

Capítulo 4

Modulación y detección en canales gaussianos

La función básica de un sistema de comunicaciones digitales consiste en el envío y la correspondiente recuperación de manera fiable de una secuencia digital de información enviada a través de un canal de naturaleza analógica. Este es el problema que se aborda en este capítulo.

El capítulo comienza introduciendo la conveniencia de una representación geométrica de las señales para el diseño y análisis de un sistema de comunicaciones digitales, y se presentan las características básicas de dicha representación. A continuación se estudia el problema de la transmisión en canales de ruido aditivo blanco y gaussiano, lo que se denomina comúnmente como *canal gaussiano*; en este modelo se supone que la única distorsión que sufre la señal de comunicaciones durante su transmisión es la suma de ruido térmico, que se modela mediante un proceso aleatorio estacionario, ergódico, blanco, gaussiano, con densidad espectral de potencia $N_0/2$ W/Hz. Se analizan los elementos funcionales del transmisor y del receptor, abordando el diseño óptimo de cada elemento, y se presenta la metodología para obtener las prestaciones que se obtienen con un cierto sistema.

4.1. Introducción

Este capítulo se centra en el análisis de sistemas de comunicaciones digitales. Conviene por tanto recordar cuáles son las características de un sistema de comunicaciones digitales y cuál es el modelo de un sistema de comunicaciones digitales.

4.1.1. Ventajas de los sistemas de comunicaciones digitales

Los sistemas de comunicaciones digitales predominan claramente sobre los sistemas digitales porque tienen múltiples ventajas sobre estos. Las más importantes son las siguientes:

- Capacidad de “*regeneración*”. Esta es sin duda la principal ventaja. Bajo ciertas circunstancias es posible transmitir sin errores, o en general, con una probabilidad de error arbitrariamente baja. Esto es posible debido a que en un cierto instante de tiempo, el símbolo transmitido es uno entre un conjunto finito, lo que supone que la señal puede tener una

forma entre un conjunto finito, y si la distorsión no es muy severa se puede buscar la más parecida, y “regenerar” la señal, como se ilustra en la Figura 4.1. Como se ve en la figura, en cada intervalo se transmite una entre dos posibles formas de onda, nivel alto y nivel, bajo, por lo que si la distorsión es moderada en el receptor se puede recuperar la señal en cada intervalo y recuperar de forma exacta la señal transmitida.

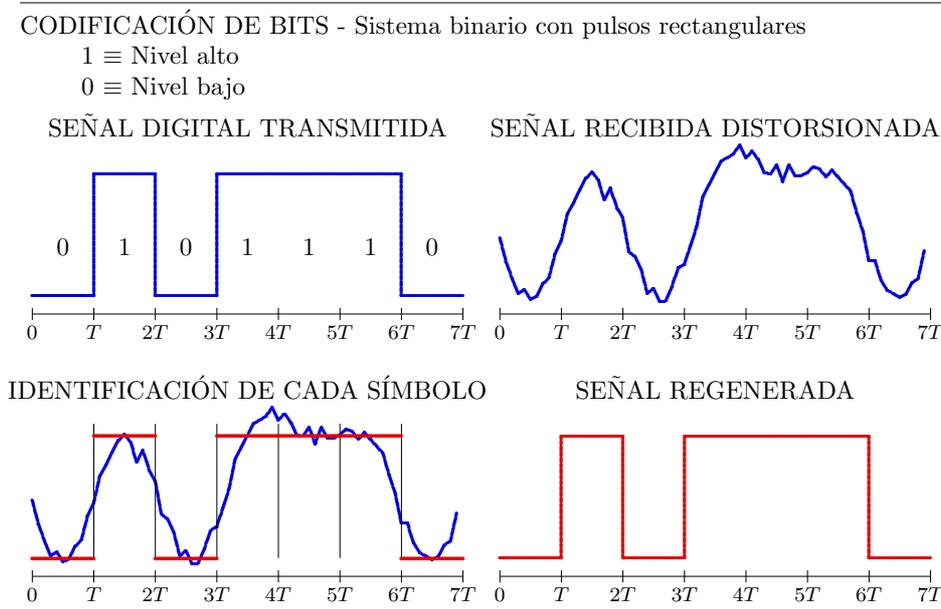


Figura 4.1: Ilustración de la capacidad de regeneración en un sistema de comunicaciones digitales.

- Existen técnicas de detección y corrección de errores. En el ejemplo de la Figura 4.1 la distorsión que sufre la señal transmitida es moderada, por lo que se recupera la señal transmitida (secuencia binaria) sin errores. En casos reales se producen errores, pero existen técnicas que permiten detectar y corregir errores.
- La información se puede encriptar (proteger) de forma relativamente sencilla. Basta con aplicar operaciones aritméticas sobre la secuencia binaria de información.
- Permite corregir la distorsión introducida por el canal (igualación de canal) de una forma mucho más sencilla que en sistemas analógicos. El hecho de saber qué posibles valores puede tomar la señal en un cierto intervalo, hace más fácil estimar el efecto del canal sobre la señal transmitida para invertirlo.
- El formato de la información transmitida es independiente del tipo de información (voz, datos, TV, etc.). En cualquier caso, la información está contenida en una secuencia de símbolos binarios (bits), por lo que al diseñar un sistema, hay que tratar de transmitir bits de forma eficiente, independientemente del tipo de información que contengan.
- Permite utilizar TDM/TDMA y CDM/CDMA (además de FDM/FDMA) como mecanismo de multiplexación/acceso al medio.
- Los circuitos son, en general
 - Más fiables
 - De menor coste

- Más flexibles (programables)

Evidentemente no todo son ventajas en la comparación entre sistemas analógicos y sistemas digitales. Aunque las ventajas claramente predominan sobre las desventajas, es conveniente conocer también estas. Las más importantes son las siguientes:

- La necesidad de sincronismo en el receptor. En el ejemplo de la Figura 4.1, están identificados los instantes donde comienza y finaliza la transmisión de cada bit, lo que permite comparar en cada intervalo la señal recibida con un nivel alto o bajo para realizar la regeneración de la señal (o lo que es lo mismo, la identificación de la secuencia binaria transmitida). En un sistema real, el receptor digital debe generar la señal de sincronismo que permite identificar estos instantes. Esto no es necesario en sistemas analógicos.
- En general, una señal de información para una fuente del mismo tipo (por ejemplo una señal de voz) requiere un mayor ancho de banda para su transmisión en formato digital que en formato analógico. Esta desventaja se puede compensar utilizando técnicas de compresión de las señales en formato digital, lo que permite transmitir las con un menor ancho de banda.
- Muchas fuentes de información son de naturaleza analógica. En realidad, este es un inconveniente menor, ya que como ya se ha visto en el Capítulo 1, la conversión de señales analógicas a formato digital (y la correspondiente reconversión de formato digital a analógico) permite la transmisión de fuentes de naturaleza analógica a través de sistemas de comunicaciones digitales. La tecnología de conversión analógico/digital (A/D) y digital/analógico (D/A) es a día de hoy una tecnología madura y económica.

4.1.2. Descripción general de un sistema de comunicaciones digitales

La principal característica de un sistema de comunicaciones digitales es que la información a transmitir está contenida en una secuencia de símbolos pertenecientes a un alfabeto finito de símbolos, en oposición a los sistemas de comunicaciones analógicos, donde la información está contenida en la forma de una determinada señal moduladora $m(t)$. Esta característica dota a este tipo de sistemas de una serie de ventajas que hacen que hayan desbancado a los sistemas de comunicaciones analógicos, como se ha visto en el capítulo introductorio de la asignatura.

Aunque en general se puede codificar la información en distintos alfabetos, este capítulo se va a centrar en el caso más frecuente, en el que la información está contenida en una secuencia binaria, con lo que el objetivo del sistema será la transmisión eficiente de bits entre la fuente y el destino de la transmisión.

Las Figuras 4.2 y 4.3 muestran los diagramas de bloques de un transmisor y de un receptor digital, respectivamente. Estos diagramas incluyen los elementos funcionales básicos que aparecen en cualquier sistema.

El primer elemento del transmisor, el codificador de fuente, se encarga de reducir la redundancia de la fuente de información (compresión de la información) para reducir la tasa binaria de la información a transmitir. El codificador de canal se encarga de introducir redundancia de forma controlada para permitir la detección y corrección de un cierto número de errores de bit. El ejemplo más sencillo de codificadores de canal son los codificadores de repetición. Estos códigos repiten varias veces cada bit a transmitir. Un código de repetición de orden 1 cuando transmite un cierto

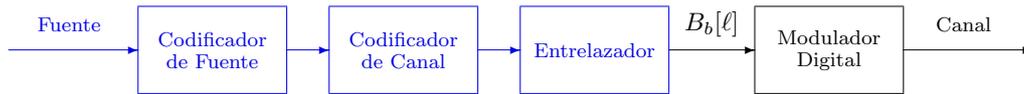


Figura 4.2: Diagrama de bloques funcionales del transmisor de un sistema de comunicaciones digitales.

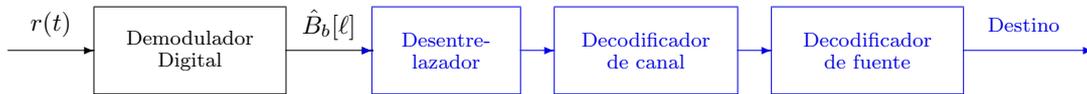


Figura 4.3: Diagrama de bloques funcionales del receptor de un sistema de comunicaciones digitales.

bit lo repite otra vez, de forma que se envía cada bit dos veces. Un código de repetición de orden 2 repite cada bit dos veces, con lo que al final se transmite cada bit tres veces, realizando un proceso de codificación

$$0 \rightarrow 000$$

$$1 \rightarrow 111$$

En cada bloque de tres bits codificados, en el receptor es posible la detección de hasta dos errores; adoptando la estrategia de decisión por mayoría, en el receptor se podrá corregir un error sobre cada bloque de tres bits transmitidos. Si hay dos o tres errores, con este esquema de decisión por mayoría se tomaría una decisión errónea sobre el bloque. Así, cada código de canal tiene una cierta capacidad de detección y corrección.

En muchas ocasiones los errores de bit en la secuencia transmitida en un sistema no se producen de forma aislada, sino que se producen en ráfagas de errores. La aplicación directa de las técnicas de codificación de canal sobre ráfagas de errores no es útil, ya que funcionan bien cuando sobre cada bloque de bits codificados (3 en el caso del ejemplo anterior) hay un número limitado de errores. El papel del entrelazador es convertir ráfagas de errores de bit en errores aislados. Para ello se reordenan los bits antes de la transmisión (entrelazado) y se devuelven al orden original en el receptor (desentrelazado), de forma que errores de ráfaga antes del desentrelazado se convierten en errores aislados a la salida del desentrelazador, tal y como se muestra en el ejemplo de la Figura 4.4.

En cuanto al transmisor, a partir de la fuente del sistema de comunicaciones, y después de realizar los procesos de codificación de fuente, codificación de canal, y del entrelazado, se genera una secuencia binaria $B_b[\ell]$ que contiene la información que se quiere transmitir. En este capítulo se va a estudiar el último elemento del transmisor, el modulador digital. La principal función de este modulador digital es la transmisión de la secuencia de bits, $B_b[\ell]$, a través de un canal de comunicaciones analógico, para lo que debe convertir dicha secuencia en una señal electromagnética $s(t)$ adecuada para su transmisión por el medio analógico utilizado para la transmisión (un cable, fibra óptica, el espectro radioeléctrico, etc.).

En cuanto al receptor, el demodulador digital se encargará de la recuperación de la secuencia de bits que se ha transmitido a partir de la señal recibida través del canal de comunicaciones, $r(t)$. Idealmente se debería recuperar de forma exacta la secuencia transmitida, pero la distorsión sufrida por la señal durante su transmisión hará que esto no sea en general posible, por lo que el demodulador digital deberá proporcionar una estima de la secuencia transmitida, $\hat{B}_b[\ell]$, intentando minimizar el número de errores de la estima.

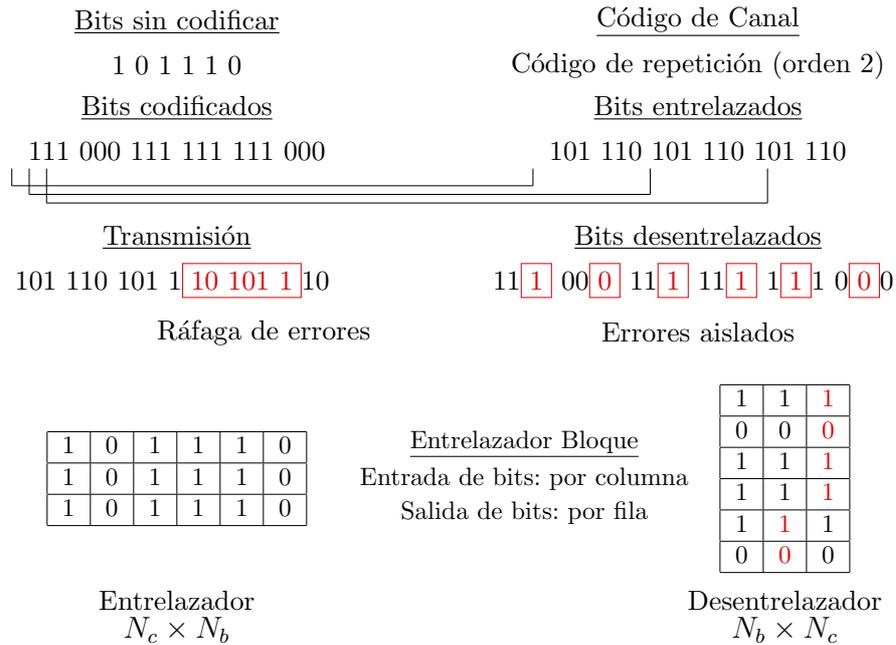


Figura 4.4: Ejemplo de entrelazador matricial, en el que se reordenan los bits introduciéndolos en una estructura matricial por columnas y sacándolos leyendo por filas.

En este capítulo se estudiarán el modulador digital del transmisor y el demodulador digital del receptor. Es decir, que los problemas básicos que se abordan son la transmisión de una secuencia de bits a través de un canal analógico, y la recuperación de dicha secuencia a partir de la señal recibida a la salida del canal, que habrá sufrido cierto tipo de distorsiones durante su transmisión a través del canal analógico.

4.1.3. Diseño genérico de un modulador digital y notación básica

Como se acaba de ver, la función del modulador digital es la transmisión de bits (secuencia $B_b[\ell]$) a una tasa binaria $R_b = \frac{1}{T_b}$ bits/s a través de un medio de transmisión (canal) de naturaleza analógica, para lo que deberá convertir la secuencia de bits en una señal electromagnética $s(t)$.

En general, la transmisión de la secuencia de bits no se realizará bit a bit, sino que se realizará por bloques de m bits. Se segmentará la secuencia de bits $B_b[\ell]$ en bloques de m bits, y se enviará cada uno de estos bloques. A cada bloque de m bits se le denominará *símbolo*, de modo que el alfabeto de los símbolos tendrá $M = 2^m$ posibles valores: las M posibles combinaciones de m bits. Los elementos de este alfabeto se denotarán como b_i , con $i \in \{0, 1, \dots, M - 1\}$, es decir

$$B[n] \in \{b_i\}_{i=0}^{M-1}.$$

Así pues, lo primero que se realizará en el modulador digital será la conversión de la secuencia de símbolos $B_b[\ell]$ a una secuencia de símbolos $B[n]$. Un aspecto importante en esta conversión es el cambio de velocidades, ya que se pasará de la velocidad binaria o de bit R_b bits/s a la velocidad de símbolo R_s símbolos/s o baudios. La relación entre estas dos velocidades o tasas de transmisión es evidente. Dado que cada símbolo contiene m bits, la relación es

$$R_b = m \times R_s.$$

Asociada a cada velocidad hay una duración temporal, el tiempo de bit, T_b y el tiempo de símbolo, T , respectivamente, que indicará cada cuantos segundos se transmite un bit o un símbolo, respectivamente. La relación de estos tiempos con las velocidades de transmisión y entre ellos es también evidente

$$T_b = \frac{1}{R_b}, \quad T = \frac{1}{R_s}, \quad T = m \times T_b.$$

En este capítulo se presentará la forma más sencilla de transmisión de la información, con el objetivo de introducir las técnicas básicas de modulación y de demodulación. La secuencia de información digital ha de convertirse en una señal eléctrica. Esta conversión se realizará por tramos. Se va a realizar la transmisión de un símbolo cada T segundos, de forma que se asociará un intervalo temporal a la transmisión de cada símbolo. El primer símbolo de la secuencia, $B[0]$, se comenzará a transmitir en el instante $t = 0$, y durará hasta $t = T$. A continuación se transmitirá el segundo símbolo de la secuencia, $B[1]$, entre los instantes $t = T$ y $t = 2T$, y así sucesivamente. En general, el símbolo $B[n]$ tendrá asociado el intervalo $nT \leq t < (n + 1)T$. De este modo, la forma de la señal en cada intervalo de símbolo estará asociada al símbolo que se transmite en dicho intervalo.

Como se ha explicado anteriormente, la secuencia de símbolos puede tomar uno entre M posible valores para cada instante discreto n

$$B[n] \in \{b_0, b_1, \dots, b_{M-1}\}.$$

La forma más simple de realizar la conversión de símbolos a señal $s(t)$ es definiendo un conjunto de M señales de duración T segundos

$$\{s_0(t), s_1(t), \dots, s_{M-1}(t)\}, \text{ definidas en } 0 \leq t < T$$

y realizando una asociación símbolo / forma de onda:

$$b_i \leftrightarrow s_i(t).$$

Cuando en el instante n la secuencia de símbolos toma un cierto valor, por ejemplo $B[n] = b_j$, la forma de la señal $s(t)$ en el intervalo asociado a este símbolo, $nT \leq t < (n + 1)T$, será la forma de onda asociada a b_j , que es $s_j(t)$. Obviamente, hay que trasladar la forma de onda $s_j(t)$ al intervalo correspondiente, por lo que se traslada la forma de onda asociada a b_j al intervalo, quedando

$$s(t) = s_j(t - nT), \text{ en } nT \leq t < (n + 1)T.$$

A continuación se ilustra este procedimiento con un ejemplo sencillo. En este caso se van a transmitir los bits en bloques de tamaño $m = 2$ bits, por lo que el sistema tiene un alfabeto de $M = 2^2 = 4$ símbolos. Por ejemplo se puede hacer la siguiente asociación

$$b_0 \equiv 00, \quad b_1 \equiv 01, \quad b_2 \equiv 10, \quad b_3 \equiv 11,$$

aunque se podría haber hecho otra diferente. Hay que elegir un conjunto de $M = 4$ señales, y asociar una señal a cada símbolo. En este ejemplo se han elegido las cuatro señales de la Figura 4.5. La asociación realizada está implícita en los subíndices ($s_i(t)$ está asociada a b_i).

Se quiere realizar la transmisión de la siguiente secuencia binaria

$$B_b[\ell] = 011110001101 \dots$$

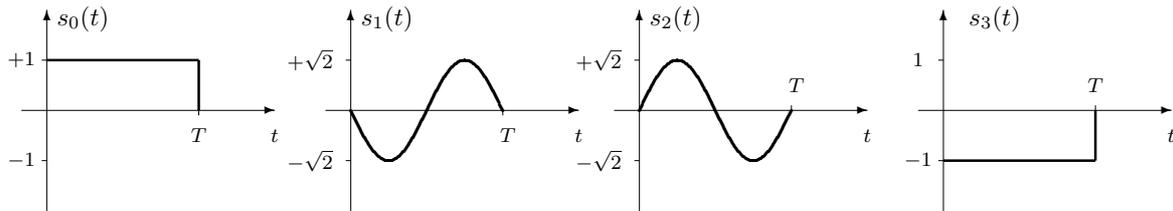


Figura 4.5: Conjunto de 4 señales elegidas para el ejemplo.

Se segmenta esta secuencia en bloques de $m = 2$ bits

$$B_b[\ell] : 01 \mid 11 \mid 10 \mid 00 \mid 11 \mid 01 \mid \dots$$

y se realiza la conversión de secuencia binaria a secuencia de símbolos identificando cada bloque de 2 bits de acuerdo a la asignación especificada anteriormente

$$B[n] = b_1 \mid b_3 \mid b_2 \mid b_0 \mid b_3 \mid b_1 \mid \dots$$

Ahora la señal transmitida se generará por intervalos de símbolo. Como el primer símbolo es b_1 , en el primer intervalo de símbolo ($0 \leq t < T$) se colocará la forma de onda asociada, $s_1(t)$. Como el segundo símbolo es b_3 , en el segundo intervalo de símbolo ($T \leq t < 2T$) se colocará la forma de onda asociada, $s_3(t)$, que se retardará T segundos para colocarla en ese intervalo. Como el tercer símbolo es b_2 , en el tercer intervalo de símbolo ($2T \leq t < 3T$) se colocará la forma de onda asociada, $s_2(t)$, que se retardará $2T$ segundos para colocarla en ese intervalo. De este modo, la señal modulada $s(t)$ se generará por intervalos de la forma siguiente

$$s(t) = \{s_1(t) \mid s_3(t - T) \mid s_2(t - 2T) \mid s_0(t - 3T) \mid s_3(t - 4T) \mid s_1(t - 5T) \mid \dots\}$$

La Figura 4.6 muestra la señal resultante al aplicar este proceso de generación por intervalos de símbolo.

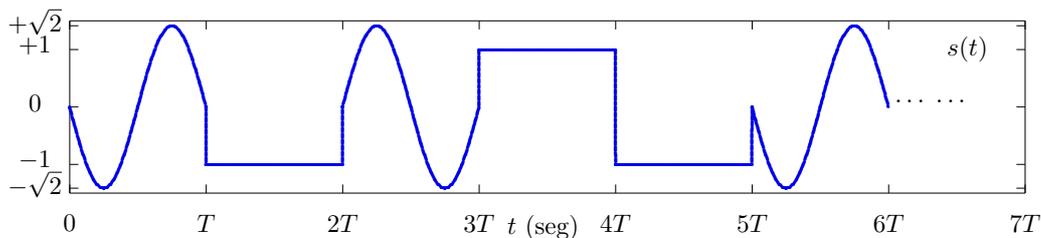


Figura 4.6: Señal generada por la secuencia binaria del ejemplo.

A la luz de este ejemplo, se puede ver claramente que el diseño del modulador digital consiste en la elección de M señales para asociarlas a cada uno de los M posibles valores de la secuencia de símbolo. En este ejemplo la elección ha sido arbitraria. ¿Es la elección del ejemplo una buena elección? ¿Qué criterios hay que tener en cuenta a la hora de elegir el conjunto de M señales? Esta es la pregunta a la que se tratará de dar respuesta a lo largo de este capítulo.

4.1.4. Transmisión a través de un canal de comunicaciones

Una vez generada la señal modulada $s(t)$ que contiene la información digital a enviar, esta se transmite a través de un medio físico o canal de comunicaciones, como podría ser un cable, fibra

óptica, el espectro radioeléctrico o cualquier otro medio que permita la transmisión de una señal. La señal sufre distorsiones durante la transmisión de forma que la señal recibida a la entrada del receptor, $r(t)$, no coincide con la señal transmitida: $r(t) \neq s(t)$.

Existen varios tipos de distorsión que afectan a una señal durante su transmisión por un medio físico. En esta asignatura se considerarán únicamente los dos más importantes: distorsión lineal y ruido. La forma habitual de modelar la distorsión lineal es mediante un modelo de sistema lineal e invariante, caracterizado por una cierta respuesta impulsiva $h(t)$ y su correspondiente respuesta en frecuencia $H(j\omega)$, respuestas relacionadas a través de la transformada de Fourier. El ruido se modela como un proceso aditivo cuyas características estadísticas se asocian en la mayoría de los casos al modelo habitual de ruido térmico: un proceso aleatorio estacionario, ergódico, blanco, gaussiano, de media nula y densidad espectral de potencia $N_0/2$ W/Hz, donde la constante N_0 se obtiene mediante el producto de la constante de Boltzmann y la temperatura expresada en grados Kelvin. Por tanto, el modelo de canal de comunicaciones que se utilizará en este capítulo será el definido por la relación

$$r(t) = s(t) * h(t) + n(t)$$

que se ilustra en la Figura 4.7.

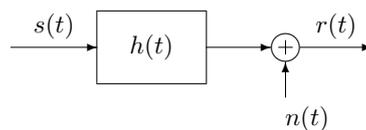


Figura 4.7: Modelo del canal de comunicaciones.

4.1.5. Diseño genérico de un demodulador digital y notación básica

La función del demodulador digital es la recuperación de la secuencia de bits $B_b[\ell]$ a partir de la señal recibida a través del canal, $r(t)$. La señal sufre distorsiones durante su transmisión a través del canal de comunicaciones, de modo que la señal recibida será diferente de la señal transmitida

$$r(t) \neq s(t).$$

El procedimiento básico para recuperar la señal sería el siguiente. Se procesará la señal recibida por tramos, en concreto por intervalos de símbolo. En el intervalo de símbolo asociado al índice discreto n , intervalo $nT \leq t < (n + 1)T$, se observará la señal recibida y se comparará con las M formas de onda del sistema. Se elegirá la señal con la que tenga un mayor “parecido”, y si esta es $s_j(t)$, la estima para el símbolo asociado a ese intervalo será el símbolo asociado a esa forma de onda $\hat{B}[n] = b_j$.

Volviendo al ejemplo anterior, tras la transmisión de la señal, esta sufre una cierta distorsión, dando lugar a la señal de la Figura 4.8.

Tras segmentar la señal en intervalos de símbolo, se buscará para cada intervalo de símbolo cuál de las M posibles señales del sistema (señales de la Figura 4.5 en este ejemplo) es más parecida a la señal recibida en el intervalo. En este caso, una sencilla inspección visual permite identificar la señal correspondiente a cada intervalo

- $n = 0$, intervalo $0 \leq t < T$ - Señal “más parecida”: $s_1(t) \rightarrow \hat{B}[0] = b_1$

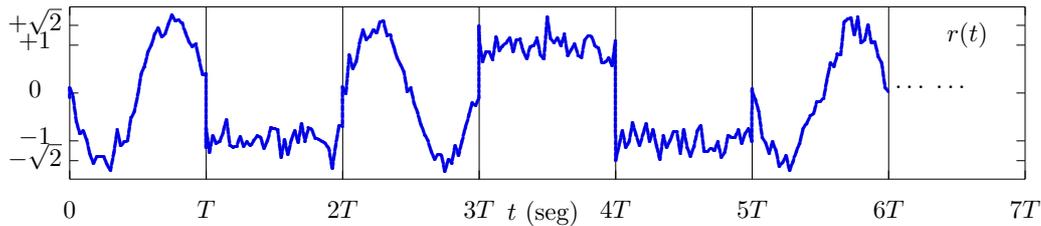


Figura 4.8: Señal recibida a la entrada del receptor en el ejemplo.

- $n = 1$, intervalo $T \leq t < 2T$ - Señal “*más parecida*”: $s_3(t) \rightarrow \hat{B}[1] = b_3$
- Siguiendo el mismo proceso: $\hat{B}[2] = b_2, \hat{B}[3] = b_0, \hat{B}[4] = b_3, \hat{B}[5] = b_1$

De este modo, la secuencia estimada sería

$$\hat{B}[n] = b_1 \mid b_3 \mid b_2 \mid b_0 \mid b_3 \mid b_1 \mid \dots$$

que tras la conversión de símbolos a bits proporcionaría

$$\hat{B}_b[\ell] : 01 \mid 11 \mid 10 \mid 00 \mid 11 \mid 01 \mid \dots$$

En el ejemplo la distorsión de la señal es moderada, de forma que mediante una inspección visual se recupera la secuencia original sin problemas. Pero el sistema real deberá medir de algún modo el “*parecido*” o “*diferencia*” entre señales. Es necesario definir una medida que cuantifique la diferencia de la señal recibida en el intervalo asociado al instante discreto n con la forma de onda $s_i(t)$, que podríamos denotar como

$$\text{diferencia}(n, i).$$

De este modo, el criterio para realizar una decisión sería decidir el símbolo asociado a la señal con una menor diferencia (mayor parecido) respecto a la señal recibida en el intervalo

$$\hat{B}[n] = b_i \text{ si } \text{diferencia}(n, i) < \text{diferencia}(n, j) \text{ para todo } j \neq i.$$

Por ejemplo, se nos podría ocurrir una medida que sume la diferencia en módulo entre la señal observada en el intervalo de símbolo y cada una de las posibles señales de referencia a lo largo del intervalo de símbolo (la suma sobre cada instante de tiempo se convierte en una integral)

$$\text{diferencia}_A(n, i) = \int_{nT}^{(n+1)T} |s(t) - s_i(t - nT)| dt.$$

Pero también se nos podría ocurrir medir directamente la energía de la señal diferencia, que vendría dada por

$$\text{diferencia}_B(n, i) = \int_{nT}^{(n+1)T} |s(t) - s_i(t - nT)|^2 dt.$$

Intuitivamente ambas medidas cuantifican de algún modo la diferencia entre las señales, dando valores más bajos cuanto más se parezcan (darían 0 sólo si las señales son idénticas). Y se nos podrían ocurrir muchas más medidas que podrían servir para cuantificar el parecido o diferencia entre señales. ¿Cuál sería la mejor medida para minimizar la probabilidad de equivocarse? El diseño del demodulador digital consistirá precisamente en eso, en buscar la mejor forma de cuantificar ese parecido para que la secuencia estimada sea la que minimice el número de estimas erróneas.

4.1.6. Factores a considerar a la hora de elegir las M formas de onda

Acabamos de ver qué función básica han de realizar el modulador y el demodulador digital en un sistema de comunicaciones digitales, qué procedimiento se utilizará para realizar dicha función, y en qué consiste su diseño. Resumiendo:

- Modulador digital

- Función: Conversión de una secuencia de bits $B_b[\ell]$ en una señal electromagnética $s(t)$.
- Procedimiento: agrupar bits en bloques de m bits (símbolos), asociar una forma de onda a cada uno de los M posibles valores de un bloque (símbolo), y transmitir en el intervalo asociado a cada símbolo la forma de onda que lo representa.
- Diseño: Elección de un conjunto apropiado de M señales para transportar cada uno de los M símbolos posibles (bloques de m bits).

- Demodulador digital

- Función: Estima de la secuencia de bits enviada, $\hat{B}_b[\ell]$ a partir de la señal recibida $r(t)$.
- Procedimiento: Decisión para cada intervalo de símbolo, de cuál de las M señales se transmitió en cada intervalo de símbolo a la vista de la señal recibida en dicho intervalo, comparando la forma de la señal en el intervalo con las M posibles señales y eligiendo una de ellas.
- Diseño: Determinar la medida óptima para realizar la comparación y la elección, de forma que se minimicen los errores cometidos teniendo en cuenta la distorsión que sufre la señal en su transmisión a través del canal.

A la hora de realizar el diseño del sistema, lo primero es tener claros los factores que habrá que considerar. En particular, a la hora de diseñar el modulador y el demodulador digital habrá que tener en cuenta principalmente tres factores:

1. Prestaciones: probabilidad de equivocarse en el receptor (P_e)

- Si en el receptor para cada intervalo se decide la señal más “parecida”, obviamente la probabilidad de error dependerá del “parecido” entre señales, por lo que habrá que buscar señales que sean lo más diferentes que sea posible.
- Otro problema es determinar la medida de “parecido” o “diferencia” que lleva a minimizar la probabilidad de error teniendo en cuenta el tipo de distorsión que sufre la señal en su transmisión.

2. Energía/potencia de la señal transmitida

- La potencia de la señal transmitida está limitada en la práctica, ya que los amplificadores utilizados en el transmisor tienen una potencia nominal que no se puede exceder por varios tipos de razones. Como la señal transmitida está compuesta por “trozos” en cada intervalo de símbolo, y en cada intervalo aparece una de las M formas de onda $s_i(t)$ que definen el modulador, en la práctica una medida razonable de la energía de la señal es el promedio de la energía de cada una de las señales que la componen, que se denominará energía media por símbolo transmitido (E_s), y que en la práctica estará acotada. Esta energía media por símbolo se obtiene ponderando la energía de cada símbolo (más propiamente, de la señal asociada a cada símbolo) teniendo en cuenta la probabilidad con la que se transmite dicho símbolo:

- Probabilidad de cada símbolo: $p_B(b_i) = P(B[n] = b_i)$
- Energía del símbolo $b_i \equiv$ energía de la señal $s_i(t)$
- Energía media por símbolo: promedio de la energía de los M símbolos

$$E_s = \sum_{i=0}^{M-1} p_B(b_i) \cdot \mathcal{E}\{s_i(t)\}, \text{ siendo } \mathcal{E}\{s_i(t)\} = \int_{-\infty}^{\infty} |s_i(t)|^2 dt.$$

3. Características del canal de comunicaciones ($h(t)$)

- A la hora de diseñar el sistema habrá que tratar de minimizar la distorsión que sufre la señal en la transmisión. Como se vió en la Sección 4.1.4, el efecto del canal habitualmente se modela a través de la relación

$$r(t) = s(t) * h(t) + n(t),$$

es decir, que la señal transmitida sufre fundamentalmente dos efectos: una distorsión lineal dada por la respuesta del canal (descrita por la respuesta al impulso $h(t)$ o respuesta en frecuencia $H(j\omega)$) y la adición de ruido (normalmente ruido térmico). El ruido siempre va a estar presente, y lo único que se puede hacer es filtrar a la entrada del receptor en el ancho de banda de la señal para minimizar su efecto sin distorsionar la señal transmitida. Pero la elección de las señales $s_i(t)$, que determinan las características de la señal transmitida, sí es relevante para minimizar la distorsión lineal; teniendo en cuenta la respuesta del canal, habrá señales que sufran una mayor distorsión, y señales que tengan una distorsión menor. La situación ideal sería que distorsión lineal introducida por el canal fuera nula, lo que teóricamente se podría conseguir si se elige un conjunto apropiado de M señales. Ese conjunto apropiado sería uno que cumpliera que

$$s_i(t) * h(t) = s_i(t) \text{ para } i \in \{0, 1, \dots, M-1\},$$

o lo que es lo mismo, tal vez sea más fácil de interpretar esta condición en el dominio frecuencial, si se cumple

$$S_i(j\omega) \times H(j\omega) = S_i(j\omega) \text{ para } i \in \{0, 1, \dots, M-1\}.$$

Si se logra el cumplimiento de esta condición para las M señales, la única distorsión que sufrirá la señal transmitida será la suma de ruido, lo que se denomina modulo de ruido aditivo blanco y gaussiano o modelo de canal gaussiano para abreviar

$$r(t) = s(t) + n(t).$$

Seleccionar M señales atendiendo de forma conjunta a estos tres factores trabajando con las señales en el dominio temporal es complicado, ya que algunos de los factores van incluso en oposición. Por ejemplo, es fácil aumentar la diferencia entre señales con señales de mayor amplitud, pero estas tienen mayor energía. Buscar M señales de una energía limitada, lo más diferentes posibles, y que al mismo tiempo tengan una transformada de Fourier adecuada a la respuesta en frecuencia del canal, es un problema complicado de resolver trabajando con la representación temporal de las señales. Además, está también el problema de determinar la medida de “*diferencia*” que lleva a tener la menor posibilidad de equivocarse en el demodulador a la hora de estimar la señal transmitida en cada intervalo de símbolo. Este problema puede tratarse de una forma más sencilla y eficiente manejando una representación geométrica de las señales a través de un espacio vectorial. Esta representación se presentará en la siguiente sección.

4.2. Representación geométrica de las señales

A la hora de diseñar el modulador, este debe asociar una forma de onda a cada símbolo del alfabeto de $B[n]$, lo que indirectamente supone que a cada b_i se le asocia una señal $s_i(t)$. Si el objetivo de diseño es tener la menor probabilidad de error, intuitivamente es lógico pensar que esas señales han de ser lo más diferentes que sea posible. Pero no es tan fácil buscar una medida de diferencia entre señales, cuando estas están definidas en el dominio temporal, que sea fácil de manejar analíticamente y que sea apropiada para tomar decisiones en un sistema de comunicaciones.

Esta es la razón para emplear una representación geométrica de las señales. Supongamos que cada señal se puede representar dentro de un espacio vectorial, es decir, que cada señal se puede representar por un vector, como se muestra en la Figura 4.9, en la que se representan en línea continua 4 señales en formato vectorial. La señal $s_i(t)$ estará representada por el vector \mathbf{a}_i (asociación basada en los subíndices). Por un lado, si se utiliza una medida de distancia entre vectores, es fácil determinar lo lejos o cerca que están los vectores asociados a dos señales. Y esa medida de distancia puede ser una buena medida del parecido entre señales. Por otro lado, el ruido será también representado como un vector, que se sumará a la señal, y tendremos otro vector resultante (en este caso el vector \mathbf{q} de la figura). Con una medida de distancia entre vectores es fácil decir de qué señal está más cerca el nuevo vector de la señal recibida (en este ejemplo, de la señal $s_0(t)$, al estar el vector \mathbf{q} más próximo al vector \mathbf{a}_0 que a los otros tres vectores).

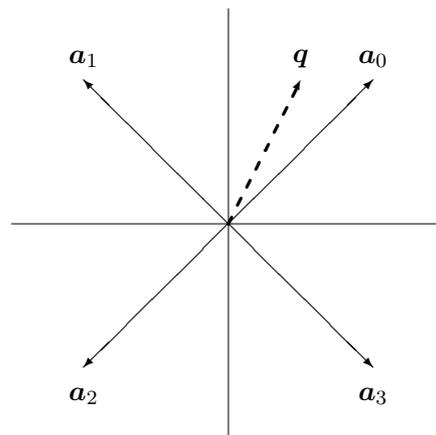


Figura 4.9: Representación geométrica de señales (ejemplo en un espacio bidimensional).

En esta sección se presentará una representación vectorial de las señales que resulta apropiada para el propósito que se persigue en el capítulo, que no es otro que el diseño y análisis de sistemas de comunicaciones digitales. En primer lugar se revisarán las principales características de los espacios vectoriales, y de los espacios de Hilbert, se verá cómo obtener una base ortonormal para la representación vectorial de un conjunto de M señales, y cómo calcular energías y medidas de distancia sobre las representaciones vectoriales de las señales.

4.2.1. Espacios vectoriales

Las señales admiten una representación como vectores de un espacio vectorial. Como vamos a ver, esta representación permite aplicar todas las herramientas de análisis y síntesis de este tipo

de espacios para trabajar con las señales.

Analicemos en primer lugar si una señal puede ser considerada como un vector dentro de un espacio vectorial. Un espacio vectorial \mathbb{V} es un conjunto de elementos, que denominamos vectores, que poseen las siguientes propiedades:

1. Existe una ley de composición interna, que se denomina suma y se representa con el símbolo $+$ que, aplicada a dos vectores $(\mathbf{x}, \mathbf{y}, \in \mathbb{V})$ de la forma $\mathbf{x} + \mathbf{y}$ da como resultado otro vector del espacio $(\mathbf{x} + \mathbf{y} \in \mathbb{V})$, cumpliendo además las siguientes propiedades
 - a) Conmutativa: $\forall \mathbf{x}, \mathbf{y} \in \mathbb{V}; \mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x}$.
 - b) Asociativa: $\forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{V}; \mathbf{x} + (\mathbf{y} + \mathbf{z}) = (\mathbf{x} + \mathbf{y}) + \mathbf{z}$.
 - c) Existencia de elemento neutro: $\exists \mathbf{0} \in \mathbb{V} : \forall \mathbf{x} \in \mathbb{V}; \mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x}$.
 - d) Existencia de elemento inverso: $\forall \mathbf{x} \in \mathbb{V} \exists (-\mathbf{x}) : \mathbf{x} + (-\mathbf{x}) = \mathbf{0}$.
2. Existe una ley de composición externa que denominamos producto con un conjunto \mathbb{C} de elementos denominados escalares (que deben tener la estructura de cuerpo) que, aplicada a un escalar α ($\alpha \in \mathbb{C}$) y a un vector \mathbf{x} ($\mathbf{x} \in \mathbb{V}$) de la forma $\alpha\mathbf{x}$ da como resultado otro vector del espacio ($\alpha\mathbf{x} \in \mathbb{V}$), cumpliendo además las siguientes propiedades:
 - a) Asociativa: $\forall \alpha, \beta \in \mathbb{C}; \forall \mathbf{x} \in \mathbb{V}; \alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x}$.
 - b) Existencia de elemento neutro: $\exists \mathbf{1} \in \mathbb{C} : \forall \mathbf{x} \in \mathbb{V}; \mathbf{1}\mathbf{x} = \mathbf{x}$.
 - c) Distributiva respecto a la suma: $\forall \alpha \in \mathbb{C}; \forall \mathbf{x}, \mathbf{y} \in \mathbb{V}; \alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y}$.
 - d) Distributiva respecto al producto por un escalar:

$$\forall \alpha, \beta \in \mathbb{C}; \forall \mathbf{x} \in \mathbb{V}; (\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x}.$$

Si se considera el caso general de una señal compleja (tanto en tiempo continuo como en tiempo discreto), esta señal cumple todas las condiciones de un espacio vectorial. La operación suma del espacio vectorial es la suma punto a punto de la señal, que es conmutativa, asociativa, tiene un elemento neutro ($x(t) = 0$) y un elemento inverso, la propia señal invertida (multiplicada por -1).

Si los escalares son el cuerpo de los números complejos, la ley de composición externa es la multiplicación de la señal por un complejo, y es fácil verificar que se cumplen todas las propiedades.

El caso de las señales reales y los número reales también lo cumplen, ya que de hecho no dejan de ser un caso particular de las señales y números complejos.

Sin embargo esta estructura de espacio vectorial genérico es muy simple como para ser de utilidad. Una estructura más elaborada la encontramos en los espacios vectoriales de Hilbert o simplemente espacios de Hilbert.

4.2.2. Espacios de Hilbert para señales de energía finita

Un espacio de Hilbert es, básicamente, un espacio vectorial con producto escalar¹. El producto escalar es una aplicación de pares de vectores en el campo de los escalares (complejos o reales en

¹Estrictamente, es un espacio vectorial con producto escalar que cumple la propiedad de completitud. La completitud se cumple cuando toda sucesión de Cauchy es convergente en la métrica inducida por el producto escalar. Si no posee esta propiedad el espacio vectorial recibe el nombre de espacio de pre-Hilbert.

nuestro caso),

$$f : (\mathbb{V}, \mathbb{V}) \rightarrow \mathbb{C},$$

que se denota como $\langle \mathbf{x}, \mathbf{y} \rangle$, y que cumple las siguientes propiedades:

1. $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^*$
2. $\langle \alpha \mathbf{x} + \beta \mathbf{y}, \mathbf{z} \rangle = \alpha \langle \mathbf{x}, \mathbf{z} \rangle + \beta \langle \mathbf{y}, \mathbf{z} \rangle$
3. $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$
4. $\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$ (vector elemento neutro)

A partir del producto escalar se define una norma para el espacio vectorial como

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle},$$

y a partir de la norma una medida de distancia entre vectores

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|.$$

El ángulo entre dos vectores se mide como

$$\theta = \cos^{-1} \left(\frac{\operatorname{Re}\{\langle \mathbf{x}, \mathbf{y} \rangle\}}{\|\mathbf{x}\| \|\mathbf{y}\|} \right).$$

Para las señales y, en general, para cualquier espacio vectorial genérico no existe una única posibilidad de definición de producto escalar: se puede escoger como producto escalar cualquier función que cumpla los requisitos anteriormente establecidos. Cada definición de producto escalar da lugar a un espacio de Hilbert distinto, con métrica y norma distinta. A continuación se presenta la estructura de dos espacios de Hilbert para señales de energía, que son los más apropiados para su aplicación en sistemas de comunicaciones

1. L_2 : espacio de Hilbert para señales de energía en tiempo continuo.
2. ℓ_2 : espacio de Hilbert para señales de energía en tiempo discreto.

Espacio de Hilbert para señales de energía en tiempo continuo

El espacio L_2 se define mediante el siguiente producto escalar

$$\langle \mathbf{x}, \mathbf{y} \rangle = \int_{-\infty}^{\infty} x(t) \cdot y^*(t) dt.$$

La norma inducida por este producto escalar es

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\int_{-\infty}^{\infty} |x(t)|^2 dt} = \sqrt{\mathcal{E}\{x(t)\}},$$

es decir, que es la raíz cuadrada de la energía de la señal. Esto es interesante porque la norma de un vector define la distancia de la representación vectorial al origen de coordenadas, por lo que

la energía de una señal se podrá evaluar de forma sencilla como la distancia al cuadrado de su representación vectorial con respecto al origen de coordenadas.

En cuanto a la distancia entre la representación vectorial de dos señales, viene dada por

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\int_{-\infty}^{\infty} |x(t) - y(t)|^2 dt} = \sqrt{\mathcal{E}\{x(t) - y(t)\}}.$$

En este caso se tiene la raíz cuadrada de la energía de la señal diferencia, que intuitivamente parece una medida cuantitativa razonable de la diferencia entre dos señales (posteriormente veremos que en ciertos casos es la medida óptima para decidir con la mínima probabilidad de equivocarse en un demodulador digital).

Espacio de Hilbert para señales de energía en tiempo discreto

El espacio ℓ_2 se define mediante el siguiente producto escalar

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=-\infty}^{\infty} x[n] \cdot y^*[n].$$

La norma inducida por este producto escalar es

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{\sum_{n=-\infty}^{\infty} |x[n]|^2} = \sqrt{\mathcal{E}\{x[n]\}},$$

y la distancia es la bien conocida distancia euclídea

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{\sum_{n=-\infty}^{\infty} |x[n] - y[n]|^2} = \sqrt{\mathcal{E}\{x[n] - y[n]\}}.$$

Como se puede ver, las definiciones son equivalentes en ambos espacios, con la única diferencia de la naturaleza continua o discreta del índice temporal (t o n , respectivamente).

Aspectos de interés

En cualquier caso, el producto escalar proporciona una medida del parecido o similitud entre dos señales. El producto escalar de dos señales cuya variación con el tiempo sea similar será “grande”, mientras que el de dos señales cuya variación con el tiempo sea distinta será “pequeño”.

Cuando el producto escalar de dos señales es igual a cero, se dice que esas señales son *ortogonales*. Esto significa que las correspondientes representaciones vectoriales tienen un ángulo de 90° entre ellas.

Hay una relación importante entre las normas de los vectores y el módulo del producto escalar. La desigualdad de Cauchy-Schwarz hace uso de esta relación y establece que

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \cdot \|\mathbf{y}\|.$$

La igualdad sólo se cumple si el vector \mathbf{y} es una versión escalada del vector \mathbf{x} , lo que en el caso de señales significa que $y(t) = \alpha \cdot x(t)$ o $y[n] = \alpha \cdot x[n]$. Esta desigualdad toma la forma siguiente para los espacios L_2 y ℓ_2 , respectivamente

$$\left| \int_{-\infty}^{\infty} x(t) \cdot y^*(t) dt \right| \leq \sqrt{\int_{-\infty}^{\infty} |x(t)|^2 dt} \cdot \sqrt{\int_{-\infty}^{\infty} |y(t)|^2 dt}.$$

$$\left| \sum_{n=-\infty}^{\infty} x[n] \cdot y^*[n] \right| \leq \sqrt{\sum_{n=-\infty}^{\infty} |x[n]|^2} \cdot \sqrt{\sum_{n=-\infty}^{\infty} |y[n]|^2}.$$

4.2.3. Representación de una señal en una base del espacio vectorial

El producto escalar también permite encontrar de forma sencilla la representación de una señal en una base del espacio vectorial.

Una base para un espacio de Hilbert \mathbb{H} de dimensión D es un subconjunto de D elementos $\{\mathbf{b}_n\} \in \mathbb{H}$, que determinan un conjunto de D coeficientes únicos $\{c_n(\mathbf{x})\}$, $n \in \{0, 1, \dots, D-1\}$, para cualquier elemento \mathbf{x} del espacio, de modo que el elemento se puede representar de forma unívoca como

$$\mathbf{x} = \sum_{n=0}^{D-1} c_n(\mathbf{x}) \cdot \mathbf{b}_n.$$

A los D coeficientes que definen unívocamente al vector \mathbf{x} , $\{c_n(\mathbf{x})\}_{n=0}^{D-1}$, se les suelen denominar coordenadas del vector en la base.

Se dice que una base es una *base ortogonal* cuando se cumple que los elementos de la base son ortogonales entre ellos, lo que significa que su producto escalar es nulo

$$\langle \mathbf{b}_n, \mathbf{b}_m \rangle = 0, \quad \forall n \neq m.$$

Se dice que una base es una *base ortonormal* cuando es una base ortogonal, es decir, se cumple que los elementos de la base son ortogonales, y además se cumple que cada uno de ellos tiene norma unidad, que teniendo en cuenta la definición de norma a través del producto escalar equivale en este caso a decir que el producto escalar de cada elemento de la base consigo mismo es uno

$$\|\mathbf{b}_n\| = 1 \rightarrow \langle \mathbf{b}_n, \mathbf{b}_n \rangle = 1.$$

En espacios de Hilbert, una de las ventajas de trabajar con una base ortonormal es que los coeficientes de la expansión en términos de la base (coordenadas) se obtienen mediante el producto escalar del vector con los distintos elementos que forman la base ortonormal

$$c_n(\mathbf{x}) = \langle \mathbf{x}, \mathbf{b}_n \rangle.$$

En el caso del espacio L_2 , un ejemplo de base ortonormal trivial lo constituye el conjunto de funciones delta de Dirac $\{\delta(t - \tau), \tau \in \mathbb{R}\}$. Si se denota $\mathbf{b}_i = \delta(t - \tau_i)$,

$$\langle \mathbf{b}_i, \mathbf{b}_j \rangle = \int_{-\infty}^{\infty} \delta(t - \tau_i) \delta(t - \tau_j) dt = \begin{cases} 1, & \tau_i = \tau_j \\ 0, & \tau_i \neq \tau_j \end{cases}.$$

La representación de la señal en términos de esta base es

$$\mathbf{x} \equiv x(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau.$$

Los coeficientes de la expansión se obtienen como

$$c_i(\mathbf{x}) = \langle \mathbf{x}, \delta(t - \tau_i) \rangle = \int_{-\infty}^{\infty} x(t) \cdot \delta(t - \tau_i) dt = x(\tau_i).$$

Lo mismo se aplica para el espacio ℓ_2 , donde el conjunto $\{\delta[n - k], k \in \mathbb{Z}\}$, forma una base ortonormal. En este caso se puede denotar $\mathbf{b}_k = \delta[n - k]$

$$\langle \mathbf{b}_k, \mathbf{b}_i \rangle = \sum_{n=-\infty}^{\infty} \delta[n - k] \delta[n - i] = \delta[k - i] = \begin{cases} 1, & k = i \\ 0, & k \neq i \end{cases}.$$

La representación de la señal en términos de esta base es

$$\mathbf{x} \equiv x[n] = \sum_{k=-\infty}^{\infty} x_k \delta[n - k].$$

Los coeficientes de la expansión se obtienen como

$$c_k(\mathbf{x}) = \langle \mathbf{x}, \mathbf{b}_k \rangle = \sum_{n=-\infty}^{\infty} x[n] \cdot \delta[n - k] = x[k].$$

4.2.4. Procedimiento de ortogonalización de Gram-Schmidt

Las dos bases ortonormales presentadas con anterioridad no son prácticas ya que involucran infinitos coeficientes, lo que supone en la práctica conocer la señal completa. Es posible obtener una base ortonormal de dimensión finita que permita representar un conjunto finito de elementos del espacio vectorial (en el caso particular que afecta al diseño y análisis de sistemas de comunicaciones, una base de dimensión finita que permita representar M señales). Una base adecuada para este tipo de representación se puede obtener mediante el denominado procedimiento de *ortogonalización de Gram-Schmidt*. Este procedimiento permite, a partir de M elementos del espacio vectorial que se quieren representar, obtener una base ortonormal de $N \leq M$ elementos a través de la que representar cada uno de los M elementos como una combinación lineal de los elementos de la base, donde los coeficientes (coordenadas) de la combinación lineal determinan la representación vectorial de cada elemento. De este modo es posible representar cada elemento como un punto, de dimensión N , en el espacio vectorial definido por esa base.

En el caso que nos interesa, los M elementos son señales de duración T segundos, y la correspondiente representación vectorial es una representación compacta del conjunto de señales a transmitir por un sistema de comunicaciones que simplifica el diseño y el análisis de su funcionamiento. A continuación se formaliza este problema y se particulariza el procedimiento de Gram-Schmidt para su aplicación en dicho problema.

En un sistema de comunicaciones se tiene un conjunto de M señales $\{s_i(t), i = 0, \dots, M - 1\}$, de duración T segundos y definidas en el intervalo $0 \leq t < T$, que se van a emplear para transmitir información a través del canal de comunicaciones. A partir de este conjunto se va a construir otro

conjunto de N señales ortonormales $\{\phi_j(t), j = 0, \dots, N-1\}$, con $N \leq M$, que permitirá obtener cada una de las señales $s_i(t)$ como una combinación lineal de las mismas.

$$s_i(t) = \sum_{j=0}^{N-1} a_{i,j} \cdot \phi_j(t).$$

Según la definición de una base ortonormal, cualquier señal $s_i(t)$ se puede expresar como combinación lineal de los elementos de la base donde los coeficientes de la expansión, o coordenadas de la representación vectorial para la señal $s_i(t)$, $a_{i,j}$, se obtienen a través del producto escalar

$$a_{i,j} = \langle s_i(t), \phi_j(t) \rangle = \int_{-\infty}^{\infty} s_i(t) \cdot \phi_j^*(t) \cdot dt.$$

Esto da lugar a una representación vectorial (N dimensional) de cada señal a través del vector de coordenadas

$$s_i(t) \rightarrow \mathbf{a}_i = \begin{bmatrix} a_{i,0} \\ a_{i,1} \\ \vdots \\ a_{i,N-1} \end{bmatrix}.$$

Aunque a simple vista puede parecer que este cambio de representación no supone un gran avance, hay que tener en cuenta que de este modo N es la dimensión del espacio de señales, y cada señal se representa con N coordenadas, lo que es más cómodo que manejar la expresión de la señal en tiempo continuo. Además, también se simplifica el cálculo de la energía de las señales $s_i(t)$ y de su diferencia. Es importante recordar que como se trata de una base ortonormal se cumplirá que el producto escalar de dos elementos distintos de la base será nulo, y que el producto escalar de un elemento de la base consigo mismo valdrá uno, es decir

$$\langle \phi_i(t), \phi_j(t) \rangle = \int_{-\infty}^{\infty} \phi_i(t) \cdot \phi_j^*(t) \cdot dt = \delta[i - j].$$

El producto escalar entre dos señales $s_i(t)$ y $s_k(t)$ se calcula como

$$\begin{aligned} \langle s_i(t), s_k(t) \rangle &= \int_{-\infty}^{\infty} s_i(t) \cdot s_k^*(t) dt \\ &= \int_{-\infty}^{\infty} \left(\sum_{j=0}^{N-1} a_{ij} \cdot \phi_j(t) \right) \cdot \left(\sum_{\ell=0}^{N-1} a_{k\ell}^* \cdot \phi_\ell^*(t) \right) dt \\ &= \int_{-\infty}^{\infty} \left(\sum_{j=0}^{N-1} \sum_{\ell=0}^{N-1} a_{ij} \cdot a_{k\ell}^* \cdot \phi_j(t) \cdot \phi_\ell^*(t) \right) dt \\ &= \sum_{j=0}^{N-1} a_{ij} \cdot a_{kj}^* \int_{-\infty}^{\infty} \phi_j^2(t) dt + \sum_{j=0}^{N-1} \sum_{\substack{\ell=0 \\ \ell \neq j}}^{N-1} a_{ij} a_{k\ell}^* \int_{-\infty}^{\infty} \phi_j(t) \phi_\ell^*(t) dt, \end{aligned}$$

que aplicando la propiedad de ortonormalidad de la base se reduce a

$$\langle \mathbf{s}_i, \mathbf{s}_k \rangle = \langle s_i(t), s_k(t) \rangle = \sum_{j=0}^{N-1} a_{ij} \cdot a_{kj}^*.$$

La energía de una señal, que en el dominio temporal involucra una integral

$$\mathcal{E}_i = \mathcal{E} \{s_i(t)\} = \int_{-\infty}^{\infty} |s_i(t)|^2 dt$$

se puede calcular por tanto a partir de la representación vectorial de la señal de una forma más sencilla

$$\mathcal{E}_i = \mathcal{E} \{s_i(t)\} = \langle \mathbf{s}_i, \mathbf{s}_i \rangle = \|s_i(t)\|^2 = \sum_{j=0}^{N-1} |a_{ij}|^2.$$

El cálculo a partir de integrales de las señales en el dominio temporal ha sido reemplazado por una suma de N coordenadas al cuadrado, mucho más fácil de realizar.

Lo mismo se puede aplicar al cálculo de la diferencia entre señales. Antes se dijo que una medida intuitivamente razonable (luego veremos además que en muchos casos es la medida óptima) es la energía de la señal diferencia, que en el dominio temporal de nuevo requiere de una integral

$$\int_{-\infty}^{\infty} |s_i(t) - s_k(t)|^2 dt.$$

Con la definición de producto escalar que se ha utilizado, la raíz cuadrada de esta cantidad corresponde a la norma del vector diferencia, es decir, a la distancia entre la representación vectorial de las señales. Resumiendo, la distancia entre señales se puede obtener a partir de la definición de norma como la norma del vector diferencia (de la diferencia de señales)

$$d(s_i(t), s_k(t)) = \|s_i(t) - s_k(t)\| = \sqrt{\int_{-\infty}^{\infty} |s_i(t) - s_k(t)|^2 dt}.$$

Teniendo en cuenta la expansión lineal que tienen ambas señales, las coordenadas de la señal diferencia son la diferencia de coordenadas de cada una de las señales, y por tanto, la norma de la misma se obtiene como la raíz cuadrada del producto escalar de la señal diferencia consigo misma

$$d(s_i(t), s_k(t)) = \sqrt{\langle \mathbf{s}_i - \mathbf{s}_k, \mathbf{s}_i - \mathbf{s}_k \rangle} = \sqrt{\sum_{j=0}^{N-1} |a_{ij} - a_{kj}|^2}.$$

De nuevo se han sustituido integrales sobre señales por sumas sobre N coordenadas. Por eso resulta conveniente poder representar las señales como vectores obtenidos a partir de combinaciones lineales de un conjunto de señales que forman una base ortonormal.

Para la obtención de una base ortonormal que permita representar M señales, el proceso de ortogonalización de Gram-Schmidt comienza seleccionando una forma de onda de energía no nula. Si $s_0(t)$ es de energía nula, se ha de reordenar el conjunto de señales para tener una nueva señal $s_0(t)$ de energía $\mathcal{E}_0 > 0$. El primer elemento de la base se calcula simplemente normalizando la primera señal

$$\phi_0(t) = \frac{s_0(t)}{\sqrt{\mathcal{E}_0}}.$$

Esto asegura que la energía de $\phi_0(t)$, $\mathcal{E}\{\phi_0(t)\} = 1$, o lo que es lo mismo, que tenemos el primer elemento de una base ortonormal.

Para obtener el segundo elemento de la base, primero se calcula la proyección de $s_1(t)$ sobre $\phi_0(t)$, es decir, la coordenada de dicha señal sobre el primer elemento de la base, que es

$$a_{1,0} = \int_{-\infty}^{\infty} s_1(t) \cdot \phi_0^*(t) dt.$$

Entonces se sustrae esta proyección de la segunda función

$$d_1(t) = s_1(t) - a_{1,0}\phi_0(t).$$

Ahora $d_1(t)$ es ortogonal a $\phi_0(t)$, pero no tiene energía unidad. Si \mathcal{E}_1 denota la energía de $d_1(t)$, es decir

$$\mathcal{E}_1 = \mathcal{E}\{d_1(t)\} = \int_{-\infty}^{\infty} |d_1(t)|^2 dt,$$

el segundo elemento de la base ortonormal es

$$\phi_1(t) = \frac{d_1(t)}{\sqrt{\mathcal{E}_1}},$$

que también tiene energía unidad.

En general, la obtención del elemento $k + 1$ de la base, $\phi_k(t)$ se obtiene como

$$\phi_k(t) = \frac{d_k(t)}{\sqrt{\mathcal{E}_k}},$$

donde

$$d_k(t) = s_k(t) - \sum_{j=0}^{k-1} a_{k,j} \cdot \phi_j(t),$$

$$\mathcal{E}_k = \mathcal{E}\{d_k(t)\} = \int_{-\infty}^{\infty} |d_k(t)|^2 dt,$$

y

$$a_{k,j} = \int_{-\infty}^{\infty} s_k(t) \cdot \phi_j^*(t) dt.$$

A continuación se presenta un ejemplo para el cálculo de una base que permita representar 4 señales.

Ejemplo

Se aplica el procedimiento de Gram-Schmidt al conjunto de señales de la Figura 4.10.

Para obtener el primer elemento de la base calculamos la energía de $s_0(t)$. Es fácil comprobar que $\mathcal{E}_0 = 2$. Por tanto

$$\phi_0(t) = \frac{s_0(t)}{\sqrt{2}}.$$

Luego se calcula la proyección de $s_1(t)$ sobre $\phi_0(t)$.

$$a_{1,0} = \int_{-\infty}^{\infty} s_1(t) \cdot \phi_0^*(t) dt = 0.$$

Por tanto, en este caso $d_1(t) = s_1(t)$, que es ortogonal a $\phi_0(t)$ pero no es ortonormal. Hay que dividir por la energía

$$\mathcal{E}_1 = 2.$$

$$\phi_1(t) = \frac{s_1(t)}{\sqrt{2}}.$$

Para calcular $\phi_2(t)$ hay que calcular las proyecciones de $s_2(t)$ sobre $\phi_0(t)$ y $\phi_1(t)$.

$$a_{2,0} = \int_{-\infty}^{\infty} s_2(t) \cdot \phi_0^*(t) dt = 0.$$

$$a_{2,1} = \int_{-\infty}^{\infty} s_2(t) \cdot \phi_1^*(t) dt = -\sqrt{2}.$$

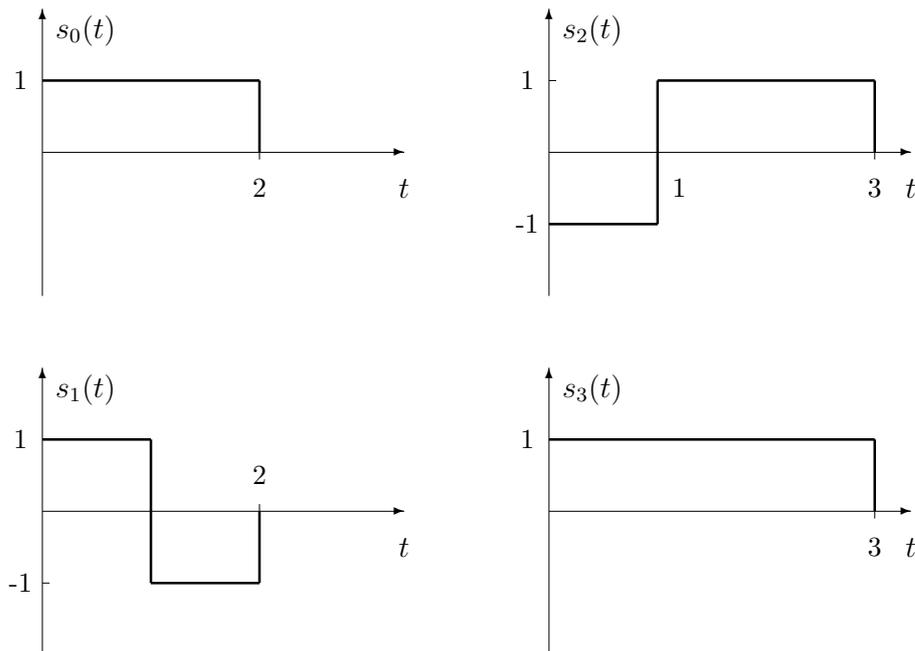


Figura 4.10: Conjunto original de señales.

Por tanto

$$\begin{aligned} d_2(t) &= s_2(t) - a_{20} \cdot \phi_0(t) - a_{21} \cdot \phi_1(t) \\ &= s_2(t) + \sqrt{2}\phi_1(t). \end{aligned}$$

La energía de $d_2(t)$ es

$$\mathcal{E}_2 = 1.$$

$$\phi_2(t) = d_2(t) = s_2(t) + \sqrt{2}\phi_1(t).$$

Ahora aplicamos el procedimiento sobre la última de las señales, $s_3(t)$. En primer lugar vamos a calcular las proyecciones de esta señal sobre los 3 elementos de la base ya disponibles:

$$a_{3,0} = \int_{-\infty}^{\infty} s_3(t) \cdot \phi_0^*(t) dt = \sqrt{2}.$$

$$a_{3,1} = \int_{-\infty}^{\infty} s_3(t) \cdot \phi_1^*(t) dt = 0.$$

$$a_{3,2} = \int_{-\infty}^{\infty} s_3(t) \cdot \phi_2^*(t) dt = 1.$$

Por tanto, ahora

$$d_3(t) = s_3(t) - a_{3,0} \cdot \phi_0(t) - a_{3,1} \cdot \phi_1(t) - a_{3,2} \cdot \phi_2(t) = 0.$$

Esto quiere decir que no es necesario ningún elemento adicional en la base para poder representar $s_3(t)$, ya que esta puede obtenerse como una combinación lineal de los 3 elementos de la base ya obtenidos anteriormente. Así pues, en este caso, la dimensión de este espacio de señales es 3, y las 3 funciones base se muestran en la Figura 4.11.

En resumen, una vez calculadas las N señales de la base ortonormal

$$\{\phi_j(t), j = 0, \dots, N - 1\}$$

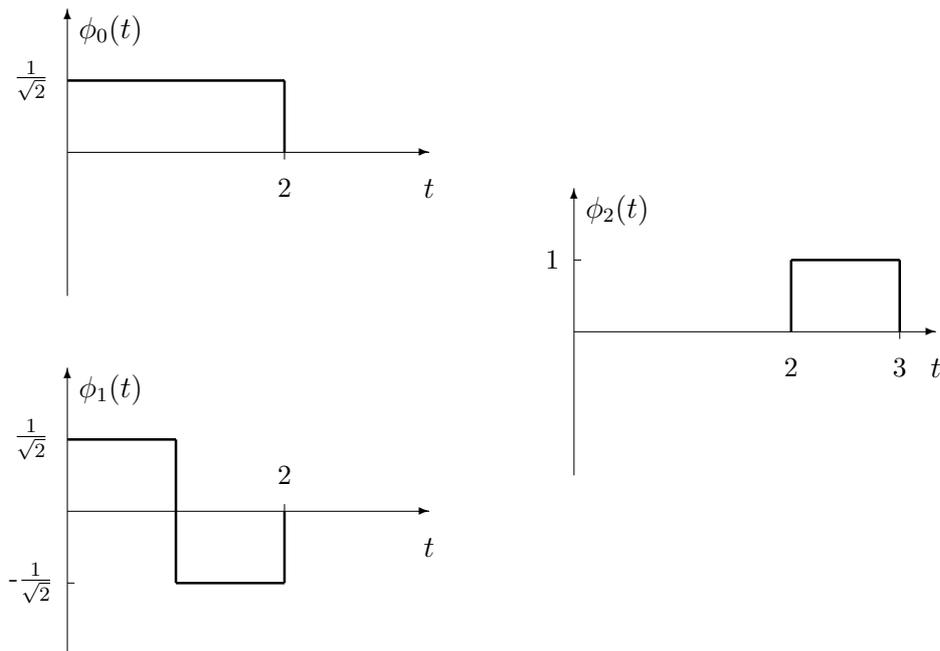


Figura 4.11: Conjunto de señales ortonormales.

cada señal se puede obtener como una combinación lineal de los N elementos de la base

$$s_i(t) = \sum_{j=0}^{N-1} a_{ij} \phi_j(t), \quad i = 0, \dots, M-1.$$

Por tanto, cada señal puede ser representada mediante un vector

$$\mathbf{a}_i = \begin{bmatrix} a_{i,0} \\ a_{i,1} \\ \vdots \\ a_{i,N-1} \end{bmatrix}.$$

o equivalentemente, como un punto en un espacio de señales N -dimensional con coordenadas $(a_{i,0}, a_{i,1}, \dots, a_{i,N-1})$.

Por otro lado, la potencia de la señal $s_i(t)$, \mathcal{E}_i , se puede calcular de forma sencilla a partir de dichas coordenadas

$$\mathcal{E}_i = \int_{-\infty}^{\infty} |s_i(t)|^2 dt,$$

$$\mathcal{E}_i = \langle \mathbf{s}_i, \mathbf{s}_i \rangle = \sum_{j=0}^{N-1} |a_{i,j}|^2.$$

Ejemplo

Las coordenadas de las 4 señales del ejemplo anterior (que se han ido calculando durante el desarrollo del procedimiento de ortogonalización) son

$$\left. \begin{array}{l} a_{0,0} = \sqrt{2} \\ a_{0,1} = 0 \\ a_{0,2} = 0 \end{array} \right\} \rightarrow \mathbf{a}_0 = [\sqrt{2}, 0, 0]^T.$$

$$\left. \begin{aligned} a_{1,0} &= 0 \\ a_{1,1} &= \sqrt{2} \\ a_{1,2} &= 0 \end{aligned} \right\} \rightarrow \mathbf{a}_1 = [0, \sqrt{2}, 0]^T.$$

$$\left. \begin{aligned} a_{2,0} &= 0 \\ a_{2,1} &= -\sqrt{2} \\ a_{2,2} &= 1 \end{aligned} \right\} \rightarrow \mathbf{a}_2 = [0, -\sqrt{2}, 1]^T.$$

$$\left. \begin{aligned} a_{3,0} &= \sqrt{2} \\ a_{3,1} &= 0 \\ a_{3,2} &= 1 \end{aligned} \right\} \rightarrow \mathbf{a}_3 = [\sqrt{2}, 0, 1]^T.$$

Estas señales se pueden representar como puntos en un espacio 3D, tal y como se muestra en la Figura 4.12

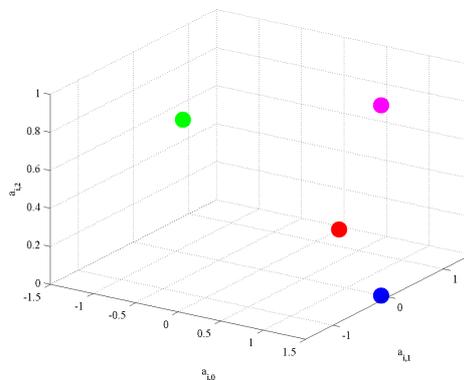


Figura 4.12: Señales representadas en la base ortonormal.

Es fácil comprobar que

$$\mathcal{E}_0 = 2, \mathcal{E}_1 = 2, \mathcal{E}_2 = 3, \mathcal{E}_3 = 3.$$

También se pueden calcular de forma sencilla las distancias entre señales, que en este caso son

$$d(\mathbf{s}_0, \mathbf{s}_1) = 2, \quad d(\mathbf{s}_0, \mathbf{s}_2) = \sqrt{5}, \quad d(\mathbf{s}_0, \mathbf{s}_3) = 1$$

$$d(\mathbf{s}_1, \mathbf{s}_2) = \sqrt{9}, \quad d(\mathbf{s}_1, \mathbf{s}_3) = \sqrt{5}, \quad d(\mathbf{s}_2, \mathbf{s}_3) = 2$$

Es importante hacer notar que la base para representar un conjunto de M señales no es única. Si se realiza el procedimiento de Gram-Schmidt cambiando el orden de las señales, la base obtenida, y las correspondientes coordenadas son otras, pero las energías y distancias entre señales se mantienen. Es decir, el resultado, si se realiza una representación de las señales, es que los nuevos puntos estarán rotados.

También es posible ver para un conjunto de señales una base adecuada para su representación. Por ejemplo, para estas señales se podrían utilizar como bases las siguientes señales:

$$\phi_0(t) = \begin{cases} 1, & \text{si } 0 \leq t < 1 \\ 0, & \text{en otro caso} \end{cases}.$$

$$\phi_1(t) = \begin{cases} 1, & \text{si } 1 \leq t < 2 \\ 0, & \text{en otro caso} \end{cases}.$$

$$\phi_2(t) = \begin{cases} 1, & \text{si } 2 \leq t < 3 \\ 0, & \text{en otro caso} \end{cases}$$

En este caso, es trivial ver que

$$\mathbf{a}'_0 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix} \quad \mathbf{a}'_1 = \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix} \quad \mathbf{a}'_2 = \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} \quad \mathbf{a}'_3 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$$

Es fácil comprobar como se mantienen energías (que se relacionan con la distancia de cada punto al origen) y distancias entre las representaciones vectoriales utilizando esta nueva representación.

Por tanto, la representación vectorial correspondiente supone una rotación de la representación utilizando la otra base ortonormal, tal y como se puede ver a en en la Figura 4.13, que muestra la representación vectorial de las 4 señales en la nueva base ortonormal.

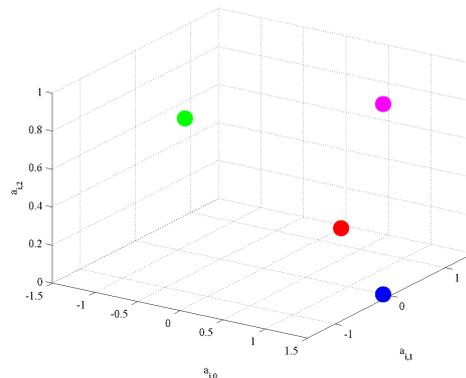


Figura 4.13: Señales representadas en la base ortonormal alternativa.

4.3. Modelo de comunicación digital

La representación vectorial de las señales a través de una base ortonormal de dimensión N permite utilizar un modelo de comunicación digital basado en esta representación que simplifica notablemente el diseño y análisis del sistema. El esquema de este modelo consta de 4 elementos funcionales (además del canal de comunicaciones), y se muestra en la Figura 4.14.

En este modelo se dividen tanto el modulador como el demodulador digital en dos módulos:

- División del modulador digital en dos módulos
 - Codificador
 - Modulador
- División del demodulador digital en dos módulos
 - Demodulador
 - Decisor

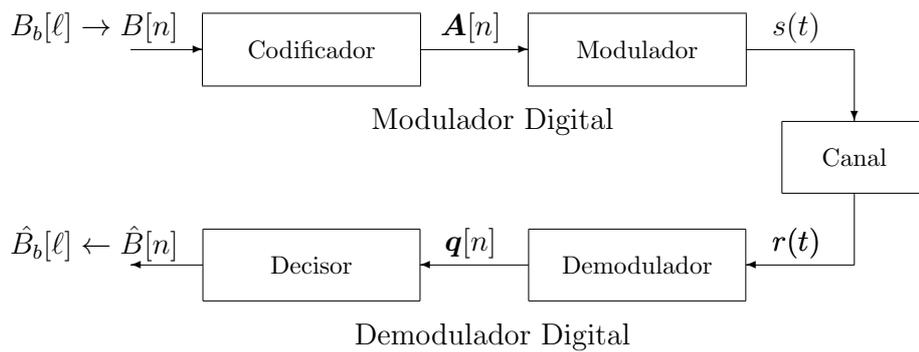


Figura 4.14: Modelo básico de un sistema de comunicaciones digitales.

Hay que recordar que la tarea del modulador y demodulador digital era convertir bits (agrupados en símbolos) en señales y hacer la conversión inversa, respectivamente. Con esta división en módulos se ha introducido en ambos casos una representación vectorial de las señales intermedia, $\mathbf{A}[n]$ en el transmisor y $\mathbf{q}[n]$ en el receptor. Van a ser precisamente estas representaciones vectoriales intermedias las que faciliten notablemente el diseño y análisis del sistema.

En el transmisor, ahora la asignación de una señal $s_i(t)$ a cada valor del alfabeto de símbolos b_i , se realiza ahora en dos pasos:

- Se hace una asignación de la representación vectorial de $s_i(t)$, \mathbf{a}_i , para b_i (codificador).
- Se convierte la representación vectorial de la señal, \mathbf{a}_i , a señal $s_i(t)$ (modulador). Para ello hay que definir una base ortonormal de dimensión N , ya que

$$s_i(t) = \sum_{j=0}^{N-1} a_{i,j} \times \phi_j(t).$$

Esta división simplifica notablemente el problema de la elección de las M señales para tener en cuenta los 3 factores que se especificaron en la Sección 4.1.6: prestaciones (parecido/diferencia entre señales), energía y características del canal. Esto es así porque los dos primeros factores se desacoplan del tercero. Como se ha visto, la energía y el parecido entre señales se pueden evaluar a partir de la representación vectorial de las señales, independientemente de la base ortonormal elegida. De este modo, el diseño del codificador, que consiste en elegir la representación vectoriales de las M señales,

$$\{\mathbf{a}_0, \mathbf{a}_1, \dots, \mathbf{a}_{M-1}\},$$

que se denominará habitualmente constelación, se realizará teniendo en cuenta estos dos factores: prestaciones (medida de diferencia entre las señales) y energía.

En cuanto a las características del canal, hay que hacer que las señales tengan una respuesta en frecuencia adecuada a la respuesta en frecuencia del canal. La forma de la respuesta en frecuencia de la señal $s_i(t)$, dado que

$$s_i(t) = \sum_{j=0}^{N-1} a_{i,j} \times \phi_j(t) \leftrightarrow S_i(j\omega) = \sum_{j=0}^{N-1} a_{i,j} \times \Phi_j(j\omega),$$

depende de la elección de la base ortonormal. Elegir una base ortonormal u otra no tendrá ninguna repercusión en la energía y medida de diferencia (distancia) entre señales, ya que estos parámetros

quedan completamente definidos por la representación vectorial de las señales. Lo que se modificará será la forma de las señales, tanto en el dominio temporal como frecuencial. De este modo, el diseño del modulador, que consistirá en la elección de una base ortonormal apropiada, se realizará teniendo en cuenta sólo el tercer factor, es decir, para que las señales resulten adecuadas a las características del canal.

En cuanto al receptor, la división en dos módulos también simplifica el procedimiento. En primer lugar, el demodulador obtiene la representación vectorial de la señal recibida en cada intervalo de símbolo, $\mathbf{q}[n]$ para el intervalo $nT \leq t < T$. Y a partir de esa representación, será mucho más sencillo buscar la regla óptima para decidir cuál de las M señales se transmitió en ese intervalo (la regla que produzca menos errores) que trabajando sobre la representación en tiempo continuo de las señales.

A continuación se resume brevemente la función y principales características de cada uno de los 4 elementos funcionales del sistema:

■ Codificador

- Define la representación vectorial de la señal asociada a cada símbolo (constelación)
 - Intervalo de índice n : vector $\mathbf{A}[n]$ representando $s(t)$ en $nT \leq t < (n+1)T$
- Criterios de diseño (para seleccionar la constelación)
 - Energía
 - Distancia (“*parecido*”) entre señales (prestaciones)

■ Modulador

- Define la base ortonormal del espacio de señales
- Criterios de diseño (para seleccionar las N funciones $\phi_j(t)$, para $i = 0, 1, \dots, N-1$)
 - Adecuación al canal de comunicaciones

■ Demodulador

- Convierte la señal recibida, por intervalos de símbolo, en vectores en el espacio de señales definido por la base $\{\phi_j(t)\}_{j=0}^{N-1}$
 - Intervalo de índice n : vector $\mathbf{q}[n]$ representando $r(t)$ en $nT \leq t < (n+1)T$

■ Decisor

- Compara el “*parecido*” entre la señal recibida y las M posibles señales $s_i(t)$ para decidir símbolos
 - Medida de distancia sobre las representaciones vectoriales
 - Compara las distancias de:
 - Vector de la señal recibida en el intervalo de símbolo: $\mathbf{q}[n]$
 - Vectores de los M posibles símbolos: \mathbf{a}_i , para $i \in \{0, 1, \dots, M-1\}$

4.3.1. Ejemplo para ilustrar la ventaja de la representación vectorial en el diseño de un sistema

Empezaremos considerando un sistema de comunicación binaria en el que cada símbolo corresponde a 1 bit ($m = 1$, $M = 2^m = 2$) y los dos bits se transmiten con la misma probabilidad. Como señales para la transmisión de cada símbolo, se puede escoger de entre cuatro posibles conjuntos de señales $\{s_0(t), s_1(t)\}$ para realizar la transmisión. Estos conjuntos se representan en las Figuras 4.15, 4.16, 4.17 y 4.18 respectivamente.

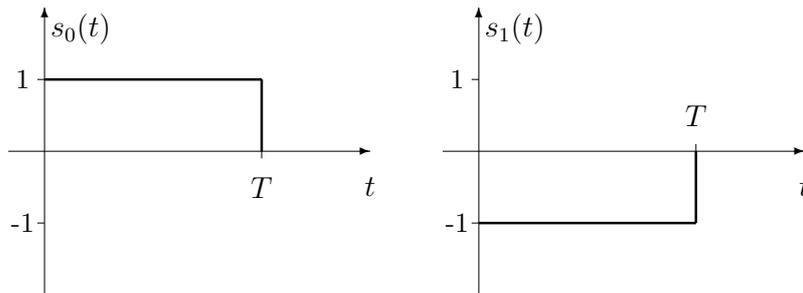


Figura 4.15: Primer conjunto de señales.

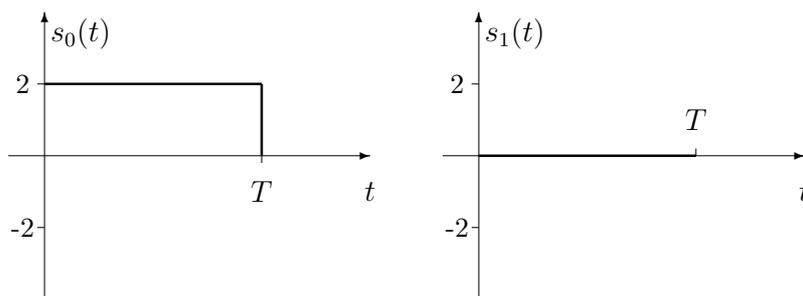


Figura 4.16: Segundo conjunto de señales.

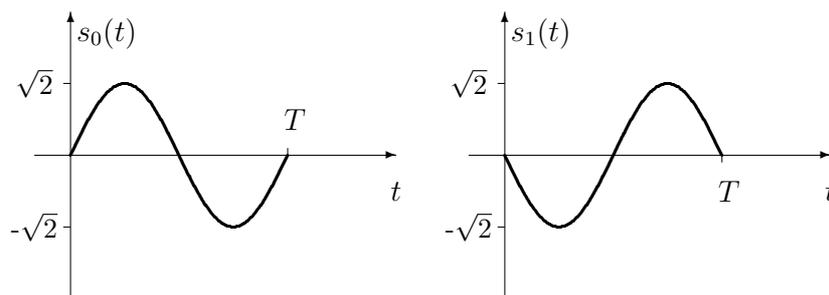


Figura 4.17: Tercer conjunto de señales

Para decidir qué conjunto de señales es el más apropiado hay que recordar que el objetivo es poder recuperar el símbolo enviado con la menor probabilidad de error teniendo en cuenta las características del canal; en este caso se suma ruido blanco gaussiano. Parece lógico pensar que si mayor es la “separación” o “diferencia” entre $s_0(t)$ y $s_1(t)$ más fácil será discernir entre ambas en el receptor pese al ruido sumado. Para medir esta separación se necesita una medida de distancia entre señales.

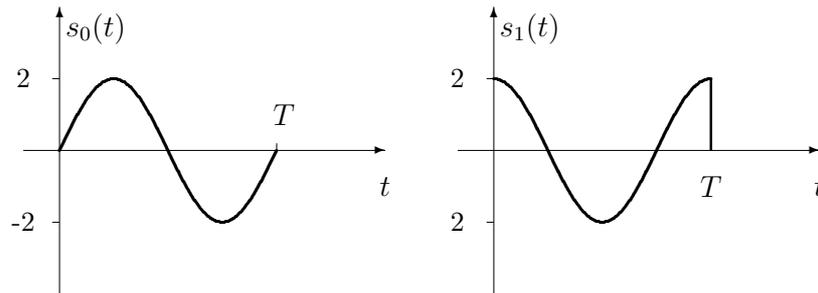


Figura 4.18: Cuarto conjunto de señales

Se va a utilizar la medida de distancia definida en el espacio de Hilbert de señales de energía finita, L_2 . En este espacio, las señales $s_i(t)$ pasan a ser vectores \mathbf{a}_i y la distancia se define como

$$d(s_i(t), s_j(t)) = \|\mathbf{a}_i - \mathbf{a}_j\| = \sqrt{\int_{-\infty}^{\infty} |s_i(t) - s_j(t)|^2 dt}.$$

Aplicando esta medida a las dos señales del primer conjunto

$$d(s_0(t), s_1(t)) = \sqrt{\int_0^T |1 - (-1)|^2 dt} = 2\sqrt{T}.$$

Para el segundo conjunto de señales tenemos

$$d(s_0(t), s_1(t)) = \sqrt{\int_0^T |2 - 0|^2 dt} = 2\sqrt{T}.$$

Para el tercer conjunto de señales

$$\begin{aligned} d(s_0(t), s_1(t)) &= \sqrt{\int_0^T \left| \sqrt{2} \sin\left(\frac{2\pi t}{T}\right) - \left(-\sqrt{2} \sin\left(\frac{2\pi t}{T}\right)\right) \right|^2 dt} \\ &= \sqrt{\int_0^T 8 \sin^2\left(\frac{2\pi t}{T}\right) dt} \\ &= \sqrt{4 \left[t - \frac{T}{2\pi} \sin\left(\frac{2\pi t}{T}\right) \cos\left(\frac{2\pi t}{T}\right) \right]_0^T} \\ &= 2\sqrt{T}. \end{aligned}$$

Y para el último conjunto de señales

$$\begin{aligned} d(s_0(t), s_1(t)) &= \sqrt{\int_0^T \left| 2 \sin\left(\frac{2\pi t}{T}\right) - \left(2 \cos\left(\frac{2\pi t}{T}\right)\right) \right|^2 dt} \\ &= \sqrt{\int_0^T 4 - 8 \sin\left(\frac{2\pi t}{T}\right) \cos\left(\frac{2\pi t}{T}\right) dt} \\ &= 2\sqrt{T}, \end{aligned}$$

ya que

$$\int_0^T 8 \sin\left(\frac{2\pi t}{T}\right) \cos\left(\frac{2\pi t}{T}\right) dt = \left[\frac{2T}{\pi} \sin^2\left(\frac{2\pi t}{T}\right) \right]_0^T = 0.$$

Así pues, los cuatro conjuntos tienen la misma separación y en principio las señales se comportarían igual ante la perturbación que introduce el canal. Pero también debemos preguntarnos si le suponen el mismo esfuerzo al transmisor enviarlos a través del canal. Y la forma de medir el esfuerzo es la *energía media por símbolo* que se define como

$$\begin{aligned} E_s &= E[\mathcal{E}\{s(t)\}] \\ &= E\left[\int_{-\infty}^{\infty} |s(t)|^2 dt\right] \\ &= \sum_{i=0}^{M-1} P(s(t) = s_i(t)) \times \mathcal{E}\{s_i(t)\} \\ &= \sum_{i=0}^{M-1} p_A(\mathbf{a}_i) \int_{-\infty}^{\infty} |s_i(t)|^2 dt. \end{aligned}$$

Si los símbolos son equiprobables, $p_A(\mathbf{a}_i) = \frac{1}{M}$, por lo que

$$E_s = \sum_{i=0}^{M-1} \frac{1}{M} \int_{-\infty}^{\infty} |s_i(t)|^2 dt.$$

Aplicando esta definición al primer conjunto de señales

$$E_s = \frac{1}{2} \int_0^T |1|^2 dt + \frac{1}{2} \int_0^T |-1|^2 dt = \frac{1}{2}T + \frac{1}{2}T = T.$$

Si se calcula esta energía para el resto de conjuntos se obtiene, respectivamente, $E_s = 2T$, $E_s = T$ y $E_s = 2T$. Se puede decir por tanto que los mejores conjuntos son el primero y el tercero, pues tienen la misma distancia con la menor energía.

Cuando se manejan conjuntos de señales con diferente distancia entre señales, es necesario emplear algún tipo de normalización que nos permita establecer su comparación. La más común de estas normalizaciones consiste en expresar la distancia entre señales en función de la energía media por símbolo. Sobre el primer conjunto de símbolos, esta distancia normalizada es

$$d(s_0(t), s_1(t)) = 2\sqrt{T} = 2\sqrt{E_s}.$$

Para el resto de conjuntos, respectivamente

$$d(s_0(t), s_1(t)) = 2\sqrt{T} = \sqrt{2E_s},$$

$$d(s_0(t), s_1(t)) = 2\sqrt{T} = 2\sqrt{E_s},$$

y

$$d(s_0(t), s_1(t)) = 2\sqrt{T} = \sqrt{2E_s}.$$

Se observa que los conjuntos de señales 1 y 3 proporcionan una mayor distancia de separación entre señales que los conjuntos 2 y 4 para un mismo valor de E_s .

Otra cuestión a plantearse es si es posible encontrar una forma de representar las señales que nos permita ver de manera más simple las prestaciones que vamos a obtener con un conjunto determinado. La respuesta la encontramos en la misma estructura de espacio vectorial que nos ha permitido medir la distancia entre señales: si somos capaces de encontrar una base ortonormal que nos permita representar un determinado conjunto de señales podremos trabajar directamente con las coordenadas en esa base y evitar cualquier tipo de cálculo sobre las formas de onda.

Expresado formalmente, se busca una representación para cada una de las señales del conjunto $\{s_i(t) : i = 0, \dots, M - 1\}$ de la forma

$$s_i(t) = \sum_{j=0}^{N-1} a_{i,j} \phi_j(t),$$

donde $\{\phi_j(t) : i = 0, \dots, N - 1\}$ son los elementos de la base ortonormal, y $\{a_{i,j} : j = 0, \dots, N - 1\}$ son las coordenadas de la señal $s_i(t)$ en dicha base, que se agrupan en el vector \mathbf{a}_i . Así pues, los símbolos \mathbf{a}_i corresponden a las coordenadas de la señal de índice i .

Si se obtiene la base para el primer conjunto de señales, se puede comprobar que

$$s_0(t) = \sqrt{T} \phi_0(t), \quad s_1(t) = -\sqrt{T} \phi_0(t),$$

donde $\phi_0(t) = s_0(t)/\sqrt{T}$, que se representa en la Figura 4.19.

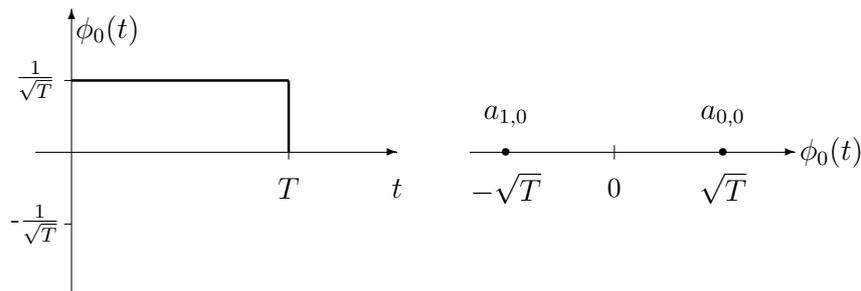


Figura 4.19: Ortogonalización para el primer conjunto de señales.

El hecho de obtener un único elemento de la base, $\phi_0(t)$, tiene una justificación muy sencilla: como $s_1(t) = -s_0(t)$, un cambio de signo en la coordenada de $s_0(t)$ nos basta para generar $s_1(t)$. La amplitud de $\phi_0(t)$ se toma para cumplir la condición de ortonormalidad, que no es otra cosa que la normalización de la energía ($\int_{-\infty}^{\infty} \phi_0^2(t) dt = 1$). Ahora es fácil calcular la distancia entre las señales y su energía

$$d(s_0(t), s_1(t)) = d(\mathbf{a}_0, \mathbf{a}_1) = \sqrt{(a_{0,0} - a_{1,0})^2} = \sqrt{(\sqrt{T} - (-\sqrt{T}))^2} = 2\sqrt{T}.$$

La energía por símbolo es

$$E_s = E[\mathcal{E}\{s(t)\}] = E[\mathcal{E}\{\mathbf{a}_i\}] = p_A(\mathbf{a}_0) a_{0,0}^2 + p_A(\mathbf{a}_1) a_{1,0}^2 = \frac{1}{2}T + \frac{1}{2}T = T.$$

El resultado, como no podía ser de otra forma, es igual al que hemos obtenido calculando sobre las formas de onda.

Para el segundo conjunto de señales se obtiene la base y coordenadas mostradas en la Figura 4.20.

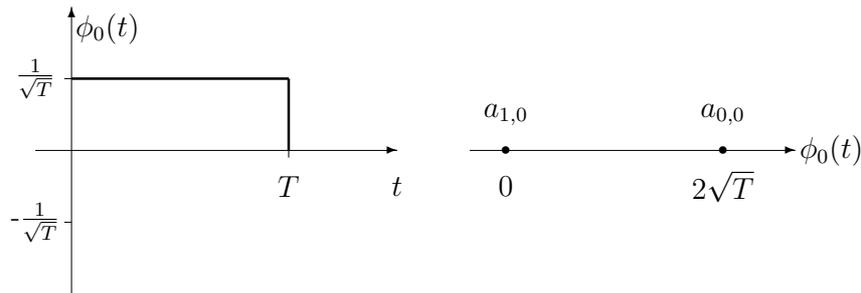


Figura 4.20: Ortogonalización para el segundo conjunto de señales.

Si se compara con el resultado obtenido para el primer conjunto, se observa que el elemento único de la base es la misma señal y que las coordenadas simplemente se han desplazado un valor $\sqrt{2}$. Este desplazamiento hace que se mantenga la distancia entre señales, de modo que frente al ruido se comportaría igual que el conjunto primero. Sin embargo, aumenta la energía media de la constelación

$$E_s = E[\mathcal{E}\{\mathbf{a}_i\}] = p_A(\mathbf{a}_0)a_{0,0}^2 + p_A(\mathbf{a}_1)a_{1,0}^2 = \frac{1}{2}4T = 2T.$$

La ortogonalización del tercer conjunto se muestra en la Figura 4.21.

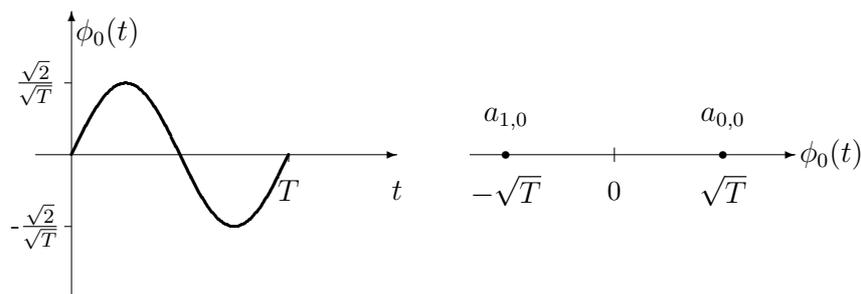


Figura 4.21: Ortogonalización para el tercer conjunto de señales.

La ortogonalización del cuarto conjunto se muestra en la Figura 4.22.

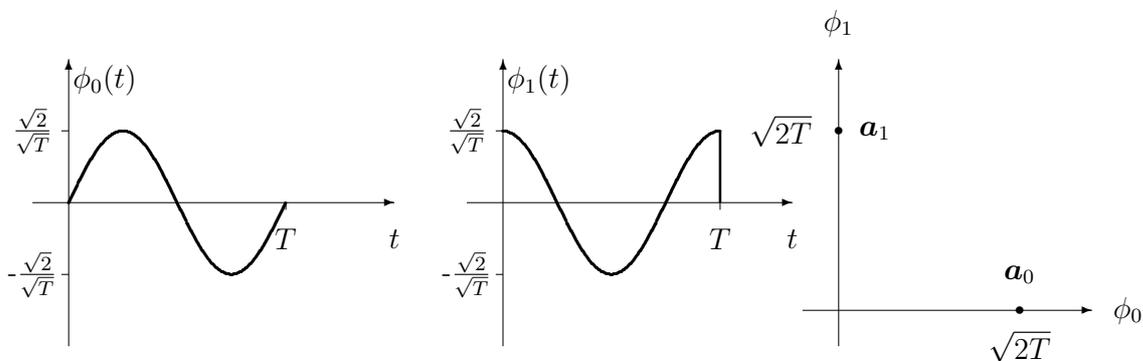


Figura 4.22: Ortogonalización para el cuarto conjunto de señales.

A diferencia de los conjuntos anteriores, aparecen dos elementos en la base del espacio vectorial. Si anteriormente los símbolos de la constelación \mathbf{a}_0 y \mathbf{a}_1 podían ser interpretados como puntos sobre

una recta, ahora se interpretan como puntos sobre un plano cuyos ejes representan las coordenadas sobre $\phi_0(t)$ y $\phi_1(t)$. Los símbolos \mathbf{a}_0 y \mathbf{a}_1 se expresan ahora como vectores de dimensión 2 cuyos valores son

$$\mathbf{a}_0 = \begin{bmatrix} a_{0,0} \\ a_{0,1} \end{bmatrix} = \begin{bmatrix} \sqrt{2T} \\ 0 \end{bmatrix} \quad \mathbf{a}_1 = \begin{bmatrix} a_{1,0} \\ a_{1,1} \end{bmatrix} = \begin{bmatrix} 0 \\ \sqrt{2T} \end{bmatrix}.$$

Ahora la distancia se calcula como la distancia sobre el plano, esto es

$$\begin{aligned} d(s_0(t), s_1(t)) &= d(\mathbf{a}_0, \mathbf{a}_1) = \|\mathbf{a}_0 - \mathbf{a}_1\| \\ &= \sqrt{(a_{0,0} - a_{1,0})^2 + (a_{0,1} - a_{1,1})^2} \\ &= \sqrt{(\sqrt{2T})^2 + (-\sqrt{2T})^2} = 2\sqrt{T}. \end{aligned}$$

Y la energía media por símbolo es

$$\begin{aligned} E_s &= E[\mathcal{E}\{\mathbf{a}_i\}] = E[\|\mathbf{a}_i\|^2] \\ &= p_A(\mathbf{a}_0)\|\mathbf{a}_0\|^2 + p_A(\mathbf{a}_1)\|\mathbf{a}_1\|^2 \\ &= p_A(\mathbf{a}_0)(a_{0,0}^2 + a_{0,1}^2) + p_A(\mathbf{a}_1)(a_{1,0}^2 + a_{1,1}^2) \\ &= \frac{1}{2}2T + \frac{1}{2}2T = 2T. \end{aligned}$$

Si se comparan el primer conjunto y el tercero, se observa que el elemento de la base de señales es distinto pero la constelación es la misma. Si la constelación es la misma, son idénticas la separación entre señales y la energía media de la constelación, o energía media por símbolo. A partir de este hecho podemos extraer una conclusión: la elección del conjunto de elementos de la base $\{\phi_j(t), j = 0, \dots, N-1\}$ no va a afectar a la fiabilidad de la comunicación ni a la energía de la señal; la fiabilidad y la energía quedarán fijadas por la constelación.

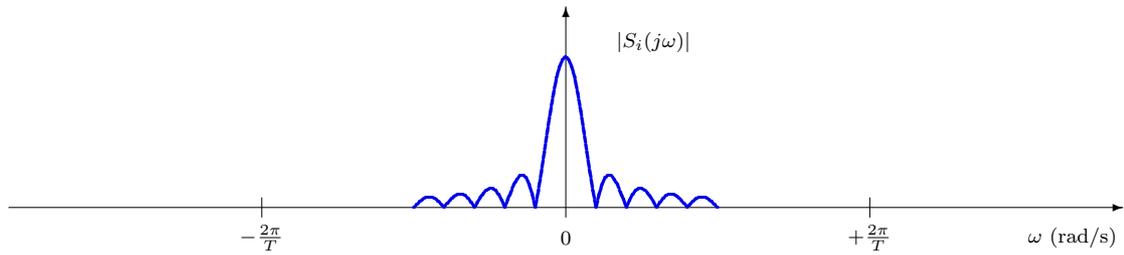
Para este ejemplo, el mejor compromiso entre prestaciones y energía se obtiene con los conjuntos 1 y 3, ya que comparten constelación, por lo que la diferencia (distancia) entre señales y la energía de cada señal es la misma. ¿Cuándo será más adecuado utilizar el conjunto 1, y cuando el conjunto 3? ¿De qué depende la elección de una u otra base ortonormal? Para responder a esta pregunta hay que considerar qué cambia al elegir entre una u otra base ortonormal. Lo que se modifica es la forma de las señales, en el dominio temporal (visto en las Figuras 4.15 y 4.17) y en el dominio frecuencial. La respuesta en este dominio para los cuatro conjuntos, que depende de la transformada de Fourier de las funciones base, se muestra en la Figura 4.23.

Se puede ver que las señales de los conjuntos 1 y 2 tienen su respuesta en frecuencia centrada en 0 Hz, con lo que son señales apropiadas para transmisión en canales con “buena respuesta” en bajas frecuencias. Las señales de los conjuntos 3 y 4 tienen su respuesta en frecuencia centrada en $\omega = \frac{2\pi}{T}$ rad/s, por lo que son señales apropiadas para transmisión en canales con “buena respuesta” en torno a la frecuencia $\frac{2\pi}{T}$ radianes/s. Es decir, que la elección de una u otra base ortonormal, dependerá de las características del canal por el que se ha de transmitir.

Esta conclusión no es específica de los conjuntos de señales que estamos evaluando, sino que es una conclusión general para el diseño de transmisores. En una situación en la que el objetivo sea diseñar el transmisor desde el inicio, la constelación y la base se diseñan prácticamente de forma independiente:

- La constelación atendiendo a criterios de fiabilidad de la comunicación y de energía de la señal, que estará limitada.

- Conjunto 1 y Conjunto 2



- Conjunto 3 y Conjunto 4

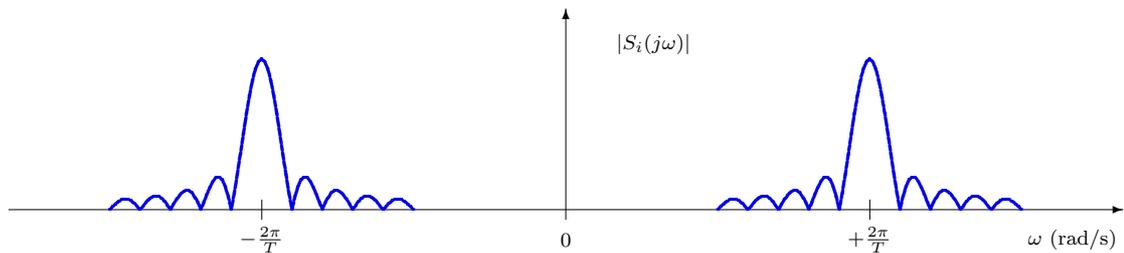


Figura 4.23: Respuesta en frecuencia de las señales de los cuatro conjuntos propuestos en el ejemplo.

- Los elementos de la base atendiendo al comportamiento del canal de comunicaciones (intendiendo que el canal físico se comporte como el modelo de ruido aditivo que hemos considerado).

La interpretación geométrica de las señales viene derivada de la estructura de espacio vectorial que poseen las señales y más concretamente de la definición de producto escalar en el espacio vectorial. Continuando con esta interpretación geométrica, la energía de un símbolo concreto es el cuadrado de la distancia del símbolo al origen de coordenadas, o lo que es lo mismo, el cuadrado de la norma del vector $E_i = \|\mathbf{a}_i\|^2$. También se puede medir el ángulo que forman los distintos vectores a través del producto escalar.

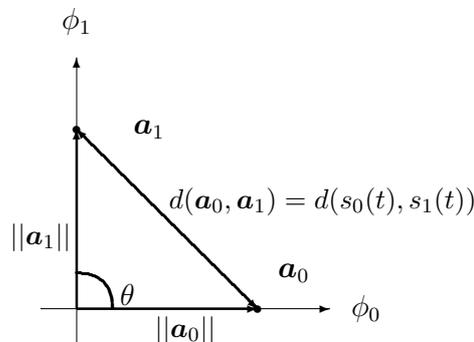


Figura 4.24: Interpretación geométrica de la constelación del cuarto conjunto de señales

La validez de todo lo planteado y discutido no se restringe a este conjunto de señales, sino que se puede aplicar al caso general de una constelación de M símbolos $\{\mathbf{a}_i, i = 0, \dots, M - 1\}$ y

una base ortonormal de N señales $\{\phi_j(t), j = 0, \dots, N - 1\}$ que generan el conjunto de señales $\{s_i(t), i = 0, \dots, M - 1\}$.

La constelación se diseñará de tal forma que se consiga una separación máxima entre los símbolos que la componen intentando al mismo tiempo que la energía media por símbolo, o energía de la constelación, esté dentro de los límites del sistema. La elección del conjunto de elementos de la base, de acuerdo con el modelo de canal que estamos considerando, es irrelevante en cuanto a prestaciones y energía. El conjunto de señales de la base se escoge para que sean adecuadas para el canal concreto sobre el que se realizará la transmisión. Y la elección de una u otra constelación no tendrá ninguna relevancia sobre la forma de las señales, en el dominio temporal y frecuencial, y por tanto en la adecuación de las señales al canal. Esto permite desacoplar el diseño de la constelación del diseño de la base ortonormal que define el espacio de señales (que se corresponden con el diseño del codificador y el diseño del modulador, respectivamente).

A continuación se analizará de forma individual cada uno de los 4 elementos funcionales del modelo de comunicación ilustrado en la Figura 4.14. Se empezará por el receptor para acabar con los módulos del transmisor.

4.4. Demodulador

El demodulador es el primer elemento funcional del receptor. Como se ilustra en la Figura 4.25, este elemento convierte la señal recibida $r(t)$ en una secuencia discreta $\mathbf{q}[n]$ de vectores de dimensión N .



Figura 4.25: Demodulador de un sistema de comunicaciones digital.

La secuencia $\mathbf{q}[n]$ contiene la representación vectorial de la señal recibida en el espacio de señales del sistema, definido por la base ortonormal de dimensión N

$$\{\phi_0(t), \phi_1(t), \dots, \phi_{N-1}(t)\}.$$

Se trata pues de una secuencia de vectores de dimensión N

$$\mathbf{q}[n] = \begin{bmatrix} q_0[n] \\ q_1[n] \\ \vdots \\ q_{N-1}[n] \end{bmatrix}$$

En concreto, la señal recibida se procesa por intervalos de símbolo, y en el instante discreto n , la secuencia $\mathbf{q}[n]$ contiene la representación vectorial del fragmento de la señal recibida en el intervalo de símbolo asociado a ese instante, es decir, en $nT \leq t < (n + 1)T$

$$\mathbf{q}[n] \equiv \text{proyección de } r(t) \text{ en } nT \leq t < (n + 1)T \text{ sobre } \{\phi_j(t)\}_{j=0}^{N-1}$$

La proyección de una señal sobre un elemento de la base, $\phi_k(t)$, se obtiene mediante el producto escalar. Como las señales que forman la base tiene por definición su soporte en el intervalo $0 \leq$

$t < T$, hacer el producto escalar del fragmento de la señal $r(t)$ en el intervalo del símbolo de índice n , en $nT \leq t < (n+1)T$, es equivalente al producto escalar de $r(t)$ con la señal $\phi_k(t)$ retardada nT segundos, es decir, que la coordenada de índice k del vector $\mathbf{q}[n]$ se puede obtener como

$$q_k[n] = \langle r(t), \phi_k(t - nT) \rangle = \int_{-\infty}^{\infty} r(t) \cdot \phi_k^*(t - nT) dt = \int_{nT}^{(n+1)T} r(t) \cdot \phi_k^*(t) dt.$$

El operador conjugado es relevante para una representación compleja de señales. En esta asignatura sólo se considerarán señales reales, aunque por completitud en la notación se incluirá la posibilidad de trabajar con señales complejas². Por tanto, para la implementación de un demodulador se necesita algún tipo de estructura capaz de obtener N productos escalares. Aquí se van a estudiar dos estructuras equivalentes para llevar a cabo esta labor:

1. Basada en operaciones de correlación.
2. Basada en operaciones de filtrado.

Por simplicidad en la notación, en los desarrollos inicialmente se considerará la obtención del primer vector de la secuencia, $\mathbf{q}[0]$, que contiene la representación vectorial de la señal $r(t)$ en el primer intervalo de símbolo, en $0 \leq t < T$. Por simplicidad, en este caso se obviará el índice temporal discreto $[0]$, de forma que se utilizará la notación

$$\mathbf{q} = \begin{bmatrix} q_0 \\ q_1 \\ \vdots \\ q_{N-1} \end{bmatrix} \equiv \mathbf{q}[0] = \begin{bmatrix} q_0[0] \\ q_1[0] \\ \vdots \\ q_{N-1}[0] \end{bmatrix}$$

La extensión para un instante discreto n arbitrario se hará de forma inmediata.

4.4.1. Demodulador por correlación

La idea del demodulador por correlación es utilizar la estructura que sale de la implementación directa de la operación producto escalar. Para el primer intervalo de símbolo

$$q_k = \langle r(t), \phi_k(t) \rangle = \int_0^T r(t) \phi_k^*(t) dt.$$

Por tanto, el demodulador tiene la estructura que se presenta en la Figura 4.26.

La notación vectorial que estamos utilizando puede dar la impresión de una estructura del demodulador excesivamente compleja para su implementación en hardware. Sin embargo, cuando particularizamos la estructura en sistemas prácticos, la estructura no suele ser en general muy compleja. Por ejemplo, para los cuatro conjuntos de señales del ejemplo de la Sección 4.3.1, la estructura es muy sencilla. Para los dos primeros conjuntos, en los que la función base toma un valor constante en todo el intervalo de símbolo, el demodulador se reduce a un escalado de la señal recibida y una integración en el intervalo de símbolo, como se muestra en la Figura 4.27 a); para el tercer conjunto, el demodulador multiplica la señal recibida por una señal sinusoidal (en hardware, un oscilador) e integra el producto en el intervalo de símbolo (Figura 4.27 b)); para el cuarto conjunto de señales se ha de multiplicar la señal recibida por dos componentes sinusoidales

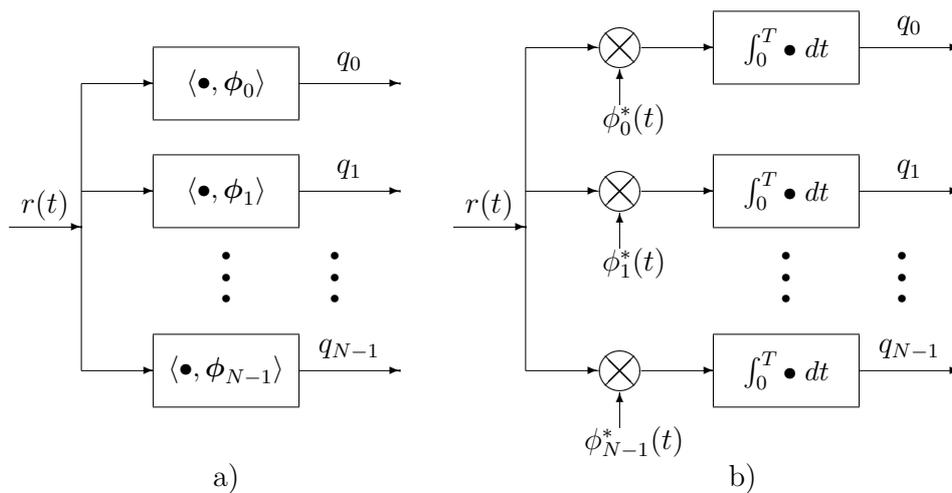


Figura 4.26: Estructura del demodulador por correlación; a) genérica, b) realización.

en cuadratura (un seno y un coseno, que en hardware se pueden obtener con un oscilador y un desfaseador de 90°) y dos integradores (Figura 4.27 c)).

4.4.2. El filtro adaptado

Otra posibilidad para el diseño del receptor es la utilización de filtros. Esta opción permite conservar y observar la característica temporal y frecuencial de las señales, y en algunos casos ofrece una alternativa más eficiente para la implementación en hardware del demodulador, dependiendo de la forma de las funciones base que definen el espacio de señales del sistema.

Si se filtra la señal recibida a la entrada del receptor, $r(t)$, con un cierto filtro con respuesta al impulso $h_k(t)$, su salida es

$$y_k(t) = r(t) * h_k(t) = \int_{-\infty}^{\infty} r(\tau) \cdot h_k(t - \tau) d\tau$$

Vamos a ver que mediante un proceso de filtrado es posible implementar la operación producto escalar que requiere el receptor. Recuerde que se busca una estructura para obtener

$$q_k = \langle r(t), \phi_k(t) \rangle = \int_0^T r(t) \phi_k^*(t) dt.$$

Si se comparan las dos expresiones anteriores, para $y_k(t)$ y q_k , respectivamente, se observa que en la segunda aparece $\phi_k^*(t)$ en lugar de $h_k(t)$, un complejo conjugado, que en la primera se integra sobre τ mientras en la segunda se integra sobre t , y que el signo sobre la variable de integración es negativo en la respuesta del filtro. Si se elige $h_k(t) = \phi_k^*(-t)$, la expresión de $y_k(t)$ se convierte en

$$y_k(t) = \int_{-\infty}^{\infty} r(\tau) \cdot \phi_k^*(-(t - \tau)) d\tau$$

Si se compara con la expresión analítica para q_k , y tenemos en cuenta que $-(t - \tau)$ se puede escribir como $(\tau - t)$ vemos que las dos expresiones son iguales si se toma el valor $t = 0$. Por tanto

²No existen señales electromagnéticas complejas, pero en algunos casos se utiliza la notación compleja para el manejo simultáneo de dos señales, una contenida en la parte real y otra en la parte imaginaria de la señal analítica.

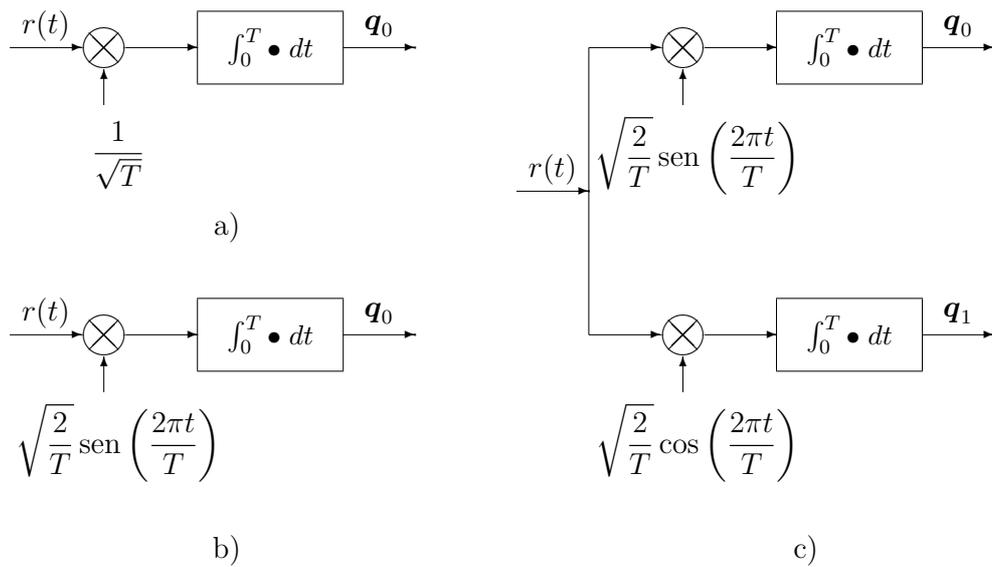


Figura 4.27: Estructura del demodulador para los cuatro conjuntos de señales analizados; a) dos primeros conjuntos, b) tercer conjunto, c) cuarto conjunto.

si se muestrea la señal de salida del filtro $y_k(t)$ en $t = 0$, tenemos

$$y_k(0) = \int_{-\infty}^{\infty} r(\tau) \cdot h_k(-\tau) d\tau = \int_{-\infty}^{\infty} r(\tau) \cdot \phi_k^*(\tau) d\tau = q_k,$$

es decir, el producto escalar de la señal recibida con $\phi_k(t)$.

En general, un filtro con respuesta al impulso $h(t) = x^*(-t)$ se dice que es un *filtro adaptado* a la señal $x(t)$. Por lo que el demodulador se puede implementar utilizando un banco de N filtros adaptados a los elementos de la base $\phi_k(t)$, con respuesta $h_k(t) = \phi_k^*(-t)$, y muestreando en $t = 0$ (al principio del intervalo de símbolo), tal y como se muestra en la Figura 4.28

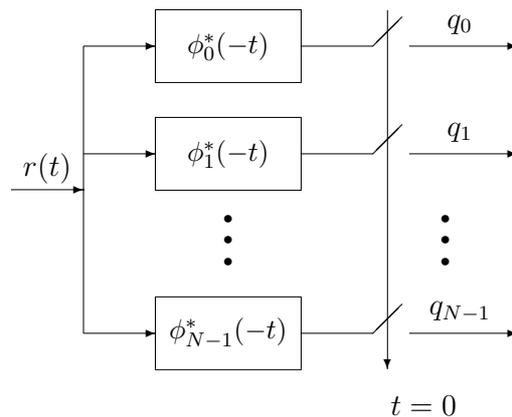


Figura 4.28: Estructura del demodulador basada en filtros adaptados.

Para la implementación práctica de esta estructura surge el problema de que los filtros son anticausales: como $\phi_k(t)$ tiene por definición como soporte $0 \leq t < T$, $\phi_k^*(-t)$ tiene como soporte $-T < t \leq 0$. Para poder tener una implementación causal, se puede retardar la respuesta $\phi_k^*(-t)$ T segundos, con lo que la salida del filtro se retarda la misma cantidad y es necesario muestrear

en $t = T$ en lugar de en cero, tal y como se muestra en la Figura 4.29. De este modo se tiene lo mismo que con el filtro anticausal muestreando en cero.

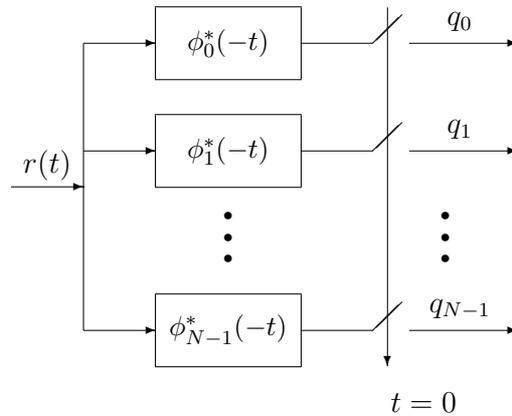


Figura 4.29: Estructura del demodulador basada en filtros adaptados causales.

Analíticamente, desplazar $\phi_k^*(-t)$ un tiempo T significa tener $\phi_k^*(-(t - T)) = \phi_k^*(T - t)$. La salida de ese filtro desplazado T , denotada $y_k^T(t)$ es entonces

$$y_k^T(t) = \int_{-\infty}^{\infty} r(\tau) \times \phi_k^*(T - (t - \tau)) d\tau.$$

Muestreando en $t = T$ se tiene

$$y_k^T(T) = \int_{-\infty}^{\infty} r(\tau) \cdot \phi_k^*(\tau) d\tau = y_k(0) = q_k.$$

Se observa por tanto que esta estructura proporciona exactamente el mismo resultado que la que utiliza filtros anticausales. Para la implementación, evidentemente hay que utilizar la estructura con filtros causales, pero como ambas son analíticamente equivalentes, para hacer cálculos se puede utilizar la expresión con filtros anticausales que es más sencilla.

Propiedades del filtro adaptado

El filtro adaptado no aporta por sí mismo ninguna novedad más que una estructura alternativa para la implementación de las operaciones de producto escalar que debe llevar a cabo el demodulador. Sin embargo, esta estructura permite comprobar de una forma sencilla algunas de las propiedades del demodulador.

Máxima relación señal a ruido Esta es seguramente la propiedad más importante. Para demostrar esta propiedad se propone el caso genérico en el que se tiene una señal conocida $s(t)$ a la que se suma ruido y que se filtra con un filtro con respuesta al impulso $h(t)$. El modelo para el término de ruido es el modelo estadístico habitual para el ruido térmico: proceso aleatorio estacionario, ergódico, blanco y gaussiano, con media nula y densidad espectral de potencia y función de autocorrelación

$$S_n(j\omega) = \frac{N_0}{2}, R_n(\tau) = \frac{N_0}{2}\delta(\tau).$$

Este esquema general se muestra en la Figura 4.30.

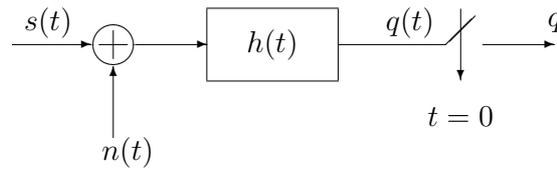


Figura 4.30: Modelo para la derivación de la propiedad de máxima relación señal a ruido del filtro adaptado.

Se va a comprobar en primer lugar que el filtro que proporciona la máxima relación señal a ruido a su salida, definida como

$$\left(\frac{S}{N}\right)_q \equiv \frac{\text{Energía en } q \text{ debida a } s(t)}{\text{Energía en } q \text{ debida a } n(t)},$$

es un filtro adaptado a la señal transmitida $s(t)$. Para simplificar la notación se considera el caso de señales reales, aunque los resultados se extienden trivialmente al caso de señales complejas. La salida del filtro, $q(t)$, se define como

$$\begin{aligned} q(t) &= (s(t) + n(t)) * h(t) = s(t) * h(t) + n(t) * h(t) \\ &= \int_{-\infty}^{\infty} s(\tau) \cdot h(t - \tau) d\tau + \int_{-\infty}^{\infty} n(\tau) \cdot h(t - \tau) d\tau. \end{aligned}$$

Su valor en $t = 0$ es

$$q = q(0) = \int_{-\infty}^{\infty} s(\tau) \cdot h(-\tau) d\tau + \int_{-\infty}^{\infty} n(\tau) \cdot h(-\tau) d\tau = s + n.$$

La relación señal a ruido de la salida se define como

$$\left(\frac{S}{N}\right)_q = \frac{E[|s|^2]}{E[|n|^2]} = \frac{|s|^2}{E[n^2]},$$

ya que el término s es determinista si la señal $s(t)$ es conocida

$$E[|s|^2] = |s|^2 = \left| \int_{-\infty}^{\infty} s(\tau) \cdot h(-\tau) d\tau \right|^2 \quad (\text{valor determinista})$$

El valor de $E[|n|^2]$ se obtiene como

$$\begin{aligned} E[|n|^2] &= E \left[\left(\int_{-\infty}^{+\infty} n(\tau) \cdot h(-\tau) d\tau \right) \left(\int_{-\infty}^{+\infty} n(\theta) \cdot h(-\theta) d\theta \right) \right] \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \underbrace{E[n(\tau) \cdot n(\theta)]}_{R_n(\tau-\theta)} \cdot h(-\tau) \cdot h(-\theta) d\tau d\theta \\ &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \underbrace{\frac{N_0}{2} \delta(\tau - \theta)}_{R_n(\tau-\theta)} \cdot h(-\tau) \cdot h(-\theta) d\tau d\theta \\ &= \frac{N_0}{2} \int_{-\infty}^{+\infty} |h(-\tau)|^2 d\tau = \frac{N_0}{2} \int_{-\infty}^{+\infty} |h(\tau)|^2 d\tau = \frac{N_0}{2} \cdot \mathcal{E}\{h(t)\} \end{aligned}$$

En el desarrollo se ha aplicado la propiedad de la integral del producto de una delta y una función

$$\int_{-\infty}^{+\infty} f(x) \cdot \delta(x - x_0) dx = f(x_0).$$

Se puede ver que la energía de la componente de ruido únicamente depende de la energía de la respuesta en frecuencia del filtro. Sustituyendo este resultado en la expresión de la relación señal a ruido se tiene

$$\left(\frac{S}{N}\right)_q = \frac{|s|^2}{E[|n|^2]} = \frac{\left|\int_{-\infty}^{\infty} s(\tau) \cdot h(-\tau) d\tau\right|^2}{\frac{N_0}{2} \mathcal{E}\{h(t)\}}.$$

Para buscar el máximo de esta relación señal a ruido respecto a la respuesta al impulso del filtro, $h(t)$, se va a hacer el siguiente razonamiento: se supone que la energía de $h(t)$ es constante, y se busca el máximo de la relación $(S/N)_q$ respecto a $h(t)$; si ese máximo no depende de $\mathcal{E}\{h(t)\}$ tendremos el máximo para cualquier valor de $\mathcal{E}\{h(t)\}$ y, por consiguiente, el máximo que estamos buscando.

Si se supone $\mathcal{E}\{h(t)\}$ constante, el máximo de la relación señal a ruido se limita a calcular el máximo del numerador. Si se particulariza la desigualdad de Cauchy-Schwarz (ver Sección 4.2.2) para las señales $s(t)$ y $h(-t)$, y elevando al cuadrado cada uno de los términos de la desigualdad se tiene

$$\left|\int_{-\infty}^{\infty} s(\tau) \cdot h(-\tau) d\tau\right|^2 \leq \left(\int_{-\infty}^{\infty} |s(\tau)|^2 d\tau\right) \left(\int_{-\infty}^{\infty} |h(-\tau)|^2 d\tau\right).$$

La igualdad, y por tanto el máximo, se produce cuando $h(-t) = \alpha \times s(t)$ para algún valor de la constante α . Introduciendo este resultado se obtiene

$$\begin{aligned} \max_{h(t)} \left(\frac{S}{N}\right)_q &= \frac{\left(\int_{-\infty}^{\infty} s(\tau) h(-\tau) d\tau\right)^2}{\frac{N_0}{2} \mathcal{E}\{h(t)\}} \Bigg|_{h(-t)=\alpha s(t)} \\ &= \frac{\left(\int_{-\infty}^{\infty} |s(\tau)|^2 d\tau\right) \left(\alpha^2 \int_{-\infty}^{\infty} |s(\tau)|^2 d\tau\right)}{\frac{N_0}{2} \alpha^2 \mathcal{E}\{s(t)\}} \\ &= \frac{2}{N_0} \mathcal{E}\{s(t)\}. \end{aligned}$$

De este resultado se obtienen dos conclusiones:

1. La relación señal a ruido a la salida se hace máxima cuando $h(t) = \alpha s(-t)$ para cualquier valor de α (exceptuando $\alpha = 0$) y, particularmente, para el filtro adaptado

$$h(t) = s(-t),$$

Para señales complejas se llega de forma trivial a la misma conclusión para

$$h(t) = s^*(-t).$$

2. La relación señal a ruido a la salida del filtro adaptado no depende de la forma específica de $s(t)$, sino únicamente de su energía y de la densidad espectral de potencia de ruido a la entrada del filtro.

Esta demostración realizada para una señal genérica se puede aplicar al conjunto de señales del transmisor, $\{s_i(t), i = 0, \dots, M-1\}$ y al conjunto de señales de la base, $\{\phi_i(t), i = 0, \dots, M-1\}$, con lo que se demuestra que la estructura de demodulador basado en filtros adaptados (o su equivalente de correlación) es la estructura que nos permite obtener la máxima relación señal a ruido en cada componente del vector \mathbf{q} .

4.4.3. Extensión de las estructuras del demodulador para cualquier instante n

Se ha estudiado en las secciones anteriores el demodulador para la obtención de la primera de las observaciones de la secuencia de vectores, $\mathbf{q} \equiv \mathbf{q}[0]$. La extensión para la obtención del vector de la secuencia en cualquier instante arbitrario, $\mathbf{q}[n]$, es trivial, ya que sólo hay que procesar $r(t)$ en el intervalo de símbolo correspondiente, $nT \leq t < (n+1)T$. En el caso de la estructura con correladores, sólo hay que retardar las funciones $\phi_k(t)$ nT segundos e integrar en el intervalo de símbolo, y en el caso de los filtros adaptados sólo hay que modificar el instante de muestreo, para que coincida con el inicio del intervalo de símbolo, $t = nT$, si se utilizan filtros anticausales, o con el final del intervalo de símbolo, $(n+1)T$, si se utilizan filtros causales, como se muestra en la Figura 4.31.

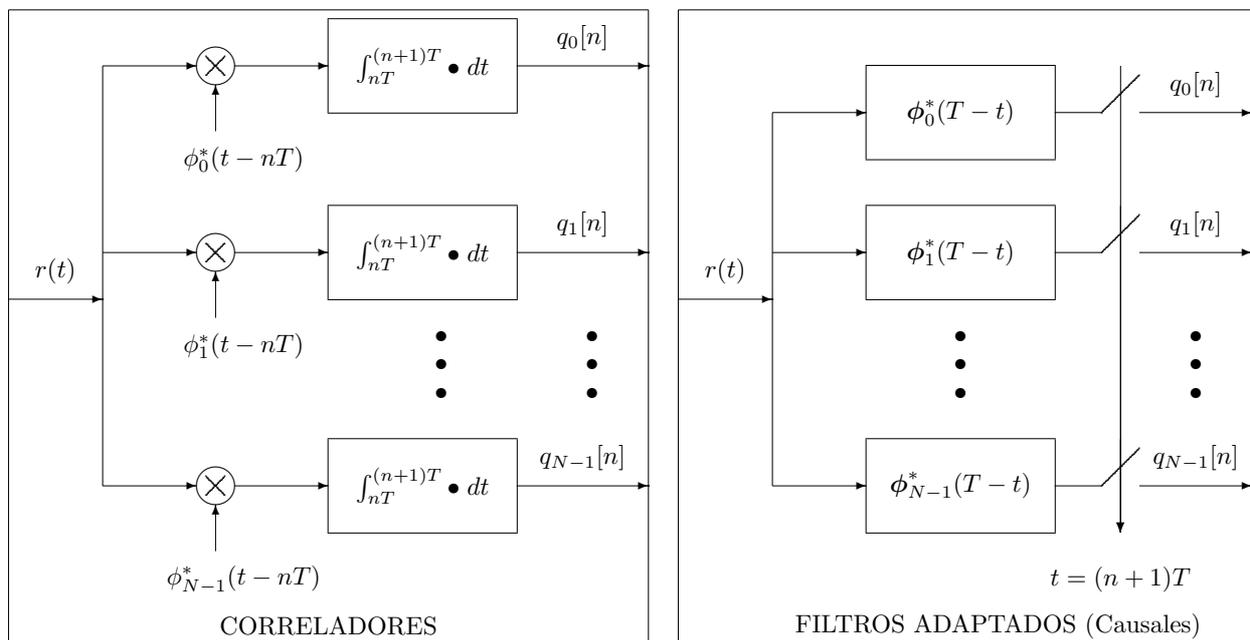


Figura 4.31: Estructura del demodulador para un flujo indefinido de símbolos.

Es interesante remarcar que ambas estructuras proporcionan el mismo resultado, ya que se trata de dos estructuras que realizan la misma operación analítica: el producto escalar de la señal recibida con cada una de las N funciones que forman la base ortonormal que define el espacio de señales del sistema.

4.4.4. Caracterización estadística de la salida del demodulador en el caso de transmisión sobre un canal gaussiano

Una vez que se han visto las estructuras que pueden utilizarse para la implementación del demodulador, a continuación se realizará la caracterización estadística de la secuencia de observaciones $\mathbf{q}[n]$. En particular se considerará el caso de transmisión de la señal sobre un canal gaussiano, cuando la señal transmitida no sufre ninguna distorsión lineal (lo que implica suponer que la base ortonormal se adapta perfectamente a las características del canal), y que la única distorsión se

debe al efecto del ruido, con lo que la señal recibida será

$$r(t) = s(t) + n(t).$$

De nuevo, para simplificar la notación se considerará el primer instante discreto, $n = 0$, y se obviará la dependencia con el índice temporal discreto n . El objetivo es obtener el modelo estadístico de la observación \mathbf{q} cuando se conoce el símbolo que se ha transmitido, que con carácter general se supondrá que es el símbolo de índice i , i.e. $\mathbf{A} \equiv \mathbf{A}[0] = \mathbf{a}_i$. Esto significa que en el intervalo de símbolo de interés, el primero, $0 \leq t < T$, la señal transmitida es

$$s(t) = s_i(t) = \sum_{j=0}^{N-1} a_{i,j} \times \phi_j(t).$$

A continuación se analiza la salida del demodulador bajo esta situación. Introduciendo la expresión analítica de $r(t)$ en la obtención de la coordenada de índice k , q_k , se tiene

$$\begin{aligned} q_k &= \langle r(t), \phi_k(t) \rangle = \int_0^T r(t) \cdot \phi_k^*(t) dt = \int_0^T (s(t) + n(t)) \cdot \phi_k^*(t) dt \\ &= \int_0^T \left(\sum_{j=0}^{N-1} a_{i,j} \cdot \phi_j(t) \right) \phi_k^*(t) dt + \underbrace{\int_0^T n(t) \cdot \phi_k^*(t) dt}_{z_k} \\ &= \sum_{j=0}^{N-1} a_{i,j} \int_0^T \phi_j(t) \cdot \phi_k^*(t) dt + z_k = \sum_{j=0}^{N-1} a_{i,j} \cdot \delta[j - k] + z_k = a_{i,k} + z_k. \end{aligned}$$

Se ha tenido en cuenta en el desarrollo que los elementos de la base son ortonormales, lo que analíticamente implica que

$$\langle \phi_j(t), \phi_k(t) \rangle = \int_0^T \phi_j(t) \cdot \phi_k^*(t) dt = \begin{cases} 0, & \text{si } k \neq j \\ 1, & \text{si } k = j \end{cases} \equiv \delta[j - k].$$

En este caso $a_{i,k}$ es la coordenada de índice k del símbolo \mathbf{a}_i y z_k es la contribución del ruido gaussiano en esa coordenada. Todas las componentes (coordenadas) de \mathbf{q} pueden expresarse de forma conjunta empleando notación vectorial como

$$\mathbf{q} = \begin{bmatrix} a_{i,0} \\ a_{i,1} \\ \vdots \\ a_{i,N-1} \end{bmatrix} + \begin{bmatrix} z_0 \\ z_1 \\ \vdots \\ z_{N-1} \end{bmatrix} = \mathbf{a}_i + \mathbf{z}$$

Los valores del vector \mathbf{a}_i son deterministas (se parte de la asunción de que el símbolo transmitido es conocido), por lo que sólo queda caracterizar el vector de ruido \mathbf{z} . Cada una de las componentes del vector de ruido se considera como una variable aleatoria. Dado que el proceso de que proviene, $n(t)$, es gaussiano, cada componente tendrá una función de densidad de probabilidad gaussiana. El vector con las coordenadas de ruido está por tanto compuesto por N variables aleatorias conjuntamente gaussianas. Para caracterizarlas, hay que encontrar su función densidad de probabilidad conjunta, lo que en el caso de v.a. gaussianas se reduce al cálculo de su vector de medias y su matriz de covarianzas. Comenzaremos calculando la media de cada una de las coordenadas del vector de ruido. Aplicando la definición de media de una variable aleatoria, la media de la coordenada de índice k es

$$m_{z_k} = E[z_k] = E \left[\int_0^T n(t) \phi_j^*(t) dt \right] = \int_0^T E[n(t)] \phi_j^*(t) dt = 0,$$

ya que $n(t)$ tiene media nula. A continuación se calcula la matriz de covarianzas. Teniendo en cuenta que todas las componentes tienen media nula, la covarianza entre dos de las coordenadas es

$$\begin{aligned} \text{Cov}(z_j, z_k) &= E[z_j \cdot z_k^*] = E \left[\left(\int_0^T n(t) \cdot \phi_j^*(t) dt \right) \left(\int_0^T n^*(\tau) \cdot \phi_k(\tau) d\tau \right) \right] \\ &= \int_0^T \int_0^T \underbrace{E[n(t) \cdot n^*(\tau)]}_{R_n(t-\tau) = \frac{N_0}{2} \cdot \delta(t-\tau)} \phi_j^*(t) \cdot \phi_k(\tau) dt d\tau \\ &= \int_0^T \int_0^T \frac{N_0}{2} \cdot \delta(t-\tau) \cdot \phi_j^*(t) \cdot \phi_k(\tau) dt d\tau \\ &= \frac{N_0}{2} \int_0^T \phi_j^*(t) \cdot \phi_k(t) dt = \frac{N_0}{2} \cdot \delta[j-k]. \end{aligned}$$

Por tanto, la covarianza entre dos coordenadas distintas es nula, y la covarianza de una coordenada consigo misma (varianza) es

$$\sigma_{z_k}^2 = \frac{N_0}{2}.$$

Esto significa que las variables aleatorias que modelan cada uno de los elementos del vector de ruido están incorreladas, lo que para variables aleatorias gaussianas es equivalente a decir que son independientes. Por tanto, la matriz de covarianzas es una matriz diagonal con la varianza de cada componente en la diagonal principal. A la luz de estos resultados se puede concluir que las N componentes de ruido son variables aleatorias gaussianas incorreladas (independientes) de media nula y varianza $N_0/2$. Como una variable aleatoria gaussiana queda unívocamente determinada a partir de su media y su varianza, la función densidad de probabilidad de cada componente z_k es

$$f_{z_k}(z_k) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{z_k^2}{N_0}},$$

y como bajo estadística gaussiana incorrelación implica independencia, la función densidad de probabilidad conjunta de \mathbf{z} se obtiene mediante el producto de las distribuciones de cada una de las componentes del vector

$$f_{\mathbf{z}}(\mathbf{z}) = \prod_{k=0}^{N-1} f_{z_k}(z_k) = \frac{1}{(\pi N_0)^{N/2}} e^{-\sum_{k=0}^{N-1} \frac{z_k^2}{N_0}} = \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{\|\mathbf{z}\|^2}{N_0}}.$$

Una vez obtenida la función de densidad de probabilidad del vector de ruido, es fácil obtener la función de densidad de probabilidad condicional de la observación \mathbf{q} . La función densidad de probabilidad de \mathbf{q} cuando se conoce el símbolo transmitido, $\mathbf{A} = \mathbf{a}_i$, es una distribución condicional: distribución de q_k condicionada a que el símbolo transmitido es $\mathbf{A} = \mathbf{a}_i$. Para cada una de las componentes del vector \mathbf{q} , q_k , se tiene también una distribución condicional: distribución de q_k condicionada a que el símbolo transmitido es $\mathbf{A} = \mathbf{a}_i$. En ese caso, $q_k = a_{i,k} + z_k$, es una variable aleatoria formada por la suma de una constante determinista $a_{i,k}$ y una variable aleatoria, z_k , por lo que tiene el mismo tipo de distribución que z_k pero con la media modificada sumando el valor constante, $a_{i,k}$, a la media de z_k , que en este caso era cero. Por tanto, la distribución es una distribución gaussiana, de media la coordenada de índice k del símbolo transmitido, $a_{i,k}$, y varianza $N_0/2$

$$f_{q_k|\mathbf{A}}(q_k|\mathbf{a}_i) = \frac{1}{\sqrt{\pi N_0}} e^{-\frac{(q_k - a_{i,k})^2}{N_0}}.$$

Debido a la independencia, la función densidad de probabilidad condicional conjunta del vector \mathbf{q} cuando se conoce que $\mathbf{A} = \mathbf{a}_i$ se obtiene mediante el producto de las distribuciones condicionales de cada una de las componentes del vector

$$\begin{aligned} f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i) &= \prod_{k=0}^{N-1} f_{q_k|A_k}(q_k|a_{i,k}) = \frac{1}{(\pi N_0)^{N/2}} e^{-\sum_{k=0}^{N-1} \frac{(q_k - a_{i,k})^2}{N_0}} \\ &= \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{\|\mathbf{q} - \mathbf{a}_i\|^2}{N_0}}. \end{aligned}$$

Como se puede ver, la distribución del vector de observación cuando se transmite un cierto símbolo es una distribución gaussiana N dimensional, de componentes independientes, media el símbolo transmitido y varianza $N_0/2$ en cada una de las N direcciones del espacio

$$f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i) = \mathcal{N}^N \left(\mathbf{a}_i, \frac{N_0}{2} \right).$$

En el desarrollo del demodulador hemos partido de la idea de recuperar las coordenadas de la señal recibida en la base del espacio de señales del transmisor, pero no nos hemos preguntado si la información que recuperamos de la señal recibida es toda la información relevante a la hora de tomar la decisión del símbolo que fue transmitido o existe información adicional que pueda ayudar a tomar la decisión. La respuesta a esta pregunta es que tal y como se ha definido el vector \mathbf{q} , contiene toda la información relevante para decidir qué símbolo fue el transmitido. Estrictamente, se dice que el vector \mathbf{q} es un *estadístico suficiente para la detección*. La demostración se puede encontrar por ejemplo en [Artés-Rodríguez et al., 2007].

4.4.5. Canal discreto equivalente

Algunas de las conclusiones obtenidas del análisis del modelo de comunicación digital son las siguientes:

1. La fiabilidad de un esquema de comunicación, por lo que respecta al proceso de modulación, viene dada por las características de la constelación de señales y no por los elementos de la base de señales que empleemos en el proceso de modulación.
2. La relación señal a ruido no depende de la forma concreta que toman cada uno de los elementos de la base de señales, sino de la energía de las señales y de la densidad espectral de potencia del ruido (la potencia del ruido).
3. La salida del demodulador es un vector \mathbf{q} que toma la forma $\mathbf{q} = \mathbf{A} + \mathbf{z}$, donde \mathbf{A} es el símbolo transmitido y \mathbf{z} es una componente de ruido con una función densidad de probabilidad conjuntamente gaussiana.

Estas conclusiones permiten obtener una simplificación del modelo general de un sistema de comunicación, englobando modulador, canal y demodulador en un único elemento que denominamos *canal discreto equivalente*. Este modelo, que se representa en la Figura 4.32, es de utilidad en el análisis del decisor porque nos permite “ocultar” la naturaleza analógica del canal y centrarnos en los aspectos que van a influir en la fiabilidad de la comunicación.

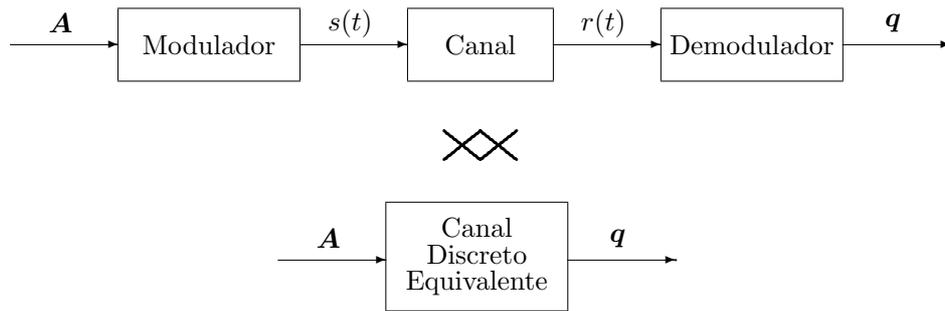


Figura 4.32: Definición de canal discreto equivalente.

El canal discreto equivalente es en general vectorial, con la dimensión dada por el espacio de señales del sistema, N . La relación entrada salida, para un modelo de transmisión sobre canal gaussiano, viene dada por

$$\mathbf{q} = \mathbf{A} + \mathbf{z},$$

tal y como se representa en la Figura 4.33

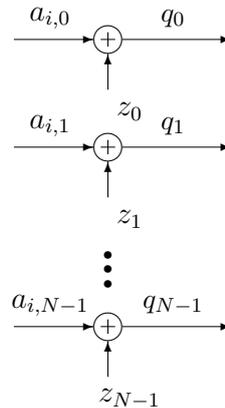


Figura 4.33: Relaciones entre representaciones vectoriales de transmisor y receptor dadas por el canal discreto equivalente, cuando se transmite el símbolo $\mathbf{A} = \mathbf{a}_i$ sobre un canal gaussiano.

4.5. Decisor

El decisor es el elemento final de un demodulador digital, y su función es proporcionar la estima del símbolo que se ha transmitido en un instante n , $\hat{B}[n]$, a partir de la representación vectorial de la señal recibida en el intervalo de símbolo asociado a dicho instante, $\mathbf{q}[n]$, tal y como se ilustra en la Figura 4.34. Con esta estima, de forma implícita se tiene también la estima de los bits transmitidos, $\hat{B}_b[\ell]$, debido a la identificación de cada posible valor del símbolo, b_i , con una m -tupla de bits específica.



Figura 4.34: Decisor de un sistema de comunicaciones.

Dada su función, es un elemento determinante para las prestaciones del sistema, parametrizadas por la probabilidad de error, definida a nivel de símbolo o a nivel de bit. Por ello, el decisor se diseña de acuerdo con el siguiente criterio: hacer mínima la probabilidad de error de símbolo.

4.5.1. Diseño del decisor - Regiones de decisión

Debido a que $\mathbf{q}[n]$ es un estadístico suficiente para la detección de $\mathbf{A}[n]$ (y por tanto de $B[n]$), la decisión se realizará símbolo a símbolo sin memoria; es decir, para cada instante temporal n , se decidirá $\hat{B}[n]$ a partir de $\mathbf{q}[n]$, aplicando siempre la misma regla, ya que la estadística de $\mathbf{q}[n]$ no varía con el tiempo.

Teniendo en cuenta que el alfabeto de símbolos tiene M posibles valores

$$B[n] \in \{b_0, b_1, \dots, b_{M-1}\},$$

el diseño del decisor consiste en establecer para cada posible valor de $\mathbf{q}[n]$, qué valor del alfabeto, de los M posibles, se decide. O visto de otra forma, establecer cuáles son los posibles valores de $\mathbf{q}[n]$ que darán lugar a la decisión de cada uno de los M posibles valores de $B[n]$. La forma de establecer esos valores es definiendo lo que se denominan “*regiones de decisión*”. Se dividirá el dominio de $\mathbf{q}[n]$ en M regiones disjuntas que formen una partición del espacio de $\mathbf{q}[n]$

$$\{I_0, I_1, \dots, I_{M-1}\}.$$

Cada región se asocia a uno de los valores del alfabeto de símbolos, identificándose la asociación a través del subíndice (I_k se asocia a b_k). Una vez definidas las regiones, a la vista de un valor de observación $\mathbf{q}[n]$, se decidirá $\hat{B}[n] = b_k$ cuando el valor de observación esté en la región de decisión de b_k , i.e. cuando $\mathbf{q}[n] \in I_k$. Por eso se denominan regiones de decisión.

Por tanto, el diseño del decisor se plantea como el problema de establecer las M regiones de decisión que minimicen la probabilidad de error de símbolo.

4.5.2. Obtención del decisor óptimo

A continuación se obtendrán las reglas a seguir para establecer de forma óptima (para minimizar la probabilidad de error de símbolo) las regiones de decisión que definen el decisor. De nuevo, para simplificar la notación se tratará analíticamente el caso del primer símbolo de la secuencia (índice discreto $n = 0$) y se obviará la dependencia con el índice temporal, para lo que se utilizará como notación

$$\hat{B} \equiv \hat{B}[n], \quad \mathbf{q} \equiv \mathbf{q}[n].$$

Cuando \mathbf{q} toma un valor determinado \mathbf{q}_0 , el decisor toma una decisión sobre el símbolo transmitido, por ejemplo $\hat{B} = b_i$; la probabilidad de error en esa decisión, que denotamos como $P_e^{(\mathbf{q}=\mathbf{q}_0 \rightarrow \hat{B}=b_i)}$ es

$$P_e^{(\mathbf{q}=\mathbf{q}_0 \rightarrow \hat{B}=b_i)} = P(B \neq b_i | \mathbf{q} = \mathbf{q}_0) = 1 - P(B = b_i | \mathbf{q} = \mathbf{q}_0) = 1 - p_{B|\mathbf{q}}(b_i | \mathbf{q}_0).$$

La probabilidad de error en una decisión ($\hat{B} = b_i$) tomada a partir de un valor de observación determinado ($\mathbf{q} = \mathbf{q}_0$) será igual a la probabilidad de que, siendo la entrada igual a \mathbf{q}_0 el símbolo transmitido no sea igual al decidido. El valor $p_{B|\mathbf{q}}(b_i | \mathbf{q}_0)$ es, obviamente, la probabilidad de acertar cuando ante un valor de entrada \mathbf{q}_0 el decisor asigna el símbolo b_i ($\hat{B} = b_i$). Esta es una probabilidad condicional, que se denomina *probabilidad a posteriori* del símbolo b_i .

Si el decisor tomara siempre la misma decisión, $\hat{B} = b_i$, independientemente del valor de observación \mathbf{q} , la probabilidad media de error, que se denotará como $P_e^{(\hat{B}=b_i, \forall \mathbf{q})}$, se obtiene calculando el promedio sobre el conjunto de todos los posibles valores de la observación \mathbf{q} , que estadísticamente es la esperanza matemática bajo la distribución de \mathbf{q}

$$P_e^{(\hat{B}=b_i, \forall \mathbf{q})} = E_{f_{\mathbf{q}}(\mathbf{q}_0)} \left[P_e^{(\mathbf{q}=\mathbf{q}_0 \rightarrow \hat{B}=b_i)} \right] = \int_{-\infty}^{\infty} [1 - p_{B|\mathbf{q}}(b_i|\mathbf{q}_0)] \cdot f_{\mathbf{q}}(\mathbf{q}_0) d\mathbf{q}_0 \quad (4.1)$$

$$= \int_{-\infty}^{\infty} f_{\mathbf{q}}(\mathbf{q}_0) d\mathbf{q}_0 - \int_{-\infty}^{\infty} p_{B|\mathbf{q}}(b_i|\mathbf{q}_0) \cdot f_{\mathbf{q}}(\mathbf{q}_0) d\mathbf{q}_0 \quad (4.2)$$

$$= 1 - \int_{-\infty}^{\infty} p_{B|\mathbf{q}}(b_i|\mathbf{q}_0) \cdot f_{\mathbf{q}}(\mathbf{q}_0) d\mathbf{q}_0 \quad (4.3)$$

Cuando el decisor toma decisiones distintas cuando la observación \mathbf{q} toma valores en cada una de las M regiones de decisión, la probabilidad de error del sistema, teniendo en cuenta que se trata de regiones disjuntas que forman una partición del dominio de \mathbf{q} , es

$$P_e = 1 - \sum_{i=0}^{M-1} \int_{I_i} p_{B|\mathbf{q}}(b_i|\mathbf{q}_0) \cdot f_{\mathbf{q}}(\mathbf{q}_0) d\mathbf{q}_0$$

Para encontrar el mínimo de esta expresión hay que tener en cuenta lo siguiente:

- El mínimo se obtiene cuando se maximiza el segundo término.
- La función dentro de las integrales de este término, $p_{B|\mathbf{q}}(b_i|\mathbf{q}_0)f_{\mathbf{q}}(\mathbf{q}_0)$, es siempre mayor o igual que cero por ser el producto de funciones no negativas ($0 \leq p_{B|\mathbf{q}}(b_i|\mathbf{q}_0) \leq 1$, $0 \leq f_{\mathbf{q}}(\mathbf{q}_0) \leq 1$). Esto implica que la probabilidad de error se minimiza cuando se maximiza este argumento.
- $f_{\mathbf{q}}(\mathbf{q}_0)$ es independiente tanto del símbolo transmitido como del recibido. Por tanto, el máximo del sumatorio se obtiene cuando se maximiza el valor de $p_{B|\mathbf{q}}(b_i|\mathbf{q}_0)$ en cada una de las regiones de decisión.
- Por consiguiente, la región de decisión I_i , para la cual se asigna como salida del decisor el valor $\hat{B} = b_i$, es aquella para la que cumple que

$$p_{B|\mathbf{q}}(b_i|\mathbf{q}_0) > p_{B|\mathbf{q}}(b_j|\mathbf{q}_0), \quad \forall j \neq i.$$

Es decir, que ante un valor de entrada $\mathbf{q} = \mathbf{q}_0$, el decisor ha de calcular el conjunto de probabilidades $\{p_{B|\mathbf{q}}(b_j|\mathbf{q}_0), j = 0, \dots, M-1\}$ y la salida del decisor será el símbolo b_i que cumple la relación anterior. En el caso en que dos símbolos distintos b_i, b_k , obtengan el máximo valor, es decir

$$p_{B|\mathbf{q}}(b_i|\mathbf{q}_0) = p_{B|\mathbf{q}}(b_k|\mathbf{q}_0) > p_{B|\mathbf{q}}(b_j|\mathbf{q}_0), \quad \forall j \neq i, k,$$

el decisor podrá optar por cualquiera de ellos de forma arbitraria sin que esta decisión tenga efecto alguno sobre la probabilidad de error.

Este criterio se conoce con el nombre de *máximo a posteriori* o MAP (de “*maximum a posteriori*”). Su nombre viene de la denominación de $p_{B|\mathbf{q}}(b_i|\mathbf{q}_0)$ como probabilidades *a posteriori* de B dado \mathbf{q} , ya que representa la probabilidad de los símbolos una vez realizada la transmisión, en lugar de las probabilidades *a priori* $p_B(b_i)$. La Figura 4.35 muestra un ejemplo de como podrían

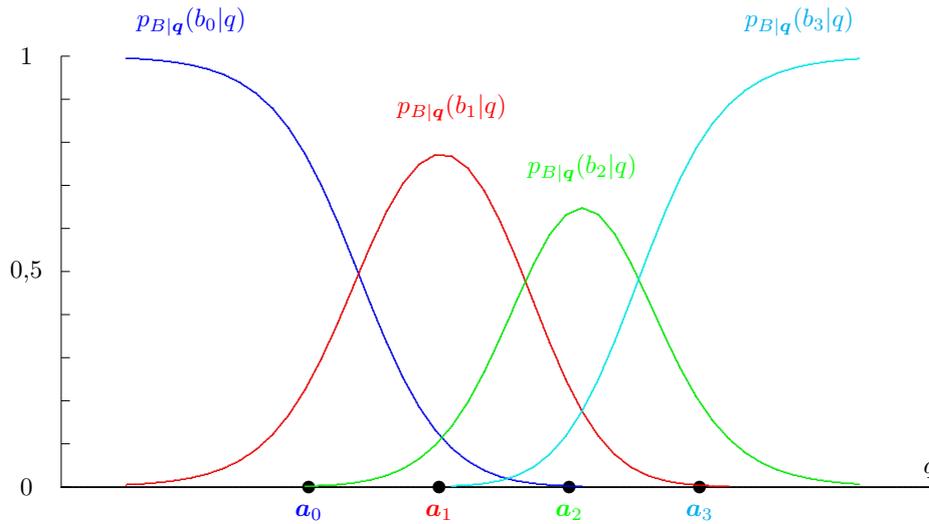


Figura 4.35: Un ejemplo de las probabilidades a posteriori para una constelación de 4 símbolos en un espacio unidimensional.

ser esas probabilidades a posteriori para una constelación de 4 símbolos en un espacio unidimensional. Se puede comprobar que para cada posible valor de observación q , la suma de las cuatro probabilidades a posteriori es uno.

Para obtener las regiones de decisión, hay que buscar cuál de los 4 símbolos tiene asociado el valor máximo de probabilidad a posteriori para cada valor de q . El máximo valor de probabilidad a posteriori, para cada posible valor de observación, se realza en negro en la Figura 4.36. La región de decisión de un símbolo estará formada por el conjunto de valores de observación para los que la probabilidad a posteriori de ese símbolo es la mayor de todas. La región de decisión de b_0 (o de \mathbf{a}_0), es el conjunto de puntos para los que $p_{B|q}(b_0|q)$ es mayor que las probabilidades a posteriori de los otros 3 símbolos; la región de decisión de b_1 (o de \mathbf{a}_1), es el conjunto de puntos para los que $p_{B|q}(b_1|q)$ es mayor que las probabilidades a posteriori de los otros 3 símbolos, y así sucesivamente. En la figura se ilustra esta forma de obtener las regiones de decisión para la constelación del ejemplo.

La expresión analítica de las probabilidades *a posteriori* puede obtenerse mediante la regla de Bayes

$$p_{B|q}(b_j|\mathbf{q}_0) = \frac{p_B(b_j) \cdot f_{q|B}(\mathbf{q}_0|b_j)}{f_q(\mathbf{q}_0)}$$

Teniendo en cuenta que $B = b_j$ implica que $\mathbf{A} = \mathbf{a}_j$ y viceversa,

$$f_{q|B}(\mathbf{q}_0|b_j) = f_{q|\mathbf{A}}(\mathbf{q}_0|\mathbf{a}_j)$$

Sustituyendo esta distribución en la expresión de las probabilidades *a posteriori*, el criterio MAP se reduce a encontrar el símbolo b_i que cumple

$$\frac{p_B(b_i) \cdot f_{q|\mathbf{A}}(\mathbf{q}_0|\mathbf{a}_i)}{f_q(\mathbf{q}_0)} > \frac{p_B(b_j) \cdot f_{q|\mathbf{A}}(\mathbf{q}_0|\mathbf{a}_j)}{f_q(\mathbf{q}_0)} \quad j = 0, \dots, M - 1, j \neq i.$$

Como $f_q(\mathbf{q}_0)$ es una cantidad no negativa independiente del símbolo transmitido, esta condición es equivalente a la condición

$$p_B(b_i) \cdot f_{q|\mathbf{A}}(\mathbf{q}_0|\mathbf{a}_i) > p_B(b_j) \cdot f_{q|\mathbf{A}}(\mathbf{q}_0|\mathbf{a}_j) \quad j = 0, \dots, M - 1, j \neq i,$$

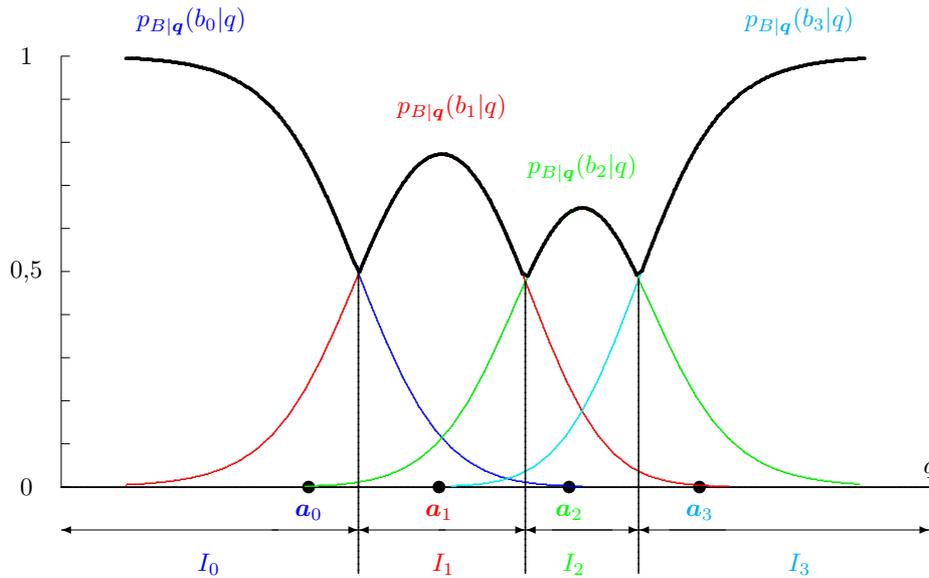


Figura 4.36: Obtención de las regiones de decisión mediante el criterio de máximo a posteriori para un ejemplo con una constelación de 4 símbolos en un espacio unidimensional.

o la expresión equivalente

$$p_A(\mathbf{a}_i) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_i) > p_A(\mathbf{a}_j) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_j) \quad j = 0, \dots, M - 1, j \neq i.$$

A continuación se muestra un ejemplo que ilustra la aplicación de este criterio bajo las condiciones analizadas en la Sección 4.4.4, i.e. transmisión sobre un canal gaussiano. Recuerde que en ese caso las distribuciones condicionales de la observación eran gaussianas de la dimensión del espacio de señales, N , de media el símbolo transmitido y varianza $N_0/2$ en todas las direcciones del espacio

$$f_{q|A}(\mathbf{q}|\mathbf{a}_i) = \mathcal{N}^N \left(\mathbf{a}_i, \frac{N_0}{2} \right).$$

Ejemplo

Se considera un sistema binario ($M = 2$) que emplea un espacio de señal unidimensional ($N = 1$). En este caso tanto \mathbf{A} como \mathbf{q} son escalares, y las probabilidades de símbolos son $p_B(b_1) = \frac{2}{3}$ y $p_B(b_0) = \frac{1}{3}$. Es decir, el símbolo b_1 es el doble de probable. La figura 4.37 muestra las curvas de $p_B(b_1) \cdot f_{q|A}(\mathbf{q}|\mathbf{a}_1)$ y $p_B(b_0) \cdot f_{q|A}(\mathbf{q}|\mathbf{a}_0)$, respectivamente cuando la distribución de \mathbf{q} condicionada a cada símbolo es gaussiana de media el símbolo transmitido y varianza $N_0/2$.

Si el valor $\mathbf{q} = \mathbf{q}_0$ en el receptor es mayor que \mathbf{q}_2 , el punto en que se cortan las curvas, entonces se cumple

$$p_B(b_0) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_0) > p_B(b_1) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_1),$$

por lo que tales puntos formarán parte de la región de decisión I_0 , mientras que si es menor que \mathbf{q}_2

$$p_B(b_1) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_1) > p_B(b_0) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_0),$$

por lo que tales puntos formarán parte de la región de decisión I_1 . Por tanto, las regiones de decisión son

$$I_0 = (\mathbf{q}_2, \infty)$$

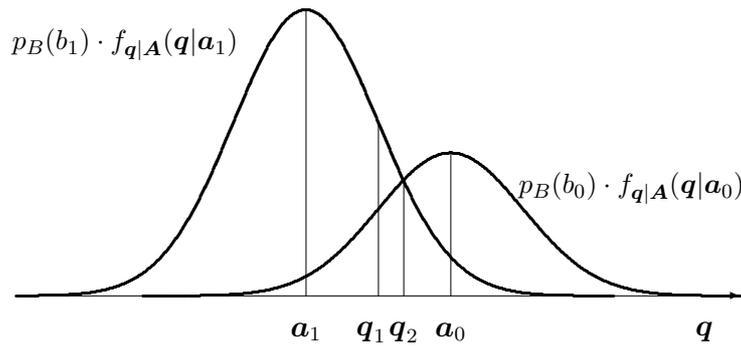


Figura 4.37: Aplicación del criterio MAP con símbolos no equiprobables.

$$I_1 = (-\infty, \mathbf{q}_2)$$

El valor que define el cambio de región de decisión es el punto $\mathbf{q} = \mathbf{q}_2$ en que se cortan ambas curvas, es decir

$$p_B(b_1) \cdot f_{q|A}(\mathbf{q}_2|\mathbf{a}_1) = p_B(b_0) \cdot f_{q|A}(\mathbf{q}_2|\mathbf{a}_0).$$

Naturalmente, el decisor queda unívocamente determinado mediante la relación

$$\mathbf{q} \in I_i \Leftrightarrow \hat{B} = b_i.$$

En un gran número de sistemas, todos los símbolos tienen la misma probabilidad de ser transmitidos. Bajo la hipótesis de símbolos equiprobables, $p_B(b_j) = 1/M$, las probabilidades *a priori* de los símbolos pasan a ser irrelevantes en la comparación, por lo que la regla de decisión para un valor de observación $\mathbf{q} = \mathbf{q}_0$ se reduce a encontrar el símbolo b_i que cumple

$$f_{q|A}(\mathbf{q}_0|\mathbf{a}_i) > f_{q|A}(\mathbf{q}_0|\mathbf{a}_j) \quad j = 0, \dots, M - 1, j \neq i,$$

es decir, el símbolo que maximiza la función $f_{q|A}(\mathbf{q}_0|\mathbf{a}_i)$. Esta particularización del criterio MAP para símbolos equiprobables recibe el nombre de *criterio de máxima verosimilitud* o criterio ML (del inglés “*maximum likelihood*”). El nombre viene de la propia función $f_{q|A}(\mathbf{q}_0|\mathbf{a}_i)$, denominada función de verosimilitud ya que representa una medida de certidumbre o verosimilitud de que la hipótesis cierta sea que el símbolo transmitido fuese \mathbf{a}_i cuando la entrada del decisor es $\mathbf{q} = \mathbf{q}_0$.

Ejemplo

Se considera el mismo sistema binario ($M = 2$) en un espacio de señal unidimensional ($N = 1$) del ejemplo anterior, pero en este caso con símbolos equiprobables. La figura 4.38 muestra las curvas de $p_B(b_1) \cdot f_{q|A}(\mathbf{q}|\mathbf{a}_1)$ y $p_B(b_0) \cdot f_{q|A}(\mathbf{q}|\mathbf{a}_0)$ respectivamente. Ahora, eliminar las probabilidades a priori de cada símbolo en la comparación no modifica la representación más que en un factor de escala, ya que al ser los símbolos equiprobables $p_B(b_0) = p_B(b_1)$.

Si el valor de $\mathbf{q} = \mathbf{q}_0$ en el receptor es mayor que \mathbf{q}_1 , entonces se cumple

$$p_B(b_0) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_0) > p_B(b_1) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_1),$$

o lo que es lo mismo

$$f_{q|A}(\mathbf{q}_0|\mathbf{a}_0) > f_{q|A}(\mathbf{q}_0|\mathbf{a}_1),$$

por lo que tales puntos formarán parte de la región de decisión I_0 , mientras que si es menor que \mathbf{q}_1

$$p_B(b_1) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_1) > p_B(b_0) \cdot f_{q|A}(\mathbf{q}_0|\mathbf{a}_0),$$

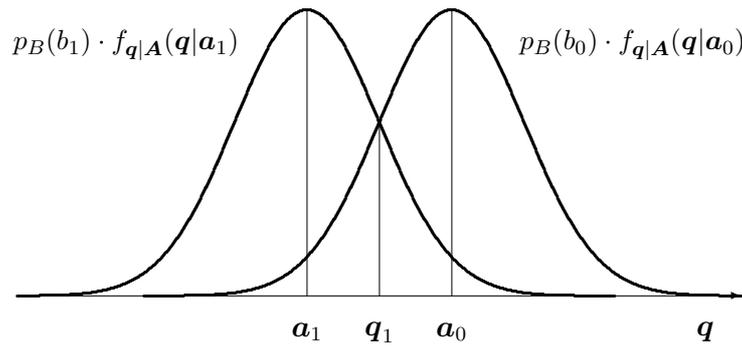


Figura 4.38: Aplicación del criterio MAP con símbolos equiprobables, en cuyo caso se particulariza en el criterio de máxima verosimilitud.

o lo que es lo mismo

$$f_{q|A}(q_0|a_1) > f_{q|A}(q_0|a_0),$$

por lo que tales puntos formarán parte de la región de decisión I_1 . Esto nos lleva a ver que las regiones de decisión son

$$I_0 = (q_1, \infty)$$

$$I_1 = (-\infty, q_1).$$

Intuitivamente se ve que en este caso que como ambas funciones gaussianas tienen la misma varianza y factor de escala, el umbral es el punto medio entre sus centros, $q_1 = \frac{a_0 + a_1}{2}$.

Cuando el espacio de señales es unidimensional, como en los dos ejemplos vistos anteriormente, los valores que separan las regiones de decisión se suelen denominar *umbrales de decisión*. Si el espacio de señales es multidimensional, las curvas ($N = 2$) o superficies que separan las regiones de decisión se denominan *fronteras de decisión*. En los dos ejemplos anteriores el umbral de decisión es q_2 para el primer ejemplo, y q_1 , el punto medio entre los símbolos a_0 y a_1 , en el segundo ejemplo, cuando los símbolos son equiprobables. Se puede ver que cuando los símbolos no se transmiten con la misma probabilidad, el decisor óptimo tiende a aumentar la región de decisión de los símbolos que se transmiten con mayor probabilidad, lo que parece bastante razonable de forma intuitiva (además de verse refrendado por los desarrollos analíticos pertinentes). Al modificar las probabilidades *a priori* para dar más probabilidad al símbolo b_1 , el umbral se ha desplazado de q_1 a q_2 , aumentando la región de decisión del símbolo más probable con respecto al caso de símbolos igualmente probables.

Para finalizar, se particularizará la regla de decisión para el criterio ML (o MAP con símbolos equiprobables) para el caso en que las distribuciones condicionales de la observación son gaussianas, como en el caso visto en la Sección 4.4.4. Introduciendo la función densidad de probabilidad a la entrada del receptor en la formulación del decisor ML se obtiene que la decisión para una observación $\mathbf{q} = \mathbf{q}_0$ será el símbolo b_i (o equivalentemente \mathbf{q}_0 pertenece a I_i) si se cumple

$$\frac{1}{(\pi N_0)^{N/2}} e^{-\frac{\|\mathbf{q}_0 - \mathbf{a}_i\|^2}{N_0}} > \frac{1}{(\pi N_0)^{N/2}} e^{-\frac{\|\mathbf{q}_0 - \mathbf{a}_j\|^2}{N_0}} \quad j = 0, \dots, M - 1, j \neq i.$$

Multiplicando ambos términos de la desigualdad por $(\pi N_0)^{N/2}$ se obtiene

$$e^{-\frac{\|\mathbf{q}_0 - \mathbf{a}_i\|^2}{N_0}} > e^{-\frac{\|\mathbf{q}_0 - \mathbf{a}_j\|^2}{N_0}} \quad j = 0, \dots, M - 1, j \neq i.$$

Teniendo en cuenta que la función exponencial es una función monótona creciente y, por tanto, cumple

$$e^a > e^b \Leftrightarrow a > b,$$

la expresión anterior es equivalente a decir

$$-\frac{\|\mathbf{q}_0 - \mathbf{a}_i\|^2}{N_0} > -\frac{\|\mathbf{q}_0 - \mathbf{a}_j\|^2}{N_0} \quad j = 0, \dots, M-1, j \neq i,$$

y multiplicando por N_0 y teniendo en cuenta el signo negativo se llega a la condición

$$\|\mathbf{q}_0 - \mathbf{a}_i\|^2 < \|\mathbf{q}_0 - \mathbf{a}_j\|^2 \quad j = 0, \dots, M-1, j \neq i.$$

Aplicando la definición de la norma de un vector

$$\|\mathbf{q}_0 - \mathbf{a}_i\|^2 = \sum_{k=0}^{N-1} |q_{0,k} - a_{i,k}|^2 = |d(\mathbf{q}_0, \mathbf{a}_i)|^2.$$

Por tanto, finalmente se establece que la regla de decisión se limita a escoger el símbolo más cercano al vector de observación $\mathbf{q} = \mathbf{q}_0$. Alternativamente, se puede decir que la región de decisión de un símbolo, I_i , estará formada por todos los puntos del espacio de \mathbf{q} que se encuentran más cerca del símbolo, \mathbf{a}_i , que de cualquier otro símbolo de la constelación. Un esquema que ilustra el decisor resultante de este criterio, que se denomina *criterio de mínima distancia euclídea*, se muestra en la Figura 4.39.

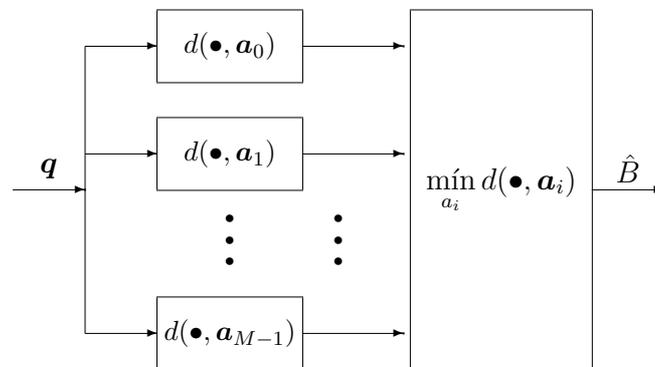


Figura 4.39: Estructura del decisor de mínima distancia euclídea.

En este punto es necesario realizar las siguientes precisiones:

- Este último desarrollo que nos ha llevado a formular el decisor ML como un decisor de mínima distancia euclídea está basado en una función densidad de probabilidad gaussiana a la entrada del decisor, y esta función densidad de probabilidad gaussiana viene dada por la naturaleza del ruido térmico presente habitualmente en canales de comunicaciones. Cuando el ruido que aparece en el canal no tenga una distribución gaussiana, como ocurre por ejemplo en algunos sistemas de comunicaciones basados en fibra óptica, el desarrollo del decisor óptimo conducirá a reglas de decisión diferentes.
- La definición de producto escalar que se adoptó en la Sección 4.2.2, no es la única posible, y se adoptó en su momento sin una justificación estricta. Esta definición da lugar a que la medida

de distancia sobre el espacio de Hilbert resultante sea la de distancia euclídea. Ahora es posible justificar que para el caso de estadística gaussiana para el ruido, esta definición resulta conveniente al permitir simplificar notablemente la regla de decisión para el caso de símbolos equiprobables, que por otra parte es el caso más frecuente en sistemas de comunicaciones digitales.

4.5.3. Cálculo de las probabilidades de error

En el apartado anterior se han obtenido las reglas de diseño de un decisor que permiten obtener la menor probabilidad de error. En esta sección se estudiará como evaluar dicha probabilidad de error en diferentes tipos de sistemas. Como se ha visto con anterioridad, en un sistema de comunicaciones las prestaciones vienen determinadas por la constelación transmitida (junto con la probabilidad con la que se transmite cada símbolo de la constelación), y son independientes de la base ortonormal que define el espacio de señales del sistema, siempre y cuando esta se elija de forma apropiada a las características del canal. Por ello, en esta sección tras plantear de forma general el problema de evaluar la probabilidad de error para un sistema, se particularizará el estudio para distintos tipos de constelaciones.

Cálculo exacto de la probabilidad de error de símbolo

La probabilidad de error de símbolo de un sistema de comunicaciones digitales se define cómo la probabilidad de que la decisión de un cierto símbolo en un instante n no sea correcta

$$P_e = P(\hat{B}[n] \neq b_i | B[n] = b_i).$$

El cálculo de esta probabilidad de error se obtiene promediando las probabilidades de error condicionales, es decir, las probabilidades de error condicionadas a que se ha transmitido cada uno de los símbolos de la constelación. Por notación, y teniendo en cuenta la relación unívoca entre un símbolo y la representación vectorial de su señal asociada, se representarán estas probabilidades condicionales como

$$P_{e|B[n]=b_i} = P_{e|\mathbf{A}[n]=\mathbf{a}_i} \equiv P_{e|\mathbf{a}_i}.$$

La probabilidad de error de símbolo se obtendrá promediando estas probabilidades condicionales teniendo en cuenta la probabilidad con la que se transmitirá cada símbolo

$$P_e = \sum_{i=0}^{M-1} p_{\mathbf{A}}(\mathbf{a}_i) \cdot P_{e|\mathbf{a}_i}.$$

Por tanto, el problema se reduce al cálculo de las probabilidades de error condicionales. Para ello sólo hay que analizar bajo qué circunstancias se produce un error cuando se ha transmitido un símbolo, $\mathbf{A}[n] = \mathbf{a}_i$, y evaluar la probabilidad de que se produzcan tales circunstancias. De nuevo, por simplicidad en la notación se obviará el índice temporal n (lo que es posible debido a la independencia de los símbolos y observaciones en distintos instantes temporales). Cuando se ha transmitido el símbolo $\mathbf{A} = \mathbf{a}_i$, se produce una decisión errónea cuando se decide $\hat{B} = b_j \neq b_i$, y esto sucede cuando el valor de observación no está en la región de decisión del símbolo, I_i . Esto significa que la probabilidad de error condicional para \mathbf{a}_i es la probabilidad de que habiendo transmitido dicho símbolo, el valor de observación no esté en la región de decisión I_i . Como la

distribución de la observación cuando se transmite \mathbf{a}_i es $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i)$, esta probabilidad se obtiene como

$$P_{e|\mathbf{a}_i} = \int_{\mathbf{q} \notin I_i} f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i) d\mathbf{q}.$$

Comentar finalmente que para poder obtener la probabilidad de error de símbolo que tiene un sistema, como se deduce claramente de las expresiones anteriores, hay que conocer los siguientes parámetros:

- probabilidades a priori de cada símbolo, $p_B(b_i) = p_{\mathbf{A}}(\mathbf{a}_i)$;
- regiones de decisión de cada símbolo, I_i ;
- distribuciones condicionales de la observación para cada símbolo, $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i)$;

en todos los casos para $i \in \{0, 1, \dots, M-1\}$. A continuación se realizará el cálculo de la probabilidad de error para distintos tipos de constelaciones. Por defecto, si no se indica de forma explícita lo contrario, se asumirán probabilidades a priori idénticas (símbolos equiprobables), y distribuciones condicionales de la observación gaussianas, como las obtenidas en la transmisión sobre un canal gaussiano (ver Sección 4.4.4).

Constelación binaria ($M = 2$) en espacio unidimensional ($N = 1$)

Como ejemplo inicial se considera una constelación de dos símbolos en un espacio unidimensional, el símbolo \mathbf{a}_0 tiene como coordenada $+A$ y el símbolo \mathbf{a}_1 , $-A$. Bajo la asunción de símbolos equiprobables y transmisión sobre canal gaussiano, la región de decisión I_1 estará formada por todos los valores de \mathbf{q} más cercanos a \mathbf{a}_1 que a \mathbf{a}_0 y al revés para la región de decisión I_0 . El umbral de decisión es, por tanto, cero, y las regiones de decisión

$$\text{Umbral } q_u = 0 \rightarrow I_0 = [0, \infty), I_1 = (-\infty, 0),$$

tal y como se muestra en la Figura 4.40.

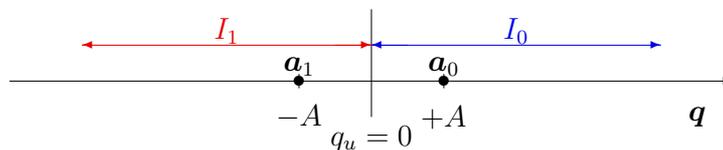


Figura 4.40: Constelación y regiones de decisión para el ejemplo de constelación binaria en espacio unidimensional.

Calculemos ahora las probabilidades de error condicionales, y empecemos por el símbolo \mathbf{a}_0 . En este caso, la distribución condicional de la observación para este símbolo es la que se muestra en la Figura 4.41.

Se trata de una distribución $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_0)$ gaussiana de media $+A$, la representación vectorial del símbolo transmitido, y varianza $N_0/2$. La probabilidad de error condicional para \mathbf{a}_0 es la integral

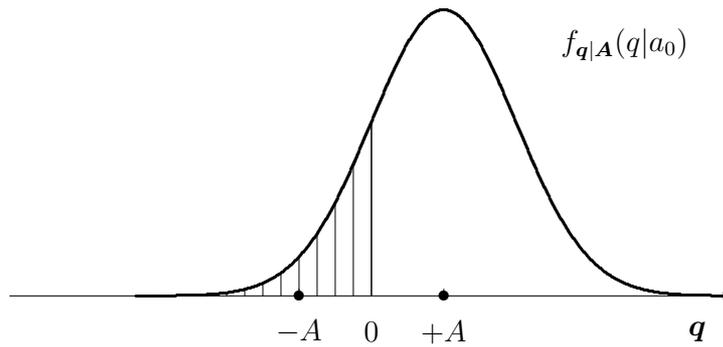


Figura 4.41: Distribución condicional de la observación para el símbolo \mathbf{a}_0 .

de esta distribución condicional fuera de la región de decisión de \mathbf{a}_0 , que en este caso es el área resaltada en la figura sobre la distribución

$$P_{e|\mathbf{a}_0} = \int_{q \notin I_0} f_{q|\mathbf{A}}(\mathbf{q}|\mathbf{a}_0) dq = Q\left(\frac{A}{\sqrt{N_0/2}}\right).$$

Como cuando se transmite el símbolo $\mathbf{A} = \mathbf{a}_i$ la distribución de la observación es $f_{q|\mathbf{A}}(\mathbf{q}|\mathbf{a}_0)$, y el error en la decisión se produce cuando la observación \mathbf{q} toma valores que caen fuera de I_0 (en este caso binario eso significa que cae en I_1), la probabilidad de que esto ocurra se calcula integrando la distribución de \mathbf{q} fuera de I_0 .

Si se calcula la probabilidad de error condicional para el símbolo \mathbf{a}_1 , en este caso la distribución condicional de la observación para este símbolo es la que se muestra en la Figura 4.42.

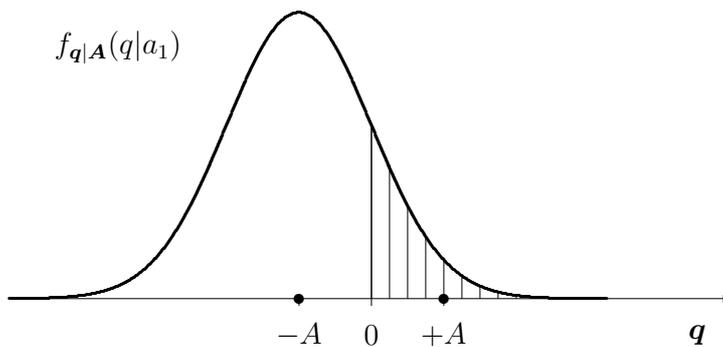


Figura 4.42: Distribución condicional de la observación para el símbolo \mathbf{a}_1 .

Se trata de una distribución $f_{q|\mathbf{A}}(\mathbf{q}|\mathbf{a}_1)$ gaussiana de media $-A$ y varianza $N_0/2$. La probabilidad de error condicional es la integral de esta distribución condicional fuera de su región de decisión, que en este caso es el área resaltada en la figura sobre la distribución

$$P_{e|\mathbf{a}_1} = \int_{q \notin I_1} f_{q|\mathbf{A}}(\mathbf{q}|\mathbf{a}_1) dq = Q\left(\frac{A}{\sqrt{N_0/2}}\right).$$

Ahora, como cuando se transmite el símbolo $\mathbf{A} = \mathbf{a}_i$ la distribución de la observación es $f_{q|\mathbf{A}}(\mathbf{q}|\mathbf{a}_1)$, y el error en la decisión se produce cuando la observación \mathbf{q} toma valores que caen fuera de I_1

(en este caso binario eso significa que cae en I_0), la probabilidad de que esto ocurra se calcula integrando la distribución de q fuera de I_1 .

Una vez calculadas las probabilidades de error condicionales, la probabilidad de error de símbolo se obtiene promediandolas, lo que en este caso significa

$$P_e = \frac{1}{2}P_{e|a_0} + \frac{1}{2}P_{e|a_1} = \frac{1}{2} \int_{q \notin I_0} f_{q|A}(q|a_0) dq + \frac{1}{2} \int_{q \notin I_1} f_{q|A}(q|a_1) dq$$

La Figura 4.43 muestra la interpretación gráfica del significado de esta probabilidad de error, y cómo la modificación del umbral de decisión haría aumentar dicha probabilidad de error.

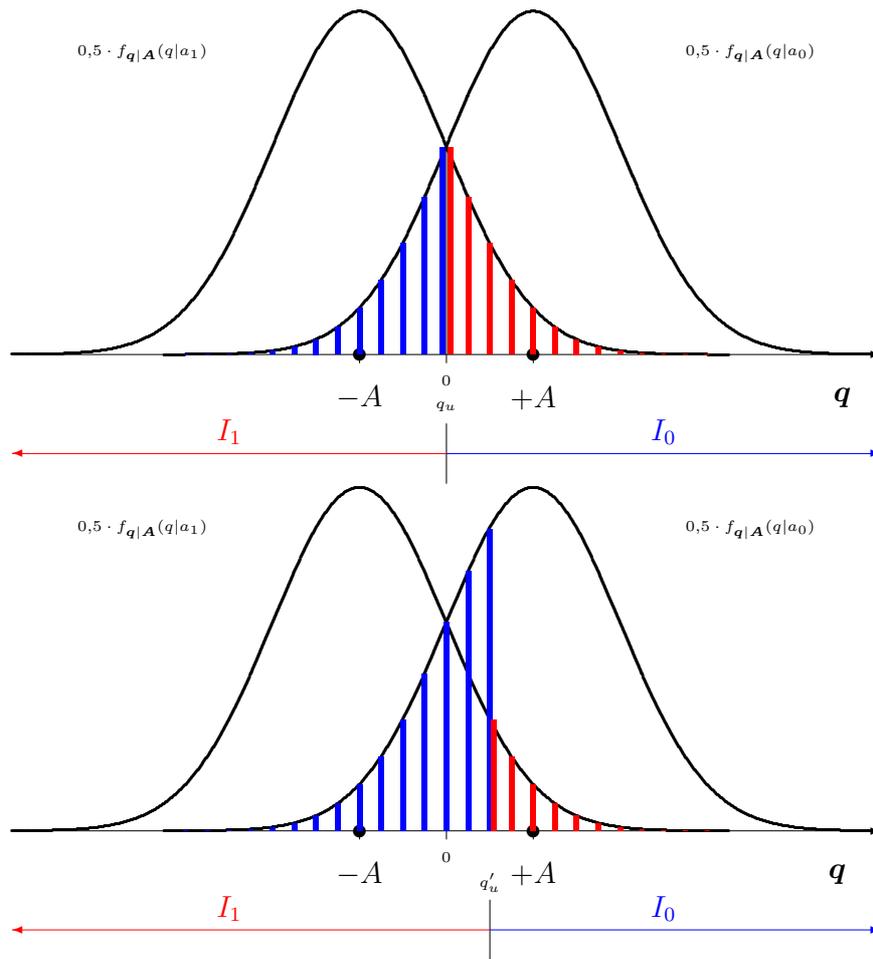


Figura 4.43: Interpretación gráfica de la probabilidad de error de símbolo para el decisor óptimo y para otro decisor.

De este ejemplo se pueden extrapolar de forma inmediata el mismo resultado cuando los valores de la representación vectorial de los dos símbolos es otra arbitraria, a_0 y a_1 . Independientemente de los valores de los símbolos, para símbolos equiprobables y distribuciones condicionales gaussianas el umbral está en el punto medio entre ambos

$$q_u = \frac{a_0 + a_1}{2},$$

de modo que la distancia de cada símbolo al umbral, que define el argumento de la función $Q(x)$ para evaluar la integral de la distribución gaussiana, es la mitad de la distancia entre los dos

símbolos

$$d(\mathbf{a}_0, \mathbf{q}_u) = d(\mathbf{a}_1, \mathbf{q}_u) = \frac{d(\mathbf{a}_0, \mathbf{a}_1)}{2}$$

por lo que las probabilidades de error condicionales son iguales y la probabilidad de error total es

$$P_e = Q\left(\frac{d(\mathbf{a}_0, \mathbf{a}_1)}{2\sqrt{N_0/2}}\right).$$

Decisor M -ario en espacio unidimensional

Para ilustrar este caso se considera un ejemplo de un sistema con una constelación de $M = 4$ símbolos que se transmiten con la misma probabilidad sobre un canal gaussiano. Las coordenadas de la representación vectorial de los símbolos son

$$\mathbf{a}_0 = -3, \mathbf{a}_1 = -1, \mathbf{a}_2 = +1, \mathbf{a}_3 = +3$$

En este caso, al poder aplicarse el criterio de mínima distancia euclídea para el diseño del decisor, los umbrales del decisor aparecen en los puntos medios entre cada dos símbolos de la constelación

$$q_{u1} = -2, q_{u2} = 0, q_{u3} = +2,$$

con lo que las regiones de decisión son

$$I_0 = (-\infty, -2], I_1 = (-2, 0], I_2 = (0, +2], I_3 = (+2, +\infty)$$

tal y como se ilustra en la Figura 4.44.

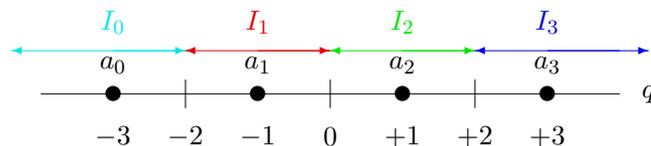


Figura 4.44: Constelación y regiones de decisión para el ejemplo de una constelación de 4 símbolos en espacio unidimensional.

La probabilidad de error condicional para el símbolo \mathbf{a}_0 , requiere la utilización de la distribución condicional de la observación para este símbolo, que es la que se muestra en la Figura 4.45.

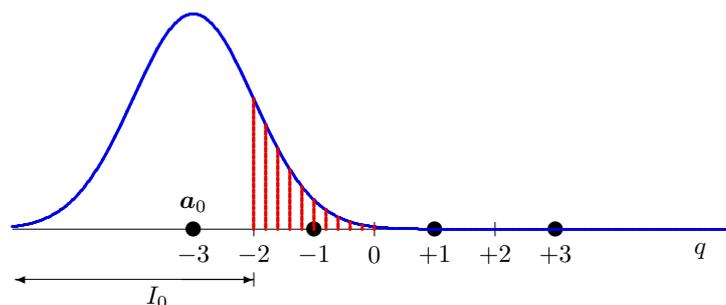


Figura 4.45: Distribución condicional de la observación para el símbolo \mathbf{a}_0 .

Se trata de una distribución $f_{q|A}(q|\mathbf{a}_0)$ gaussiana de media -3 , la representación vectorial del símbolo transmitido, y varianza $N_0/2$. La probabilidad de error condicional es la integral de esta distribución condicional fuera de su región de decisión, que en este caso es el área resaltada en la figura sobre la distribución

$$P_{e|a_0} = \int_{q \notin I_0} f_{q|A}(q|a_0) dq = Q\left(\frac{1}{\sqrt{N_0/2}}\right).$$

La probabilidad de error condicional para el símbolo \mathbf{a}_1 , se basa en la distribución condicional de la observación para este símbolo, que es la que se muestra en la Figura 4.46.

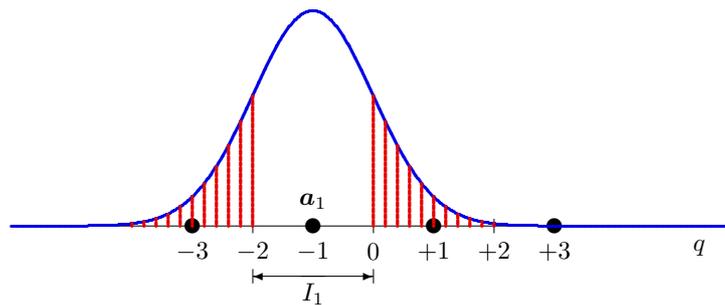


Figura 4.46: Distribución condicional de la observación para el símbolo \mathbf{a}_1 .

Se trata de una distribución $f_{q|A}(q|\mathbf{a}_1)$ gaussiana de media -1 , la representación vectorial del símbolo transmitido, y varianza $N_0/2$. La probabilidad de error condicional es la integral de esta distribución condicional fuera de su región de decisión, que en este caso es el área resaltada en la figura sobre la distribución

$$P_{e|a_1} = \int_{q \notin I_1} f_{q|A}(q|a_1) dq = 2Q\left(\frac{1}{\sqrt{N_0/2}}\right).$$

La distribución condicional de la observación para el símbolo \mathbf{a}_2 se muestra en la Figura 4.47.

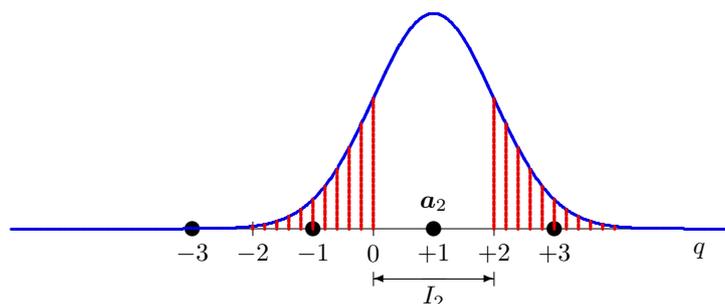


Figura 4.47: Distribución condicional de la observación para el símbolo \mathbf{a}_2 .

Se trata de una distribución $f_{q|A}(q|\mathbf{a}_2)$ gaussiana de media $+1$, la representación vectorial del símbolo transmitido, y varianza $N_0/2$. La probabilidad de error condicional es la integral de esta distribución condicional fuera de su región de decisión, que en este caso es el área resaltada en la figura sobre la distribución

$$P_{e|a_2} = \int_{q \notin I_2} f_{q|A}(q|a_2) dq = 2Q\left(\frac{1}{\sqrt{N_0/2}}\right).$$

Finalmente, para el símbolo \mathbf{a}_3 , la distribución condicional de la observación para este símbolo, que es la que se muestra en la Figura 4.48.

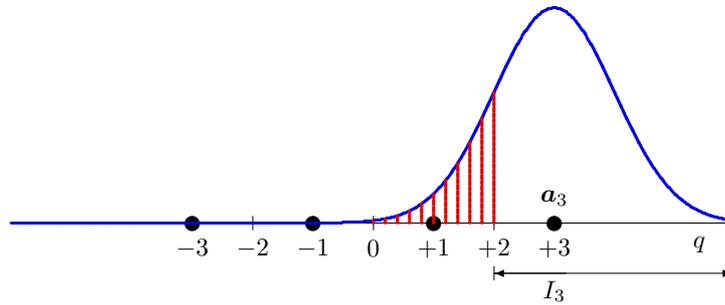


Figura 4.48: Distribución condicional de la observación para el símbolo \mathbf{a}_3 .

Se trata de una distribución $f_{q|\mathbf{A}}(\mathbf{q}|\mathbf{a}_3)$ gaussiana de media +3, la representación vectorial del símbolo transmitido, y varianza $N_0/2$. La probabilidad de error condicional es la integral de esta distribución condicional fuera de su región de decisión, que en este caso es el área resaltada en la figura sobre la distribución

$$P_{e|\mathbf{a}_3} = \int_{q \notin I_3} f_{q|\mathbf{A}}(q|\mathbf{a}_3) dq = Q\left(\frac{1}{\sqrt{N_0/2}}\right).$$

En este ejemplo, dada la simetría de las regiones de decisión, claramente se puede ver que

$$P_{e|\mathbf{a}_0} = P_{e|\mathbf{a}_3} \text{ y } P_{e|\mathbf{a}_1} = P_{e|\mathbf{a}_2},$$

lo que se podría haber aprovechado para simplificar el cálculo.

Una vez calculadas las probabilidades de error condicionales, la probabilidad de error de símbolo se obtiene promediándolas

$$P_e = \sum_{i=0}^{M-1} p_A(\mathbf{a}_i) P_{e|\mathbf{a}_i} = \frac{1}{4} \sum_{i=0}^{M-1} P_{e|\mathbf{a}_i} = \frac{3}{2} Q\left(\frac{1}{\sqrt{N_0/2}}\right).$$

Sistema binario ($M = 2$) en espacio multidimensional ($N > 1$)

Para ilustrar este caso se considera la siguiente constelación

$$\mathbf{a}_0 = \begin{bmatrix} A \\ 0 \end{bmatrix}, \quad \mathbf{a}_1 = \begin{bmatrix} 0 \\ A \end{bmatrix}.$$

En ella, los símbolos son vectores de dimensión 2. Para el caso por defecto de símbolos equiprobables y bajo transmisión sobre un canal gaussiano, se puede aplicar de nuevo el criterio de mínima distancia euclídea, con lo que la región de decisión de cada símbolo está formada por los puntos del espacio de \mathbf{q} (en este caso el plano bidimensional formado por las dos coordenadas, q_0 y q_1) que están más cerca de ese símbolo que del otro. La frontera de decisión que separa los puntos más cercanos de \mathbf{a}_0 de los de \mathbf{a}_1 es la recta $q_0 = q_1$. La región de decisión I_0 es el semiplano de la derecha, y la región de decisión I_1 es el semiplano de la izquierda

$$I_0 = \left\{ \mathbf{q} = \begin{bmatrix} q_0 \\ q_1 \end{bmatrix} \mid q_0 \geq q_1 \right\} \quad I_1 = \left\{ \mathbf{q} = \begin{bmatrix} q_0 \\ q_1 \end{bmatrix} \mid q_0 < q_1 \right\}$$

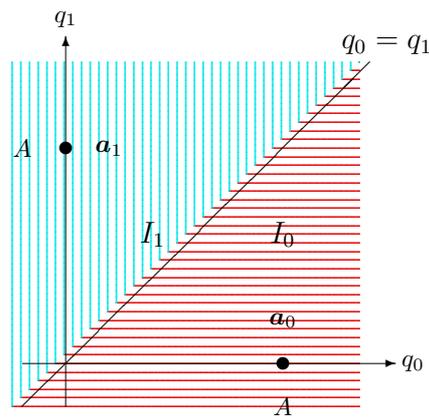


Figura 4.49: Frontera de decisión y regiones de decisión para el ejemplo de sistema binario en espacio bidimensional.

tal y como se representa en la Figura 4.49.

Si el símbolo transmitido es \mathbf{a}_0 la función de densidad de probabilidad de la entrada al receptor $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_0)$ es una gaussiana de media $\mathbf{a}_0 = \begin{bmatrix} A \\ 0 \end{bmatrix}$, componentes independientes y varianza de cada una de esas componentes igual a $N_0/2$, y para el símbolo \mathbf{a}_1 $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_1)$ es una gaussiana de media $\mathbf{a}_1 = \begin{bmatrix} 0 \\ A \end{bmatrix}$, de nuevo con componentes independientes y varianza $N_0/2$

$$f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_i) = \frac{1}{\pi N_0} e^{-\frac{\|\mathbf{q}-\mathbf{a}_i\|^2}{N_0}} = \frac{1}{\pi N_0} e^{-\frac{(q_0 - \sqrt{2T})^2 + q_1^2}{N_0}},$$

tal y como se muestra en la Figura 4.51 (en este caso se representan para $A = 1$, por un lado cada una por separado, y en la figura inferior ambas juntas, cada una dentro de su región de decisión).

Para calcular las probabilidades de error condicionales, $P_{e|\mathbf{a}_i}$, hay que integrar la función densidad de probabilidad de cada símbolo \mathbf{a}_i fuera de su región de decisión, I_i , en este caso el semiplano definido por $q_0 > q_1$ ó $q_0 < q_1$, respectivamente para \mathbf{a}_0 y \mathbf{a}_1 , tal y como se muestra en la Figura 4.51.

Como es bien sabido, la integral de una función gaussiana no tiene una expresión analítica definida. En el caso unidimensional se recurre a la función tabulada $Q(x)$, que se puede calcular de forma numérica, para obtener el valor de la integral de una distribución gaussiana unidimensional. Pero ahora, para espacios bidimensionales, no hay ninguna función tabulada que calcule integrales de distribuciones gaussianas sobre semiplanos. En lugar de resolver el cálculo de forma directa, aprovechando que si se tienen dos símbolos siempre se puede hacer pasar una recta por esos puntos, se va a hacer una transformación del sistema de coordenadas para convertir el problema en otro problema unidimensional equivalente. Para ello se realiza el siguiente cambio de variables:

$$\begin{aligned} q'_0 &= \frac{1}{\sqrt{2}}(q_0 - q_1), \\ q'_1 &= \frac{1}{\sqrt{2}}(q_0 + q_1 - A). \end{aligned}$$

La transformación gira la constelación 45° y desplaza el sistema resultante $\frac{A}{\sqrt{2}}$ hacia abajo, tal y como se ilustra en el Figura 4.52, de forma que los puntos de la constelación pasan ahora a

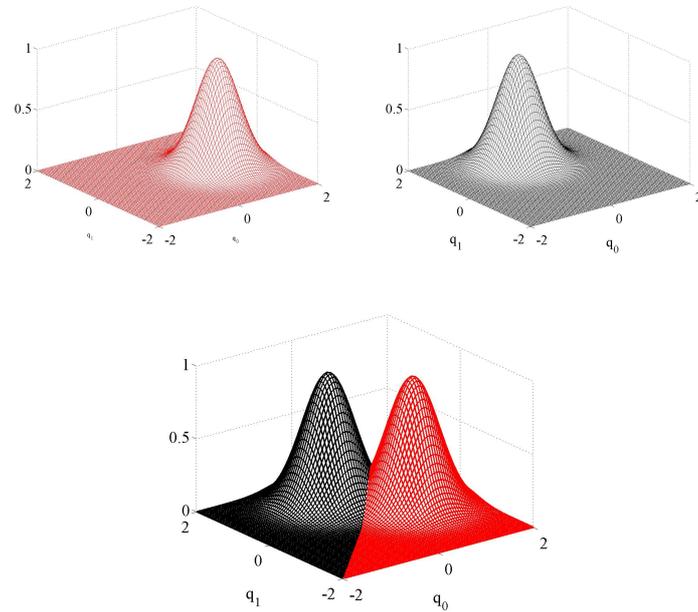


Figura 4.50: Funciones densidad de probabilidad condicionales para el ejemplo de sistema binario en espacio bidimensional. Cada función por separado, y ambas representadas cada una dentro de su región de decisión.

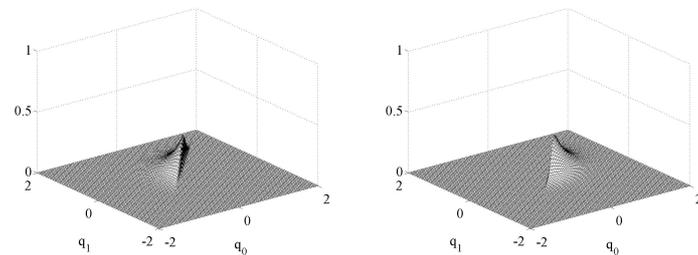


Figura 4.51: Interpretación gráfica de las probabilidades de error condicionales para el ejemplo de sistema binario en espacio bidimensional.

estar sobre un espacio unidimensional, ya que su coordenada sobre el segundo eje, q'_1 , es nula para ambos símbolos,

$$\mathbf{a}'_0 = \begin{bmatrix} +\frac{A}{\sqrt{2}} \\ 0 \end{bmatrix}, \mathbf{a}'_1 = \begin{bmatrix} -\frac{A}{\sqrt{2}} \\ 0 \end{bmatrix}.$$

Podríamos plantearnos eliminar por tanto la segunda coordenada y resolver el problema para la nueva constelación unidimensional resultante, tal y como se muestra en la Figura 4.53. Para poder hacer esto, es necesario que las componentes de ruido sobre los nuevos ejes sean independientes, y hay que comprobar la distribución del ruido en la coordenada q'_0 al realizar la transformación, y ver si sigue siendo una distribución gaussiana de media nula y varianza $N_0/2$. Vamos a ver qué estas condiciones sí se cumplen.

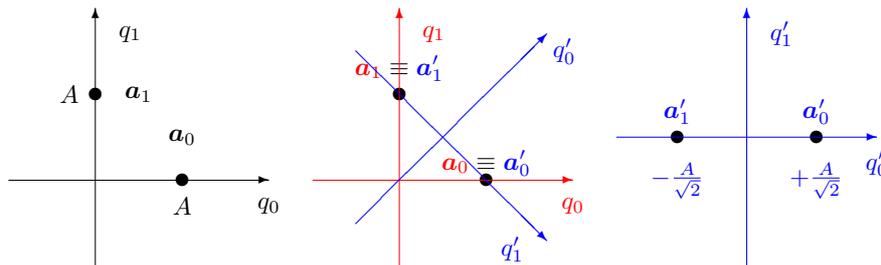


Figura 4.52: Ilustración de la transformación geométrica implícita en el cambio de coordenadas planteado.

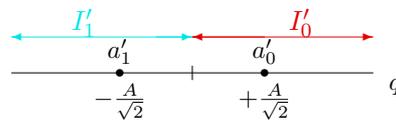


Figura 4.53: Regiones de decisión sobre el nuevo sistema de referencia tras el cambio de coordenadas.

Empezemos por recordar que cuando se transmite el símbolo $\mathbf{A} = \mathbf{a}_i$, se tiene que

$$q_0 = a_{i,0} + z_0, \quad q_1 = a_{i,1} + z_1.$$

Denotaremos la observación recibida como

$$q'_0 = a'_{i,0} + z'_0, \quad q'_1 = a'_{i,1} + z'_1.$$

Si el símbolo transmitido es \mathbf{a}_0 , q'_0 toma la forma

$$\begin{aligned} q'_0|_{\mathbf{A}=\mathbf{a}_0} &= \frac{1}{\sqrt{2}} ((a_{0,0} + z_0) - (a_{0,1} + z_1)) \\ &= \frac{1}{\sqrt{2}} ((A + z_0) - (0 + z_1)) \\ &= \underbrace{\frac{A}{\sqrt{2}}}_{a'_{0,0}} + \underbrace{\frac{1}{\sqrt{2}}(z_0 - z_1)}_{z'_0} \end{aligned}$$

y q'_1

$$\begin{aligned} q'_1|_{\mathbf{A}=\mathbf{a}_0} &= \frac{1}{\sqrt{2}} ((a_{0,0} + z_0) + (a_{0,1} + z_1) - A) \\ &= \frac{1}{\sqrt{2}} ((A + z_0) + (0 + z_1) - A) \\ &= \underbrace{\frac{1}{\sqrt{2}}(z_0 + z_1)}_{z'_1} \end{aligned}$$

De forma análoga, si el símbolo transmitido es \mathbf{a}_1 , q'_0 y q'_1 toman la forma

$$\begin{aligned} q'_0|_{\mathbf{A}=\mathbf{a}_1} &= \frac{1}{\sqrt{2}} ((a_{1,0} + z_0) - (a_{1,1} + z_1)) \\ &= \frac{1}{\sqrt{2}} ((0 + z_0) - (A + z_1)) \\ &= \underbrace{-\frac{A}{\sqrt{2}}}_{a'_{1,0}} + \underbrace{\frac{1}{\sqrt{2}}(z_0 - z_1)}_{z'_0} \end{aligned}$$

y

$$\begin{aligned} q'_1|_{\mathbf{A}=\mathbf{a}_1} &= \frac{1}{\sqrt{2}} ((a_{1,0} + z_0) + (a_{1,1} + z_1) - A) \\ &= \frac{1}{\sqrt{2}} ((0 + z_0) + (A + z_1) - A) \\ &= \underbrace{\frac{1}{\sqrt{2}}(z_0 + z_1)}_{z'_1} \end{aligned}$$

De las dos componentes, sólo q'_0 contiene información sobre el símbolo transmitido; q'_1 contiene ruido. Esto implica que se ha reducido la dimensión del espacio de señal de 2 a 1. Las coordenadas de los nuevos elementos de la constelación respecto a q'_0 , que denotaremos a'_0 y a'_1 , son

$$\begin{aligned} a'_0 &= +\frac{A}{\sqrt{2}}, \\ a'_1 &= -\frac{A}{\sqrt{2}}, \end{aligned}$$

es decir, exactamente las mismas que en la constelación unidimensional considerada anteriormente.

Respecto al ruido, es fácil demostrar que los términos z'_0 y z'_1 son independientes. El primero es proporcional a $z_0 - z_1$ y el segundo a $z_0 + z_1$, ya que la suma (o resta) de dos variables aleatorias con función densidad de probabilidad gaussiana es otra variable aleatoria gaussiana y

$$E[(z_0 - z_1)(z_0 + z_1)] = E[z_0^2 - z_1^2] = \frac{N_0}{2} - \frac{N_0}{2} = 0.$$

Esto implica que el valor de q'_1 es irrelevante para la decisión y que q'_0 es un estadístico suficiente para la detección.

Para determinar la probabilidad de error debemos conocer los estadísticos de la componente de ruido, $\frac{1}{\sqrt{2}}(z_0 - z_1)$, ya que por ahora sólo sabemos que es gaussiana. El valor medio es

$$E[z'_0] = E\left[\frac{1}{\sqrt{2}}(z_0 - z_1)\right] = \frac{1}{\sqrt{2}}E[z_0] - \frac{1}{\sqrt{2}}E[z_1] = 0.$$

Y la varianza vale

$$\begin{aligned} \text{Var}(z'_0) &= E\left[\left(\frac{1}{\sqrt{2}}(z_0 - z_1)\right)^2\right] = \frac{1}{2}E[z_0^2] + \frac{1}{2}E[z_1^2] - E[z_0z_1] \\ &= \frac{1}{2}\frac{N_0}{2} + \frac{1}{2}\frac{N_0}{2} - 0 \\ &= \frac{N_0}{2}. \end{aligned}$$

Finalmente, si la nueva constelación es la misma que en el caso unidimensional y los estadísticos del ruido también (media cero y varianza $N_0/2$), también lo será la probabilidad de error,

$$P_e = Q\left(\frac{d(\mathbf{a}'_0, \mathbf{a}'_1)}{2\sqrt{N_0/2}}\right).$$

La distancia entre los símbolos es la misma que en la constelación original, $d(\mathbf{a}'_0, \mathbf{a}'_1) = d(\mathbf{a}_0, \mathbf{a}_1)$, por lo que es indistinto realizar la medida en el espacio original que en el transformado. De hecho, el transformado ni siquiera hay que calcularlo para evaluar la probabilidad media de error en este caso.

El desarrollo realizado puede extenderse a cualquier constelación binaria en un espacio de dimensión arbitraria, ya que siempre es posible definir una dirección del espacio a través de la recta que pasa por los dos puntos. Así pues, de forma general, la probabilidad de error para sistemas binarios con símbolos equiprobables y con transmisión sobre un canal gaussiano es

$$P_e = Q\left(\frac{d(\mathbf{a}_0, \mathbf{a}_1)}{2\sqrt{N_0/2}}\right),$$

independientemente de la dimensión del espacio de señales.

Decisor M -ario en espacio multidimensional

El caso general de cálculo de la probabilidad media de error en constelaciones multidimensionales es un problema complejo por las formas que pueden tomar las regiones de decisión. Para constelaciones binarias siempre es posible transformar el problema en unidimensional simplificando así la resolución de las integrales necesarias en el cálculo de la probabilidad de error de cada símbolo. Sin embargo, para constelaciones de más de dos símbolos esto no es siempre posible. Como ejemplo, baste considerar las constelaciones de la Figura 4.54, que presenta dos ejemplos de constelaciones bidimensionales de uso común en sistemas de comunicaciones y sus regiones de decisión asociadas. En algunos casos será posible calcular la probabilidad de error exacta, y en otros casos esto no será posible de forma analítica, y habrá que recurrir a cálculos numéricos, a aproximaciones o a cotas de la probabilidad de error.

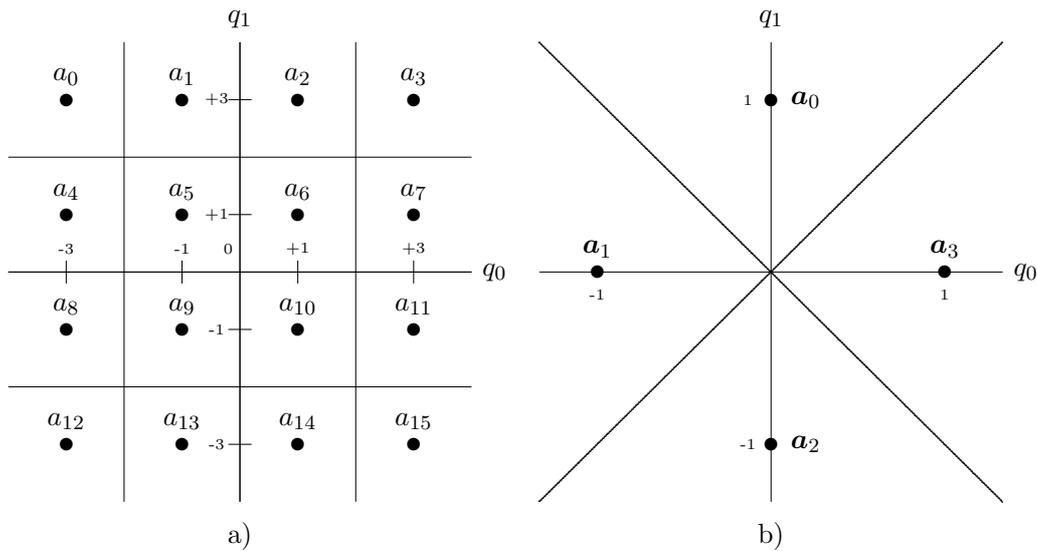


Figura 4.54: Ejemplos de constelaciones bidimensionales y sus fronteras de decisión.

Para sistemas con constelaciones como las de la figura es aún posible calcular de manera sencilla la probabilidad media de error. Veremos que esto es posible cuando las fronteras de las regiones de decisión forman una retícula sobre el espacio, como ocurre con las dos constelaciones de la figura.

Por ejemplo, primero se considera la constelación de la Figura 4.54 a). Veamos cómo se calcularía la probabilidad de error condicional para uno de los símbolos, por ejemplo \mathbf{a}_6 . Este símbolo tiene coordenadas

$$\mathbf{a}_6 = \begin{bmatrix} +1 \\ +1 \end{bmatrix},$$

por lo que la distribución condicional de la observación, $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_6)$, es una distribución gaussiana bidimensional de media \mathbf{a}_6 y varianzas $\sigma^2 = N_0/2$, tal y como se muestra en la Figura 4.55.

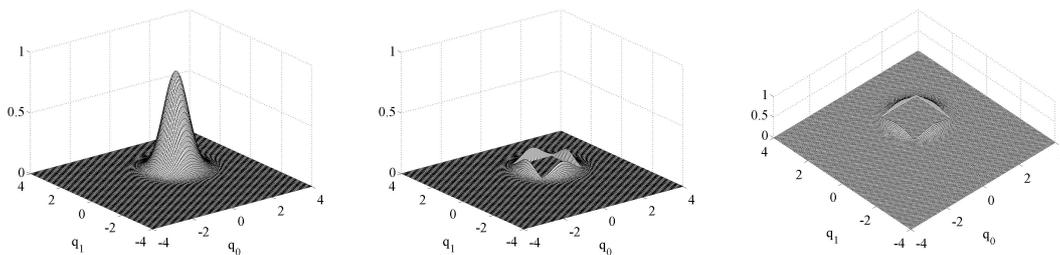


Figura 4.55: Distribución condicional de la observación para el símbolo \mathbf{a}_6 , $f_{\mathbf{q}|\mathbf{A}}(\mathbf{q}|\mathbf{a}_6)$: distribución completa, y distribución fuera de su región de decisión (dos vistas desde diferentes perspectivas).

La región de decisión de este símbolo es una cuadrícula, que en este caso, al estar alineada con los dos ejes de la representación se puede escribir como dos condiciones independientes (una por cada eje) que deben darse simultáneamente; en este caso

$$0 \leq q_0 < 2 \text{ y } 0 \leq q_1 < 2.$$

No hay expresiones analíticas para calcular directamente la integral de una gaussiana fuera de un cuadrado o un rectángulo, pero se puede transformar este problema 2D en dos problemas

unidimensionales acoplados. Cuando la región de decisión se puede parametrizar en dos condiciones independientes, una por cada dimensión del espacio, que han de cumplirse simultáneamente, la probabilidad de error condicional de un símbolo se puede escribir como

$$P_{e|\mathbf{a}_i} = 1 - P_{a|\mathbf{a}_i} = 1 - P_{a|a_{i,0}} \times P_{a|a_{i,1}} = 1 - [(1 - P_{e|a_{i,0}}) \times (1 - P_{e|a_{i,1}})],$$

donde se han seguido los siguientes pasos:

- La probabilidad de error condicional se puede escribir como 1 menos la probabilidad de acierto condicional, $P_{a|\mathbf{a}_i}$.
- La probabilidad condicional de acierto $P_{a|\mathbf{a}_i}$ se puede escribir como el producto de la probabilidad de acertar en cada una de las dos direcciones del espacio $P_{a|a_{i,0}} \times P_{a|a_{i,1}}$ (ya que la región de decisión se establece con dos condiciones independientes, una sobre cada dirección del espacio).
- La probabilidad de acierto en una de las direcciones del espacio ($P_{a|a_{i,0}}$ ó $P_{a|a_{i,1}}$) se puede escribir como 1 menos la probabilidad de error en esa dirección ($P_{e|a_{i,0}}$ ó $P_{e|a_{i,1}}$). Y esa probabilidad es igual a las probabilidades de error en espacios unidimensionales que ya se ha aprendido a calcular.

Esto permite calcular la probabilidad condicional de error para los símbolos si sus regiones de decisión forman una retícula alineada con los ejes del espacio de observación \mathbf{q} .

Veamos ahora con ejemplos numéricos cómo se calcularía la probabilidad de error total para la constelación anterior. Se pueden agrupar las regiones de decisión en tres tipos, dependiendo del número de regiones de decisión con las que limitan.

- Tipo 1: $\{I_0, I_3, I_{12}, I_{15}\}$
 - Una única frontera de decisión en cada dirección del espacio
- Tipo 2: $\{I_5, I_6, I_9, I_{10}\}$
 - Dos fronteras de decisión en cada dirección del espacio
- Tipo 3: $\{I_1, I_2, I_4, I_7, I_8, I_{11}, I_{13}, I_{14}\}$
 - Una frontera en una de las direcciones del espacio
 - Dos fronteras en la otra dirección

Todos los símbolos de un mismo tipo tienen la misma probabilidad condicional de error, ya que hay que integrar la distribución gaussiana centrada en el símbolo fuera de una región de las mismas dimensiones. Por tanto, es posible tomar un ejemplo de cada tipo, y extrapolar los resultados. Lo ejemplos elegidos pueden ser, por ejemplo: \mathbf{a}_0 (Tipo 1), \mathbf{a}_5 (Tipo 2), \mathbf{a}_7 (Tipo 3). De este modo, la probabilidad de error total será

$$\begin{aligned} P_e &= \sum_{i=0}^{M-1} p_{\mathbf{A}}(\mathbf{a}_i) \cdot P_{e|\mathbf{a}_i} \\ &= 4 \times \frac{1}{16} P_{e|\mathbf{a}_0} + 4 \times \frac{1}{16} P_{e|\mathbf{a}_5} + 8 \times \frac{1}{16} P_{e|\mathbf{a}_7}, \end{aligned}$$

que ahora se comprobará que vale

$$P_e = 3Q\left(\frac{1}{\sqrt{N_0/2}}\right) - \frac{9}{4}Q^2\left(\frac{1}{\sqrt{N_0/2}}\right).$$

Empezaremos el cálculo con el símbolo que ejemplifica las regiones de tipo 1, que es \mathbf{a}_0 . El procedimiento de cálculo para este caso, que se ilustra en la Figura 4.56, es el siguiente:

■ Dirección del espacio q_0

- Media de la distribución gaussiana 1-D: $a_{0,0} = -3$
- Región de decisión : $-\infty < q_0 < -2$

$$P_{a|a_{0,0}} = 1 - P_{e|a_{0,0}} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

■ Dirección del espacio q_1

- Media de la distribución gaussiana 1-D: $a_{0,1} = +3$
- Región de decisión : $+2 \leq q_1 < +\infty$

$$P_{a|a_{0,1}} = 1 - P_{e|a_{0,1}} = 1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

■ Probabilidad de error condicional

$$P_{e|a_0} = 1 - \left[1 - Q\left(\frac{1}{\sqrt{N_0/2}}\right)\right]^2 = 2Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q^2\left(\frac{1}{\sqrt{N_0/2}}\right)$$

A continuación realizaremos el cálculo para el símbolo que ejemplifica las regiones de tipo 2, que es \mathbf{a}_5 . El procedimiento de cálculo para este caso, que se ilustra en la Figura 4.57, es el siguiente:

■ Dirección del espacio q_0

- Media de la distribución gaussiana 1-D: $a_{5,0} = -1$
- Región de decisión : $-2 \leq q_0 < 0$

$$P_{a|a_{5,0}} = 1 - P_{e|a_{5,0}} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

■ Dirección del espacio q_1

- Media de la distribución gaussiana 1-D: $a_{5,1} = +1$
- Región de decisión : $0 \leq q_1 < +2$

$$P_{a|a_{5,1}} = 1 - P_{e|a_{5,1}} = 1 - 2Q\left(\frac{1}{\sqrt{N_0/2}}\right)$$

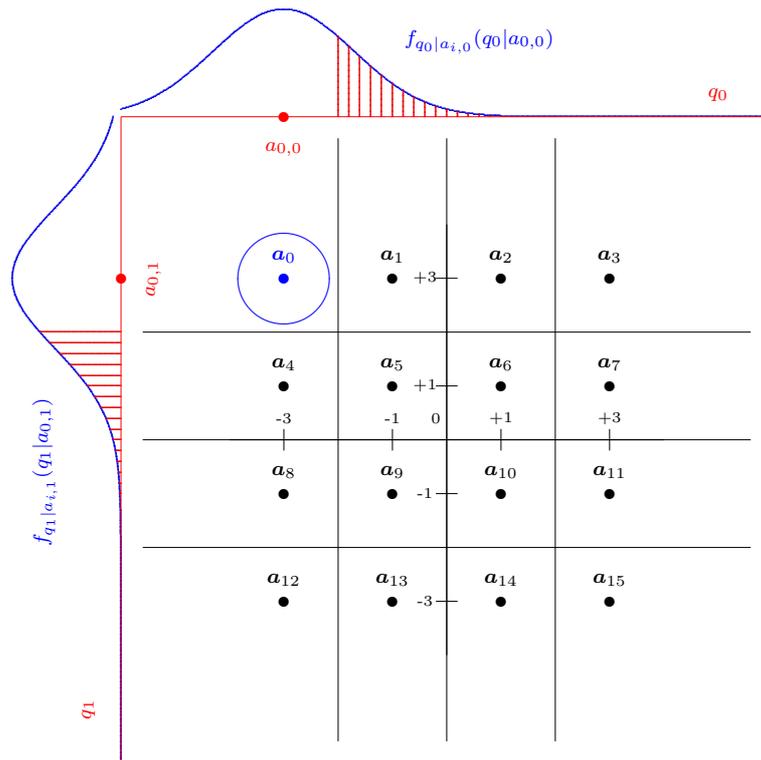


Figura 4.56: Distribuciones condicionales marginales para el símbolo a_0 .

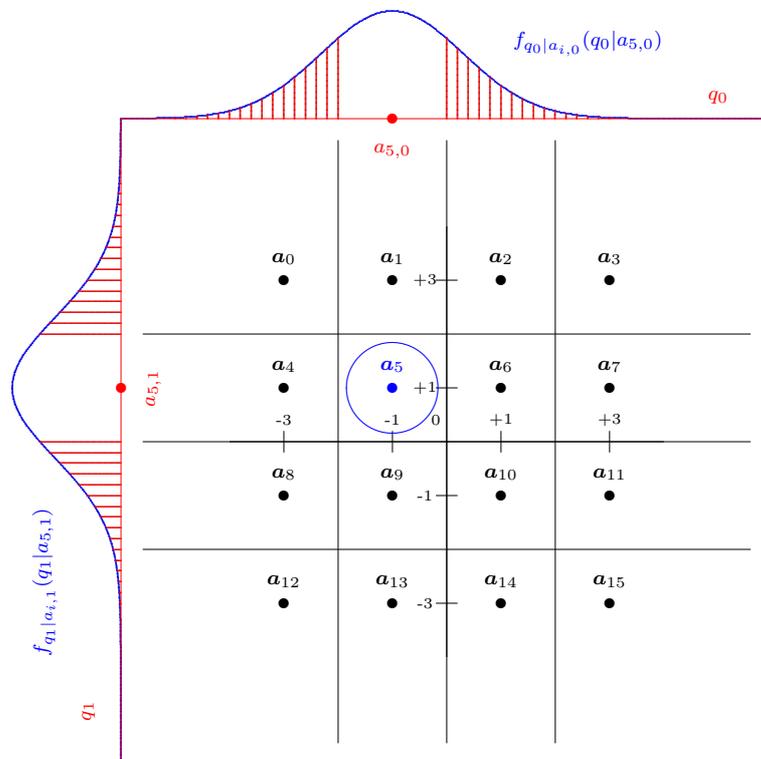


Figura 4.57: Distribuciones condicionales marginales para el símbolo a_5 .

- Probabilidad de error condicional

$$P_{e|a_5} = 1 - \left[1 - 2Q \left(\frac{1}{\sqrt{N_0/2}} \right) \right]^2 = 4Q \left(\frac{1}{\sqrt{N_0/2}} \right) - 4Q^2 \left(\frac{1}{\sqrt{N_0/2}} \right)$$

Finalizaremos con el cálculo para el símbolo que ejemplifica las regiones de tipo 3, que es a_7 . El procedimiento de cálculo para este caso, que se ilustra en la Figura 4.56, es el siguiente:

- Dirección del espacio q_0
 - Media de la distribución gaussiana 1-D: $a_{7,0} = +3$
 - Región de decisión : $+2 \leq q_0 < +\infty$

$$P_{a|a_{7,0}} = 1 - P_{e|a_{7,0}} = 1 - Q \left(\frac{1}{\sqrt{N_0/2}} \right)$$

- Dirección del espacio q_1
 - Media de la distribución gaussiana 1-D: $a_{7,1} = +1$
 - Región de decisión : $0 \leq q_1 < +2$

$$P_{a|a_{7,1}} = 1 - P_{e|a_{7,1}} = 1 - 2Q \left(\frac{1}{\sqrt{N_0/2}} \right)$$

- Probabilidad de error condicional

$$P_{e|a_7} = 1 - \left[1 - Q \left(\frac{1}{\sqrt{N_0/2}} \right) \right] \cdot \left[1 - 2Q \left(\frac{1}{\sqrt{N_0/2}} \right) \right] = 3Q \left(\frac{1}{\sqrt{N_0/2}} \right) - 2Q^2 \left(\frac{1}{\sqrt{N_0/2}} \right)$$

Finalmente, promediando las distribuciones condicionales para los 16 símbolos se obtiene el resultado final que se mostró anteriormente.

Se ha visto cómo se puede calcular la probabilidad de error para regiones de decisión que formen una retícula alineada con los ejes del espacio \mathbf{q} . El problema 2D puede convertirse en dos problemas 1D acoplados que pueden resolverse de forma independiente. Para la constelación de la Figura 4.54 b), las regiones de decisión forman una retícula, pero esta no está alineada con los ejes del espacio de \mathbf{q} , q_0 y q_1 en este ejemplo bidimensional. Sin embargo, es fácil ver que si se aplica un cambio de variables que produzca un giro de 45° , tal y como se hizo en el ejemplo de constelación binaria en espacio bidimensional, las regiones de decisión pasan a estar alineadas con los nuevos ejes y se puede aplicar el mismo procedimiento. En este caso, todas las regiones son del tipo 1, y como la distancia entre símbolos en este caso es $\sqrt{2}$, es fácil comprobar que la probabilidad de error de cada símbolo es la misma, con lo que su valor coincide con la probabilidad media de error y vale

$$P_e = 1 - \left(1 - Q \left(\frac{1}{\sqrt{N_0}} \right) \right)^2 = 2Q \left(\frac{1}{\sqrt{N_0}} \right) - Q^2 \left(\frac{1}{\sqrt{N_0}} \right).$$

En las constelaciones representadas en la Figura 4.54 se pueden calcular las probabilidades de error de forma sencilla debido a que o bien se tienen regiones de decisión que forman una retícula

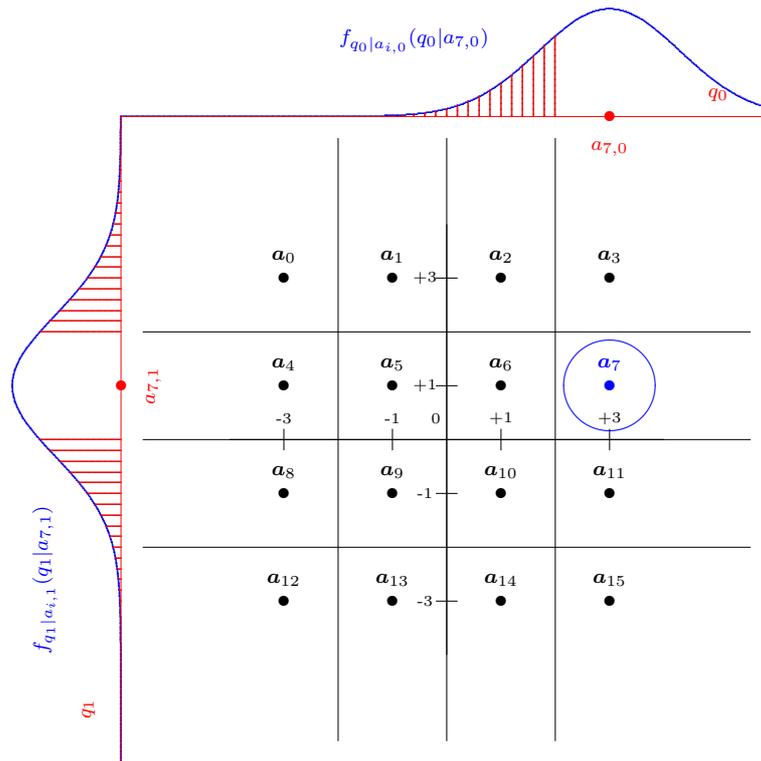


Figura 4.58: Distribuciones condicionales marginales para el símbolo a_7 .

alineada con los ejes del espacio, o bien existe una transformación sencilla que las convierte en ese tipo de regiones. Pero en otros casos esto puede no ser posible, como en el caso de un sistema con la constelación de la Figura 4.59, donde se muestran las ocho regiones de decisión resultantes. Y no resulta posible evaluar de forma analítica la integral de una distribución gaussiana fuera de ese tipo de regiones de decisión. En este caso, para calcular la probabilidad media de error es preciso resolver la integral de la función densidad de probabilidad condicional de la observación fuera de la región de decisión mediante otros procedimientos. En algunos casos será posible calcularla de forma analítica, pero en otros, como el de este ejemplo, no será posible. En casos como este, si se desea calcular la probabilidad de error de forma exacta, habrá que recurrir a cálculos numéricos, con programas como MATLAB que permiten evaluar numéricamente este tipo de resultados. Si se desea utilizar expresiones analíticas, habrá que recurrir a aproximaciones o cotas de la probabilidad de error. A continuación se verán estas últimas herramientas analíticas.

4.5.4. Aproximación y cotas para la probabilidad de error

Las aproximaciones y cotas son útiles cuando no es posible evaluar de forma exacta la probabilidad de error de forma analítica, o cuando se requiere una idea de la magnitud de la probabilidad de error, sin necesidad de conocer el valor exacto.

Aproximación de la probabilidad de error

La probabilidad de error entre símbolos depende de la distancia entre los mismos. Lo más probable, en caso de error, es decidir de forma errónea un símbolo que está a mínima distancia del

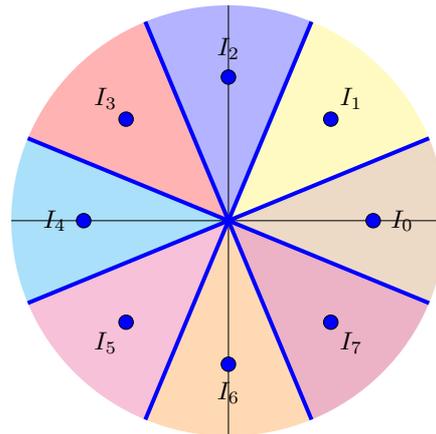


Figura 4.59: Ejemplo de una constelación en un espacio de señales bidimensional en la que no es posible calcular la probabilidad de error de forma analítica.

transmitido, mientras que la probabilidad de que se produzca un error con símbolos más alejados es considerablemente mayor. La aproximación más utilizada asume que sólo se cometerán errores con símbolos que están a mínima distancia, y que todos los símbolos tienen el mismo número de símbolos a mínima distancia, siendo este número el mayor posible para la constelación. Estas consideraciones llevan a la aproximación

$$P_e \approx k \cdot Q \left(\frac{d_{min}}{2\sqrt{N_0/2}} \right),$$

donde los dos parámetros que aparecen en la expresión son:

- d_{min} : mínima distancia entre dos símbolos de la constelación;
- k : máximo número de símbolos a mínima distancia de un símbolo de la constelación.

Estos dos parámetros son muy fáciles de calcular para cualquier constelación. Por ejemplo, para la constelación de 16 símbolos de la Figura 4.54, figura (a), estos parámetros serían

$$d_{min} = 2, k = 4.$$

Para el parámetro k , en la constelación hay símbolos que tienen 2 símbolos a mínima distancia (los de las esquinas), símbolos que tienen 4 símbolos a mínima distancia (los 4 del centro) y otros que tienen a 3 símbolos a mínima distancia (el resto). El valor que hay que incluir en el parámetro es el máximo número de símbolos que hay a mínima distancia de un cierto símbolo, en este caso son 4. Para la constelación de la figura (b), los parámetros serían

$$d_{min} = \sqrt{2}, k = 2.$$

Cota de la unión

La aproximación de la probabilidad de error es útil cuando se quiere tener una idea de la magnitud de la probabilidad de error, y no es relevante que la probabilidad de error total sea

mayor o menor que el valor aproximado. Sin embargo, en ciertas ocasiones es necesario tener una idea aproximada de la probabilidad de error, pero con la certeza de que la probabilidad de error está por debajo del valor especificado (que lo acota). En este caso se recurre a cotas de la probabilidad de error, valores que cumplen que

$$P_e \leq \text{Cota.}$$

Aquí se verán dos cotas: la *cota de la unión*, que acota la probabilidad de error de forma relativamente ajustada, es decir, que la probabilidad de error no es lejana al valor de la cota, especialmente para relaciones señal a ruido altas, aunque su cálculo se complica para constelaciones con muchos símbolos; y la *cota holgada*, una cota con una expresión analítica mucho más sencilla, pero que proporciona valores algo más alejados de la probabilidad de error exacta que la cota de la unión.

Comenzaremos con la cota de la unión. Se trata de una cota que se expresa como la suma de probabilidades de error de sistemas binarios. La idea es que acota la probabilidad condicional de error de un símbolo por la suma de las probabilidades de error de los $M - 1$ sistemas binarios resultantes de utilizar siempre ese símbolo y cada uno de los $M - 1$ símbolos restantes de la constelación, es decir

$$P_{e|\mathbf{a}_i} \leq \sum_{\substack{j=0 \\ j \neq i}}^{M-1} Q \left(\frac{d(\mathbf{a}_i, \mathbf{a}_j)}{2\sqrt{N_0/2}} \right),$$

lo que lleva a acotar la probabilidad de error total como

$$P_e \leq \sum_{i=0}^{M-1} p_{\mathbf{A}}(\mathbf{a}_i) \sum_{\substack{j=0 \\ j \neq i}}^{M-1} Q \left(\frac{d(\mathbf{a}_i, \mathbf{a}_j)}{2\sqrt{N_0/2}} \right).$$

Para ilustrar el procedimiento, se utilizará como ejemplo la constelación de la Figura 4.54 b). En esta constelación, la probabilidad de error cuando el símbolo transmitido es \mathbf{a}_0 , $P_{e|\mathbf{a}_0}$, se obtiene integrando $f_{q|\mathbf{A}}(\mathbf{q}|\mathbf{a}_0)$ fuera de la región de decisión I_0 , lo que corresponde al área rayada de la Figura 4.60 a).

En lugar de calcular esta integral, se calcula el error que se obtendría empleando tres decisores binarios para decidir entre el símbolo \mathbf{a}_0 y cada uno de los otros tres símbolos, es decir, considerar los casos binarios

- \mathbf{a}_0 y \mathbf{a}_1 ;
- \mathbf{a}_0 y \mathbf{a}_2 ;
- \mathbf{a}_0 y \mathbf{a}_3 ;

Esto supone integrar la función densidad de probabilidad en cada una de las tres regiones mostradas en las Figuras 4.60 b), c) y d), respectivamente. Comenzando con el símbolo \mathbf{a}_1 , denotamos la probabilidad de error del decisor binario entre \mathbf{a}_0 y \mathbf{a}_1 como $P_e(\mathbf{a}_0, \mathbf{a}_1)$, y se obtiene integrando el área rayada en la Figura 4.60 b). Para un decisor binario

$$P_e(\mathbf{a}_0, \mathbf{a}_1) = Q \left(\frac{d(\mathbf{a}_0, \mathbf{a}_1)}{2\sqrt{N_0/2}} \right) = Q \left(\frac{\sqrt{2}}{2\sqrt{N_0/2}} \right) = Q \left(\frac{1}{\sqrt{N_0}} \right).$$

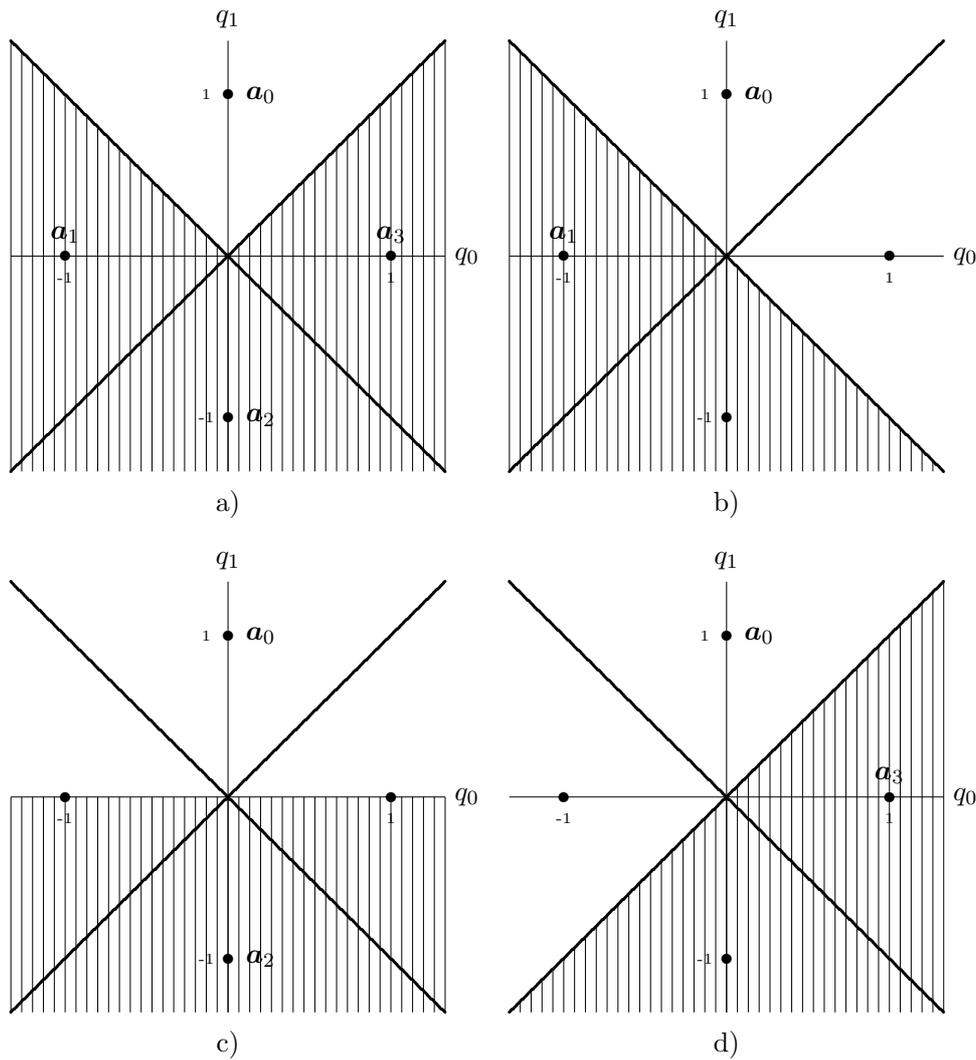


Figura 4.60: Cota de la unión.

Procediendo de igual modo con los otros dos símbolos

$$P_e(\mathbf{a}_0, \mathbf{a}_2) = Q\left(\frac{d(\mathbf{a}_0, \mathbf{a}_2)}{2\sqrt{N_0/2}}\right) = Q\left(\frac{2}{2\sqrt{N_0/2}}\right) = Q\left(\frac{\sqrt{2}}{\sqrt{N_0}}\right).$$

$$P_e(\mathbf{a}_0, \mathbf{a}_3) = Q\left(\frac{d(\mathbf{a}_0, \mathbf{a}_3)}{2\sqrt{N_0/2}}\right) = Q\left(\frac{\sqrt{2}}{2\sqrt{N_0/2}}\right) = Q\left(\frac{1}{\sqrt{N_0}}\right).$$

A partir de estas probabilidades de error con decisores binarios se define la cota de la unión para

$P_{e|a_0}$ como

$$\begin{aligned} P_{e|a_0} &\leq \sum_{j=1}^{M-1} P_e(\mathbf{a}_0, \mathbf{a}_j) \\ &= P_e(\mathbf{a}_0, \mathbf{a}_1) + P_e(\mathbf{a}_0, \mathbf{a}_2) + P_e(\mathbf{a}_0, \mathbf{a}_3) \\ &= 2Q\left(\frac{1}{\sqrt{N_0}}\right) + Q\left(\frac{2}{2\sqrt{N_0/2}}\right). \end{aligned}$$

Procediendo de igual modo con el resto de símbolos de la constelación obtenemos la denominada **cota de la unión para la probabilidad media de error**, que para una constelación genérica de M símbolos se escribe como

$$P_e \leq \sum_{i=0}^{M-1} p_A(\mathbf{a}_i) \sum_{\substack{j=0 \\ j \neq i}}^{M-1} P_e(\mathbf{a}_i, \mathbf{a}_j) = \sum_{i=0}^{M-1} p_A(\mathbf{a}_i) \sum_{\substack{j=0 \\ j \neq i}}^{M-1} Q\left(\frac{d(\mathbf{a}_i, \mathbf{a}_j)}{2\sqrt{N_0/2}}\right)$$

Para constelaciones con símbolos equiprobables

$$P_e \leq \frac{1}{M} \sum_{i=0}^{M-1} \sum_{\substack{j=0 \\ j \neq i}}^{M-1} Q\left(\frac{d(\mathbf{a}_i, \mathbf{a}_j)}{2\sqrt{N_0/2}}\right)$$

Particularizando para el ejemplo que nos ocupa, dada la simetría de la constelación, da como resultado

$$P_e \leq \frac{1}{4} \left(2Q\left(\frac{1}{\sqrt{N_0}}\right) + Q\left(\frac{2}{2\sqrt{N_0/2}}\right) \right)$$

Para comprender por qué la cota de la unión es una cota superior hay que tener en cuenta que $P_{e|a_0}$ se obtiene integrando $f_{a|A}$ en la región sombreada de la Figura 4.60 a), esto es, fuera de I_0 . En lugar de proceder a calcular esa integral, se han sumado tres integrales de la misma función sobre las regiones mostradas por las Figuras 4.60 b), c) y d). La superposición de estas áreas comprende todo el espacio fuera de I_0 , con lo que se asegura que como mínimo se obtiene esa probabilidad de error buscada. Sin embargo, en algunas regiones se está integrando varias veces. Este aspecto se muestra en la Figura 4.61. En las zonas etiquetadas como I sólo se ha integrado una vez, en las zonas etiquetadas como II se ha integrado dos veces y en la zona etiquetada como III se ha integrado tres veces.

Se ha comprobado que la cota de la unión es una cota superior de la probabilidad de error, pero no sabemos si la cota está cerca o lejos del valor verdadero de probabilidad media de error. Para la constelación anterior se han calculado tanto la probabilidad exacta como la cota de la unión. La Figura 4.62 compara ambos valores como una función de la relación E_s/N_0 . Se puede ver que la cota de la unión proporciona un valor superior a la probabilidad de error exacta que además es bastante ajustado al valor exacto, especialmente para valores altos de relación señal a ruido (parametrizada por E_s/N_0).

Cota holgada

En constelaciones con un número de símbolos elevado, la evaluación de la cota de error puede ser engorrosa debido al número de términos a evaluar, que crece aproximadamente con el cuadrado

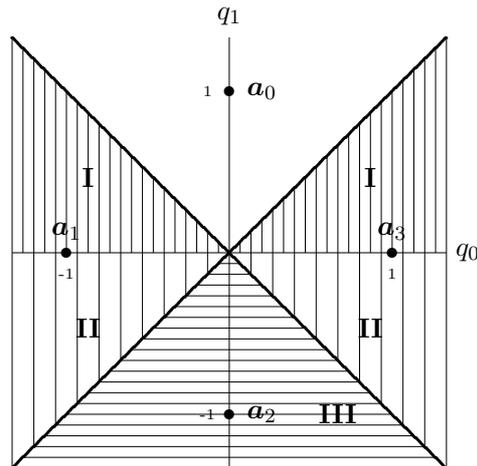


Figura 4.61: Cota de la unión: Cota superior

del número de símbolos de la constelación. Para una constelación de 4 símbolos hay que evaluar 6 distancias (aprovechando que $d(\mathbf{a}_i, \mathbf{a}_j) = d(\mathbf{a}_j, \mathbf{a}_i)$ y hacer 6 evaluaciones de la función $Q(x)$. En general, el número de distancias mínimo a considerar en la expresión, la simetría $d(\mathbf{a}_i, \mathbf{a}_j) = d(\mathbf{a}_j, \mathbf{a}_i)$, es

$$N_{distancias} = \sum_{k=1}^{M-1} k$$

Esto, para algunos ejemplos numéricos supone

- $M = 4$ significa $N_{distancias} = 6$.
- $M = 8$ significa $N_{distancias} = 28$.
- $M = 16$ significa $N_{distancias} = 120$.
- $M = 64$ significa $N_{distancias} = 2016$.

Para constelaciones con un número elevado de símbolos, el cálculo analítico sin utilizar un computador es relativamente costoso, por lo que cabe buscar una cota más compacta analíticamente. Una cota con una expresión analítica compacta y que es comúnmente utilizada es la denominada *cota holgada* (del inglés “*looser bound*”). Esta aproximación utiliza como base la cota de la unión, pero asume que todos los símbolos están a una distancia d_{min} del resto, con lo que es una aproximación pesimista que para constelaciones con un número grande de símbolos puede estar muy alejada de la probabilidad de error exacta. Con esta asunción, la expresión de la probabilidad de error queda acotada por

$$P_e \leq (M - 1)Q\left(\frac{d_{min}}{2\sqrt{N_0/2}}\right).$$

La Figura 4.63 compara la probabilidad de error exacta con la cota holgada para la constelación de cuatro símbolos de la Figura 4.54 (b), como una función de la relación E_s/N_0 . Se puede ver que la cota de la unión proporciona un valor superior a la probabilidad de error exacta que ahora es es menos ajustado al valor exacto que en el caso de la cota de la unión.

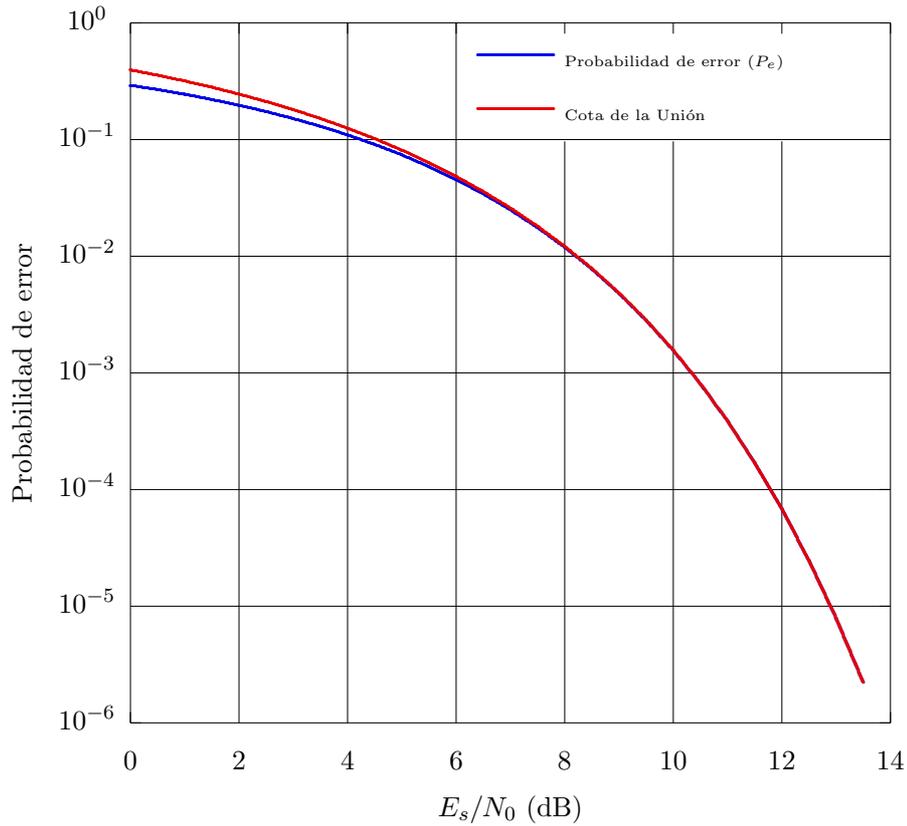


Figura 4.62: Probabilidades de error exacta y cota de la unión para la constelación de la Figura 4.54 (b), como una función de la relación E_s/N_0 .

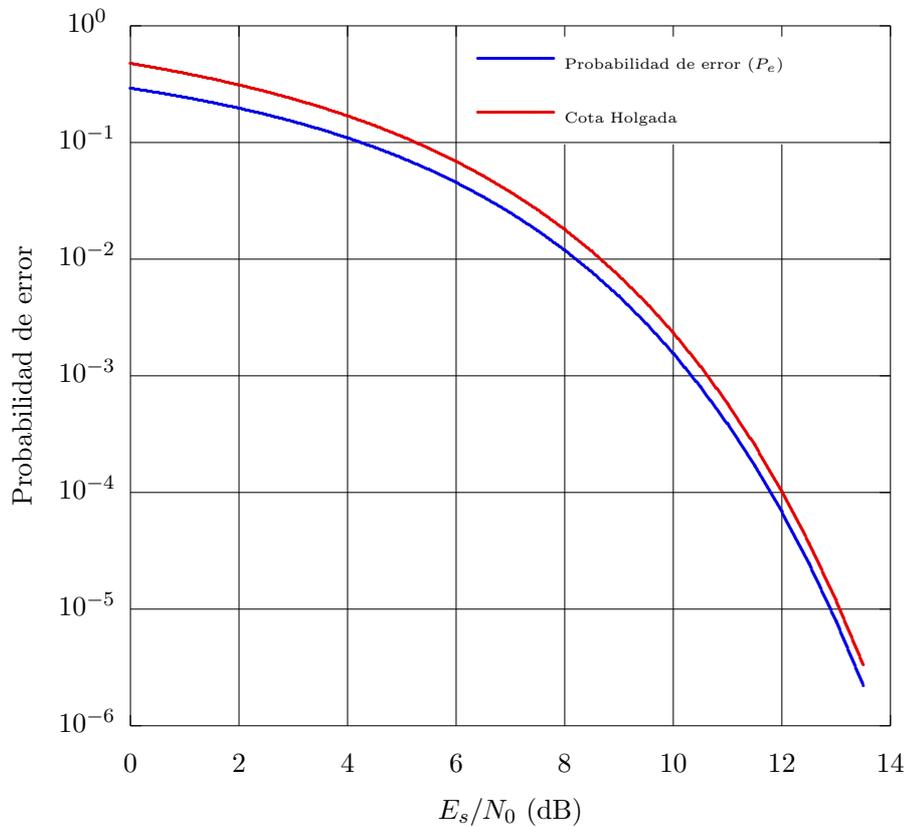


Figura 4.63: Probabilidades de error exacta y cota de la holgada para la constelación de la Figura 4.54 (b), como una función de la relación E_s/N_0 .

4.5.5. Expresiones de la probabilidad de error en función de E_s/N_0

En muchas ocasiones, como ocurre en las Figuras 4.62 o 4.63, resulta interesante expresar la probabilidad de error (bien el valor exacto, una aproximación o una cota), como una función de la relación señal a ruido del sistema, parametrizada por la relación entre energía media por símbolo de la señal transmitida, y densidad espectral de potencia del ruido, i.e. E_s/N_0 . Obtener expresiones de la probabilidad de error en función de esta relación es muy sencillo en la mayoría de los casos. Los términos que aparecen habitualmente en las probabilidades de error se suelen escribir como

$$Q\left(\frac{d_{min}}{2\sqrt{N_0/2}}\right) \text{ ó } Q\left(\frac{A}{\sqrt{N_0/2}}\right).$$

En cualquier caso, multiplicando y dividiendo por $\sqrt{E_s}$, se pueden reescribir como

$$Q\left(v \cdot \sqrt{\frac{E_s}{N_0}}\right), \text{ donde } Q\left(\underbrace{\frac{d_{min}}{\sqrt{2}\sqrt{E_s}}}_v \sqrt{\frac{E_s}{N_0}}\right) \text{ ó } Q\left(\underbrace{\frac{A\sqrt{2}}{\sqrt{E_s}}}_v \sqrt{\frac{E_s}{N_0}}\right)$$

El factor v es un valor constante, que depende de la constelación y que puede evaluarse de forma sencilla: basta con conocer el valor numérico de E_s y sustituir en la expresión anterior. Este factor se puede ver como una medida de eficiencia de la constelación, ya que cuanto más alto es v , menor es el valor de la función $Q(x)$ para ese argumento, y más eficiente es la constelación. A continuación se introduce un ejemplo en que se calcula la probabilidad de error en función de E_s/N_0 para dos constelaciones binarias diferentes.

Ejemplo

Se ha visto que para cualquier dimensión N , cualquier sistema binario con símbolos equiprobables que transmita sobre un canal gaussiano tiene una probabilidad de error

$$P_e = Q\left(\frac{d(\mathbf{a}_0, \mathbf{a}_1)}{2\sqrt{N_0/2}}\right)$$

Se compararán dos casos

- Caso (a): Constelación binaria simétrica ($N = 1$)

$$\mathbf{a}_0 = +A, \mathbf{a}_1 = -A$$

- Caso (b): Constelación ortogonal ($N = 2$)

$$\mathbf{a}_0 = \begin{bmatrix} A \\ 0 \end{bmatrix}, \mathbf{a}_1 = \begin{bmatrix} 0 \\ A \end{bmatrix}$$

Sólo hay que calcular las distancias y energías medias por símbolo para cada constelación

- Caso (a): $E_s = A^2$, $d(\mathbf{a}_0, \mathbf{a}_1) = 2 \cdot A$
- Caso (b): $E_s = A^2$, $d(\mathbf{a}_0, \mathbf{a}_1) = \sqrt{2} \cdot A$

A partir de estas expresiones, el factor de eficiencia v se calcula como

- Caso (a): Constelación binaria simétrica

$$v = \frac{d(\mathbf{a}_0, \mathbf{a}_1)}{\sqrt{2}\sqrt{E_s}} = \sqrt{2} \rightarrow P_e = Q\left(\sqrt{2\frac{E_s}{N_0}}\right)$$

- Caso (b): Constelación ortogonal (se extiende para $N > 2$)

$$v = \frac{d(\mathbf{a}_0, \mathbf{a}_1)}{\sqrt{2}\sqrt{E_s}} = 1 \rightarrow P_e = Q\left(\sqrt{\frac{E_s}{N_0}}\right)$$

Se puede ver que la constelación binaria simétrica es más eficiente. Para un valor de E_s dado, tiene mejores prestaciones que la constelación ortogonal.

4.6. Codificador

Hasta ahora, la descripción que se ha hecho del codificador se reduce a decir que realiza una transformación de una secuencia de símbolos, $B[n]$, a una representación vectorial de las señales que estarán asociadas a cada símbolo de la secuencia, $\mathbf{A}[n]$. Esta función se ilustra en la Figura 4.64. Cada uno de los símbolos estará formado por un bloque de m bits.



Figura 4.64: Codificador en el transmisor de un sistema de comunicaciones digital.

Es ahora cuando, una vez analizado el receptor y a la vista de las expresiones obtenidas en el cálculo de las prestaciones que se pueden obtener con un sistema de comunicaciones, se establecerán las condiciones que debe cumplir el codificador para conseguir una comunicación de la forma más eficiente que sea posible.

4.6.1. Diseño del codificador

El diseño del codificador consiste en dos partes:

1. Diseño o elección de la constelación a transmitir.
2. Asignación binaria para cada uno de los símbolos.

La primera parte consiste en elegir la constelación de M puntos en un espacio de dimensión N que define la representación vectorial de las M señales que se utilizan en el transmisor para transportar cada uno de los $M = 2^m$ valores del alfabeto de símbolos (bloques de m bits). De modo que cada posible combinación de m bits, símbolo b_i , tendrá asociada una señal $s_i(t)$, y el codificador lo que hace es definir la representación vectorial de la señal, \mathbf{a}_i . Luego, el modulador convertirá esta representación vectorial en una señal continua en el tiempo $s_i(t)$. Es importante recordar que la representación vectorial de un conjunto de M señales determina dos factores importantes

1. La energía de cada señal, y por tanto, la energía media de las M señales.
2. La distancia entre cada par de señales, que se relaciona con la energía de la señal diferencia, y que hemos visto que determina las prestaciones del sistema.

Por tanto, la elección de la constelación determina estos dos factores: energía y prestaciones. Y de hecho, el diseño de la constelación se hará en base a estos dos factores: se buscará el mejor compromiso entre energía y prestaciones.

La segunda parte del diseño del codificador es la asignación binaria. Esta asignación consiste en adjudicar a cada uno de los posibles valores del alfabeto de \mathbf{A} (o equivalentemente de $B[n]$, ya que hay una asignación biunívoca $b_i \leftrightarrow \mathbf{a}_i$) una de las M posibles combinaciones de m bits. Como veremos en su momento, la asignación determina la probabilidad de error a nivel de bit, es decir, las prestaciones del sistema. Por tanto, los criterios que han de guiar el diseño del codificador son

1. Prestaciones (probabilidades de error de símbolo y de bit).
2. Energía.

A continuación se analizará cómo se realiza el diseño de la constelación para el caso de transmisión de símbolos equiprobables cuando el ruido es gaussiano, que es el caso más frecuente en sistemas de comunicaciones.

4.6.2. Diseño de la constelación

El diseño de la constelación depende de dos factores: prestaciones y energía. Y las prestaciones están relacionadas con la distancia entre símbolos, dependiendo fundamentalmente de la mínima distancia entre dos símbolos de la constelación. Recuerde la aproximación de la probabilidad de error

$$P_e \approx k \cdot Q\left(\frac{d_{min}}{2\sqrt{N_0/2}}\right).$$

Así que dependiendo del tipo de limitaciones que a priori se tengan en el sistema, el problema de diseño se puede plantear de dos formas. Si existe una limitación en energía, para conseguir la menor probabilidad de error posible, el codificador ha de generar una constelación que tenga la mayor distancia entre símbolos, pero con la limitación que ponga el valor máximo de energía de símbolo o de energía media por símbolo, E_s que sea admisible teniendo en cuenta las limitaciones en energía que tenga el sistema. Si lo que se limita es la probabilidad de error máxima admisible, habrá que buscar una constelación que la consiga utilizando la menor cantidad de energía, de modo que se minimize el coste del sistema transmisor.

En cualquier caso, los dos factores que determinan la elección de la constelación van en sentidos contrapuestos: si se aumenta la distancia entre símbolos de una constelación para disminuir la probabilidad de error, en general aumentará la energía necesaria para transmitir cada símbolo. Es pues necesario encontrar un compromiso entre ambos requerimientos. Conviene también recordar que la energía de una señal, en cuanto a la representación vectorial de la misma, está relacionada con la distancia que tiene respecto del origen dicha representación vectorial, ya que

$$\mathcal{E}\{\mathbf{a}_i\} = \|\mathbf{a}_i\|^2 = \sum_{k=0}^{N-1} |a_{i,k}|^2,$$

es decir, que la energía de una señal es la distancia al cuadrado de su representación vectorial al origen de coordenadas. De este modo, se puede plantear el problema del diseño óptimo de un codificador como el problema de colocar M puntos en un espacio de dimensión N de forma que tengan la mayor separación mínima entre ellos que sea posible, al mismo tiempo que estén todos lo más cerca posible del origen de coordenadas, buscando un compromiso entre las dos cosas.

Relacionada con la idea de colocar los puntos lo más cerca posible del origen, está la siguiente propiedad, que determina una de las condiciones que debe cumplir un codificador óptimo desde el punto de vista de compromiso prestaciones/energía: Para unas distancias entre símbolos dadas, se minimiza la energía media por símbolo cuando la media de la constelación es nula

$$E[\mathbf{a}_i] = \begin{bmatrix} E[a_{i,0}] \\ E[a_{i,1}] \\ \vdots \\ E[a_{i,N-1}] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

Esta propiedad se ilustrará con un sencillo ejemplo de sistema binario en un espacio unidimensional, que es fácilmente extensible a cualquier otra situación. Se consideran dos símbolos en un espacio 1D, en concreto los símbolos

$$\mathbf{a}_0 = B - A, \quad \mathbf{a}_1 = B + A$$

que se representan en la Figura 4.65

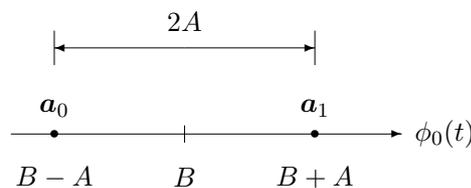


Figura 4.65: Ejemplo para un caso binario y unidimensional de la relación de la energía de una constelación con su media y con la distancia entre símbolos de la constelación.

Los dos parámetros de los símbolos, A y B , parametrizan por separado la distancia entre símbolos y la media de los mismos:

- Media: B
- Distancia entre símbolos: $2A$

Si se calcula la energía media por símbolo (asumiendo símbolos equiprobables) para esta constelación, se tiene

$$\begin{aligned} E_s &= \frac{1}{2} \cdot \mathcal{E}\{\mathbf{a}_0\} + \frac{1}{2} \cdot \mathcal{E}\{\mathbf{a}_1\} = \frac{1}{2} \cdot (B - A)^2 + \frac{1}{2} \cdot (B + A)^2 \\ &= \frac{1}{2} \cdot (B^2 + A^2 - 2AB) + \frac{1}{2} \cdot (B^2 + A^2 + 2AB) = B^2 + A^2. \end{aligned}$$

Se puede ver que aparecen dos términos independientes, uno relacionado con la media de la constelación, y otro con la distancia entre símbolos.

- Contribución de la media: B^2
- Contribución de la distancia entre símbolos: A^2

Por tanto, para cualquier distancia entre símbolos, la mínima energía por símbolo se obtiene cuando la media es nula

$$\text{Media nula } (B = 0) \rightarrow E_s = A^2.$$

Resulta trivial extender este desarrollo para espacios de mayor dimensión o para un mayor número de símbolos. Así que un diseño óptimo siempre contará con esta característica: la media de los símbolos de la constelación será nula (siendo la media de dimensión N).

Técnica de empaquetado de esferas

Teniendo en cuenta todas estas consideraciones, el problema de diseñar una constelación óptima, entendiéndolo por óptima el que tenga el mejor compromiso posible entre prestaciones y consumo de energía, se puede plantear como la búsqueda de constelaciones que para una mínima distancia entre símbolos dada, requieran la mínima energía, que implica que los símbolos estén lo más cerca posible del origen (teniendo una media nula). Atendiendo a esta descripción, el problema de diseño de la constelación se puede resolver mediante la denominada *técnica de empaquetado de esferas*: modelar cada símbolo como una esfera de diámetro d_{min} , de modo que dos esferas que se toquen estén a distancia d_{min} , como se muestra en la Figura 4.66, y buscar como empaquetarlas en un espacio N -dimensional ocupando el menor volumen posible.

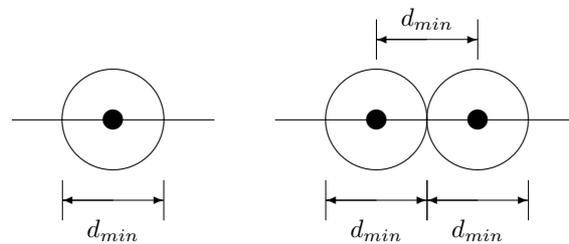


Figura 4.66: Modelo de un símbolo en la técnica de empaquetado de esferas.

Técnica de empaquetado de esferas - Espacio 1D

En un espacio unidimensional, la aplicación de esta metodología es muy simple. Se busca colocar M símbolos sobre una recta, separados como mínimo d_{min} , y que estén lo más cercanos que sea posible del origen. Esto se consigue colocando los puntos equiespaciados, centrados en el origen, con la mitad a cada lado, para conseguir que la media sea nula, con lo que las coordenadas de los M símbolos serán

$$\mathbf{a}_i \in \left\{ \pm \frac{d_{min}}{2}, \pm 3 \cdot \frac{d_{min}}{2}, \dots, \pm (M - 1) \cdot \frac{d_{min}}{2} \right\},$$

es decir, símbolos con coordenadas \pm números impares por la mitad de la distancia mínima entre símbolos. La Figura 4.67 muestra un ejemplo para $M = 4$.

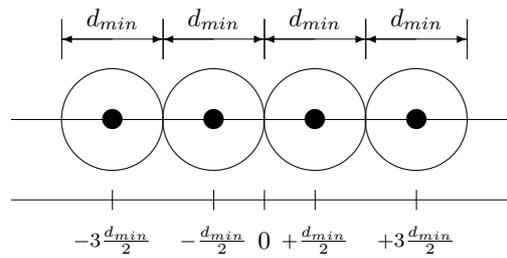


Figura 4.67: Ejemplo de aplicación de la técnica de empaquetado de esferas para un espacio 1D.

Técnica de empaquetado de esferas - Espacio 2D

En un espacio 2D, la mejor forma de empaquetar esferas, tal y como se ilustra en la Figura 4.68, es siguiendo una configuración hexagonal, más eficiente que una configuración en una retícula.

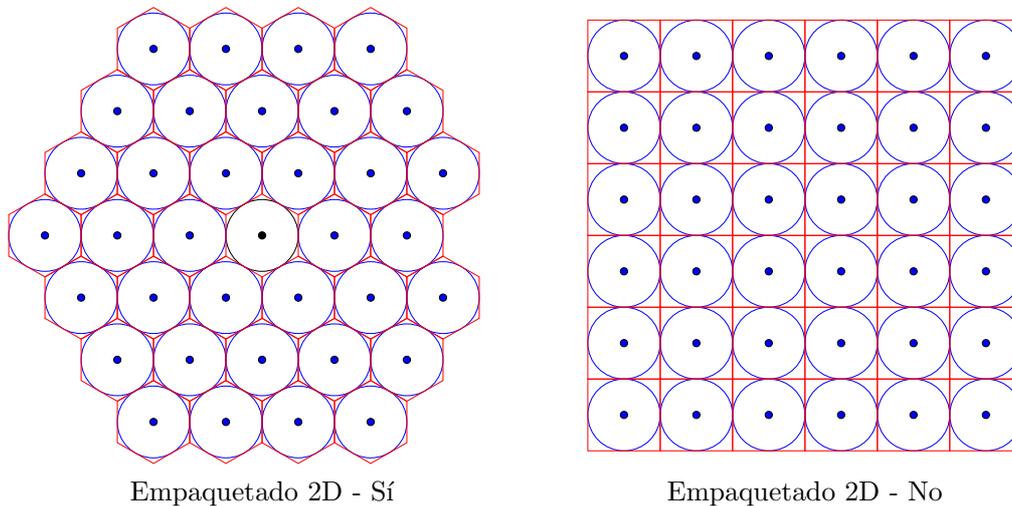


Figura 4.68: Ejemplo de aplicación de la técnica de empaquetado de esferas para un espacio 2D, en la que se compara la forma más eficiente de colocar las esferas, utilizando una configuración hexagonal, frente a un modo menos eficiente en una configuración reticular (se compara el empaquetado de 36 esferas con ambas configuraciones).

La opción de colocar una esfera en el centro, y colocar un círculo de esferas tocándola a su alrededor, provoca una configuración hexagonal mucho más eficiente que una configuración rectangular con los puntos formando una retícula. Si hay más símbolos, se colocará un segundo anillo sobre el primero, teniendo también una configuración hexagonal, y se seguirá así hasta completar M esferas. Como M es una potencia de 2, es posible que el último anillo no se complete, como ocurre en el ejemplo de la constelación de 32 símbolos ($m = 5$ bits por símbolo) de la Figura 4.69. En ese caso, si la esfera inicial se mantiene en el origen, la media de la constelación no será nula, por lo que habrá que “recolocar la constelación” para que tenga media nula, como se ha hecho ya para la constelación de la figura.

La forma en que ha de modificarse una constelación con media no nula para que manteniendo la relación de distancias entre símbolos (lo que geoméricamente equivale a hacer un desplazamiento sin escalado) pase a tener media nula es restando a cada símbolo de la constelación la media de

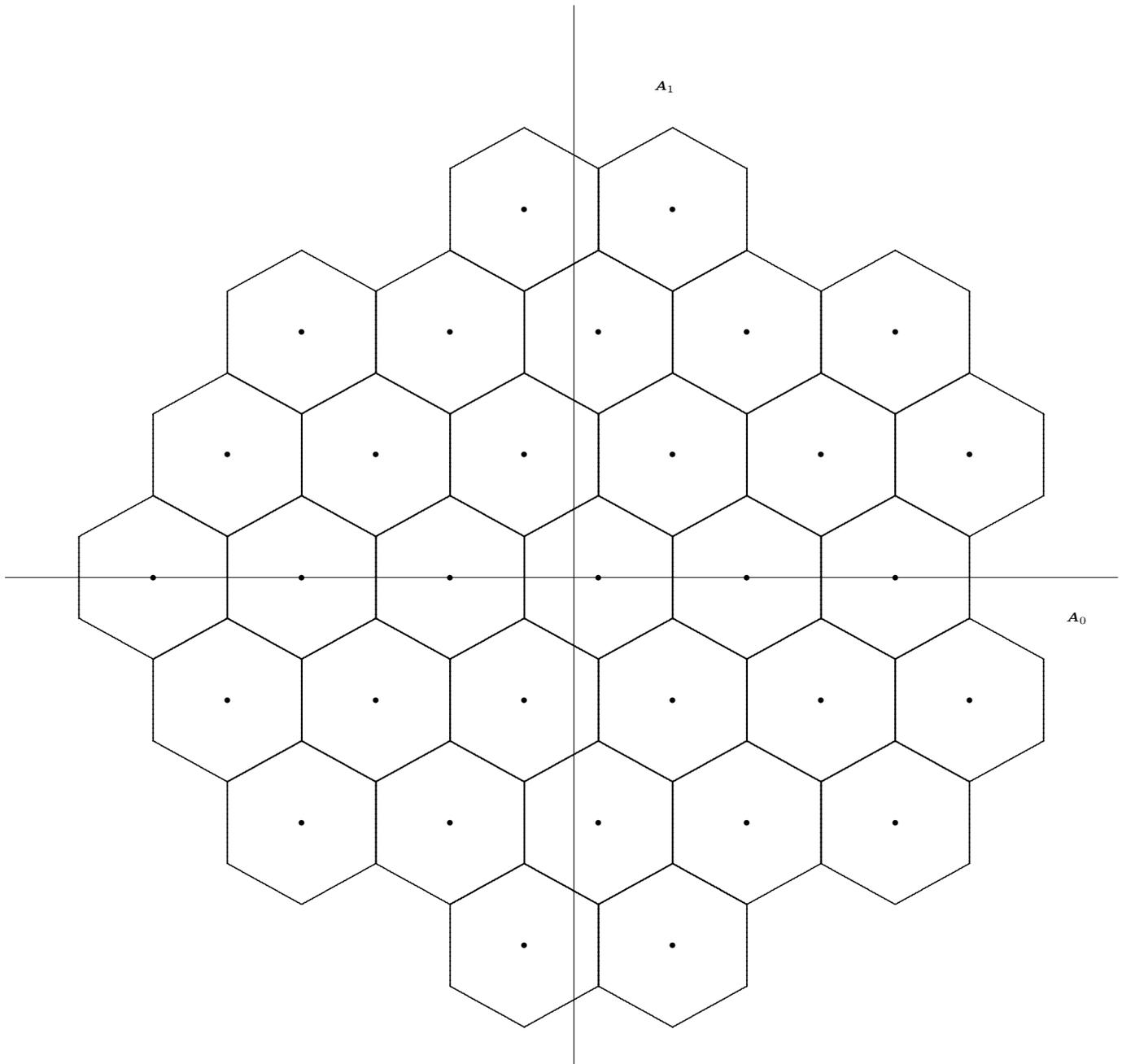


Figura 4.69: Ejemplo de aplicación de la técnica de empaquetado de esferas para un espacio 2D, para una constelación de 32 puntos, en la que se han sustituido las esferas por hexágonos para resaltar la estructura hexagonal resultante.

la misma, es decir, hay que calcular la media de la constelación

$$E[\mathbf{a}_i] = \sum_{i=0}^{M-1} p_A(\mathbf{a}_i) \cdot \mathbf{a}_i.$$

y la nueva constelación modificada para tener media nula tendrá símbolos con coordenadas

$$\mathbf{a}'_i = \mathbf{a}_i - E[\mathbf{a}_i]$$

de forma que la media de la nueva constelación es

$$E[\mathbf{a}'_i] = \sum_{i=0}^{M-1} \underbrace{p_{A'}(\mathbf{a}'_i)}_{p_A(\mathbf{a}_i)} \cdot \mathbf{a}'_i = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \mathbf{0}$$

A continuación se muestra un ejemplo de como modificar una constelación para, manteniendo la relación de distancias entre símbolos, la media de la nueva constelación sea nula, y por tanto, la energía media por símbolo sea mínima.

Ejemplo

En concreto, se considera una constelación con $M = 4$ símbolos, $p_A(\mathbf{a}_i) = \frac{1}{4}$, $\forall i$, con coordenadas

$$\mathbf{a}_0 = \begin{bmatrix} -\frac{1}{2} \\ 0 \end{bmatrix}, \mathbf{a}_1 = \begin{bmatrix} +\frac{3}{2} \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} -\frac{1}{2} \\ +2 \end{bmatrix}, \mathbf{a}_3 = \begin{bmatrix} +\frac{3}{2} \\ +2 \end{bmatrix}$$

La constelación se representa en la Figura 4.70 muestra esta constelación.

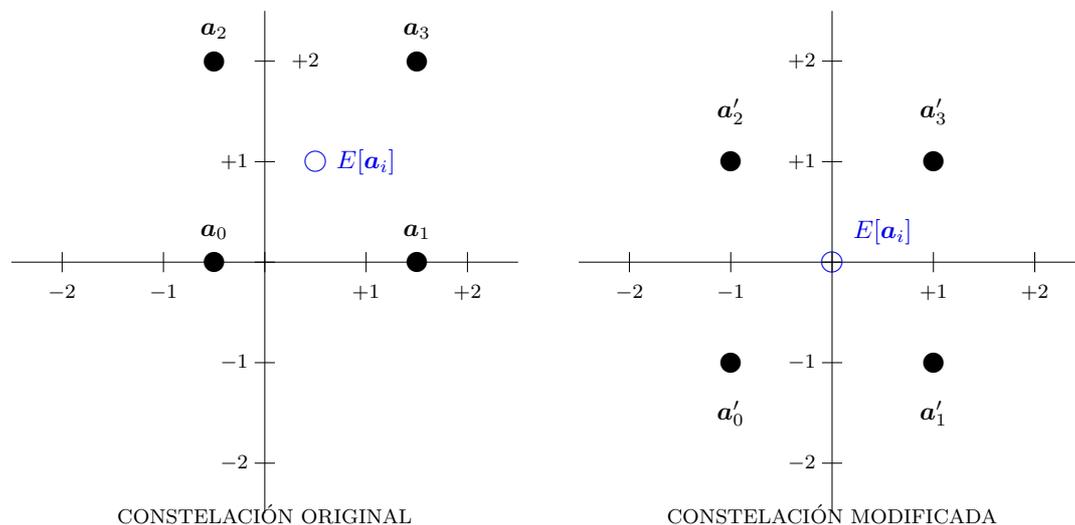


Figura 4.70: Ejemplo de modificación de una constelación para minimizar la energía media por símbolo manteniendo las prestaciones (distancias relativas entre símbolos).

La energía media por símbolo de la constelación en este caso es

$$E_s = \frac{1}{4} \cdot \left[\left(-\frac{1}{2} \right)^2 + 0^2 \right] + \frac{1}{4} \cdot \left[\left(\frac{3}{2} \right)^2 + 0^2 \right] + \frac{1}{4} \cdot \left[\left(-\frac{1}{2} \right)^2 + 2^2 \right] + \frac{1}{4} \cdot \left[\left(\frac{3}{2} \right)^2 + 2^2 \right] = \frac{13}{4} = 3,25$$

Para minimizar la energía media por símbolo, primero se calcula la media de la constelación, que en este caso vale

$$E[\mathbf{a}_i] = \frac{1}{4} \cdot \begin{bmatrix} -\frac{1}{2} \\ 0 \end{bmatrix} + \frac{1}{4} \cdot \begin{bmatrix} +\frac{3}{2} \\ 0 \end{bmatrix} + \frac{1}{4} \cdot \begin{bmatrix} -\frac{1}{2} \\ +2 \end{bmatrix} + \frac{1}{4} \cdot \begin{bmatrix} +\frac{3}{2} \\ +2 \end{bmatrix} = \begin{bmatrix} +\frac{1}{2} \\ +1 \end{bmatrix}.$$

Esta media se ilustra con un círculo en la figura. Una vez calculada la media de la constelación, a cada símbolo se le resta dicha media, de modo que se obtiene la siguiente constelación modificada

$$\mathbf{a}'_0 = \mathbf{a}_0 - E[\mathbf{a}_i] = \begin{bmatrix} -1 \\ -1 \end{bmatrix}, \quad \mathbf{a}'_1 = \mathbf{a}_1 - E[\mathbf{a}_i] = \begin{bmatrix} +1 \\ -1 \end{bmatrix}, \quad \mathbf{a}'_2 = \begin{bmatrix} -1 \\ +1 \end{bmatrix}, \quad \mathbf{a}'_3 = \begin{bmatrix} +1 \\ +1 \end{bmatrix},$$

que se representa también en la Figura 4.70, y que tiene media nula. Si se calcula ahora la energía media por símbolo de la constelación modificada, se tiene

$$E'_s = \frac{1}{4} \cdot [(-1)^2 + (-1)^2] + \frac{1}{4} \cdot [(+1)^2 + (-1)^2] + \frac{1}{4} \cdot [(-1)^2 + (+1)^2] + \frac{1}{4} \cdot [(+1)^2 + (+1)^2] = 2.$$

Se puede ver que es menor que en la constelación original, tal y como era de esperar.

4.6.3. Constelaciones utilizadas en sistemas de comunicaciones

Aunque esta es la forma óptima de las constelaciones desde el punto de vista del mejor compromiso entre prestaciones y consumo de energía, existen otros factores a tener en cuenta a la hora de escoger una constelación que hacen que no siempre se elija esta estrategia. De hecho, las constelaciones hexagonales no son las constelaciones que se usan con más frecuencia en los sistemas de comunicaciones prácticos. Entre estos factores que llevan a elegir otro tipo de constelaciones destacan los siguientes:

Facilidad de realización del transmisor Las coordenadas de los distintos símbolos deben poder expresarse con números de precisión finita (no irracionales) pues es necesario almacenarlos en el hardware o software del transmisor. En un empaquetamiento hexagonal aparecen coordenadas de símbolos del tipo irracional. Si se trunca el valor de las coordenadas la posición de los símbolos se modifica, teniendo un cambio en las prestaciones.

Relación energía de pico/energía media Existen tecnologías de amplificación que funcionan bien cuando la energía de las señales no varía entre intervalos de símbolo, pero que generan serias distorsiones cuando hay cambios bruscos en la energía de un intervalo a otro. Constelaciones como la de la Figura 4.69 tienen símbolos con energías relativamente diferentes (muy baja para los símbolos próximos al origen, y varios órdenes de magnitud mayor para los símbolos más alejados). Constelaciones en las que la relación entre la energía de pico (energía máxima de un símbolo), y la energía media por símbolo es alta, no son adecuadas para su uso con este tipo de amplificadores.

Facilidad de realización del receptor Existen aplicaciones en las que el coste del receptor es determinante, hasta tal punto que es preferible sacrificar una menor probabilidad de error en recepción para conseguir una simplificación de la circuitería necesaria en el receptor. Dependiendo de la tecnología empleada, este puede ser un factor determinante para utilizar constelaciones que aunque no tengan el mejor compromiso prestaciones/energía, permitan implementaciones más económicas para el receptor. Un ejemplo de problema que puede aparecer en algunas tecnologías es la recuperación del signo de las coordenadas del símbolo

recibido (implementar la función signo en un ASIC es relativamente complicado) y esto fuerza a utilizar constelaciones donde las coordenadas no sean negativas, denominadas comúnmente *constelaciones unipolares*. Otro ejemplo es la implementación del decisor. En constelaciones hexagonales, las regiones de decisión son hexagonales, y determinar si un punto está dentro de una determinada región de decisión, implica comparar dicho punto con 6 rectas y comprobar que está simultáneamente en el lado correcto de las seis rectas. Constelaciones con regiones de decisión más simples, pueden ser en algunos casos más convenientes.

Por razones de orden práctico, especialmente las mencionadas con anterioridad, las constelaciones utilizadas con mayor frecuencia son las que se citan a continuación.

Constelaciones QAM

Estas constelaciones, cuyo nombre viene del acrónimo inglés para “*Quadrature Amplitude Modulation*”, tienen como principal característica que los puntos de la constelación forman una retícula sobre un espacio 2D, tal y como los ejemplos que se muestran en la Figura 4.71. Para ello, tienen la particularidad de que el número de bits por símbolo m es par.

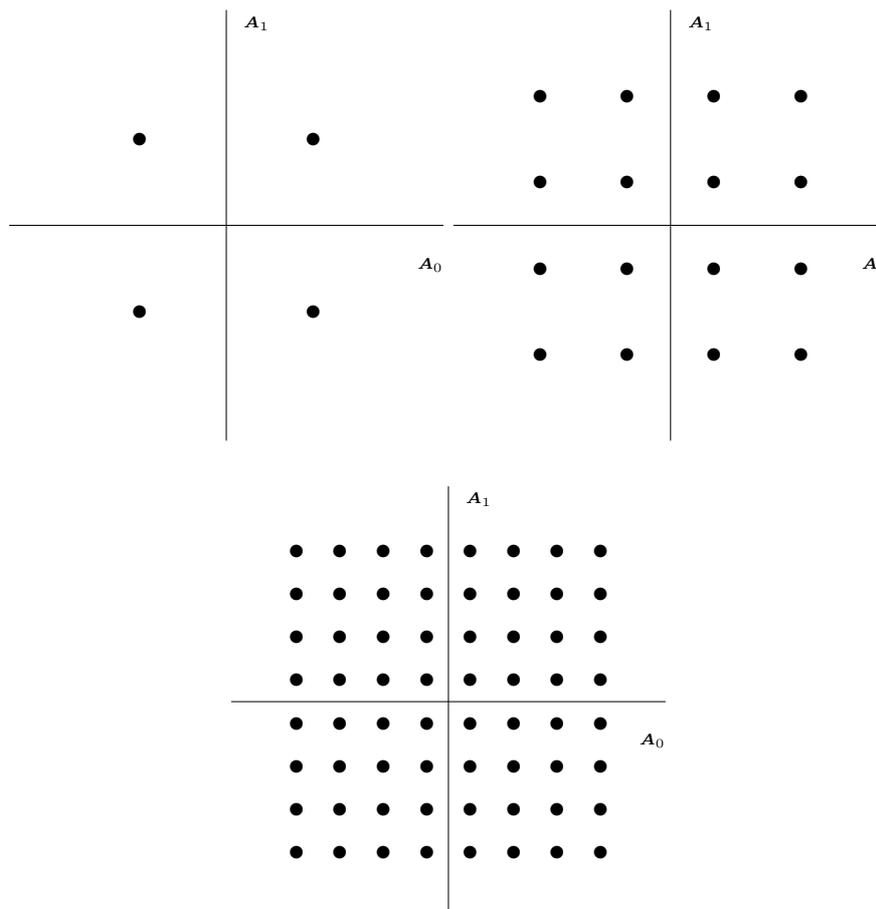


Figura 4.71: Ejemplo de constelaciones QAM de 4, 16 y 64 símbolos.

Este tipo de constelaciones tienen la ventaja de que la implementación de transmisor y receptor es relativamente simple. Por ejemplo, los valores de coordenadas toman los mismos valores en

cada dirección del espacio, y las regiones de decisión de cada símbolo se pueden expresar como condiciones independientes para cada una de las direcciones del espacio, por lo que el procesamiento se puede hacer de forma independiente para cada dirección del espacio tanto en el transmisor como en el receptor.

Constelaciones PSK

Estas constelaciones, cuyo nombre viene del acrónimo inglés para “*Phase Shift Keying*”, tienen como principal característica que los puntos de la constelación forman una circunferencia sobre un espacio 2D, tal y como los ejemplos que se muestran en la Figura 4.72.

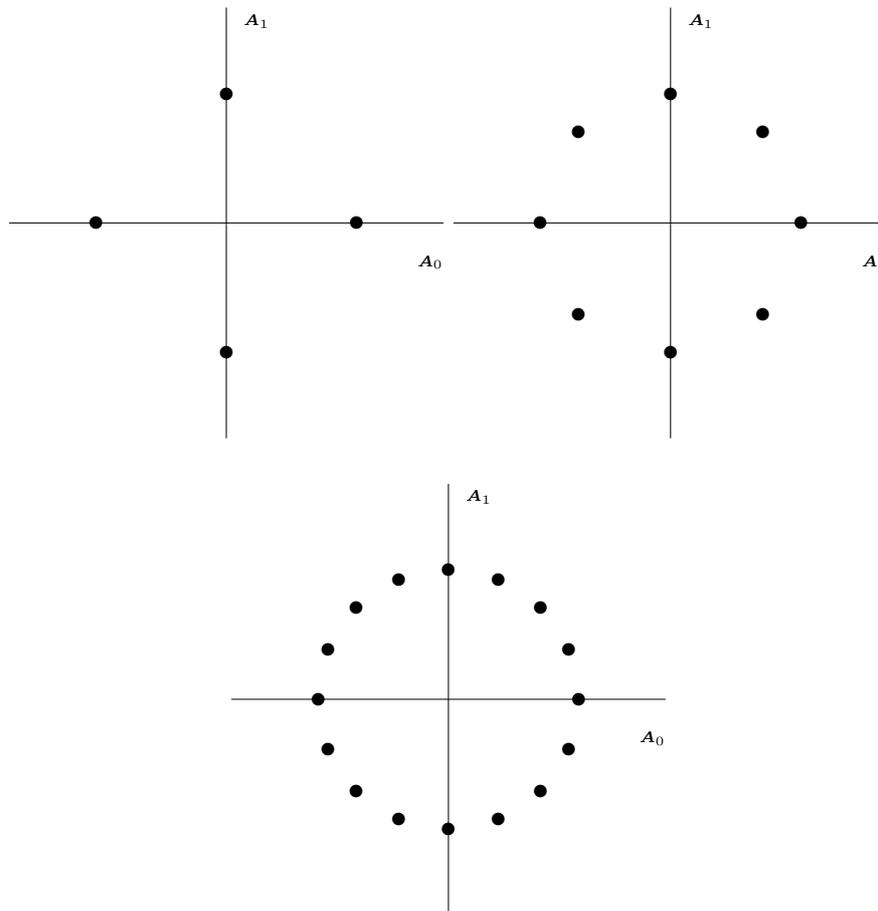


Figura 4.72: Ejemplo de constelaciones QAM de 4, 8 y 16 símbolos.

Su principal ventaja es que como los símbolos tienen todos la misma distancia al origen, todos tienen la misma energía, por lo que la energía permanece constante en todos los intervalos de símbolo durante la transmisión. La implementación de transmisor y receptor tampoco es compleja, ya que los símbolos se pueden codificar como

$$\mathbf{A}[n] = \sqrt{E_s} \times e^{j\phi[n]},$$

es decir, que la información de cada símbolo está en una fase. En el decisor, las regiones de decisión se definen a partir de umbrales sobre las fases, por lo que lo único que hay que estimar es la fase de la observación recibida y compararla con los valores que definen las regiones de decisión.

Constelaciones ortogonales unipolares

Estas constelaciones son constelaciones en las que la dimensión del espacio de señales coincide con el número de señales, de modo que se transmiten M señales ortogonales cuya representación vectorial tiene una única componente no nula idéntica, es decir

$$\mathbf{a}_0 = \begin{bmatrix} \sqrt{E_s} \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{a}_1 = \begin{bmatrix} 0 \\ \sqrt{E_s} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{a}_2 = \begin{bmatrix} 0 \\ 0 \\ \sqrt{E_s} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \mathbf{a}_{M-1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \vdots \\ \sqrt{E_s} \end{bmatrix}.$$

En este caso, todas las coordenadas son positivas, la energía de todos los símbolos es la misma, y el receptor es muy sencillo, ya que basta con buscar la posición con la mayor componente.

4.6.4. Asignación binaria - Codificación de Gray y cálculo de la BER

Una vez seleccionada una constelación, es necesario realizar la asignación binaria: asignar a cada uno de los M símbolos de la constelación, una de las M posibles combinaciones de m bits. El objetivo al hacer la asignación será minimizar la probabilidad de error de bit, que denotaremos como BER (del inglés “*Bit Error Rate*”), y que se define como

$$BER = P(\hat{B}_b[\ell] \neq B_b[\ell]).$$

Para obtener una regla que permita minimizar esta probabilidad de error, lo primero es saber de qué depende, y como le afecta el tener una cierta asignación binaria en el codificador. El cálculo de esta probabilidad de error se hace de forma similar al de la probabilidad de error de símbolo. Hay que calcular las probabilidades de error de bit condicionales, BER_{a_i} , y promediarlas teniendo en cuenta la probabilidad con la que se transmite cada símbolo de la constelación

$$BER = \sum_{i=0}^{M-1} p_A(a_i) \cdot BER_{a_i}.$$

En cuanto a las probabilidades condicionales, hay que promediar la probabilidad de error que se tiene cuando se comete una equivocación de símbolo decidido teniendo en cuenta con qué probabilidad se producen estos errores, lo que significa evaluar la siguiente expresión

$$BER_{a_i} = \sum_{\substack{j=0 \\ j \neq i}}^{M-1} P_{e|a_i \rightarrow a_j} \times \frac{m_{e|a_i \rightarrow a_j}}{m},$$

donde los parámetros involucrados se definen de la forma siguiente:

- $P_{e|a_i \rightarrow a_j}$: probabilidad de transmitiendo $\mathbf{A} = \mathbf{a}_i$, decidir $\hat{\mathbf{A}} = \mathbf{a}_j$

$$P_{e|a_i \rightarrow a_j} = \int_{\mathbf{q}_0 \in I_j} f_{\mathbf{q}|A}(\mathbf{q}_0 | \mathbf{a}_i) d\mathbf{q}_0$$

- $m_{e|a_i \rightarrow a_j}$: número de errores de bit que conlleva esa decisión
- m : número de bits por símbolo de la constelación

Un ejemplo de cálculo de BER para un sistema unidimensional

A continuación se calcula la BER para una constelación unidimensional de 4 símbolos equiprobables con coordenadas

$$\mathbf{a}_0 = -3, \mathbf{a}_1 = -1, \mathbf{a}_2 = +1, \mathbf{a}_3 = +3.$$

Las regiones de decisión están definidas por los umbrales $q_{u1} = -2, q_{u2} = 0, q_{u3} = +2$

$$I_0 = (-\infty, -2], I_1 = (-2, 0], I_2 = (0, +2], I_3 = (+2, +\infty),$$

y la asignación binaria elegida (de momento de forma arbitraria es)

$$\mathbf{a}_0 \equiv 01, \mathbf{a}_1 \equiv 00, \mathbf{a}_2 \equiv 10, \mathbf{a}_3 \equiv 11.$$

La Figura 4.73 ilustra estas características del sistema bajo evaluación.

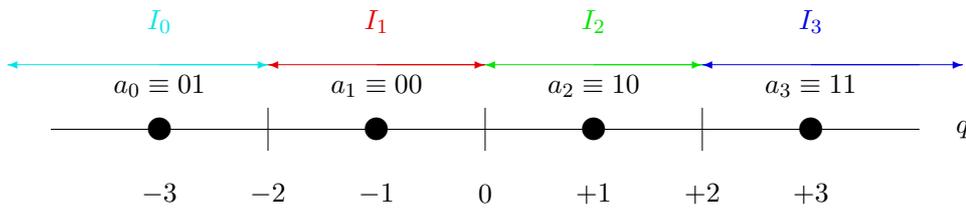


Figura 4.73: Constelación, regiones de decisión, y asignación binaria para el cálculo de la BER.

A continuación se evaluarán las BER condicionales para cada uno de los 4 símbolos de la constelación. Empezamos con el primer símbolo, \mathbf{a}_0 . Para este símbolo, la distribución $f_{q|A}(\mathbf{q}|\mathbf{a}_0)$ es gaussiana de media \mathbf{a}_0 y varianza $N_0/2$, por lo que la BER condicional viene dada por

$$BER_{\mathbf{a}_0} = \underbrace{\left[Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_1}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_1}}{m}} + \underbrace{\left[Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_2}} \times \underbrace{\frac{2}{2}}_{\frac{m_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_2}}{m}} + \underbrace{\left[Q\left(\frac{5}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_3}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_3}}{m}}$$

La Figura 4.74 representa graficamente el significado de las probabilidades $P_{e|\mathbf{a}_0 \rightarrow \mathbf{a}_j}$ en distintos colores.

Para el segundo símbolo, \mathbf{a}_1 , la distribución $f_{q|A}(\mathbf{q}|\mathbf{a}_1)$ es gaussiana de media \mathbf{a}_1 y varianza $N_0/2$, por lo que la probabilidad de error de bit condicional es

$$BER_{\mathbf{a}_1} = \underbrace{\left[Q\left(\frac{1}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_0}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_0}}{m}} + \underbrace{\left[Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_2}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_2}}{m}} + \underbrace{\left[Q\left(\frac{3}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_3}} \times \underbrace{\frac{2}{2}}_{\frac{m_{e|\mathbf{a}_1 \rightarrow \mathbf{a}_3}}{m}}$$

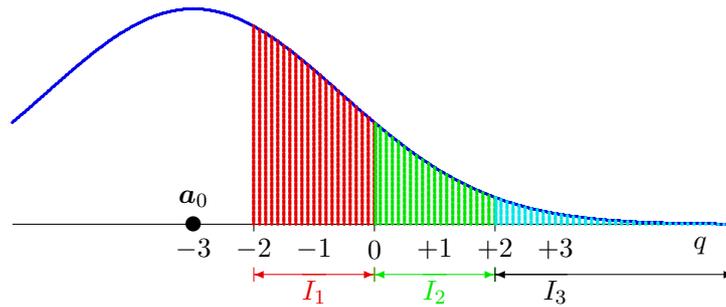


Figura 4.74: Ilustración gráfica de las probabilidades de error entre símbolos, $P_{e|a_0 \rightarrow a_j}$, para $\mathbf{A} = \mathbf{a}_0$.

La Figura 4.75 representa gráficamente el significado de las probabilidades $P_{e|a_1 \rightarrow a_j}$ en distintos colores.

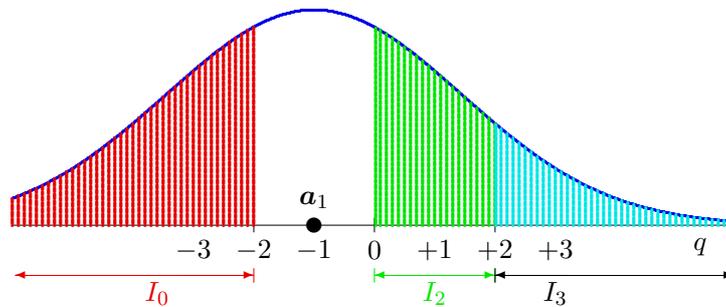


Figura 4.75: Ilustración gráfica de las probabilidades de error entre símbolos, $P_{e|a_1 \rightarrow a_j}$, para $\mathbf{A} = \mathbf{a}_1$.

Para el tercer símbolo, la distribución $f_{q|\mathbf{A}}(q|\mathbf{a}_2)$ es gaussiana de media \mathbf{a}_2 y varianza $N_0/2$, por lo que

$$\begin{aligned}
 BER_{\mathbf{a}_2} &= \underbrace{\left[Q\left(\frac{3}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_0}} \times \underbrace{\frac{2}{2}}_{\frac{m_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_0}}{m}} + \underbrace{\left[Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_1}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_1}}{m}} \\
 &+ \underbrace{\left[Q\left(\frac{1}{\sqrt{N_0/2}}\right) \right]}_{P_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_3}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|\mathbf{a}_2 \rightarrow \mathbf{a}_3}}{m}}
 \end{aligned}$$

La Figura 4.76 representa gráficamente el significado de las probabilidades $P_{e|\mathbf{a}_2 \rightarrow a_j}$ en distintos colores.

Finalmente, para el último símbolo de la constelación, \mathbf{a}_3 , la distribución condicional de la

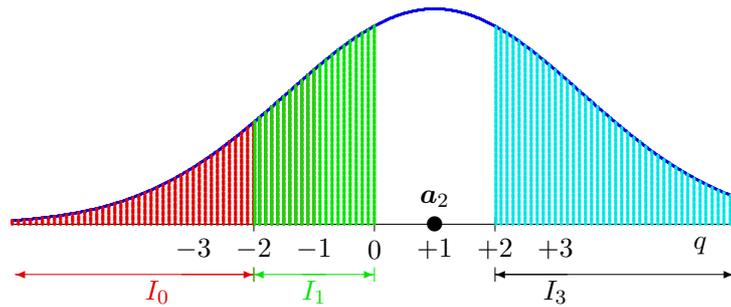


Figura 4.76: Ilustración gráfica de las probabilidades de error entre símbolos, $P_{e|a_2 \rightarrow a_j}$, para $\mathbf{A} = a_2$.

observación, $f_{q|\mathbf{A}}(q|a_3)$, es gaussiana de media a_3 y varianza $N_0/2$, lo que significa que

$$\begin{aligned}
 BER_{a_3} &= \underbrace{\left[Q\left(\frac{5}{\sqrt{N_0/2}}\right) \right]}_{P_{e|a_3 \rightarrow a_0}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|a_3 \rightarrow a_0}}{m}} + \underbrace{\left[Q\left(\frac{3}{\sqrt{N_0/2}}\right) - Q\left(\frac{5}{\sqrt{N_0/2}}\right) \right]}_{P_{e|a_3 \rightarrow a_1}} \times \underbrace{\frac{2}{2}}_{\frac{m_{e|a_3 \rightarrow a_1}}{m}} \\
 &+ \underbrace{\left[Q\left(\frac{1}{\sqrt{N_0/2}}\right) - Q\left(\frac{3}{\sqrt{N_0/2}}\right) \right]}_{P_{e|a_3 \rightarrow a_2}} \times \underbrace{\frac{1}{2}}_{\frac{m_{e|a_3 \rightarrow a_2}}{m}}
 \end{aligned}$$

La Figura 4.77 representa graficamente el significado de las probabilidades $P_{e|a_3 \rightarrow a_j}$ en distintos colores.

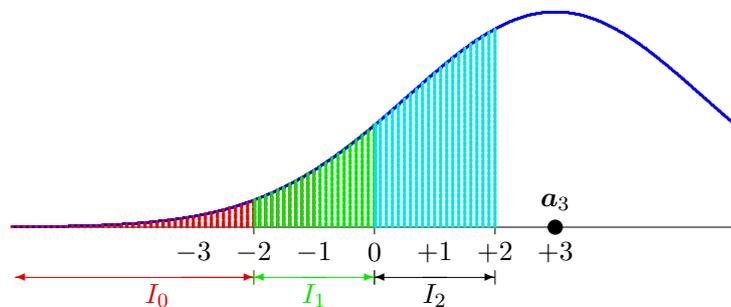


Figura 4.77: Ilustración gráfica de las probabilidades de error entre símbolos, $P_{e|a_3 \rightarrow a_j}$, para $\mathbf{A} = a_3$.

Una vez calculadas las BER condicionales, la BER total se obtiene promediándolas, lo que para este ejemplo lleva a

$$\begin{aligned}
 BER &= \frac{1}{4} \times BER_{a_0} + \frac{1}{4} \times BER_{a_1} + \frac{1}{4} \times BER_{a_2} + \frac{1}{4} \times BER_{a_3} \\
 &= \frac{3}{4} Q\left(\frac{1}{\sqrt{N_0/2}}\right) + \frac{1}{2} Q\left(\frac{3}{\sqrt{N_0/2}}\right) - \frac{1}{4} Q\left(\frac{5}{\sqrt{N_0/2}}\right).
 \end{aligned}$$

Veamos qué ocurre si se modifica la asignación binaria. Si se utilizara ahora, por ejemplo, la asignación binaria

$$a_0 \equiv 11, a_1 \equiv 00, a_2 \equiv 10, a_3 \equiv 01,$$

si se repitiera el proceso anterior para calcular la BER , se llegaría a otro resultado distinto, en concreto

$$BER = \frac{5}{4}Q\left(\frac{1}{\sqrt{N_0/2}}\right) - \frac{1}{4}Q\left(\frac{3}{\sqrt{N_0/2}}\right).$$

Se puede ver que este valor es mayor (el término dominante es el de menor argumento de la función $Q(x)$, en este caso $Q\left(\frac{1}{\sqrt{N_0/2}}\right)$, que en el primer caso está multiplicado por $\frac{3}{4}$, y en el segundo por $\frac{5}{4}$. ¿Qué es lo que hace que una asignación sea mejor o peor respecto a la BER que produce? Si se observa el procedimiento de cálculo, veremos que modificar la asignación binaria no modifica los términos de probabilidad de error entre un símbolo y otro, es decir

No se modifican los términos $P_{e|a_i \rightarrow a_j}$.

Sin embargo, si que cambia la probabilidad de error de bit asignada a cada error de símbolo

Sí se modifican los términos $m_{e|a_i \rightarrow a_j}$.

Esto significa, que dependiendo de la asignación binaria elegida, los términos

$$\frac{m_{e|a_i \rightarrow a_j}}{m}$$

pueden tomar uno de dos valores: $\frac{1}{2}$ ó $\frac{2}{2} = 1$. Cambiar la asignación binaria implica cambiar los términos que tienen asociado uno u otro valor. Lo que interesa aquí es asociar esos valores de forma que se minimice la probabilidad de error de bit. Eso implica tratar de asignar el menor valor, $1/2$, a aquellos términos que más pesen al ser ponderados por $P_{e|a_i \rightarrow a_j}$. Y los términos que tienen un mayor valor de probabilidad de error de símbolo $P_{e|a_i \rightarrow a_j}$ son los asociados a símbolos que están a mínima distancia en la constelación. De aquí surge la regla que optimiza la asignación binaria, y que se denomina *codificación de Gray*.

Codificación de Gray La asignación binaria de símbolos que están a mínima distancia, debe diferir únicamente en 1 bit, de modo que los términos que más pesan en la BER (para valores de $P_{e|a_i \rightarrow a_j}$ más elevados) se multipliquen por el valor $\frac{1}{m}$.

Dicho de otra forma. Cuando se produce un error de símbolo, lo más probable es que sea un error con un símbolo que está a mínima distancia del transmitido. Por eso, lo mejor es hacer en que el tipo de error más frecuente, sólo se produzca un error de bit sobre los m bits codificados. De esta forma, utilizando un codificación de Gray, la probabilidad de error de bit se puede aproximar para relaciones señal a ruido razonablemente altas como

$$BER \approx \frac{1}{m} \cdot P_e,$$

siendo $m = \log_2(M)$ el número de bits por símbolo del sistema.

En todo codificador deber utilizarse una codificación de Gray, o si no es posible, una codificación pseudo-Gray, donde la regla se haga cumplir para tantos símbolos a mínima distancia como sea posible. Para los tipos de constelaciones más utilizadas en la práctica, QAM y PSK, es siempre posible encontrar una codificación de Gray. En el caso de QAM esto es sencillo porque la codificación bidimensional se puede hacer estableciendo una codificación de Gray unidimensional con la mitad de bits asignados de forma independiente en cada dirección del espacio, tal y como se muestra en la Figura 4.78.

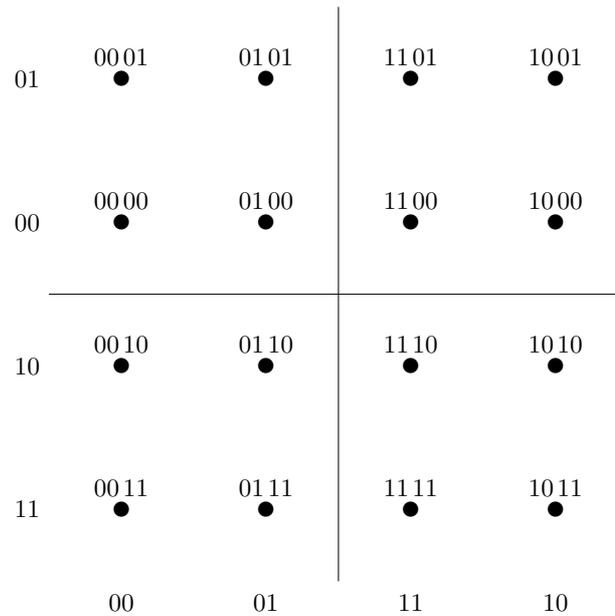


Figura 4.78: Ejemplo de codificación de Gray para una constelación 16-QAM.

Para las constelaciones PSK, al estar los puntos en un círculo, también es relativamente sencillo encontrar una codificación de Gray. Basta plantear una codificación de Gray unidimensional en la que los símbolos de los extremos difieran en un único bit, y convertirla en una asignación circular en el espacio 2D, como en el ejemplo de la Figura 4.79.

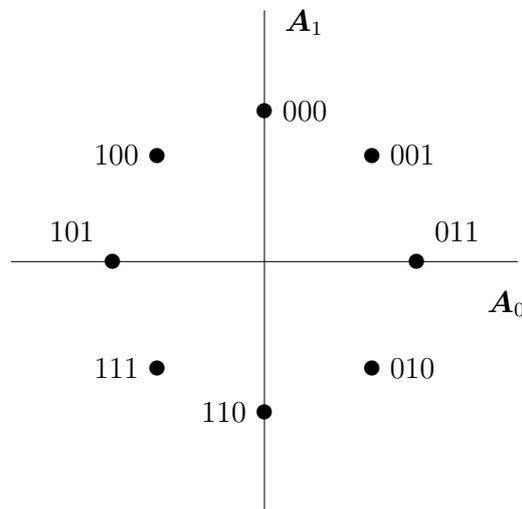


Figura 4.79: Ejemplo de codificación de Gray para una constelación 8-PSK.

4.6.5. Relación entre tasa de bit y tasa de símbolo

Para finalizar con el codificador, recordar únicamente la relación entre tasa de símbolo y tasa de bit. Dado que cada símbolo transporta m bits, la relación entre velocidades de transmisión o tasas de símbolo y de bit es obvia

$$R_b = m \times R_s \text{ bits/s}$$

M (símbolos)	m (bits/símbolo)	E_s con niveles normalizados ($d_{min} = 2$)	d_{min} con $E_s = 2$
4	2	2	2
16	4	10	0,8944
64	8	42	0,4364
256	16	170	0,2169

Tabla 4.1: Transmisión con constelaciones M -QAM.

ó

$$R_s = \frac{R_b}{m} \text{ baudios (símbolos/s).}$$

Esto significa que un sistema con un mayor número de bits por símbolo (o lo que es lo mismo, con un mayor número de símbolos, $M = 2^m$), tendrá para la misma velocidad de símbolo una mayor capacidad de transmisión de bits. Lo que podría invitar a considerar la idea de transmitir constelaciones muy densas (con muchos símbolos). Sin embargo, hay que tener en cuenta otro factor a la hora de elegir una cierta densidad de la constelación. Si el sistema tiene una limitación en energía, constelaciones más densas implicarán menores distancias entre símbolos, lo que llevará a su vez a peores prestaciones. Para ilustrarlo con un ejemplo, analicemos un sistema con constelaciones M -QAM. La Tabla 4.1 muestra para distintos tamaños de la constelación, como aumenta la energía necesaria si se mantiene la distancia mínima (niveles normalizados) o como equivalentemente se reduce la distancia mínima si se mantiene constante la energía media por símbolo de la constelación. En la Figura 4.80, se muestran a escala las constelaciones para ese nivel de energía media por símbolo constante, $E_s = 2$ en este caso. Se puede ver que al elegir constelaciones más densas, aunque el número de bits por símbolo se incrementa, disminuye la relación entre prestaciones y energía, por lo que en la práctica de nuevo hay que buscar un compromiso a la hora de establecer la densidad óptima de la constelación.

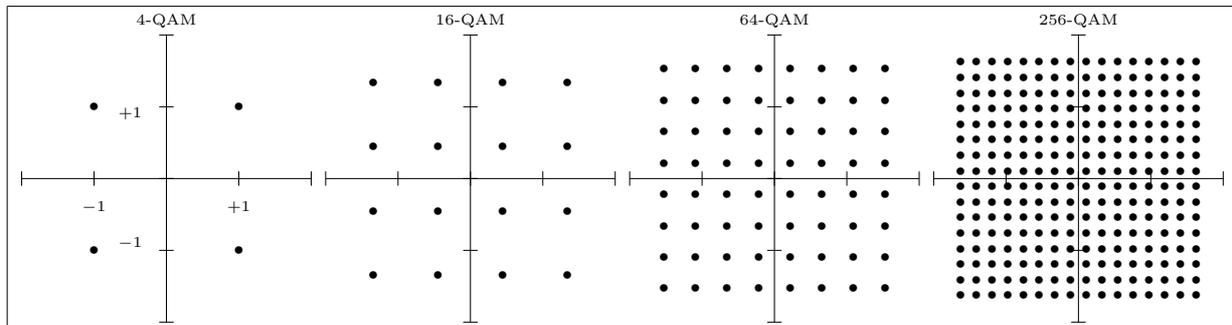


Figura 4.80: Distintas constelaciones M -QAM con la misma energía media por símbolo ($E_s = 2$).

4.7. Modulador

El modulador es el segundo elemento funcional del sistema de transmisión, tras el codificador. Su función consiste en transformar la secuencia de representaciones vectoriales de las señales a transmitir, $\mathbf{A}[n]$, en la señal modulada $s(t)$ que contiene la información a transmitir. Esta función se ilustra en la Figura 4.81. La forma de hacerlo será transformando cada vector de la secuencia $\mathbf{A}[n]$ en una señal de duración T que definirá la forma de señal en el intervalo de símbolo corres-

pendiente, $nT \leq t < (n + 1)T$, de tal modo que si $\mathbf{A}[n] = \mathbf{a}_k$ entonces $s(t) = s_k(t - nT)$ en $nT \leq (n + 1)T$, formando así por intervalos de símbolo la señal modulada $s(t)$.



Figura 4.81: Codificador en el transmisor de un sistema de comunicaciones digital.

4.7.1. Diseño del modulador

La función básica del modulador, como se ha dicho, es generar una señal a partir de su representación discreta. Como la relación entre señal y representación discreta viene dada por la expresión

$$s_i(t) = \sum_{j=0}^{N-1} a_{i,j} \times \phi_j(t),$$

lo que hace falta para realizar esta conversión es definir una base ortonormal de N elementos (dimensión del espacio de señales). Y en eso consiste el diseño del modulador, en la elección de esa base ortonormal de dimensión N

$$\text{Diseño del modulador: } \{\phi_0(t), \phi_1(t), \dots, \phi_{N-1}(t)\}.$$

Así, si $\mathbf{A}[n] = \mathbf{a}_k$ entonces $s(t) = s_k(t - nT)$ en $nT \leq (n + 1)T$, con lo que la expresión analítica de la señal completa es

$$s(t) \sum_n \sum_{j=0}^{N-1} A_j[n] \times \phi_j(t - nT),$$

donde $A_j[n]$ denota la coordenada de índice j del símbolo $\mathbf{A}[n]$, es decir

$$\mathbf{A}[n] = \begin{bmatrix} A_0[n] \\ A_1[n] \\ \vdots \\ A_{N-1}[n] \end{bmatrix}.$$

Así como el diseño del modulador se hacía teniendo en cuenta dos factores, prestaciones y energía, el diseño del modulador se hace en base a un único factor: las características del canal. Se debe buscar una base ortonormal cuyos elementos minimizen la distorsión lineal que sufre la señal durante su transmisión a través del canal. Si la respuesta del canal es $h(t)$ en el dominio temporal, o equivalentemente su transformada de Fourier $H(j\omega)$ en el dominio frecuencial, idealmente se desea que la base sea tal que no exista ninguna distorsión lineal, para lo que debe cumplirse la siguiente condición, expresada en el dominio temporal y en el frecuencial

$$\phi_i(t) * h(t) = \phi_i(t) \Leftrightarrow \Phi_i(j\omega) \times H(j\omega) = \Phi_i(j\omega).$$

Aunque en muchos casos en esta asignatura se considerará que se puede lograr una adecuación perfecta a las condiciones del canal, en la práctica esto no será posible, con lo que habrá que conformarse con seleccionar señales cuya respuesta en frecuencia esté en la banda de paso del canal de transmisiones, lo que da lugar a una distinción entre canales en banda base y canales paso banda.

4.7.2. Algunos ejemplos de moduladores y señales moduladas que generan

En esta sección se muestran algunos ejemplos de moduladores, y del tipo de señales moduladas que se generan con los mismos cuando se desea transmitir una cierta secuencia de bits de información.

El primer ejemplo es un sistema cuya constelación (CODIFICADOR) y base ortonormal (MODULADOR), son los mostrados en la Figura 4.82

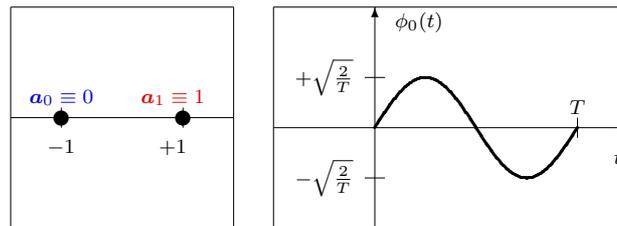


Figura 4.82: Constelación (CODIFICADOR) y base ortonormal (MODULADOR) de un sistema de comunicaciones. Ejemplo A.

Teniendo en cuenta que se trata de un sistema unidimensional, las señales asociadas a cada uno de los símbolos son

$$s_0(t) = -1 \times \phi_0(t), \quad s_1(t) = +1 \times \phi_0(t).$$

Al tratarse de un espacio de dimensión $N = 1$, las dos señales son réplicas escaladas, con distintos factores de escala, de la misma señal, la base ortonormal que define el modulador, $\phi_0(t)$. Las dos señales se muestran en la Figura 4.83

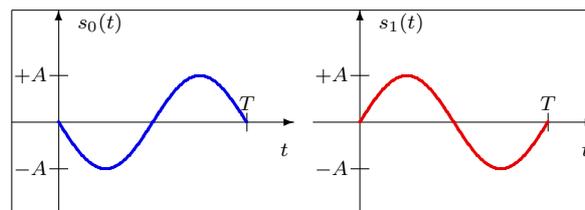


Figura 4.83: Señales asociadas a cada símbolo en el Ejemplo A.

Para la siguiente secuencia binaria de información

$$B_b[\ell] = 00101101100 \dots$$

la señal modulada correspondiente es la que muestra la Figura 4.84

En cada intervalo de símbolo se transmite un bit, que se transporta en una de las dos formas de onda, $s_0(t)$ (en azul) transporta el bit 0, y $s_1(t)$ (en rojo) transporta el bit 1.

En el segundo ejemplo, se considera un sistema binario pero ahora en un espacio bidimensional. La constelación (CODIFICADOR) y base ortonormal (MODULADOR) empleados en este caso son los mostrados en la Figura 4.85

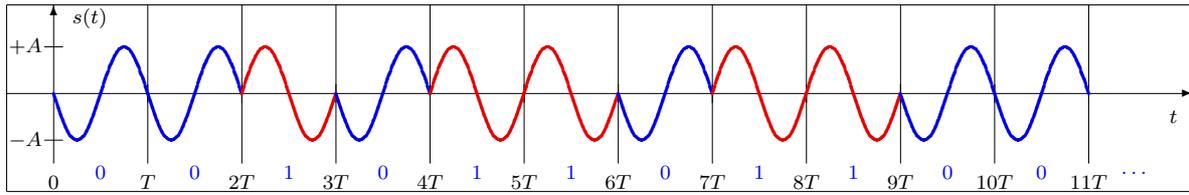


Figura 4.84: Señal modulada para la secuencia de información transmitida en el Ejemplo A.

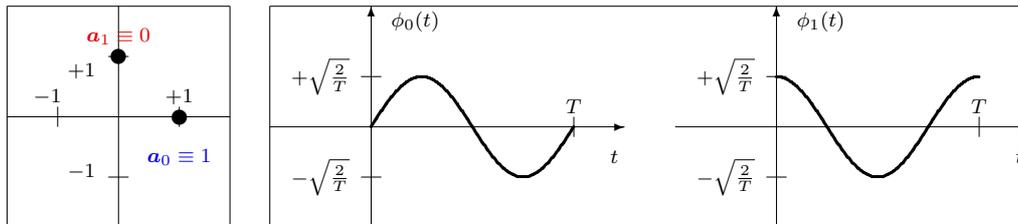


Figura 4.85: Constelación (CODIFICADOR) y base ortonormal (MODULADOR) de un sistema de comunicaciones. Ejemplo B.

Teniendo en cuenta que se trata de un sistema bidimensional, las señales asociadas a cada uno de los símbolos son

$$s_0(t) = +1 \times \phi_0(t) + 0 \times \phi_1(t), \quad s_1(t) = 0 \times \phi_0(t) + 1 \times \phi_1(t),$$

que a diferencia del caso anterior ahora se forman a partir de la combinación lineal de dos señales, los dos elementos de la base ortonormal que define el modulador. Las dos señales se muestran en la Figura 4.86

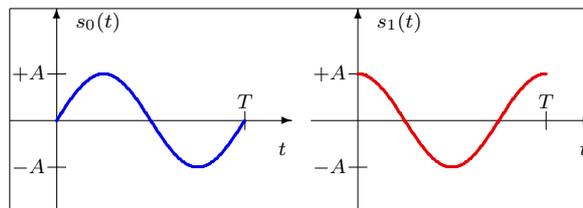


Figura 4.86: Señales asociadas a cada símbolo en el Ejemplo B.

Para la misma secuencia binaria de información que en el caso anterior

$$B_b[\ell] = 0 \ 0 \ 1 \ 0 \ 1 \ 1 \ 0 \ 1 \ 1 \ 0 \ 0 \ \dots ,$$

la señal modulada correspondiente es la que muestra la Figura 4.87

Al igual que en el ejemplo anterior, en cada intervalo de símbolo se transmite un bit, que se transporta en una de las dos formas de onda, $s_0(t)$ (en azul) transporta el bit 0, y $s_1(t)$ (en rojo) transporta el bit 1. Lo que cambia con respecto al ejemplo anterior es la forma que tienen ahora las señales que se utilizan para transmitir cada símbolo.

En el tercer ejemplo, utilizando el mismo modulador que en el caso anterior, ahora se utilizará una constelación de $M = 4$ símbolos. La constelación (CODIFICADOR) y base ortonormal (MODULADOR) empleados en este caso son los mostrados en la Figura 4.88

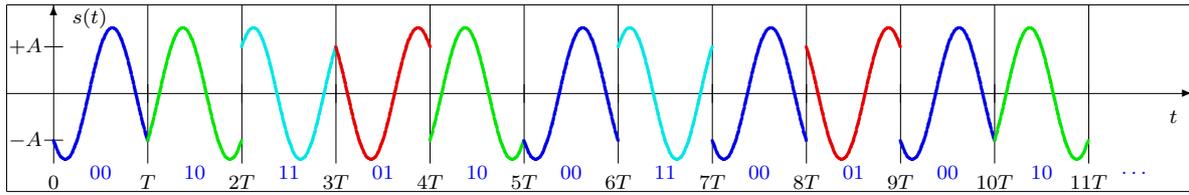


Figura 4.90: Señal modulada para la secuencia de información transmitida en el Ejemplo C.

verde) transporta la dupla 10, y $s_3(t)$ (en cyan) transporta la dupla 11. Ahora, al tratarse de una constelación de $M = 4$ símbolos, cada símbolo transporta $m = \log_2(M) = 2$ bits de información.

Finalmente, el cuarto ejemplo utiliza la misma constelación de cuatro símbolos que el ejemplo anterior, con idéntica asignación binaria, pero con otro modulador de dimensión $N = 2$, tal y como se muestra en la Figura 4.91

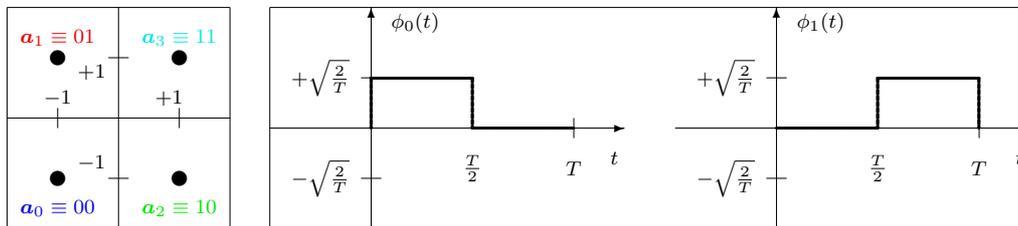


Figura 4.91: Constelación (CODIFICADOR) y base ortonormal (MODULADOR) de un sistema de comunicaciones. Ejemplo D.

La expresión analítica de las cuatro señales en función de los elementos de la base, al no cambiar la constelación, es la misma que en el ejemplo anterior

$$\begin{aligned}
 s_0(t) &= -1 \times \phi_0(t) - 1 \times \phi_1(t), \\
 s_1(t) &= -1 \times \phi_0(t) + 1 \times \phi_1(t), \\
 s_2(t) &= +1 \times \phi_0(t) - 1 \times \phi_1(t),
 \end{aligned}$$

y

$$s_3(t) = +1 \times \phi_0(t) + 1 \times \phi_1(t),$$

pero con distintos valores para las dos funciones que forman la base ortonormal, por lo que las cuatro señales resultantes son ahora las que se muestran en la Figura 4.92

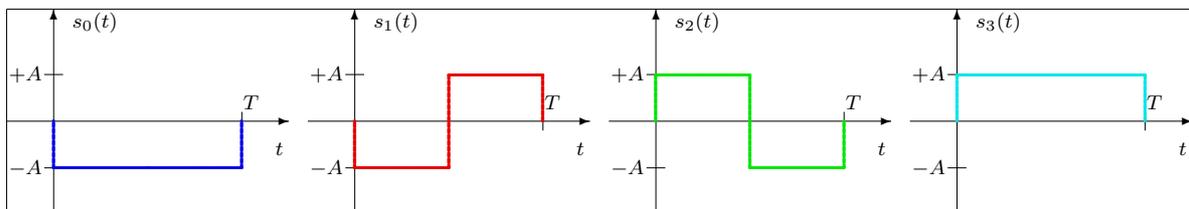


Figura 4.92: Señales asociadas a cada símbolo en el Ejemplo D.

Si se transmite la misma secuencia binaria de información que en el ejemplo anterior

$$B_b[\ell] = 00\ 10\ 11\ 01\ 10\ 00\ 11\ 00\ 01\ 00\ 10\ \dots,$$

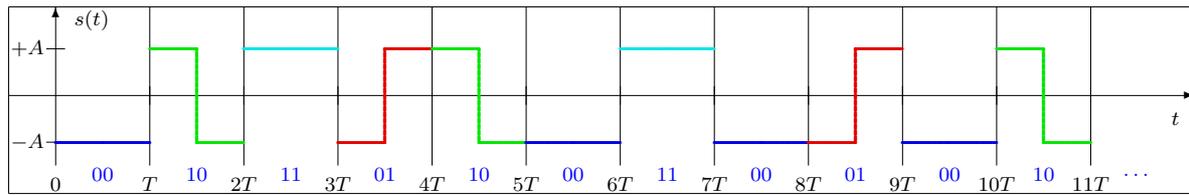


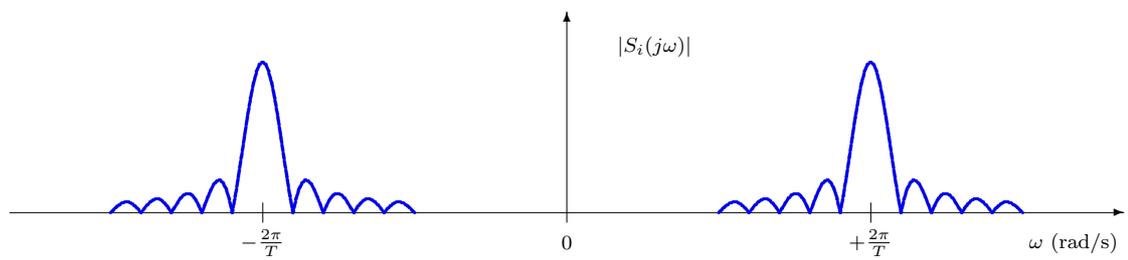
Figura 4.93: Señal modulada para la secuencia de información transmitida en el Ejemplo D.

la señal modulada correspondiente es la que muestra la Figura 4.93

De nuevo, en cada intervalo de símbolo se transmiten dos bits, que se transportan en una de las cuatro formas de onda, $s_0(t)$ (en azul) transporta la dupla 00, $s_1(t)$ (en rojo) transporta la dupla 01, $s_2(t)$ (en verde) transporta la dupla 10, y $s_3(t)$ (en cyan) transporta la dupla 11. Al tratarse de la misma constelación de $M = 4$ símbolos, cada símbolo transporta de nuevo $m = \log_2(M) = 2$ bits de información. Lo que cambia es la forma de onda que transporta cada dupla de bits, al haber cambiado la base ortonormal que define el modulador.

Si comparamos los sistemas de los dos últimos ejemplos, que comparten codificador, las prestaciones si se transmite sobre un canal gaussiano son idénticas, y las energías de los cuatro símbolos también lo son, ya que estas vienen determinadas por la elección del codificador (que contiene la representación discreta de las señales). Lo que cambia del tercer ejemplo al cuarto es la forma de onda de las señales y por tanto su respuesta en frecuencia. En el cuarto ejemplo tenemos señales cuadradas, cuya respuesta en frecuencia es banda base (centrada en 0 Hz), mientras que en el tercer ejemplo tenemos señales cuya respuesta en frecuencia está centrada en la frecuencia de las sinusoides que forman la base, por lo que es una respuesta paso banda, tal y como se muestra en la Figura 4.94. La elección de una base ortonormal (modulador) y otra, dependerá de las características del canal por el que queramos transmitir, y en particular en este caso, de la banda en la que se desee realizar la transmisión. Si el canal tiene respuesta paso bajo, será mejor emplear el modulador del cuarto ejemplo, que genera señales adecuadas para este tipo de canales. Si el canal tiene una respuesta paso banda, será mejor emplear el modulador del tercer ejemplo, que genera señales más adecuadas para este tipo de respuesta.

■ Ejemplo C



■ Ejemplo D

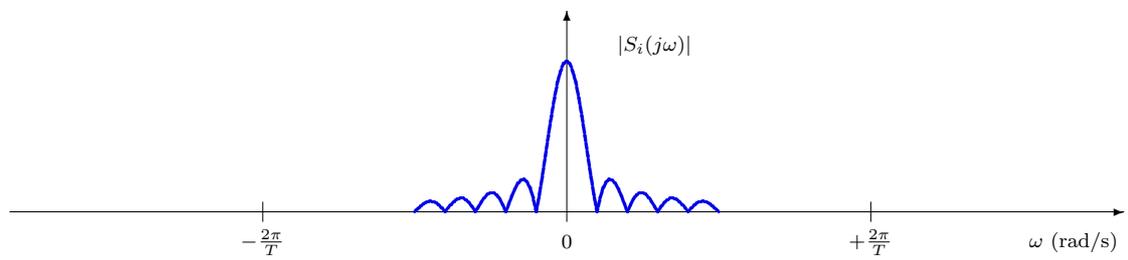


Figura 4.94: Respuesta en frecuencia de las señales de los ejemplos C y D.