los vectores de steering de las fuentes, el error cuadrático medio mínimo contempla tanto términos de ruido como residuos de señal. En este sentido, en modulaciones digitales, puede apreciarse claramente este efecto. La "contracción" que sufre la constelación de una modulación digital, refleja la polarización que introducen las técnicas adaptativas de mínimo error cuadrático medio en la posición de los respectivos símbolos y que demuestran el compromiso entre nivel de ruido y de señal residual (ISI). En la sección siguiente realizamos un estudio estadístico detallado de este efecto.

### 2.3 Características de la solución de Wiener.

El objetivo de este apartado será calcular el vector de coeficientes óptimo que minimiza una función de coste cuadrática del error con respecto a la referencia y examinar su comportamiento con respecto a la correlación del ruido y al número y potencia de las interferentes. Estableceremos con esto un patrón respecto al que comparar la aplicación de funciones de coste no lineales basadas en el conocimiento previo de la estadística de la señal deseada en capítulos subsiguientes. Supondremos conocida la secuencia transmitida por una de las fuentes,  $d_k$  (de potencia unidad) y consideraremos al resto como interferencias incorreladas con la fuente deseada. Definimos la función de coste de la forma habitual utilizando únicamente la referencia del usuario deseado,

(2.3.1)

$$J = \frac{1}{\sigma_o^2} E |d_k - \mathbf{w}_k^H \mathbf{r}_k|^2 + \ln \pi \sigma_o^2$$

obtenida como el logaritmo neperiano de la función de densidad condicionada a la referencia  $d_k$  conocida, donde  $\sigma_o$  representa la varianza del error de filtrado a la salida. Aplicando el gradiente e igualando a cero llegamos a la conocida solución de Wiener,

$$\nabla_{\mathbf{w}_k^H} J = -E(d_k - \mathbf{w}_k^H \mathbf{r}_k)^H \mathbf{r}_k = \mathbf{0} \Rightarrow \mathbf{w}_k = \mathbf{R}_{rr}^{-1} E d_k^* \mathbf{r}_k = \mathbf{R}_{rr}^{-1} \rho_{rd} \quad (2.3.2)$$

Vemos pues que el gradiente impone correlación nula (ortogonalidad) entre la señal de error y los datos  $\mathbf{r}_k$ . Si denominamos  $\epsilon_k$  a la señal de error, vemos que existe correlación no nula entre la señal no ruidosa  $\mathbf{S}_k \mathbf{a}_k$  y la señal de error,  $\epsilon_k$ ,

$$E \epsilon_k^* \mathbf{S}_k \mathbf{a}_k = -E \epsilon_k^* \mathbf{n}_k \neq \mathbf{0} \quad (2.3.3)$$

Es decir, pretendemos demostrar que el término de error  $\epsilon_k$  contiene términos residuales de señal y no únicamente términos del ruido de entrada al sistema,  $\mathbf{n}_k$ . Procederemos mediante reducción al absurdo. Actuando bajo el supuesto de que la señal de error sólo contiene ruido llegamos a una contradicción, puesto que en ese caso la correlación del término de la izquierda de la ecuación anterior debería ser cero por la condición de correlación nula entre señal,  $\mathbf{a}_k$ , y ruido,  $\mathbf{n}_k$ . Es decir, en la señal de error que minimiza el criterio aparece un término residual de señal deseada. Más adelante deduciremos el valor exacto de esta contribución.

Podemos demostrar esta afirmación de una forma más rigurosa. En efecto, supongamos en el caso simple de un único usuario que la señal de error sólo contiene términos de ruido y evaluemos las dos correlaciones cruzadas de la ecuación 2.3.3 (a este efecto definimos el error de "prueba" como  $\epsilon'_k$ ). Veremos que se llega a una contradicción respecto a la igualdad expuesta en 2.3.3, con lo cuál habremos demostrado que el error  $\epsilon_k$  contiene términos de  $\mathbf{a}_k$ . Tenemos entonces que,

$$\epsilon'_k \stackrel{def}{=} \mathbf{w}_k^H \mathbf{n}_k = \rho_{rd}^H \mathbf{R}_{rr}^{-1} \mathbf{n}_k \Rightarrow E \epsilon_k^* \mathbf{n}_k = E \mathbf{n}_k \mathbf{n}_k^H \mathbf{R}_{rr}^{-1} \rho_{rd} = \mathbf{R}_{nn} (\mathbf{R}_{rr}^{-1} \rho_{rd}) \quad (2.3.4)$$

Donde la correlación cruzada entre el error de prueba y el ruido de entrada es igual al vector de Wiener multiplicado por la matriz de correlación del ruido de entrada. Ahora bien, si la matriz  $\mathbf{R}_{rr}$  es de rango completo (lo es siempre que la matriz de correlación del ruido lo sea), también lo es su inversa, por consiguiente, el Kernel o subespacio nulo de la aplicación lineal consistente en multiplicar por la inversa de  $\mathbf{R}_{rr}$  es nulo (contiene únicamente el vector todo ceros). Es decir, el vector  $\mathbf{w}_k$  es no nulo siempre que exista correlación entre la

secuencia de referencia y los datos. Puesto que suponemos que  $\mathbf{R}_{nn}$  es de rango completo, utilizando el mismo razonamiento llegamos a la conclusión que la correlación cruzada entre  $\varepsilon'_k$  y  $\mathbf{n}_k$  no puede ser nula, es decir, igual al vector todo ceros. En segundo lugar, considerando el término de la izquierda de la ecuación 2.3.3, la correlación cruzada entre  $\varepsilon'_k$  y los datos no ruidosos  $\mathbf{S}_k \mathbf{a}_k$ , llegamos a que ésta sí debe ser cero. Evidentemente, esto es contradictorio con la premisa inicial de que el error de prueba  $\varepsilon'_k$  sólo contiene términos de ruido. Se concluye entonces que aparece un término residual de señal en el error  $\varepsilon_k$  proporcionado por la solución de Wiener.

Podemos ver cuál es la magnitud de la correlación cruzada entre la señal de error  $\varepsilon_k$  y la referencia del usuario. En efecto, sea la siguiente descomposición del vector de coeficientes que minimiza el criterio, definida como el escalado mediante  $\lambda$  de un vector  $\mathbf{w}_0$  paralelo a  $\mathbf{w}_k$ ,

$$\mathbf{w}_k = \lambda \mathbf{w}_0 \quad , \quad \mathbf{w}_0^H \mathbf{s}_k = 1 \quad \Rightarrow \quad z_k = \mathbf{w}_k^H \mathbf{r}_k = \lambda d_k + \lambda \mathbf{w}_0^H \mathbf{n}_k \quad (2.3.5)$$

donde  $\mathbf{s}_k$  es el vector de steering correspondiente a la señal que deseamos recuperar. Esta restricción del nuevo vector  $\mathbf{w}_0$  es válida puesto que sin ella no sería posible recuperar la señal de interés y nos especifica qué  $\mathbf{w}_0$  nos permite recuperar la señal  $\mathbf{a}_k$  con la amplitud correcta. La señal filtrada, denominada  $z_k$ , se obtiene entonces según la expresión anterior. Calculando entonces el error  $\varepsilon_k$  respecto a la referencia obtenemos,

$$\varepsilon_k = z_k - d_k = (\lambda - 1)d_k + \lambda \mathbf{w}_0^H \mathbf{n}_k \quad (2.3.6)$$

donde aparece una ligadura mediante el factor  $\lambda$  entre el término de referencia y el término de ruido. Veremos entonces que esta ligadura tiene como consecuencia la aparición del término residual de señal en el error  $\varepsilon_k$  de mínima potencia. Podemos hallar las expresiones exactas para  $\lambda$  y  $\mathbf{w}_0$  minimizando la siguiente función Lagrangiana,

$$\Lambda = E|\varepsilon_k|^2 - \lambda'(\mathbf{w}_0^H \mathbf{s}_k - 1) = |\lambda - 1|^2 E|d_k|^2 + |\lambda|^2 \mathbf{w}_0^H \mathbf{R}_m \mathbf{w}_0 - \lambda'(\mathbf{w}_0^H \mathbf{s}_k - 1) \quad (2.3.7)$$

donde el multiplicador de Lagrange viene dado por  $\lambda'$ . La minimización del Lagrangiano conduce a las siguientes expresiones para  $\lambda$  y  $\mathbf{w}_0$ ,

$$\nabla_{\lambda, \mathbf{w}_0} \Lambda = \mathbf{0} \quad \Rightarrow \quad \lambda = \frac{E|d_k|^2}{E|d_k|^2 + \mathbf{w}_0^H \mathbf{R}_m \mathbf{w}_0} \quad , \quad \mathbf{w}_0 = \frac{\lambda'}{|\lambda|^2} \mathbf{R}_m^{-1} \mathbf{s}_k \quad (2.3.8)$$

Forzando el cumplimiento de la restricción hallamos la expresión final para el multiplicador de Lagrange, el término de escalado  $\lambda$  y el término de filtrado  $w_0$ ,

$$w_0^H s_k = 1 \Rightarrow \lambda' = \frac{|\lambda|^2}{s_k^H R_{nn}^{-1} s_k}, \quad \lambda = \frac{E|d_k|^2}{E|d_k|^2 + s_k^H R_{nn}^{-1} s_k}, \quad w_0 = \frac{R_{nn}^{-1} s_k}{s_k^H R_{nn}^{-1} s_k} \quad (2.3.10)$$

Vemos pues que el vector de coeficientes que minimiza el criterio es simplemente un escalado (menor que la unidad) del vector de mínima varianza no polarizado (MVUE: "Minimum Variance Unbiased Estimator", el cuál coincide con el vector solución del criterio de referencia espacial o ML). Es decir, mediante este criterio imponemos una minimización de la potencia conjunta del bias o polarización (primer sumando de la ecuación 2.3.11) y del término de ruido de mínima varianza,  $\sigma_{MV}$ , ponderado mediante  $|\lambda|^2$  (segundo sumando de la ecuación 2.3.11), consiguiendo de este modo una varianza menor que la que se conseguiría en el caso de mínima varianza con la restricción  $\lambda=1$ .

$$E|\varepsilon_k|^2 = |\lambda - 1|^2 E|d_k|^2 + |\lambda|^2 w_0^H R_{nn} w_0 = |\lambda - 1|^2 \sigma_d^2 + |\lambda|^2 \sigma_{MV}^2 \quad (2.3.11)$$

El criterio de mínimo error cuadrático medio aplicado a la entrada del sistema (ecuación 2.2.1), o de referencia espacial, introducido anteriormente, es equivalente a un criterio de minimización de potencia de ruido con restricciones proporcionadas por el conocimiento de los vectores de steering. En efecto, supongamos la minimización de la siguiente potencia del error a la salida del sistema según el modelo de señal de la ecuación 2.1 con las siguientes restricciones de proyección sobre la matriz  $S_k$  (forzamos así que el criterio conduzca a un estimador no polarizado),

$$J(W_k) = E\|W_k^H n_k\|_2^2 = E n_k^H W_k W_k^H n_k, \quad W_k^H S_k = I \quad (2.3.12)$$

donde  $W_k$  constituye la matriz de coeficientes que recupera todas las señales del modelo. de donde planteamos el siguiente Lagrangiano,

$$L = E n_k^H W_k W_k^H n_k - tr(\Lambda^H (W_k^H S_k - I)) = tr(W_k^H R_{nn} W_k - \Lambda^H (W_k^H S_k - I)) \quad (2.3.13)$$

siendo  $\Lambda$  la matriz asociada de multiplicadores de Lagrange. Igualando a cero el gradiente con respecto a la matriz de coeficientes hermítica llegamos a la siguiente relación,

$$\nabla_{W_k^H} L = 0 \Rightarrow R_{nn} W_k - S_k \Lambda^H = 0 \Rightarrow W_k = R_{nn}^{-1} S_k \Lambda^H \quad (2.3.14)$$

Podemos hallar la expresión exacta de la matriz Lagrangiana y por ende de  $W_k$ , forzando el cumplimiento de las restricciones de la matriz de coeficientes,

$$W_k^H S_k = I \Rightarrow \Lambda S_k^H R_{nn}^{-1} S_k = I \Rightarrow \Lambda = (S_k^H R_{nn}^{-1} S_k)^{-1} \Rightarrow W_k = R_{nn}^{-1} S_k (S_k^H R_{nn}^{-1} S_k)^{-1} \quad (2.3.15)$$

donde la matriz Lagrangiana resulta ser la matriz de correlación del ruido de estimación. Vemos pues que esta solución para la matriz de coeficientes coincide con la solución de máxima verosimilitud restringida, con lo cual se demuestra la equivalencia entre ambos criterios

## 2.4 Eliminación de interferencias

A partir de la solución presentada en la ecuación 2.3.14 y suponiendo el modelo de señal expuesto anteriormente queremos determinar si la solución de Wiener consigue eliminar las interferencias. Para ello necesitaremos demostrar si es cierta la siguiente igualdad,

$$w_k^H S_k = [\alpha, 0, 0, \dots, 0] \quad , \quad \alpha \neq 0 \quad (2.4.1)$$

Necesitaremos pues expresar el vector de coeficientes en función de la matriz de steering  $S_k$  y de la matriz de correlación del ruido  $R_{nn}$ . Examinaremos primero el vector de correlación cruzada entre la referencia y los datos  $r_k$ ,

$$\rho_{rd} = E d_k^* (S_k a_k + n_k) = S_k E d_k^* a_k = S_k p_1 \pi_1 = p_1 s_k^1 \quad (2.4.2)$$

donde  $p_1$  representa la amplitud relativa de la fuente con respecto a la referencia y  $\pi_1$  el vector de "pinning" con su primera componente a uno y el resto ceros. La correlación cruzada resulta ser igual al vector de steering de la fuente uno modulado por su amplitud.

Examinemos a continuación la matriz de correlación de los datos de llegada,  $R_{rr}$ ,

$$R_{rr} = E r_k r_k^H = E n_k n_k^H + S_k (E a_k a_k^H) S_k^H = \sigma_n^2 R_{nn} + S_k P_k^2 S_k^H \quad (2.4.3)$$

donde  $P_k$  es una matriz diagonal que contiene las amplitudes de cada una de las fuentes. Vemos que la matriz de correlación de los datos se descompone en las correlaciones de las señales y del ruido, bajo la hipótesis de incorrelación entre ambos. El rango de la primera

matriz es precisamente igual al número de señales siempre que consideremos que la matriz de steering es de rango completo, esto es, que está formada por vectores columna linealmente independientes, y que las señales están incorreladas entre sí. Sustituyendo ambas expresiones en la ecuación de los coeficientes, vemos que en principio éstos dependen de las amplitudes de las señales y de la potencia del ruido a diferencia de la solución de máxima verosimilitud,

$$\mathbf{w}_k = \left( \sigma_n^2 \mathbf{R}_{nn} + \mathbf{S}_k \mathbf{P}_k^2 \mathbf{S}_k^H \right)^{-1} p_1 \mathbf{s}_k^1 \quad (2.4.4)$$

Comprobaremos a continuación si este vector es ortogonal a las señales interferentes. Para ello debemos estudiar la estructura del siguiente vector que proporciona las amplitudes relativas de cada fuente a la salida del filtro,

$$\mathbf{w}_k^H \mathbf{S}_k = \mathbf{s}_k^{1H} p_1 \left( \sigma_n^2 \mathbf{R}_{nn} + \mathbf{S}_k \mathbf{P}_k^2 \mathbf{S}_k^H \right)^{-1} \mathbf{S}_k \quad (2.4.5)$$

o lo que es equivalente, la primera fila de la matriz  $\mathbf{K}$  definida como sigue,

$$\mathbf{K} = \mathbf{P}_k \mathbf{S}_k^H \left( \sigma_n^2 \mathbf{R}_{nn} + \mathbf{S}_k \mathbf{P}_k^2 \mathbf{S}_k^H \right)^{-1} \mathbf{S}_k \Rightarrow \pi_1^T \mathbf{K} = \mathbf{w}_k^H \mathbf{S}_k \quad (2.4.6)$$

puesto que ésta es proporcional a la expresión 2.4.1. Al estar  $\mathbf{K}$  expresada en función de matrices deficientes en rango recurrimos al Lema de Inversión para llegar a una expresión más manejable de  $\mathbf{K}$ ,

$$\left( \sigma_n^2 \mathbf{R}_{nn} + \mathbf{S}_k \mathbf{P}_k^2 \mathbf{S}_k^H \right)^{-1} = \frac{1}{\sigma_n^2} \mathbf{R}_{nn}^{-1} - \frac{1}{\sigma_n^2} \mathbf{R}_{nn}^{-1} \mathbf{S}_k \mathbf{P}_k \left( \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \mathbf{P}_k \right)^{-1} \mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \frac{1}{\sigma_n^2} \quad (2.4.7)$$

pre- y post-multiplicando por la matriz de steering llegamos a la siguiente igualdad,

$$\mathbf{K} = \mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \left( \frac{1}{\sigma_n^2} \mathbf{I} - \frac{1}{\sigma_n^2} \mathbf{P}_k \left( \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \mathbf{P}_k \right)^{-1} \mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \frac{1}{\sigma_n^2} \right) \quad (2.4.8)$$

Aplicando ahora el Lema de Inversión en sentido inverso, llegamos a la siguiente igualdad,

$$\mathbf{K} = \mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \left( \sigma_n^2 \mathbf{I} + \mathbf{P}_k^2 \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \right)^{-1} \quad (2.4.9)$$

donde ahora todas las matrices involucradas en la parte inversa son de rango completo. Resulta fácil comprobar que la primera fila de esta matriz no presenta la estructura supuesta en la ecuación 2.4.1 excepto en el límite  $\sigma_n \rightarrow 0$ ,

$$\lim_{\sigma_n \rightarrow 0} \mathbf{K} = \mathbf{P}_k^{-1} \Rightarrow \lim_{\sigma_n \rightarrow 0} \pi_1^T \mathbf{K} = \lim_{\sigma_n \rightarrow 0} \mathbf{w}_k^H \mathbf{S}_k = \mathbf{p}_1^{-1} \pi_1^T \quad (2.4.10)$$

Es decir, la función de coste definida como la minimización de la potencia del error con respecto a la referencia no consigue eliminar en general la interferencia de otras fuentes excepto en el caso asintótico de relación señal a ruido infinita o cuando se cumple la siguiente condición de ortogonalidad de los vectores de steering,

$$\mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \equiv \text{diagonal} \quad (2.4.11)$$

Podemos encontrar una expresión más compacta para el vector de coeficientes en función de las ecuaciones 2.4.5 y 2.4.8,

$$\begin{aligned} \mathbf{w}_k &= \left( \frac{1}{\sigma_n^2} \mathbf{I} - \frac{1}{\sigma_n^2} \mathbf{R}_{nn}^{-1} \mathbf{S}_k \mathbf{P}_k \left( \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \mathbf{P}_k \right)^{-1} \mathbf{P}_k \mathbf{S}_k^H \frac{1}{\sigma_n^2} \right) \mathbf{R}_{nn}^{-1} \mathbf{p}_1 s_k^1 \\ &= \left( \sigma_n^2 \mathbf{I} + \mathbf{P}_k^2 \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \right)^{-1} \mathbf{R}_{nn}^{-1} \mathbf{p}_1 s_k^1 \end{aligned} \quad (2.4.12)$$

Podemos ver de una forma más intuitiva el efecto de las interferencias si consideramos el comportamiento en situaciones de relación señal a ruido elevada. Esto es equivalente a desarrollar la expresión de la matriz  $\mathbf{K}$  en serie de Taylor en un entorno de la variable  $\sigma_n$  cercano a cero, quedándonos únicamente con los términos de hasta primer orden,

$$\mathbf{K} = \sum_{i=0}^{+\infty} \mathbf{K}_i \sigma_n^{2i} \approx \mathbf{K}_0 + \mathbf{K}_1 \sigma_n^2, \quad \mathbf{K}_1 = \left. \frac{d}{d\sigma_n^2} \mathbf{K} \right|_{\sigma_n^2=0} \quad (2.4.13)$$

El primer término de la expresión es simplemente el límite asintótico de  $\mathbf{K}$  ya deducido en la ecuación 2.4.10. La primera derivada de  $\mathbf{K}$  puede calcularse fácilmente de la siguiente forma,

$$\mathbf{K}_1 = -\mathbf{P}_k \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \left. \frac{d}{d\sigma_n^2} \left( \sigma_n^2 \mathbf{I} + \mathbf{P}_k^2 \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \right)^{-1} \right|_{\sigma_n^2=0} = -\mathbf{P}_k^{-1} \left( \mathbf{P}_k^2 \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k \right)^{-1} \quad (2.4.14)$$

de donde la aproximación final de la matriz  $\mathbf{K}$  aparece como,

$$(2.4.15)$$

fenómeno de conoce como **inversión de potencia**. Vemos que es función también de la correlación que exista entre los vectores de steering de las fuentes respectivas.

El examen del caso particular de una sola fuente en presencia de ninguna interferencia y ruido coloreado proporciona una comprensión intuitiva de los efectos del algoritmo. Particularizando el vector de coeficientes para este caso llegamos a la siguiente expresión,

$$\mathbf{w}_k = \frac{\mathbf{R}_{nn}^{-1} p \mathbf{s}_k}{\sigma_n^2 + p^2 \mathbf{s}_k^H \mathbf{R}_{nn}^{-1} \mathbf{s}_k} \Rightarrow \alpha = \frac{p^2 \mathbf{s}_k^H \mathbf{R}_{nn}^{-1} \mathbf{s}_k}{\sigma_n^2 + p^2 \mathbf{s}_k^H \mathbf{R}_{nn}^{-1} \mathbf{s}_k} \tag{2.4.16}$$

de donde hemos deducido el factor  $\alpha$  de contracción de la constelación (acotado en el intervalo [0,1] en función de la relación señal a ruido existente). Sustituyendo  $\alpha$  en la función de coste definida anteriormente, vemos que la potencia del error se descompone en los siguientes términos en función del factor de contracción,

$$|e|^2 = \frac{\sigma_n^2}{\sigma_n^2 + p^2 \mathbf{s}_k^H \mathbf{R}_{nn}^{-1} \mathbf{s}_k} E|d_k|^2 + \frac{p^2 \mathbf{s}_k^H \mathbf{R}_{nn}^{-1} \mathbf{s}_k}{\sigma_n^2 + p^2 \mathbf{s}_k^H \mathbf{R}_{nn}^{-1} \mathbf{s}_k} \sigma_n^2 = (1 - \alpha) E|d_k|^2 + \alpha \sigma_n^2 \tag{2.4.17}$$

Apareciendo los términos de ruido y de potencia residual de la señal.

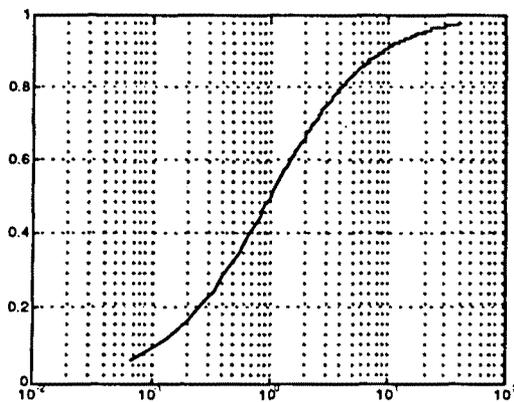


Figura 2.2.1: evolución del factor de contracción de la constelación en función de la relación señal a ruido.

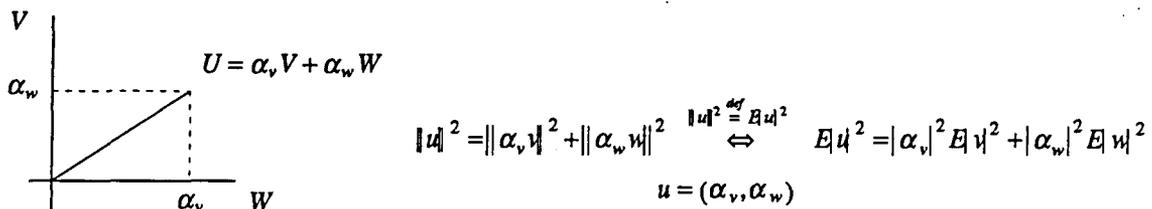
En la figura 2.2.1 representamos la evolución del factor de contracción en función de SNR (consideramos para ello que la norma del vector de steering en la norma de la inversa de la correlación del ruido es igual a la unidad sin pérdida de generalidad).

En la figura 2.2.1 representamos la evolución del factor de contracción en función de SNR (consideramos para ello que la norma del vector de steering en la norma de la inversa de la correlación del ruido es igual a la unidad sin pérdida de generalidad).

Llegados a este punto, hemos realizado un estudio detallado de la solución de Wiener en presencia de interferencias. En el siguiente apartado pasamos a realizar una interpretación geométrica en un espacio multi-dimensional de dicha solución. Igualmente, estableceremos la relación existente entre el criterio de referencia espacial y temporal en términos geométricos.

### 2.5 Espacio Estadístico Normado

En este apartado proponemos una interpretación geométrica de los criterios anteriores en base a un espacio estadístico normado (este tipo de desarrollo se debe originalmente a Kotelnikov, quién lo desarrolló por vez primera en el campo de la detección). La idea básica en la concepción de este espacio consiste en determinar una variable aleatoria como un punto o un vector del espacio. Suponemos unos ejes de coordenadas ortogonales donde cada eje se asocia a una variable aleatoria. Se entiende en lo sucesivo si los ejes son ortogonales, que las variables aleatorias asociadas son independientes e incorreladas:



Es decir, se define un producto escalar entre dos vectores en este espacio como la correlación cruzada de las variables aleatorias asociadas:  $\langle x, y \rangle = Exy$ . La potencia de la variable aleatoria o norma del vector asociado se calcula entonces como  $\langle x, x \rangle = E|x|^2$ . No existe restricción en cuanto a las densidades marginales de probabilidad de cada una de las variables aleatorias coordenadas (la extensión al caso multidimensional es directa). Cada vector de este espacio,  $u = (\alpha_1, \alpha_2, \dots, \alpha_N)$  tiene por tanto asociada una función de densidad de probabilidad igual a la convolución de las funciones de densidad de las componentes proyectadas sobre cada eje. La versatilidad de este concepto se basa en la dualidad entre combinación lineal de variables aleatorias y combinación lineal de vectores. Es importante

destacar que un punto de este espacio es una variable aleatoria y no una realización particular de esta y que por tanto las distancias o normas en este espacio equivalen a potencias promediadas y no a módulos instantáneos. Nótese que dos ejes coordenados de este espacio pueden estar asociados a variables aleatorias con la misma función de densidad, la ortogonalidad de estos ejes sólo está relacionada con su función de densidad conjunta. Veremos que la operación de filtrado se representa en este espacio en base a conceptos geométricos básicos.

Supongamos que los datos recibidos se presentan como un vector de  $N$  componentes (por ejemplo un "snapshot" en un array de comunicaciones) mientras que la dimensión del espacio estadístico está determinada por el número de fuentes independientes e incorreladas que contribuyen a la formación de la señal,  $N'$  (señal deseada más número de señales interferentes más número de términos de ruido en cada sensor). La variable aleatoria snapshot se representará entonces en este espacio como una constelación de puntos, donde cada uno se corresponde con la variable aleatoria asociada a cada componente del vector de snapshot,

$$\mathbf{r}_k = a_k^1 s_k^1 + \sum_i a_k^i s_k^i + \mathbf{n}_k, \quad \dim(\mathbf{r}_k) = N \Leftrightarrow \mathbf{u}^T = [a_k^1, a_k^2, \dots, a_k^{N'}, \mathbf{n}_k^T], \quad \dim(\mathbf{u}) = N' \quad (2.5.1)$$

donde el vector  $\mathbf{u}$  contiene a todas las variables aleatorias generadoras del espacio. El proceso de filtrado del snapshot se entiende entonces estadísticamente como una combinación lineal de los puntos de esta constelación o equivalentemente como el subespacio generado por los vectores asociados a cada punto de la constelación. Cada punto (vector) de este subespacio es una posible variable aleatoria a la salida del array dada una determinada combinación lineal (conformación),

$$z_k = \mathbf{w}_k^H \mathbf{r}_k \Leftrightarrow U(z_k) = \sum_{i=1}^N w_k^i U(r_k^i) \quad (2.5.2)$$

donde mediante  $U(z)$  denotamos el vector en el espacio estadístico asociado a la variable aleatoria  $z$ . Si no imponemos restricciones sobre los coeficientes vemos que el origen de coordenadas forma siempre parte del plano de filtrado, en el caso de un vector de coeficientes nulo.

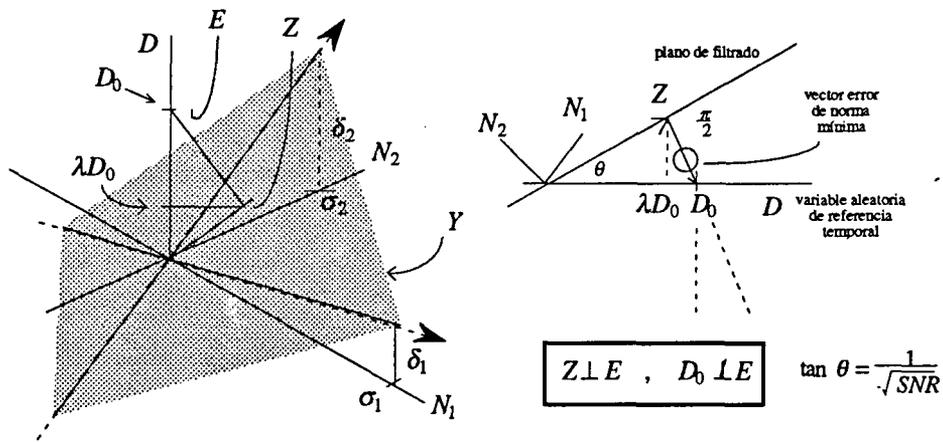


Figura 2.5.1: Representación vectorial del método de referencia temporal en el espacio estadístico normado

En la figura anterior hemos realizado una representación esquemática del método de referencia temporal en el espacio estadístico normado. Hemos supuesto un snapshot de dos componentes y una única señal útil. La dimensión del espacio es por tanto igual a 3: una componente de señal más dos componentes de ruido (que no son necesariamente las componentes de ruido en cada sensor. Puesto que podemos tener ruido coloreado en los sensores, en la representación de la figura debemos considerar ruido incoloreado, obtenido después de multiplicar el vector de snapshot por una matriz blanqueadora de ruido). El eje D representa el eje coordenado de la señal útil, mientras que los ejes  $N_1$  y  $N_2$  son los ejes correspondientes al ruido de sensor. El plano de filtrado (o de conformación) se construye a partir de los vectores generadores  $(\delta_1, \sigma_1, 0)$  y  $(\delta_2, 0, \sigma_2)$ , es decir, la representación vectorial de las variables aleatorias presentes en cada sensor. Debemos entonces hallar cual es el vector contenido en el plano de filtrado más próximo a la variable aleatoria de referencia  $D_0$ . Este vector, Z, resulta ser la proyección de  $D_0$  sobre el plano de filtrado, tal que  $D_0 = Z + E$ , donde vemos que Z es perpendicular a E, es decir, ambos se encuentran incorrelados. No obstante, E no es ortogonal a  $D_0$ , es decir, existe correlación no nula entre la señal de error resultante y la referencia. La magnitud de esta correlación se obtiene proyectando el vector E sobre el eje coordenado D según la dualidad proyección-correlación.

Podemos hallar también una interpretación geométrica de la ecuación 2.3.5 en el contexto de la Figura 2.5.1. En efecto, dada la descomposición del vector de coeficientes  $w_k = \lambda w_0$ , y considerando a  $w_0$  fijo e igual a la solución de mínima varianza y al factor de escalado  $\lambda$  variable, el conjunto de variables aleatorias filtradas por  $w_k$  se encuentran todas contenidas en la recta que contiene al vector Z de la figura. En otras palabras, la recta construida

mediante la intersección del plano de filtrado y un plano perpendicular a éste que contenga el eje coordenado de la referencia,  $D$ , genera todas las variables aleatorias producidas por un vector de coeficientes proporcional a la solución de mínima varianza,  $w_0$ . El escalado correcto de la variable  $Z$  que minimiza la norma del error se encuentra por consiguiente asociado a la componente ortogonal al plano de filtrado con respecto a la referencia  $D_0$ . Cabe destacar que el ángulo que forma el eje de la referencia con la recta  $Z$  es igual a la raíz cuadrada de la inversa de la relación señal a ruido a la salida del filtro. Es decir, ángulos pequeños se corresponden con relaciones señal a ruido elevadas o potencias de ruido (componente ortogonal) pequeñas. El esquema representado en la Figura 2.5.1 se corresponde por lo tanto con una situación de SNR baja debido a la magnitud del ángulo  $\theta$ , lo cual facilita su representación gráfica. Obsérvese que a medida que  $\theta$  aumenta, la magnitud de la proyección de  $Z$  sobre el eje  $D$  de referencia disminuye a su vez, decrece el factor de escalado  $\lambda$  con lo cual tenemos una contracción de la constelación de símbolos. Podemos relacionar fácilmente el factor  $\lambda$  con el ángulo  $\theta$ ,

$$\sin \theta = \frac{(1-\lambda)\|D_0\|}{\|E\|} = \frac{1-\lambda}{\sin \theta} \Rightarrow \sin^2 \theta = 1-\lambda \quad , \quad \lambda = \cos^2 \theta \tag{2.5.3}$$

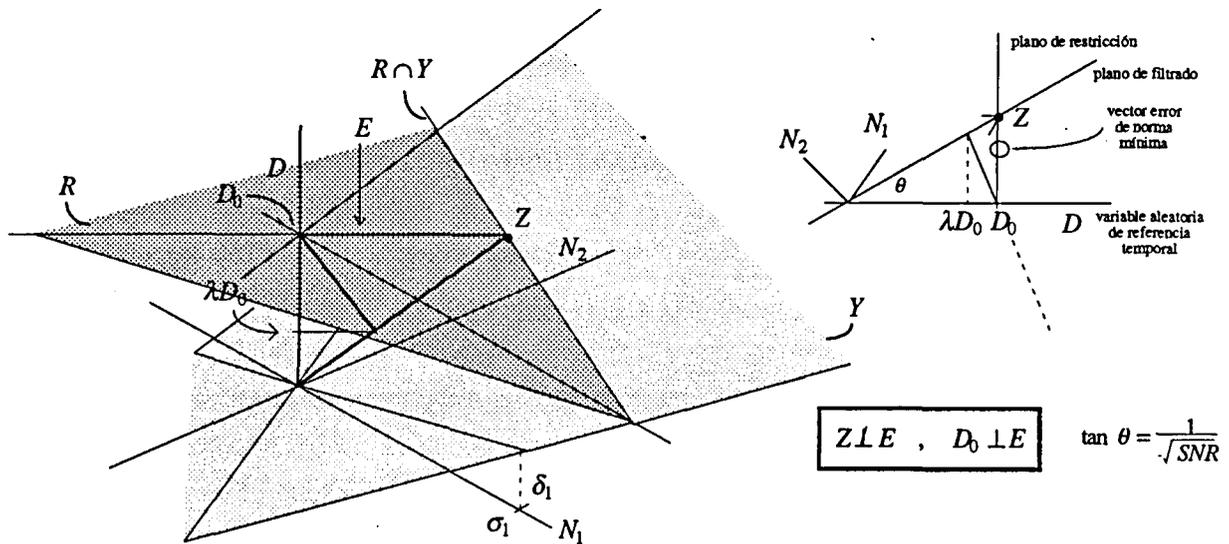


Figura 2.5.2: Representación vectorial del método de referencia espacial en el espacio estadístico normado

En la figura 2.5.2 representamos la interpretación geométrica del método de referencia espacial. En ella,  $Y$  constituye el plano de filtrado ya citado en la ecuación 2.5.2,  $R$  es el

plano de restricciones perpendicular al eje de referencia  $D$  y que pasa por  $D_0$ , y  $R \cap Y$  constituye su intersección, es decir, la recta que contiene todas las variables aleatorias de salida posibles del filtro dada una estadística de entrada definida por el plano  $Y$  y tales que cumplen la restricción del plano  $R$ . De todos los puntos de la recta  $R \cap Y$ , nos interesará aquél cuya distancia al punto  $D_0$  sea mínima, al cuál denominamos  $Z$  en la figura 2.5.2. Éste corresponderá a la variable aleatoria que minimiza la potencia del error dadas unas restricciones sobre el filtro proporcionadas por la referencia espacial. Por construcción, este punto se encuentra en la misma línea definida  $Z$  de la figura de referencia temporal.

En el siguiente apartado realizaremos el estudio de un caso particular, pero no por ello marginal, de referencia espacial: la decorrelación de señales CDMA mediante un array de antenas. Veremos como del estudio de este caso se derivarán resultados teóricos aplicables al problema de la ecualización, del que nos ocuparemos en capítulos subsiguientes desde una nueva perspectiva.

## 2.6 Decorrelación en Array

En esta sección analizaremos en detalle el problema de la detección multi-usuario aplicando en concepto de referencia espacial ya introducido en las secciones 2.1 y 2.2. Más adelante, en la sección 2.7 analizaremos el mismo problema para el caso de usuarios síncronos bajo la óptica de referencia temporal. Deduciremos un algoritmo para recuperar las secuencias transmitidas por más de un usuario en un terminal receptor conjunto. El modelo que se considera es el de un canal DS-CDMA (Direct Sequence Code Division Multiple Access). Modelos de receptor más simples como el ecualizador para un único usuario pueden obtenerse mediante particularización de las ecuaciones deducidas en esta sección. Se considera el caso general de un receptor en array para la detección de las señales CDMA y se evalúan las prestaciones del detector propuesto. Para el análisis se ha considerado que la distribución del ruido en cada uno de los sensores es Gaussiana y blanca. Se supone ruido incorrelado sensor a sensor. Supondremos también que la sincronización de código con respecto a cada uno de los usuarios se ha realizado previamente, así como la adquisición de los vectores de enfoque o de steering correspondientes. Propondremos un algoritmo lineal al que denominamos el Detector por Decorrelación en Array (DDA o ADD: Arrayed Decorrelating Detector). El principio básico consiste en efectuar un procesamiento espacio-temporal conjunto que satisface un criterio de mínimo error cuadrático medio. El detector que se propone es robusto frente al efecto "near-far" y sus prestaciones son independientes de las potencias de los usuarios. Veremos que el ADD es una generalización del decorrelador de sensor único. Veremos también que la probabilidad de error del ADD presenta una mejora en economía de escala respecto al decorrelador de sensor único debido al mayor número de dimensiones que se pueden considerar en este nuevo entorno. Las conclusiones teóricas se verificarán por simulación.

### 2.6.1 Introducción

Las redes de comunicaciones basadas en Acceso Múltiple por División de Código permiten el intercambio simultáneo de información entre usuarios que comparten la misma banda de frecuencias. Esto se consigue en base a la asignación de códigos de secuencias pseudo-aleatorias diferentes a cada uno de los usuarios para la modulación de los datos correspondientes. El usuario final puede recuperar la información transmitida en base a la correlación de los datos (muestras) recibidos con la secuencia PN conocida. Un nivel

mínimo de fiabilidad del sistema exige que las secuencias asignadas presenten un grado de correlación cruzada tan pequeño como sea posible a fin de minimizar el nivel de interferencia inter-usuario. Un grave problema siempre presente en las redes de comunicación CDMA lo constituye el efecto "near-far", consecuencia de los diferentes niveles de potencia de cada usuario y de los coeficientes de correlación cruzada entre secuencias. La interferencia de múltiple acceso reduce considerablemente las prestaciones de los detectores basados en la correlación con el filtro adaptado a la secuencia PN para canales AWGN ya que el usuario de interés puede quedar completamente enmascarado por otros usuarios. Un cuidadoso diseño de las secuencias PN no es suficiente para asegurar una calidad aceptable del enlace para todas las posibles combinaciones de sincronismos entre usuarios. Demostraremos en este capítulo que la sensibilidad near-far de los detectores puede mejorarse cuando recurrimos al concepto de Subspace Matched Filtering. Ya se había demostrado previamente [Verdú] que el efecto near-far puede solucionarse completamente cuando se conocen los sincronismos de cada usuario y las secuencias asociadas. No obstante, sus prestaciones se ven limitadas por la posible distribución de sincronismos y códigos de "spread", es decir, en cuán linealmente independientes sean las secuencias PN de los usuarios. Demostraremos que el nivel de decorrelación entre usuarios puede aumentar cuando consideramos entornos de dimensionalidad más elevada, es decir, cuando disponemos de varias realizaciones acopladas de la misma secuencia. Deduciremos la expresión del ADD óptimo para un número de sensores no necesariamente equiespaciados. Las expresiones serán válidas para cualquier distribución de sensores. Cabe destacar, que un procesado espacio-temporal conjunto redundará en una mejor decorrelación entre usuarios.

A lo largo de toda la demostración supondremos que se ha realizado previamente una estimación de DOA o de adquisición de los pesos del array, siéndole así conocidos al receptor los ángulos de llegada de cada usuario, así como sus respectivos sincronismos. Es importante destacar que el número de sensores del array no vendrá determinado por el número de usuarios que deseamos demodular. De hecho, se pretende que los algoritmos que aquí se desarrollan sean aplicables a la detección de más usuarios que sensores (en particular para un bajo número de sensores) ya que el procesado espacio-temporal conjunto debiera ser suficiente para decorrelar a los usuarios. De esta forma pueden relajarse las especificaciones de la etapa de procesado espacial. El detector por decorrelación de un solo sensor es un caso particular del detector que aquí presentamos. Puede considerársele como un array omnidireccional de ancho de haz  $2\pi$ . Para un array de-facto, el ancho de haz disminuye. Demostraremos no obstante que usuarios que se encuentren espacialmente separados por debajo de la resolución del haz pueden aún decorrelarse a un nivel mayor que en el caso de un único sensor debido al procesado espacio-temporal. Los diferentes desfases

de los frentes de onda que inciden sobre el array desde direcciones diferentes contribuyen a hacer la matrix de correlación de usuarios más "parecida" a una matriz diagonal. La complejidad del detector es lineal tanto en el número de sensores como en el número de usuarios.

La utilización de técnicas de arrays en entornos CDMA no se limita únicamente al problema de la detección. Ayuda también a combatir la degradación en la calidad de las señales debido a multicamino e interferencias. En particular, un array de sensores omnidireccionales resulta ser un método eficiente y flexible para implementar la sectorización en una red de comunicaciones móviles.

El modelo de señal se presenta en la Sección 2.6.2. En la sección 2.6.3 se derivan las ecuaciones del ADD a partir de la función de probabilidad conjunta de los datos para un intervalo finito. La probabilidad de error del detector propuesto se analiza en la Sección 2.6.4, y sus prestaciones se comparan con las del detector convencional de un único sensor. La convergencia y la probabilidad de error asintótica cuando la longitud del intervalo tiende a infinito se analizan en la sección 2.6.5. En los apéndices, secciones 2.6.6, 2.6.7 y 2.6.8, se analizan los criterios de detección así como algunas variantes del ADD en entornos de ruido coloreado. Finalmente, en la Sección 2.6.9 se estudia el problema de decorrelación adaptativa basada en técnicas de referencia temporal para el caso de acceso síncrono, en contraste con las técnicas de referencia espacial aplicadas en las secciones anteriores.

### **2.6.2 Modelo de Señal**

Antes de definir nuestro modelo de señal, es conveniente determinar si nos enfrentamos a un problema de banda estrecha o de banda ancha desde el punto de vista de procesado de arrays. Las señales de espectro ensanchado por secuencia directa del tipo utilizado en comunicaciones CDMA se consideran generalmente pertenecientes al segundo tipo, puesto que la banda de información se ha extendido sobre un ancho de banda relativamente mayor, aproximadamente igual a la inversa del tiempo de duración del chip. En términos de procesado de arrays, se considera que nos encontramos en un entorno de banda ancha cuando el mayor retardo de propagación entre sensores es del orden del inverso del ancho de banda temporal de la señal. Con esto queremos decir que el punto hasta el que se considera que una señal es de banda ancha depende de la aplicación y no es por tanto un concepto absoluto. Para el problema que nos concierne, y desde el punto de vista de

procesado de arrays, supondemos que el array es de banda estrecha, es decir, el tiempo de duración del chip es mucho mayor que el tiempo de retardo de propagación entre sensores. En términos de las variables físicas en cuestión, esto se reduce para un array lineal a,

$$\frac{1}{R_c} \gg \frac{d \sin \theta}{c} \Rightarrow R_c \ll f_c < \frac{c}{d} \quad (2.6.2.1)$$

donde  $R_c$ ,  $d$  y  $c$  representan la velocidad de chip, el espaciado entre sensores y la velocidad de propagación en el medio, respectivamente. La frecuencia portadora  $f_c$  está ligada a las dimensiones del array para evitar una ambigüedad en la determinación de los ángulos de llegada de los usuarios y debe ser mucho mayor que el ancho de banda de la señal.

Dos suposiciones se han llevado a cabo en cuanto al modelo de señal:

- no se ha supuesto multicamino debido a las especiales características de propagación del medio.
- se ha supuesto ruido espacial y temporalmente incorrelado.

Esta última suposición representa las condiciones más duras a las que debe hacer frente el array dado que constituyen una evaluación de peor caso de la probabilidad de error con respecto a ruido correlado de la misma potencia.

Definamos la señal transmitida por cada usuario de la siguiente forma,

$$s_k(n) = \sum_{m=-\infty}^{m=+\infty} a_k(m) p'_k(n - mN - \varepsilon_k) = \sum_{m=-\infty}^{m=+\infty} a_k(m) p_k(n - mN) \quad , \quad 1 \leq k \leq N_u \quad (2.6.2.2)$$

Donde  $A_k = \{a_k(m), -\infty \leq m \leq +\infty\}$  constituye la secuencia de información transmitida por cada usuario en un total de  $N_u$  usuarios indexados por el subíndice  $k$ . El conjunto de retardos entre usuarios lo hemos representado por  $\{\varepsilon_k, 1 \leq k \leq N_u\}$ . El conjunto de pulsos de conformación  $P = \{p_k(n - mN), -\infty \leq m \leq +\infty\}$  constituye el soporte de la información transmitida por los usuarios. Nótese que la información relativa al sincronismo se encuentra implícita en este conjunto supuesto conocido. Estas señales son recogidas por un array no necesariamente lineal, donde supondremos que se cumplen las condiciones de banda estrecha. En consecuencia, la señal recibida por cada uno de los  $N_s$  sensores se puede expresar como,

$$r_i(n) = w_i(n) + \sum_{k=1}^{k=N_u} s_{ki} s_k(n) \quad , \quad 1 \leq i \leq N_s \quad (2.6.2.3)$$

El conjunto  $\mathbb{R} = \{r_i(n), 1 \leq i \leq N_s\}$  constituye las  $N_s$  secuencias de datos de las que se deben estimar los símbolos  $A_k$ .  $\mathbb{W} = \{w_i(n), 1 \leq i \leq N_s\}$  constituyen los procesos de ruido blanco Gaussiano aditivo en cada sensor, incorrelados entre sí.  $S_{ki}$  denotan las componentes del vector de steering asociado a cada usuario. Las componentes de la matriz de steering  $S$  se pueden expresar en términos de los vectores de steering de cada usuario como,

$$S = [s_1, s_2, \dots, s_{N_u}] \quad , \quad s_k^T = [e^{j\theta_{k,1}}, e^{j\theta_{k,2}}, \dots, e^{j\theta_{k,N_u}}] \quad (2.6.2.4)$$

donde  $\theta_{ki}$  representan los desfases de los diversos frentes de onda incidentes en el array. Podemos ahora agrupar la señales  $r_i(n)$  y  $w_i(n)$  de la ecuación 2.6.2.3 recibidas en cada sensor en los vectores  $r(n)$  y  $w(n)$  y las secuencias de símbolos  $a_k(n)$  y los pulsos  $p_k(n)$  de la ecuación 2.6.2.2 en los vectores  $a(n)$  y  $p(n)$  respectivamente. Esto nos permite expresar  $r(n)$  en función de  $a(n)$  de forma muy compacta,

$$r(n) = w(n) + S \sum_{m=-\infty}^{m=+\infty} p(n-mN) \bullet a(m) \quad (2.6.2.5)$$

donde  $\bullet$  denota el producto de Schur o de Hadamard (componente a componente) entre matrices rectangulares. El proceso de ruido  $w(n)$  se considera espacial y temporalmente incorrelado con matriz de covarianza  $E\{w(n)w^H(m)\} = \sigma^2 I_{nm}$ .

El algoritmo lineal óptimo para la decorrelación de  $N_u$  usuarios del tipo definido en la ecuación 2.6.2.2 se deduce a partir de un criterio de mínimo error cuadrático. Una condición fundamental para hacer posible la decorrelación es que las secuencias de  $P$  sean linealmente independientes. De esta forma se garantiza que las secuencias de símbolos de cada usuario pueden recuperarse mediante un algoritmo lineal. A modo de justificación basta considerar cada símbolo asociado a un vector determinado en el espacio de los datos.

Se aprovecha la decorrelación en el dominio espacial (índice  $i$  en la ecuación 2.6.2.3) y la decorrelación en el dominio temporal (índice  $n$  en la ecuación 2.6.2.2). En este sentido, un número de señales superior a la dimensión del vector de observación (igual al número de sensores si consideramos un array),  $N_u \geq N_s$ , puede extraerse de los datos disponibles. Aunque para este caso, el procesado espacial no basta para decorrelar a todos los usuarios, el hecho de que a cada usuario le haya sido asignada una secuencia diferente permite un procesado temporal que garantiza una correcta decorrelación.

La estructura de la matriz de steering tiene una influencia importante en la capacidad del sistema para decorrelar los usuarios. Se puede demostrar que la bondad de la decorrelación está en función de cuán ortogonales sean las columnas de la matriz de steering. La ortogonalidad o independencia lineal entre las columnas de  $\mathbf{S}$  se cuantifica según la magnitud relativa con respecto a los elementos de la diagonal de los componentes del Grammiano de la matriz de steering,  $\mathbf{S}^H\mathbf{S}$ . En general, una correlación no nula entre señales, temporal y espacialmente, conduce a un decremento en las prestaciones de decorrelación del algoritmo óptimo. Este hecho puede explicarse en términos de la equivalencia entre correlación cruzada y superposición parcial de los subespacios asociados a cada señal. En el caso límite en que la correlación cruzada entre dos señales atañe su máximo valor, los subespacios asociados son completamente paralelos y la decorrelación es deficiente en rango. Es decir, no existe ningún algoritmo lineal capaz de decorrelar dichas señales.

### 2.6.3 El ADD

En esta sección definimos el criterio de optimización para la detección de los símbolos transmitidos por cada usuario. Se considerarán dos enfoques basados en la dimensión del vector de datos utilizado para la detección y se comparará la estructura de ambos. Primero, el criterio se define como la minimización de la esperanza estadística de la siguiente norma del error,

$$L = E\{\mathbf{e}_r^H(n)\mathbf{e}_r(n)\} \quad , \quad \mathbf{e}_r(n) = \mathbf{r}(n) - \mathbf{r}_a(n) \quad (2.6.3.1)$$

donde  $\mathbf{r}(n)$  denota el vector de snapshots en un determinado instante de tiempo i  $\mathbf{r}_a(n)$  el modelo de la señal sin ruido. Aunque se demostrará que este criterio proporciona una probabilidad de error mínima, de momento trabajaremos con segmentos finitos de datos a expensas de una disminución en las prestaciones. Un criterio más apropiado para este caso lo constituye la norma del error descrita en la ecuación 2.6.3.2. Si los procesos involucrados son ergódicos, ambos criterios son equivalentes cuando la dimensión de la ventana de datos tiende a infinito.

$$L_{N'} \cong \frac{1}{2N'+1} \sum_{n=-N'}^{n=N'} \mathbf{e}_r^H(n)\mathbf{e}_r(n) \quad (2.6.3.2)$$

La minimización de esta expresión se lleva a cabo con respecto a la secuencia  $\mathbf{a}(m)$ . Aunque los símbolos transmitidos pertenecen a un conjunto finito determinado por la constelación (BPSK, QPSK), la utilización de esta información lateral conduce a un algoritmo tipo Viterbi con complejidad exponencial en el número de usuarios. Por lo tanto, supondremos que los componentes de  $\mathbf{a}(m)$  pueden tomar cualquier valor en el plano complejo. De este modo, la minimización se puede abordar derivando con respecto a  $\mathbf{a}(m)$  en vez de minimizar la norma del error sobre una retícula multidimensional definida por todas las combinaciones posibles de los símbolos transmitidos por cada usuario. Para la minimización de la expresión 2.6.3.2 tomamos derivadas con respecto a  $\mathbf{a}^H(m')$  e igualamos a zero,

$$(2N'+1) \frac{\partial}{\partial \mathbf{a}^H(m')} L_{N'} = - \sum_{n=-N'}^{n=+N'} \frac{\partial \mathbf{r}_a^H(n)}{\partial \mathbf{a}^H(m')} \mathbf{e}_r(n) = 0 \quad (2.6.3.3)$$

Utilizando algunas propiedades del producto de Schur y la ecuación 2.6.2.5 para obtener un modelo de  $\mathbf{r}_a$ , el término de la derivada y  $\mathbf{r}_a$  pueden expandirse y resulta,

$$\begin{aligned} \frac{\partial \mathbf{r}_a^H(n)}{\partial \mathbf{a}^H(m')} &= \text{diag}(\mathbf{p}^H(n-m'N)) \mathbf{S}^H \Rightarrow \\ \mathbf{b}(m') &= \sum_n \text{diag}(\mathbf{p}^H(n-m'N)) \mathbf{S}^H \mathbf{r}(n) \\ &= \sum_n \text{diag}(\mathbf{p}^H(n-m'N)) \mathbf{S}^H \mathbf{S} \sum_m \mathbf{p}(n-mN) \bullet \mathbf{a}(m) \end{aligned} \quad (2.6.3.4)$$

La secuencia  $\mathbf{b}(m')$  de la ecuación 2.6.3.4 constituye un conjunto de estadísticos suficientes para la detección de los símbolos. Si refundimos  $\mathbf{b}(m')$  del siguiente modo, podemos ver que  $\mathbf{b}(m')$  puede obtenerse al pasar por un filtro adaptado la secuencia resultante de realizar el producto escalar de los vector de snapshot con el vector de steering,

$$b_k(m') = \sum_{n=-N'}^{n=+N'} (\mathbf{s}_k^H \mathbf{r}(n)) p_k(n-m'N) \quad , \quad 1 \leq k \leq N_u \quad (2.6.3.5)$$

Volviendo a la ecuación 2.6.3.4, el término de la derecha se puede simplificar como sigue,

$$\begin{aligned} &\sum_n \text{diag}(\mathbf{p}^H(n-m'N)) \mathbf{S}^H \mathbf{S} \sum_m \mathbf{p}(n-mN) \bullet \mathbf{a}(m) \\ &= \sum_m \mathbf{a}(m) \sum_n \text{diag}(\mathbf{p}^H(n-m'N)) \mathbf{S}^H \mathbf{S} \text{diag}(\mathbf{p}(n-mN)) \\ &= \sum_m \mathbf{a}(m) \sum_n (\mathbf{p}^*(n-m'N) \mathbf{p}^T(n-mN)) \bullet (\mathbf{S}^H \mathbf{S}) \end{aligned} \quad (2.6.3.6)$$

En su forma final, tenemos que la secuencia de estadísticos suficientes  $\mathbf{b}(m')$  se encuentra relacionada con la secuencia de símbolos estimados en función de las siguientes matrices de correlación cruzada,

$$\mathbf{b}(m') = \sum_{m=-\infty}^{m'+\infty} (\mathbf{R}(m, m') \bullet (\mathbf{S}^H \mathbf{S})) \mathbf{a}(m) \quad (2.6.3.7)$$

Los componentes de la matriz  $\mathbf{R}(m, m')$  son simplemente las correlaciones cruzadas entre los pulsos  $p_k(n-m'N)$  y  $p_k(n-mN)$ ,

$$[\mathbf{R}(m, m')]_{kk'} = \sum_{n=-N'}^{n'+N'} p_k^*(n-m'N) p_{k'}(n-mN) \quad (2.6.3.8)$$

Nótese que debido a que los límites del sumatorio de la ecuación 2.6.3.8 son finitos no se cumple la propiedad de estacionariedad de las matrices de correlación:  $\mathbf{R}(m, m') = \mathbf{R}(m-m')$ . Los componentes de  $\mathbf{R}(m, m')$  se ponderan mediante el producto de Schur con el Grammiano de la matriz de steering  $\mathbf{S}^H \mathbf{S}$ , relacionando la correlación global entre usuarios con su correlación temporal y espacial, respectivamente. La correlación global se cuantifica según la magnitud relativa de los componentes fuera de la diagonal de las siguientes matrices con respecto a sus términos diagonales,

$$\bar{\mathbf{R}}(m, m') = \mathbf{R}(m, m') \bullet (\mathbf{S}^H \mathbf{S}) \quad (2.6.3.9)$$

Debido a que la correlación global se expresa a partir del producto de correlaciones parciales, la correlación global será siempre menor que cualquiera de las parciales, temporal o espacial. Cuando cualquier término fuera de la diagonal de  $\mathbf{R}$  o de  $\mathbf{S}^H \mathbf{S}$  sea zero, es decir, tenemos correlación temporal nula (las secuencias temporales son ortogonales) o correlación espacial nula (los vectores de steering son ortogonales), la decorrelación global queda asegurada. Debido a las especiales propiedades del Grammiano  $\mathbf{S}^H \mathbf{S}$  tendremos que las matrices de correlación modificadas se han vuelto más diagonales como se demuestra por la relación entre los componentes diagonales y no-diagonales,

$$[\bar{\mathbf{R}}(m, m')]_{kk} / [\bar{\mathbf{R}}(m, m')]_{ki} = \left( \frac{s_k^H s_k}{s_k^H s_i} \right) [\mathbf{R}(m, m')]_{kk} / [\mathbf{R}(m, m')]_{ki} \quad , \quad \frac{s_k^H s_k}{s_k^H s_i} \geq 1 \quad (2.6.3.10)$$

Nótese que para un intervalo de observación finito las matrices  $\mathbf{R}$  no cumplen la propiedad de estacionariedad. Tal condición sólo se cumple cuando suponemos un intervalo de dimensión infinita. El objetivo es ahora el de procesar el conjunto finito de estadísticos suficientes  $\mathbf{b}(m')$  desde  $-N'$  a  $+N'$  para obtener las estimaciones de los símbolos

transmitidos. Aunque los límites del sumatorio van de menos infinito a más infinito, sólo un número finito de componentes de  $\mathbf{a}$  contribuyen a  $\mathbf{b}$  (debido a la finitud de las secuencias de spread, cada símbolo sólo está interferido por los símbolos inmediatamente pasados o futuros del resto de usuarios). Por lo tanto, el problema de detección se reduce a resolver un sistema lineal de ecuaciones. Organizemos los componentes de interés de las secuencias  $\mathbf{a}$  y  $\mathbf{b}$  de la ecuación 2.6.3.7 en la siguiente forma matricial,

$$\tilde{\mathbf{b}} = \Gamma_N \tilde{\mathbf{a}} \tag{2.6.3.11}$$

$$\tilde{\mathbf{a}}^T = [a_2(-N), \dots, a_{N_u}(-N), a_1(-N+1), \dots, a_{N_u}(-N+1), \dots, a_1(N), \dots, a_{N_u-1}(N)]$$

Donde la matriz  $\Gamma$  es una matriz de bloques quasi-diagonal, debido al spill-over de las colas de los símbolos en ambos límites de la ventana de observación (ver figura 2.6.3.1). El ADD se define entonces a partir de la matriz de decorrelación  $\Gamma^{-1}$  que recupera los símbolos a la salida del banco de filtros adaptados (ver figure 2.6.3.2). cuando la longitud de la ventana de observación tiende a infinito, la dimensión de la matriz de correlación  $\Gamma$  también tiende a infinito, aproximándose a la estructura de una matriz de Toeplitz ya que en el límite se cumplirá la propiedad de estacionariedad ( $\mathbf{R}(m, m') = \mathbf{R}(m - m')$ ).

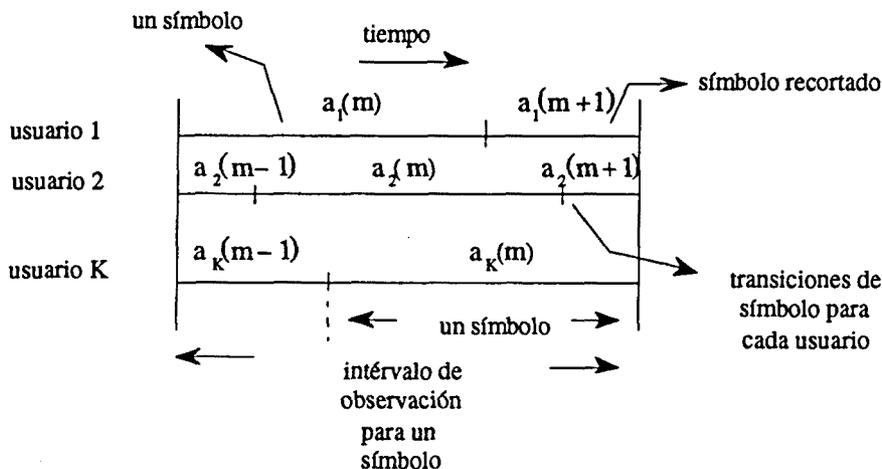


Figura 2.6.3.1: Snapshot conteniendo todos los símbolos que se deben considerar para el procedimiento de detección. Nótese que  $a_1(m-1)$  y  $a_K(m+1)$  no están presentes en la ventana de observación y que  $a_2(m-1)$  hasta  $a_K(m-1)$  así como  $a_1(m+1)$  hasta  $a_{K-1}(m+1)$  se han recortado según los límites de la ventana.

Introduciremos ahora un artificio matemático para el subsiguiente análisis de prestaciones del ADD. Tomemos la ecuación 2.6.3.9 y reordenemos los productos definiendo las nuevas matrices de correlación modificadas por la respuesta del array. Resulta fácil ver que el producto de Schur de la matriz de correlación temporal con el Grammiano de la matriz de

steering se puede expresar en función de las siguientes matrices diagonales de steering como,

$$\mathbf{R}(m, m') \bullet (\mathbf{S}^H \mathbf{S}) = \sum_{q=1}^{q=N_s} \mathbf{D}_q^H \mathbf{R}(m, m') \mathbf{D}_q \quad , \quad \mathbf{D}_q = \text{diag}(s_q) \tag{2.6.3.12}$$

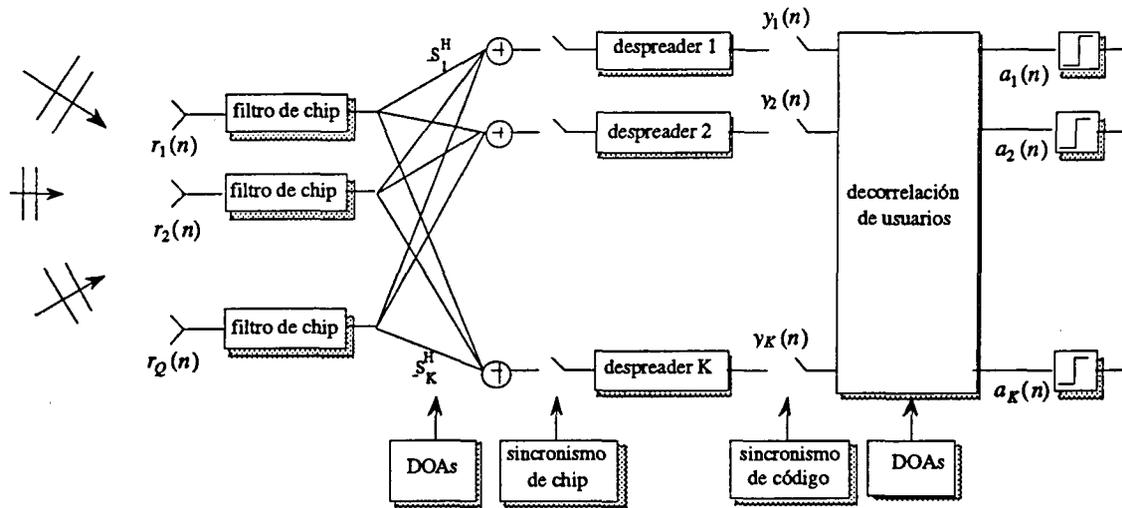


Figura 2.6.3.2: Representación de las tres etapas involucradas en el ADD: combinación espacial, banco de filtros adaptados y decorrelación.

### 2.6.4 Análisis de Prestaciones

En este apartado demostraremos que la mejora en la decorrelación conseguida por el ADD resulta en una mejora de la probabilidad de error asociada a cada usuario. Por tanto, necesitamos deducir la matriz de covarianza del detector.

Cabe recalcar que la utilización de diversos sensores incrementa la relación  $E_b/N_o$  gracias a un simple promediado (coherente con la fase de los vectores de steering, ver figura 2.6.4.1.b) incluso cuando todas las señales inciden en el array desde la misma dirección. Sólo en este caso, ninguna mejora es posible con respecto al ADD(1). Llámemos  $\sigma_{1k}$  a la varianza del ruido a la entrada del detector de símbolo del usuario k-ario. Tendremos entonces que la probabilidad de error  $p(\epsilon)$  viene dada por,

$$\sigma_{N_s, k}^2 = \frac{1}{N_s} \sigma_{1, k}^2 \quad , \quad p_{N_s, k}(\epsilon) = Q\left(\frac{A_k}{\sigma_{N_s, k}}\right) \leq p_{1, k}(\epsilon) = Q\left(\frac{A_k}{\sigma_{1, k}}\right) \tag{2.6.4.1}$$

Esto se deduce fácilmente de la ecuación 2.6.3.12 cuando las matrices diagonales de steering se convierten en la matriz identidad. En el caso más general de usuarios distribuidos aleatoriamente, se demostrará que la probabilidad de error mejora con respecto a la de 2.6.4.1 a causa de una disminución de la varianza del ruido de detección,

$$\sigma_{N_s, k}^2 \leq \frac{1}{N_s} \sigma_{1, k}^2, \quad p_{N_s, k}(\varepsilon) = Q\left(\frac{A_k}{\sigma_{N_s, k}}\right) \leq p_{1, k}(\varepsilon) = Q\left(N_s \frac{A_k}{\sigma_{1, k}}\right) \quad (2.6.4.2)$$

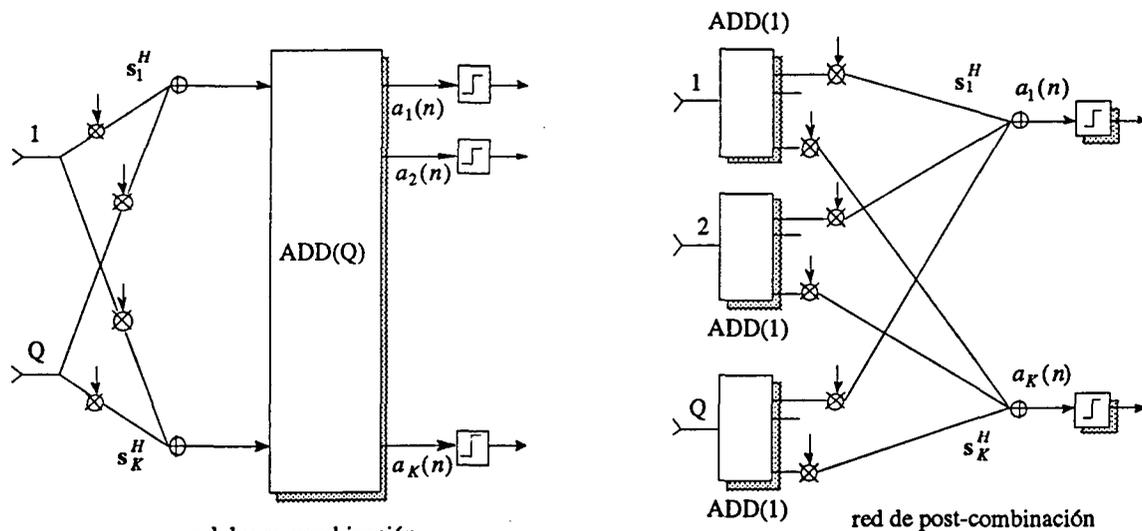
La demostración sigue por construcción. Las ecuaciones del ADD se han deducido directamente del principio de máxima verosimilitud. Por consiguiente, la matriz de covarianza asociada llega al límite de Cramer-Rao. Cualquier estimador diferente al ADD poseerá una matriz de covarianza mayor en términos de su definición positiva. Aquí propondremos un estimador ad-hoc subóptimo tal que sus elementos de la diagonal sean iguales a  $(\sigma_{1k})^2/N_s$  demostrando así la afirmación en 2.6.4.2.

El detector ad-hoc en cuestión será tal que solo aproveche la decorrelación temporal entre usuarios, en contraste al ADD, basado también en la decorrelación espacial. Ambas estrategias de detección se representan en las figuras 2.6.4.1.a-b. Mientras que en el decorrelador óptimo, la combinación espacial se realiza con anterioridad a la decorrelación temporal, en el caso ad-hoc, los papeles se invierten y la combinación espacial se realiza con posterioridad a la decorrelación temporal. El carácter subóptimo del decorrelador ad-hoc es palpable por el hecho de que no puede decorrelar señales que utilicen el mismo código y sean completamente sincronas cuando éstas llegan por direcciones diferentes. En cambio, el ADD no presenta este problema.

A continuación calculamos la matriz de covarianza del estimador. Consideremos uno de los sensores, de tal forma que las mejores estimaciones posibles de los símbolos basados sólo en el conocimiento de la señal recibida en el sensor escogido vienen dados por,

$$\tilde{\mathbf{a}}_q = \Gamma_{ADD(1)}^{-1} \tilde{\mathbf{y}}_q \quad (2.6.4.3)$$

Cuando este procedimiento se aplica a la salida de cada sensor, el decorrelador final se puede construir combinando adecuadamente los símbolos estimados de un determinado usuario proporcionados por los sensores individuales. De acuerdo con la figura 2.6.3.1,



red de pre-combinación  
 Figura 2.6.4.1: [a] ADD de Ns sensores [b] Decorrelador ad-hoc Ns sensores

este método puede interpretarse como un promediado de las maximizaciones individuales de las funciones de densidad de probabilidad de cada sensor, siendo el procedimiento óptimo la maximización de la función de densidad conjunta de todos los sensores. Los componentes de  $\mathbf{a}_q$  se obtienen de,

$$\min_{\alpha_k(m) \in \tilde{\mathbf{a}}_q} \sum_{n=-N}^{n=+N} \hat{e}_q^* \hat{e}_q, \quad e_q(n) = r_q(n) - \sum_{k=1}^{k=N_s} \sum_{m=-\infty}^{m=+\infty} \alpha_k(m) p_k(n - mT) \quad (2.6.4.4)$$

con  $\alpha_k(m) = [\mathbf{s}_k]_q a_k(m)$

Debe tenerse en cuenta que ahora los símbolos estimados a la salida de los ADD(1)'s colocados en cada sensor están afectados por la fase de los componentes de los vectores de steering correspondientes. Es decir, los símbolos estimados para el usuario k-ario por el ADD(1) del sensor q-ario están afectados por la fase de  $[\mathbf{s}_k]_q$  debido al desfase del frente de onda para ese sensor. Por tanto, el promediado de los estimandos se debe realizar coherentemente con estas fases residuales de la siguiente forma,

$$a'_k(m) = \frac{1}{N_s} \sum_{q=1}^{q=N_s} [\mathbf{s}_k]_q^* \alpha_k(m) \quad (2.6.4.5)$$

o equivalentemente,

$$\tilde{\mathbf{a}}' = \frac{1}{N_s} \sum_{q=1}^{q=N_s} \tilde{\mathbf{D}}_q^H \tilde{\mathbf{a}}_q, \quad \tilde{\mathbf{D}}_q = \text{diag}\{\mathbf{D}_q, \dots, \mathbf{D}_q\} \Rightarrow \tilde{\mathbf{a}}' = \frac{1}{N_s} \sum_{q=1}^{q=N_s} \tilde{\mathbf{D}}_q^H \Gamma_{ADD(1)}^{-1} \tilde{\mathbf{y}}_q \quad (2.6.4.6)$$

Después de estas consideraciones, la matriz de covarianza asociada puede deducirse a partir de la matriz de correlación del ruido de estimación  $\{v\}$ . Definimos  $\{\eta_q\}$  como el ruido a la salida del banco de filtros adaptados del sensor q-ario y  $\{v_q\}$  como el ruido de estimación marginal. Obtenemos,

$$v_q = \tilde{D}_q^H \Gamma_{ADD(1)}^{-1} \eta_q \Rightarrow v' = \frac{1}{N_s} \sum_{q=1}^{q=N_s} v_q \tag{2.6.4.7}$$

Basándonos en la correlación nula de los procesos de ruido marginales, la matriz de covarianza del estimador puede deducirse a partir de la matriz de correlación del ruido de estimación como,

$$\begin{aligned} \Gamma'^{-1} = E v' v'^H &= \frac{1}{N_s^2} \sum_{k,k'=1}^{N_s} \tilde{D}_q^H \Gamma_{ADD(1)}^{-1} (E \eta_k \eta_{k'}^H) \Gamma_{ADD(1)}^{-1} \tilde{D}_q = \\ &= \frac{1}{N_s^2} \sum_{k=1}^{k=N_s} \tilde{D}_q^H \Gamma_{ADD(1)}^{-1} (E \eta_k \eta_k^H) \Gamma_{ADD(1)}^{-1} \tilde{D}_q = \frac{1}{N_s^2} \sum_{k=1}^{k=N_s} \tilde{D}_q^H \Gamma_{ADD(1)}^{-1} \tilde{D}_q \end{aligned} \tag{2.6.4.8}$$

Cuando evaluamos los elementos de la diagonal de la matriz de covarianza, encontramos que son iguales a  $1/N_s$  veces la varianza de ruido para el caso de un único sensor,

$$\Gamma'^{-1} = \frac{1}{N_s^2} \sum_{k=1}^{k=N_s} \tilde{D}_q^H \Gamma_{ADD(1)}^{-1} \tilde{D}_q \Rightarrow [\Gamma'^{-1}]_{kk} = \sigma_k'^2 = \frac{1}{N_s} \sigma_{1,k}^2 \tag{2.6.4.9}$$

Por Teoría de la Estimación sabemos que la matriz de covarianza de este detector ad-hoc es más definida positiva que la del ADD (detector óptimo). Por tanto, su diferencia debe ser definida positiva y los elementos de la diagonal de dicha matriz diferencia deben ser mayores que cero para mantener la positividad. En consecuencia, la aseveración 2.6.4.2 queda demostrada. La desigualdad entre la matriz de covarianza del ADD (la cual es igual a la matriz de decorrelación como puede demostrarse partiendo de la condición de ruido blanco) y la del detector ad-hoc conduce a,

$$\Gamma_{ADD(N_s)}^{-1} = \left( \sum_{q=1}^{q=N_s} \tilde{D}_q^H \Gamma_{ADD(1)}^{-1} \tilde{D}_q \right)^{-1} \leq \frac{1}{N_s^2} \sum_{q=1}^{q=N_s} \tilde{D}_q^H \Gamma_{ADD(1)}^{-1} \tilde{D}_q \Rightarrow [\Gamma_{ADD(N_s)}^{-1}]_{kk} = \sigma_{N_s,k}^2 \leq \frac{1}{N_s} \sigma_{1,k}^2 \tag{2.6.4.10}$$

El carácter óptimo del ADD sigue del hecho que combina la información proporcionada por cada sensor de la mejor forma posible. Una posible justificación proviene del hecho que la matriz de información de Fisher asociada es expresable como suma de las matrices de

información individuales. Por otra parte, la matriz de información del detector ad-hoc cumple una relación diferente con las matrices de información individuales,

$$\Gamma_{CR}^{-1} = \left( \sum_{q=1}^{q=N_s} \Gamma_q^{-1} \right)^{-1} \leq \Gamma_{ad-hoc}^{-1} = \frac{1}{N_s^2} \sum_{q=1}^{q=N_s} \Gamma_q^{-1} \quad (2.6.4.11)$$

Nótese que la ecuación 2.6.3.11 resulta ser una generalización de la ecuación 2.6.3.10. El único requisito es que el conjunto formado por  $\{\Gamma_q, 1 \leq q \leq N_s\}$  sean matrices de información de Fisher (las cuales tienen la estructura de una matriz de Gram).

Resulta ilustrativo considerar cuales son las prestaciones del ADD comparadas con las del detector de usuario único en ausencia de otros usuarios en el canal. La decorrelación de la interferencia de múltiple acceso con el usuario de interés eleva la potencia de ruido a la entrada del detector. Se ha demostrado en esta sección que la mejora en la decorrelación proporcionada por un detector en array produce una varianza de ruido menor a la entrada del detector. La completa ortogonalidad entre los códigos de usuario representa la máxima decorrelación posible y conduce a un desacoplo de las ecuaciones. En ese caso, ningún usuario se ve afectado por la presencia simultánea de otros en el canal y las prestaciones son máximas. Para este caso particular, el ADD se reduce a un banco de detectores de usuario único. Por lo tanto, y como puede colegirse de la ecuación 2.6.2.10, el ADD presenta unas prestaciones mas cercanas al caso de interferencia de múltiple acceso inexistente.

Consideremos la matriz de información de Fisher obtenida para el ADD. Es nuestra intención ahora establecer una cota inferior para los elementos de la diagonal de  $\Gamma^{-1}$ .  $\Gamma^{-1}$  puede considerarse como el Grammiano de un conjunto equivalente de formas de onda. A su vez, podemos suponer un array equivalente donde las señales correspondientes llegan de direcciones tales que sus vectores de steering forman un conjunto ortonormal, es decir,  $S^H S = I$ . La aplicación de la ecuación 2.6.3.11 a este caso conduce a,

$$\Gamma'' \stackrel{def}{=} \frac{1}{N_s} \sum_{q=1}^{q=N_s} \tilde{D}_q''^H \Gamma \tilde{D}_q'' \Rightarrow \Gamma''^{-1} \leq \frac{1}{N_s} \sum_{q=1}^{q=N_s} \tilde{D}_q''^H \Gamma^{-1} \tilde{D}_q'' \quad (2.6.4.12)$$

pero como el conjunto  $\{\tilde{D}_q\}$  se ha obtenido de un conjunto ortogonal de vectores de steering y el sumatorio de la ecuación 2.6.3.12 se puede expresar como el producto de Schur con el Grammiano de la matriz de steering, se sigue que  $\Gamma''$  se diagonaliza,

$$\Gamma''^{-1} = I \leq \text{diag}(\Gamma^{-1}) \Rightarrow [\Gamma^{-1}]_{kk} \geq 1 \quad (2.6.4.13)$$

como deseábamos demostrar. La igualdad se cumple para Grammiano que ya son ellos mismos diagonales. Con esta última demostración podemos acotar inferiormente y superiormente la varianza del ruido de estimación,

$$1 \leq \sigma_{N_s, k}^2 \leq \frac{1}{N_s} \sigma_{1, k}^2 \tag{2.6.4.14}$$

donde tanto  $\sigma_{N_s, k}$  como  $\sigma_{1, k}$  se han normalizado a la varianza de ruido del detector en el caso de usuario único.

### 2.6.5 El Caso Límite. Probabilidad de Error Asintótica

En el caso de usuarios síncronos, los símbolos estimados pueden desacoplarse símbolo a símbolo. En estas circunstancias, la matriz  $\Gamma_{ADD(N_s)}$  se convierte en diagonal a bloques. En el caso más general de usuarios asíncronos, la influencia de los símbolos pasados y futuros queda representada por la tri-diagonalidad de  $\Gamma_{ADD(N_s)}$ . Se puede utilizar un número arbitrario de observaciones para la estimación de un símbolo dado,  $a_k(m)$ , puesto que no es posible desacoplar las ecuaciones. Podemos preguntarnos entonces si la calidad de los estimandos aumenta con la longitud del intervalo de observación. Tal aseveración puede demostrarse como sigue: sea  $\Gamma^{-1}(1, N)$  el decorrelador de  $N$  símbolos que produce las estimaciones  $\hat{a}(1)$ ,  $\hat{a}(2)$  hasta  $\hat{a}(N)$ . Análogamente sea  $\Gamma^{-1}(N+1, M)$  el decorrelador de los  $M$  símbolos consecutivos  $\hat{a}(N+1)$ ,  $\hat{a}(N+2)$  to  $\hat{a}(M)$ . Por consiguiente, podemos construir un posible decorrelador de la siguiente forma,

$$\Gamma_1^{-1} = \begin{cases} \Gamma^{-1}(1, N) & 1 \leq m \leq N \\ \Gamma^{-1}(N+1, M) & N+1 \leq m \leq M \end{cases} \tag{2.6.5.1}$$

Sea  $\Gamma_2^{-1} = \Gamma^{-1}(1, N+M)$  el decorrelador del conunto completo de  $N+M$  símbolos que produce estimaciones desde  $\hat{a}(1)$  thasta  $\hat{a}(M)$ . Por construcción,  $\Gamma_2^{-1}$  es el mejor decorrelador de los símbolos contenidos en el intervalo  $[1, M]$  y la varianza asociada atañe el límite de Cramer-Rao. Por tanto, debemos tener que,

$$[\Gamma_2^{-1}]_{kk} \leq [\Gamma_1^{-1}]_{kk} \tag{2.6.5.2}$$

lo cual demuestra que cuando se amplía el intervalo, por ejemplo de  $[1, N]$  hasta  $[1, N+M]$ , la varianza del ruido asociado a cada símbolo debe disminuir. En particular, podemos construir una secuencia de estimadores para un símbolo dado, cuyas respectivas covarianzas forman una sucesión monótonamente decreciente. Puesto que la covarianza es no-negativa, dicha sucesión está acotada inferiormente por cero y por tanto existe un límite de la sucesión cuando  $N$  tiende a infinito. Dicho límite determina la mínima probabilidad de error obtenible bajo esta clase de estimadores.

Se demostró en [Verdú] que cuando el intervalo de observación tiende a  $(-\infty, +\infty)$ , el detector de decorrelación puede implementarse como un filtro vectorial no-causal invariante en el tiempo cuya matriz de transferencia viene dada por,

$$\mathbf{K}_{N_s}^{-1}(z^{-1}) = \left[ \sum_{m=-1}^{m=+1} \bar{\mathbf{R}}(m) z^{-m} \right]^{-1} \quad (2.6.5.3)$$

donde el subíndice de  $\mathbf{K}$  denota el número de sensores. Aunque la implementación exacta de este detector es irrealizable debido a su carácter no-causal y de longitud infinita, su análisis proporciona la probabilidad de error límite que puede conseguirse con técnicas de decorrelación. Las simulaciones han demostrado que para códigos con buenas propiedades de correlación, la convergencia a este límite se logra en el orden de unos pocos símbolos.

La existencia y estabilidad del decorrelador asintótico se puede formular en función de las raíces del determinante de la matriz de transferencia,  $\det \mathbf{K}_{N_s}(z^{-1})$  [Verdú]. Resulta ilustrativo observar cómo se estiman los símbolos de un usuario individual. Sean  $\mathbf{y}(z^{-1})$  y  $\mathbf{a}(z^{-1})$  las transformadas  $z$  de la salida del banco de filtros adaptados y de los símbolos estimados, respectivamente. Entonces, en términos de los determinantes adjuntos de  $\mathbf{K}_{N_s}$  tenemos que,

$$a_k(z^{-1}) = \frac{1}{\det \mathbf{K}_{N_s}(z^{-1})} \sum_{i=1}^{i=N_s} \text{adj}_{k,i} \mathbf{K}_{N_s}(z^{-1}) y_i(z^{-1}) \quad (2.6.5.4)$$

como puede obtenerse fácilmente de la descomposición de la ecuación 2.6.5.3 en determinantes. Cuando interpretamos esta ecuación como un filtro, debe cumplirse que  $\det \mathbf{K}_{N_s}(e^{-j\omega}) \neq 0$ , para todo  $\omega$  en  $[-\pi, +\pi]$ . Es decir, para garantizar la estabilidad, ninguna raíz puede hallarse en el círculo unidad. Ya que estamos manejando secuencias no-causales, no es necesario imponer ninguna otra restricción sobre la localización de los polos de la función de transferencia. Si los polos se encuentran fuera del círculo unidad, la región de

convergencia de  $[\det \mathbf{K}_{N_s}(z^{-1})]^{-1}$  estará limitada a una corona circular que contenga el círculo unidad.

- Proposición 2.6.5.1: si  $\lambda_i$  es raíz de  $\det \mathbf{K}_{N_s}(z^{-1})$ ,  $1/\lambda_i^*$  también lo es.

◊ Demostración:

$$\det \mathbf{K}_{N_s}(\lambda_i^{-1}) = 0 \Rightarrow [\det \mathbf{K}_{N_s}(\lambda_i^{-1})]^H = \det \mathbf{K}_{N_s}(\lambda_i^{-1})^H = \det \mathbf{K}_{N_s}(\lambda_i^*) = 0 \quad (2.6.5.5)$$

donde las raíces en el interior del círculo unidad corresponden a la parte causal del filtro y las raíces exteriores a la parte no-causal. ◊

- Proposición 2.6.5.2: si el decorrelador asintótico para un único sensor,  $\mathbf{K}_1(z^{-1})$ , existe y es estable,  $\mathbf{K}_{N_s}(z^{-1})$  también existe y es estable para el caso de  $N_s$  sensores.

◊ Demostración: La estabilidad de  $\mathbf{K}_{N_s}(z^{-1})$  es equivalente a la condición de que  $\mathbf{K}_{N_s}(e^{-j\omega})$  sea estrictamente definida positiva en el círculo unidad:  $\mathbf{K}_{N_s}(e^{-j\omega}) > 0$  para todo  $\omega$  en  $[-\pi, +\pi]$ . ( Podemos asegurar que es positiva  $\mathbf{K}_{N_s}(e^{-j\omega}) \geq 0$ , ya que tiene la estructura de una matriz de Gram). Si para algún  $\omega_0$ , tenemos que  $\det \mathbf{K}_{N_s}(e^{-j\omega_0}) = 0$ , como mínimo un valor propio de  $\mathbf{K}_{N_s}(e^{-j\omega_0})$  es cero y el decorrelador asintótico no existe. Pasamos a demostrar ahora que  $\mathbf{K}_{N_s}(e^{-j\omega})$  es estrictamente definida positiva. Tenemos que,

$$\mathbf{K}_{N_s}(e^{-j\omega}) = \sum_{k=1}^{k=N_s} \mathbf{D}_k^H \mathbf{K}_1(e^{-j\omega}) \mathbf{D}_k, \quad \mathbf{n}^H \mathbf{S}(e^{-j\omega}) \mathbf{n} > 0 \quad \forall \mathbf{n}, \omega \quad (2.6.5.6)$$

lo cual conduce a,

$$\mathbf{n}^H \mathbf{K}_{N_s}(e^{-j\omega}) \mathbf{n} = \sum_{k=1}^{k=N_s} [\mathbf{D}_k \mathbf{n}]^H \mathbf{K}_1(e^{-j\omega}) [\mathbf{D}_k \mathbf{n}] > 0 \quad \forall \mathbf{n}, \omega \quad (2.6.5.7)$$

quedando así demostrada la Proposición 2.6.5.2. ◊

No obstante, la inversa a la Proposición 2.6.5.2 no se cumple. Tomemos un sistema hipotético donde dos de los usuarios comparten la misma secuencia PN y transmiten sincronamente. Un detector de un único sensor resultaría inútil ya que el decorrelador asociado sería deficiente en rango, aunque la información podría recuperarse suponiendo que los usuarios se encontrasen en direcciones ortogonales.

• Proposición 2.6.5.3: la velocidad de convergencia del ADD al límite asintótico en términos de la longitud de la ventana de observación queda determinada por la raíz de  $\det \mathbf{K}_{N_s}(z^{-1})$  más próxima al círculo unidad.

◊ Demostración: cuando  $[\det \mathbf{K}_{N_s}(z^{-1})]^{-1}$  se interpreta como un filtro, la longitud efectiva de su respuesta impulsional está determinada por el polo más cercano al círculo unidad. Esto es evidente a partir de la descomposición en fracciones simples de la respuesta del filtro. cuando calculamos la expansión en serie de Laurent de cada fracción simple en la zona de convergencia, la que se atenúa más lentamente corresponde al polo más cercano al círculo unidad. De este modo, símbolos que se encuentren separados una longitud superior que la de la longitud efectiva de la respuesta impulsional del símbolo de interés, tendrán poca influencia en su estimación. ◊

Nótese que para esta formulación, la función de correlación del vector de ruido se convierte en una secuencia de matrices en la variable compleja  $z^{-1}$  (tomada aquí como un operador de retardo entre vectores). Por consiguiente, la función de correlación del ruido a continuación del banco de filtros adaptados,  $\mathbf{R}_{\eta\eta}$ , se puede expresar como,

$$\mathbf{R}_{\eta\eta}(z^{-1}) = \sum_{m=-1}^{m=+1} \bar{\mathbf{R}}(m)z^{-m} \tag{2.6.5.8}$$

la correlación del ruido a la salida del ADD,  $\mathbf{R}_{vv}$ , se expresa como,

$$\mathbf{R}_{vv}(z^{-1}) = \mathbf{K}_{N_s}^{-1}(z^{-1})\mathbf{R}_{\eta\eta}(z^{-1})\mathbf{K}_{N_s}^{-H}(z^{-1}) = \mathbf{K}_{N_s}^{-1}(z^{-1}) \tag{2.6.5.9}$$

donde los elementos de la diagonal constituyen la secuencia de correlación del ruido de estimación de los símbolos dados. El elemento central de la secuencia constituye la covarianza de la estimación de los símbolos. La aplicación de la teoría de la transformada  $z$  conduce a,

$$\sigma_k^2 = \frac{1}{2\pi j} \oint_{C_1} [\mathbf{K}_{N_s}^{-1}(z^{-1})]_{kk} z^{-1} dz = \frac{1}{2\pi} \oint_{C_1} [\mathbf{K}_{N_s}^{-1}(e^{-j\omega})]_{kk} d\omega \tag{2.6.5.10}$$

• Proposición 2.6.5.4:

$$\mathbf{K}_{N_s}^{-1}(e^{-j\omega}) \leq \frac{1}{N_s^2} \sum_{k=1}^{k=N_s} \mathbf{D}_k^H \mathbf{K}_1^{-1}(e^{-j\omega}) \mathbf{D}_k = \frac{1}{N_s^2} \mathbf{K}_1^{-1}(e^{-j\omega}) \bullet (\mathbf{S}^H \mathbf{S}) , \quad \forall \omega \in [-\pi, +\pi] \tag{2.6.5.11}$$

◊ Demostración: véase la ecuación 2.6.3.11. Las matrices de dicha ecuación tienen la misma estructura para todo  $z$  en el círculo unidad. ◊

Esta propiedad puede utilizarse para demostrar la siguiente proposición:

• Proposición 2.6.5.5: las densidades espectrales de potencia del ruido de estimación del usuario  $k$ -ario para los casos de sensor único y múltiples sensores están relacionadas según la siguiente desigualdad.

$$[\mathbf{K}_1^{-1}(e^{-j\omega})]_{kk} \geq N_s [\mathbf{K}_{N_s}^{-1}(e^{-j\omega})]_{kk} \geq \frac{1}{2\pi}, \quad \forall \omega \in [-\pi, +\pi] \quad (2.6.5.12)$$

◊ Demostración: demostremos primero la cota superior. La evaluación de la ecuación 2.6.5.11 en la diagonal de  $\mathbf{K}_{N_s}^{-1}(e^{-j\omega})$  produce,

$$[\mathbf{K}_{N_s}^{-1}(e^{-j\omega})]_{kk} \leq \left[ \frac{1}{N_s^2} \sum_{k=1}^{k=N_s} \mathbf{D}_k^H \mathbf{K}_1^{-1}(e^{-j\omega}) \mathbf{D}_k \right]_{kk} = \frac{1}{N_s^2} \sum_{k=1}^{k=N_s} [\mathbf{K}_1^{-1}(e^{-j\omega})]_{kk} = \frac{1}{N_s} [\mathbf{K}_1^{-1}(e^{-j\omega})]_{kk} \quad (2.6.5.13)$$

La cota inferior corresponde a la densidad espectral de potencia del ruido de estimación para el caso de usuarios síncronos con secuencias ortogonales (ruido blanco de varianza normalizada 1 como puede verse en la ecuación 2.6.4.14). Por lo tanto, la cota inferior se obtiene cuando dividimos esta varianza por el ancho de banda total normalizado  $2\pi$ . La ecuación 2.6.5.12 resulta ser una desigualdad más estricta que la ecuación 2.6.4.12. El  $\text{ADD}(N_s)$  mejora las características espectrales del ruido de estimación en todas las frecuencias con respecto al  $\text{ADD}(1)$ . ◊

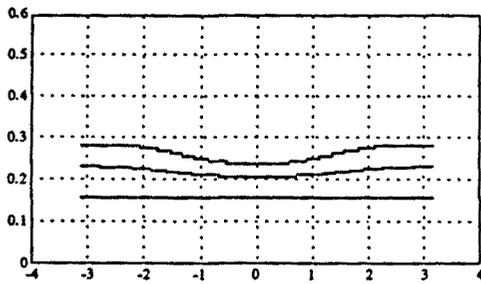


Figura 2.6.5.2.a: Usuario 1

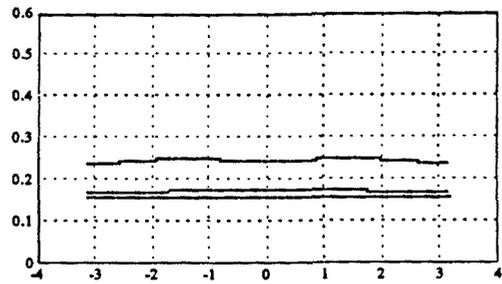


Figura 2.6.5.2.b: Usuario 2

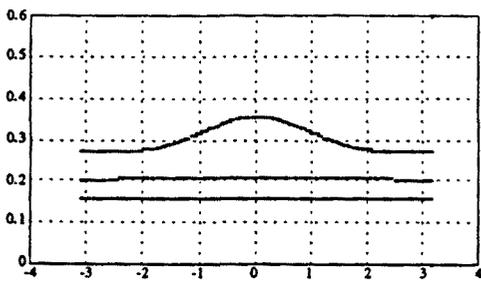


Figura 2.6.5.2.c: Usuario 3

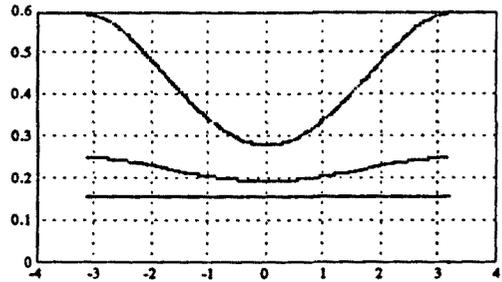


Figura 2.6.5.2.d: Usuario 4

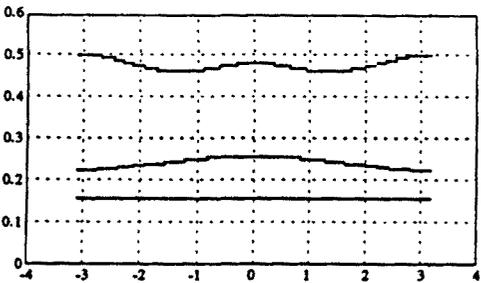


Figura 2.6.5.2.e: Usuario 5

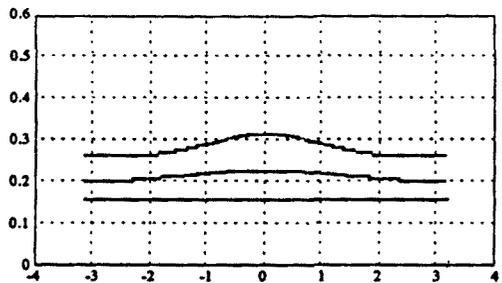


Figura 2.6.5.2.f: Usuario 6

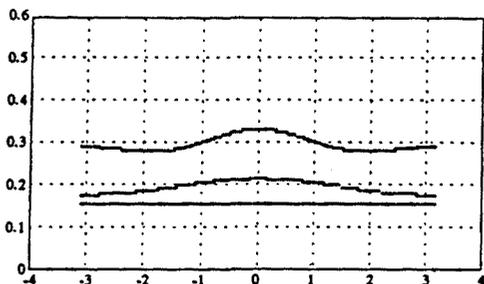


Figura 2.6.5.2.g: Usuario 7

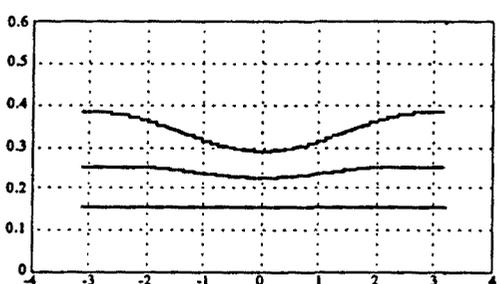


Figura 2.6.5.2.h: Usuario 8

Figura 2.6.5.2: Comparación de los espectros de potencia de ruido en detección para los 8 usuarios de la figura 2.6.5.1 a la salida del decorrelador. Cada uno de los espectros corresponde a: 1) Decorrelador de sensor único. 2) Decorrelador de dos sensores (el espectro de potencia se ha multiplicado por dos). 3) Un único sensor y un único usuario. Puede comprobarse como se cumplen las igualdades deducidas en la sección 2.6.5.

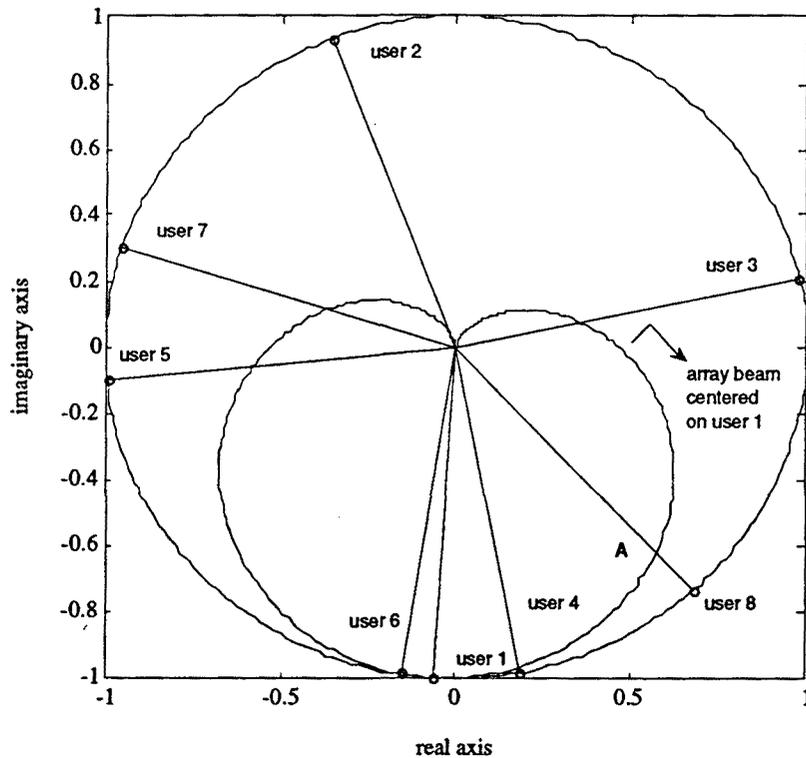


Figura 2.6.5.1: Representación de las direcciones de llegada (DOA) de 8 usuarios para un escenario determinado. Las DOAs se representan como un fasor cuyo ángulo es el desfase entre sensores experimentado por el frente de onda que incide en el array. Las DOAs se han representado mediante "o"s sobre el círculo unidad. Los ángulos entre 0 y  $\pi$  corresponden a las direcciones broadside y endfire, respectivamente. Se ha superpuesto el diagrama de haz con respecto al desfase entre sensores, con la máxima ganancia normalizada a uno y enfocada sobre el usuario 1. La longitud de la intersección de las líneas de DOA con el diagrama de array centrado en el usuario de interés, es igual al producto escalar entre los vectores de steering respectivos, normalizado al número de sensores (la distancia entre el punto A y el origen, para el usuario 8), o los componentes del Grammiano de la matriz  $S$ . Según puede apreciarse en la figura, cuando el array enfoca al suuario 1, únicamente el usuario 2 queda razonablemente filtrado en la etapa de combinación espacial. El ADD se encargará de eliminar la influencia de los usuarios restantes.

### 2.6.6 Apéndice A. Criterios de Detección.

En este apéndice demostraremos como los criterios de detección perfilados en la sección 2.6.2 se deducen a partir del criterio de Máxima Verosimilitud. Sea  $p_n(\mathbf{n};N)$  la función de densidad conjunta de los procesos de ruido Gaussiano blanco recibidos en cada uno de los sensores para un intervalo temporal de  $N$  muestras. Supondremos que el nivel de ruido en cada uno de los sensores es el mismo. De este modo, la función de densidad puede expresarse de la siguiente forma,

$$(2.6.6.1)$$

$$p_N(\mathbf{n}) = \prod_{m=-N'}^{m=+N'} \frac{1}{\pi^{N_s} \sigma^{2N_s}} \exp\left(-|\mathbf{n}_m|^2 / \sigma^2\right)$$

a la que llamaremos la factorización temporal de la función de densidad. La factorización espacial se deduce de la condición de incorrelación entre los procesos de ruido en los sensores y consiste en,

$$p_N(\mathbf{n}) = \prod_{q=1}^{q=N_s} \frac{1}{\pi^N \sigma^{2N}} \exp\left(-\sum_{m=-N'}^{m=+N'} |n_{m,q}|^2 / \sigma^2\right), \quad N = 2N' + 1 \quad (2.6.6.2)$$

El criterio de detección consiste entonces en la maximización de la siguiente función,

$$\{\mathbf{a}\} = \arg \max p_N(\mathbf{r} - \mathbf{s}(\mathbf{a})) \quad (2.6.6.3)$$

donde  $\mathbf{s}(\mathbf{a})$  constituye el modelo de la señal recibido por el array para una secuencia determinada de símbolos. El criterio de maximización de la función de densidad evaluada en los datos se reduce a la minimización del exponente,

$$\max_{\mathbf{a}} \prod_{m=-N'}^{m=+N'} \frac{1}{\pi^{N_s} \sigma^{2N_s}} \exp\left(-|\mathbf{r}_m - \mathbf{s}_m(\mathbf{a})|^2 / \sigma^2\right) = \max_{\mathbf{a}} \frac{1}{\pi^{N_s N} \sigma^{2N_s N}} \exp\left(-\sum_{m=-N'}^{m=+N'} |\mathbf{r}_m - \mathbf{s}_m(\mathbf{a})|^2 / \sigma^2\right) \quad (2.6.6.4)$$

o sea,

$$\min_{\mathbf{a}} \sum_{m=-N'}^{m=+N'} |\mathbf{r}_m - \mathbf{s}_m(\mathbf{a})|^2 \quad (2.6.6.5)$$

Este criterio es invariante si se normaliza la norma del error por la dimensión de la ventana temporal de observación. Es fácil ver entonces que el criterio óptimo de decisión para una longitud infinita de datos viene dado según el operador esperanza de la función de error,

$$\lim_{N \rightarrow \infty} \min_{\mathbf{a}} \frac{1}{N} \sum_{m=-N'}^{m=+N'} |\mathbf{r}_m - \mathbf{s}_m(\mathbf{a})|^2 = \min_{\mathbf{a}} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{m=-N'}^{m=+N'} |\mathbf{r}_m - \mathbf{s}_m(\mathbf{a})|^2 = \min_{\mathbf{a}} E|\mathbf{r} - \mathbf{s}(\mathbf{a})|^2 \quad (2.6.6.6)$$

con el supuesto de que puedan intercambiarse el límite y el proceso de minimización (siempre que éste sea continuo).

A partir de las factorizaciones anteriormente mencionadas pueden derivarse algoritmos de detección subóptimos. Consideremos la ecuación de la factorización temporal. Podemos ver que,

$$p_N(\mathbf{r}-\mathbf{s}) = \prod_{q=1}^{N_s} p(r_q - s_q) \tag{2.6.6.7}$$

el algoritmo subóptimo se define entonces como una combinación lineal de los argumentos que maximizan cada una de las funciones de densidad marginales de cada sensor. Los coeficientes de dicha combinación lineal dependerán de los retardos de fase de cada frente de onda respectivo.

$$\mathbf{a} = \sum_{q=1}^{q=N_s} w_q \underset{\mathbf{a}}{\arg \max} p(r_q - s_q(\mathbf{a})) \tag{2.6.6.8}$$

Puede demostrarse que este algoritmo es simplemente la suma de componentes del vector de datos proyectado sobre subespacios del espacio global de señal.

### 2.6.7 Apéndice B. Ruido espacialmente coloreado

En la determinación de los criterios de detección en el Apéndice A se supuso que los procesos de ruido estaban incorrelados tanto espacial como temporalmente. Un caso más general y acorde con la realidad consiste en suponer el ruido espacialmente coloreado con matriz de correlación  $\mathbf{R}$ . La factorización espacial de la función de densidad del ruido viene entonces dada por,

$$p_N(\mathbf{n}) = \prod_{m=-N'}^{m=+N'} \frac{1}{\pi^{N_s} \det \mathbf{R}_n} \exp(-\mathbf{n}_m^H \mathbf{R}_n^{-1} \mathbf{n}_m) \tag{2.6.7.1}$$

El criterio asintótico equivalente de detección se expresa entonces de la siguiente forma,

$$\underset{\mathbf{a}}{\min} E(\mathbf{r}-\mathbf{s}(\mathbf{a}))^H \mathbf{R}_n^{-1} (\mathbf{r}-\mathbf{s}(\mathbf{a})) \tag{2.6.7.2}$$

Podemos ver que es equivalente a la minimización de la potencia del error a la salida de un filtro. Si consideramos un enventanado finito de los datos, entonces el criterio de minimización nos conduce a la siguiente expresión,

$$(2.6.7.3) \quad \frac{\partial}{\partial \mathbf{a}^H(m')} L_{N'} = - \sum_{n=-N'}^{n=+N'} \frac{\partial \mathbf{r}_a^H(n)}{\partial \mathbf{a}^H(m')} \mathbf{R}_n^{-1} \mathbf{e}_r(n) = 0$$

con lo cual y siguiendo los pasos ya expuestos en la sección 2.6.2 llegamos a la siguiente expresión para la secuencia de estadísticos suficientes,

$$(2.6.7.4) \quad \mathbf{b}(m') = \sum_{n=-N'}^{n=+N'} (\mathbf{S}^H \mathbf{R}_n^{-1} \mathbf{r}(n)) \bullet \mathbf{p}(n - m'N)$$

los cuales estan relacionados con los símbolos estimados según la ecuación,

$$(2.6.7.5) \quad \mathbf{b}(m') = \sum_{m=-\infty}^{m=+\infty} (\mathbf{R}(m, m') \bullet (\mathbf{S}^H \mathbf{R}_n^{-1} \mathbf{S})) \mathbf{a}(m)$$

Obsérvese ahora que la etapa de conformación de haz presenta una multiplicación previa por la inversa de la matriz de correlación del ruido a fin de decorrelar (blanquear) el ruido espacial.  $\mathbf{R}_n^{-1}$  vuelve a aparecer en la expresión del decorrelador tras el banco de filtros adaptados debido a que se ha modificado la ley de combinación de las señales en los sensores. De ello se deduce que la bondad en la decorrelación de los usuarios no depende ya únicamente de las propiedades de correlación cruzada de los vectores de steering, sino también de las características espaciales del ruido. Es decir, de cuan ortogonales sean entre sí los vectores de steering en la norma del producto escalar respecto de  $\mathbf{R}_n^{-1}$ . Cabe resaltar que si el ruido sigue siendo temporalmente incorrelado, la estructura básica del decorrelador descrito en las secciones anteriores no varía.

La ecuación del decorrelador asintótico vendrá dada por simple sustitución,

$$(2.6.7.6) \quad \Delta(z^{-1}) = \left[ \sum_{n=-\infty}^{n=+\infty} \bar{\mathbf{R}}(n) z^{-n} \right]^{-1}, \quad \bar{\mathbf{R}}(n) = \mathbf{R}(n) \bullet (\mathbf{S}^H \mathbf{R}_n^{-1} \mathbf{S})$$

### 2.6.8 Apéndice C. Ruido temporal y espacialmente coloreado.

En los decorreladores que hemos estudiado hasta ahora, la estructura básica no ha variado debido a la premisa de incorrelación temporal. En este apéndice deduciremos cómo se modifica esa estructura cuando pasamos a considerar correlación temporal no nula en los procesos de ruido. Tal hecho tiene una cierta relevancia ya que modela un entorno de

interferencias de banda estrecha superpuestas en la banda de las señales CDMA. El desarrollo de las ecuaciones del decorrelador se llevará a cabo mediante aplicación del criterio de mínimos cuadrados en el dominio  $z$  ya que conlleva una mayor simplicidad que su equivalente temporal cuando se introduce la hipótesis de correlación temporal del ruido. Primeramente definamos la expresión de la función de correlación espacial-temporal del ruido en los sensores. Sea  $\mathbf{n}_k$  la secuencia vectorial del ruido de sensor (supuesto estacionario), entonces su función de correlación se define como la siguiente secuencia de matrices,

$$\mathbf{R}_{nn}(k) = E\mathbf{n}_m\mathbf{n}_{m+k}^H \tag{2.6.8.1}$$

con transformada  $z$  definida como,

$$\mathbf{R}_{nn}(z^{-1}) = \sum_{k=-\infty}^{k=+\infty} \mathbf{R}_{nn}(k)z^{-k} \tag{2.6.8.2}$$

El modelo de señal es el mismo que ya se expuso en la Sección 2.6.2. Su equivalente en el dominio transformado viene dado por,

$$\mathbf{r}_a(z^{-1}) = \mathbf{s}(a(z^{-N_s}) \bullet \mathbf{p}(z^{-1})) \tag{2.6.8.3}$$

definiendo el error en el dominio  $z$  como la transformada de la secuencia temporal del error entre el modelo y la señal recibida en los sensores,

$$\mathbf{e}(z^{-1}) = \mathbf{r}(z^{-1}) - \mathbf{r}_a(z^{-1}) \tag{2.6.8.4}$$

Definimos la secuencia de las estimaciones de los símbolos de cada usuario como aquella que minimiza la siguiente expresión del error en el dominio transformado,

$$\begin{aligned} \phi^2 &= \frac{1}{2\pi j} \oint_{C_1} \mathbf{e}(z^{-1})^H \mathbf{R}_{nn}^{-1}(z^{-1}) \mathbf{e}(z^{-1}) z^{-1} dz \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} \mathbf{e}(e^{-j\omega})^H \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{e}(e^{-j\omega}) d\omega \end{aligned} \tag{2.6.8.5}$$

La minimización se realiza mediante igualación a cero del gradiente de la función de coste con respecto a los términos hermíticos de la secuencia  $\mathbf{a}$ ,

$$\tag{2.6.8.6}$$

$$\nabla_{\mathbf{a}^H(m)} \phi^2 = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \left( \nabla_{\mathbf{a}^H(m)} \left( \mathbf{e}(e^{-j\omega})^H \right) \right) \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{e}(e^{-j\omega}) d\omega = 0$$

Denominando al integrando  $i(\omega)$ , su gradiente de expresa como,

$$\begin{aligned} \nabla_{\mathbf{a}^H(m)} i(\omega) &= -\nabla_{\mathbf{a}^H(m)} \left( \mathbf{a}(e^{-j\omega N_s})^H \bullet \mathbf{p}(e^{-j\omega})^H \right) \mathbf{S}^H \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{e}(e^{-j\omega}) \\ &= -e^{j\omega m N_s} \mathbf{p}(e^{-j\omega})^* \bullet \left( \mathbf{S}^H \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{e}(e^{-j\omega}) \right) \end{aligned} \quad (2.6.8.7)$$

Sustituyendo en la integral y expandiendo el término del error, llegamos a la siguiente igualdad,

$$\begin{aligned} \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{j\omega m N_s} \mathbf{p}(e^{-j\omega})^* \bullet \left( \mathbf{S}^H \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{r}(e^{-j\omega}) \right) d\omega &= \mathbf{b}(m') = \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{j\omega m N_s} \mathbf{p}(e^{-j\omega})^* \bullet \left( \mathbf{S}^H \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{S} \left( \mathbf{a}(e^{-j\omega N_s}) \bullet \mathbf{p}(e^{-j\omega}) \right) \right) d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{j\omega m N_s} \text{diag} \left( \mathbf{p}(e^{-j\omega})^* \right) \mathbf{S}^H \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{S} \text{diag} \left( \mathbf{p}(e^{-j\omega}) \right) \mathbf{a}(e^{-j\omega N_s}) d\omega \\ &= \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{j\omega m N_s} \left[ \left( \mathbf{p}(e^{-j\omega}) \mathbf{p}(e^{-j\omega})^H \right) \bullet \left( \mathbf{S}^H \mathbf{R}_{nn}^{-1}(e^{-j\omega}) \mathbf{S} \right) \right] \mathbf{a}(e^{-j\omega N_s}) d\omega \end{aligned} \quad (2.6.8.8)$$

donde el término de la izquierda representa la secuencia de estadísticos suficientes  $\mathbf{b}(m')$  y el término de la derecha establece su relación con las estimaciones de los símbolos. Pasando al dominio  $z$ , obtenemos las siguientes ecuaciones para la estimación de los símbolos,

$$\mathbf{b}(z^{-1}) = \left[ \mathbf{p}(z^{-1})^* \bullet \left( \mathbf{S}^H \mathbf{R}_{nn}^{-1}(z^{-1}) \mathbf{r}(z^{-1}) \right) \right]_{\downarrow N_s} \quad (2.6.8.9)$$

donde los estadísticos suficientes se obtienen por decimación de la secuencia filtrada. El proceso de decorrelación de usuarios viene dado por,

$$\Delta(z^{-1}) \stackrel{def}{=} \left[ \left[ \left( \mathbf{p}(z^{-1}) \mathbf{p}(z^{-1})^H \right) \bullet \left( \mathbf{S}^H \mathbf{R}_{nn}^{-1}(z^{-1}) \right) \right]^{-1} \right]_{\downarrow N_s} \quad (2.6.8.10)$$

con,

$$\mathbf{a}(z^{-1}) = \Delta(z^{-1}) \mathbf{b}(z^{-1}) \quad (2.6.8.11)$$

Podemos desglosar la ecuación 2.6.8.9 en una etapa previa de conformación de haz más una etapa de filtrado adaptado. Contrariamente a desarrollos del decorrelador en otros entornos, observamos en este caso que la conformación de haz incluye un filtrado espacio-temporal conjunto debido al acoplo de la correlación de la interferencia entre sensores distintos y instantes de tiempo diferentes. La inversa de la matriz de correlación elimina las interferencias en base a su correlación como etapa previa al filtrado adaptado (véase el desarrollo matemático de la sección 2.9: Modelado y Eliminación de Interferencias)

A fin de calcular la probabilidad de error de este correlador, necesitamos evaluar la función de correlación a la salida del banco de filtros adaptados. Debemos también modelar la función de correlación de los procesos de ruido en los sensores para el caso de interferencias de banda estrecha (sección 2.9) y ver cómo influye en la estructura del conformador de haz.

Debemos también evaluar el algoritmo óptimo de detección bajo el conocimiento de la estadística del ruido a la salida del decorrelador. Denominemos  $\mathbf{R}_{aa}(z^{-1})$  a la función de correlación del ruido pre-detección, el algoritmo asintótico óptimo de detección es entonces el que minimiza la siguiente norma del vector de error,

$$\hat{\mathbf{a}}(z^{-1}) = \arg \min_{\mathbf{a} \in \mathbb{C}} \frac{1}{2\pi} \int_{-\pi}^{+\pi} \boldsymbol{\varepsilon}(e^{-j\omega})^H \mathbf{R}_{vv}^{-1}(e^{-j\omega}) \boldsymbol{\varepsilon}(e^{-j\omega}) d\omega \quad , \quad \boldsymbol{\varepsilon} = \mathbf{a} - \hat{\mathbf{a}} \quad (2.6.8.12)$$

sobre el conjunto de todas las secuencias de símbolos posibles. Al igual que se había mencionado anteriormente, la determinación del signo del símbolo a la salida del decorrelador constituye un algoritmo subóptimo aunque sencillo puesto que no utilizamos toda la información a nuestra disposición.

En el ámbito de esta sección hemos estudiado la aplicación del criterio de referencia espacial a la decorrelación en array de señales CDMA bajo varios aspectos: ruido aditivo en los sensores sin y con correlación espacial y temporal. En la siguiente sección aplicaremos el criterio de referencia temporal al caso síncrono y destacaremos qué efectos se producen en el proceso de decorrelación de los usuarios, en contraste al algoritmo expuesto en esta sección: el ADD.

## 2.7 Decorrelación Síncrona.

En este apartado estudiaremos el algoritmo de decorrelación aplicado a señales CDMA síncronas bajo el criterio de referencia temporal. Veremos como bajo este criterio no se consigue una decorrelación total de los usuarios debido a razones análogas a las expuestas en la sección 2.4. De hecho, el análisis realizado en dicha sección puede considerarse como un caso particular del que aquí llevaremos a término. Observaremos como aparece de nuevo el concepto de inversión de potencia. El estudio del caso de usuarios CDMA síncronos, más sencillo que el asíncrono, proporciona conceptos intuitivos sobre el comportamiento de la interferencia de múltiple acceso y su eliminación. En secciones anteriores se han hallado los algoritmos lineales óptimos de decorrelación basados en conocimiento previo de los sincronismos de cada usuario. En el entorno CDMA síncrono supondremos que los usuarios sólo pueden transmitir según el sincronismo marcado por una referencia externa. Podemos asumir que aproximadamente, todos los usuarios se muestrean en el mismo instante de tiempo.

La conexión o desconexión de usuarios modifica el carácter de la interferencia de múltiple acceso que afecta a cada uno de los usuarios activos. Para proceder a su eliminación de forma óptima necesitamos conocer cuáles son los usuarios que están activos a fin de calcular el decorrelador correspondiente. El cálculo del decorrelador óptimo requiere la inversión de una matriz, con lo cual, si el número de usuarios es suficientemente elevado, puede resultar engorroso. Un método optimizado para la inversión de la matriz de correlación de usuarios (activos) lo constituye la descomposición de Schur de una matriz de Gram (tal es el carácter de la matriz de correlación de usuarios pues ésta es simplemente una matriz de producto escalar de las secuencias de spread de cada usuario). En otra línea, se puede solventar el problema de recalcular la matriz de decorrelación si se efectúa una decorrelación global de todos los usuarios, estén o no activos. Este procedimiento es sin embargo subóptimo al introducir un exceso de ruido en el algoritmo adaptativo proveniente de las señales a la salida del banco de filtros adaptados correspondientes a usuarios no activos. Aunque en esas salidas también observamos una contribución de señal del usuario deseado, tal información es redundante puesto que ya disponemos de un conjunto mínimo de estadísticos suficientes en las salidas del banco de filtros correspondiente a usuarios activos. A continuación justificaremos matemáticamente los conceptos presentados en esta pequeña introducción.

En este modelo supondremos que la señal recibida es un proceso escalar que contiene las señales transmitidas por cada usuario más un proceso de ruido Gaussiano aditivo blanco. Obtendremos un conjunto de estadísticos suficientes de los símbolos transmitidos por cada usuario a la salida de un banco de filtros adaptados a cada secuencia. Denotamos a esta salida muestreada en el sincronismo correcto con el vector  $\mathbf{y}_k$ , donde el subíndice denota el vector de símbolos  $k$ -ésimo. La matriz de correlación (o mezcla) de usuarios se denota con  $\Phi_k$ , y el ruido a la salida del banco de filtros con  $\mathbf{n}_k$ , cuya matriz de correlación es proporcional a la matriz de correlación de usuarios en base a la suposición de ruido blanco a la entrada del banco de filtros. Supondremos también amplitudes diferentes para cada uno de los usuarios. En estas condiciones,  $\mathbf{y}_k$  puede expresarse como sigue,

$$\mathbf{y}_k = \Phi_k \mathbf{P}_k \mathbf{a}_k + \mathbf{n}_k = \Phi_k (\mathbf{p}_k \bullet \mathbf{a}_k) + \mathbf{n}_k \quad , \quad E_{\mathbf{n}_k} \mathbf{n}_k \mathbf{n}_k^H = \sigma_n^2 \Phi_k \quad (2.7.1)$$

$$\mathbf{P}_k = \text{diag}(\mathbf{p}_k)$$

Hemos supuesto que  $\Phi_k$  constituye la matriz de correlación de *todos* los usuarios del sistema. La matriz diagonal  $\mathbf{P}_k$  contiene las amplitudes de los usuarios, al igual que el vector  $\mathbf{p}_k$ . La activación o desactivación de usuarios se simula poniendo a cero los componentes correspondientes del vector de amplitudes  $\mathbf{p}_k$  (y/o también poniendo a cero las componentes correspondientes de  $\mathbf{a}_k$ ). Denominaremos a este modelo de señal el modelo global puesto que agrupa a todos los usuarios posiblemente activos. Definimos paralelamente el modelo reducido como aquél que agrupa únicamente a los usuarios activos y utilizaremos el símbolo super-barra para identificarlos. Las ecuaciones correspondientes se obtienen super-barrando todos los términos involucrados en el modelo global, es decir,

$$\bar{\mathbf{y}}_k = \bar{\Phi}_k \bar{\mathbf{P}}_k \bar{\mathbf{a}}_k + \bar{\mathbf{n}}_k = \bar{\Phi}_k (\bar{\mathbf{p}}_k \bullet \bar{\mathbf{a}}_k) + \bar{\mathbf{n}}_k \quad , \quad E_{\bar{\mathbf{n}}_k} \bar{\mathbf{n}}_k \bar{\mathbf{n}}_k^H = \bar{\Phi}_k \quad (2.7.2)$$

$$\bar{\mathbf{P}}_k = \text{diag}(\bar{\mathbf{p}}_k)$$

Las relación entre los vectores de ambos modelos se logra en base a matrices de bloqueo (matrices con un número de filas iguales al orden del modelo y número de columnas igual al orden del modelo global que contienen un único uno en cada fila para seleccionar al usuario correspondiente, diferente para cada fila). En estas condiciones,

$$\bar{\mathbf{p}}_k = \Pi_k \mathbf{p}_k \quad , \quad \bar{\mathbf{a}}_k = \Pi_k \mathbf{a}_k \quad , \quad \bar{\mathbf{n}}_k = \Pi_k \mathbf{n}_k \quad (2.7.3)$$

La relación entre las matrices de correlación de usuarios se obtiene en base a la matriz de secuencias de spread,

$$(2.7.4)$$

$$\bar{\mathbf{S}}_k = \mathbf{S}_k \Pi_k^T \Rightarrow \bar{\Phi}_k = \bar{\mathbf{S}}_k^H \bar{\mathbf{S}}_k = \Pi_k \mathbf{S}_k^H \mathbf{S}_k \Pi_k^T = \Pi_k \Phi_k \Pi_k^T$$

Es decir, la matriz de correlación reducida, al igual que el resto de vectores reducido se obtiene por diezmado de las matrices y vectores globales. La decorrelación se realiza en ambos modelos con la inversa de la matriz de correlación. Denominamos  $\mathbf{u}_k$  al vector resultado,

$$\mathbf{u}_k = \Phi_k^{-1} \mathbf{y}_k, \quad \bar{\mathbf{u}}_k = \bar{\Phi}_k^{-1} \bar{\mathbf{y}}_k \quad (2.7.5)$$

que se considera sin normalizar por las amplitudes de cada usuario.

### 2.7.1 Decorrelación Asistida (I)

Estudiaremos a continuación el algoritmo óptimo en el sentido de mínimo error cuadrático que nos permita recuperar los usuarios activos a partir del modelo global. Es nuestro objetivo ver si coincide con la solución del modelo reducido según el criterio ML utilizando la información de qué usuarios se encuentran activos. La función de coste supuestas conocidas las secuencias de datos de los usuarios activos se plantea de la siguiente forma (en referencia a la Sección 2.2),

$$J = E_{\mathbf{a}_t, \mathbf{n}_t} (\bar{\mathbf{a}}_k - \mathbf{B}_k^H \mathbf{y}_k)^H \mathbf{C}_k^{-1} (\bar{\mathbf{a}}_k - \mathbf{B}_k^H \mathbf{y}_k) - \ln(\pi^{-N_s} \det \mathbf{C}_k^{-1}) \quad (2.7.1.1)$$

donde  $\mathbf{C}_k$  constituye la matriz de correlación del ruido a la salida del decorrelador. La solución se obtiene mediante igualación a cero del gradiente matricial de  $J$  con respecto a la matriz de decorrelación  $\mathbf{B}_k^H$  (Nótese que esta matriz no es únicamente de decorrelación sino que también incluye normalización de la potencia de todos los usuarios),

$$\begin{aligned} \nabla_{\mathbf{B}_k^H} J &= -\nabla_{\mathbf{B}_k^H} E_{\mathbf{a}_t, \mathbf{n}_t} (\bar{\mathbf{a}}_k - \mathbf{B}_k^H \mathbf{y}_k)^H \mathbf{C}_k^{-1} \mathbf{B}_k^H \mathbf{y}_k = -\nabla_{\mathbf{B}_k^H} E_{\mathbf{a}_t, \mathbf{n}_t} \text{tr} \left[ \mathbf{B}_k^H \mathbf{y}_k (\bar{\mathbf{a}}_k - \mathbf{B}_k^H \mathbf{y}_k)^H \mathbf{C}_k^{-1} \right] \\ &= -\nabla_{\mathbf{B}_k^H} \text{tr} \left[ \mathbf{B}_k^H E_{\mathbf{a}_t, \mathbf{n}_t} \mathbf{y}_k (\bar{\mathbf{a}}_k - \mathbf{B}_k^H \mathbf{y}_k)^H \mathbf{C}_k^{-1} \right] = -\nabla_{\mathbf{B}_k^H} \text{tr} \left[ \mathbf{B}_k^H \mathbf{C}_k^{-1} (\mathbf{R}_{y\bar{\mathbf{a}}} - \mathbf{R}_{yy} \mathbf{B}_k) \right] \end{aligned} \quad (2.7.1.2)$$

donde hemos definido las matrices de correlación cruzada de la salida del decorrelador con los datos del modelo global y la matriz de autocorrelación del modelo global. Efectuando la diferenciación de la traza, obtenemos la solución óptima,

$$(2.7.1.3)$$

$$\nabla_{\mathbf{B}_k^H} J = \left[ \mathbf{C}_k^{-1} (\mathbf{R}_{y\bar{a}} - \mathbf{R}_{yy} \mathbf{B}_k) \right]^H = \mathbf{0} \Rightarrow \mathbf{B}_k = \mathbf{R}_{yy}^{-1} \mathbf{R}_{y\bar{a}}$$

Observamos que la solución  $\mathbf{B}_k$  es independiente de la matriz de correlación  $\mathbf{C}_k$ . Debemos comprobar a continuación si dicha matriz cumple la siguiente igualdad,

$$\mathbf{B}_k \stackrel{?}{=} \bar{\Phi}_k^{-1} \bar{\mathbf{P}}_k^{-1} \tag{2.7.1.4}$$

Con sideremos en primer lugar la matriz de correlación cruzada  $\mathbf{R}_{y\bar{a}}$ ,

$$\mathbf{R}_{y\bar{a}} = E_{\mathbf{a}_t, \mathbf{n}_t} \mathbf{y}_k \bar{\mathbf{a}}_k^H = E_{\mathbf{a}_t, \mathbf{n}_t} (\Phi_k \mathbf{P}_k \mathbf{a}_k + \mathbf{n}_k) \bar{\mathbf{a}}_k^H = \Phi_k \mathbf{P}_k E_{\mathbf{a}_t} \mathbf{a}_k \bar{\mathbf{a}}_k^H = \Phi_k \mathbf{P}_{a\bar{a}} \quad , \quad \mathbf{P}_{a\bar{a}} = \begin{bmatrix} \bar{\mathbf{P}}_k \\ \mathbf{0} \end{bmatrix} \tag{2.7.1.5}$$

Donde hemos supuesto incorrelación entre los datos conocidos y el ruido  $\mathbf{n}_k$ . Sin pérdida de generalidad suponemos que los usuarios activos se encuentran agrupados en la parte superior del banco de filtros adaptados y que los usuarios no activos se suponen con amplitud cero. Observemos que la matriz  $\mathbf{P}_{aa}$  presenta una estructura de bloqueo de componentes. Consideremos en segundo lugar la matriz de autocorrelación del modelo global,

$$\mathbf{R}_{yy} = E_{\mathbf{a}_t, \mathbf{n}_t} \mathbf{y}_k \mathbf{y}_k^H = E_{\mathbf{a}_t, \mathbf{n}_t} (\Phi_k \mathbf{P}_k \mathbf{a}_k + \mathbf{n}_k) (\Phi_k \mathbf{P}_k \mathbf{a}_k + \mathbf{n}_k)^H = \Phi_k \mathbf{P}_k^2 \Phi_k + \sigma_n^2 \Phi_k \tag{2.7.1.6}$$

Sustituyendo las expresiones de ambas matrices en la ecuación 2.7.1.3, llegamos a la siguiente expresión para el decorrelador,

$$\mathbf{B}_k = (\Phi_k \mathbf{P}_k^2 \Phi_k + \sigma_n^2 \Phi_k)^{-1} \Phi_k \mathbf{P}_{a\bar{a}} \tag{2.7.1.7}$$

Consideremos ambos términos de  $\mathbf{R}_{yy}$ . La primera matriz corresponde a la correlación entre usuarios y su rango es igual al número de usuarios activos mientras que el segundo correspondiente a la correlación del ruido es de rango completo e igual al número global de usuarios,

$$\text{rank}(\Phi_k \mathbf{P}_k^2 \Phi_k) = \bar{N}_u \quad , \quad \text{rank}(\Phi_k) = N_u \tag{2.7.1.8}$$

Vemos pues que la solución para el decorrelador según este criterio depende explícitamente de la varianza del ruido, hecho que no sucede con la expresión del decorrelador óptimo según se definió en la sección 2.6, dado que su derivación se basaba en un criterio de referencia espacial. Obsérvese que dicha solución es formalmente equivalente al filtrado de

Wiener. Resulta de interés observar si esta matriz efectúa un bloqueo de los usuarios no activos y deducir su comportamiento asintótico cuando la varianza del ruido tiende a cero. Apliquemos el Lema de Inversión de Matrices a la inversa de la autocorrelación  $\mathbf{R}_{yy}$ ,

$$\begin{aligned} (\sigma_n^2 \Phi_k + \Phi_k \mathbf{P}_k^2 \Phi_k)^{-1} &= \frac{1}{\sigma_n^2} \Phi_k^{-1} - \frac{1}{\sigma_n^2} \Phi_k^{-1} \Phi_k \mathbf{P}_k \left( \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{P}_k \Phi_k \Phi_k^{-1} \Phi_k \mathbf{P}_k \right)^{-1} \mathbf{P}_k \Phi_k \Phi_k^{-1} \frac{1}{\sigma_n^2} \\ &= \frac{1}{\sigma_n^2} \Phi_k^{-1} - \frac{1}{\sigma_n^2} \mathbf{P}_k \left( \mathbf{I} + \frac{1}{\sigma_n^2} \mathbf{P}_k \Phi_k \mathbf{P}_k \right)^{-1} \mathbf{P}_k \frac{1}{\sigma_n^2} \end{aligned} \quad (2.7.1.9)$$

Ahora bien, podemos descomponer  $\mathbf{P}_k$  y  $\Phi_k$  en base a una partición en usuarios activos y usuarios no-activos,

$$\mathbf{P}_k = \begin{bmatrix} \bar{\mathbf{P}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \Phi_k = \begin{bmatrix} \bar{\Phi}_k & \times \\ \times & \times \end{bmatrix} \quad (2.7.1.10)$$

con lo cual,

$$\begin{aligned} \mathbf{R}_{yy}^{-1} &= \frac{1}{\sigma_n^2} \Phi_k^{-1} - \frac{1}{\sigma_n^2} \begin{bmatrix} \bar{\mathbf{P}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{I} + \frac{1}{\sigma_n^2} \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{P}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \frac{1}{\sigma_n^2} \\ &= \frac{1}{\sigma_n^2} \Phi_k^{-1} - \frac{1}{\sigma_n^2} \begin{bmatrix} \frac{1}{\sigma_n^2} \bar{\mathbf{P}}_k \left( \mathbf{I} + \frac{1}{\sigma_n^2} \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k \right)^{-1} \bar{\mathbf{P}}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \end{aligned} \quad (2.7.1.11)$$

Si incluimos ahora la postmultiplicación por la matriz de correlación cruzada  $\mathbf{R}_{y\bar{a}}$ , obtenemos,

$$\mathbf{R}_{yy}^{-1} \mathbf{R}_{y\bar{a}} = \frac{1}{\sigma_n^2} \begin{bmatrix} \bar{\mathbf{P}}_k \\ \mathbf{0} \end{bmatrix} - \frac{1}{\sigma_n^2} \begin{bmatrix} \frac{1}{\sigma_n^2} \bar{\mathbf{P}}_k \left( \mathbf{I} + \frac{1}{\sigma_n^2} \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k \right)^{-1} \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{B}}_k \\ \mathbf{0} \end{bmatrix} \quad (2.7.1.12)$$

donde podemos confirmar que el decorrelador diseñado según el criterio de la ecuación 2.7.1.1 efectivamente bloquea a los usuarios no activos como se deduce de la presencia de la submatriz  $\mathbf{0}$  en la matriz de decorrelación. El hecho de que se bloqueen los usuarios no activos depende de la relación existente entre la matriz de correlación de los usuarios y la matriz de correlación del ruido. Una eliminación de los usuarios conlleva una minimización de la potencia de ruido debido a la dependencia entre la matriz de correlación del ruido y la matriz de correlación de los usuarios.

Si intentamos calcular ahora el límite cuando  $\sigma_n \rightarrow 0$  de la expresión anterior llegamos a una indeterminación del tipo 0/0 por lo que resulta aconsejable hallar una expresión más

compacta para el subdecorrelador aplicando el Lema de Inversión en sentido contrario, de donde,

$$\bar{\mathbf{B}}_k = \bar{\mathbf{P}}_k \left( \frac{1}{\sigma_n^2} \mathbf{I} - \frac{1}{\sigma_n^2} \left( \mathbf{I} + \frac{1}{\sigma_n^2} \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k \right)^{-1} \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k \frac{1}{\sigma_n^2} \right) = \bar{\mathbf{P}}_k \left( \sigma_n^2 \mathbf{I} + \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k \right)^{-1} \quad (2.7.1.13)$$

Nótese que la expresión final del subdecorrelador no coincide con la solución de máxima verosimilitud obtenida en la sección 2.6. La solución óptima constituye una mezcla de la solución obtenida según el criterio ML, excepto en el límite  $\sigma_n \rightarrow 0$  (lo cual resulta una aproximación para entornos donde el término dominante es la interferencia de múltiple acceso y no la potencia de ruido Gaussiano aditivo) donde coinciden,

$$\lim_{\sigma_n \rightarrow 0} \bar{\mathbf{B}}_k = \bar{\mathbf{P}}_k \left( \bar{\mathbf{P}}_k \bar{\Phi}_k \bar{\mathbf{P}}_k \right)^{-1} = \bar{\Phi}_k^{-1} \bar{\mathbf{P}}_k^{-1} \Rightarrow \lim_{\sigma_n \rightarrow 0} \bar{\mathbf{z}}_k = \bar{\mathbf{P}}_k^{-1} \bar{\mathbf{u}}_k \quad (2.7.1.14)$$

Vemos pues que asintóticamente, el decorrelador consigue eliminar completamente la interferencia de múltiple acceso según el criterio 2.7.1.1. De hecho, en el caso límite de potencia de ruido nula, existe una infinidad de decorreladores que eliminan la interferencia de otros usuarios. Podemos ver que el conjunto de todos los decorreladores posibles puede generarse a partir de un subespacio de rango igual al número de usuarios no activos. Como no existe la restricción de minimizar la potencia de ruido, cualquier proyección de la que se obtengan los símbolos de cada usuario libres de interferencia pertenecen al conjunto de la solución. En estas condiciones podemos suponer que el conjunto de usuarios activos barren un subespacio del espacio global  $\mathbf{y}_k$ . A partir de la ecuación 2.7.1.1 expresamos el modelo sin ruido,

$$\mathbf{y}_k = \Phi_k \mathbf{P}_k \mathbf{a}_k = \Phi_k \begin{bmatrix} \bar{\mathbf{P}}_k \bar{\mathbf{a}}_k \\ \mathbf{0} \end{bmatrix} \quad (2.7.1.15)$$

Introduciendo la siguiente partición de la matriz  $\Phi_k$ , llegamos al conjunto de vectores generadores del subespacio  $\mathbf{y}_k$ ,

$$\Phi_k = \begin{bmatrix} \bar{\Phi}_k & \Phi_{12}^H \\ \Phi_{12} & \Phi_{22} \end{bmatrix} \Rightarrow \mathbf{y}_k = \begin{bmatrix} \bar{\Phi}_k \\ \Phi_{12} \end{bmatrix} \bar{\mathbf{P}}_k \bar{\mathbf{a}}_k = \mathbf{F}_k \bar{\mathbf{P}}_k \bar{\mathbf{a}}_k \quad (2.7.1.16)$$

Llegando a una sistema sobredeterminado de ecuaciones para obtener el vector de símbolos de los usuarios activos. Los vectores generadores son precisamente las columnas correspondiente de la matriz  $\Phi_k$ ,  $\mathbf{F}_k$ . El conjunto de todos los decorreladores vendrá

parametrizado por la matriz rectangular  $\mathbf{M}_k$ , que combina todos los vectores del complemento ortogonal a  $\mathbf{F}_k$  para cada usuario determinado.

$$\bar{\mathbf{a}}_k = \bar{\mathbf{P}}_k^{-1}(\mathbf{F}_k^\# + \mathbf{M}_k \mathbf{F}_k^\perp) \mathbf{y}_k \quad (2.7.1.17)$$

donde el superíndice # denota pseudo-inversa.

### 2.7.2 Decorrelación Asistida (II)

En este apartado consideraremos un entorno ligeramente diferente al del apartado anterior. Supondremos desconocido el número de usuarios activos pero consideraremos conocidas las secuencias de símbolos transmitidas por cada usuario en el caso de que el usuario estuviese activo. Llamaremos a la secuencia vectorial de referencia  $\mathbf{d}_k$  para distinguirla del vector de símbolos de los usuarios activos (ambos coincidían en el apartado anterior). Supondremos que la secuencia  $\mathbf{d}_k$  puede agrupar cualquier conjunto posible de usuarios. La función de coste a minimizar sigue siendo el error respecto a la referencia,

$$J = E_{\mathbf{a}_k, \mathbf{n}_k} (\mathbf{d}_k - \mathbf{B}_k^H \mathbf{y}_k)^H \mathbf{C}_k^{-1} (\mathbf{d}_k - \mathbf{B}_k^H \mathbf{y}_k) - \ln(\pi^{-N_k} \det \mathbf{C}_k^{-1}) \quad (2.7.2.1)$$

Efectuando la minimización de forma análoga llegamos a la expresión formalmente equivalente a la deducida en el apartado anterior,

$$\mathbf{B}_k = \mathbf{R}_{yy}^{-1} \mathbf{R}_{yd} \quad , \quad \mathbf{R}_{yd} = \Phi_k \mathbf{P}_k \mathbf{P}_{ad} = \Phi_k \mathbf{P}_k E_{\mathbf{a}_k} \mathbf{a}_k \mathbf{d}_k^H \quad (2.7.2.2)$$

Suponiendo también incorrelación estadística entre secuencias, llegamos a la misma identidad que en el apartado anterior donde agrupamos en la parte inferior de la matriz a aquellos usuarios no activos,

$$\mathbf{P}_k E_{\mathbf{a}_k} \mathbf{a}_k \mathbf{d}_k^H = \begin{bmatrix} \bar{\mathbf{P}}_k^d \\ \mathbf{0} \end{bmatrix} \quad , \quad \bar{\mathbf{P}}_k = \begin{bmatrix} \bar{\mathbf{P}}_k^d & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{P}}_k^+ \end{bmatrix} \quad (2.7.2.3)$$

donde  $\mathbf{P}_k^d$  es una matriz diagonal no nula que contiene las amplitudes de los usuarios activos dentro del subconjunto de usuarios de la secuencia  $\mathbf{d}_k$  y la matriz  $\mathbf{P}_k^+$  contiene al resto de usuarios activos. Según lo expuesto en el apartado anterior, vemos que la función de coste bloquea automáticamente aquellos usuarios no activos en la salida correspondiente

del filtro adaptado sin conocimiento de qué usuarios se encuentran activos. En cualquier caso, la matriz  $\mathbf{B}_k$  solución del criterio sigue difiriendo de la matriz óptima de decorrelación en presencia de ruido no nulo. El único caso en que no ocurre este efecto es cuando la secuencia de referencia  $\mathbf{d}_k$  contiene a un único usuario. La diferencia respecto al decorrelador óptimo radica entonces en una simple constante de proporcionalidad.

En la siguiente sección consideraremos un nuevo problema al que también resulta aplicable el criterio de referencia espacial: la ecualización de canal en comunicaciones digitales. Utilizaremos básicamente resultados de la sección 2.6, estableciendo un paralelismo entre las ecuaciones que rigen la ecualización y la decorrelación de usuarios CDMA.

## 2.8 Ecuación

En esta sección se trata de las propiedades del ecualizador óptimo bajo el criterio de referencia espacial, análogamente a como se dedujo la expresión del decorrelador óptimo en la sección 2.6. Veremos que tal ecualizador no es sino un caso particular del decorrelador de símbolos CDMA. Supondremos que disponemos de la información lateral necesaria, es decir, el instante óptimo de sincronismo y la respuesta impulsional del canal. A semejanza de la sección 2.6, supondremos también condiciones estacionarias en la propagación.

Nos centraremos exclusivamente en la ecualización de señales moduladas linealmente para entornos de sensor único y de arrays. Una característica general de los ecualizadores lineales es el aumento de la varianza del ruido a la entrada del detector con respecto al caso de interferencia intersimbólica nula (estableciendo una analogía con respecto a la decorrelación de usuarios en un entorno CDMA, donde los términos *interferencia de múltiple acceso* y *interferencia intersimbólica* son intercambiables). Aunque existe una infinitud de ecualizadores que eliminan la interferencia intersimbólica, nos interesaremos sólo por la estructura del ecualizador con comportamiento óptimo frente al ruido. Una derivación alternativa basada en la expansión de Karthunen-Loève de la secuencia de datos puede hallarse en [Proakis], apartado 6.3. La aquí obtenida está basada en una particularización del decorrelador CDMA, es decir, utilizamos la filosofía de referencia espacial. Aunque existen algoritmos que realizan las funciones de ecualización y detección de forma conjunta por minimización de métricas (algoritmos de tipo Viterbi), en este capítulo consideramos que se realizan separadamente en una etapa previa de ecualización seguida de una etapa de detección.

### 2.8.1 Modelo de la señal

Sea una secuencia de símbolos  $a$  modular  $a$ , pertenecientes a una constelación  $\mathcal{C}$ , la expresión matemática de la señal equivalente en banda base a la salida del modulador viene dada por,

$$s(t) = \sum_{m=-\infty}^{m=+\infty} a_m p_o(t - mT) \quad (2.8.1.1)$$

donde T representa el tiempo de símbolo. Dicha señal es filtrada por un canal lineal y estacionario, de respuesta al impulso  $h_c(t)$ . La señal filtrada vendrá expresada por,

$$r_c(t) = \sum_{m=-\infty}^{m=+\infty} a_m p(t - mT) \quad , \quad p(t) = h_c(t) * p_o(t) \tag{2.8.1.2}$$

donde  $p(t)$  representa el pulso equivalente resultado de la propagación del pulso del modulador por el canal. El propósito del ecualizador es entonces recuperar los símbolos modulados a partir del conocimiento de  $p(t)$ . Para una implementación digital resulta más adecuado trabajar con un modelo discreto de la señal y de la propagación donde,

$$s(n) = \sum_{m=-\infty}^{m=+\infty} a_m p_o(n - mN_{ss}) \tag{2.8.1.3}$$

y,

$$p(n) = \sum_{k=-\infty}^{k=+\infty} h_c(k) p_o(n - k) \tag{2.8.1.4}$$

donde ahora  $N_{ss}$  representa el número de muestras por símbolo. En base a este modelo lineal, el efecto de rotación de fase de la portadora y del sincronismo también puede asociarse a la propagación  $h_c(n)$ . Aunque físicamente esto no ocurra así.

### 2.8.2 Ecualizador óptimo

La expresión del ecualizador óptimo se deduce por simple sustitución de las ecuaciones del decorrelador obtenidas en la sección 2.6. Por comodidad reproducimos aquí la ecuación del cálculo de los estadísticos suficientes,

$$b(m') = \sum_{n=-\infty}^{n=+\infty} (S^H r(n)) \cdot p(n - m'N_{ss}) \tag{2.8.2.1}$$

Puesto que estamos considerando un único usuario y un único sensor, la ecuación previa puede reducirse a una ecuación escalar mediante  $S=1$ ,  $r=r$  y  $p=p$ ,

$$b(m') = \sum_{n=-\infty}^{n=+\infty} r(n) p(n - m'N_{ss}) \tag{2.8.2.2}$$

donde podemos observar que el conjunto de estadísticos suficientes para la detección de los símbolos se obtiene después de filtrar los datos recibidos  $r(n)$  por un filtro igual a la convolución del pulso emitido por el modulador con la respuesta impulsional del canal. No obstante, esto no eliminará la interferencia intersimbólica ya que la respuesta impulsional global hasta la obtención de la secuencia  $b(m')$  no presentará en general nullos en múltiplos del tiempo de símbolo. Sin embargo, se permite diezmar la secuencia a la salida del filtro  $p(n)$  en los instantes de sincronismo sin ninguna pérdida de información por el hecho de tratarse de estadísticos suficientes. Una etapa de equalización posterior a la velocidad de símbolo se encargará de eliminar la interferencia intersimbólica residual. Nótese que esta etapa posterior es equivalente a la decorrelación de usuarios CDMA, aunque en este caso se trata de una decorrelación de símbolos. A semejanza de la sección 2.6, el equalizador óptimo de símbolos es también no-causal.

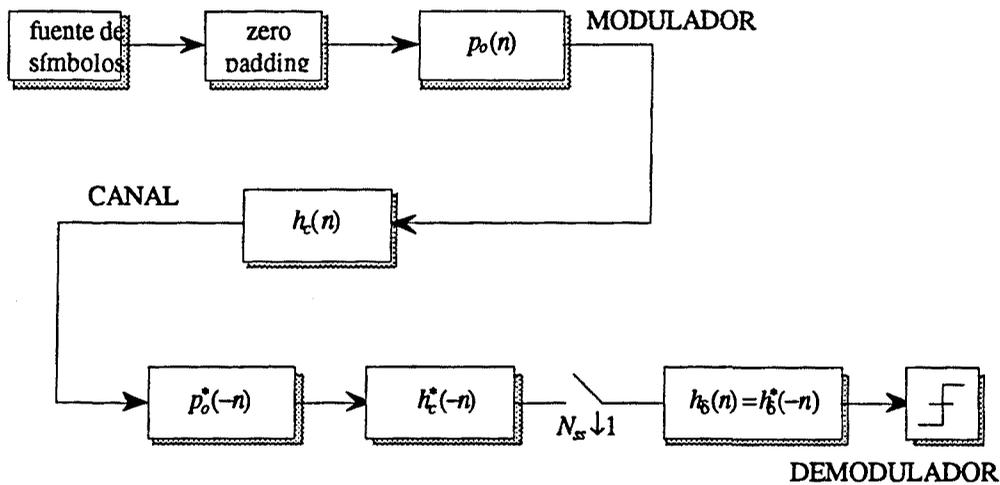


Figura 2.8.2.1: modelo equivalente banda base del equalizador óptimo. El modulador está compuesto por una fuente de símbolos, los cuales se interpolan mediante el pulso conformador  $p_o(n)$ . El canal está modelado por un filtro  $h_c(n)$  invariante. El demodulador está constituido por un filtro adaptado al pulso conformador en cascada con un filtro adaptado al canal. La salida de la concatenación de filtros adaptados se diezma en el instante óptimo de sincronismo y los símbolos se decorrelan con el inverso de la respuesta impulsional del canal equivalente visto desde la etapa de diezmo,  $h_g(n)$ .

La etapa de diezmo introduce un aliasing en el espectro que no conlleva ninguna pérdida de información excepto en el caso en que los filtros no estén exactamente adaptados. La

pérdida de información que introduce entonces el diezmado se puede interpretar paralelamente como un aliasing no controlado (subóptimo).

La expresión del decorrelador de símbolos viene dada por,

$$\delta(z^{-1}) = \left[ \sum_{k=-\infty}^{k=+\infty} \rho(k) z^{-k} \right]^{-1}, \quad \rho(k) = \sum_{n=-\infty}^{n=+\infty} p(n) p(n - kN_{ss}) \quad (2.8.2.3)$$

donde la secuencia  $r$  es precisamente la correlación cruzada entre réplicas desplazadas de la respuesta impulsional a la salida del filtro adaptado. Es importante destacar que el decorrelador de símbolos (al igual que el decorrelador de usuarios) es básicamente un inversor del canal equivalente observado a la salida del filtro adaptado (al pulso del modulador más el canal) después de diezmar por el tiempo de símbolo. En general, una inversión de canal no se corresponde con el ecualizador óptimo, excepto en el caso de que deba invertir un conjunto mínimo de estadísticos suficientes. En este sentido, puede afirmarse que el ecualizador óptimo pseudo-invierte el canal para producir varianza mínima de ruido a la salida. A lo largo de los años se han desarrollado multitud de algoritmos para la identificación de canal. Sin embargo, una vez identificado el modelo de canal y sus coeficientes, debe utilizar la información de forma eficiente, no invirtiendo el canal, sino pseudo-invirtiéndolo.

La transformada  $z$  del ecualizador óptimo puede factorizarse de la siguiente forma,

$$h_{eq}(z^{-1}) = h_c(z^{-1}) * h_g(z^{-N_{ss}}) \quad (2.8.2.4)$$

entendiendo como ecualizador el filtro a la salida del filtro adaptado al pulso modulador. Esta factorización equivale a un ecualizador de muestras o fraccionado (fractionally-spaced equalizer) en cascada con un ecualizador de símbolos (symbol-spaced equalizer). El ecualizador de muestras resulta ser la respuesta impulsional del canal.

Podemos demostrar por construcción que  $h_g(z^{-1})$  es un filtro estable (remitámonos simplemente a las demostraciones de estabilidad de la sección 2.6). La estabilidad del decorrelador CDMA venía dada en función de las raíces del determinante de la matriz de correlación de símbolos. Por sustitución, dicha matriz es en este entorno una secuencia escalar y las raíces del determinante son simplemente las soluciones de  $\rho(z^{-1}) = 0$ . La única condición para la estabilidad del ecualizador consiste en que ninguna de las raíces (o polos del decorrelador de símbolos) se encuentren en el círculo unidad. De hecho existe un número igual de raíces dentro y fuera del círculo unidad debido al carácter hermfítico del

ecualizador. No existe ninguna raíz en el círculo unidad por construcción, ya que la secuencia  $r$  se ha obtenido correlando réplicas desplazadas de la respuesta impulsional.

### 2.8.3 Detección óptima

Una vez que se ha eliminado la interferencia intersimbólica introducida por el canal, debe determinarse cuál es el algoritmo de detección a utilizar. Para ello centrémonos en la función de densidad del ruido a la salida del ecualizador. Sea  $v_k$  la secuencia del ruido de detección. Su función de densidad conjunta vendrá unívocamente determinada por sus estadísticos de segundo orden. Es suficiente por tanto determinar la función de correlación,

$$\rho_{vv}(n) = E v(k) v^*(k+n) \tag{2.8.3.1}$$

donde el ruido de detección se obtiene mediante dos pasos: la convolución del ruido blanco de entrada con la respuesta impulsional del filtro adaptado a  $p(n)$  (pulso conformador más canal), el cual denominamos  $\eta_k$ , y la convolución del ruido diezmado a la salida del filtro adaptado con la respuesta impulsional del decorrelador de símbolos,

$$\begin{aligned} \eta(k) &= n(k) * p_o^*(-k) * h_c^*(-k) = n(k) * p^*(-k) \\ v(k) &= \eta(kN_{ss}) * h_d(k) \end{aligned} \tag{2.8.3.2}$$

La función de densidad conjunta del ruido de detección para  $N$  muestras vendrá dada por la función de densidad Gaussiana multivariada con matriz de correlación  $\mathbf{R}_{vv}$ ,

$$[\mathbf{R}_{vv}]_{ij} = \rho_{vv}(i-j) \tag{2.8.3.3}$$

es decir,

$$p_v(\mathbf{v}) = \frac{1}{\pi^N \det \mathbf{R}_{vv}} \exp(-\mathbf{v}^H \mathbf{R}_{vv}^{-1} \mathbf{v}) \tag{2.8.3.4}$$

Si denominamos  $\mathbf{a}$  a la salida del ecualizador y  $\hat{\mathbf{a}}$  a los símbolos decididos, el criterio óptimo de decisión será aquél que maximice la función de distribución del ruido evaluada en  $\mathbf{a}-\hat{\mathbf{a}}$  sobre todos los símbolos  $\hat{\mathbf{a}}$  pertenecientes a la constelación  $\mathcal{C}$ . Por lo tanto, el criterio de proximidad vendrá dado según la distancia de Mahalanobis,

$$\tag{2.8.3.5}$$

$$\hat{\mathbf{a}} = \arg \min_{\mathbf{a} \in \mathcal{C}} \{(\mathbf{a} - \hat{\mathbf{a}})^H \mathbf{R}_{\mathbf{v}\mathbf{v}}^{-1} (\mathbf{a} - \hat{\mathbf{a}})\}$$

La detección símbolo a símbolo constituye por tanto una estrategia subóptima. Debido a la correlación cruzada debe realizarse una detección sobre conjuntos de símbolos. La bondad de la detección aumentará en función de la longitud del intervalo considerado, obteniendo un resultado óptimo en el caso asintótico cuando  $N \rightarrow \infty$ . El criterio se modifica en el caso límite ya que debe interpretarse como la minimización de la potencia de la secuencia de error a la salida de un filtro. El término de ruido a la salida de este filtro será ruido blanco puesto que la inversa de la matriz de correlación del ruido actúa como un filtro blanqueador.

#### 2.8.4 Ecuador óptimo en array

En este apartado desarrollaremos las ecuaciones del ecualizador óptimo en array. Consideramos de momento que no existen interferencias y que la función del array es simplemente una combinación de las señales recibidas en los sensores para maximizar la relación señal ruido. Si suponemos ruido espacialmente correlado, la secuencia de estadísticos suficientes puede obtenerse mediante particularización de la ecuación 2.2.1 como,

$$b(m') = \sum_{n=-\infty}^{n=+\infty} (\mathbf{s}^H \mathbf{R}_{esp}^{-1} \mathbf{r}(n)) p(n - m'N) \quad (2.8.4.1)$$

que como vemos, será equivalente a utilizar el mismo ecualizador en todos los sensores. El decorrelador de símbolos es en este caso simplemente proporcional al decorrelador de sensor único, con la siguiente constante de proporcionalidad para normalizar la etapa de combinación espacial,

$$\delta_s(z^{-1}) = [\mathbf{s}^H \mathbf{R}_{esp}^{-1} \mathbf{s}]^{-1} \delta(z^{-1}) \quad (2.8.4.2)$$

Es decir, el hecho de que el ruido no esté correlado también temporalmente permite un desacoplo del procesado espacial y del temporal. El receptor óptimo se construye en base a un combinador espacial seguido en cascada por un ecualizador temporal.

## 2.9 Modelado de las Interferencias

El objetivo de esta sección es ver como la multiplicación por la matriz inversa de correlación realiza un bloqueo de las interferencias superpuestas a una señal. Modelaremos las características estadísticas de señales interferentes superpuestas a la señal CDMA como procesos Gausianos espacialmente localizados en espacio y en frecuencia. Prestaremos especial atención a interferentes de banda estrecha. Definimos cada una de las señales interferentes de la siguiente forma,

$$i_l(n) = i_l(n)s_l, \quad 1 \leq l \leq N_I \quad (2.9.1)$$

donde  $i_l(n)$  representa un proceso escalar y  $s_l$  es el vector de steering de la interferencia. Pueden darse dos casos en cuanto al modelado de la estadística del vector de steering: o bien consideramos que sus componentes son procesos estocásticos independientes del proceso escalar  $i_l(n)$  o bien lo consideramos invariante en el tiempo en cuyo caso la matriz de correlación del proceso  $i_l(n)$  será de rango unidad, es decir, podremos definir las estadísticas de segundo orden de la interferencia a partir de un menor número de parámetros. Podemos también modelar vectores de steering variantes en el tiempo suponiendo una función de correlación del vector de steering altamente correlada en tiempo (si suponemos que su estadística varía lentamente). Nos centraremos en el caso fijo de interferencias espacialmente localizadas. La función de correlación espacio-temporal de las secuencias interferentes más el ruido de sensor  $\mathbf{n}$  (que supondremos incorrelado) viene dada por,

$$\mathbf{R}_I(m) = E\mathbf{r}_I(n)\mathbf{r}_I^H(n+m), \quad \mathbf{r}_I(n) = \mathbf{n}(n) + \sum_{l=1}^{N_I} i_l(n)s_l \quad (2.9.2)$$

Suponiendo que las interferentes se encuentran incorreladas dos a dos, obtenemos la siguiente expresión para  $\mathbf{R}_I(m)$ ,

$$\mathbf{R}_I(m) = \sigma^2 \mathbf{I}\delta(m) + \sum_{l=1}^{N_I} (E i_l(n) i_l^*(n+m)) s_l s_l^H = \sigma^2 \mathbf{I}\delta(m) + \sum_{l=1}^{N_I} \rho_l(m) s_l s_l^H \quad (2.9.3)$$

donde  $\rho_l(m)$  es la secuencia de correlación temporal de cada una de las interferentes. En situaciones reales de comunicaciones, dichas interferentes suelen ser cicloestacionarias, lo cual permite un modelado más adecuado. En este estudio se suponen simplemente estacionarias. Podemos representar la anterior ecuación de forma más compacta si

agrupamos los vectores de enfoque de las interferentes en una matriz de steering y los procesos escalares asociados en un vector, llegando a la siguiente expresión,

$$\mathbf{R}_I(m) = \sigma^2 \mathbf{I} \delta(m) + \mathbf{S}_I \mathbf{D}_\rho(m) \mathbf{S}_I^H \quad (2.9.4)$$

De lo deducido en la sección 2.6.7 (ruido espacialmente coloreado) sabemos que la primera etapa de la conformación de haz se realiza en base a la multiplicación por la inversa de la transformada z de la matriz de correlación de la interferencia, es decir,

$$\mathbf{R}_I^{-1}(z^{-1}) = [\sigma^2 \mathbf{I} + \mathbf{S}_I \mathbf{D}_\rho(z^{-1}) \mathbf{S}_I^H]^{-1} \quad (2.9.5)$$

Resulta interesante comprobar si dicho filtrado anula los procesos interferentes. Para ello consideremos primeramente el caso en que aparece una única interferente. La matriz de correlación inversa puede calcularse entonces a partir de la identidad de Woodbury,

$$\mathbf{R}_I^{-1}(z^{-1}) = [\sigma^2 \mathbf{I} + \rho(z^{-1}) \mathbf{s} \mathbf{s}^H]^{-1} = \frac{1}{\sigma^2} \left( \mathbf{I} - \frac{\rho(z^{-1}) / \sigma^2}{1 + \mathbf{s}^H \mathbf{s} \rho(z^{-1}) / \sigma^2} \mathbf{s} \mathbf{s}^H \right) \quad (2.9.6)$$

donde observamos que para aquellas frecuencias donde la relación potencia interferente a ruido de fondo es elevada, la matriz de correlación inversa tiende a comportarse como un operador de proyección sobre el subespacio ortogonal a la interferencia,

$$\mathbf{R}_I^{-1}(z^{-1}) \xrightarrow{|\rho(z^{-1})| / \sigma^2 \gg 1} \frac{1}{\sigma^2} \left( \mathbf{I} - \mathbf{s} (\mathbf{s}^H \mathbf{s})^{-1} \mathbf{s}^H \right) \quad (2.9.7)$$

No obstante, aunque la interferencia esté espacialmente localizada, podemos observar que el algoritmo óptimo no la elimina totalmente en la etapa espacial. Definiendo el factor de polarización  $\beta(z^{-1})$  como sigue,

$$\beta(z^{-1}) = \frac{\mathbf{s}^H \mathbf{s} \rho(z^{-1}) / \sigma^2}{1 + \mathbf{s}^H \mathbf{s} \rho(z^{-1}) / \sigma^2} \quad (2.9.8)$$

la matriz de decorrelación se expresa como,

$$\mathbf{R}_I^{-1}(z^{-1}) = \frac{1}{\sigma^2} \left( \mathbf{I} - \mathbf{s} (\mathbf{s}^H \mathbf{s})^{-1} \mathbf{s}^H \right) + \frac{1}{\sigma^2} (1 - \beta(z^{-1})) \mathbf{s} (\mathbf{s}^H \mathbf{s})^{-1} \mathbf{s}^H \quad (2.9.9)$$

En este caso, la matriz de decorrelación se descompone en la suma de dos operadores de proyección sobre los subespacios ortogonal y paralelo a la interferencia, respectivamente. La norma del operador de proyección paralelo tiende a cero con relación al proyector ortogonal a medida que disminuye la varianza del ruido de fondo con respecto a la potencia de la interferencia. Nótese que la norma del operador paralelo depende de la frecuencia en función del espectro de potencia de la interferencia. La matriz de decorrelación se aproxima tanto más al proyector ortogonal sólo en aquellos márgenes de frecuencia donde el espectro de potencia se encuentra a un nivel suficientemente alto respecto al de ruido de fondo. En el resto del espectro, el factor de polarización tiende a cero, dejando pasar toda la señal. Una inspección más detallada de la descomposición revela que el proyector ortogonal es independiente de la frecuencia, realizando simplemente un bloqueo espacial de la señal interferente. el proyector paralelo es en cambio dependiente de la frecuencia y realiza la etapa de bloqueo frecuencial sobre la señal no bloqueada espacialmente. Obsérvese que puede descomponerse en un proyector espacial paralelo seguido de un filtrado temporal, igual para todos los sensores, donde la expresión del factor de polarización coincide exactamente con la del filtro de Wiener adaptado a la interferencia. El factor complementario  $1-\beta(z^{-1})$  se encarga entonces de realizar el bloqueo temporal. El diseño de la matriz de decorrelación de la interferencia a partir solamente de bloqueo espacial conduce a una solución subóptima ya que se desaprovecha la información disponible. El procesado espacio-temporal conjunto permite además eliminación de un número de interferencias superior al número de sensores gracias a los grados de libertad añadidos por el procesado temporal.

La deducción de la matriz de decorrelación de la interferencia para un número de señales interferentes superior a la unidad requiere ya de la aplicación del Lema de Inversión. Los resultados a los que se llega son parecidos. Puede llegarse también a la misma expresión de la matriz de decorrelación mediante aplicación recursiva de la identidad de Woodbury. Definamos las siguientes matrices de correlación acumulativas,

$$\mathbf{R}_k(z^{-1}) = \mathbf{R}_{k-1}(z^{-1}) + \rho_k(z^{-1}) \mathbf{s}_k \mathbf{s}_k^H \quad , \quad \mathbf{R}_0(z^{-1}) = \sigma^2 \mathbf{I} \tag{2.9.10}$$

según Woodbury,

$$\tag{2.9.11}$$

$$\begin{aligned}
\mathbf{R}_k^{-1}(z^{-1}) &= \mathbf{R}_{k-1}^{-1}(z^{-1}) - \frac{\rho_k(z^{-1})}{1 + \rho_k(z^{-1}) \mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1}) \mathbf{s}_k} \mathbf{R}_{k-1}^{-1}(z^{-1}) \mathbf{s}_k \mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1}) \\
&= \mathbf{R}_{k-1}^{-1}(z^{-1}) \left( \mathbf{I} - \frac{\rho_k(z^{-1})}{1 + \rho_k(z^{-1}) \mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1}) \mathbf{s}_k} \mathbf{s}_k \mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1}) \right) \\
&= \mathbf{R}_{k-1}^{-1}(z^{-1}) \left( \mathbf{I} - \beta_k(z^{-1}) \mathbf{s}_k \frac{\mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1})}{\mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1}) \mathbf{s}_k} \right)
\end{aligned}$$

Donde el factor de polarización  $\beta_k$  se ha definido de forma análoga a la ecuación 2.9.8,

$$\beta_k(z^{-1}) = \frac{\rho_k(z^{-1})}{\frac{1}{\mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1}) \mathbf{s}_k} + \rho_k(z^{-1})} = \frac{\rho_k(z^{-1})}{v_k(z^{-1}) + \rho_k(z^{-1})} \quad (2.9.12)$$

Siendo  $v_k(z^{-1})$  el espectro del ruido (señal CDMA) filtrado. La factorización final de la matriz de decorrelación de la interferencia en operadores de proyección se reduce a lo siguiente a partir de la ecuación 2.9.11,

$$\mathbf{R}_I^{-1}(z^{-1}) = \frac{1}{\sigma^2} \prod_{k=1}^{N_I} \left( \mathbf{I} - \beta_k(z^{-1}) \mathbf{s}_k \frac{\mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1})}{\mathbf{s}_k^H \mathbf{R}_{k-1}^{-1}(z^{-1}) \mathbf{s}_k} \right) \quad (2.9.13)$$

Es fácil ver que cada uno de los operadores va eliminando sucesivamente cada una de las interferencias. La cancelación óptima de  $N_I$  interferencias se consigue a partir de la concatenación de  $N_I$  etapas como la representada en la figura 2.9.1. Un conformador de haz de mínima varianza con memoria se utiliza para recuperar una de las interferencias. Nótese que se utiliza como matriz de correlación la de las  $k-1$  interferencias no canceladas previamente. A continuación un filtrado de Wiener realiza la interferencia  $k$ -ésima frente al ruido coloreado introducido por el conformador. Modulando el vector de steering de la interferente con la señal a la salida del filtro se reconstruye la contribución de la interferente a la entrada del bloque y se cancela para propagar la señal resultante a la siguiente etapa.

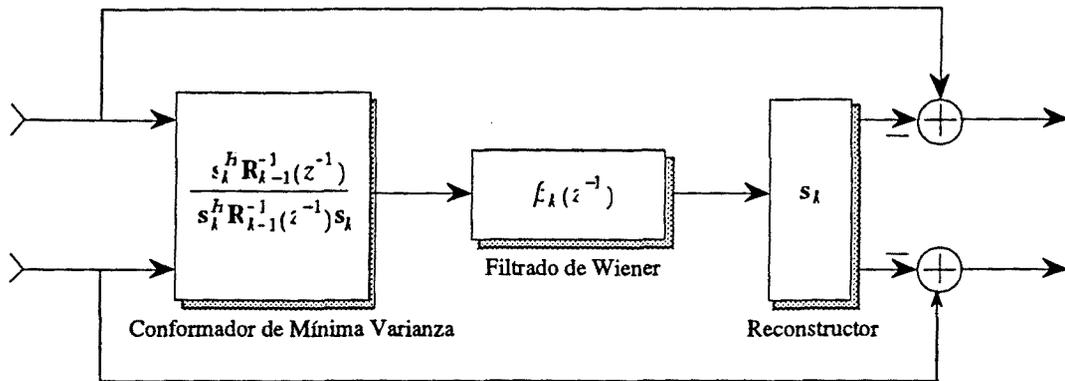


Figura 2.9.1: Representación de una etapa de cancelación de interferencia

Podemos llegar a una formulación equivalente en términos del Lema de Inversión. La matriz de decorrelación viene dada entonces por,

$$R_I^{-1}(z^{-1}) = \frac{1}{\sigma^2} (\mathbf{I} - \mathbf{S}(\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H) + \frac{1}{\sigma^2} (\mathbf{I} - \mathbf{B}(z^{-1})) \mathbf{S}(\mathbf{S}^H \mathbf{S})^{-1} \mathbf{S}^H \quad (2.9.14)$$

con una definición análoga de la matriz de polarización  $\mathbf{B}(z^{-1})$ ,

$$\mathbf{B}(z^{-1}) = \mathbf{S} \frac{1}{\sigma} \mathbf{D}_\rho(z^{-1})^{\frac{1}{2}} \left[ \mathbf{I} + \frac{1}{\sigma^2} \mathbf{S}^H \mathbf{D}_\rho(z^{-1}) \mathbf{S} \right]^{-1} \frac{1}{\sigma} \mathbf{D}_\rho(z^{-1})^{\frac{1}{2}} \mathbf{S}^H \quad (2.9.15)$$

Equivalente al filtrado vectorial de Wiener.

Supongamos a continuación un modelo de interferencia de banda estrecha localizado a una frecuencia portadora, su modelo vendrá dado por la siguiente ecuación,

$$i_l(n) = e^{j\omega_l n} \tilde{i}_l(n) \quad , \quad i_l(z^{-1}) = \tilde{i}_l(e^{j\omega_l} z^{-1}) \quad (2.9.16)$$

si la banda de la interferencia es muy estrecha, el operador de proyección equivalente presenta un nulo a la frecuencia portadora cuya profundidad dependerá del ruido de fondo, como puede deducirse de su filtro de Wiener equivalente, dejando pasar el resto de frecuencias temporales y espaciales.

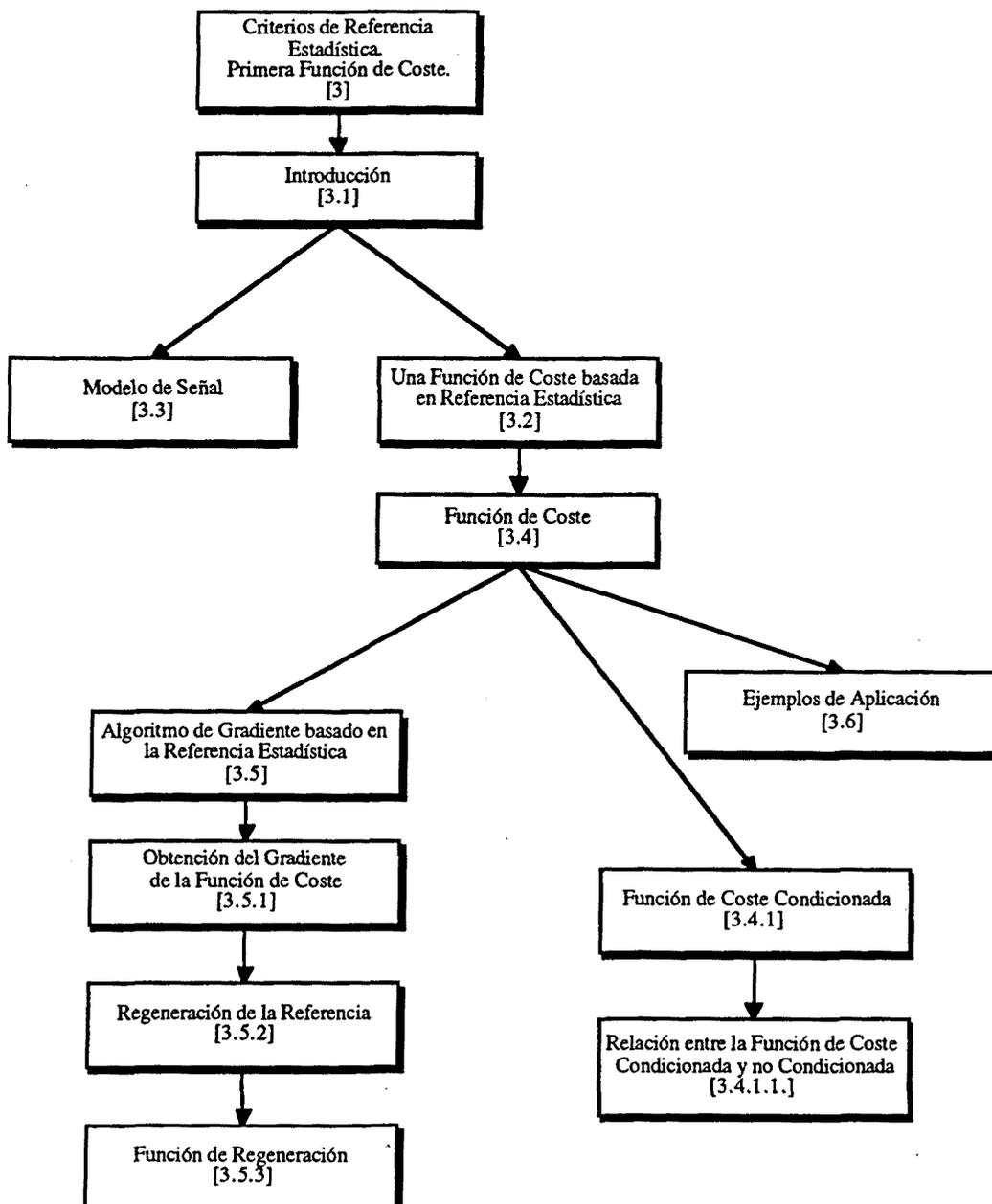


## Capítulo III

### Criterios de Referencia Estadística.

#### Primera Función de Coste

---





### 3.1 Introducción

En el capítulo anterior hemos examinado algoritmos lineales para la recuperación de secuencias de datos a partir de un conocimiento preciso de la estructura del modelo de la señal recibida en cuanto a efectos tales como correlación del ruido, sincronismos, interferencias, etc., para varios escenarios de interés en comunicaciones digitales como la decorrelación de señales CDMA o el problema de la ecualización. Hemos presentado dos criterios que permiten la recuperación de señales dependiendo del tipo de conocimiento que se posee del entorno: el criterio de referencia espacial, donde conocemos la estructura de los datos, y el criterio de referencia temporal, donde conocemos la secuencia transmitida. Dichos algoritmos han resultado ser estimadores no polarizados de mínima varianza en el caso de referencia espacial, no así en referencia temporal, donde aparecen efectos como la contracción de la constelación.

En este capítulo introduciremos un nuevo criterio para el planteamiento de funciones de coste en algoritmos adaptativos: el criterio de **referencia estadística**, donde utilizaremos conocimientos previos sobre las funciones de densidad de probabilidad presentes en el modelo de señal, en contraste con el método de referencia temporal, donde no existe ambigüedad en lo que se refiere a una realización particular de la señal o proceso estocástico que deseamos recuperar.

Los estudios teóricos realizados en el capítulo anterior se han centrado básicamente en el planteamiento de una función de coste y en el cálculo del vector de coeficientes que minimiza el criterio. Procederemos igualmente en este capítulo, haciendo además hincapié en el algoritmo adaptativo utilizado para la actualización de los coeficientes. Veremos que debido al carácter, ahora no-lineal, de las funciones de coste, no siempre puede llegarse a una expresión completamente cerrada del vector de coeficientes. No obstante, la estructura especial de las funciones de coste que propondremos admiten una riqueza considerable de interpretaciones, y conducen en muchos casos, a pesar de su naturaleza no-lineal, a resultados cerrados.

En este capítulo analizaremos diversos aspectos de la primera función de coste (una nueva función se presentará más adelante, en el capítulo V). En la sección 3.2 plantearemos la hipótesis inicial de la primera función de coste a partir de la discriminación de Kullback-Leibler como un ajuste de funciones de densidad de probabilidad. En la sección 3.3 introduciremos los modelos de señal para una y varias fuentes respectivamente. Estos

modelos se utilizarán también para el capítulo IV para el estudio de las soluciones (vectores de coeficientes) que minimizan la función de coste.

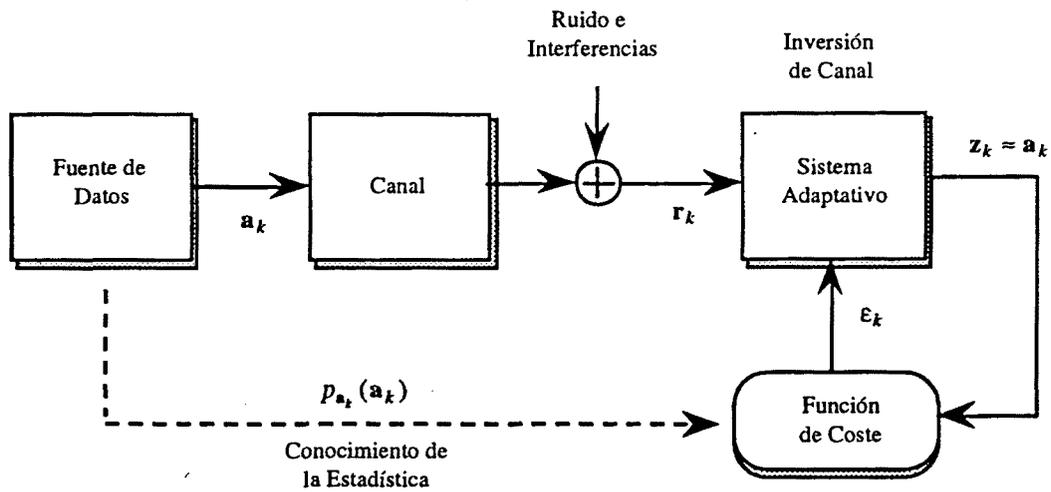


Figura 5.1.1: el objetivo del diseño es hallar una función de coste cuya minimización asegure que obtenemos una versión fiable de los datos emitidos por la fuente. Para ello utilizamos el conocimiento apriorístico de la estadística de los datos.

A partir de la propuesta para una función de coste en la sección 3.2, investigaremos también en la sección 3.4 una formulación alternativa en términos de las funciones de densidad condicionadas y veremos cómo se relaciona con un criterio de referencia temporal. Expresaremos la función de coste mediante un promediado no-lineal de la norma del error entre la variable aleatoria a la salida del sistema adaptativo y la variable aleatoria que deseáramos obtener. En la sección 3.5 analizaremos el algoritmo de gradiente de la función de coste y veremos cómo de él se deriva el concepto de la función de regeneración, la cual nos conduce a la definición de un error generalizado como proceso que controla la actualización de los coeficientes del filtro adaptativo. El planteamiento de ajuste de densidades de probabilidad se plantea en términos generales, independientemente de si la función de densidad deseada es continua o discreta. A modo ilustrativo de los conceptos anteriormente presentados, deducimos un algoritmo de módulo constante en la sección 3.6 para una función de densidad continua, así como un algoritmo para la recuperación de señales QPSK, para una función de densidad discreta.

### 3.2 Una Función de Coste basada en Referencia Estadística

En este apartado proponemos una posible función de coste basada en el concepto de referencia estadística introducido en la sección anterior. En un entorno real de filtrado adaptativo, el conocimiento "a priori" que tenemos del comportamiento del canal y de la estadística del ruido es prácticamente nulo. A fin de garantizar una robustez mínima de los algoritmos, pretendemos que el sistema adaptativo pueda capturar la señal deseada a partir de información lateral de la estadística de la fuente. La única información "apriorística" que pretendemos utilizar es el conocimiento previo de la estadística de la señal de entrada al canal. Nuestro propósito puede formularse como sigue:

- *Conocemos la distribución de la señal de entrada. Por lo tanto, queremos que la salida del sistema adaptativo presente una distribución de probabilidad o función de densidad tan cercana como sea posible a esta distribución previa. En consecuencia, debemos encontrar un algoritmo que fuerce una función de densidad dada a la salida del sistema.*

Para cumplir este propósito partiremos de una funcional proporcionada por la Teoría de la Información. A semejanza del criterio de referencia temporal, en el que definimos la "distancia" o función de coste entre la secuencia a la salida del sistema y la referencia como la potencia del error entre ambas señales (error cuadrático medio), necesitamos utilizar en el contexto de referencia estadística un criterio que cuantifique el error entre funciones de densidad de probabilidad. Consideremos la discriminación de Kullback-Leibler [Bla87] definida como,

$$D(p_1 \| p_0) = \int p_1(x) \ln p_1(x) dx - \int p_1(x) \ln p_0(x) dx \geq 0 \quad (3.2.1)$$

Aplicando este funcional al entorno de filtrado adaptativo, entendemos que  $p_1()$  es la función de densidad de probabilidad a la salida del sistema adaptativo,  $p_Y()$ , y que  $p_0()$  es la función de densidad objetivo o referencia. Aunque éste resulta un posible criterio óptimo, aparecen algunos problemas asociados a esta elección y que conducirán a una modificación de 3.2.1 como función de coste del algoritmo. Primero, debemos conocer cuál es la función de densidad óptima,  $p_0()$ , para cada escenario dado. Consideremos un ejemplo: debe tenerse en cuenta que a la salida del sistema aparecerá el ruido de entrada filtrado y con una determinada varianza, sumado a la señal de interés. Por lo tanto, deberemos modelar este efecto en la función de densidad  $p_0()$  pero el valor exacto de la varianza de ruido que

debemos utilizar es a priori desconocido. Por otra parte, nos encontramos en dificultades al intentar evaluar el primer término de la ecuación 3.2.1, menos la entropía del proceso aleatorio a la salida del estimador:  $-H(Y)$ , ya que su función de densidad nos es desconocida.

Aunque vemos en este primer enfoque del criterio de referencia estadística mediante la discriminación de Kullback que la posible realización de un algoritmo adaptativo se ve obstaculizada por el desconocimiento de las funciones de densidad de 3.2.1, podemos solventar este problema mediante modificaciones de 3.2.1 que nos conduzcan a algoritmos realizables. La función de coste por la que optamos es, por tanto, la minimización del segundo término de la ecuación 3.2.1, sin considerar el primer sumando (ver ecuación 3.2.2). En secciones posteriores analizaremos qué efectos tiene el hecho de despreciar el primer término en la ecuación de la discriminación.

En este trabajo proponemos dos funciones de coste para procesado adaptativo. La primera, la cual introducimos en esta misma sección constituye la primera parte del capítulo. La segunda requiere ciertos conocimientos previos que sólo pueden adquirirse realizando un estudio de esta primera función de coste. Por consiguiente posponemos su presentación hasta el capítulo 5. En lo sucesivo y por brevedad, nos referiremos a la primera función de coste como función de coste únicamente, explicitando el término segunda función de coste cuando sea necesaria la diferenciación. Veremos que algunos problemas presentados por la primera función de coste quedan resueltos por la segunda. En lo sucesivo nos referiremos a la primera y segunda función de coste como  $J$  y  $J_2$  respectivamente. Veremos cómo definir una estrategia en el caso  $J_2$  que nos permita implementar una pseudo-discriminación para el ajuste de densidades de probabilidad, y retomando plenamente la filosofía de la ecuación 3.2.1 aunque desde una óptica diferente.

El criterio propuesto se formaliza de la siguiente forma:

• Sea  $Y^{(w)}$  la variable aleatoria de salida del sistema que depende del vector de coeficientes,  $w$ , y sea  $\ln p_\phi()$  la familia de funciones de log-likelihood parametrizadas por el vector  $\phi$  que caracterizan la función de densidad deseada a la salida del sistema. El criterio de optimización consiste en minimizar la funcional definida por,

$$J(w, \phi) = -E_{Y^{(w)}} \ln p_\phi(y^{(w)}) = - \int_{-\infty}^{+\infty} p_{Y^{(w)}}(y) \ln p_\phi(y) dy \quad (3.2.2)$$

en el conjunto de parámetros  $[w, \phi]$ .

El vector de coeficientes  $w$  controla el sistema adaptativo y, por tanto, la función de densidad de probabilidad a su salida, mientras que el vector de parámetros  $\phi$  controla la forma de la no-linealidad sobre la que se evalúan los datos de salida.

Hemos hablado hasta ahora de "sistema" adaptativo y no de "filtro" adaptativo. Queremos recalcar que las funciones de coste que proponemos en este capítulo no son solamente aplicables a ecualización, sino que su campo de aplicación es absolutamente general. Es por lo tanto necesario diferenciar entre la función de coste de un sistema adaptativo, la arquitectura de este sistema y el modelo de señal involucrado. A medida que prosigamos veremos la aplicación de estas funciones de coste al campo del procesado de array y a la decorrelación de señales (temas que tratamos en esta tesis). De hecho, el criterio de forzar una función de densidad (conjunta) a la salida de un sistema ya no escalar sino MIMO (Multiple-Input Multiple-Output) con memoria, es absolutamente general. No obstante, los estudios teóricos y simulaciones que hemos realizado en esta tesis se circunscriben solamente al análisis de sistemas SISO (single-input single-output) MISO (multiple-input single-output) y SIMO (single-input multiple-output) adaptativos.

En la siguiente sección pasaremos a ocuparnos del modelo de señal que seguiremos en el resto de la tesis. Básicamente, introduciremos la notación para el modelo de una fuente y de varias fuentes.

### 3.3 Modelo de Señal

En esta breve sección introduciremos los modelos de señal que utilizaremos en el resto del capítulo. Consideraremos dos modelos posibles: el modelo de una fuente (3.3.1) y el modelo de varias fuentes (3.3.2). La distinción entre estos dos modelos resultará útil en los análisis teóricos que realizaremos en secciones posteriores. Un análisis general del comportamiento de la función de coste no resulta posible ya que su comportamiento depende en gran medida de la estadística de la señal de entrada. Por ello, desglosaremos el conjunto de casos posibles, considerando el modelo de una y varias fuentes en las secciones 4.2 y 4.3 respectivamente.

#### 3.3.1 Modelo de una Fuente

En esta sección presentamos el modelo de señal utilizado para tratar un simple problema de aplicación de la función de coste ya propuesta: la recuperación de un proceso aleatorio de distribución desconocida (en términos de su función de densidad de probabilidad) que ha experimentado una transformación lineal desconocida y que se encuentra inmerso en ruido Gaussiano de correlación también desconocida (en el apartado 3.3.2 generalizaremos el modelo de señal al caso de varias fuentes). Este proceso modula a un vector de steering de modo que la señal recibida se puede expresar mediante el siguiente modelo,

$$\mathbf{r}_k = a_k \mathbf{s}_k + \mathbf{n}_k \quad (3.3.1.1)$$

donde  $k$  denota el índice temporal,  $a_k$  son las muestras transmitidas por una fuente de distribución conocida y  $\mathbf{s}_k$  es el vector de steering (que se supone variante en el tiempo por defecto) y  $\mathbf{n}_k$  representa el ruido Gaussiano aditivo recibido en cada uno de los componentes. Nótese que este modelo simula la señal recibida por un array en ausencia de interferencias. La distribución espacial del ruido es desconocida así como las componentes del vector de steering generalizado. Por lo tanto, el modelo es aplicable a un array de geometría arbitraria. Como modelo de la estadística de la fuente podemos tomar la función de densidad de probabilidad de cualquier modulación digital: ASK, QAM, CPM.

Definamos ahora la arquitectura del sistema adaptativo (arquitectura lineal) que se utiliza para la recuperación de la señal como la multiplicación escalar de la señal recibida  $\mathbf{r}_k$  por un vector de coeficientes  $\mathbf{w}_k$ ,

$$y_k = \mathbf{w}_k^H \mathbf{r}_k \quad (3.3.1.2)$$

El objetivo que perseguimos es ahora definir un algoritmo que imponga una función de densidad de probabilidad similar (debido al ruido) a la de la fuente de señal,

$$y_k = a_k + n_k \Leftrightarrow p_{y_k} = p_{a_k + n_k} \quad (3.3.1.3)$$

de tal forma que la potencia del término de ruido se minimice. Demostraremos en la sección 4,2 según el modelo estadístico de la ecuación 3.3.1.1 que esto ocurre excepto en un factor de escala.

### 3.3.2 Modelo de Varias Fuentes

En esta sección consideraremos un modelo de señal más elaborado, incluyendo varias fuentes a diferencia de la sección 3.3.1. Definiremos la señal recibida de la siguiente forma,

$$\mathbf{r}_k = \sum_{i=1}^{i=N_u} a_k^i \mathbf{s}_k^i + \mathbf{n}_k = \mathbf{S}_k \mathbf{a}_k + \mathbf{n}_k \quad (5.3.1)$$

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 5.2, 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  $\mathbf{r}_k$ ).

### 3.4 Función de Coste

Según hemos visto en la sección 3.2, la función de coste propuesta consiste en la minimización de la esperanza de la función log-likelihood basada en la función de densidad de probabilidad deseada a la salida del sistema y evaluada a la salida del filtro adaptativo. Más adelante, en las secciones 4.2 y 4.3 consideraremos el caso más general de una no-linealidad en lugar de la función log-likelihood. En esta sección presentamos la primera función de coste en base a la notación que utilizaremos en el resto del trabajo. Introduciremos también una sencilla reformulación de la función de coste, basada en el Teorema Fundamental de la Esperanza y que nos resultará de utilidad en varios desarrollos teóricos posteriores.

El conjunto óptimo de coeficientes que minimiza la función de coste  $J$  se obtiene como sigue,

$$w_k = \arg \min \left\{ J = -E_{y_k} \ln p_{a_k+n_t}(y_k) \right\} = \arg \min \left\{ \int_C p_{y_k}(y_k) \ln p_{a_k+n_t}(y_k) dy_k \right\} \tag{3.4.1}$$

Donde  $a_k$  representa la variable aleatoria deseada en ausencia de ruido y  $n_t$  el término de ruido siempre presente a la salida del sistema, necesario para modelar con mayor detalle la función de densidad real a la salida del sistema. La función de densidad objetivo resulta ser entonces la correspondiente a la suma de ambas variables:  $a_k+n_t$ . En un caso más general, consideraremos que la función de densidad deseada es parametrizable (por ejemplo en la varianza del ruido  $n_t$ :  $\sigma_t$ ), de forma que el proceso de minimización se extiende también a los parámetros, a los cuales nos referimos mediante el vector  $\phi_k$ ,

$$\{w_k, \phi_k\} = \arg \min \left\{ J = -E_{y_k} \ln p_\phi(y_k) \right\} \tag{3.4.2}$$

Esta parametrización resulta útil para conseguir mejores prestaciones del sistema adaptativo y podemos entenderla como una adaptación de la no-linealidad log-likelihood a la estadística de la señal de entrada. Antes de proceder conviene expresar la función de coste en un formato más conveniente para los propósitos de minimización. Si aplicamos ahora el gradiente con respecto a los coeficientes, la ecuación 3.4.1 sugiere que el gradiente no puede intercambiarse con el operador esperanza ya que éste se define sobre la variable aleatoria a la salida del filtro, la cual depende a su vez de los coeficientes. Resulta más ilustrativo expresar la esperanza en función de variables aleatorias que no dependan de la

cantidad que se desea minimizar. Por lo tanto, recurrimos al Teorema Fundamental de la Esperanza,

$$J = -E_{y_k} \ln p_\phi(y_k) = -E_{a_k, n_k} \ln p_\phi(y_k(a_k, n_k)) = -E_{a_k, n_k} \ln p_\phi(y_k) \tag{3.4.3}$$

donde vemos que el gradiente y la esperanza sí pueden intercambiarse según esta nueva formulación puesto que las variables aleatorias sobre las que definimos el operador esperanza no son ya variables presentes a la salida, sino a la entrada, e independientes por tanto del vector de coeficientes a partir del cual se define el gradiente de la función de coste.

En la siguiente sección trataremos de cómo enlazar la primera función de coste con el criterio de referencia temporal. Es decir, veremos cómo mediante una sencilla manipulación podemos expresar la función de coste en términos de funciones de densidad condicionadas por una referencia conocida y promediando sobre el conjunto de todas las posibles referencias.

### 3.4.1 Función de Coste Condicionada

Hasta ahora hemos abordado el problema del diseño de una función de coste para algoritmos ciegos basados en el conocimiento de la estadística de la señal que se desea recuperar. Intuitivamente, cuanto mayor sea el conocimiento de las características de la señal transmitida, mejor comportamiento podremos obtener del algoritmo adaptativo. En particular esto será cierto cuanto menor sea la incertidumbre con respecto a la secuencia transmitida. En este apartado veremos cómo se modifica la función de coste cuando disponemos de información lateral sobre la secuencia de llegada. Plantearemos el criterio, no sobre la función de densidad deseada de los datos, sino sobre la función de densidad condicionada a la secuencia de referencia. En forma matemática, el criterio se convierte en,

$$J = -E_{z_k} \ln p_{a_k + \eta_k | d_k}(z_k | d_k) \tag{3.4.1.1}$$

donde  $d_k$  denota a la secuencia de referencia conocida. El hecho de conocer la secuencia transmitida nos evita el utilizar la función de densidad de la variable aleatoria deseada, promediada sobre todas las secuencias posibles. La función de densidad condicionada viene expresada en función de la densidad del ruido puesto que el conocimiento previo de la

secuencia de referencia  $d_k$  nos informa sobre la media estadística que debemos esperar en cada instante de tiempo de la variable aleatoria de salida del sistema adaptativo, es decir,

$$p_{a_k + \eta_k | d_k}(z_k | d_k) = p_{\eta_k}(z_k - d_k) = \frac{1}{\pi \sigma_{\eta_k}^2} e^{-|z_k - d_k|^2 / \sigma_{\eta_k}^2} \tag{3.4.1.2}$$

con lo cual la función de coste se convierte en,

$$J = -E_{z_k} \ln \frac{1}{\pi \sigma_{\eta_k}^2} e^{-|z_k - d_k|^2 / \sigma_{\eta_k}^2} = \ln \pi \sigma_{\eta_k}^2 + \frac{1}{\sigma_{\eta_k}^2} E_{z_k} |z_k - d_k|^2 \tag{3.4.1.3}$$

Vemos pues que en el caso de conocer la secuencia de llegada, se llega al criterio equivalente de mínimos cuadrados en lo que respecta a la estimación de los coeficientes. La estimación del parámetro varianza no afecta en absoluto a la estimación del vector de coeficientes  $w_k$ , con lo cual el mínimo de esta función de coste nos conduce a la solución de Wiener. Basta con calcular el gradiente respecto  $w_k$  para destacar esta independencia,

$$\nabla_{w_k^H} J = \frac{1}{\sigma_{\eta_k}^2} E_{z_k} (z_k - d_k)^* \nabla_{w_k^H} z_k \tag{3.4.1.4}$$

donde observamos que al igualar a cero, desaparece la dependencia con respecto al parámetro varianza. La obtención del parámetro varianza se deriva a partir de la igualación a cero de su gradiente como,

$$\nabla_{\sigma_{\eta_k}^2} J = \frac{1}{\sigma_{\eta_k}^2} - \frac{1}{\sigma_{\eta_k}^4} E_{z_k} |z_k - d_k|^2 = 0 \Rightarrow \sigma_{\eta_k}^2 = E_{z_k} |z_k - d_k|^2 \tag{3.4.1.5}$$

El valor óptimo de la función de coste se obtiene tras sustitución de la varianza en la ecuación 3.4.1.3,

$$J_{\min} = \ln(\pi e E_{z_k} |z_k - d_k|^2) \tag{3.4.1.6}$$

Vemos que el valor mínimo de la función de coste es formalmente equivalente a la entropía de una variable aleatoria Gaussiana de varianza igual a la expresada en la ecuación 3.4.1.5. Dicha relación es solo formal puesto que en general no ocurrirá que el error  $z_k - d_k$  en el mínimo de la función de coste sea Gaussiano ya que aparece un término residual de potencia de señal deseada como hemos visto en el capítulo anterior al estudiar la solución de Wiener y como de hecho ocurre con esta función de coste como se verá en la sección 4.2.

### 3.4.1.1 Relación entre la función de coste condicionada y no condicionada

En este apartado obtendremos una relación general entre la función de coste condicionada y la no condicionada con información lateral. Introduciremos esta información lateral mediante el parámetro  $\theta$  de forma que la disponibilidad de este conocimiento viene dado por la variable aleatoria  $b_k|\theta$ , donde  $b_k$  es la variable aleatoria deseada a la salida del sistema adaptativo. Las funciones de coste condicionada y no condicionada se expresan entonces como,

$$J_\theta = -E_{z_k} \ln p_{b_k|\theta}(z_k) \quad , \quad J = -E_{z_k} \ln p_{b_k}(z_k) \quad (3.4.1.1.1)$$

podemos relacionar de forma sencilla ambas densidades de probabilidad según la siguiente igualdad,

$$p_{b_k}(z_k) = E_\theta p_{b_k|\theta}(z_k) \quad (3.4.1.1.2)$$

con lo cual la función de coste no condicionada se obtiene mediante la función de densidad condicionada promediada sobre la variable aleatoria  $\theta$ . En el contexto de utilización de una referencia, llegamos a la solución de Wiener mediante utilización de la función condicionada, o en el caso de carecer de dicha referencia utilizamos en la construcción de la función de coste un promediado o ponderación sobre el conjunto de referencias posibles. Observemos que la derivación es válida tanto para distribuciones continuas como discretas.

$$J_\theta = -E_{z_k} \ln p_{b_k|\theta}(z_k) \quad , \quad J = -E_{z_k} \ln E_\theta p_{b_k|\theta}(z_k) \quad (3.4.1.1.3)$$

Bajo la hipótesis de ruido Gausiano, podemos expresar la función de densidad condicionada de la forma siguiente,

$$p_{b_k|\theta}(z_k) = \frac{1}{\pi\sigma_i^2} e^{-|z_k - a'_k|^2 / \sigma_i^2} \quad (3.4.1.1.4)$$

donde  $a'_k$  es un valor correspondiente a la distribución de la variable aleatoria deseada (pasa a jugar el papel del parámetro  $\theta$ ). Podemos entonces expresar la función de coste condicionada de la siguiente forma,

$$(3.4.1.1.5)$$

$$J = -E_{z_k} \ln E_{a'_k} \frac{1}{\pi\sigma_i^2} e^{-|z_k - a'_k|^2 / \sigma_i^2} = \ln \pi\sigma_i^2 - E_{z_k} \ln E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2}$$

mediante la introducción de la siguiente potencia del error, podemos llegar a una expresión de la ecuación anterior formalmente parecida al caso de la función de coste condicionada a la secuencia transmitida conocida, la cual denotamos como  $d_k$ ,

$$|\epsilon_J|^2 \stackrel{def}{=} -\sigma_i^2 \ln E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} \Rightarrow J = \ln \pi\sigma_i^2 + \frac{1}{\sigma_i^2} E_{z_k} |\epsilon_J|^2 \tag{3.4.1.1.6}$$

Notemos que la anterior definición es válida puesto que la expresión de  $|\epsilon_J|^2$  resulta ser definida positiva por construcción. En efecto,

$$0 \leq e^{-|z_k - a'_k|^2 / \sigma_i^2} \leq 1 \Rightarrow 0 \leq E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} \leq 1 \Rightarrow +\infty > -\ln E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} \geq 0 \tag{3.4.1.1.7}$$

Destacamos el parecido formal de la nueva expresión de J con la entropía de una variable aleatoria Gaussiana. Vemos que la potencia del error  $\epsilon_J$  se obtiene como un promedio no lineal de los errores  $z_k - a'_k$  sobre la distribución de la variable aleatoria  $a'_k$ . La definición anterior de  $|\epsilon_J|^2$  realiza la contribución de aquellos valores de  $a'_k$  próximos al valor  $z_k$  mediante la exponenciación de la diferencia y su promedio sobre la distribución. Valores poco probables de  $a'_k$  dado  $z_k$  contribuyen apreciablemente menos a la potencia del error  $\epsilon_J$ .

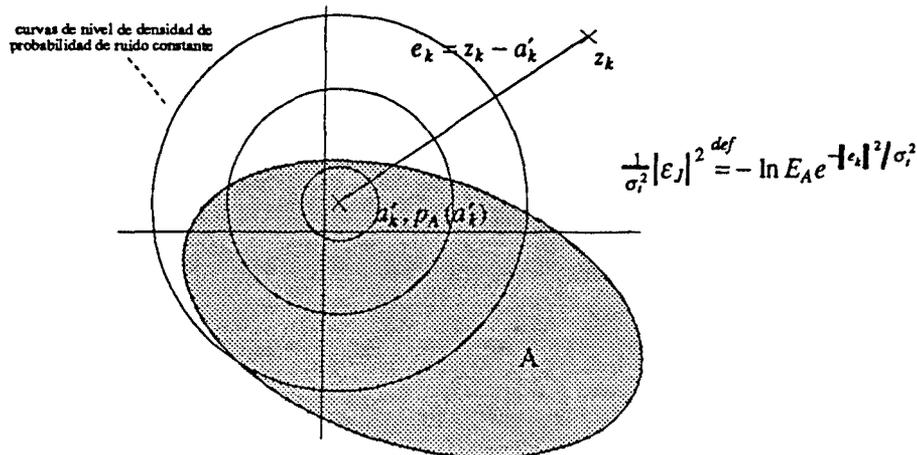


Figura 3.4.1.1.1: representación conceptual de la definición de la norma del error  $|\epsilon_J|^2$ . Hemos representado en trazo sombreado el conjunto de valores que puede tomar la variable aleatoria A deseada en el plano complejo. Cada punto de la región A contribuye al error en función de su densidad de probabilidad según A y de su distancia a la muestra  $z_k$  recibida según el promedio no lineal representado en la figura.

Más adelante cuantificaremos este efecto mediante lo que llamaremos funciones  $q(\cdot)$  o funciones de calidad, las cuales están relacionadas también con estimaciones de probabilidad de  $a'_k$  dado  $z_k$ .

Podemos ver de la definición del error  $\epsilon_j$  que éste se reduce al caso de mínimos cuadrados cuando consideramos conocida la secuencia  $a_k$ , puesto que entonces,

$$\begin{aligned} p_{a'_k}(a'_k) = \delta(a'_k - d_k) &\Rightarrow E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} = \int_C \delta(a'_k - d_k) e^{-|z_k - a'_k|^2 / \sigma_i^2} da'_k = e^{-|z_k - d_k|^2 / \sigma_i^2} \\ &\Rightarrow |\epsilon_j|^2 = |z_k - d_k|^2 \end{aligned} \tag{3.4.1.1.8}$$

A partir de esta nueva formulación resulta fácil relacionar la potencia del error  $\epsilon_j$  con el error generalizado  $\epsilon_k$ . Consideremos la derivada de la potencia  $|\epsilon_j|^2$  respecto a la variable  $z_k$ ,

$$\begin{aligned} \frac{d}{dz_k} |\epsilon_j|^2 &= -\sigma_i^2 \frac{d}{dz_k} \ln E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} = \frac{E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} (z_k - a'_k)^*}{E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2}} = E_{a'_k} (z_k - a'_k)^* q(z_k, a'_k) \\ &= \epsilon_k^*(z_k) \end{aligned} \tag{3.4.1.1.9}$$

siendo igual al error generalizado conjugado. Vemos así que  $|\epsilon_j|^2$  puede jugar también el papel de función de coste siempre que no tengamos en cuenta la minimización respecto al parámetro varianza  $\sigma_i$ . En la ecuación 3.4.1.1.9 hemos introducido por vez primera la expresión de las funciones  $q(\cdot)$  como un cociente de exponenciales. El pleno significado de esta definición se halla expuesto en la sección 3.5.

En las siguientes figuras hemos representado la función  $|\epsilon_j|^2$  para el caso particular de una función de densidad de una modulación BPSK de símbolos equiprobables dada por (supondremos variables aleatorias reales para este ejemplo),

$$p_{a_k}(a_k) = \frac{1}{2} \delta(a_k - 1) + \frac{1}{2} \delta(a_k + 1) \tag{3.4.1.1.10}$$

también representamos su derivada (error generalizado) y su segunda derivada, la cual se encuentra relacionada con el Hessiano y nos proporciona información sobre la estabilidad del algoritmo en un punto determinado del espacio de coeficientes. Podemos observar en la primera figura como en los máximos de densidad de probabilidad de la distribución deseada, la norma del error  $|\epsilon_j|^2$  exhibe una cuenca parabólica, análoga a la que tendríamos en el caso de disponer de una referencia de los símbolos enviados. La curvatura de la función cambia para diversos valores del parámetro varianza. Observando el error generalizado vemos como cruza el eje de abscisas con pendiente positiva en los puntos

correspondientes a los máximos de densidad de probabilidad. De este modo resulta coherente su definición como error generalizado.

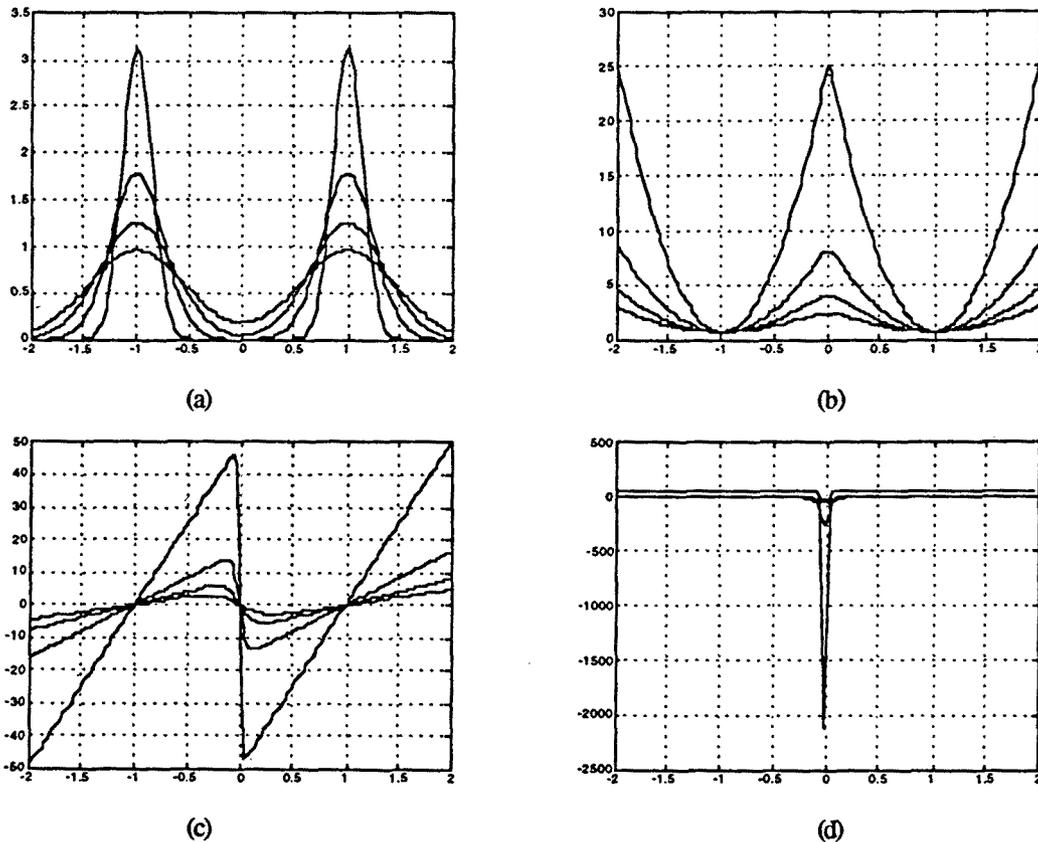


Figura (3.4.1.1.2): Representación de (a) la función de densidad de probabilidad de una variable aleatoria BPSK para diferentes valores de la varianza de ruido Gaussiano superpuesto; (b) norma del error  $|\epsilon_j|^2$  normalizado a la potencia de ruido; (c) error generalizado, derivada de (b); (d) curvatura de la norma del error normalizado. Se han escogido las varianzas 0.2, 0.35, 0.5 y 0.65. Nótese que las curvas más abruptas corresponden siempre a las varianzas menores.

Podemos ver que en aquellos puntos correspondientes a mínimos de densidad de probabilidad, se invierte la pendiente del error generalizado, convirtiéndolos en puntos inestables. Observamos que el error generalizado presenta una característica bastante lineal excepto en los puntos de transición correspondientes a los mínimos de densidad, donde su pendiente presenta un valor negativo (algoritmos basados en el gradiente tienden a diverger cuando evaluamos la no-linealidad de la función de coste sobre ellos). La pendiente depende del parámetro varianza que estemos considerando. Vemos en las gráficas de la segunda derivada como las zonas cercanas a los mínimos de densidad presentan un valor negativo (el Hessiano no estará definido positivo en estos puntos).

La justificación matemática de la característica lineal del error generalizado es fácilmente deducible de la expresión de la norma del error, puesto que para valores cercanos a alguno de los símbolos posibles, la exponencial asociada al otro símbolo presenta un valor relativo pequeño (en relación a la otra exponencial) para valores pequeños del parámetro varianza,

$$\frac{1}{\sigma^2} |\varepsilon_J|^2 = -\ln \left( \frac{1}{2} e^{-(z-1)^2/\sigma^2} + \frac{1}{2} e^{-(z+1)^2/\sigma^2} \right) \quad (3.4.1.1.11)$$

El error es por consiguiente aproximadamente igual al exponente de la exponencial más importante.

Nos hemos dedicado en esta sección a analizar la naturaleza de la primera función de coste, introduciendo también conceptos importantes como las funciones  $q(\cdot)$ , cuyas implicaciones trataremos en detalle en la siguiente sección. En la siguiente sección estudiaremos también el algoritmo de algoritmo de gradiente basado en el concepto de referencia estadística.

### 3.5 Algoritmo de Gradiente basado en la Referencia Estadística

En esta sección realizaremos el estudio del gradiente de la función de coste. Conceptos importantes como la función de regeneración de la referencia y el error generalizado aparecen de forma natural a partir de las ecuaciones del gradiente. Introduciremos también las funciones de calidad, o funciones  $q(\cdot)$ , fundamentales en las dos funciones de coste basadas en Referencia Estadística que presentamos en este trabajo.

#### 3.5.1 Obtención del Gradiente de la Función de Coste. Interpretación

Tras proponer una función de coste para algoritmos adaptativos ciegos en secciones anteriores, en esta sección analizaremos el algoritmo de gradiente de la función de coste. En los algoritmos de referencia temporal, el gradiente de los coeficientes se construye en base al error lineal entre la señal de salida del sistema y la señal de referencia. En este apartado daremos una interpretación al gradiente desde el punto de vista de regeneración ciega de la referencia a partir de los datos de llegada. Veremos también una formulación del criterio de ortogonalidad error-datos para un sistema ciego.

Calculemos el gradiente de la función de coste respecto a los coeficientes. Debido a que estamos trabajando con funciones de variable compleja, es necesario tener en cuenta antes de evaluar un gradiente con respecto a un vector complejo, si la función es o no analítica. Refiriéndonos a la reformulación de la función de log-likelihood en la ecuación 3.4.1.1.5, vemos que ésta se expresa en función del módulo al cuadrado de una variable compleja. Por lo tanto, podemos afirmar que la función de coste es no-analítica, es decir, admite derivadas tanto respecto a una variable determinada como respecto a su conjugada. Por lo tanto, el gradiente de la función de coste se expresa según la regla de la cadena como,

$$\nabla_{\mathbf{w}_k^H} J = -E_{a_k, n_k} \nabla_{\mathbf{w}_k^H} \ln p_{a_k+n_k}(y_k) = -E_{a_k, n_k} \left( \left( \frac{\partial}{\partial y_k} \ln p_{a_k+n_k} \right) \nabla_{\mathbf{w}_k^H} y_k + \left( \frac{\partial}{\partial y_k^*} \ln p_{a_k+n_k} \right) \nabla_{\mathbf{w}_k^H} y_k^* \right) \quad (3.5.1.1)$$

apareciendo por una parte la derivada de la función de coste respecto a  $y_k$  y por otra el gradiente de la arquitectura o gradiente de  $y_k$  con respecto al hermítico del vector de coeficientes,  $\mathbf{w}_k$ . Sabemos que la salida del sistema se expresa como  $y_k = \mathbf{w}_k^H \mathbf{r}_k$ , por

consiguiente, el segundo sumando de la ecuación 3.5.1.1 es nulo al no ser  $y_k$  conjugado función del  $w_k$  hermítico. La expresión simplificada del gradiente aparece entonces como,

$$\nabla_{w_k^H} J = -E_{a_k, n_k} \left( \frac{\partial}{\partial y_k} \ln p_{a_k+n_k} \right) \nabla_{w_k^H} y_k = -E_{a_k, n_k} \left( \frac{\partial}{\partial y_k} \ln p_{a_k+n_k} \right) r_k \quad (3.5.1.2)$$

donde hemos sustituido el gradiente de  $y_k$  por su valor de  $r_k$ , el vector de datos a la entrada del sistema. Observamos que en el mínimo (al igualar el gradiente a cero) existe incorrelación entre los datos de llegada  $r_k$ , y una función no lineal de la salida del filtro,  $y_k$ , igual a la derivada de la función de coste. Estableciendo un paralelo con el método de mínimos cuadrados, podemos asociar esta derivada a una función de error generalizada como justificaremos más adelante,

$$\varepsilon^*(y_k) = \frac{\partial}{\partial y_k} \ln p_{a_k+n_k}(y_k) = \frac{\partial}{\partial y_k} \ln E_{a_k} \frac{1}{\pi\sigma_i^2} e^{-\frac{1}{\sigma_i^2}|y_k-a_k'|^2} \quad (3.5.1.2)$$

Tras efectuar el cálculo de la derivada anterior, obtenemos,

$$\begin{aligned} \varepsilon^*(y_k) &= -\frac{\partial}{\partial y_k} \ln E_{a_k} \frac{1}{\pi\sigma_i^2} e^{-|y_k-a_k'|^2/\sigma_i^2} = \frac{1}{\sigma_i^2} \frac{E_{a_k'} e^{-|y_k-a_k'|^2/\sigma_i^2} (y_k-a_k')^*}{E_{a_k'} e^{-|y_k-a_k'|^2/\sigma_i^2}} \\ &= \frac{1}{\sigma_i^2} E_{a_k'} \frac{e^{-|y_k-a_k'|^2/\sigma_i^2}}{E_{a_k'} e^{-|y_k-a_k'|^2/\sigma_i^2}} (y_k-a_k')^* = \frac{1}{\sigma_i^2} E_{a_k'} q(y_k, a_k') (y_k-a_k')^* \end{aligned} \quad (3.5.1.3)$$

en términos de las funciones  $q(\cdot)$  ya definidas. Vemos pues que la función  $\varepsilon(\cdot)$  puede representar como una ponderación del error a la salida del filtro respecto cada uno de los elementos posibles de la constelación a través de los factores de peso  $q(\cdot)$ . Dichos factores son en realidad estimadores de la probabilidad de que  $y_k$  provenga de un  $a_k'$  determinado. La función  $\varepsilon(\cdot)$  juega el papel de un error medio o esperado cuando no disponemos de la referencia real de los símbolos transmitidos por la fuente, y normalizado con respecto al parámetro varianza de la función de coste. Operando un poco más con la ecuación 3.5.1.3 obtenemos una nueva interpretación de la función error,

$$\begin{aligned} \varepsilon(y_k) &= -\frac{1}{\sigma_i^2} y_k E_{a_k'} q(y_k, a_k') + \frac{1}{\sigma_i^2} E_{a_k'} a_k' q(y_k, a_k') = -\frac{1}{\sigma_i^2} (y_k - \hat{a}_k) \\ \hat{a}_k &= E_{a_k'} a_k' q(y_k, a_k') \end{aligned} \quad (3.5.1.4)$$

donde hemos utilizado la propiedad de que la esperanza de la función  $q(\cdot)$  respecto la variable  $a_k$  es igual a la unidad,

(3.5.1.5)

$$E_{a_k} q(y_k, a'_k) = 1$$

como resulta inmediato demostrar por simple sustitución de  $q(\cdot)$  por su definición en la ecuación 3.5.1.3. Vemos pues que el error puede interpretarse como la salida del filtro menos una estimación o regeneración no-lineal del símbolo transmitido a través de la función de regeneración  $\hat{a}_k(y_k)$ . Es decir, utilizamos como referencia regenerada un promedio de todos los valores posibles que puede tomar la distribución de la variable aleatoria  $a_k$ , ponderando cada valor según sea la probabilidad de que la realización particular  $y_k$  provenga de un  $a_k$  determinado. Dicha ponderación se lleva a cabo mediante las funciones  $q(\cdot)$  de la ecuación 3.5.1.4.

Volvamos a la ecuación 3.5.1.1. Tenemos entonces que el gradiente puede expresarse como,

(3.5.1.6)

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

siempre que la arquitectura del sistema adaptativo sea la de un filtro lineal. En la ecuación anterior, el gradiente será distinto de cero siempre que haya correlación entre el error regenerado, obtenido como diferencia entre  $y_k$  y la referencia regenerada, y los datos.

A continuación, pasamos a realizar en el siguiente apartado un estudio detallado de la función de regeneración de la referencia derivada en la ecuación 3.5.1.4.

### 3.5.2 Regeneración de la Referencia

La función de error generalizado en un sistema adaptativo ciego se ha relacionado en la sección precedente con una función de regeneración de los datos transmitidos. Hemos visto anteriormente que el método de regeneración de referencia es independiente de la arquitectura del sistema adaptativo. En esta sección consideraremos algunos de los efectos ya deducidos de la función de coste desde el punto de vista de regeneración de la referencia. En particular estudiaremos con detalle la estructura de las funciones  $q(\cdot)$ , la función de regeneración y el problema de la contracción de la constelación. Veremos primero que la regeneración de referencia puede considerarse como una proyección no lineal de los datos sobre el espacio de señal (los términos proyección y espacio de señal o de ruido no deben

interpretarse en un sentido riguroso en este contexto puesto que se utilizan solamente de modo intuitivo. Es decir, la referencia regenerada se obtiene pasando las muestras recibidas por una no-linealidad sin memoria). Consideremos una vez más la expresión del regenerador de referencia,

$$\hat{a}_k = E_{a'_k} a'_k q(z_k, a'_k) \tag{3.5.2.1}$$

Podemos interpretar dicha fórmula como un producto escalar entre la distribución de la señal deseada sin ruido y la función  $q(\cdot)$  inducida por una salida determinada del filtro  $z_k$ . El Kernel del producto escalar lo constituye precisamente la función de densidad de  $a'_k$  como se puede deducir recurriendo a la formulación integral de la ecuación 3.5.2.1,

$$\hat{a}_k = \int_C p_{a'_k}(a'_k) a'_k q(z_k, a'_k) da'_k \tag{3.5.2.2}$$

Consideraremos algunos ejemplos prácticos de regeneración de referencia para una modulación real discreta, tipo 4-ASK (Amplitude Shift Keying). La fórmula anterior de regeneración, cuando se aplica a distribuciones discretas conduce a un conjunto de  $M$  funciones (donde  $M$  es la dimensión del alfabeto) indexadas por la variable  $a_k$  de  $q(z_k, a_k)$ . Dicho conjunto constituye un grupo de funciones indicatrices del símbolo transmitido a partir del dato  $z_k$  recibido. Con función indicatriz queremos significar una función que realiza un intervalo de valores sobre el resto. La influencia del parámetro  $\sigma_t$  en las funciones  $q(\cdot)$  determinará la fiabilidad de los datos regenerados. Observemos la figura 3.5.2.1 donde se representa el conjunto  $q(\cdot)$  para valores extremos del parámetro varianza.

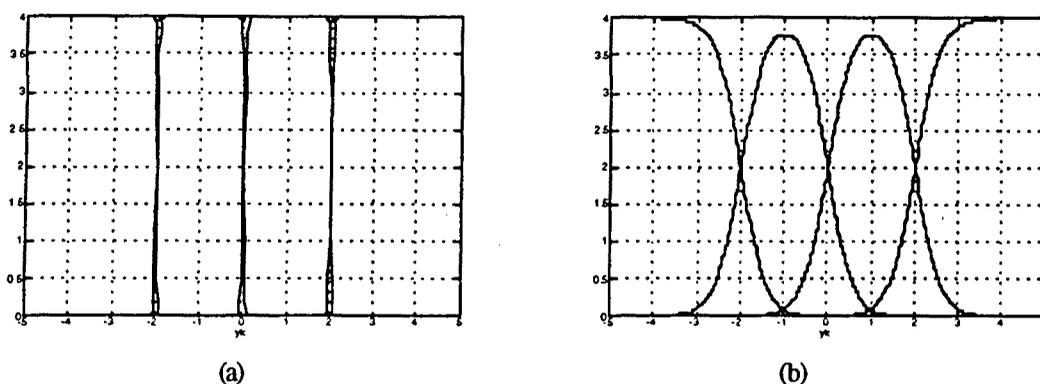


Figura 3.5.2.1: comparación de las funciones  $q(\cdot)$  para el caso particular de una modulación 4-ASK de símbolos equiprobables. Se ha considerado el alfabeto  $\{\pm 1, \pm 3\}$ . Se han considerado dos valores diferentes para el parámetro varianza: (a) 0.2 y (b) 0.75.

Observamos que las regiones delimitadas por cada función son más 'difusas' a medida que aumenta el valor  $\sigma_t$ , es decir, existe una mayor incertidumbre en la habilidad del algoritmo para discernir los datos. En la siguiente figura consideraremos la evolución de una de las funciones  $q(.)$  a medida que aumenta el parámetro varianza.

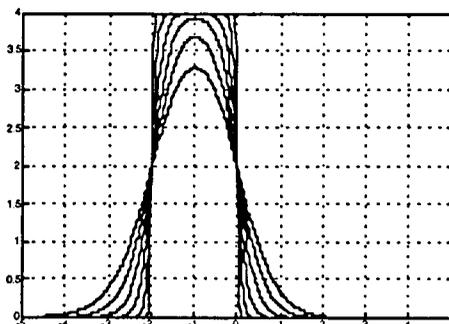


Figura 3.5.2.2: representación de la función  $q(.)$  correspondiente al símbolo '-1' para valores crecientes del parámetro varianza, ( $\sigma_t$ : 0.2,0.35,0.5,0.65,0.8,0.95).

Obsérvese en la figura 3.5.2.2 que cuando  $\sigma_t \rightarrow 0$ , la función  $q(-1)$  tiende a una función rectangular centrada en -1 y de amplitud 2, de tal forma que sus flancos coinciden con las fronteras de decisión del símbolo '-1' bajo la condición de equiprobabilidad. La amplitud de la función rectangular es precisamente el inverso de la probabilidad de ocurrencia del símbolo, o sea,  $1/0.25 = 4$ , o el número de niveles de la modulación. Demostraremos más adelante que cuando  $\sigma_t \rightarrow 0$ , las funciones  $q(.)$  tienden a pasar por 2, la mitad del inverso de la probabilidad, en la zona frontera de la decisión. Fuera de esta región, su valor es inferior a la mitad de la amplitud.

En la siguiente figura pasaremos a considerar un ejemplo diferente donde dejamos de lado la equiprobabilidad de los símbolos. De este modo ganamos cierta intuición de cómo se comportan las funciones  $q(.)$  cuando consideramos no ya modulaciones discretas sino modulaciones con densidad de probabilidad continua. Podemos interpretar esta simplificación como un muestreo de la densidad continua. Consideramos igualmente una modulación 4-ASK donde la matriz de probabilidad de los símbolos viene dada por  $P=\{1/8,2/8,4/8,1/8\}$  a modo ilustrativo (el hecho de que sea una modulación de media no nula es aquí irrelevante).

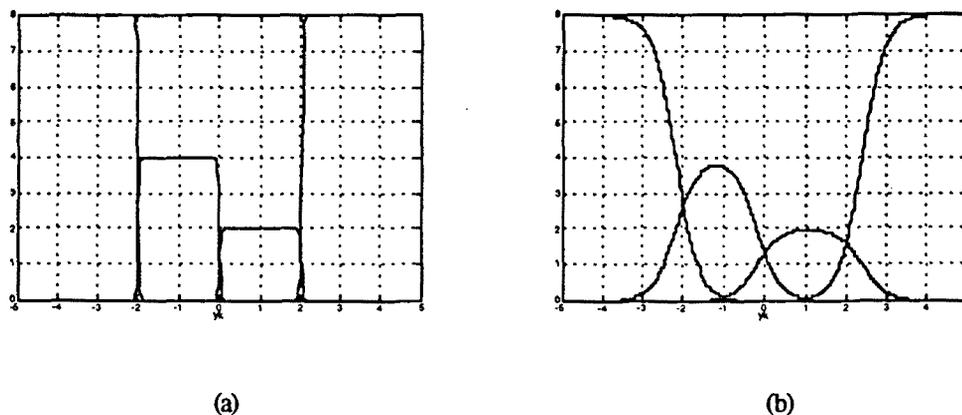


Figura 3.5.2.3: comparación de las funciones  $q(.)$  para el caso particular de una modulación 4-ASK de símbolos con probabilidades  $\{1/8, 2/8, 4/8, 1/8\}$ . Se han considerado dos valores diferentes para el parámetro varianza: (a) 0.2 y (b) 0.75.

Podemos observar como ahora ha variado la amplitud de las funciones  $q(.)$  correspondientes. Su amplitud viene modulada por la inversa de la probabilidad de cada símbolo. Es decir, símbolos con mayor probabilidad obtienen menor amplitud (este efecto se justificará más adelante). Vemos que las fronteras de los flancos no han variado con respecto al caso equiprobable. Pasamos a considerar ahora la evolución de la función  $q(-1)$  con respecto al parámetro varianza en la figura 3.5.2.4.

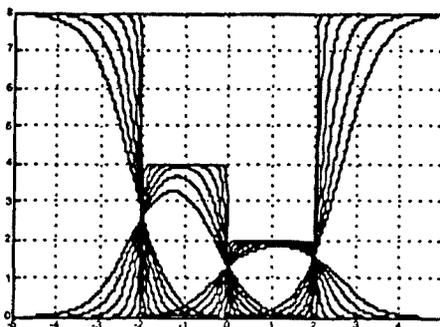


Figura 3.5.2.4: representación de la función  $q(.)$  correspondientes a cada símbolo para valores crecientes del parámetro varianza,  $\{\sigma_t: 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$ .

Obsérvese ahora como las funciones  $q(-1)$  siguen cortándose en un mismo punto para diferentes valores del parámetro varianza para un mismo símbolo y para sus símbolos vecinos. No obstante, debido a que los símbolos contiguos son de diferente probabilidad, la altura del punto de corte ya no es la mitad de la amplitud, sino un valor proporcional a la relación de probabilidades con el símbolo contiguo. Podemos observar también una

polarización del pico de la función hacia el símbolo de menor probabilidad. Esto se debe simplemente a un efecto de compensación.

Matemáticamente es fácil demostrar la cota superior de las funciones  $q(\cdot)$  cuando tenemos en cuenta modulaciones discretas. Según la definición de  $q(\cdot)$ , éstas se pueden expresar como,

$$p(a_k) = \sum_{i=1}^{i=M} p_i \delta(a_k - a_k^i) \Rightarrow q(z_k, a_k) = \frac{e^{-|z_k - a_k|^2 / \sigma_t^2}}{E_{a_k^i} e^{-|z_k - a_k^i|^2 / \sigma_t^2}} = \frac{e^{-|z_k - a_k|^2 / \sigma_t^2}}{\sum_{i=1}^{i=M} p_i e^{-|z_k - a_k^i|^2 / \sigma_t^2}} \quad (3.5.2.3)$$

Donde M es la dimensión del alfabeto. Puesto que todos los términos del denominador son positivos, podemos acotarlo inferiormente por cualquiera de sus sumandos, lo cual conduce inmediatamente a una cota superior de cada una de las M funciones  $q(\cdot, a_k)$ ,

$$\sum_{i=1}^{i=M} p_i e^{-|z_k - a_k^i|^2 / \sigma_t^2} \geq p_j e^{-|z_k - a_k^j|^2 / \sigma_t^2} \forall j \Rightarrow q(z_k, a_k^j) \leq \frac{1}{p_j} \quad (3.5.2.4)$$

lo cual concuerda con los resultados obtenidos en las figuras anteriores. La medida en que  $q(z_k, a_k^j)$  es menor que  $1/p_j$  viene dada por la importancia relativa de los términos que hemos despreciado en el denominador. El resto de términos será tanto mayor cuanto más grande sea  $\sigma_t$ , como puede verse en las figuras. La igualdad de la ecuación 3.5.2.4 sólo se cumple en un intervalo determinado cuando  $\sigma_t \rightarrow 0$ , puesto que entonces sólo una de las exponenciales del denominador es claramente dominante frente a las otras. Consideremos la expresión de  $q(\cdot)$  cuando pasamos el numerador dividiendo al denominador,

$$q(z_k, a_k^j) = \frac{1}{p_j + \sum_{i \neq j} p_i e^{-\left(|z_k - a_k^i|^2 - |z_k - a_k^j|^2\right) / \sigma_t^2}} \quad (3.5.2.5)$$

resulta fácil observar a partir de esta expresión cómo el comportamiento de las funciones  $q(\cdot)$  viene definido por las distancias Euclídeas de la muestra  $z_k$  con respecto a los centroides de la constelación. A medida que  $\sigma_t$  tiende a cero aquellos intervalos más cercanos en términos de la distancia Euclídea al centroide j, harán que las exponenciales asociadas tiendan a cero de tal forma que  $q(\cdot)$  sea igual al inverso de la probabilidad del centroide en ese intervalo determinado. Fuera de dicho intervalo se invierte el signo del exponente de la integral, por lo cual  $q(\cdot)$  tiende a cero. En la frontera del intervalo, existirán dos exponenciales contiguas de exponente nulo, de tal forma que  $q(\cdot)$  tiene a la inversa de la suma de ambas probabilidades. Resumiendo, tenemos que,

$$\lim_{\sigma_i \rightarrow 0} q(z_k, a_k^j) = \begin{cases} p_j^{-1} & , \min_{i \neq j} |z_k - a_k^i|^2 > |z_k - a_k^j|^2 \\ 0 & , \min_{i \neq j} |z_k - a_k^i|^2 < |z_k - a_k^j|^2 \\ (p_j + p_i)^{-1} & , \min_{i \neq j} |z_k - a_k^i|^2 = |z_k - a_k^j|^2 \end{cases} \quad (3.5.2.6)$$

lo cual concuerda con las figuras anteriores. Un punto interesante a tener en cuenta es la acotación inferior de las funciones  $q(\cdot)$ . Consideremos el denominador,

$$p_j + e^{+|z_k - a_k^j|^2 / \sigma_i^2} \sum_{i \neq j} p_i e^{-|z_k - a_k^i|^2 / \sigma_i^2} < p_j + e^{+|z_k - a_k^j|^2 / \sigma_i^2} \sum_{i \neq j} p_i = p_j + e^{+|z_k - a_k^j|^2 / \sigma_i^2} (1 - p_j) \quad (3.5.2.7)$$

obteniendo finalmente límites inferiores y superiores para el margen de variación de las funciones  $q(\cdot)$ ,

$$\left[ p_j + e^{+|z_k - a_k^j|^2 / \sigma_i^2} (1 - p_j) \right]^{-1} < q(z_k, a_k^j) \leq p_j^{-1} \quad (3.5.2.8)$$

Nos es posible no obstante obtener una desigualdad aún más estricta definiendo el símbolo más cercano en distancia Euclídea al conjunto de símbolos diferentes de  $a_k^j$ ,

$$a_k^0 = \arg \min_{a_k^i, i \neq j} |z_k - a_k^i|^2 \quad (3.5.2.9)$$

Volviendo al denominador de la función  $q(\cdot)$  tenemos,

$$a_k^0 p_j + e^{+|z_k - a_k^j|^2 / \sigma_i^2} \sum_{i \neq j} p_i e^{-|z_k - a_k^i|^2 / \sigma_i^2} < p_j + e^{+|z_k - a_k^j|^2 / \sigma_i^2} e^{-|z_k - a_k^0|^2 / \sigma_i^2} (1 - p_j) \quad (3.5.2.10)$$

con lo cual obtenemos el nuevo margen de variación,

$$\left[ p_j + e^{(|z_k - a_k^j|^2 - |z_k - a_k^0|^2) / \sigma_i^2} (1 - p_j) \right]^{-1} < q(z_k, a_k^j) \leq p_j^{-1} \quad (3.5.2.11)$$

La cota inferior es suficientemente estricta para un margen de variación de  $\sigma_i$  como para que pueda utilizarse como una aproximación de la función  $q(\cdot)$ . Nótese que así logramos una reducción importante en la complejidad de cálculo de la no-linealidad en vistas a una implementación práctica. En el caso de una constelación compleja, existen zonas del plano fase-cuadratura donde la muestra recibida  $z_k$ , puede hallarse a similar distancia de varios

centroides simultáneamente. En este caso podemos refinar aún más la cota inferior teniendo en cuenta el número de vecinos,

$$q(z_k, a_k^j) \approx \left[ p_j + e^{\left( |z_k - a_k^j|^2 - |z_k - a_k^0|^2 \right) / \sigma_i^2} \left( 1 - \sum_{N, \text{vecinos}} p_i \right) \right]^{-1} \tag{3.5.2.12}$$

Cuando pasamos a considerar distribuciones continuas, las acotaciones anteriores requieren mayor complejidad. Consideremos primero el comportamiento de  $q(\cdot)$  en el límite  $\sigma_i \rightarrow 0$ . Normalizaremos nominador y denominador previamente por el factor  $1/\pi\sigma_i^2$ . Siempre que el límite del denominador sea finito y no nulo, podemos expresar el límite del cociente como el cociente de límites,

$$\lim_{\sigma_i \rightarrow 0} q(z_k, a_k) = \lim_{\sigma_i \rightarrow 0} \frac{\frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k|^2 / \sigma_i^2}}{E_{a_k'} \frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k'|^2 / \sigma_i^2}} = \frac{\lim_{\sigma_i \rightarrow 0} \frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k|^2 / \sigma_i^2}}{\lim_{\sigma_i \rightarrow 0} E_{a_k'} \frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k'|^2 / \sigma_i^2}} \tag{3.5.2.13}$$

Consideremos ahora en detalle el denominador,

$$\begin{aligned} \lim_{\sigma_i \rightarrow 0} E_{a_k'} \frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k'|^2 / \sigma_i^2} &= \int_C p_{a_k'}(a_k') \lim_{\sigma_i \rightarrow 0} \frac{1}{\pi\sigma_i^2} e^{-|z_k - a_k'|^2 / \sigma_i^2} da_k' \\ &= \int_C p_{a_k'}(a_k') \delta(z_k - a_k') da_k' = p_{a_k'}(y_k) \end{aligned} \tag{3.5.2.14}$$

puesto que una gaussiana tiende asintóticamente a una función delta cuando su varianza tiende a zero. El denominador es puesto la función de densidad de la fuente evaluada en  $z_k$ . Es decir, no se supone incertidumbre en los datos. Considerando el límite del denominador, también una función delta, llegamos a la siguiente igualdad asintótica,

$$\lim_{\sigma_i \rightarrow 0} q(z_k, a_k) = p_{a_k}^{-1}(z_k) \delta(z_k - a_k) \tag{3.5.2.15}$$

donde notamos que es precisamente el caso límite de las figuras anteriores cuando la varianza tiende a cero, es decir, las funciones rectangulares tienden a convertirse en deltas. La anchura de las funciones rectangulares va disminuyendo asintóticamente a medida que pasamos de una distribución discreta a una distribución continua. Observamos en el límite la relación con respecto al inverso de la función de densidad tal como habíamos remarcado anteriormente. La acotación superior de la función  $q(\cdot)$  realizada en el caso discreto para cualquier valor del parámetro varianza no tiene sentido en el caso continuo puesto que la

amplitud de  $q(\cdot)$  puede tender a infinito. En cambio, sí que resulta posible llegar a cotas inferiores. Análogamente, consideremos la expresión del denominador,

$$E_{a_i} e^{-|z_k - a_i|^2 / \sigma_i^2} < 1 \Rightarrow e^{-|z_k - a_i|^2 / \sigma_i^2} \leq q(z_k, a_k) \leq +\infty \tag{3.5.2.16}$$

En el siguiente apartado pasamos a estudiar el comportamiento de la función de regeneración.

### 3.5.3 Función de regeneración

A continuación nos ocuparemos del estudio del comportamiento de la función de regeneración de referencia  $\hat{a}(z_k)$  expresada en la ecuación 3.5.2.1. A modo de ejemplo consideraremos también una modulación 4-ASK de símbolos no equiprobables para ganar intuición en cuanto al comportamiento de la regeneración para funciones de densidad continuas. Consideraremos primeramente una distribución equiprobable de símbolos para valores extremos del parámetro varianza.

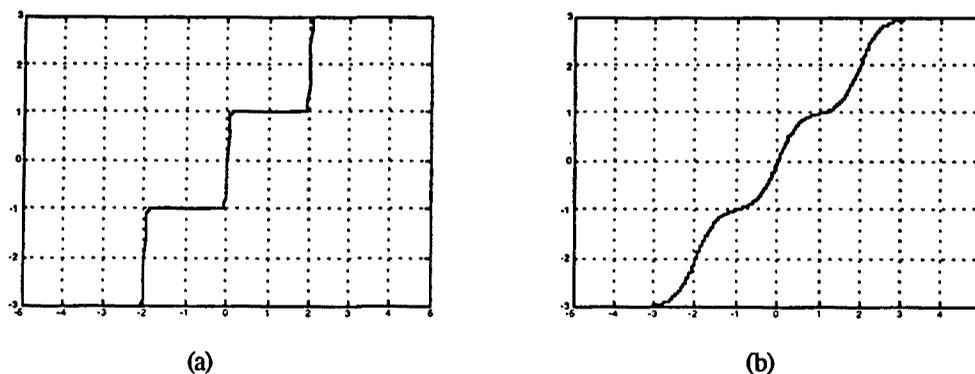


Figura 3.5.3.1: comparación de las funciones  $\hat{a}$  de regeneración de referencia para el caso particular de una modulación 4-ASK de símbolos equiprobables. Se ha considerado el alfabeto  $\{\pm 1, \pm 3\}$ . Se han considerado dos valores diferentes para el parámetro varianza: (a) 0.2 y (b) 0.7.

Observamos que la función de regeneración se comporta como una función de cuantificación para valores pequeños del parámetro varianza. A medida que éste aumenta, tendemos a un comportamiento más suave, como en la figura 3.5.2.5.a debido a la mayor incertidumbre que asignamos a los datos. En la siguiente figura consideramos la evolución de la función de regeneración respecto a  $\sigma_t$ .

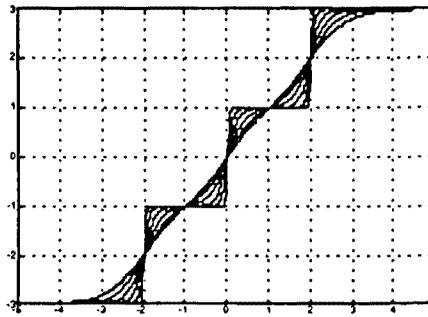


Figura 3.5.3.2: representación de la función de regeneración para valores crecientes del parámetro varianza,  $\{\sigma_t: 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$ .

Podemos observar que no todos los símbolos se ven afectados de la misma forma. La amplitud de los símbolos externos  $\{\pm 3\}$  se ve reducida a medida que aumenta  $\sigma_t$  de una forma apreciablemente mayor que los símbolos internos  $\{\pm 1\}$ . De esta forma justificamos experimentalmente el efecto de contracción de la constelación. Este efecto de distorsión resulta apreciable sobretodo cuando los símbolos no ocurren con equiprobabilidad. Consideraremos a continuación una modulación 4-ASK con matriz de probabilidad de símbolos  $\{1/13, 1/13, 10/13, 1/13\}$  y veremos qué efecto tiene sobre la función de regeneración. Obsérvense en la siguiente figura la representación de dicha función para valores extremos del parámetro varianza.

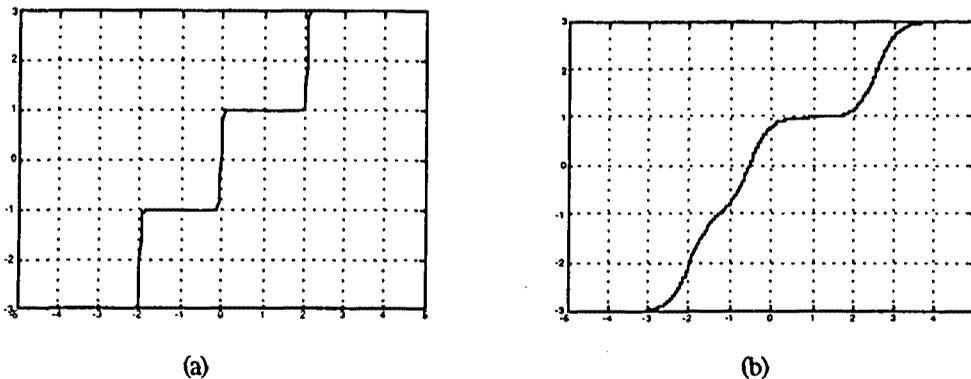


Figura 3.5.3.3: comparación de las funciones  $\hat{a}$  de regeneración de referencia para el caso particular de una modulación 4-ASK de símbolos no-equiprobables. La matriz de probabilidad correspondiente es  $\{1/13, 1/13, 10/13, 1/13\}$ . Se han considerado dos valores diferentes para el parámetro varianza: (a) 0.2 y (b) 0.7.

Obsevamos que para valores pequeños de  $\sigma_t$ , la función de regeneración es básicamente la misma que en el caso de equiprobabilidad de símbolo. No obstante, a medida que aumenta  $\sigma_t$ , aparece un comportamiento diferente de  $\hat{a}(z_k)$  en aquellos intervalos correspondientes a

probabilidades de símbolo diferentes. Vemos cómo el intervalo correspondiente al símbolo de mayor probabilidad presenta una transferencia mucho más plana que la del resto de símbolos, al mismo tiempo que su amplitud de transferencia apenas se ve afectada. En cambio, la amplitud de la función de regeneración correspondiente a símbolos menos probables presenta una contracción. Vemos pues que la distorsión introducida por el regenerador de referencia aumenta con la carácter no equiprobable de la distribución. En la siguiente figura consideramos la evolución de  $\hat{a}(z_k)$  para varios valores de  $\sigma_t$ .

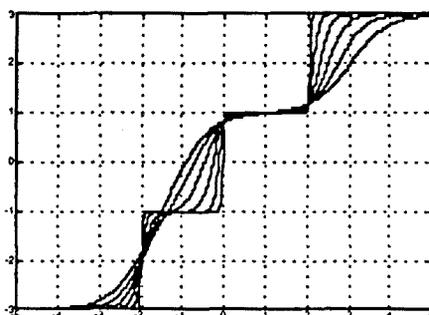


Figura 3.5.3.4: representación de la función de regeneración para valores crecientes del parámetro varianza,  $\{\sigma_t: 0.2, 0.35, 0.5, 0.65, 0.8, 0.95\}$ .

Nótese cómo el intervalo correspondiente al símbolo de mayor probabilidad presenta una variación mucho menor que el resto de los símbolos frente a la variación de  $\sigma_t$ .

Consideraremos a continuación el límite asintótico de la función de regeneración cuando  $\sigma_t \rightarrow 0$ . Recurramos a la ecuación 3.5.2.1,

$$\begin{aligned} \lim_{\sigma_t \rightarrow 0} \hat{a}(z_k) &= \lim_{\sigma_t \rightarrow 0} E_{a'_k} a'_k q(z_k, a'_k) = \lim_{\sigma_t \rightarrow 0} \int_C p_{a_k}(a'_k) a'_k q(z_k, a'_k) da'_k \\ &= \int_C p_{a_k}(a'_k) a'_k p_{a'_k}^{-1}(z_k) \delta(z_k - a_k) da'_k = z_k \end{aligned} \tag{3.5.3.1}$$

Es decir, cuando parametrizamos la función de coste con valores muy pequeños del parámetro varianza, la función de regeneración tiende a comportarse como la función identidad (para funciones de densidad de probabilidad continuas) debido a que consideramos los datos muy fiables. En el caso límite tenemos pues que el error generalizado tiende a cero para cualquier valor de la variable de salida  $z_k$ . Es aconsejable inicializar el algoritmo adaptativo con un valor de  $\sigma_t$  suficientemente grande para que permita la adaptación generando un error no nulo. En el caso de relaciones señal a ruido altas, el algoritmo ya ajustará por sí sólo el parámetro  $\sigma_t$  a un valor pequeño apropiado. Para el caso

de modulaciones discretas, podemos ver como la igualdad 3.5.3.1 se cumple sólo para aquellos valores correspondientes al alfabeto aunque  $\hat{a}(z_k)$  sea una función continua. Trás inspección de la figura 3.5.3.4 vemos que el error generalizado (la diferencia entre  $z_k$  i  $\hat{a}(z_k)$ ) es tanto mayor cuanto menos probable es el valor  $a_k$  correspondiente de la distribución. Valores muy verosímiles correspondientes a picos de la función de densidad generan error generalizados instantaneos muy pequeños.

Podemos ver en todas las figuras de la función de regeneración hasta ahora representadas, que ésta parece ser una función monótona de  $z_k$ . Para la demostración supongamos la derivada de  $\hat{a}(z_k)$  con respecto a  $z_k$ ,

$$\frac{\partial}{\partial z_k} \hat{a}_k = E_{a'_k} a'_k \frac{\partial}{\partial z_k} q(z_k, a'_k) \tag{3.5.3.2}$$

quedando expresada en términos de la derivada de  $q(\cdot)$ . Sin embargo, el cálculo de esta derivada se realiza en sección 4.4, por lo cual sustituimos aquí el resultado,

$$\frac{\partial}{\partial z_k} \hat{a}_k = E_{a'_k} a'_k (a'_k - \hat{a}_k) \frac{\partial}{\partial z_k} q(z_k, a'_k) = E_{a'_k} |a'_k|^2 q(z_k, a'_k) - |\hat{a}_k|^2 \tag{3.5.3.3}$$

El término de la derecha de la ecuación anterior aparece en otros contextos en este trabajo (véase el siguiente apartado) donde se demuestra su positividad. En la sección 4.4 se muestra su relación con una potencia de error y se denomina  $\sigma_d^2(z_k)$  o  $-\gamma(z_k)$ . Queda por tanto demostrada la monotonicidad de la función de regeneración,

$$\frac{\partial}{\partial y_k} \hat{a}_k = \sigma_d^2(z_k) \geq 0 \tag{3.5.3.4}$$

La explicación física de esta propiedad descansa en el hecho de que la función de regeneración intenta estimar unos datos perturbados por ruido aditivo. En el caso de que la distorsión que experimenten las muestras que deseamos regenerar no sean aditivas sino no-lineales (con ambigüedad en la inversión de la transformación no-lineal) podemos afirmar que la propiedad anterior no se cumplirá.

### 3.5.3.1 Contracción de la constelación

En este apartado consideraremos la contracción de la constelación para señales de amplitud constante. El hecho de recurrir a una formulación de producto escalar nos permite utilizar todas las propiedades de éste. Consideremos por tanto la desigualdad de Cauchy-Schwartz aplicada a la ecuación 3.5.2.1,

$$|\hat{a}_k|^2 = \left| E_{a'_k} a'_k \sqrt{q(z_k, a'_k)} \sqrt{q(z_k, a'_k)} \right|^2 \leq E_{a'_k} |a'_k|^2 q(z_k, a'_k) E_{a'_k} q(z_k, a'_k) = E_{a'_k} |a'_k|^2 q(z_k, a'_k) \quad (3.5.3.1.1)$$

La presente desigualdad se deriva también en la sección 4.4 por una vía alternativa. Vemos pues que la energía instantánea de la referencia es menor que la energía de la señal promediada a través de las funciones  $q(\cdot)$ . Si consideramos el caso particular de una distribución de señal de amplitud constante e igual a  $A$ , se cumple que,

$$|\hat{a}_k|^2 \leq A^2 \quad (3.5.2.1.4)$$

con desigualdad estricta, siempre que el parámetro varianza sea no nulo. Hemos demostrado pues la contracción de la constelación para señales de amplitud constante (Ej. QPSK, BPSK, CPM) independientemente de la arquitectura.

### 3.5.4 Estimación de la Varianza del Ruido en el Modelo

En los desarrollos que hemos seguido hasta ahora, hemos considerado únicamente optimización con respecto al vector de coeficientes. La función de coste es también función del parámetro varianza  $\sigma_t$ , tal como se expresa en la ecuación 3.4.1.1.5. El propósito del criterio de Referencia Estadística consiste en forzar una determinada distribución a la salida de un sistema. En una situación real desconocemos la potencia del ruido de filtrado (la cual es función por una parte de la potencia de ruido a la entrada y por otra del vector de coeficientes). En principio desconocemos qué  $\sigma_t$  hemos de utilizar, El propósito de esta sección consiste en deducir una relación para la estimación de la varianza de ruido en el modelo. Es decir, en el proceso de minimización, realizaremos una estimación conjunta del vector de coeficientes  $w_k$  y del parámetro  $\sigma_t$ . El gradiente respecto a los coeficientes no varía según este nuevo enfoque. Debemos únicamente calcular el gradiente con respecto al parámetro varianza. Recurriendo a la ecuación 3.4.1.1.6, se evalúa el gradiente,

(3.5.4.1)

$$\nabla_{\sigma_i^2} J = \frac{\partial}{\partial \sigma_i^2} \left( \ln \pi \sigma_i^2 + E_{r_k} \frac{|e_J|^2}{\sigma_i^2} \right)$$

Expandiendo la norma del error,  $|e_J|^2$ , se obtiene,

(3.5.4.2)

$$\nabla_{\sigma_i^2} J = \frac{1}{\sigma_i^2} - \frac{\partial}{\partial \sigma_i^2} E_{r_k} \ln E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} = \frac{1}{\sigma_i^2} - E_{r_k} \frac{E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2} |z_k - a'_k|^2 \frac{1}{\sigma_i^4}}{E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2}}$$

donde agrupando términos podemos expresar la anterior igualdad en términos de las funciones  $q(\cdot)$ ,

(3.5.4.3)

$$\nabla_{\sigma_i^2} J = \frac{1}{\sigma_i^2} \left( 1 - \frac{1}{\sigma_i^2} E_{r_k} E_{a'_k} \frac{e^{-|z_k - a'_k|^2 / \sigma_i^2}}{E_{a'_k} e^{-|z_k - a'_k|^2 / \sigma_i^2}} |z_k - a'_k|^2 \right) = \frac{1}{\sigma_i^2} \left( 1 - \frac{1}{\sigma_i^2} E_{r_k} E_{a'_k} q(z_k, a'_k) |z_k - a'_k|^2 \right)$$

Igualando el gradiente a cero obtenemos la estimación del parámetro varianza óptimo,

(3.5.4.4)

$$\sigma_i^2 = E_{r_k} E_{a'_k} q(z_k, a'_k) |z_k - a'_k|^2$$

Intuitivamente podemos ver que la varianza  $\sigma_i$  se estima teniendo en cuenta la norma del error definido como la muestra a la salida del filtro,  $z_k$ , respecto a la variable  $a'_k$ . Dicha norma se promedia sobre todos los valores posibles de  $a'_k$ , ponderada por las funciones  $q(\cdot)$  que cuantifican la probabilidad de que  $z_k$  proviniese de un valor determinado de  $a'_k$ . Finalmente, se toma el promedio estadístico con respecto a la variable  $z_k$ , o en su lugar, con respecto a la variable de entrada  $r_k$ .

### 3.6 Ejemplos de Aplicación

En este apartado consideraremos un par de ejemplos que ilustran el comportamiento de la función de coste para dos tipos de estadística de la señal deseada y veremos como se relacionan con dos criterios de optimización conocidos en el margen de relaciones señal-ruido elevadas y suponiendo que el sistema adaptativo ha convergido: supondremos el caso de una señal de envolvente constante y veremos que el criterio se reduce a un algoritmo alternativo al conocido algoritmo CMA (Constant Modulus Algorithm). Veremos que este caso admite una fácil generalización al caso de señales multi-amplitud, en las que se pueden englobar un amplio espectro de modulaciones. El segundo caso consiste en la recuperación de una señal de estadística QPSK y veremos qué relación existe con el algoritmo DD (Decision Directed).

#### 3.6.1 Señal de envolvente constante

Hemos elegido este caso como validación del algoritmo para funciones de densidad continuas. Definimos la función de densidad compleja de una señal de envolvente constante como aquella donde el módulo de la variable aleatoria es fijo y su fase sigue una distribución que supondremos uniforme (entendemos que el algoritmo aquí derivado es también aplicable a aquellas distribuciones que aunque estén caracterizadas por amplitud constante, su distribución de fase no sea uniforme en toda la corona. Aunque pueden obtenerse mejores comportamientos parametrizando exactamente la función de densidad, el proceder así conlleva ciertas ventajas computacionales. En general, puede aplicarse el algoritmo a señales con modulación discreta tipo QPSK, MSK,...). Expresamos la función de densidad en el formato real-imaginario y en el formato polar, respectivamente,

$$p_{a_k}(a_k) = \frac{1}{2\pi} \delta(a_k a_k^* - A^2) \quad , \quad p_{a_k}(r e^{j\theta}) = \frac{1}{2\pi} \delta(r - A) \tag{3.6.1.1}$$

A fin de hallar la función  $q(\cdot)$  correspondiente a este problema, necesitamos calcular la función de densidad correspondiente a la variable aleatoria señal más ruido Gausiano, la cual viene dada por la convolución de ambas densidades o según aplicación del Teorema Fundamental de la Esperanza como,

$$p_{z_k = a_k + n_k}(z_k) = E_{a_k} \frac{1}{\pi\sigma^2} e^{-|z_k - a_k|^2 / \sigma^2} = \int_0^\infty \int_{-\pi}^{+\pi} \frac{1}{2\pi} \delta(r - A) \frac{1}{\pi\sigma^2} e^{-|z_k - r e^{j\theta}|^2 / \sigma^2} dr d\theta$$

$$= \frac{1}{2\pi^2 \sigma^2} \int_{-\pi}^{+\pi} e^{-|z_k - A e^{j\theta}|^2 / \sigma^2} d\theta = \frac{1}{2\pi^2 \sigma^2} e^{-|z_k|^2 / \sigma^2} e^{-A^2 / \sigma^2} \int_{-\pi}^{+\pi} e^{2 \operatorname{Re}(z_k^* A e^{j\theta}) / \sigma^2} d\theta \tag{3.6.1.2}$$

Consideremos una representación polar de la señal  $z_k$ ,

$$z_k = \rho e^{j\alpha} \Rightarrow p_{z_k}(\rho e^{j\alpha}) = \frac{1}{2\pi^2\sigma^2} e^{-\rho^2/\sigma^2} e^{-A^2/\sigma^2} \int_{-\pi}^{+\pi} e^{2\rho A \cos(\theta-\alpha)/\sigma^2} d\theta \quad (3.6.1.3)$$

Definiendo la integral  $I_A(\rho, \alpha)$  podemos ver que ésta presenta simetría de revolución, es decir, depende solamente de la variable  $\rho$  debido a la naturaleza de la función de densidad,

$$I_A(\rho, \alpha) = \int_{-\pi}^{+\pi} e^{2\rho A \cos(\theta-\alpha)/\sigma^2} d\theta = \int_{-\pi}^{+\pi} e^{2\rho A \cos \theta/\sigma^2} d\theta = I_A(\rho) \quad (3.6.1.4)$$

La función  $I_A$  se representa fácilmente (ver apéndice) en términos de la función de Bessel modificada de primera especie y orden cero, con lo cual obtenemos una expresión final compacta para la función de densidad deseada,

$$I_A(\rho) = 2\pi I_0\left(\frac{2A}{\sigma^2}\rho\right) \Rightarrow p_{z_k}(\rho e^{j\alpha}) = \frac{1}{\pi\sigma^2} e^{-\rho^2/\sigma^2} e^{-A^2/\sigma^2} I_0\left(\frac{2A}{\sigma^2}\rho\right) \quad (3.6.1.5)$$

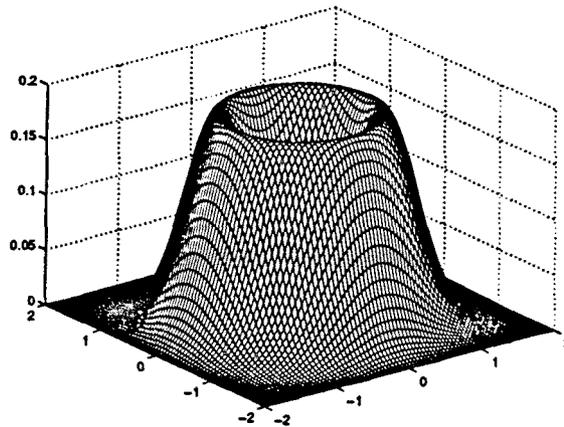


Figura 3.6.1.1: representación de la función de densidad de una señal de amplitud constante más ruido Gaussiano.

Una vez calculada la función de densidad, podemos expresar la función de coste de forma explícita como,

$$J = E_\rho \left\{ \frac{1}{\sigma^2} (\rho^2 - A^2) - \ln I_0\left(\frac{2A}{\sigma^2}\rho\right) \right\} \quad (3.6.1.6)$$

Vemos por una parte que la función de coste sólo se ve afectada por la estadística del módulo de la variable aleatoria a la salida del sistema adaptativo,  $z_k = \rho e^{j\alpha}$ . Observamos.

también la aparición de un término de error de módulo,  $\rho^2 - A^2$ , modificado por un logaritmo, que justifica el hecho de que la función de coste actúe como un ajuste no cuadrático de módulos.

En estas condiciones, las funciones  $q(\cdot)$  sobre las que se define la regeneración de referencia siguen la expresión,

$$q(z_k, r e^{j\theta}) = \frac{e^{-|z_k - r e^{j\theta}|^2 / \sigma^2}}{E_{r,\theta} e^{-|z_k - r e^{j\theta}|^2 / \sigma^2}} = \frac{e^{-\rho^2 / \sigma^2} e^{-r^2 / \sigma^2} e^{2\rho r \cos(\theta - \alpha) / \sigma^2}}{\frac{1}{2\pi} e^{-\rho^2 / \sigma^2} e^{-A^2 / \sigma^2} I_A(\rho)} \tag{3.6.1.7}$$

$$= \frac{1}{\frac{1}{2\pi} I_A(\rho)} e^{-(r^2 - A^2) / \sigma^2} e^{2\rho r \cos(\theta - \alpha) / \sigma^2}$$

En la figura 3.6.1.2 representamos cualitativamente la dependencia angular de la función  $q(\cdot)$  en el argumento  $\theta$ . Observamos que la proyección del vector  $(A, \theta)$  sobre el vector  $(\rho, \alpha)$  nos da el exponente de la dependencia angular de  $q(\cdot)$ , es decir, cuanto mayor sea la distancia entre ambos puntos, tanto menor es la probabilidad de que  $(\rho, \alpha)$  proviniese originariamente de  $(A, \theta)$  más un término de ruido Gaussiano.

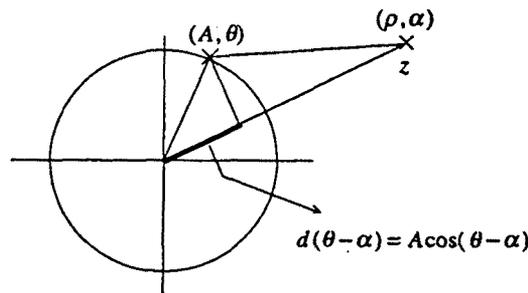


Figura 3.6.1.2: representación vectorial de la dependencia angular de  $q(\cdot)$  con respecto a los puntos de la corona circular de la distribución sin ruido:  $q(z_k, A e^{j\theta})$

La función de regeneración de referencia viene dada por,

$$\hat{a}_k = E_{r,\theta} r e^{j\theta} q(z_k, r e^{j\theta}) = \int_0^{\infty} \int_{-\pi}^{+\pi} \frac{1}{2\pi} \delta(r - A) r e^{j\theta} \frac{1}{\frac{1}{2\pi} I_A(\rho)} e^{-(r^2 - A^2) / \sigma^2} e^{2\rho r \cos(\theta - \alpha) / \sigma^2} dr d\theta \tag{3.6.1.8}$$

$$= A \int_{-\pi}^{+\pi} e^{j\theta} \frac{1}{I_A(\rho)} e^{2\rho A \cos(\theta - \alpha) / \sigma^2} d\theta = A e^{j\alpha} \int_{-\pi}^{+\pi} e^{j\theta} \frac{1}{I_A(\rho)} e^{2\rho A \cos \theta / \sigma^2} d\theta = A e^{j\alpha} \zeta_A(\rho)$$

donde observamos que la señal regenerada es simplemente una proyección sobre el círculo de radio  $A$ , pesada por un factor real que cuantifica la distancia de  $z_k$  a este círculo,

$$\tag{3.6.1.9}$$

$$\zeta_A(\rho) = \frac{1}{I_A(\rho)} \int_{-\pi}^{+\pi} \cos \theta e^{2\rho A \cos \theta / \sigma^2} d\theta = \frac{\sigma^2}{2A} \frac{d}{d\rho} \ln I_A(\rho)$$

El hecho de que la función de regeneración preserve el argumento de los datos es debido a la simetría de revolución de la función de distribución, ya que ningún ángulo se ve favorecido con respecto a otros. El error generalizado mantiene también el argumento de los datos, es decir, es simplemente un error de módulo.

La derivada de la función  $I_A$  se representa fácilmente también en términos de la función de Bessel modificada de primera especie,

$$I_A(\rho) = 2\pi I_0\left(\frac{2A}{\sigma^2}\rho\right) \quad , \quad \frac{d}{d\rho} I_A(\rho) = 2\pi \frac{d}{d\rho} I_0\left(\frac{2A}{\sigma^2}\rho\right) = \frac{4\pi A}{\sigma^2} I_1\left(\frac{2A}{\sigma^2}\rho\right) \quad (3.6.1.10)$$

Expresándose la función de regeneración finalmente como,

$$\hat{a}_k = A e^{j\alpha} \zeta_A(\rho) = \frac{I_1\left(\frac{2A}{\sigma^2}\rho\right)}{I_0\left(\frac{2A}{\sigma^2}\rho\right)} A e^{j\alpha} \quad (3.6.1.11)$$

En la figura 3.6.1.1 hemos representado el factor  $\zeta_A$  para diversos valores de la relación señal a ruido definida por  $K=2A/\sigma^2$ , particularizando para el círculo unidad. Obsérvese el efecto de contracción de la constelación para valores sucesivamente decrecientes del parámetro  $K$ . Para todo  $\rho$ ,  $\zeta_A(\rho)$  es siempre menor que uno y decrece en función de  $K$ . Vemos que cuando  $K$  tiende a infinito (o  $\sigma$  tiende a cero), la función de regeneración tiende a la saturación del módulo del dato  $z_k$ .

La expresión para la curva S del algoritmo o función de error generalizado viene dada según la siguiente ecuación,

$$\varepsilon_k = z_k - \hat{a}_k = \rho e^{j\alpha} - A e^{j\alpha} \zeta_A(\rho) = \left( \rho - A \frac{I_1\left(\frac{2A}{\sigma^2}\rho\right)}{I_0\left(\frac{2A}{\sigma^2}\rho\right)} \right) e^{j\alpha} \quad (3.6.1.12)$$

Vemos que el error conserva el argumento de los datos, por lo cual resulta suficiente representar el error de módulo. Obsérvese en la figura 3.6.1.1 la función de error generalizada. Podemos observar como para factores  $K$  crecientes, los cortes por cero se van desplazando hacia la izquierda hasta un valor límite donde ya no existe tal corte.

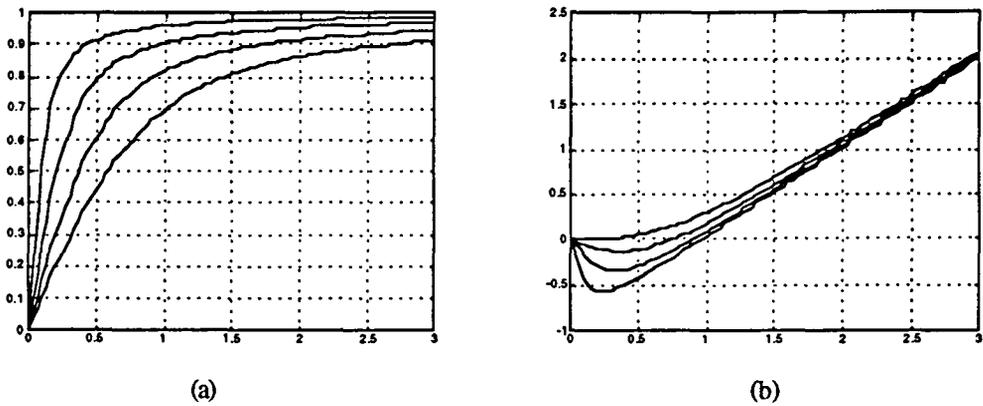


Figura 3.6.1.3: (a) representación de la función de regeneración para una corona circular de radio unidad y varianzas de ruido 0.4, 0.6, 0.8 y 1 (b) representación equivalente de la función de error generalizado.

Resulta ilustrativo comparar la curva S del algoritmo CMA clásico (que exponemos a continuación) con la función de error generalizado obtenida en este apartado,

$$\epsilon_k^{CMA} = \rho(\rho^2 - A^2)e^{j\alpha} \tag{3.6.1.13}$$

Teniendo en cuenta el comportamiento asintótico vemos que para valores del módulo muy alejados del radio A, la función de error del CMA tiende a una rama parabólica de orden  $\rho^3$  mientras que el error generalizado varía linealmente con el módulo  $\rho$ . Dicho comportamiento parece justificar la mayor robustez del nuevo algoritmo en las actualizaciones mediante el gradiente, sobre todo para amplitudes de la señal de entrada alejadas de la situación de equilibrio.

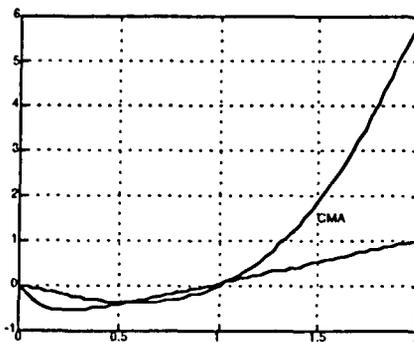


Figura 3.6.1.4: comportamiento entre las funciones de error generalizado o curvas S del algoritmo CMA clásico y el deducido en este apartado en función del módulo de los datos. Obsérvense los comportamientos claramente parabólico y lineal.

### 3.6.2 Señales multiamplitud

Definimos las señales multiamplitud a partir de su función de distribución como aquellas cuya amplitud toma valores en un conjunto discreto y su distribución de fase es continua (pueden utilizarse para modelar modulaciones del tipo CPM multiamplitud),

$$p_{a_k}(a_k) = \frac{1}{2\pi} \sum_{i=1}^N p_i \delta(a_k - A_i) \quad , \quad \sum_{i=1}^N p_i = 1 \tag{3.6.2.1}$$

donde el conjunto  $\{p_i\}$  representa las probabilidades de que la señal esté contenida en la corona circular  $i$ -ésima. Reproduciendo los pasos expuestos en el apartado anterior, construimos la función de densidad de la variable aleatoria multiamplitud más ruido Gaussiano como,

$$p_{z_k}(z_k) = \int_0^{+\infty} \int_{-\pi}^{+\pi} \frac{1}{2\pi} \sum_i p_i \delta(r - A_i) \frac{1}{\pi\sigma^2} e^{-|z_k - re^{j\theta}|^2 / \sigma^2} dr d\theta = \frac{1}{2\pi\sigma^2} e^{-|z_k|^2 / \sigma^2} \sum_i p_i e^{-A_i^2 / \sigma^2} \int_{-\pi}^{+\pi} e^{2\text{Re}\{z_k^* A_i e^{j\theta}\} / \sigma^2} d\theta \tag{3.6.2.2}$$

de donde hallamos la expresión final de las funciones  $q(\cdot)$  y de la función de regeneración,

$$q(z_k, re^{j\theta}) = \frac{e^{-r^2/\sigma^2} e^{2\rho r \cos(\theta - \alpha) / \sigma^2}}{\frac{1}{2\pi} \sum_i p_i e^{-A_i^2/\sigma^2} I_{A_i}(\rho)} \Rightarrow \hat{a}_k = \left( \sum_i \bar{p}_i A_i \right) e^{j\alpha} = \bar{A} e^{j\alpha} \tag{3.6.2.3}$$

donde las probabilidades barradas representan la probabilidad de que el dato recibido provenga de una corona circular determinada. La amplitud de los datos regenerados se obtiene entonces como el promedio de las amplitudes de cada corona con respecto a su probabilidad estimada. Dichas probabilidades vienen dadas por la siguiente ecuación,

$$\bar{p}_i = \frac{p_i e^{-A_i^2/\sigma^2} I_1\left(\frac{2A_i}{\sigma^2} \rho\right)}{\sum_{i'} p_{i'} e^{-A_{i'}^2/\sigma^2} I_0\left(\frac{2A_{i'}}{\sigma^2} \rho\right)} \tag{3.6.2.4}$$

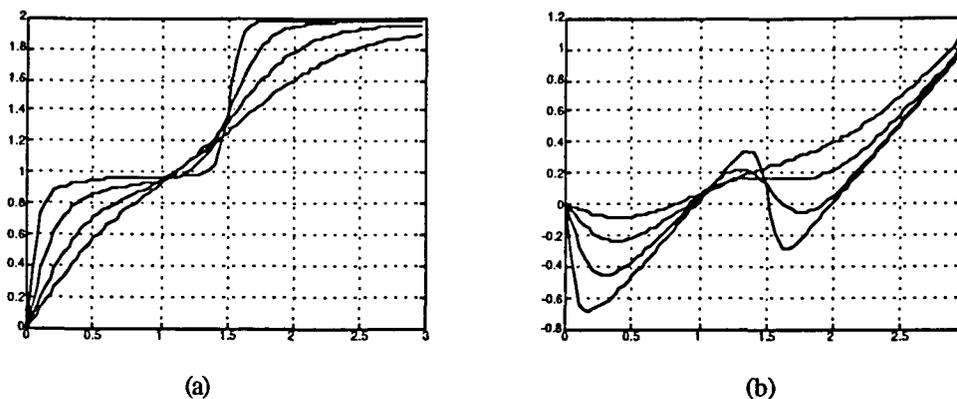


Figura 3.6.2.1: (a) representación de la función de regeneración para dos coronas circulares equiprobables de radio uno y dos y varianzas de ruido 0.3, 0.5, 0.7 y 0.9 (b) representación equivalente de la función de error generalizado.

Obsérvese en la figura 3.6.2.1b como para valores excesivamente elevados del parámetro varianza desaparece el cruce por cero de la curva S correspondiente a la corona de mayor amplitud. Este efecto es debido a que la convolución con la función de densidad Gaussiana ha enmascarado el pico menor de la función de densidad (si consideramos equiprobabilidad de las coronas, tendremos que aquella correspondiente a un radio mayor presenta una densidad de probabilidad menor debido a que abarca un mayor perímetro), apareciendo únicamente uno correspondiente al pico de mayor altura (menor radio) de la función de densidad.

### 3.6.3 Señal QPSK

Hemos elegido este caso particular por tratarse de un ejemplo de señal discreta correspondiente a una modulación lineal. Definimos la señal QPSK como aquella que toma los valores  $\{e^{j\pi(1+2k)/4}, k=0,1,2,3\}$  con igual probabilidad. Su función de densidad viene dada por,

$$p_{a_k}(a_k) = \frac{1}{4} \sum_{i=0}^{i=3} \delta(a_k - e^{j\pi(1+2i)/4}) \tag{3.6.3.1}$$

La función de densidad correspondiente a la variable aleatoria señal más ruido Gaussiano viene dada por la convolución de ambas densidades o según la notación de esperanzas como,

$$p_{z_k = a_k + n_k}(z_k) = E_{a_k} \frac{1}{\pi\sigma^2} e^{-|z_k - a_k|^2 / \sigma^2} = \frac{1}{4\pi\sigma^2} \sum_{i=0}^{i=3} e^{-|z_k - e^{j\pi(1+2i)/4}|^2 / \sigma^2} \tag{3.6.3.2}$$

Las funciones  $q(\cdot)$  se calculan a partir de la ecuación 3.6.2 conduciendo a,

$$q(z_k, a_k) = \frac{e^{-|z_k - e^{j\pi(1+2i)/4}|^2 / \sigma^2}}{\sum_{i=0}^{i=3} \frac{1}{4} e^{-|z_k - e^{j\pi(1+2i)/4}|^2 / \sigma^2}} \quad (3.6.3.3)$$

En el margen de relaciones señal-ruido altas evaluamos entonces el comportamiento de la función de coste y de la función  $q(\cdot)$  cuando la varianza del ruido tiende a cero manteniendo constante la amplitud de la señal. En ese caso, tendremos que la muestra recibida se encuentra con gran probabilidad muy cercana a alguno de los símbolos de la constelación. Dado que la función de densidad viene dada por la suma de los errores exponenciados con respecto a cada símbolo de la constelación, sólo aquella exponencial correspondiente al símbolo más cercano tendrá un valor significativo. Denotemos el símbolo más cercano como  $a_k^+$ . Entonces, la función de coste tiende a,

$$J \rightarrow -\ln 4\pi\sigma^2 - \frac{1}{\sigma^2} E_{a_k} |z_k - a_k^+(a_k)|^2, \quad a_k^+ = \arg \min_{a_k \in C} \{|z_k - a_k|^2\} \quad (3.6.4)$$

Vemos pues que la función de coste tiende a minimizar el error cuadrático entre la muestra recibida y el símbolo más cercano. El valor estimado del símbolo emitido por la fuente se calcula también según la ecuación 3.6.4 como el símbolo más cercano,

$$\hat{a}_k = E_{a_k} q(z_k, a_k) a_k \rightarrow E_{a_k} \delta(z_k - a_k^+) a_k = a_k^+ \quad (3.6.5)$$

### 3.7 Resumen del Algoritmo de Referencia Estadística Primera Función de Coste

En la sección 3.5 hemos planteado la estimación del gradiente del algoritmo de referencia estadística basado en la primera función de coste. Hemos considerado por una parte el gradiente de la función de coste con respecto a los coeficientes del filtro  $w_k$  y con respecto al parámetro varianza  $\sigma_i$ . A modo de resumen presentamos el gráfico de la figura 3.7.1 donde aparece el procedimiento completo.

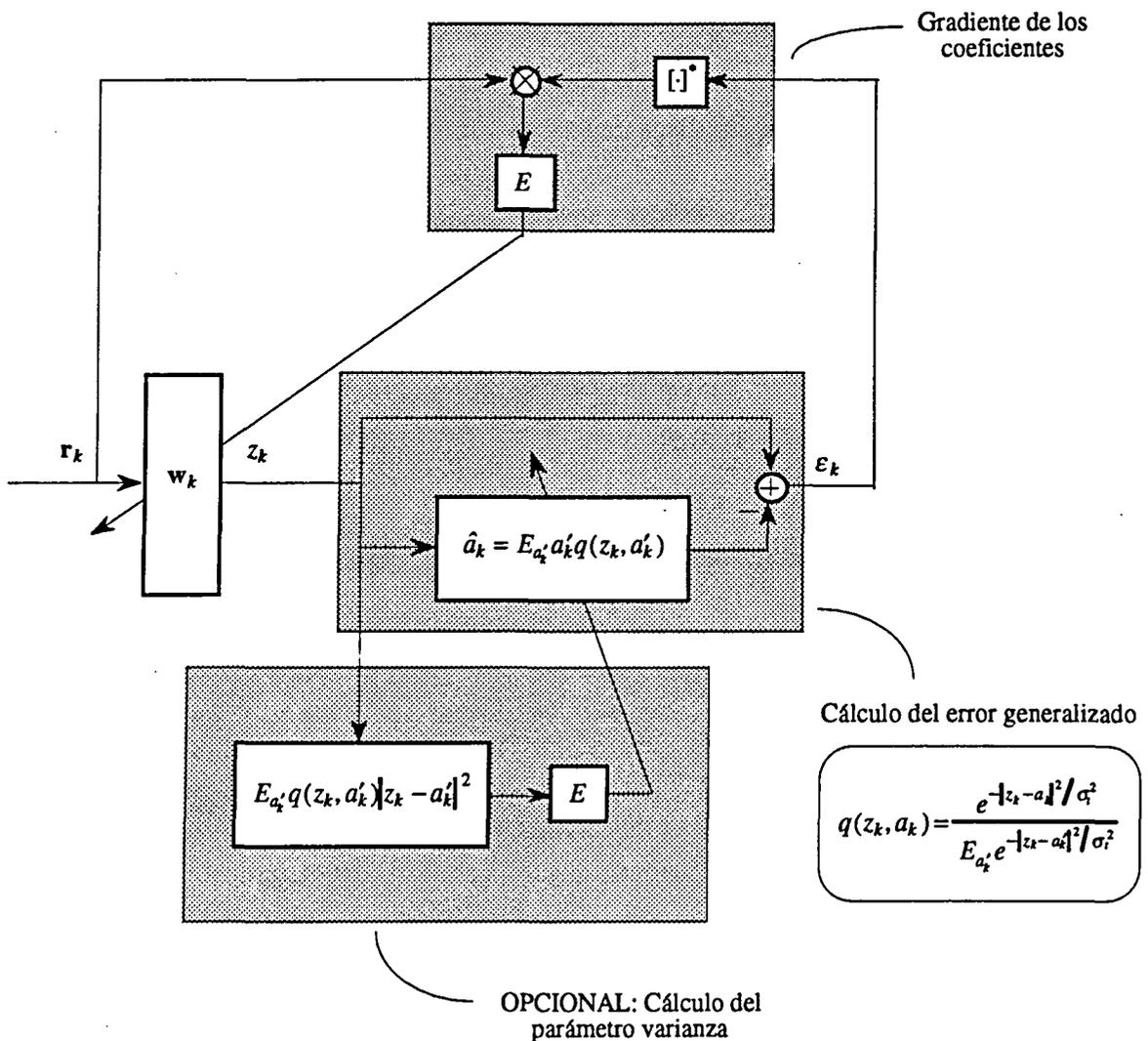


Figura 3.7.1: Algoritmo de Referencia Estadística . Primera Función de Coste.

### 3.8 Conclusiones

En este capítulo hemos introducido los conceptos generales de Referencia Estadística mediante el estudio de una primera función de coste. Hemos visto que a partir de un planteamiento de ajuste de densidades de probabilidad como la discriminación de Kullback-Leibler, podemos llegar mediante ciertas simplificaciones a un algoritmo de adaptación ciega del tipo expuesto en la figura 3.7.1. La función de coste resultante puede expresarse en términos de una norma no-lineal de error que introduce información sobre la densidad de probabilidad deseada, tal como se ha demostrado en la sección sobre la función de coste condicionada. A partir del estudio del gradiente aparece de forma natural un concepto capital en los algoritmos de referencia estadística, el de la función de regeneración de la referencia. Este concepto establece una cierta relación con los algoritmos de referencia temporal. Al no disponer de una realización de los datos transmitidos, el algoritmo de Referencia Estadística intenta regenerarlos de la forma más inteligente posible a fin de generar una señal de error, la cual hemos llamado error generalizado en este capítulo. Podemos ver cierta relación en este sentido con los algoritmos de Busgang para ecualización ciega. Veremos más adelante, en el capítulo 4, como también puede definirse una función análoga para la segunda función de coste. Una propiedad básica de la función de regeneración de la primera función de coste es que ésta puede calcularse previamente en función de la densidad de probabilidad deseada y no depende de la estadística de la señal a la salida del sistema adaptativo. Veremos como esta propiedad no se cumple en el caso de la segunda función de coste, obligando a la actualización continua de la función de regeneración.

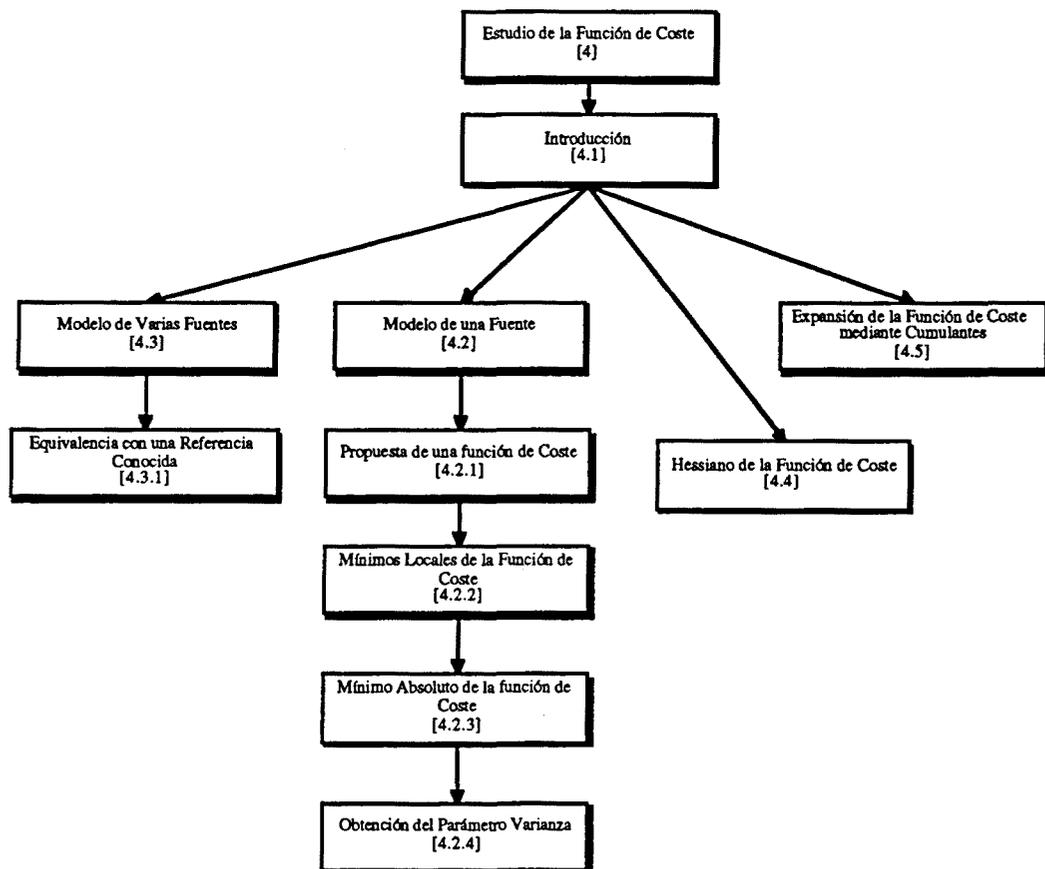
Aunque en la mayoría de los casos que nos interesan en el campo de las comunicaciones digitales, tratamos de recuperar una señal deseada proveniente de una distribución discreta, es decir, la correspondiente a los símbolos de una constelación, hemos realizado el desarrollo para el caso general de cualquier distribución. En la sección titulada ejemplos de aplicación hemos desarrollado un par de casos, basados respectivamente en el ajuste a una distribución de módulo constante (variable aleatoria con densidad de probabilidad continua) y a una constelación QPSK (variable aleatoria con densidad de probabilidad discreta).

Básicamente, hemos dedicado este capítulo al planteamiento de la primera función de coste y al estudio de su gradiente. En el siguiente capítulo realizaremos un estudio de la estructura de las soluciones de esta función de coste, donde veremos la relación con las soluciones de mínima varianza y de Wiener en términos de la función de regeneración.

## Capítulo IV

### Estudio de la Función de Coste

---





## 4.1 Introducción

En el capítulo anterior hemos identificado los conceptos básicos de un algoritmo de referencia estadística basado en la primera función de coste. El objetivo de este capítulo es presentar resultados teóricos a un nivel más fundamental. En esta introducción resumiremos los resultados más importantes a fin de que, en una primera lectura, el lector pueda evitar el aparato matemático que se desarrolla sin perder información de su contenido. Aconsejamos que en el caso de querer seguir las demostraciones con detalle, se deje para una segunda lectura.

Hemos estructurado este capítulo en cuatro secciones. dejando aparte esta introducción:

### 4.1.1 Modelo de una Fuente

Estudiamos el comportamiento de la función de coste según el modelo de señal de la sección 3.3.1. Se generaliza la función de coste a la esperanza de una no-linealidad arbitraria de la siguiente forma,

$$J = -E_{z_k} \ln p_{\alpha_k + \eta_k}(z_k) \Rightarrow J = -E_{z_k} \psi(z_k) \quad (4.1.1.1)$$

El enfoque básico de esta sección se basa en el cálculo de los mínimos locales y del mínimo absoluto de la función de coste definida en la ecuación anterior. En concreto, se llega a los siguientes resultados:

- Todos los mínimos locales de  $J(\mathbf{w}_k)$  resultan ser proporcionales a la solución de mínima varianza:

$$\mathbf{w}_k = \alpha_k \mathbf{R}_{nn}^{-1} \mathbf{s}_k \quad (4.1.1.2)$$

donde  $\mathbf{R}_{nn}$  representa la matriz de correlación del ruido  $\mathbf{n}_k$  del modelo de señal,  $\mathbf{s}_k$  es el vector de steering de la única fuente presente y  $\alpha_k$  es el término de escalado. Es decir, todos los mínimos locales se hallan en la misma recta.

- Para el caso particular de que la no-linealidad de la ecuación 4.1.1.1 sea la función de log-likelihood, el mínimo absoluto de la función de coste se atañe para un valor del escalado menor del que conseguiríamos según el criterio de mínima varianza:

$$\alpha_k \leq \frac{1}{\mathbf{s}_k^H \mathbf{R}_{nn}^{-1} \mathbf{s}_k} \quad (4.1.1.3)$$

hecho que se traduce en algoritmos adaptativos en la contracción de la constelación a la salida del ecualizador.

#### 4.1.2 Modelo de Varias Fuentes.

Estudiamos el comportamiento de la función de coste según el modelo de señal de la sección 3.3.1. Generalizamos la función de coste de la misma forma que en la ecuación 4.1.1.1. El enfoque de esta sección se basa por una parte en el análisis de los mínimos locales, y por otra, en cómo relacionar el vector de coeficientes óptimo según el criterio 4.1.1.1 con la función de regeneración de la referencia. Los resultados que se obtienen son:

- Todos los mínimos locales de  $J(\mathbf{w}_k)$  resultan ser generados por la siguiente ecuación análoga a 4.1.1.2:

$$\mathbf{w}_k = \mathbf{R}_{nn}^{-1} \mathbf{S}_k \alpha_k \quad (4.1.2.1)$$

donde  $\mathbf{R}_{nn}$  representa la matriz de correlación del ruido  $\mathbf{n}_k$  del modelo de señal,  $\mathbf{S}_k$  es el matriz de steering que contiene los vectores de steering de todas las fuentes y  $\alpha_k$  es el término de escalado. Es decir, todos los mínimos locales se hallan contenidos en un determinado subespacio de dimensión igual al número de fuentes.

- Consideremos la siguiente transformación de los datos  $\mathbf{r}_k$  a  $\mathbf{u}_k$ :

$$\mathbf{u}_k = (\mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{S}_k)^{-1} \mathbf{S}_k^H \mathbf{R}_{nn}^{-1} \mathbf{r}_k \quad (4.1.2.2)$$

tenemos entonces que la salida del filtro,  $z_k$ , puede expresarse según la ecuación  $z_k = \xi_k^H \mathbf{u}_k$ , donde los coeficientes  $\xi_k$  que minimizan la función de coste se obtienen mediante la siguiente ecuación,

$$\xi_k = \mathbf{R}_{uu}^{-1} E \hat{a}_k^* \mathbf{u}_k \Leftrightarrow \xi_{Wiener} = \mathbf{R}_{uu}^{-1} E d_k^* \mathbf{u}_k \quad (4.1.2.3)$$

en función de la correlación cruzada de la referencia regenerada,  $\hat{a}_k$ , con el vector de datos  $\mathbf{u}_k$ . Obsérvese el paralelismo con el criterio de Wiener en el que disponemos de la referencia real de la secuencia transmitida,  $d_k$ .

- El vector de coeficientes óptimo,  $\mathbf{w}_k$ , puede expresarse también de forma análoga a la solución de Wiener, mediante la matriz de correlación de los datos con

$$\mathbf{w}_k = \mathbf{R}_{rr}^{-1} E_{r_k} \hat{a}_k^* \mathbf{r}_k \tag{4.1.2.4}$$

Se demuestra en el capítulo que esta forma es coherente con las anteriores.

### 4.1.3 Hessiano de la función de coste

En esta sección calculamos el Hessiano de la función de coste para determinar bajo qué condiciones se producirá un mínimo local. Se considera únicamente el caso en que la no-linealidad coincide con la función de log-likelihood. De forma paralela se deducen relaciones importantes entre las diferentes varianzas que caracterizan los mínimos de la función de coste.

- El Hessiano de la función de coste puede expresar mediante la siguiente ecuación,

$$\mathbf{H}_{\mathbf{w}_k} J = E_{\mathbf{a}_k, n_k} \left( 1 - \frac{\sigma_d^2(z_k)}{\sigma_r^2} \right) \mathbf{r}_k \mathbf{r}_k^H \tag{4.1.3.1}$$

en función de la matrix de correlación de los datos, del parámetro varianza de la función de coste,  $\sigma_r$ , y de un nuevo parámetro  $\sigma_d$ . La condición de mínimo local que se deduce es que el Hessiano debe ser de definición positiva menor que la matriz de correlación de los datos.

- Se deduce también la siguiente ecuación de ligadura entre diferentes términos de potencia,

$$\sigma_r^2 + E_{z_k} \sigma_d^2(z_k) + E_{z_k} |z_k|^2 = E_{z_k} |\hat{a}_k|^2 \tag{4.1.3.2}$$

donde se relaciona el parámetro varianza,  $\sigma_r$ , con el parámetro  $\sigma_d$ , la potencia de  $z_k$  y la potencia de la referencia regenerada,  $\hat{a}_k$ .

#### 4.1.4 Expansión de la Función de Coste mediante Cumulantes

Partiendo de una relación entre la función de coste y una función característica de una variable aleatoria, se llega a una expansión mediante cumulantes de la función de coste.

- Definiendo la función característica de una variable aleatoria  $u$  según la siguiente ecuación,

$$\Phi_u(s) = E_u e^{su} \tag{4.1.4.1}$$

tenemos la siguiente expresión de la función de coste en términos de la función característica de una variable aleatoria  $u$  igual a la norma  $|z_k - a_k|^2$ ,

$$J = -E_{z_k} \ln p_{a_k + \eta_k}(z_k) = -E_{z_k} \ln \frac{1}{\pi \sigma_k^2} \Phi_{u(a_k) = |z_k - a_k|^2} \left( \frac{-z_k}{\sigma_k^2} \right) \tag{4.1.4.2}$$

De esta forma resulta posible relacionar la función de coste en términos de los cumulantes de la variable aleatoria  $|z_k - a_k|^2$ , puesto que estos son los coeficientes del desarrollo en serie de potencias del logaritmo natural de la función característica.

## 4.2 Modelo de una Fuente. Introducción

Un análisis general del comportamiento de la función de coste no resulta posible ya que su comportamiento depende en gran medida de la estadística de la señal de entrada. Por lo tanto hemos elegido estructurar su estudio en unos pocos casos representativos. En esta sección trataremos un simple problema de aplicación de la función de coste ya propuesta: la recuperación de un proceso aleatorio de distribución desconocida (en términos de su función de densidad de probabilidad) que ha experimentado una transformación lineal desconocida y que se encuentra inmerso en ruido Gaussiano de correlación también desconocida (ver sección 3.3.1). En secciones posteriores refinaremos el modelo de señal según la sección 3.3.2 y veremos qué influencia tiene la aparición de varias fuentes en el vector de coeficientes solución del criterio. A diferencia del capítulo anterior, plantearemos la función de coste no en términos de la esperanza de la función log-likelihood sino que consideraremos una no-linealidad arbitraria. Veremos que muchos de los resultados que obtenemos son aplicables sin necesidad de la premisa de la función de likelihood.

En esencia la función de coste es multimodal por lo cual realizaremos un estudio detallado de los mínimos locales y el mínimo absoluto. El modelo de señal que utilizaremos en esta sección aparece expuesto en la ecuación 3.3.1.1. Demostraremos para el modelo estadístico de la ecuación 3.3.1.1 que podemos recuperar la señal de distribución deseada excepto en un factor de escala, y que el vector de coeficientes es proporcional al obtenido mediante el criterio de mínima varianza o de referencia espacial. Definiendo la función de coste como la esperanza de una no-linealidad evaluada en la variable de salida,  $y_k$ , veremos que existe una amplia clase de no-linealidades que pueden forzar una distribución dada a la salida del sistema adaptativo. Deduciremos la función de densidad óptima que minimiza el criterio y demostraremos que es una versión polarizada de la función de densidad deseada, cuya polarización o bias depende de la potencia de ruido a la entrada.

### 4.2.1 Propuesta de una función de coste

Según hemos visto en secciones anteriores, la función de coste propuesta consiste en la minimización de la esperanza de la función log-likelihood evaluada a la salida del filtro adaptativo. En este apartado consideraremos el caso más general de una no-linealidad en lugar de la función log-likelihood. Veremos que el criterio también presenta mínimos

deseables aún cuando no parametrizemos exactamente la función de densidad de probabilidad a la salida del filtro adaptativo.

El conjunto óptimo de coeficientes que minimiza la función de coste  $J$  se obtiene como sigue,

$$\mathbf{w}_k = \arg \min \{ J = - E_{y_k} \psi(y_k) \} \quad (4.2.1.1)$$

En un caso más general, consideraremos que la no-linealidad es parametrizable, de forma que el proceso de minimización se extiende también a los parámetros,

$$\{ \mathbf{w}_k, \phi_k \} = \arg \min \{ J = - E_{y_k} \psi_\phi(y_k) \} \quad (4.2.1.2)$$

Esta parametrización resulta útil para conseguir mejores prestaciones del sistema adaptativo en ciertos casos, como cuando se utiliza la función likelihood. En general, para cualquier no-linealidad, no podemos afirmar que 4.2.1.2 sea mejor que 4.2.1.1. Podemos entender 4.2.1.2 como una adaptación de la no-linealidad a la estadística de la señal de entrada. Antes de proceder conviene expresar la función de coste en un formato más conveniente para los propósitos de minimización. Si aplicamos ahora el gradiente con respecto a los coeficientes, la ecuación 4.2.1.1 sugiere que el gradiente no puede intercambiarse con el operador esperanza ya que éste se define sobre la variable aleatoria a la salida del filtro, cuya estadística depende a su vez de los coeficientes. Por lo tanto, siguiendo el procedimiento expuesto en la ecuación 3.4.3,

$$J = - E_{y_k} \psi_\phi(y_k) = - E_{a_k, \mathbf{n}_k} \psi_\phi(y_k(a_k, \mathbf{n}_k)) = - E_{a_k, \mathbf{n}_k} \psi_\phi(y_k) \quad (4.2.1.3)$$

donde vemos que el gradiente y la esperanza sí pueden intercambiarse según esta nueva formulación. Nótese que al igualar el gradiente a cero obtendremos los mínimos locales y puntos de silla de la función de coste. La optimización dentro de este conjunto para llegar a un mínimo absoluto se realizará posteriormente.

### 4.2.2 Mínimos locales de la función de coste

En este apartado calcularemos la expresión del gradiente de la función de coste respecto a los coeficientes del filtro para el caso de una no-linealidad general. En apartados posteriores particularizaremos dicha no-linealidad para el caso de la función de log-likelihood.

Calculemos el gradiente de la función de coste respecto a los coeficientes. Debido a que estamos trabajando con funciones de variable compleja, es necesario tener en cuenta antes de evaluar un gradiente con respecto a un vector complejo, si la función es o no analítica. Puesto que estamos considerando un caso general, supondremos que la no-linealidad es no-analítica, es decir, admite derivadas tanto respecto a una variable determinada como respecto a su conjugada. Más adelante deberá particularizarse la no-linealidad por la función de log-likelihood, siendo la función de densidad de una variable aleatoria compleja en general una función no-analítica como se muestra en el apartado 3.5.1. Por lo tanto, el gradiente de la función de coste se expresa como,

$$\nabla_{\mathbf{w}_k^H} J = -E_{a_k, n_k} \nabla_{\mathbf{w}_k^H} \Psi_\phi(y_k) = -E_{a_k, n_k} \left( \left( \frac{\partial}{\partial y_k} \Psi_\phi \right) \nabla_{\mathbf{w}_k^H} y_k + \left( \frac{\partial}{\partial y_k^*} \Psi_\phi \right) \nabla_{\mathbf{w}_k^H} y_k^* \right) \quad (4.2.2.1)$$

Ahora bien, recurriendo a la definición de  $y_k$  en función de  $\mathbf{w}_k$  y  $\mathbf{r}_k$ , el segundo término del gradiente anterior es zero ya que  $y_k$  sólo depende del hermítico del vector de coeficientes. Obtenemos entonces la siguiente igualdad,

$$\nabla_{\mathbf{w}_k^H} J = -E_{a_k, n_k} \left( \frac{\partial}{\partial y_k} \Psi_\phi \right) \nabla_{\mathbf{w}_k^H} y_k = -E_{a_k, n_k} \left( \frac{\partial}{\partial y_k} \Psi_\phi \right) (a_k \mathbf{s}_k + \mathbf{n}_k) \quad (4.2.2.2)$$

Igualando a cero obtenemos una igualdad para el vector de steering<sup>†</sup>,

$$\mathbf{s}_k = -\frac{E_{a_k, n_k} \dot{\Psi}_\phi \mathbf{n}_k}{E_{a_k, n_k} \dot{\Psi}_\phi a_k} = -\frac{E_{n_k} (E_{a_k} \dot{\Psi}_\phi) \mathbf{n}_k}{E_{a_k} (E_{n_k} \dot{\Psi}_\phi) a_k} \quad (4.2.2.3)$$

<sup>†</sup> donde a partir de ahora utilizaremos la siguiente notación abreviada para la derivada de la no-linealidad,

$$\dot{\Psi}_\phi = \frac{\partial}{\partial y_k} \Psi_\phi(y_k)$$

Necesitamos una expresión diferente para J a fin de obtener un resultado explícito para el vector de coeficientes. Consideremos  $\eta_k$ , el ruido a la salida del filtro adaptado y desglosemos la esperanza conjunta en concatenación de esperanzas,

$$J = -E_{a_k, n_k} \psi_\phi(y_k) = -E_{a_k} E_{n_k} \psi_\phi(y_k) = -E_{a_k} E_{n_k} \psi_\phi(a_k w_k^H s_k + \eta_k) \quad (4.2.2.4)$$

No obstante, la esperanza de una no-linealidad evaluada sobre dos variables aleatorias independientes puede reescribirse como la esperanza respecto una de las variables de la no-linealidad convolucionada con la función de densidad de la otra variable cambiada de signo, tal como se expresa en la Sección 8.1, es decir,

$$J = -E_{a_k} (\psi_\phi * p_{-\eta_k})(a_k w_k^H s_k) = -E_{a_k} (\psi_\phi * p_{-\eta_k})(y_k^a = a_k w_k^H s_k) \quad (4.2.2.5)$$

donde se ha utilizado la simetría de la distribución de  $\eta_k$ . Nótese que el argumento de la nueva no-linealidad se evalúa sobre la variable aleatoria equivalente al caso no ruidoso. Calculando el gradiente con respecto a los coeficientes llegamos a,

$$\begin{aligned} \nabla_{w_k^H} J &= -E_{a_k} \left( (\nabla_{w_k^H} \psi_\phi) * p_{\eta_k} + \psi_\phi * (\nabla_{w_k^H} p_{\eta_k}) \right) = -E_{a_k} \left( (\dot{\psi}_\phi * p_{\eta_k}) \nabla_{w_k^H} y_k^a + \left( \psi_\phi * \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right) \nabla_{w_k^H} \sigma_{\eta_k}^2 \right) \\ &= -E_{a_k} (\dot{\psi}_\phi * p_{\eta_k}) a_k s_k - E_{a_k} \left( \psi_\phi * \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right) R_k w_k \end{aligned} \quad (4.2.2.6)$$

donde en el segundo término hemos desglosado el gradiente de la función de densidad del ruido según la regla de la cadena en términos de la derivada parcial con respecto a la varianza y el gradiente de la varianza (la cual es el único parámetro que caracteriza tal función de densidad). Teniendo en cuenta además que la varianza del ruido a la salida se expresa mediante la siguiente ecuación,

$$\sigma_{\eta_k}^2 = w_k^H R_k w_k \quad (4.2.2.7)$$

trás igualación a cero y suponiendo que ninguno de los factores escalares es cero llegamos a la siguiente expresión para el vector de coeficientes,

$$\begin{aligned} w_k &= -\frac{E_{a_k} (\dot{\psi}_\phi * p_{\eta_k}) a_k}{E_{a_k} \left( \psi_\phi * \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right)} R_k^{-1} s_k = -\frac{E_{a_k, n_k} a_k \dot{\psi}_\phi(y_k)}{E_{a_k} \left( \psi_\phi * \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right)} R_k^{-1} s_k \\ &= \alpha_k R_k^{-1} s_k \end{aligned} \quad (4.2.2.8)$$

donde hemos utilizado la siguiente propiedad: para  $\eta$  Gaussiano, sabemos que,

$$\psi^* p_{\eta_k} = E_{\eta_k} \psi(y_k^a - \eta) = E_{\eta_k} \psi(y_k^a + \eta) \Rightarrow E_{a_k} \psi^* p_{\eta_k} = E_y \psi(y) \tag{4.2.2.9}$$

donde una simple sustitución completa la demostración.

Resulta sorprendente ver que para una variable aleatoria de distribución arbitraria y escogiendo entre una amplia variedad de no-linealidades, llegamos a un esquema de mínima varianza, donde, en ausencia de interferencias equivale a máxima relación señal-ruido a la salida del filtro (el factor de escalado es independiente en este sentido).

Para que este criterio tenga realmente una solución única, debe cumplirse que tanto el numerador como el denominador sean no-nulos y finitos. Esto determina algunas condiciones que debe cumplir la no-linealidad.

$$\left. \begin{aligned} E_{a_k} \left( \psi_\phi^* \frac{\partial}{\partial \sigma_{\eta_k}^2} p_{\eta_k} \right) & \begin{cases} < \infty \\ \neq 0 \end{cases} \\ E_{a_k, \eta_k} a_k \psi_\phi(y_k) & \neq 0 \end{aligned} \right\} \tag{4.2.2.10}$$

lo cual significa que  $\psi(y \rightarrow \infty) \rightarrow 0$  para garantizar la finitud de la cantidad anterior. Si tal condición no se cumple, la esperanza de la función no lineal no estaría acotada. Hemos demostrado pues que: *la minimización del criterio propuesto presenta una clara similitud al estimador de mínima varianza cuando la no-linealidad cumple las condiciones anteriormente expuestas (nótese que son las condiciones que cumple cualquier función de densidad o de log-likelihood).*

La no-linealidad debe cumplir condiciones adicionales para que la solución derivada en la ecuación 4.2.2.9 sea efectivamente un mínimo y no un punto de inflexión. Condiciones que veremos más adelante cuando consideremos el Hessiano de la función de coste. No obstante, hemos demostrado que cuando el parámetro  $\phi$  de la no-linealidad es fijo, todos los mínimos locales se encuentran en la misma línea, correspondiendo a diferentes escalados del parámetro  $\alpha_k$ . En la sección siguiente profundizaremos más sobre las características del escalado del mínimo absoluto e intentaremos delimitar en qué rango de valores se encuentra el escalado correspondiente al mínimo absoluto cuando utilizamos la no-linealidad log-likelihood.

### 4.2.3 Mínimo absoluto de la función de coste

En el apartado anterior hemos analizado la distribución de los mínimos locales para una no-linealidad general, viendo que estos se encontraban sobre una misma línea, en puntos correspondientes a los escalados  $\alpha_k$  de cada mínimo local. En esta sección, y utilizando resultados del apartado anterior, analizaremos el mínimo absoluto de la función de coste para la no-linealidad específica log-likelihood. Cabe destacar entonces que los resultados aquí obtenidos no son generalizables a cualquier no-linealidad, sino que ara obtenerlos nos hemos basado en las propiedades especiales de la esperanza de una función log-likelihood. Consideramos de momento una no-linealidad fija ( $\phi$ =constante). Nuestro objetivo consistirá en determinar qué rango de valores puede tomar el escalado correspondiente al mínimo absoluto.

Como se ha demostrado, el vector de coeficientes óptimo debe ser proporcional a la solución de mínima varianza, tal como exponemos en la siguiente ecuación. Introducimos algunas nuevas variables en nuestra notación,

$$\bar{w}_k = R_k^{-1} s_k \quad , \quad w_k = \alpha_k \bar{w}_k \quad (4.2.3.1)$$

donde utilizamos el vector  $w_k$  super-barrado como referencia de escalado, y entendiendo al vector  $w_k$  como el correspondiente al mínimo absoluto. Definimos también las siguientes variables aleatorias a la salida del sistema adaptativo,

$$\bar{y}_k = \bar{w}_k^H r_k \quad , \quad z_k = \alpha_k \bar{y}_k \quad (4.2.3.2)$$

donde  $\alpha_k$  representa el escalado correspondiente al mínimo absoluto de la función de coste, y la variable  $y_k$  super-barrada corresponde a la salida del filtro cuando utilizamos el vector de coeficientes definido en la ecuación 4.2.3.1. Se cumple, por consiguiente, que,

$$-E_{z_k} \psi_\phi(z_k) \leq -E_{x=\alpha \bar{y}_k} \psi_\phi(x) = -E_{\bar{y}_k} \psi_\phi(\alpha' \bar{y}_k) \quad (4.2.3.3)$$

al tener la variable aleatoria  $z_k$  el escalado correspondiente al mínimo absoluto. Cualquier otro escalado  $\alpha'$  se traducirá en un valor mayor de la función de coste. Nos preguntamos ahora si el mínimo absoluto se alcanza para aquel escalado  $\alpha'$  tal que se cumplan las siguientes expresiones,

$$(4.2.3.4)$$

$$\alpha' \bar{y}_k = z_k = a_k + \frac{\mathbf{s}_k^H \mathbf{R}_k^{-1} \mathbf{n}_k}{\mathbf{s}_k^H \mathbf{R}_k^{-1} \mathbf{s}_k} = a_k + \eta_k, \quad \alpha' = \alpha_k$$

en cuyo caso, el escalado de la función de densidad sería el correcto con respecto al criterio de mínima varianza puesto que recuperamos la señal deseada  $a_k$  (más términos de ruido) con la amplitud correcta de valor 1. Primero, a fin de demostrar si  $J(\alpha')$  presenta un mínimo (local), en el escalado correcto de mínima varianza, debemos evaluar la siguiente derivada en ese punto (nótese que aunque resultase ser cero, esto no demostraría que fuese un mínimo absoluto, aunque el mínimo absoluto se da para algún valor de  $\alpha'$ ),

(4.2.3.5)

$$\left. \frac{d}{d\alpha'} J(\alpha') \right|_{\alpha' = \frac{1}{\mathbf{s}_k^H \mathbf{R}_k^{-1} \mathbf{s}_k}} = 0$$

donde el escalado correspondiente al criterio de mínima varianza viene dado por la expresión  $[\mathbf{s}_k^H \mathbf{R}_k^{-1} \mathbf{s}_k]^{-1}$ . Llegados a este punto, debemos explicitar cuál es la forma de la no-linealidad para poder continuar con la demostración. Tomémosla como la función log-likelihood parametrizada,

(4.2.3.6)

$$\psi_\phi(x) = \ln p_{a_k + \eta_k}(x), \quad \phi = E|\eta_k|^2$$

donde tomamos los parámetros  $\phi$  como la potencia del término de ruido  $\eta_k$  de la función de densidad. Para llegar al mínimo absoluto, minimizaremos la función de coste basada en la no-linealidad anterior en los parámetros  $\alpha'$  y  $\phi$ . Es decir, consideremos la siguiente expresión de la función de coste,

(4.2.3.7)

$$J(\alpha', \sigma_\eta^2) = -E_{z_k = \alpha' \bar{y}_k} \ln p_{a_k + \eta_k}(z_k)$$

donde mediante  $\sigma_\eta$  nos referimos a la varianza del término de ruido presente en la variable  $a_k + \eta_k$ :  $\eta_k$ . Definamos el escalado de mínima varianza como sigue,

(4.2.3.8)

$$\beta_k = \frac{1}{\mathbf{s}_k^H \mathbf{R}_k^{-1} \mathbf{s}_k}$$

y evaluemos en un primer paso la función de coste de la ecuación 4.2.3.7 para la variable aleatoria  $z_k = \beta_k \mathbf{y}_k = a_k + \eta'_k$  correspondiente al criterio de mínima varianza, considerando

únicamente variación en el parámetro  $\sigma_\eta$ . Debido a las propiedades<sup>†</sup> de la discriminación de Kullback tenemos que,

$$-E_{z_k} \ln p_{a_k + \eta_k}(z_k) \leq -E_{z_k} \ln p_{a_k + \eta'_k}(z_k) \quad , \quad E|\eta'_k|^2 \neq \sigma_\eta^2 = E|\eta_k|^2 \tag{4.2.3.9}$$

es decir, conseguimos minimizar la función de coste con respecto al parámetro varianza sólo cuando este coincide con la varianza real del ruido a la salida del filtro. Por lo tanto, suponiendo dado el escalado correcto, puede estimarse la varianza del ruido sin polarización. Queda demostrar si lo mismo sucede para el parámetro de escala.

Definamos ahora la variable aleatoria  $z'_k = \alpha'_k y_k$  y supongamos que el parámetro varianza de la función de coste es correcto, es decir, coincide con la varianza real, para determinar si tenemos un mínimo (al menos local) en  $\alpha'_k = \beta_k$ . Con esto comprobaremos si existe un mínimo en la solución de mínima varianza. Simplemente debemos calcular la derivada con respecto al parámetro de escala. La función de coste puede expresarse ahora de la siguiente forma.

$$J(\alpha'_k) = -E_{x=\alpha'_k \bar{y}_k} \ln p_{a_k + \eta_k}(x) = -E_{\bar{y}_k} \ln p_{a_k + \eta_k}(\alpha'_k \bar{y}_k) \tag{4.2.3.10}$$

su derivada se expresa por tanto como,

$$\begin{aligned} \left. \frac{d}{d\alpha'} J(\alpha') \right|_{\alpha' = \beta_k} &= -E_{\bar{y}_k} \left. \frac{\dot{p}_{a_k + \eta_k}(\alpha'_k \bar{y}_k)}{p_{a_k + \eta_k}(\alpha'_k \bar{y}_k)} \bar{y}_k \right|_{\alpha' = \beta_k} = - \int \bar{y}_k \dot{p}_{a_k + \eta_k}(\beta_k \bar{y}_k) d\bar{y}_k \\ &= - \frac{1}{\beta_k^2} \int z_k \dot{p}_{a_k + \eta_k}(z_k) dz_k \end{aligned} \tag{4.2.3.11}$$

Podemos encontrar un valor más explícito de la expresión anterior considerando la siguiente igualdad en función de la discriminación de Kullback,

$$\left. \frac{d}{d\alpha} D(X||\alpha X) \right|_{\alpha=1} = 0 \tag{4.2.3.12}$$

puesto que en  $\alpha=1$ , la discriminación presenta un mínimo al ser iguales ambas distribuciones. Continuando obtenemos,

$$\tag{4.2.3.13}$$

---

<sup>†</sup>  $D(z_k = a_k + \eta'_k || a_k + \eta_k) = E_{z_k} \ln \frac{p_{a_k + \eta'_k}(z_k)}{p_{a_k + \eta_k}(z_k)} \geq 0$

$$D(X||\alpha X) = E_x \ln \frac{p_X(x)}{p_X(x/\alpha)/|\alpha|^2} = E_x \ln p_X(x) - E_x \ln p_X(x/\alpha) + \ln|\alpha|^2$$

la derivada con respecto a  $\alpha$  conduce a,

$$\frac{d}{d\alpha} D(X||\alpha X) = E_x \frac{\dot{p}_X(x/\alpha)}{p_X(x/\alpha)} \frac{x}{\alpha^2} + \frac{1}{\alpha} \stackrel{\alpha=1}{=} \int_C x \dot{p}_X(x) dx + 1 \tag{4.2.3.14}$$

Por lo tanto, para cualquier variable aleatoria se cumple que,

$$\int_C x \dot{p}_X(x) dx = -1 \tag{4.2.3.15}$$

La derivada de la ecuación 4.2.3.11 es entonces igual a,

$$\left. \frac{d}{d\alpha'} J(\alpha') \right|_{\alpha'=\beta_k} = \frac{1}{\beta_k^2} \tag{4.2.3.16}$$

con lo cual demostramos que la densidad de probabilidad que minimiza el criterio se encuentra polarizada en escala y que existe un mínimo para  $\alpha' < \beta_k$  como puede determinarse a partir del signo de la derivada anterior. Aunque así se demuestra que existe un mínimo a la izquierda de  $\beta_k$  aún resta por probar que el mínimo absoluto se encuentra también en esta región. Recordemos que de momento estamos considerando  $\phi$  fijo y de valor igual a la mínima varianza en el caso óptimo. Destacamos que  $J(\alpha')$  también puede expresarse de la siguiente forma, renormalizando el escalado con respecto al de mínima varianza,

$$J(\alpha'_1) = J(\alpha'/\beta_k) = -E_{z_k} \ln p_{z_k}(\alpha'_1 z_k) \tag{4.2.3.17}$$

Por construcción formamos la siguiente igualdad,

$$-E_{z_k} \ln p_{z_k}(\alpha'_1 z_k) = -E_{z_k} \ln |\alpha'_1|^2 p_{z_k}(\alpha'_1 z_k) + \ln |\alpha'_1|^2 = -E_{z_k} \ln p_{z_k/\alpha'_1}(z_k) + \ln |\alpha'_1|^2 \tag{4.2.3.18}$$

Insertando la discriminación de Kullback en la ecuación anterior obtenemos que,

$$J(\alpha'_1) = -E_{z_k} \ln p_{z_k}(z_k) + E_{z_k} \ln \frac{p_{z_k}(z_k)}{p_{z_k/\alpha'_1}(z_k)} + \ln |\alpha'_1|^2 = H(z_k) + D(z_k||z_k/\alpha'_1) + \ln |\alpha'_1|^2 \tag{4.2.3.19}$$

donde  $H(\cdot)$  denota la entropía de la variable aleatoria considerada. Por lo tanto, considerando el margen  $|\alpha_1| \geq 1$  y recurriendo al carácter no-negativo de la discriminación, encontramos que,

$$J(\alpha_1) \stackrel{|\alpha_1| \geq 1}{\geq} H(z_k) \quad , \quad J(\alpha_1) \stackrel{|\alpha_1| = 1}{=} H(z_k) \tag{4.2.3.20}$$

lo cual, junto con la derivada anterior demuestra que el mínimo absoluto se encuentra en algún punto de la región  $|\alpha_1| \leq 1$ .

Se ha demostrado que el mínimo absoluto no se da en la función de densidad deseada, excepto en un término de escala. La razón de esto está basado en la ecuación 4.2.3.18. Podemos ver que la elección de este criterio es equivalente a la minimización de la discriminación de Kullback con respecto a la función de densidad deseada más una función de los parámetros. Entonces, si esta función de los parámetros no presenta un mínimo en el mismo punto que la discriminación, obtenemos una función de densidad polarizada en el mínimo del criterio.

Consideraremos ahora la localización del mínimo absoluto cuando el parámetro varianza no es igual al óptimo. Tomemos un valor fijo del escalado y examinemos la variación de la función de coste con respecto al parámetro varianza. Definamos la siguiente variable aleatoria,

$$z'_k = a_k + \eta'_k \quad , \quad \sigma_1^2 = E|\eta'_k|^2 \neq E|\eta_k|^2 \tag{4.2.3.21}$$

y estudiemos la minimización de  $J(\alpha_1, \sigma_1^2)$ . Utilizando la misma estrategia que en las ecuaciones 4.2.3.18-19, tenemos que,

$$J(\alpha_1, \sigma_1^2) = -E_{z_k} \ln p_{z_k}(z_k) + E_{z_k} \ln \frac{p_{z_k}(z_k)}{p_{z'_k/\alpha_1}(z_k)} + \ln|\alpha_1|^2 = H(z_k) + D(z_k || z'_k/\alpha_1) + \ln|\alpha_1|^2 \tag{4.2.3.22}$$

con el resultado neto que,

$$J(\alpha_1, \sigma_1^2) \stackrel{|\alpha_1| \geq 1}{\geq} H(z_k) \quad , \quad J(\alpha_1, \sigma_1^2) \stackrel{\alpha_1=1, \sigma_1^2 = E|\eta_k|^2}{=} H(z_k) \tag{4.2.3.23}$$

Por consiguiente, el mínimo absoluto, sin fijar el parámetro varianza, se encuentra también en el margen  $|\alpha_1| \leq 1$ . Con esto se demuestra que la utilización de esta no-linealidad resulta

en una disminución del radio de la constelación a la salida del filtro con relación al radio de la función de densidad deseada. Debemos demostrar ahora si la distribución que minimiza el criterio también está polarizada con respecto al parámetro varianza. Para ello debemos calcular la derivada parcial respecto a este parámetro y evaluarla en la varianza óptima. Si no es cero, se demuestra que no es un mínimo absoluto.

#### 4.2.4 Obtención del parámetro varianza

El propósito de este apartado es doble: por una parte consideraremos la obtención del parámetro varianza como minimización global de la función de coste y por otra relacionaremos el término de escalado con el parámetro varianza mediante una ecuación de ligadura. Aunque debido al carácter no lineal de la función de coste no podemos hallar una ecuación cerrada que nos exprese tanto el parámetro varianza como el escalado en función de parámetros del modelo de señal como la varianza de ruido Gaussiano a la entrada, sí que podemos, en cambio, hallar una ecuación cerrada o de ligadura que involucra a ambos términos.

A fin de calcular la derivada de la función de coste respecto al parámetro varianza, resulta útil expresar la función de densidad de una distribución en términos de la esperanza estadística de una función de dos variable aleatorias independientes tal como sigue† ,

$$p_{z=x+y}(z) = E_x p_y(z-x) = E_y p_x(z-y) \tag{4.2.4.1}$$

de tal forma que la no-linealidad se puede expresar como,

$$\psi(x) = \ln E_{a'_k} p_{\eta'_k}(x - a'_k) , \quad p_{\eta'_k}(x) = \frac{1}{\pi \sigma_i^2} e^{-|x|^2 / \sigma_i^2} \tag{4.2.4.2}$$

Según la ecuación 4.2.2.8, vimos que el vector óptimo de coeficientes se obtenía como el escalado de  $\mathbf{R}_k^{-1} \mathbf{s}_k$ . Por lo tanto, hemos de evaluar el mínimo conjunto respecto a los parámetros de escala y de varianza del siguiente criterio estadístico equivalente,

$$J(\alpha, \sigma_i^2) = -E_{\bar{y}_k} \ln E_{a'_k} p_{\eta'_k}(\alpha \bar{y}_k - a'_k) = \ln \pi \sigma_i^2 - E_{\bar{y}_k} \ln E_{a'_k} e^{-|\alpha \bar{y}_k - a'_k|^2 / \sigma_i^2} \tag{4.2.4.3}$$

---

† En la sección 3.4 ya vimos una derivación alternativa basada en funciones de densidad condicionadas

Es decir, consideramos una minimización de la función de coste restringida a una recta en el espacio de coeficientes, puesto que sabemos que en esa recta se hallan los mínimos locales.

(a) Derivando con respecto al parámetro de escala obtenemos la siguiente igualdad expresada también en función de la no-linealidad,

$$\frac{\partial}{\partial \alpha} J = E_{\bar{y}_k} e^{-\psi} E_{a'_k} \frac{1}{\sigma_i^2} (\alpha \bar{y}_k - a'_k) \bar{y}_k^* e^{-|\alpha \bar{y}_k - a'_k|^2 / \sigma_i^2} = 0 \quad (4.2.4.4)$$

(b) Derivando con respecto al parámetro varianza obtenemos,

$$\frac{\partial}{\partial \sigma_i} J = \frac{2}{\sigma_i} - E_{\bar{y}_k} e^{-\psi} E_{a'_k} \frac{2}{\sigma_i^3} |\alpha \bar{y}_k - a'_k|^2 e^{-|\alpha \bar{y}_k - a'_k|^2 / \sigma_i^2} = 0 \quad (4.2.4.5)$$

De la última igualdad despejamos la varianza que minimiza el criterio,

$$\sigma_i^2 = E_{\bar{y}_k} e^{-\psi} E_{a'_k} |\alpha \bar{y}_k - a'_k|^2 e^{-|\alpha \bar{y}_k - a'_k|^2 / \sigma_i^2} \quad (4.2.4.6)$$

De la derivada con respecto al parámetro de escala obtenemos la siguiente relación,

$$\alpha E_{\bar{y}_k} |\bar{y}_k|^2 = E_{\bar{y}_k} e^{-\psi} E_{a'_k} \bar{y}_k^* a'_k e^{-|\alpha \bar{y}_k - a'_k|^2 / \sigma_i^2} \quad (4.2.4.7)$$

de donde podemos deducir una expresión explícita para el escalado. Sustituyendo la derivada con respecto al parámetro varianza resulta finalmente,

$$\sigma_i^2 + |\alpha|^2 E_{\bar{y}_k} |\bar{y}_k|^2 = E_{\bar{y}_k} e^{-\psi} E_{a'_k} |a'_k|^2 e^{-|\alpha \bar{y}_k - a'_k|^2 / \sigma_i^2} \quad (4.2.4.8)$$

donde el término de la derecha se puede interpretar como una estimación de la potencia de la señal a la salida del sistema. Para el caso particular de señales de amplitud constante podemos demostrar que el término de la derecha es exactamente igual a uno (si consideramos amplitud normalizada a uno). Por lo tanto, llegamos a una expresión que nos relaciona la varianza con el escalado que minimizan el criterio,

$$\sigma_i^2 + |\alpha|^2 E_{\bar{y}_k} |\bar{y}_k|^2 = 1 \quad (4.2.4.9)$$

Resulta fácil ahora relacionar ambos parámetros con la varianza del ruido a la entrada. A partir de la definición de  $y_k$  tenemos,

$$(4.2.4.10)$$

$$E_{\bar{y}_k} |\bar{y}_k|^2 = 1 + \frac{1}{\mathbf{s}_k^H \mathbf{R}_k^{-1} \mathbf{s}_k}$$

con la amplitud de la señal normalizada a uno. Suponiendo ruido blanco a la entrada llegamos a la siguiente ecuación de ligadura,

$$\sigma_i^2 + |\alpha|^2 \left( 1 + \frac{1}{SNR_{in}} \right) = 1 \quad (4.2.4.11)$$

A partir de la ecuación de ligadura anterior podemos obtener fácilmente una cota superior para el parámetro de escalado. Puesto que el cuadrado de la varianza es positivo obtenemos,

$$\sigma_i^2 + |\alpha|^2 \left( 1 + \frac{1}{SNR_{in}} \right) = 1 \geq |\alpha|^2 \left( 1 + \frac{1}{SNR_{in}} \right) \Rightarrow |\alpha|^2 \leq \frac{1}{1 + \frac{1}{SNR_{in}}} \quad (4.2.4.12)$$

Llegando así a una cota superior para el parámetro de escala. Vemos que se admite la solución  $\alpha=1$  para el caso de relación señal a ruido infinita. A medida que disminuye la relación señal a ruido disminuye también la cota superior del escalado.

En las últimas secciones hemos considerado la aplicación de la primera función de coste al modelo de una fuente, llegando a un conocimiento más íntimo de la función de coste. Hemos visto como resultado más importante, cuál es la relación entre el vector de coeficientes óptimo según este criterio y según el criterio de referencia espacial o de mínima varianza. En la siguiente sección realizaremos un estudio parecido e introduciremos nuevas propiedades para el caso de varias fuentes, extendiendo algunos de los resultados aquí obtenidos a estadísticas más generales de la señal de entrada.