Indice

Prefazione

Queste brevi note sono state scritte per gli studenti del corso di Trasmissione numerica da me tenuto presso la Facoltà di Ingegneria del Politecnico di Milano. Contengono materiale che si può trovare, in gran parte, su vari testi facilmente reperibili. Lo scopo principale è di fornirne una sintesi, senza che si debba estrarre l'informazione da più fonti e spesso con notazioni diverse. Un risultato collaterale è quello di indicare quali sono, a mio giudizio, gli argomenti che meritano maggior attenzione tra i moltissimi possibili sotto il titolo di Trasmissione numerica.

Il primo capitolo offre molti richiami di teoria dei segnali, selezionati e presentati in vista dell'applicazione alla trasmissione numerica. Viene introdotta la notazione, sono presentati i fondamenti della rappresentazione geometrica di segnali e rumore e sono indicati i risultati più significativi ai fini sia degli sviluppi teorici sia delle elaborazioni da effettuare nel ricevitore, con particolare attenzione alla realizzazione in forma numerica. I richiami sui segnali passa banda intendono soprattutto mostrare quanto sia sintetico e conveniente l'uso sistematico dell'equivalente passa basso. Non è invece previsto alcun "ripasso" di teoria della probabilità, che si dà per acquisita.

Nel secondo capitolo sono sintetizzati gli elementi fondamentali della teoria della decisione necessari per il progetto del ricevitore, nel caso di rumore additivo gaussiano e bianco nella banda dei segnali. Sono presentati e discussi i metodi per il calcolo della probabilità d'errore. Sono mostrati in modo informale alcuni esempi di insiemi di segnali, relativamente facili da analizzare ma con prestazioni già di un qualche interesse, allo scopo di introdurre quanto prima la nozione che i sistemi numerici efficienti vanno ricercati in spazi aventi un grande numero di dimensioni.

Dopo un breve capitolo di raffinamenti e complementi teorici, tra cui la ricezione in presenza di parametri indeterminati ed in particolare la ricezione non coerente, l'indagine sistematica dei sistemi efficienti di trasmissione numerica viene condotta nel quarto e quinto capitolo. Questi trattano rispettivamente i limiti teorici alle prestazioni raggiungibili quando non si ponga alcun limite alla complessità (capacità di canale) ed i sistemi codificati realizzabili in pratica, insieme ai metodi per il calcolo delle relative prestazioni.

Il sesto capitolo è dedicato ai canali con risposta impulsiva variabile nel tempo, in modo non prevedibile, ed alle varie tecniche sviluppate per rendere possibile la trasmissione anche su tali mezzi, assai importanti per le applicazioni alle telecomunicazioni.

Il successivo capitolo introduce gli elementi fondamentali della modulazione numerica di frequenza a fase continua, che è praticamente l'unica forma di modulazione di frequenza meritevole di considerazione nella trasmissione numerica. La continuità di fase equivale ad una qualche forma di codifica, che il ricevitore deve saper utilizzare per ottenere prestazioni degne di nota.

Preceduto da un capitolo sulle tecniche per la stima di parametri continui come fase, frequenza o posizione temporale di un segnale, l'ultimo capitolo considera la sincronizzazione di simbolo e di portante, cioè l'estrazione di tali fondamentali parametri dallo stesso segnale ricevuto. Fra i moltissimi metodi disponibili vengono privilegiati quelli che meglio si prestano alla realizzazione in forma numerica, anche a costo di lasciare un po' in ombra le tecniche analogiche, che peraltro lasciano sempre più spazio a quelle digitali.

Mancano non pochi argomenti, ed anche quelli selezionati potrebbero essere sviluppati in modo più approfondito. Ho preferito fornire una buona base che consenta, spero con una certa facilità, la comprensione della vasta letteratura scientifica disponibile. Del resto è inutile proporre a tutti gli studenti l'assimilazione di ogni dettaglio, ed è anzi pericoloso illudersi che ciò sia per loro conveniente. A mio parere c'è già molto materiale in questo testo, e la frazione di studenti che ne avrà veramente bisogno nella ormai imminente attività professionale non è così elevata come forse ci si potrebbe attendere in un paese industrializzato. Pochissimi comunque hanno idea di quale sarà la loro attività futura. Oso sperare che anche chi non utilizzerà appieno ciò che qui si tenta di insegnare ricorderà con piacere questi argomenti, come a me è capitato (da studente, in tempi ormai remoti) anche per materie che sapevo certamente "inutili". L'esperienza successiva non ha fatto che confermare che nel tempo dello studio si accumula, da qualsiasi fonte, una insospettata capacità di analisi dei problemi che rende poi soddisfacente qualunque attività di progetto, in ogni ramo.

Da quanto detto sopra si comprende che questo testo non è pensato prevalentemente per una rapida consultazione, ma piuttosto per un apprendimento più sistematico dei fondamenti della trasmissione numerica. Non si può escludere tuttavia che anche chi abbia già studiato, in passato, l'argomento possa qui trovare nuovi punti di vista ed anche utili riferimenti alla realizzazione

numerica del ricevitore.

Vi sono numerosi testi sulla trasmissione numerica, che è consigliabile consultare per approfondire argomenti specifici o anche solo per trovarvi una trattazione più adatta a esigenze particolari: c'è infatti chi ama la sintesi e chi le ripetizioni; chi pretende una strutturazione tutta fatta di definizioni, teoremi, corollari e lemmi, e chi invece preferisce uno stile discorsivo. È giusto che ognuno trovi il modo più efficace per apprendere. Fra i tanti, segnalo i seguenti testi

- Benedetto S., Biglieri E., Castellani V., 1987: *Digital Transmission Theory*, Prentice-Hall, 639 pp., 6 appendici, esercizi e ricca bibliografia (disponibile anche in traduzione italiana). Molto accurato, rivolto un po' più alla teoria che alla pratica (per fare onore al titolo), contiene anche un capitolo dedicato alle nonlinearità.

- Proakis, J. G., 1989: *Digital Communications*, McGraw-Hill (seconda edizione; in arrivo la terza), 905 pp., 10 appendici, esercizi e ricca bibliografia. Contiene moltissimo materiale, anche sulla trasmissione in diversità su canali *multipath* e sui sistemi *spread spectrum*; ampia la parte dedicata all'equalizzazione.

- Tartara, G., 1986: *Teoria dei sistemi di comunicazione*, Boringhieri, 198 pp., bibliografia. Sintetica e chiara la trattazione della rappresentazione geometrica dei segnali, della teoria della ricezione ottima e della stima di parametri e l'introduzione ai sistemi codificati; contiene anche materiale relativo alla rappresentazione geometrica delle variabili casuali.

Per chi volesse approfondire la teoria dei codici suggerisco

- Clark, G. C., Cain, J. B., 1981: *Error-Correction Coding for Digital Communications*, Plenum, 422 pp., 2 appendici, esercizi e abbondante bibliografia. Uno dei più noti testi sui codici; vi si trovano un'ottima sintesi dell'algebra dei campi di *Galois*, senza troppi teoremi, e molte indicazioni pratiche.

- Michelson, A. M., Levesque, A. H., 1985: *Error-Control Techniques for Digital Communication*, Wiley, 465 pp., 2 appendici e ricca bibliografia. Forse meno noto del precedente, ma altrettanto raccomandabile; teoria un po' più completa, ma anche molti esempi numerici e indicazioni pratiche.

Tornando a questo testo, è molto difficile quando ci si accinge ad un'opera così coinvolgente, pur nella sua limitatezza, resistere alla tentazione di includere qualche divagazione che forse potrebbe essere trascurata senza troppo rimpianto, ma che tuttavia può incuriosire qualcuno in grado di trarne vantaggio; naturalmente occorre poi nelle lezioni dare ad ogni parte il suo peso.

È bene avvertire il lettore che per semplificare la notazione sono frequentemente sottintesi gli estremi di integrali e somme, qualora siano infiniti oppure chiaramente deducibili dal contesto.

Numerosi sono gli esercizi proposti. Alcuni richiedono tempo, fantasia, curiosità o strumenti numerici e non sono raccomandabili per il lettore frettoloso, che comunque è invitato a cimentarsi perlomeno con i seguenti:
Cap. 1 - 5, 8, 11, 15, 19, 20
Cap. 2 - 5, 10, 12, 14
Cap. 3 - 1, 7, 8, 11, 15
Cap. 4 - 1, 5
Cap. 5 - 6, 7, 10, 12, 15, 17, 22, 23, 31, 35
Cap. 6 - 6, 12, 13, 14, 21, 22
Cap. 7 - 1, 7, 8
Cap. 8 - 2, 4, 5, 6, 7
Cap. 9 - 2, 4, 6, 13, 14

Inevitabili naturalmente gli errori. Posso solo sperare che non siano troppo numerosi ed invocare la comprensione di chi legge, chiedendo la cortesia di segnalarmeli.

Come tutti i miei lavori precedenti, piccoli o meno, anche questo è dedicato a Ilia, mia moglie.

Sandro Bellini

Capitolo 1

Rappresentazione dei segnali

1.1 Introduzione

Un *segnale* per la trasmissione a distanza dell'informazione consiste in una forma d'onda, cioè nell'andamento temporale di una qualche grandezza fisica $s(t)$, interpretabile dal destinatario senza ambiguità non solo nel caso ideale di ricezione senza deformazioni ma anche in presenza dei disturbi tipici del collegamento. Questi sono inevitabili, e quindi la scelta delle forme d'onda dovrà essere tale da consentire, con elevata probabilità, l'interpretazione del segnale *ricevuto*.

Più in generale si potrebbe definire segnale ogni forma di rappresentazione, intenzionale o meno, dell'informazione: segnali di fumo, radiazioni emesse da corpi celesti, atteggiamenti di un corpo umano, e così via. Sarà presto evidente che le tecniche descritte nel seguito non sono affatto appropriate per tali situazioni. Piuttosto esse sono mirate al caso di trasmissione dell'informazione mediante forme d'onda.

Si considererà quasi esclusivamente la trasmissione di informazione di tipo numerico, cioè discreto (*trasmissione numerica*). Il caso limite è la trasmissione di un solo bit d'informazione, un *sì* o un *no* che risponda ad una domanda che non ammette altre alternative. Si dovranno prevedere *due* forme d'onda, $s_1(t)$ ed $s_2(t)$, e si trasmetterà quella associata al messaggio da inviare. Il ricevitore avrà il compito di *decidere* quale forma d'onda è stata trasmessa, e quindi quale messaggio è stato inviato. Naturalmente nulla poi vieta di trasmettere una sequenza, anche molto lunga, di bit mediante una successione di forme d'onda.

Informazioni più variegate (come le quattro alternative: sì, no, forse, non so) sono rappresentabili sia con quattro diverse forme d'onda $s_1(t), \ldots, s_4(t)$ (trasmissione *quaternaria*) sia con una coppia di bit (trasmissione *binaria*, ad un *ritmo doppio* nel caso di trasmissione continua).

È quasi inutile ricordare che anche i segnali *analogici*, cioè continui nelle ampiezze come quello telefonico, televisivo, ecc. sono sempre più spesso rappresentati in forma numerica, mediante sequenze binarie. Uno dei motivi per questa scelta è che il sistema di trasmissione numerico è in certo senso universale. Per molti aspetti sapere quale tipo si segnale si stia trasmettendo è richiesto solo alle apparecchiature terminali che eseguono codifica di sorgente e multiplessaggio dei segnali, e le operazioni inverse in ricezione.

Per il progetto del sistema di trasmissione si traducono i requisiti di qualità sui segnali analogici in semplici specifiche sulla probabilità d'errore ed eventualmente sul ritardo tollerabili nella trasmissione numerica.

I supporti fisici più classici (le onde elettromagnetiche utilizzate per ponti radio, satelliti, sistemi radiomobili e diffusione di segnali vocali e televisivi; i cavi coassiali e le linee bifilari per telefonia e servizi collegati; ecc.) presentano per la trasmissione numerica un insieme di problemi abbastanza omogeneo. Tutte le tecniche discusse nel seguito sono, in maggiore o minor misura, applicabili a questi sistemi.

Un caso a sé, almeno in qualche misura, è quello delle comunicazioni ottiche, che non viene qui trattato. In particolare perlomeno nei sistemi non coerenti, che sono i più diffusi, è diversa la statistica del rumore che si sovrappone al segnale nel rivelatore ottico. Certamente non si vuol dire che il contenuto di un corso di Trasmissione numerica è irrilevante per le comunicazioni ottiche, ma piuttosto che l'argomento merita una trattazione specifica, anche per le particolari tecnologie impiegate.

Un altro caso un po' particolare e per ora non trattato in questo testo è quello dei sistemi a spettro espanso (*spread spectrum*), un tempo presi in considerazione solo per applicazioni militari con speciali requisiti di riservatezza ma ora assai promettenti per i sistemi radiomobili civili. In tali sistemi l'interferenza principale è quella reciproca tra i vari utenti, mentre almeno in prima approssimazione è trascurabile il rumore additivo indipendente dai segnali.

È ben noto che se disturbo deve esserci è di gran lunga preferibile che sia gaussiano[1]. L'analisi diventa quasi facile anche in situazioni complesse, senza

[1] senza disturbo non c'è divertimento; la singola realizzazione del rumore, pur imprevedibile, non prende di sorpresa chi sa dominare il calcolo del comportamento *medio* del rumore

dover invocare troppe approssimazioni dalla validità incerta. In un certo senso conviene imparare prima a maneggiare il caso del rumore gaussiano, per cui si trovano strumenti potenti e sintetici. Del resto molto spesso il disturbo è così gentile da adattarsi a questo modello.

1.2 Rappresentazione geometrica dei segnali

Scomporre una generica forma d'onda $s(t)$ in somma di opportune *funzioni base* è operazione assai comune, particolarmente utile nel caso di sistemi di trasmissione lineari ma talvolta anche in presenza di nonlinearità. Molto noto è ad esempio, per una forma d'onda di durata limitata all'intervallo $(0, T_0)$, lo sviluppo in serie di Fourier nella forma esponenziale

$$s(t) = \sum_{k=-\infty}^{\infty} s_k \exp(j2\pi kt/T_0) \tag{1.1}$$

o, limitatamente al caso di segnali reali, nella corrispondente forma trigonometrica.

La generalità della serie di Fourier, che è in grado di rappresentare praticamente tutte le forme d'onda possibili in natura, è pagata quasi sempre con un numero teoricamente infinito di funzioni base $\Phi_k(t) = \exp(j2\pi kt/T_0)$. Ciò anche per rappresentare *una sola* funzione $s(t)$, o un insieme *finito* di funzioni $s_i(t)$; in quest'ultimo caso il coefficiente k-esimo del segnale i-esimo sarà indicato con s_{ik}.

Se la (1.1) vale, è ben nota l'espressione dei coefficienti

$$s_k = \frac{1}{T_0} \int_0^{T_0} s(t) \exp(-j2\pi kt/T_0) dt \tag{1.2}$$

che è ottenibile moltiplicando la (1.1) per $\Phi_n^*(t) = \exp(-j2\pi nt/T_0)$ e integrando da 0 a T_0. Infatti risulta determinante l'*ortogonalità* tra le funzioni base, facilmente verificabile,

$$\int_0^{T_0} \Phi_k(t)\Phi_n^*(t)dt = \begin{cases} T_0 & k = n \\ 0 & k \neq n \end{cases} \tag{1.3}$$

per cui il risultato dell'integrale si riduce al solo termine $T_0 s_n$. Ridenominando n in k e dividendo per T_0 si ottiene la (1.2).

Gli esponenziali, o i seni e coseni, non hanno un ruolo specifico. Espressioni come la (1.1) e (1.2) si otterrebbero con un *diverso* insieme di funzioni base

ortogonali. Limitandosi al caso di funzioni base *reali*, naturalmente atte a rappresentare solo forme d'onda $s_i(t)$ reali, si ha

$$s_i(t) = \sum_k s_{ik}\Phi_k(t) \tag{1.4}$$

$$s_{ik} = \int s_i(t)\Phi_k(t)dt \tag{1.5}$$

dove si è ottenuta una piccola semplificazione imponendo che l'energia delle funzioni base sia unitaria:

$$\int \Phi_k^2(t)dt = 1 \tag{1.6}$$

La normalizzazione, peraltro del tutto inessenziale, si ottiene scalando le funzioni $\Phi_k(t)$; dopo tale operazione le funzioni base sono dette *ortonormali*: ortogonali e normalizzate.

Si noti che si usa uno stesso insieme di funzioni base $\Phi_k(t)$ per tutti i segnali $s_i(t)$. Le (1.4) e (1.5) costituiscono le espressioni per la *sintesi* della forma d'onda (somma di ingredienti elementari, in quantità opportune) e l'*analisi* (determinazione della quantità richiesta di ciascun ingrediente).

Il numero di funzioni $s_i(t)$ da rappresentare, che dipende dalla quantità di informazione da trasmettere, verrà indicato nel seguito con M.

Gli estremi della somma nella (1.4), e quindi il numero di coefficienti s_{ik} richiesti per rappresentare il segnale $s_i(t)$, verranno solitamente sottintesi. Le funzioni base possono essere in numero finito, nel qual caso gli indici saranno numerati da 1 ad N, o infinito (e si potrà comunque numerare a partire da 1). Si vedrà nel seguito, indicando una procedura costruttiva, che è sempre possibile rappresentare un numero finito M di segnali con un numero finito $N \leq M$ di funzioni base. In non pochi casi la (1.4) non è solo una espansione lecita per la forma d'onda $s_i(t)$, ma corrisponde al modo in cui essa viene effettivamente generata in trasmissione.

La durata delle forme d'onda (eventualmente infinita, in teoria, ma in pratica sempre finita) e gli estremi dell'intervallo saranno normalmente sottintesi. In ogni caso si supporranno finite le energie.

Una conseguenza quasi immediata dell'espansione (1.4) in somma di funzioni ortonormali è

$$\int s_i(t)s_j(t)dt = \int \sum_k s_{ik}\Phi_k(t) \sum_n s_{jn}\Phi_n(t)dt = \sum_k s_{ik}s_{jk} \tag{1.7}$$

L'ultima espressione è ottenuta scambiando integrale e somme, e tenendo solo i termini con $k = n$. La proprietà ha un analogo, ben noto, nel caso della serie di Fourier.

L'integrale del prodotto di due funzioni viene detto *correlazione* o anche *prodotto scalare* delle due funzioni. Infatti $\sum_k s_{ik} s_{jk}$ è l'espressione del prodotto scalare di due vettori in N dimensioni con componenti cartesiane rispettivamente s_{ik} e s_{jk}. Dunque ai fini del calcolo della correlazione le funzioni si comportano come *vettori* con componenti pari ai coefficienti (1.5) dello sviluppo.

Nel caso $i = j$ la correlazione o prodotto scalare è l'energia della forma d'onda, ed è pari al quadrato della lunghezza del vettore.

Il vettore con componenti s_{ik} verrà indicato nel seguito con $\mathbf{s}_i$, ed il generico prodotto scalare con $\mathbf{s}_i \cdot \mathbf{s}_j$. Il prodotto scalare $\mathbf{s} \cdot \mathbf{s}$, pari al quadrato della lunghezza del vettore, e all'energia della forma d'onda, verrà normalmente indicato con $|\mathbf{s}|^2$, e quindi $|\mathbf{s}|$ sarà la lunghezza del vettore. Infine $|\mathbf{s}_i - \mathbf{s}_j|$ è la distanza tra gli estremi dei vettori $\mathbf{s}_i$ ed $\mathbf{s}_j$, che viene anche detta distanza tra le forme d'onda $s_i(t)$ e $s_j(t)$. Il suo quadrato è l'energia della differenza tra le forme d'onda.

Si noti che per calcolare prodotti scalari e distanze non è affatto necessario scegliere una base e calcolare le componenti dei vettori. Tutti i calcoli si possono effettuare come integrali nella variabile t. In un certo senso la rappresentazione geometrica è uno strumento concettualmente potente, come si vedrà, ma non necessariamente il più comodo mezzo di calcolo.

Le funzioni base $\Phi_k(t)$ hanno componenti cartesiane tutte nulle, eccetto la k-esima pari a uno; infatti volendo sintetizzare la funzione $\Phi_k(t)$ basta sommare la sola funzione base $\Phi_k(t)$, con peso uno! Il prodotto scalare di funzioni base diverse è nullo. I corrispondenti vettori, di lunghezza unitaria (in accordo con l'energia unitaria) e diretti secondo gli assi possono essere considerati i versori di un sistema di assi cartesiani ortogonali[2]. L'espressione (1.5) per il calcolo della k-esima componente del vettore $\mathbf{s}_i$ può essere interpretata come il prodotto scalare tra il vettore $\mathbf{s}_i$ ed il k-esimo versore Φ_k.

Ovviamente si potrebbe decidere di utilizzare un *diverso* insieme di funzioni base $\Phi'_k(t)$. I coefficienti s'_{ik} e s'_{jk}, cioè le componenti dei vettori, avrebbero valori diversi; resterebbe però immutato il prodotto scalare, pari all'integrale del prodotto delle due funzioni $s_i(t)$ e $s_j(t)$. In realtà non sarebbero cambiati i vettori, ma solo ruotato il sistema di assi cartesiani di riferimento, essendo

[2] ciò spiega perché le funzioni base sono dette ortogonali

evidente che le nuove funzioni base $\Phi'_k(t)$ non sono altro che combinazioni lineari (ortogonali) delle precedenti $\Phi_k(t)$, e viceversa.

Da quanto visto sembra di poter concludere che le forme d'onda si comportano come vettori, anche senza che si sia scelta esplicitamente una base. Effettivamente è pressoché immediato verificare che la correlazione tra due generiche funzioni $x(t)$ e $y(t)$ a energia finita esiste e soddisfa tutte le proprietà richieste ad un prodotto scalare (es. 1.1). È quindi lecito assegnare alle funzioni a energia finita tutte le *proprietà geometriche* dei vettori. Per fare solo due esempi è ben noto che il modulo del prodotto scalare tra vettori non può superare il prodotto dei moduli

$$|\mathbf{s}_i \cdot \mathbf{s}_j| \leq |\mathbf{s}_i||\mathbf{s}_j| \tag{1.8}$$

e che vale la diseguaglianza triangolare

$$||\mathbf{s}_i| - |\mathbf{s}_j|| \leq |\mathbf{s}_i \pm \mathbf{s}_j| \leq |\mathbf{s}_i| + |\mathbf{s}_j| \tag{1.9}$$

Tradotte in chiaro esplicitando i prodotti scalari tra funzioni, tali proprietà diventano rispettivamente

$$\Big| \int s_i(t) s_j(t) dt \Big| \leq \sqrt{\int s_i^2(t) dt} \sqrt{\int s_j^2(t) dt} \tag{1.10}$$

(nota anche come diseguaglianza di Schwartz) e

$$\begin{aligned} &\Big| \sqrt{\int s_i^2(t) dt} - \sqrt{\int s_j^2(t) dt} \Big| \leq \sqrt{\int (s_i(t) \pm s_j(t))^2 dt} \leq \\ &\leq \sqrt{\int s_i^2(t) dt} + \sqrt{\int s_j^2(t) dt} \end{aligned} \tag{1.11}$$

1.3 Ortogonalizzazione di Gram-Schmidt

Dato un insieme $\{s_i(t)\}$ di segnali ($i = 1, \ldots, M$) è utile mostrare come sia sempre possibile ottenere un insieme *finito* di funzioni base atte a rappresentarli. Lo scopo non è tanto di dare un procedimento pratico, raramente utilizzato per gli insiemi di segnali tipici della trasmissione numerica, quanto di giustificare *in ogni caso* il ragionare in termini geometrici.

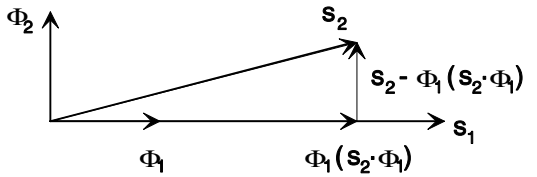

Fig. 1.1 - Ortogonalizzazione di Gram-Schmidt

Se si dovesse rappresentare una sola forma d'onda $s_1(t)$ verrebbe naturale scegliere come funzione base la funzione stessa, normalizzata

$$\Phi_1(t) = \frac{s_1(t)}{\sqrt{\int s_1^2(t)dt}} \tag{1.12}$$

relazione esprimibile molto più sinteticamente in forma vettoriale:

$$\Phi_1 = \frac{\mathbf{s}_1}{|\mathbf{s}_1|} \tag{1.13}$$

Può poi accadere che la seconda funzione $s_2(t)$ sia proporzionale a $s_1(t)$, e quindi a $\Phi_1(t)$. In tal caso non occorre introdurre una seconda funzione base. Altrimenti si cerca un secondo versore, ortogonale al primo e giacente nel piano di $\mathbf{s}_1$ e $\mathbf{s}_2$. Geometricamente si procede sottraendo a $\mathbf{s}_2$ la sua proiezione sull'asse Φ_1. Il vettore differenza, non nullo, viene poi normalizzato.

Osservando che la lunghezza della proiezione di $\mathbf{s}_2$ su Φ_1 è pari a $\mathbf{s}_2 \cdot \Phi_1$ e la direzione è quella di Φ_1 (fig. 1.1), si ha quindi

$$\Phi_2 = \frac{\mathbf{s}_2 - \Phi_1(\mathbf{s}_2 \cdot \Phi_1)}{|\mathbf{s}_2 - \Phi_1(\mathbf{s}_2 \cdot \Phi_1)|} \tag{1.14}$$

In termini di forme d'onda si ha evidentemente

$$\Phi_2(t) = \frac{s_2(t) - \Phi_1(t) \int s_2(t)\Phi_1(t)dt}{\sqrt{\int \left(s_2(t) - \Phi_1(t) \int s_2(t)\Phi_1(t)dt\right)^2 dt}} \tag{1.15}$$

Per un terzo segnale $s_3(t)$ si può procedere allo stesso modo, ottenendo

$$\Phi_3 = \frac{\mathbf{s}_3 - \Phi_1(\mathbf{s}_3 \cdot \Phi_1) - \Phi_2(\mathbf{s}_3 \cdot \Phi_2)}{|\mathbf{s}_3 - \Phi_1(\mathbf{s}_3 \cdot \Phi_1) - \Phi_2(\mathbf{s}_3 \cdot \Phi_2)|} \tag{1.16}$$

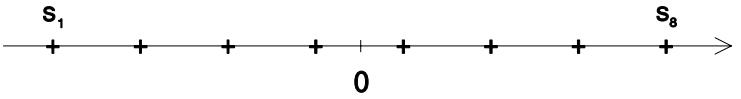

Fig. 1.2 - $M = 8$ segnali in una sola dimensione

naturalmente solo se il numeratore non è nullo. Il numeratore eventualmente nullo indica che $\mathbf{s}_3$ è esprimibile come combinazione lineare di Φ_1 e Φ_2 e che quindi non occorre un altro asse.

Procedendo fino all'ultimo segnale, naturalmente saltando i passi in cui il numeratore risulta nullo, si determinano $N \leq M$ funzioni base. Le funzioni base ottenute *dipendono* dall'ordinamento dei segnali, come è evidente già nel caso dei due soli vettori non ortogonali di fig. 1.1. La differenza consiste in una rotazione di assi. Si potrebbe mostrare, ma è talmente intuitivo da non giustificarne la fatica, che il numero di assi richiesti non dipende dall'ordinamento dei segnali.

Il procedimento appena descritto per la costruzione delle funzioni base mostra che non solo i segnali sono combinazioni lineari delle funzioni base, ma anche viceversa le funzioni base sono combinazioni lineari dei segnali. Se dunque si è interessati, come quasi sempre accade, a segnali ben limitati in banda, tali sono anche le funzioni base[3]. In non pochi casi le N funzioni base vengono scelte a priori, con i requisiti di banda e durata richiesti, ed il progetto dell'insieme dei segnali $s_i(t)$ equivale alla scelta della loro disposizione geometrica nello spazio ad N dimensioni.

Prima ancora di vedere qualche esempio di funzioni base utilizzate in pratica, può essere opportuna una spiegazione intuitiva dei motivi che portano ad utilizzare spazi con un gran numero di dimensioni. Occorre solo premettere il risultato, giustificato nel Cap. 2, che in presenza di rumore bianco nella banda dei segnali la probabilità d'errore di un sistema di trasmissione numerica dipende dalle distanze relative tra i segnali, e in prima approssimazione solo dalla distanza minima.

Avendo a disposizione una sola dimensione non si può fare altro che disporre gli estremi dei vettori $\mathbf{s}_i$ *equispaziati* sull'unico asse. Ad esempio in

[3] non è detto che ciascuna funzione base occupi tutta la banda B e tutto l'intervallo T_0, ma solo che l'insieme delle loro combinazioni lineari richiede tutta la banda e tutto l'intervallo di tempo; del resto non sempre ciascun segnale $s_i(t)$ occupa l'intera banda e l'intera durata

fig. 1.2 sono mostrati otto segnali per la modulazione d'ampiezza multilivello (PAM: *Pulse Amplitude Modulation*)[4].

Si noti che i segnali sono tutti proporzionali all'unica funzione base. Fissata la distanza minima, l'unico parametro disponibile è il numero M dei segnali, ciò che non consente una grande varietà di soluzioni. Ma soprattutto è evidente che aumentando il numero di segnali l'energia di quelli di ampiezza maggiore diventa molto elevata; è infatti pari al *quadrato* della lunghezza del vettore.

Potendo disporre l'insieme degli M segnali in più dimensioni si ottiene innanzitutto il risultato che l'energia dei segnali, non più allineati su un asse, può essere considerevolmente minore a parità di distanza minima. Inoltre c'è una maggior varietà di disposizioni possibili. Se il numero di dimensioni non avesse un costo sarebbe quindi naturale disperdere i segnali in un gran numero di dimensioni. Si vedrà nel seguito che i sistemi di trasmissione numerica efficienti occupano effettivamente un grande numero di dimensioni, anche se la geometria dei segnali in molte dimensioni è per gli esseri umani poco intuitiva.

1.4 Relazioni tra durata, banda e numero di dimensioni

I primi parametri fondamentali di un sistema di trasmissione numerica sono il ritmo di trasmissione dell'informazione R (bit/s) e la banda occupata B. Se il segnale trasmesso $s_i(t)$ può presentarsi in M configurazioni distinte l'informazione trasmessa è pari a $m = \log_2 M$ bit. Se la trasmissione del segnale e quindi degli m bit avviene in un tempo T_0 si ha

$$R = \frac{m}{T_0} = \frac{\log_2 M}{T_0} \tag{1.17}$$

Si può dunque ottenere un prefissato ritmo di trasmissione dell'informazione in molti modi diversi. Il numero N di dimensioni dello spazio dei segnali dipende da quante funzioni ortogonali è possibile costruire rispettando i vincoli sulla durata T_0 e sulla banda B.

Lasciando perdere il risultato teorico che non esistono forme d'onda con durata e banda limitate[5], ma accettando invece il teorema pratico per cui *tutte* le forme d'onda in natura hanno durata e banda limitate, il legame

[4] nel rappresentare le *costellazioni* di segnali si indicano solo gli estremi dei vettori, per non rendere le figure incomprensibili

[5] è un ottimo esempio di *teorema inutile*, in quanto anche non facile da dimostrare

tra banda, durata e numero di funzioni ortogonali che si possono costruire è approssimabile con

$$N = 2BT_0 \tag{1.18}$$

da considerarsi valido solo per $N \gg 1$. Una semplice giustificazione intuitiva è la seguente[6]. Si considerino le funzioni

$$\Phi_k(t) = \frac{1}{\sqrt{T}} \frac{\sin \pi(t - kT)/T}{\pi(t - kT)/T} \tag{1.19}$$

Mediante il teorema di Parseval è immediato verificare che esse sono ortonormali (es. 1.3). Di una forma d'onda $s(t)$ con banda $B \leq 1/2T$ si calcolino le componenti

$$\begin{aligned} s_k &= \int s(t)\Phi_k(t)dt = \int s(t) \frac{1}{\sqrt{T}} \frac{\sin \pi(t - kT)/T}{\pi(t - kT)/T} dt = \\ &= \int s(\tau) \frac{1}{\sqrt{T}} \frac{\sin \pi(kT - \tau)/T}{\pi(kT - \tau)/T} d\tau = \sqrt{T} s(kT) \end{aligned} \tag{1.20}$$

Infatti nell'ultima convoluzione si riconosce l'uscita di un filtro passa basso ideale con banda $1/2T$ e guadagno $\sqrt{T}$, calcolata all'istante kT; poiché il segnale $s(t)$ è già limitato in banda il filtro si limita a moltiplicare per $\sqrt{T}$.

La (1.4) diventa

$$\sum s_k \Phi_k(t) = \sum s(kT) \frac{\sin \pi(t - kT)/T}{\pi(t - kT)/T} = s(t) \tag{1.21}$$

in virtù della formula interpolante fornita dal teorema del campionamento. Dunque l'espansione vale per qualunque forma d'onda a banda limitata, e le componenti del vettore sono sostanzialmente i campioni della forma d'onda[7].

In pratica tutte le forme d'onda, oltre ad avere banda limitata, hanno durata limitata. Se T_0 è la durata, il numero dei campioni non nulli e quindi il numero di funzioni base richieste è $N \approx T_0/T = 2BT_0$. Naturalmente ciò

[6] il più semplice teorema a questo proposito è talmente involuto che si fatica a capirne l'enunciato

[7] se si calcolano secondo la (1.20) i coefficienti s_k di una forma d'onda $s(t)$ con banda maggiore di $1/2T$ non si ottengono i campioni della forma d'onda, e se si utilizza l'espansione (1.21) si ottiene una replica di $s(t)$ *filtrata* con il passa basso ideale con banda $1/2T$; essa è l'approssimazione a distanza minima possibile da $s(t)$

ha senso solo se risulta $N \gg 1$, cioè se è lecito trascurare effetti di bordo. Né si può sperare di meglio di $N = 2BT_0$ perché il teorema del campionamento non consente di rappresentare la forma d'onda con un numero inferiore di campioni. Analoghe considerazioni valgono per segnali passa banda.

Altri semplici insiemi di funzioni base che possono essere proposti portano alle stesse conclusioni (es. 1.5-1.7).

Da $N = 2BT_0$ deriva che

$$B = \frac{N}{2T_0} = R\frac{N}{2\log_2 M} \tag{1.22}$$

e questa mostra che il rapporto tra ritmo di trasmissione dell'informazione e banda dipende solo dal numero di bit trasmessi per dimensione $\log_2 M/N$. Naturalmente si ha anche

$$R = 2B\frac{\log_2 M}{N} \tag{1.23}$$

e quindi il numero di bit trasmessi per unità di tempo e di banda (bit/s/Hz) è pari al doppio del numero di bit trasmessi per dimensione.

Naturalmente è possibile utilizzare un numero di funzioni base minore del massimo teorico[8]. Quindi la (1.22) deve essere vista come un limite inferiore alla banda occupata, e la (1.23) come un limite superiore al ritmo di trasmissione realizzabile.

Dunque per un confronto a parità di efficienza spettrale tra sistemi di trasmissione, un insieme di 8 segnali in una dimensione va paragonato a 64 segnali in due dimensioni, 512 in tre, e così via. Si nota immediatamente come il numero dei segnali cresca *esponenzialmente* con il numero delle dimensioni.

Si consideri la forma d'onda

$$s_i(t) = \sum_{k=1}^{N} a_k g(t - kT) \tag{1.24}$$

dove le funzioni $g(t - kT)$ sono repliche traslate di una stessa funzione $g(t)$ scelta in modo da ottenere l'ortogonalità tra le repliche, ed i livelli possibili sono $a_k = \pm 1$.

Il segnale $s_i(t)$ può presentarsi in $M = 2^N$ configurazioni, e lo spazio dei segnali ha N dimensioni. Il numero di bit per dimensione è $\log_2 M/N = 1$,

[8] in pratica c'è sempre un qualche *eccesso di banda*, solitamente compreso tra il 20% e il 50%

indipendentemente da N, e può dunque essere calcolato anche come rapporto tra il numero di *bit per simbolo*, intendendo con simbolo la singola forma d'onda $a_k g(t - kT)$, e il numero di *dimensioni per simbolo*:

$$\frac{\log_2 M}{N} = \frac{\text{bit}}{\text{dimensioni}} = \frac{\text{bit/simbolo}}{\text{dimensioni/simbolo}} \tag{1.25}$$

In altri termini la (1.24) può essere interpretata sia come la trasmissione di $\log_2 M = N$ bit in N dimensioni, sia come la trasmissione di un bit in una sola dimensione, ripetuta successivamente N volte[9]. In questo senso può essere accettabile, come peraltro comunemente si fa anche se non propriamente corretto, indicare con M sia il numero complessivo di segnali ($M = 2^N$) sia il numero di livelli utilizzati per un solo simbolo ($M = 2$). Se si considera la trasmissione di un solo simbolo per volta si hanno due segnali in una sola dimensione, opposti l'uno all'altro (segnali *antipodali*). Se invece si considerano congiuntamente gli N simboli, gli estremi dei segnali sono i 2^N vertici dell'ipercubo in N dimensioni. Si vedrà in seguito che questo non è affatto il modo migliore di disporre 2^N segnali in N dimensioni; invece, come già detto, in una sola dimensione non si saprebbe cosa altro proporre[10].

Per quanto riguarda la durata T_0, se nella (1.24) N è sufficientemente grande si può ritenere $T_0 \approx NT$ indipendentemente dalla durata di $g(t)$. La minima banda richiesta è quindi

$$B = \frac{N}{2T_0} = \frac{1}{2T} \tag{1.26}$$

Questo esempio ed altri che si potrebbero costruire, eventualmente con simboli bidimensionali, mostrano in generale che quando si trasmette una lunga successione di simboli la (1.22) vale non solo considerando il *numero di dimensioni* e la *durata* complessiva, ma anche il *numero di dimensioni per simbolo* e l'*intervallo tra i simboli*. Non ha invece alcuna importanza la *durata di un simbolo*, solitamente molto maggiore di T. La stessa informazione è comunque contenuta, più chiaramente, nelle (1.22), (1.23) e (1.25).

Si supponga ora che nella (1.24) non tutte le combinazioni dei livelli a_k siano consentite, ma ad esempio solo $2^{N/2}$ fra le 2^N siano lecite[11]. Si

[9] bisognerà verificare che non ci sia interferenza reciproca tra le forme d'onda trasmesse successivamente

[10] disporre pochi segnali in poche dimensioni produce soluzioni scadenti, ma con poca fatica: l'ideale per i pigri di mente

[11] non interessa in questo momento né *perché* possa convenire fare questa operazione, né *come* si possano selezionare le combinazioni da utilizzare

trasmettono $\log_2 M/N = 0.5$ bit per dimensione, e quindi la (1.22) mostra che è inevitabile espandere la banda[12] di un fattore 2. Naturalmente è possibile descrivere la situazione affermando che si trasmette un bit per simbolo, ma che c'è un *codice* (in altri termini, una regola) per cui si trasmette un bit effettivo d'informazione ogni due simboli, e dunque si trasmette solo mezzo bit per simbolo. In tal caso si distingue tra i bit *d'informazione* per simbolo e i bit *di canale* (o *di codice*).

1.5 Esempi in banda base

Nei sistemi pratici di trasmissione numerica risulta molto comodo, anche in vista delle elaborazioni richieste in ricezione, realizzare il segnale trasmesso come somma di contributi elementari semplici. Ad esempio quasi tutti i sistemi in banda base sono rappresentabili da una espressione come la (1.24), con un numero variabile di possibile ampiezze a_k e con regole più o meno complesse per la scelta di tali livelli.

Le forme d'onda $g(t-kT)$ possono risultare ortogonali anche se fortemente sovrapposte temporalmente[13]. L'esempio più noto è quello delle forme d'onda cosiddette a *radice di Nyquist*[14], con trasformata di Fourier il cui modulo quadro ha transizione simmetrica intorno alla frequenza $1/2T$. Nel caso, piuttosto comune, di transizione sinusoidale la trasformata $|G(f)|^2$ è detta *a coseno rialzato*. Un esempio è mostrato in fig. 1.3, insieme alla trasformata $G(f)$. Per semplicità si è posto $T = 1$ e sono mostrate le sole frequenze positive. L'eccesso di banda rispetto al minimo teorico, detto anche *roll-off*, è un parametro a disposizione del progettista. È facile dimostrare che le funzioni $g(t - kT)$ sono ortogonali (es. 1.8). L'espressione analitica di $g(t)$ si ricava abbastanza facilmente dalla trasformata, ed è data da

$$g(t) = \frac{\sin \pi(1-\alpha)t/T}{\pi t/T(1 - 16\alpha^2 t^2/T^2)} + \frac{4\alpha}{\pi} \frac{\cos \pi(1+\alpha t/T)}{1 - 16\alpha^2 t^2/T^2} \tag{1.27}$$

dove α è il *roll-off*. La fig. 1.4 mostra la funzione $g(t)$ per $T = 1$ e $\alpha = 0.4$ (*roll-off* del 40%, si dice di solito). Il caso particolare di $\alpha = 0$ corrisponde

[12] in cambio di qualche altro vantaggio, sperabilmente

[13] l'ortogonalità sarebbe evidentemente ottenuta con forme d'onda non sovrapponentesi ma la banda occupata sarebbe eccessiva per quasi tutti gli scopi pratici

[14] modo di dire un po' infelice, perché lascia credere che la forma d'onda sia la radice delle forme d'onda di Nyquist, mentre ciò vale per le trasformate di Fourier

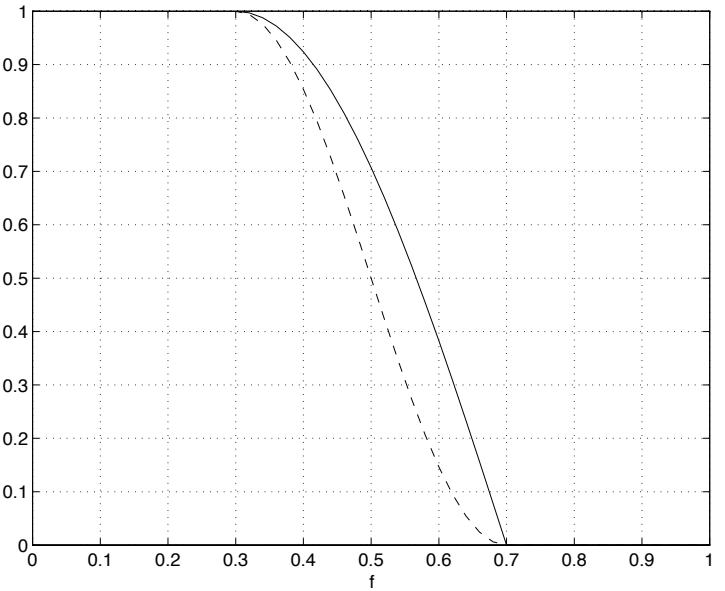

Fig. 1.3 - Trasformata di Fourier delle funzioni di Nyquist a coseno rialzato (tratteggio) e radice di Nyquist (curva continua) (*roll-off* = 40%)

alla funzione $\frac{\sin \pi t/T}{\pi t/T}$ cosiddetta *seno cardinale*, mai usata in pratica perché le sue code si estendono su un intervallo di tempo lungo in modo esasperante.

Per motivi che saranno chiari nel seguito interessa anche conoscere la forma d'onda $g(t) * g(t)$, che ha come trasformata $|G(f)|^2$. Essa è data da

$$g(t) * g(t) = \frac{\sin \pi t/T}{\pi t/T} \frac{\cos \alpha \pi t/T}{1 - 4\alpha^2 t^2/T^2} \tag{1.28}$$

ed ha zeri equispaziati ogni T secondi (escluso $t = 0$)[15]. È facile verificare, quando non sia già noto, che sia $g(t)$ sia $g(t) * g(t)$ si esauriscono tanto più rapidamente quanto più elevato è il *roll-off*.

[15] qualche anima candida ritiene che sia la forma d'onda $g(t)$ ad avere zeri equispaziati; non c'è alcun motivo perché la forma d'onda elementare *trasmessa* abbia tali zeri

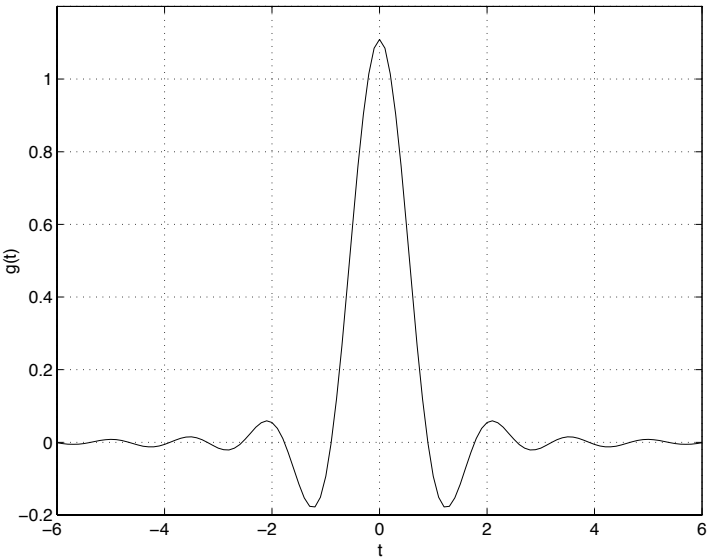

Fig. 1.4 - Andamento della funzione $g(t)$ a radice di Nyquist (*roll-off* = 40%)

1.6 Calcolo di correlazioni in banda base

Un'operazione fondamentale da eseguire in ricezione, come si vedrà, è il calcolo delle componenti del segnale ricevuto[16] $r(t)$, che consiste nel calcolo di integrali come

$$\int r(t)g(t)dt \tag{1.29}$$

dove $g(t)$ indica la generica funzione con cui correlare. Occorre quindi una circuiteria opportuna per eseguire prodotti tra funzioni del tempo e per calcolare l'area del prodotto. L'unica cosa da notare è che $g(t)$ è nota a priori mentre $r(t)$ viene osservato al momento, in tempo reale. Moltiplicatori e integratori analogici, che pure esistono, non sono tra i componenti a cui si fa ricorso a cuor leggero per cui i correlatori analogici non sono comuni in pratica.

[16] per il momento si supponga che il mezzo trasmissivo semplicemente attenui il segnale ed aggiunga del rumore

Una struttura analogica alternativa, un tempo famosissima ma che ora lascia spazio sempre più spesso alle tecniche digitali, è il cosiddetto *filtro adattato*. L'uscita di un filtro con risposta impulsiva $h(t)$ e con ingresso $r(t)$ è la forma d'onda

$$y(t) = \int r(\tau)h(t-\tau)d\tau \tag{1.30}$$

Poiché la correlazione che si vuol calcolare è un numero, e non una funzione del tempo, si può pensare di leggere l'uscita $y(t)$ in un qualche istante t_0 prefissato avendo scelto $h(t)$ in modo che risulti

$$y(t_0) = \int r(\tau)h(t_0-\tau)d\tau = \int r(\tau)g(\tau)d\tau \tag{1.31}$$

dove, per facilitare il confronto tra le due espressioni, la (1.29) è stata trascritta cambiando nome alla variabile di integrazione. Si ottiene infine, cambiando nuovamente nome alle variabili, che la risposta all'impulso richiesta è

$$h(t) = g(t_0 - t) \tag{1.32}$$

La risposta impulsiva $h(t)$ del filtro adattato deve dunque essere la $g(t)$ con asse dei tempi ribaltato, e traslata, cosa indispensabile per rendere il filtro causale e quindi (si spera) fisicamente realizzabile[17].

Se ora si volesse correlare $r(t)$ anche con $g(t-kT)$, replica traslata di $g(t)$, sarebbe assurdo mantenere inalterato t_0 e modificare la risposta impulsiva $h(t)$, che risulterebbe semplicemente traslata. Piuttosto si userà la stessa $h(t)$, cioè lo *stesso filtro*, e si campionerà nell'istante $kT + t_0$. Il grande merito del filtro adattato è quindi che con un solo filtro si ottengono successivamente le correlazioni di $r(t)$ con repliche traslate di una stessa forma d'onda. Naturalmente se le funzioni base fossero, per fare un esempio, $g_1(t)$ e $g_2(t-kT)$ occorrerebbero due diversi filtri adattati.

Passando al calcolo *numerico* di una correlazione, l'integrale sarà sostituito da una somma di prodotti di campioni[18].

[17] il problema è che non è facile progettare un filtro analogico con la risposta impulsiva desiderata; inoltre tutti i filtri analogici sono sensibili a temperatura, umidità, vibrazioni, invecchiamento, ecc.

[18] i campioni sia di $r(t)$ sia di $g(t)$ saranno rappresentati con precisione finita; nel caso del segnale ricevuto si tratta di un rumore di quantizzazione, da aggiungere al rumore all'ingresso del ricevitore; per quanto riguarda $g(t)$ si tratta di una, sia pur piccola, deformazione della forma d'onda con cui si correla, di cui si dovrà calcolare l'effetto

Conviene fare una breve digressione sul calcolo approssimato di un integrale mediante una somma

$$\int_{-\infty}^{\infty} f(t)dt \approx \tau \sum_{n=-\infty}^{\infty} f(n\tau + t_0) \tag{1.33}$$

dove i campioni, presi con passo τ, sono traslati di una generica quantità t_0. Mentre è evidente che al limite per $\tau \to 0$ le due espressioni coincidono, è fondamentale osservare che si può avere un risultato esatto anche per τ finito. A tale scopo si ricordi da un lato la cosiddetta formula di Poisson[19]

$$\sum_{n=-\infty}^{\infty} f(n\tau + t_0) = \frac{1}{\tau} \sum_{m=-\infty}^{\infty} F(\frac{m}{\tau}) \exp(j2\pi m t_0/\tau) \tag{1.34}$$

dove $F(f)$ è la trasformata di Fourier di $f(t)$, e dall'altro che

$$\int_{-\infty}^{\infty} f(t)dt = F(0) \tag{1.35}$$

Se dunque $F(m/\tau) = 0$ per ogni $m \neq 0$ la (1.33) è verificata senza errore pur essendo τ finito. A tal fine è sufficiente, ma non necessario, che la banda B_f di $f(t)$ sia minore di $1/\tau$, e quindi basta che sia[20]

$$\tau < \frac{1}{B_f} \tag{1.36}$$

Nel calcolo discreto della correlazione tra $r(t)$ e ad esempio $g(t - kT)$, cioè dell'integrale del prodotto $f(t) = r(t)g(t - kT)$, la banda B_f di $f(t)$ è pari alla somma delle bande dei due segnali. La banda B_g di $g(t - kT)$ è minore o uguale a B, banda occupata dai segnali, per cui basta porre

$$\tau < \frac{1}{B_r + B} \tag{1.37}$$

dove B_r è la banda del segnale ricevuto. Naturalmente è richiesto non solo che sia limitata la banda del prodotto $r(t)g(t - kT)$, ma che lo sia anche la durata per non dover eseguire nella (1.33) un numero infinito di prodotti e somme, operazione impensabile in una macchina digitale. Ciò viene garantito dalla durata *praticamente* limitata di $g(t-kT)$. Se T_g è tale durata, il minimo numero di moltiplicazioni (e somme) richieste è $T_g/\tau = T_g(B_r + B)$.

[19] che non è altro che lo sviluppo in serie di Fourier della ripetizione periodica di $f(t)$ con periodo τ, valutata in $t = t_0$

[20] il teorema del campionamento, evocato a sproposito, potrebbe forse suggerire $\tau < 1/2B_f$; ma qui non si tratta di *interpolare* $f(t)$, bensì solo di *calcolarne l'area*

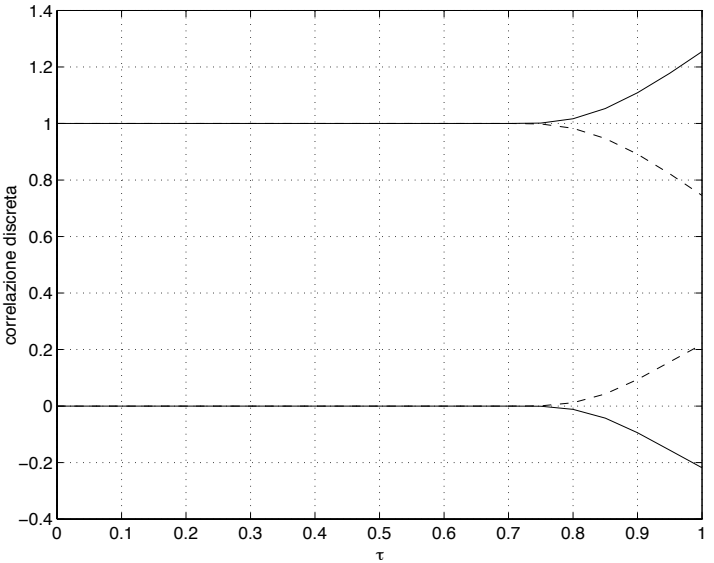

Fig. 1.5 - Risultati del calcolo di correlazioni discrete tra funzioni a radice di Nyquist (*roll-off*=40%), con $t_0 = 0$ (curva continua) e $t_0 = \tau/2$ (tratteggio)

I risultati del calcolo di semplici correlazioni numeriche sono mostrati in fig. 1.5. La funzione integranda è il prodotto della forma d'onda a radice di Nyquist di fig. 1.4, con *roll-off* del 40 % e quindi con banda 0.7, con sé stessa oppure con una replica traslata di un secondo (e quindi ortogonale). In entrambi i casi la funzione integranda ha banda $B_f = 1.4$. La figura mostra i risultati in funzione di τ, per $t_0 = 0$ (curve continue) e $t_0 = \tau/2$ (curve tratteggiate). La funzione integranda è troncata a ± 10 (intervallo sovrabbondante). Si conferma la (1.36), secondo cui il risultato è corretto fino a $\tau = 1/1.4 = 0.71$. Si noti anche che, come previsto dalla (1.34), quando per $\tau > 0.71$ interviene il termine $F(1/\tau)$ il risultato dipende da t_0.

Tornando alla correlazione tra segnale ricevuto e funzioni base, la banda di $r(t)$ sarebbe pari a B se il segnale ricevuto fosse una replica di quello trasmesso. Ma occorre tener conto del rumore che si aggiunge al segnale e che ha banda molto larga, limitata solo dai filtri analogici del ricevitore.

Se la funzione di trasferimento del filtro analogico di ricezione è ideale nella banda di $g(t)$ il risultato della correlazione è uguale a quello che si

avrebbe con banda infinita[21]. In compenso il filtro analogico limita la banda B_r di $r(t)$ e consente di realizzare il correlatore in forma numerica. Conviene tuttavia resistere alla tentazione di utilizzare un filtro analogico molto stretto per ridurre il numero di moltiplicazioni, perché è difficile tenere sotto controllo la funzione di trasferimento nella banda utile[22].

Nel calcolo delle correlazioni tra $r(t)$ e $g(t-kT)$ per i diversi valori di k non è possibile, in generale, riutilizzare risultati parziali. Ogni correlazione richiede $(B_r+B)T_g$ moltiplicazioni. Si deve eseguire una correlazione per ogni dimensione dello spazio dei segnali, e poiché $N = 2BT_0$ dimensioni equivalgono a $2B$ dimensioni al secondo si devono eseguire $2B(B_r + B)T_g$ moltiplicazioni al secondo.

Se infine si tentasse di realizzare un filtro adattato in forma numerica con un filtro FIR si scoprirebbe facilmente che non si otterrebbe altro che il correlatore numerico appena descritto (es. 1.10).

Filtri numerici IIR possono essere presi in considerazione solo se è possibile controllarne la fase. Tipicamente occorrono funzioni di trasferimento con fase nulla, per cui si può solo pensare a due filtraggi IIR uguali eseguiti successivamente una volta con l'asse dei tempi *normale* ed una volta *ribaltato*.

1.7 Richiami sui segnali passa banda

Un segnale $s(t)$ passa banda, con trasformata di Fourier non nulla solo in un intorno, non necessariamente simmetrico, della frequenza f_0 può essere rappresentato come

$$\begin{aligned} s(t) &= \mathrm{Re}\{z(t)\exp(j2\pi f_0 t)\} = \frac{1}{2}z(t)\exp(j2\pi f_0 t) + \\ &+\frac{1}{2}z^*(t)\exp(-j2\pi f_0 t) = |z(t)|\cos(2\pi f_0 t + \arg(z(t))) \end{aligned} \tag{1.38}$$

La funzione, generalmente complessa, $z(t)$ è detta *equivalente passa basso*, o anche *inviluppo complesso*, ed ha trasformata di Fourier data da

$$Z(f) = 2S(f+f_0)U(f+f_0) \tag{1.39}$$

[21] $\int r(t)g(t)dt = \int R(f)G^*(f)df = \int R(f)H(f)G^*(f)df$

[22] interessa non solo la caratteristica d'ampiezza, ma anche quella di fase; spesso i filtri con caratteristica d'ampiezza ripida maltrattano la fase

dove $U(\cdot)$ è la funzione scalino. In pratica $Z(f)$ è la sola parte a frequenze positive della trasformata del segnale passa banda, traslata intorno alla frequenza zero.

Un segnale passa banda $A(t)\cos(2\pi f_0 t + \varphi(t))$ con $A(t)$ e $\varphi(t)$ lentamente variabili rispetto al periodo della portante[23] ha equivalente passa basso $A(t)\exp(j\varphi(t))$. È anche facile verificare che un tale segnale passa banda ha area nulla (es. 1.11).

Rappresentando l'inviluppo complesso, anziché in forma polare, in quella cartesiana $z(t) = x(t)+jy(t)$ si ottiene facilmente dalla (1.38) la scomposizione nelle due componenti in fase e quadratura

$$s(t) = x(t)\cos 2\pi f_0 t - y(t)\sin 2\pi f_0 t \tag{1.40}$$

È poi quasi immediato verificare che il filtraggio di un segnale passa banda con funzione di trasferimento $H(f)$ equivale a filtrare l'equivalente passa basso dell'ingresso con la funzione di trasferimento semplicemente traslata $H(f+f_0)$.

Analogamente a quanto accade moltiplicando due sinusoidi, il prodotto di due segnali passa banda $s_1(t)$ ed $s_2(t)$ con equivalenti passa basso $z_1(t)$ e $z_2(t)$ è esprimibile come un termine passa basso ed uno passa banda a frequenza doppia[24]

$$s_1(t)s_2(t) = \frac{1}{2}\mathrm{Re}\{z_1(t)z_2^*(t)\} + \frac{1}{2}\mathrm{Re}\{z_1(t)z_2(t)\exp(j4\pi f_0 t)\} \tag{1.41}$$

Di solito interessa uno solo dei due termini, che sono facilmente separabili con filtri passa basso o passa banda.

Se le due forme d'onda passa banda che vengono moltiplicate sono $s(t)$ e $2\cos 2\pi f_0 t$, con equivalenti passa basso rispettivamente $z(t)$ e 2, la (1.41) mostra che un filtro passa basso a valle del prodotto dà la componente in fase $\mathrm{Re}\{z(t)\} = x(t)$; analogamente si può verificare che moltiplicando invece per $-2\sin 2\pi f_0 t$ si ottiene la componente in quadratura $y(t)$. Quindi parte reale e immaginaria dell'equivalente passa basso si possono ottenere mediante demodulatori d'ampiezza in fase e quadratura.

Analogamente la demodulazione con $2\cos(2\pi f_0 t + \psi)$ e $-2\sin(2\pi f_0 t + \psi)$ equivale al calcolo delle componenti in fase quadratura di $z(t)\exp(-j\psi)$, e quindi ad una rotazione di fase della portante. È importante osservare che tale

[23] più precisamente, occorre che la banda di $A(t)$ e di $\exp(j\varphi(t))$ sia minore di f_0, cosa in pratica quasi sempre verificata

[24] basta porre $s_i(t) = \frac{1}{2}z_i(t)\exp(j2\pi f_0 t) + \frac{1}{2}z_i^*(t)\exp(-j2\pi f_0 t)$ $(i = 1, 2)$, eseguire il prodotto e riunire i termini complessi coniugati

rotazione può essere ottenuta sia demodulando con fase ψ, sia demodulando con fase nulla e poi moltiplicando in banda base l'equivalente passa basso $z(t)$ per $\exp(-j\psi)$.

Se si vuol calcolare la correlazione tra le forme d'onda passa banda $s_1(t)$ ed $s_2(t)$ il termine a frequenza doppia ha integrale nullo, e quindi si ha

$$\int s_1(t)s_2(t)dt = \frac{1}{2}\mathrm{Re}\{\int z_1(t)z_2^*(t)dt\} \tag{1.42}$$

Nel caso particolare di segnali coincidenti l'energia vale

$$\int s^2(t)dt = \frac{1}{2}\int |z(t)|^2 dt \tag{1.43}$$

Ad esempio si ottengono le seguenti formule, assai utili:

$$\begin{aligned}&\int A(t)\cos(2\pi f_0 t + \varphi_1(t))\cos(2\pi f_0 t + \varphi_2(t))dt = \\ &= \frac{1}{2}\int A(t)\cos(\varphi_2(t) - \varphi_1(t))dt\end{aligned} \tag{1.44}$$

$$\int A^2(t)\cos^2(2\pi f_0 t + \varphi(t))dt = \frac{1}{2}\int A^2(t)dt \tag{1.45}$$

La correlazione tra un generico segnale ricevuto passa banda $r(t)$ e una funzione base passa banda sarà eseguita in banda base mediante la (1.42), calcolando dapprima le componenti in fase e quadratura di $r(t)$ con demodulatori in fase e quadratura e poi correlando con l'equivalente passa basso, che è noto a priori, della funzione base.

In teoria si potrebbe eseguire il calcolo della correlazione anche in banda passante, ma in pratica si incontrano varie difficoltà. Se l'operazione è fatta con un correlatore analogico il moltiplicatore e l'integratore devono avere banda larga per non distorcere, anche solo in fase, le componenti alle frequenze intorno a f_0. Un filtro adattato in banda passante presenta problemi ancora maggiori, a causa della grande precisione richiesta per l'istante di lettura (es. 1.12). Infine in una realizzazione numerica la frequenza di campionamento e il numero di moltiplicazioni sarebbero inutilmente elevati.

1.8 Funzioni base passa banda

Se i segnali $s_i(t)$ sono passa banda, tali saranno anche le funzioni base. Una generica tra queste può essere rappresentata con l'equivalente passa basso:

$$\Phi_k(t) = \mathrm{Re}\{z_k(t)\exp(j2\pi f_0 t)\} = A_k(t)\cos(2\pi f_0 t + \varphi_k(t)) \tag{1.46}$$

È immediato verificare dalla (1.38) che la funzione passa banda con equivalente passa basso $jz_k(t)$ è data da

$$\begin{aligned}\Phi_{k'}(t) &= \mathrm{Re}\{jz_k(t)\exp(j2\pi f_0 t)\} = -\mathrm{Im}\{z_k(t)\exp(j2\pi f_0 t)\} = \\ &= -A_k(t)\sin(2\pi f_0 t + \varphi_k(t))\end{aligned} \tag{1.47}$$

e dalla (1.42) che è ortogonale a $\Phi_k(t)$. Essa deve quindi essere utilizzata, se non si vuole sprecare banda. Nella gran parte dei casi $\varphi_k(t) = 0$, cioè $z_k(t)$ è reale. Ciò implica che lo spettro in banda passante sia simmetrico.

Siano s_{ik} e $s_{ik'}$ le componenti di $s_i(t)$ lungo gli assi k e k'. La somma dei due corrispondenti contributi è

$$s_{ik}\Phi_k(t) + s_{ik'}\Phi_{k'}(t) = \mathrm{Re}\{(s_{ik} + js_{ik'})z_k(t)\exp(j2\pi f_0 t)\} \tag{1.48}$$

e quindi, se lo si desidera, è lecito ragionare come se vi fosse la sola funzione base $\Phi_k(t)$ e le componenti s_{ik} del vettore $\mathbf{s}_i$ fossero complesse. In pratica chi usa tale notazione, più sintetica, indica con un unico numero complesso s_{ik} la coppia di componenti $s_{ik} + js_{ik'}$.

Ad esempio nella modulazione a quattro livelli in fase e quadratura (16QAM: *Quadrature Amplitude Modulation* con 16 punti)

$$s_i(t) = a\ g(t)\cos 2\pi f_0 t - b\ g(t)\sin 2\pi f_0 t \tag{1.49}$$

dove $a = \pm 1, \pm 3$ e $b = \pm 1, \pm 3$ la rappresentazione geometrica dei sedici possibili segnali è quella mostrata in fig. 1.6, ed è comunissimo rappresentare i dati con il numero complesso[25] $d = a + jb$. Val la pena osservare che la *stessa* geometria dei segnali si ottiene nel caso passa basso

$$s_i(t) = a\ g(t) + b\ g(t - T) \tag{1.50}$$

con a e b a quattro livelli, e con $g(t)$ e $g(t - T)$ ortogonali. In questo caso pensare alla coppia (a, b) come un numero complesso, pur lecito, è inutile

[25] del resto è normale rappresentare vettori bidimensionali con numeri complessi, e viceversa

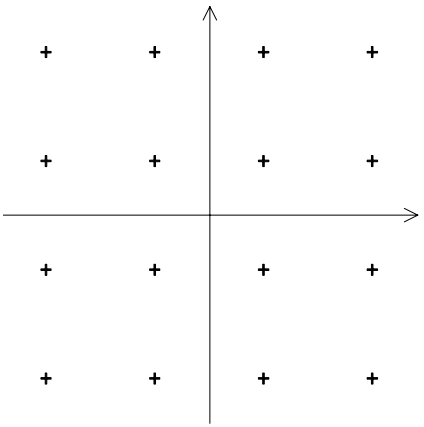

Fig. 1.6 - Costellazione bidimensionale 16QAM

e può addirittura confondere le idee. Si osservi anche, da questo banale esempio, che la *stessa geometria* dei segnali può corrispondere a forme d'onda completamente diverse. Naturalmente ciò avviene perché sono diverse le funzioni base. Per altri esempi di ugual geometria con segnali diversi si vedano gli es. 1.15-1.19.

Spesso si scelgono gli equivalenti passa basso $z_k(t)$ delle funzioni base $\Phi_k(t)$ come repliche traslate di una opportuna $z(t)$. Dalla (1.42) si ottiene facilmente che tali repliche devono essere ortogonali; più esplicitamente, trattandosi in generale di funzioni complesse, si deve avere

$$\int z(t-kT)z^*(t-nT)dt = 0 \qquad (n \neq k) \tag{1.51}$$

e ciò basta perché anche $\Phi_k(t)$ e $\Phi_{n'}(t)$ siano ortogonali.

Dunque la forma più comune di segnali passa banda utilizzati per la trasmissione numerica, assumendo come solitamente avviene $z(t) = g(t)$ reale, è

$$\begin{aligned} s_i(t) &= \mathrm{Re}\{\sum d_k g(t-kT)\exp(j2\pi f_0 t)\} = \\ &= \sum a_k g(t-kT)\cos 2\pi f_0 t - \sum b_k g(t-kT)\sin 2\pi f_0 t \end{aligned} \tag{1.52}$$

e i *livelli* trasmessi possono essere considerati complessi: $d_k = a_k + jb_k$. Si osservi anche che passando dalle funzioni base $g(t-kT)$ in banda base alle $g(t-kT)\cos 2\pi f_0 t$ e $g(t-kT)\sin 2\pi f_0 t$ in banda passante raddoppiano sia la banda occupata sia il numero di funzioni base disponibili, a conferma che resta valida la relazione $N = 2BT_0$.

In ricezione, volendo calcolare la correlazione tra $r(t)$ e $\Phi_k(t)$ mediante i rispettivi equivalenti passa basso $z(t)$ e $z_k(t)$ si avrà

$$\int r(t)\Phi_k(t)dt = \frac{1}{2}\mathrm{Re}\{\int z(t)z_k^*(t)dt\} \tag{1.53}$$

Analogamente la correlazione con la funzione base $\Phi_{k'}(t)$ è data da

$$\int r(t)\Phi_{k'}(t)dt = \frac{1}{2}\mathrm{Re}\{\int z(t)(jz_k(t))^*dt\} = \frac{1}{2}\mathrm{Im}\{\int z(t)z_k^*(t)dt\} \tag{1.54}$$

Le due componenti della forma d'onda ricevuta $r(t)$ sono quindi espresse sinteticamente da parte reale e parte immaginaria di un r_k complesso, dove

$$r_k = \frac{1}{2}\int z(t)z_k^*(t)dt \tag{1.55}$$

Ovviamente l'integrale viene discretizzato come già discusso per il caso passa basso, e le operazioni sono effettuate tra numeri reali. Nel caso particolare di $z_k(t)$ reale, cioè di spettro simmetrico, ogni prodotto complesso equivale a due moltiplicazioni reali, una per la componente in fase ed una per quella in quadratura, con un vantaggio rispetto al caso generale in cui ne occorrono quattro.

Talvolta vengono utilizzati i sistemi di modulazione in fase e quadratura con *offset*

$$s_i(t) = \sum a_k g(t-kT)\cos 2\pi f_0 t - \sum b_k g(t-kT-T/2)\sin 2\pi f_0 t \tag{1.56}$$

per i quali non è invece raccomandabile la rappresentazione con dati d_k complessi.

L'ortogonalità tra funzioni base può essere ottenuta anche in altri modi, ad esempio utilizzando forme d'onda con frequenze centrali f_0 diverse; è questo il modo tipico in cui più utenti o comunque più segnali condividono un canale radio.

1.9 Rappresentazione geometrica del rumore

Si vuol dare una rappresentazione geometrica non solo dei possibili segnali trasmessi $s_i(t)$, ma anche del rumore che ad essi si somma.

Il rumore ha diverse peculiarità da sottolineare. Innanzitutto le sue possibili realizzazioni sono infinite, contrariamente a quanto accade per gli M segnali, con M anche molto grande ma finito. Inoltre il rumore può essere noto solo in senso statistico, cioè deve essere considerato come un processo casuale.

Si supponga valida una espansione del rumore $n(t)$ come combinazione di funzioni base ortonormali opportune

$$n(t) = \sum n_k \Phi_k(t) \tag{1.57}$$

in un intervallo di tempo *prefissato* di durata *finita* T_0 in cui sono contenuti anche i segnali $s_i(t)$. Invocando l'ortogonalità delle funzioni base si ottiene che i coefficienti n_k sono dati da

$$n_k = \int n(t)\Phi_k(t)dt \tag{1.58}$$

da cui è subito evidente che n_k varia dall'una all'altra realizzazione di $n(t)$, e quindi n_k è una variabile casuale. I coefficienti n_k sono da considerare un insieme di variabili casuali, per la cui descrizione statistica occorre la densità di probabilità (*ddp*) *congiunta*. Infine poiché le possibili realizzazioni del processo $n(t)$ sono infinite non è da escludere che occorra un numero infinito di funzioni base.

Si possono facilmente calcolare valori medi e covarianze delle variabili casuali n_k. Si ha, scambiando integrale e valor medio e supponendo che il processo $n(t)$ abbia valor medio nullo,

$$E[n_k] = E[\int n(t)\Phi_k(t)dt] = \int E[n(t)]\Phi_k(t)dt = 0 \tag{1.59}$$

Le autocovarianze (o varianze, se $k = j$) sono date da

$$\begin{aligned} \sigma_{kj} &= E[n_k n_j] = E[\int n(t_1)\Phi_k(t_1)dt_1 \int n(t_2)\Phi_j(t_2)dt_2] = \\ &= \int\int R_n(t_2 - t_1)\Phi_k(t_1)\Phi_j(t_2)dt_1 dt_2 \end{aligned} \tag{1.60}$$

avendo nuovamente scambiato integrale e valor medio, e avendo indicato con $R_n(\tau)$ l'autocorrelazione del processo, che per semplicità si suppone stazionario.

L'unico caso in cui valori medi e covarianze forniscono una caratterizzazione completa della statistica delle variabili casuali n_k è quello, fortunatamente molto comune, in cui il processo $n(t)$ è gaussiano e quindi le variabili casuali n_k sono *congiuntamente* gaussiane. La *ddp* congiunta ha una espressione addirittura banale se le variabili casuali sono *incorrelate*, cioè se $\sigma_{kj} = 0$ per $j \neq k$. Infatti in tal caso esse risultano anche *indipendenti*, e la *ddp* congiunta è il prodotto delle *ddp* marginali. Occorre dunque tentare di ottenere l'incorrelazione, mediante una opportuna scelta delle funzioni base $\Phi_k(t)$.

Esaminando la (1.60) è facile constatare che il risultato è ottenuto se si soddisfa, per ogni k e ogni valore di t_2, la condizione sufficiente

$$\int R_n(t_2 - t_1)\Phi_k(t_1)dt_1 = \sigma_k^2 \Phi_k(t_2) \tag{1.61}$$

Non è difficile mostrare che la condizione è anche necessaria. Infatti, assumendo valida la (1.57), e quindi la (1.58), si ha con i soliti scambi tra integrali, somme e valor medio, e invocando l'incorrelazione delle variabili casuali n_k

$$\int E[n(t_1)n(t_2)]\Phi_k(t_1)dt_1 = E[n_k n(t_2)] = E[n_k \sum n_j \Phi_j(t_2)] = \\ = \sigma_k^2 \Phi_k(t_2) \tag{1.62}$$

Resta da esaminare se l'equazione integrale (1.61) abbia soluzioni, e inoltre se l'espansione (1.57) valga davvero. Non si dimentichi che la si è assunta valida a priori. Le conoscenze necessarie per rispondere a tali questioni vanno al di là delle nozioni elementari di analisi matematica. Ci si limiterà a richiamare i risultati:

- esistono infinite soluzioni della (1.61), ma non per ogni valore del parametro σ^2; soluzioni corrispondenti a valori diversi di σ_k^2 sono ortogonali, e possono essere normalizzate; i valori di σ^2 per cui esistono soluzioni sono detti *autovalori* e le soluzioni sono dette *autofunzioni*

- poiché l'equazione integrale è lineare, combinazioni lineari di soluzioni sono soluzioni; soluzioni diverse corrispondenti allo stesso valore di

σ^2 possono essere ortogonalizzate e normalizzate, ad esempio con la procedura di Gram-Schmidt

- si ha quindi un insieme di infinite soluzioni ortonormali $\Phi_k(t)$
- la funzione di autocorrelazione del processo è espandibile nella serie

$$R_n(t_2 - t_1) = \sum_{k=1}^{\infty} \sigma_k^2 \Phi_k(t_1)\Phi_k(t_2) \tag{1.63}$$

- le infinite funzioni $\Phi_k(t)$ sono una base *completa*, cioè in grado di rappresentare qualunque forma d'onda ad energia finita nell'intervallo di tempo T_0 considerato, e quindi anche i segnali $s_i(t)$
- le infinite funzioni $\Phi_k(t)$ sono in grado di rappresentare anche il rumore, nel senso che[26]

$$E[\big(n(t) - \sum_{k=1}^{\infty} n_k \Phi_k(t)\big)^2] = 0 \tag{1.64}$$

Dunque esiste un insieme di funzioni base che ha quasi tutte le caratteristiche desiderate: può rappresentare i segnali $s_i(t)$ e il rumore $n(t)$; inoltre le componenti n_k del rumore lungo i vari assi sono incorrelate, e quindi indipendenti nel caso gaussiano. L'unica seccatura viene dal fatto che, per colpa del rumore, occorre un numero infinito di funzioni base mentre ne basterebbe un numero finito per i segnali. Si può almeno sperare che le N funzioni base $\Phi_k(t)$ che si sceglierebbero se si dovessero rappresentare solo i segnali facciano parte delle soluzioni dell'equazione integrale (1.61).

In generale ciò non accade. Ma basta che il rumore $n(t)$ abbia densità spettrale di potenza *costante* nella banda dei segnali[27] perché la risposta sia affermativa, come si vede esaminando le trasformate di Fourier dei due membri della (1.61)[28]. Ed anzi si vede che la varianza di n_k è numericamente uguale

[26] dire che la varianza della differenza è nulla è, per tutti i fini pratici, equivalente a dire che $n(t) = \sum n_k \Phi_k(t)$

[27] si usa dire che il rumore è bianco nella banda dei segnali

[28] le N funzioni base richieste per rappresentare gli M segnali $s_i(t)$ occupano lo stesso intervallo di tempo e la stessa banda dei segnali; la (1.61) è una convoluzione, se si può assumere che durata e banda siano limitate

alla densità spettrale di potenza (bilatera) del rumore[29] $n(t)$, che nel seguito verrà indicata con $N_0/2$:

$$\sigma_k^2 = \frac{N_0}{2} \qquad (k = 1, \ldots, N) \tag{1.65}$$

Si noti che la varianza di *ciascuna* componente ortogonale del rumore, *non* la varianza del processo $\sigma^2_{n(t)}$, vale $N_0/2$ (es. 1.23). La varianza del processo dipende dalla banda, oltre che dalla densità spettrale di potenza, e quindi anche *dimensionalmente*[30] non può essere pari a $N_0/2$.

In conclusione se il rumore è bianco nella banda dei segnali le prime N funzioni base possono essere scelte come le più convenienti per rappresentare i segnali $s_i(t)$, e le corrispondenti componenti del rumore gaussiano hanno varianza $N_0/2$. Le infinite restanti funzioni base si ottengono risolvendo effettivamente l'equazione integrale (1.61), e le varianze delle relative componenti del rumore, per $k > N$, sono i corrispondenti autovalori. Si vedrà tuttavia nel seguito che in questo caso, davvero fortunato, non occorre né determinare esplicitamente le infinite funzioni base per $k > N$ né calcolare le relative componenti del rumore.

Merita infine di essere sottolineato il fatto che non fa alcuna differenza che i segnali, e quindi le funzioni base, siano di tipo passa basso oppure passa banda. Ad ogni asse della rappresentazione geometrica è comunque associata una componente del rumore con varianza $N_0/2$. Come per le componenti dei segnali, nel caso passa banda si possono rappresentare *coppie* di componenti del rumore con numeri complessi, indicando cioè con n_k la coppia di componenti $n_k + jn_{k'}$.

Il caso di rumore gaussiano non bianco nella banda dei segnali verrà considerato successivamente.

---

[29] se si è disposti a sforzare un po' la matematica, per un rumore bianco su tutto l'asse delle frequenze si può porre $R_n(\tau) = \frac{N_0}{2}\delta(\tau)$ e l'equazione $\int \frac{N_0}{2}\delta(t_2 - t_1)\Phi_k(t_1)dt_1 = \frac{N_0}{2}\Phi_k(t_2)$ è automaticamente verificata per *qualsiasi* funzione $\Phi_k(t)$

[30] quanto alle dimensioni, è quasi sempre adottata la convenzione di riferirsi alle potenze disponibili, ovvero alle tensioni su un carico adattato di impedenza unitaria; quindi le densità spettrali sono espresse in W/Hz, e ad esempio per il rumore termico si ha $N_0/2 = \frac{1}{2}kT$, dove k è la costante di Boltzmann e T è la temperatura assoluta

1.10 Equivalente passa basso di processi passa banda

Come per i segnali $s_i(t)$ e le funzioni base $\Phi_k(t)$, talvolta si ricorre anche per il rumore ad una rappresentazione mediante un equivalente passa basso complesso.

Come appena visto, ciò non è affatto necessario per rappresentare geometricamente un rumore gaussiano bianco nella banda dei segnali. Non lo è neppure, come si vedrà, nel caso di rumore gaussiano non bianco, purché si usi il ricevitore ottimo; può invece risultare utile quando si utilizzino ricevitori non ottimi.

È opportuno qualche rapido richiamo sui processi complessi. Un processo casuale $z(t) = u(t) + jv(t)$ è completamente descritto dalle *ddp* congiunte dei due processi reali $u(t)$ e $v(t)$. Nel caso gaussiano ci si può limitare ai momenti del primo e secondo ordine. Supponendo nulli i valori medi e stazionario il processo, bastano le autocorrelazioni $R_u(\tau)$ e $R_v(\tau)$, oltre alla *correlazione mutua*

$$R_{uv}(\tau) = E[u(t+\tau)v(t)] = R_{vu}(-\tau) \tag{1.66}$$

Una descrizione più sintetica è fornita dalla funzione di autocorrelazione di $z(t)$ definita come

$$R_z(\tau) = E[z(t+\tau)z^*(t)] = R_u(\tau) + R_v(\tau) + jR_{vu}(\tau) - jR_{uv}(\tau) \tag{1.67}$$

In generale $R_z(\tau)$ non determina univocamente i contributi della parte reale $u(t)$ e dell'immaginaria $v(t)$[31]. Il caso di gran lunga più interessante, ed anche il più frequente, è quello in cui le autocorrelazioni di $u(t)$ e $v(t)$ e la correlazione mutua non si modificano moltiplicando il processo per un generico $\exp(j\varphi)$. Sinteticamente la condizione può essere espressa come

$$E[z(t+\tau)z(t)] = 0 \tag{1.68}$$

Infatti svolgendo il prodotto in modo analogo alla (1.67) si ottiene immediatamente che $R_u(\tau) = R_v(\tau)$ e $R_{vu}(\tau) = -R_{uv}(\tau)$ e quindi

$$R_u(\tau) = R_v(\tau) = \frac{1}{2}\mathrm{Re}\{R_z(\tau)\} \tag{1.69}$$

[31] si pensi al caso particolare di processo $z(t)$ reale, per cui $R_v(\tau) = 0$; invece il processo immaginario $jz(t)$ ha $R_u(\tau) = 0$. Si noti che $z(t)$ e $jz(t)$ hanno la *stessa* funzione di autocorrelazione $R_z(\tau)$

$$R_{vu}(\tau) = -R_{uv}(\tau) = \frac{1}{2}\mathrm{Im}\{R_z(\tau)\} = R_{uv}(-\tau) = -R_{vu}(-\tau) \tag{1.70}$$

Il processo $y(t)$ ottenuto filtrando il processo complesso $z(t)$ con un sistema lineare con risposta impulsiva (reale o complessa) $h(t)$ ha autocorrelazione

$$R_y(\tau) = R_z(\tau) * h(\tau) * h^*(-\tau) \tag{1.71}$$

(relazione del tutto analoga a quella valida per processi reali, e dimostrabile seguendo le stesse linee). Definendo, al solito, le densità spettrali di potenza come trasformate di Fourier delle autocorrelazioni si ottiene anche nel caso di processi complessi l'usuale relazione tra le densità spettrali

$$S_y(f) = S_z(f)|H(f)|^2 \tag{1.72}$$

Non è poi difficile mostrare che le densità spettrali sono funzioni non negative; non hanno però alcun dovere di essere simmetriche, contrariamente al caso di processi reali.

Si consideri ora un processo $x(t)$ reale, a valor medio nullo, e con densità spettrale $S_x(f)$ non nulla solo in un intorno di $\pm f_0$. Come nel caso di forme d'onda deterministiche si può considerare il processo $y(t)$ ottenuto eliminando le componenti di $x(t)$ a frequenze negative. Questa operazione non è solo una *sottrazione*, che potrebbe far temere una perdita irrecuperabile, ma anche una *somma*. Infatti la funzione di trasferimento $H(f) = U(f)$, dove $U(\cdot)$ è la funzione scalino, è esprimibile come

$$H(f) = 1 + \mathrm{sgn}(f) \tag{1.73}$$

a cui corrisponde la risposta impulsiva

$$h(t) = \delta(t) + \frac{j}{\pi t} \tag{1.74}$$

La cancellazione delle frequenze negative *aggiunge* quindi al processo reale $x(t)$ una componente immaginaria $j\hat{x}(t) = x(t) * (j/\pi t)$, che si può sempre rimuovere, poiché $x(t) = \mathrm{Re}\{y(t)\}$. Se poi si definisce $z(t) = y(t)\exp(-j2\pi f_0 t)$ si ha la rappresentazione

$$x(t) = \mathrm{Re}\{z(t)\exp(j2\pi f_0 t)\} \tag{1.75}$$

o anche, se si pone $z(t) = u(t) + jv(t)$,

$$x(t) = u(t)\cos 2\pi f_0 t - v(t)\sin 2\pi f_0 t \tag{1.76}$$

in perfetta analogia con il caso deterministico. Resta solo da mostrare che $z(t)$ è un processo stazionario passa basso, e determinarne l'autocorrelazione. In modo del tutto analogo alla (1.71) si può mostrare che

$$E[y(t+\tau)y(t)] = R_x(\tau) * h(\tau) * h(-\tau) \tag{1.77}$$

La corrispondente trasformata di Fourier, data da $S_x(f)H(f)H^*(-f)$ è evidentemente nulla. Quindi si ha $E[y(t+\tau)y(t)] = 0$ e $E[z(t+\tau)z(t)] = 0$. Si ha poi

$$\begin{aligned} E[z(t+\tau)z^*(t)] &= E[y(t+\tau)\exp(-j2\pi f_0(t+\tau)y^*(t)\exp(j2\pi f_0 t)] = \\ &= \exp(-j2\pi f_0\tau)R_y(\tau) \end{aligned} \tag{1.78}$$

Infine, utilizzando anche la (1.72), la densità spettrale di potenza è data da

$$S_z(f) = S_y(f+f_0) = 4S_x(f+f_0)U(f+f_0) \tag{1.79}$$

Quest'ultima mostra che $z(t)$ è un processo passa basso, con densità spettrale proporzionale a quella di $x(t)$ traslata intorno alla frequenza zero. Infine

$$R_x(\tau) = 2\mathrm{Re}\{R_z(\tau)\exp(j2\pi f_0 t)\} \tag{1.80}$$

Se lo spettro $S_x(f)$ è simmetrico intorno alla frequenza f_0 anche $S_z(f)$ è simmetrico e quindi $R_z(\tau)$ è reale. La correlazione mutua tra le componenti in fase e quadratura $u(t+\tau)$ e $v(t)$ è dunque nulla per ogni valore di τ. In ogni caso è nulla per $\tau = 0$, perché $R_{vu}(\tau)$ è una funzione dispari di τ.

Si può infine osservare che le formule sono esteticamente più gradevoli, e ancora più simili a quelle del caso deterministico, se si modifica la definizione di autocorrelazione di un processo complesso in

$$R_z(\tau) = \frac{1}{2}E[z(t+\tau)z^*(t)] \tag{1.81}$$

Si ottiene infatti

$$R_z(\tau) = R_u(\tau) + jR_{vu}(\tau) \tag{1.82}$$

$$S_z(f) = 2S_x(f+f_0)U(f+f_0) \tag{1.83}$$

$$R_x(\tau) = \mathrm{Re}\{R_z(\tau)\exp(j2\pi f_0 t)\} \tag{1.84}$$

Questa convenzione è molto diffusa. Analogamente per una variabile casuale complessa $z = u + jv$, con u e v incorrelati e con varianza σ^2, c'è chi definisce la varianza di z come $E[|z|^2] = 2\sigma^2$ ma non pochi preferiscono $\frac{1}{2}E[|z|^2] = \sigma^2$.

1.11 Esercizi

1.1 - Si mostri che il prodotto scalare tra funzioni a energia finita soddisfa le seguenti proprietà, che sono richieste al prodotto scalare per indurre uno spazio vettoriale:

$\mathbf{x} \cdot \mathbf{y} = \mathbf{y} \cdot \mathbf{x}$ (proprietà commutativa)
$(\mathbf{x}_1 + \mathbf{x}_2) \cdot \mathbf{y} = \mathbf{x}_1 \cdot \mathbf{y} + \mathbf{x}_2 \cdot \mathbf{y}$ (proprietà distributiva)
$(a\mathbf{x}) \cdot \mathbf{y} = a\mathbf{x} \cdot \mathbf{y}$ (per ogni a reale)
$\mathbf{x} \cdot \mathbf{x} \geq 0$ ($\mathbf{x} \cdot \mathbf{x} = 0$ solo se $\mathbf{x} = \mathbf{0}$)

Per quanto riguarda l'ultimo punto si mostri che occorre escludere funzioni patologiche, peraltro di nessuna rilevanza pratica, non nulle in un insieme di misura nulla secondo Lebesgue.

1.2 - Si mostri che il prodotto scalare tra funzioni soddisfa tutte le proprietà richieste anche con la definizione

$$\mathbf{x} \cdot \mathbf{y} = \int w(t)x(t)y(t)dt \tag{1.85}$$

purché la funzione peso $w(t)$ sia positiva. Perché occorre imporre tale condizione?

1.3 - Si mostri mediante il teorema di Parseval che le forme d'onda

$$g(t - kT) = \frac{\sin \pi(t - kT)/T}{\pi(t - kT)/T}$$

sono ortogonali ed hanno energia T.

1.4 - Un numero infinito di funzioni base non garantisce di poter rappresentare qualunque forma d'onda. In tal caso si dice che l'insieme non è *completo*[32]. Si mostri che $\hat{s}(t) = \sum s_k \Phi_k(t)$ ha la distanza minima possibile da $s(t)$ se i coefficienti s_k sono dati dalla (1.5). Si mostri anche che l'errore $s(t) - \hat{s}(t)$ è ortogonale a $\Phi_k(t)$ per tutti i k, e quindi ortogonale anche a $\hat{s}(t)$. Infine si valuti l'energia dell'errore.

1.5 - Si considerino, nell'intervallo di tempo $(0, T_0)$, le funzioni $\cos 2\pi kt/T_0$ ($k = 0, \ldots, N/2$) e $\sin 2\pi kt/T_0$ ($k = 1, \ldots, N/2$), e si mostri che sono

[32] del resto per perdere la completezza basta eliminare da un insieme completo, come gli esponenziali della serie di Fourier, una o più funzioni base

ortogonali. Se $N \gg 1$ si mostri che si può ritenere la banda *complessiva* occupata dalle funzioni pari a $B \approx N/2T_0$ e quindi $N \approx 2BT_0$.

1.6 - Si consideri nell'intervallo di tempo $(0, T_0)$ l'insieme di funzioni $\cos \pi kt/T_0$ $(k = 1, \ldots, N)$. Si mostri che le funzioni sono ortogonali. Se $N \gg 1$ si mostri che si può ritenere la banda *complessiva* occupata dalle funzioni pari a $B \approx N/2T_0$ e quindi $N \approx 2BT_0$.

1.7 - Si ripetano i due esercizi precedenti con sinusoidi in banda passante $(k = N_1, \ldots, N_2)$.

1.8 - Si mostri mediante il teorema di Parseval che le repliche traslate con passo T delle forme d'onda a radice di Nyquist sono ortogonali. *Commento*: è fondamentale la transizione *dispari* intorno alla frequenza $1/2T$ di $|G(f)|^2$; non è invece necessario che la transizione sia sinusoidale.

1.9 - Si voglia realizzare un correlatore numerico in banda base, e si limiti la banda del segnale ricevuto con un filtro analogico con funzione di trasferimento $H(f)$ nota, ma non ideale nella banda dei segnali. Si mostri che è sufficiente modificare i campioni della forma d'onda $g(t)$ con cui si correla $r(t)$. Come si devono calcolare i campioni modificati?

1.10 - Si supponga di voler calcolare la correlazione tra il segnale ricevuto $r(t)$ e una funzione base $g(t - kT)$ mediante un filtro adattato, campionato all'istante $kT + t_0$ opportuno, realizzato in forma numerica (FIR). Si mostri che le operazioni da eseguire coincidono con quelle richieste dalla correlazione numerica. Si spieghi perché non occorre campionare con passo $\tau < 1/2B_r$ come sembrerebbe richiedere il teorema del campionamento, ma è invece sufficiente che sia $\tau < 1/(B_r + B)$.

1.11 - Si mostri che un segnale $s(t) = A(t)\cos(2\pi f_0 t + \varphi(t))$ che abbia come equivalente passa basso $A(t)\exp(j\varphi(t))$ ha area nulla. *Suggerimento*: $\int s(t)dt = S(0)$. Si supponga poi $A(t)$ rettangolare con durata T_0 e $\varphi(t) = 0$. Si spieghi perché il risultato non è applicabile *esattamente*. *Commento*: comunque se $f_0 \gg 1/T_0$ l'errore è trascurabile.

1.12 - Si realizzi un filtro passa banda adattato alla forma d'onda $g(t)\cos 2\pi f_0 t$,

con istante di campionamento nominale t_0. Si supponga poi che, a causa di inevitabili imperfezioni, l'istante di campionamento sia $t_0 + \varepsilon$. Si mostri che l'errore, pur con ε piccolo, può essere tale da rendere il filtro adattato inutilizzabile. *Suggerimento*: si ragioni con gli equivalenti passa basso.

1.13 - Si considerino tre segnali costituiti da rettangoli di ampiezza A rispettivamente negli intervalli $(0, T_0/2)$, $(T_0/2, T_0)$ e $(0, T_0)$. A questi si aggiunga un quarto segnale positivo in $(0, T_0/2)$ e negativo in $(T_0/2, T_0)$. Quali funzioni base possono rappresentare i quattro segnali, e quale è la geometria dei segnali? Se si aggiungono i quattro segnali opposti quale diventa la rappresentazione geometrica? Si calcolino le distanze tra coppie di segnali sia mediante le coordinate geometriche sia attraverso le forme d'onda. *Commento*: naturalmente si devono ottenere gli stessi risultati.

1.14 - Riprendendo gli otto segnali dell'esercizio precedente, si aumenti l'ampiezza delle quattro forme d'onda aventi energia maggiore fino a ottenere la massima distanza minima possibile tra i segnali. Quale è l'ampiezza delle quattro forme d'onda? Si calcoli la distanza minima sia geometricamente sia mediante le forme d'onda.

1.15 - Si considerino le quattro forme d'onda $A\cos(2\pi f_0 t \pm \pi t/T_0 + \varphi_k)$ nell'intervallo $(0, T_0)$, con $\varphi_k = 0, \pi/2$. Si mostri che le funzioni sono *ortogonali*. Quale è la rappresentazione geometrica, con un piccolo sforzo di fantasia per uscire nella quarta dimensione? Se poi si aggiungono le quattro forme d'onda opposte quale è la geometria? *Commento*: i segnali sono detti *biortogonali*; ogni forma d'onda ne ha sei ortogonali ed una opposta.

1.16 - Si considerino quattro forme d'onda ortogonali $g(t - kT)$ e le loro opposte. Si mostri che la geometria è la stessa dell'esercizio precedente.

1.17 - Si considerino le quattro forme d'onda

$$s_i(t) = \sum_{k=1}^{4} a_k g(t - kT)$$

dove le funzioni $g(t-kT)$ sono ortogonali e gli a_k sono rispettivamente (1,1,1,1), $(1, 1, -1, -1)$, $(1, -1, 1, -1)$ e $(1, -1, -1, 1)$. Ai quattro segnali si aggiungano

gli opposti. Si mostri che la geometria degli otto segnali è la stessa dell'esercizio precedente.

1.18 - Si considerino nell'intervallo $(0, T_0)$ le forme d'onda $\cos 2\pi f_k t$ con frequenze $f_k = f_0 + k/2T_0$ $(k = 1, \ldots, 4)$, e le opposte. Si mostri che la geometria è la stessa dell'esercizio precedente.

1.19 - In un intervallo T_0 suddiviso in due parti uguali si considerino le quattro forme d'onda $\pm A \cos 2\pi f_0 t; \pm A \cos 2\pi f_0 t$ (da intendersi nel senso che in ciascuno dei due intervalli la forma d'onda può essere positiva o negativa) e le altre quattro $\pm A \sin 2\pi f_0 t; \pm A \sin 2\pi f_0 t$. Si calcolino le distanze tra coppie di funzioni, e da queste distanze si verifichi che la geometria è, ancora una volta, quella di otto segnali biortogonali. Come conviene scegliere le funzioni base per mettere in evidenza questo fatto?

1.20 - Si considerino i segnali

$$\sum_{k=1}^{4} a_k g(t - kT) \cos 2\pi f_0 t \qquad \text{oppure} \qquad \sum_{k=1}^{4} a_k g(t - kT) \sin 2\pi f_0 t$$

in cui le forme d'onda $g(t - kT)$ sono ortogonali, $a_k = \pm 1$ e solo un numero *pari* di a_k può avere segno negativo. Quanti sono i segnali? Quante dimensioni occupano? Detta E_s l'energia di ciascuno dei quattro simboli, e quindi $4E_s$ l'energia di ciascuna forma d'onda, si mostri che la distanza minima tra le forme d'onda è $8E_s$. Si faccia qualche esempio di coppie di segnali a distanza minima.

1.21 - Supponendo vera la (1.63) si dimostri la (1.64). *Suggerimento*: si inizi con il mostrare che $E[n(t)n_k] = \sigma_k^2 \Phi_k(t)$; naturalmente non è lecito utilizzare $n(t) = \sum n_k \Phi_k(t)$, che è il risultato da dimostrare.

1.22 - L'energia in un intervallo T_0 di una realizzazione del rumore $n(t)$ è una variabile casuale. Si mostri che il valor medio dell'energia è dato da

$$E[\int_0^{T_0} n^2(t)dt] = \sum_{k=1}^{\infty} \sigma_k^2 \tag{1.86}$$

Suggerimento: si espanda $n(t)$ in somma di funzioni base. *Commento*: per ogni processo $\sigma_k^2 \to 0$ per $k \to \infty$; se invece si vuol considerare un rumore bianco

su tutto l'asse delle frequenze, un'idealizzazione non fisica, *tutte* le componenti hanno varianza $N_0/2$ e l'energia è infinita.

1.23 - Utilizzando il risultato dell'esercizio precedente si mostri che la varianza del processo $n(t)$ è data da

$$\sigma^2_{n(t)} = \frac{1}{T_0} \sum_{k=1}^{\infty} \sigma_k^2 \tag{1.87}$$

Suggerimento:

$$E[\int_0^{T_0} n^2(t)dt] = \int_0^{T_0} E[n^2(t)]dt$$

Commento: per ogni processo $\sigma_k^2 \to 0$ per $k \to \infty$; se invece si vuol considerare un rumore bianco su tutto l'asse delle frequenze, un'idealizzazione non fisica, *tutte* le componenti hanno varianza $N_0/2$ e $\sigma^2_{n(t)} = \infty$.

Capitolo 2

Fondamenti di trasmissione numerica

2.1 Introduzione

Sia $\{m_i\}$ $(i = 1, \ldots, M)$ l'insieme dei possibili messaggi, ed a questi siano associate le forme d'onda $s_i(t)$. L'accordo fra trasmettitore e ricevitore è che per inviare l'i-esimo messaggio si trasmette la forma d'onda corrispondente. Il ricevitore conosce l'insieme delle forme d'onda $\{s_i(t)\}$ ma non quale è stata effettivamente trasmessa, che è suo compito determinare[1].

Si dovrebbe distinguere tra la forma d'onda trasmessa, quella ricevuta (escluso il rumore), e quella effettivamente osservata a valle degli amplificatori di ricezione, in cui è incluso il rumore. Spesso la forma d'onda all'ingresso del ricevitore, ignorando il rumore, è una replica attenuata di quella trasmessa. In tali casi viene comodo sottintendere l'attenuazione del mezzo trasmissivo e chiamare segnale *trasmesso* la replica *ricevuta* della forma d'onda trasmessa. Quando ad esempio si afferma che la probabilità d'errore dipende dall'energia del segnale trasmesso, si intende ovviamente l'energia che raggiunge effettivamente il ricevitore.

[1] in generale le forme d'onda inviate sul canale di trasmissione potrebbero dipendere da parametri non perfettamente conosciuti dal ricevitore; non vi sarebbe quindi una corrispondenza biunivoca tra messaggi e segnali, e forme d'onda *diverse* ad esempio per ampiezza, fase, frequenza o posizione temporale rappresenterebbero lo stesso messaggio; può anche accadere che i segnali siano distorti dal mezzo trasmissivo in modo casuale, cioè prevedibile dal ricevitore solo in senso statistico. Questi casi, non infrequenti, verranno considerati nel seguito

Dunque spesso si scrive che il segnale ricevuto $r(t)$ è la somma di quello trasmesso e del rumore

$$r(t) = s_i(t) + n(t) \tag{2.1}$$

mentre sarebbe corretto affermare che $r(t)$ è la somma del rumore e dell'effetto, all'ingresso del ricevitore, del segnale trasmesso. Questo è attenuato ed eventualmente anche filtrato, o distorto in modo non lineare. Nella (2.1) è anche ignorato il controllo automatico di guadagno del ricevitore (AGC), che mantiene il segnale al livello più conveniente per le elaborazioni.

Per quanto riguarda il rumore in questo capitolo si supporrà che sia indipendente dal segnale, gaussiano e bianco nella banda dei segnali, con densità spettrale bilatera $N_0/2$.

Compito del ricevitore è decidere quale delle possibili forme d'onda sia stata trasmessa. Il criterio di ottimalità solitamente ritenuto più conveniente è la minimizzazione della probabilità d'errore, senza assegnare pesi diversi ai vari tipi di errore[2].

La minima probabilità di errore si ottiene evidentemente se, ricevuta la forma d'onda $r(t)$, si sceglie comunque il messaggio più probabile *a posteriori*. Ovviamente si sta supponendo che non si possa rifiutare di decidere. Se in un sistema di trasmissione binario si determina che, dato il particolare segnale $r(t)$ ricevuto, un messaggio ha probabilità 0.51 e l'altro 0.49 conviene effettivamente scegliere il primo, ma è chiaro che *questa decisione* è sbagliata con probabilità 0.49. Comunque se anche è lecito non decidere, ad esempio perché si può chiedere la *ritrasmissione* del messaggio, per decidere di non decidere occorre aver calcolato le probabilità a posteriori dei due messaggi. È dunque questo il primo problema da affrontare.

2.2 Probabilità a posteriori

Dato il segnale ricevuto $r(t) = s_i(t) + n(t)$ le probabilità a posteriori possono essere calcolate mediante la *regola di Bayes*, che è lo strumento tipico della

[2] è più grave che un sì diventi no, o viceversa? dipende dal significato del messaggio, su cui si preferisce non indagare; chi ha da trasmettere messaggi estremamente importanti si cautela chiedendo che il sistema di trasmissione abbia una probabilità d'errore molto bassa; chi progetta il sistema di trasmissione provvede a garantire la probabilità richiesta; è poco frequente che si progetti un sistema di trasmissione in modo da minimizzare una probabilità d'errore *pesata*, soprattutto per la difficoltà di assegnare i valori dei pesi

teoria della decisione. Infatti date possibili *cause* A_i ed un *effetto* osservato B, si ha

$$P(A_i/B) = \frac{P(B/A_i)P(A_i)}{P(B)} \tag{2.2}$$

Se l'effetto è rappresentabile con una variabile casuale anziché un evento, la regola di Bayes è

$$P(A_i/x) = \frac{f(x/A_i)P(A_i)}{f(x)} \tag{2.3}$$

con ovvia estensione al caso di più variabili casuali congiunte. In ogni caso la semplificazione prodotta dalla regola di Bayes deriva dal dover calcolare probabilità dell'effetto data la causa, e non viceversa.

Ai fini della decisione si può ignorare il denominatore, se non interessa calcolare le effettive probabilità a posteriori ma solo scegliere il massimo; infatti il denominatore ha un valore indipendente dall'ipotesi A_i.

Nel caso della trasmissione numerica le cause possibili sono i segnali $s_i(t)$ e l'effetto è la forma d'onda ricevuta $r(t)$. Quest'ultima non è però rappresentabile con un insieme *finito* di variabili casuali: $r(t)$ è un processo casuale, cioè un'infinità *non numerabile* di variabili casuali. Ciò rende non immediato applicare la regola di Bayes. Un primo passo verso la soluzione consiste nel rappresentare *geometricamente* $r(t)$ mediante il corrispondente vettore $\mathbf{r}$. Questo ha in generale, per colpa del rumore, infinite componenti $r_1, r_2, \ldots$ e non ha evidentemente senso scrivere la densità di probabilità congiunta di queste *infinite* variabili casuali. Tuttavia si è passati da un'infinità non numerabile di variabili casuali ad un'infinità *numerabile*, e *senza perdere nulla* perché il vettore $\mathbf{r}$ è del tutto equivalente alla forma d'onda $r(t)$. Ora si possono considerare ricevitori, forse non ottimali, che utilizzino un numero *finito* di componenti r_k del vettore ricevuto ($k = 1, \ldots, n$). Si valuterà poi quale compromesso tra complessità e prestazioni sia conveniente, cioè quale debba essere il valore di n.

Se l'insieme di funzioni base è scelto in modo opportuno, come descritto nel Cap. 1, le n componenti r_k sono incorrelate e quindi indipendenti. Dato che si sia trasmesso il vettore $\mathbf{s}_i$, r_k ha valor medio, dovuto al segnale,

$$E[r_k/\mathbf{s}_i] = \begin{cases} s_{ik} & n \leq N \\ 0 & n > N \end{cases} \tag{2.4}$$

e varianza σ_k^2, pari ad $N_0/2$ per $k \leq N$. La densità di probabilità (*ddp*) congiunta, dato che si sia trasmesso $\mathbf{s}_i$, è quindi

$$f(r_1, \ldots, r_n/\mathbf{s}_i) =$$

$$= \prod_{k=1}^{N} \frac{1}{\sqrt{\pi N_0}} \exp\Big(- \frac{(r_k - s_{ik})^2}{N_0}\Big) \prod_{k=N+1}^{n} \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp\Big(- \frac{r_k^2}{2\sigma_k^2}\Big) \tag{2.5}$$

se $n > N$; altrimenti si ha solo il primo termine, con indici da 1 ad n. Sembra comunque necessario che sia almeno $n = N$ per non trascurare alcune componenti dei segnali.

Ai fini della ricerca del massimo si possono ignorare tutti i fattori moltiplicativi che non dipendono dall'indice i. In particolare, per quanto grande sia n, non dipendono da *i tutti* i termini della (2.5) con indice $k > N$. Dunque le componenti r_k per $k > N$, cioè le componenti del rumore lungo assi che non contengono segnale, sono *irrilevanti*. Si usa dire che le componenti r_k, per $k \leq N$, costituiscono una *statistica sufficiente*. Inutile quindi per $k > n$ calcolare le correlazioni r_k del segnale ricevuto $r(t)$ con le funzioni base $\Phi_k(t)$, ed anzi inutile preoccuparsi di determinare le stesse $\Phi_k(t)$!

Nel caso quindi di rumore gaussiano bianco nella banda dei segnali sono sufficienti le N funzioni base richieste per rappresentare i segnali, e le corrispondenti componenti del vettore ricevuto.

La *ddp* condizionata $f(\mathbf{r}/\mathbf{s}_i)$ è proporzionale a

$$\exp(-\frac{1}{N_0}\sum_{k=1}^{N}(r_k - s_{ik})^2) = \exp(-\frac{1}{N_0}|\mathbf{r} - \mathbf{s}_i|^2) \tag{2.6}$$

dove la distanza al quadrato $|\mathbf{r} - \mathbf{s}_i|^2$ è calcolata nello spazio ad N dimensioni. Se si volesse calcolare la stessa distanza *non geometricamente* ma mediante un integrale nella variabile t si otterrebbe

$$|\mathbf{r} - \mathbf{s}_i|^2 = \int (r(t) - s_i(t))^2 dt = \sum_{k=1}^{\infty}(r_k - s_{ik})^2 =$$

$$= \sum_{k=1}^{N}(r_k - s_{ik})^2 + \sum_{k=N+1}^{\infty} r_k^2 \tag{2.7}$$

che differisce per un termine indipendente da i. Questo produce, nell'esponenziale, un inessenziale fattore moltiplicativo.

Tornando alla regola di Bayes, si deve ricercare il massimo di

$$f(\mathbf{r}/\mathbf{s}_i)P(\mathbf{s}_i) \equiv \exp(-\frac{1}{N_0}|\mathbf{r}-\mathbf{s}_i|^2)P(\mathbf{s}_i) \tag{2.8}$$

dove $P(\mathbf{s}_i)$ sono le probabilità *a priori*, cioè prima della trasmissione, dei messaggi emessi dalla sorgente, non necessariamente equiprobabili. Tuttavia scostarsi troppo dalla equiprobabilità è uno spreco, da evitare mediante la codifica di sorgente. Solitamente i messaggi hanno probabilità poco diverse l'una dall'altra, e spesso *non note*. È perciò assai comune che i fattori $P(\mathbf{s}_i)$ siano ignorati, e si cerchi non il massimo delle probabilità a posteriori ma il massimo delle sole *verosimiglianze* $f(\mathbf{r}/\mathbf{s}_i)$. Le due strategie sono dette rispettivamente a *massima probabilità a posteriori* (MAP) e a *massima verosimiglianza* (MV, o più spesso ML: *Maximum Likelihood*).

Ai fini della ricerca del massimo si possono considerare i logaritmi di probabilità a posteriori e verosimiglianze, moltiplicati per N_0 e cambiati di segno (naturalmente di questi si cercherà il minimo), ovvero rispettivamente $|\mathbf{r}-\mathbf{s}_i|^2 - N_0 \log P(\mathbf{s}_i)$ e $|\mathbf{r}-\mathbf{s}_i|^2$.

Nel caso ML tutto si riduce a cercare fra i possibili vettori $\mathbf{s}_i$ quello alla *minima distanza* dal vettore ricevuto $\mathbf{r}$. Si noti invece che per la decisione MAP occorre conoscere oltre alle probabilità a priori $P(\mathbf{s}_i)$ anche la densità spettrale di potenza $N_0/2$ del rumore, cioè conoscere esattamente il rapporto segnale-rumore. Questo complica il ricevitore senza un significativo miglioramento delle prestazioni, per cui è assai più frequente il ricevitore ML.

Nei casi più semplici si può determinare a priori, per ogni i, il luogo dei punti I_i più vicini al segnale $\mathbf{s}_i$ che ad ogni altro, cioè la i-esima *regione di decisione*. L'esempio più semplice è quello di due segnali $\mathbf{s}_1$ ed $\mathbf{s}_2 = -\mathbf{s}_1$ (segnali *antipodali*) con energia E_s, cioè con coordinate $\pm\sqrt{E_s}$ sull'unico asse. Il vettore $\mathbf{r}$ ricevuto ha un'unica componente rilevante r_1, ed è evidente che i punti più vicini a $\mathbf{s}_1$ ed $\mathbf{s}_2$ corrispondono a $r_1 > 0$ ed $r_1 < 0$, rispettivamente. Le regioni di decisione sono i due semiassi positivo e negativo, e nessuno si sognerebbe di calcolare le due distanze al quadrato $(r_1 - \sqrt{E_s})^2$ e $(r_1 + \sqrt{E_s})^2$ per sapere quale è minore.

Analogamente si considerino i quattro segnali in due dimensioni di fig. 2.1 (modulazione 4PSK: *Phase Shift Keying* a 4 fasi), in cui è anche indicata

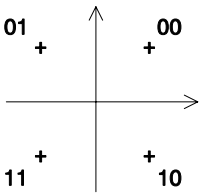

Fig. 2.1 - Costellazione 4PSK (e relativo *mapping*)

una possibile corrispondenza tra segnali e coppie di bit[3]. È evidente che le regioni di decisione del ricevitore ML sono i quattro quadranti, e che quindi la decisione è basata sui due segni delle componenti r_1 ed r_2 del vettore ricevuto. In modo analogo si determinano le regioni di decisione per le costellazioni di fig. 1.2 e 1.6.

Nei semplici casi visti finora i confini delle regioni di decisione sono paralleli agli assi, ma ciò non accade in generale (si pensi ad esempio a due o più segnali ortogonali). Confrontando *separatamente* le componenti r_k con delle soglie si possono ottenere solamente regioni di decisione con confini paralleli agli assi. Quindi in generale le componenti r_k vanno esaminate *congiuntamente*: non ha senso prendere una decisione indipendente per ogni asse, come il principiante è sempre tentato di fare.

Il quadrato della distanza tra vettore ricevuto $\mathbf{r}$ e segnale $\mathbf{s}_i$ è esprimibile come[4]

$$|\mathbf{r} - \mathbf{s}_i|^2 = |\mathbf{r}|^2 - 2\mathbf{r} \cdot \mathbf{s}_i + |\mathbf{s}_i|^2 \tag{2.9}$$

Il termine $|\mathbf{r}|^2$ non dipende dall'indice i, e può essere ignorato. Cambiando nuovamente segno e dividendo per 2 si è ricondotti alla ricerca del massimo di

$$\mathbf{r} \cdot \mathbf{s}_i - \frac{1}{2}|\mathbf{s}_i|^2 \tag{2.10}$$

a cui si deve aggiungere $(N_0/2)\log P(\mathbf{s}_i)$ nel caso di ricevitore MAP. Si noti che i termini non dipendenti da $\mathbf{r}$ sono *precalcolabili*, e solo per la correlazione $\mathbf{r} \cdot \mathbf{s}_i$ si deve attendere di ricevere $r(t)$.

Si può infine osservare che se i segnali $s_i(t)$ hanno tutti la stessa energia il segnale più verosimile è quello per cui è massima la correlazione $\mathbf{r} \cdot \mathbf{s}_i$.

[3] in gergo si chiama *mapping* l'associazione degli $m = \log_2 M$ bit ai simboli

[4] basta applicare la proprietà distributiva al prodotto scalare $(\mathbf{r} - \mathbf{s}_i) \cdot (\mathbf{r} - \mathbf{s}_i) = |\mathbf{r} - \mathbf{s}_i|^2$

Quanto al calcolo di $\mathbf{r} \cdot \mathbf{s}_i$ vale quanto già detto nel Cap. 1: la correlazione può essere valutata mediante la

$$\mathbf{r} \cdot \mathbf{s}_i = \sum_{k=1}^{N} r_k s_{ik} \tag{2.11}$$

dopo aver calcolato le componenti del segnale ricevuto

$$r_k = \int r(t)\Phi_k(t)dt \tag{2.12}$$

oppure direttamente come

$$\mathbf{r} \cdot \mathbf{s}_i = \int r(t)s_i(t)dt \tag{2.13}$$

Si tratta comunque di calcolare le correlazioni tra $r(t)$ e le funzioni base $\Phi_k(t)$ oppure i segnali $s_i(t)$, con i metodi discussi nel Cap. 1.

Non è possibile dire a priori quale soluzione sia più conveniente, ma è da vedere caso per caso quanti e quali siano i segnali e le funzioni base. Certamente se il numero M di segnali è molto elevato, come sempre accade nei sistemi efficienti di trasmissione numerica, non si può neppure pensare di calcolare *tutte* le correlazioni $\mathbf{r} \cdot \mathbf{s}_i$, scriverle in una memoria gigantesca, ed infine cercare il massimo. La ricerca del segnale più verosimile, o di quello più probabile, deve poter essere condotta limitando la ricerca ad un numero trattabile di casi. L'insieme dei segnali $\mathbf{s}_i$ deve essere scelto avendo in mente questo scopo.

Vale la pena di ricordare, come già segnalato nel Cap. 1, che nel caso di segnali passa banda la notazione diventa più sintetica se si rappresentano le componenti s_{ik} ed $s_{ik'}$ del segnale lungo gli assi $\Phi_k(t)$ e $\Phi_{k'}(t)$ con un unico numero complesso[5] s_{ik}, e analogamente le componenti del segnale ricevuto $r(t)$ con il numero complesso

$$r_k = \frac{1}{2} \int z(t)z_k^*(t)dt \tag{2.14}$$

[5]dal contesto è sempre chiaro se si stanno considerando variabili reali o complesse; nel primo caso s_{ik} indica la componente lungo l'asse $\Phi_k(t)$, nel secondo la coppia di componenti $s_{ik} + js_{ik'}$

dove $z(t)$ e $z_k(t)$ sono gli equivalenti passa basso del segnale ricevuto e della funzione base $\Phi_k(t)$. Resta solo da osservare che la somma dei due contributi alla correlazione è esprimibile come

$$\begin{aligned} &\int r(t)\big(s_{ik}\Phi_k(t) + s_{ik'}\Phi_{k'}(t)\big)dt = \\ &= \frac{1}{2}\mathrm{Re}\{\int z(t)(s_{ik} + js_{ik'})^* z_k^*(t)dt\} = \mathrm{Re}\{r_k s_{ik}^*\} \end{aligned} \tag{2.15}$$

e quindi, considerando tutte le componenti e utilizzando anche per s_{ik} la notazione complessa,

$$\mathbf{r}\cdot\mathbf{s}_i = \mathrm{Re}\{\sum r_k s_{ik}^*\} \tag{2.16}$$

2.3 Probabilità d'errore: trasmissione binaria

Si ha errore quando il vettore ricevuto cade al di fuori della regione di decisione I_i corrispondente al segnale trasmesso $\mathbf{s}_i$. Le regioni di decisione degli M segnali possono avere forma diversa, e quindi in generale i vari casi sono da trattare separatamente. La probabilità d'errore $P(E)$ non condizionata[6] è la media delle condizionate

$$P(E) = \sum_{i=1}^{M} P(\mathbf{s}_i)P(E/\mathbf{s}_i) = \sum_{i=1}^{M} P(\mathbf{s}_i)\sum_{j\neq i} P(\mathbf{s}_j/\mathbf{s}_i) \tag{2.17}$$

dove $P(\mathbf{s}_j/\mathbf{s}_i)$ è la probabilità che avendo trasmesso $\mathbf{s}_i$ si decida a favore di $\mathbf{s}_j$. In linea di principio il calcolo di $P(E/\mathbf{s}_i)$ oppure $P(\mathbf{s}_j/\mathbf{s}_i)$ richiede la valutazione dell'integrale della *ddp* congiunta del vettore ricevuto $\mathbf{r}$ in una regione opportuna. In alcuni casi ciò è facile, ma in generale ottenere risultati numerici esatti è estremamente complesso. Fortunatamente non occorre mai grande precisione[7].

[6] sempre che la media sia significativa; in certi casi può essere più appropriato determinare il caso peggiore, cioè il massimo di $P(E/\mathbf{s}_i)$

[7] un fattore moltiplicativo 1/2 o 2 viene accettato a cuor leggero, perché non interessa tanto $P(E)$ per un rapporto segnale-rumore prefissato, quanto piuttosto il rapporto segnale-rumore richiesto per una certa $P(E)$; l'andamento di $P(E)$ in funzione del rapporto segnale-rumore è molto ripido, per cui non occorre valutare $P(E)$ con grande precisione

Si può osservare fin d'ora che le probabilità $P(\mathbf{s}_j/\mathbf{s}_i)$ dipendono solo dalla posizione relativa dei segnali, e quindi non direttamente dalle forme d'onda $s_i(t)$ ma dalla disposizione geometrica dei corrispondenti vettori. Insiemi di forme d'onda diverse ma con la stessa rappresentazione geometrica danno le stesse prestazioni. Inoltre la posizione degli assi rispetto ai segnali non ha alcuna importanza: se si trasla o ruota rigidamente l'insieme dei segnali, lo stesso avviene per la *ddp* congiunta del vettore ricevuto e per le regioni di decisione, e quindi non cambia la probabilità d'errore; cambiano invece le *forme d'onda* e la loro *energia*, e di ciò conviene tener conto nel progetto.

La probabilità d'errore è calcolabile esattamente, e comodamente, nel caso di due soli segnali $\mathbf{s}_1$ ed $\mathbf{s}_2$, a distanza d. Il confine tra le due regioni di decisione è l'asse della congiungente gli estremi dei vettori. Traslando e ruotando i segnali in modo da disporre un asse lungo la congiungente dei due segnali è evidente che si deve calcolare la probabilità che una variabile casuale gaussiana con valor medio nullo e varianza $N_0/2$ superi $d/2$, e quindi

$$P(\mathbf{s}_2/\mathbf{s}_1) = P(\mathbf{s}_1/\mathbf{s}_2) = P(E) = Q\Big(\frac{d/2}{\sqrt{N_0/2}}\Big) = Q\Big(\frac{d}{\sqrt{2N_0}}\Big) \tag{2.18}$$

dove

$$Q(y) = \frac{1}{\sqrt{2\pi}} \int_y^\infty \exp(-x^2/2)dx \tag{2.19}$$

è la probabilità che una variabile casuale gaussiana normalizzata, con valor medio nullo e varianza unitaria, superi y.

Si consideri ad esempio il caso di segnali antipodali ($\mathbf{s}_2 = -\mathbf{s}_1$) con energia E_s e quindi a distanza $\sqrt{E_s}$ dall'origine. L'energia per simbolo E_s coincide con l'energia E_b spesa per trasmettere un bit d'informazione. Pertanto si ha $d = 2\sqrt{E_s} = 2\sqrt{E_b}$, e quindi

$$P(E) = Q\Big(\sqrt{\frac{2E_b}{N_0}}\Big) \tag{2.20}$$

Ad esempio in tab. 1 sono indicati i valori di E_b/N_0 (in dB) richiesti per ottenere particolari probabilità d'errore. Si noti che per ridurre $P(E)$ da 10^{-5}, valore adeguato per molte applicazioni, a 10^{-13}, praticamente nullo, occorrono quasi 5 dB. Si noti anche che E_b/N_0 non è il rapporto segnale-rumore all'uscita del filtro adattato, che risulta pari a $2E_b/N_0$ (es. 2.1). D'altra parte definire un rapporto segnale-rumore all'ingresso del decisore è facile (ma è anche utile?)

$P(E)$	10^{-3}	10^{-5}	10^{-7}	10^{-10}	10^{-13}
E_b/N_0 (dB)	6.79	9.59	11.31	13.06	14.31

Tab. 1 - Valori di E_b/N_0 (in dB) richiesti per ottenere una prefissata probabilità d'errore $P(E)$ (trasmissione binaria antipodale)

solo nel caso binario. È quindi consuetudine, in generale, dare risultati e grafici in funzione di E_b/N_0.

La funzione $Q(y)$ non ha una espressione analitica chiusa, ma si trova tabulata e ne esistono ottime approssimazioni, come

$$Q(y) \approx \frac{1}{\sqrt{2\pi}y} \exp(-y^2/2) \tag{2.21}$$

che è utilizzabile per $y > 3$, e quindi per $Q(y) < 10^{-3}$. Per tutti i valori $y \geq 0$ si commette un errore entro lo 0.27% con l'approssimazione

$$Q(y) \approx \frac{1}{\sqrt{2\pi}} \exp(-y^2/2) \frac{1}{(1-a)y + a\sqrt{y^2+b}} \tag{2.22}$$

dove $a = 0.339$ e $b = 5.51$.

Talvolta alla funzione $Q(\cdot)$ viene preferita la *funzione errore complementare*

$$erfc(y) = \frac{2}{\sqrt{\pi}} \int_y^\infty \exp(-x^2)dx = 2Q(y\sqrt{2}) \tag{2.23}$$

a cui corrisponde la relazione inversa

$$Q(y) = \frac{1}{2} erfc(\frac{y}{\sqrt{2}}) \tag{2.24}$$

per cui la probabilità d'errore nella trasmissione binaria antipodale è data da

$$P(E) = \frac{1}{2} erfc\Big(\sqrt{\frac{E_b}{N_0}}\Big) \tag{2.25}$$

Nel caso di due segnali *ortogonali* con energia $E_s = E_b$ la distanza d risulta uguale a $\sqrt{2E_s}$ ed è ridotta del fattore $\sqrt{2}$ rispetto alla trasmissione binaria

antipodale. Occorre quindi una energia doppia (3 dB in più) per ottenere la stessa probabilità d'errore:

$$P(E) = Q\Big(\sqrt{\frac{E_b}{N_0}}\Big) \tag{2.26}$$

L'ultimo caso binario di un qualche interesse è la trasmissione[8] *on-off*, in cui i segnali sono $s_1(t)$ ed $s_2(t) = 0$ (anche l'assenza di forma d'onda è un segnale, sebbene chi ha orrore del vuoto non voglia convincersene). Detta E_1 l'energia del primo segnale la distanza è $d = \sqrt{E_1}$. Si possono dare definizioni diverse dell'energia per simbolo, e quindi per bit: l'energia *media* $E_1/2$ (nel caso di segnali equiprobabili), oppure l'energia del caso peggiore E_1. Nel primo caso si troverebbero le stesse prestazioni dei segnali ortogonali, nel secondo un peggioramento di 3 dB. Non c'è una scelta *giusta*. Ciò che conta è la disponibilità, il costo ed eventualmente il consumo del dispositivo amplificatore che deve generare le forme d'onda da trasmettere. Si potrebbe essere tentati di pensare che l'energia del caso peggiore sia il parametro importante, perché l'amplificatore deve essere in grado di generarla. Ma ancora più dell'energia conta la *potenza di picco*, e questa non si calcola dall'energia del segnale dividendo per l'intervallo di tempo T_0, come sarebbe corretto se si trasmettessero forme d'onda rettangolari. Dunque persino l'energia del caso peggiore *da sola* non racconta l'intera storia.

L'importante è essere ben chiari, nel riferire i risultati, su cosa si intende per energia per simbolo, o per bit. Spesso ci si riferisce all'energia *media*, ma è solo una convenzione. Non si deve cadere nell'errore di dare troppa importanza all'energia media. Potrebbe venire la tentazione di darsi come regola di traslare un qualunque insieme di segnali $\{\mathbf{s}_i\}$ in modo da rendere minima l'energia media (es. 2.3). Benché ciò sia ragionevole in alcuni casi, si possono costruire esempi in cui sarebbe un vero delitto (es. 2.4).

Il confronto tra ricevitori MAP e ML è decisamente a favore di questi ultimi dal punto di vista della semplicità di realizzazione, ma occorre vedere se vi sia una differenza significativa di prestazioni. La fig. 2.2 mostra $P(E)$ in funzione di E_b/N_0 nel caso di segnali antipodali con probabilità a priori fortemente sbilanciate $P(\mathbf{s}_1) = 0.9$ e $P(\mathbf{s}_2) = 0.1$. Nel ricevitore MAP la soglia di decisione è leggermente spostata rispetto a $r_1 = 0$ (es. 2.5). Lo spostamento dipende da N_0, cosa fastidiosa perché occorre conoscerne il valore. La curva

[8] si dice anche *segnalazione*

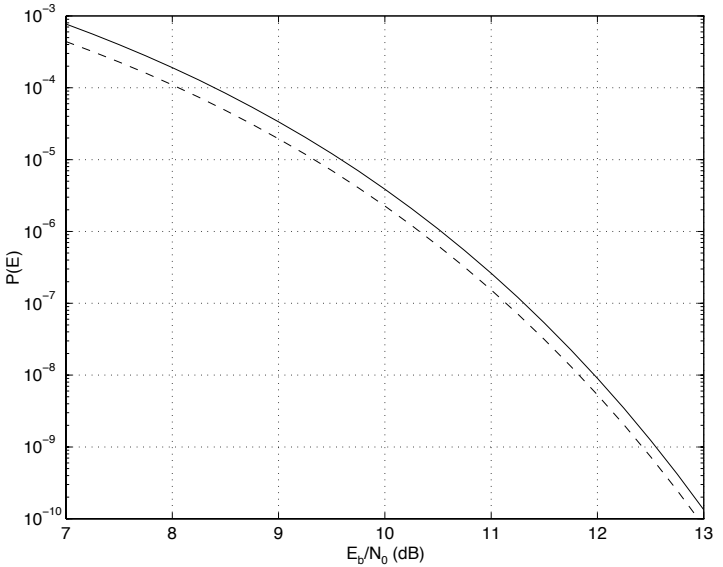

Fig. 2.2 - Probabilità d'errore con ricevitori ML (curva continua) e MAP (tratteggio) nella trasmissione binaria antipodale di segnali con probabilità a priori 0.9 e 0.1

MAP di fig. 2.2 è ottenuta ottimizzando la soglia per *ogni* valore di E_b/N_0, cosa difficilmente realizzabile in pratica, eppure il miglioramento è modesto.

Come già accennato, in questi casi occorre intervenire con la codifica di sorgente. La quantità d'informazione è $I = -0.9 \log_2 0.9 - 0.1 \log_2 0.1 = 0.47$ bit/messaggio. Ciò significa che sarebbe possibile trasmettere la *stessa informazione* con meno della metà dei bit, e quindi anche dell'energia.

Passando ad esaminare il caso di più di due segnali diventa importante distinguere tra energia per simbolo ed energia per bit d'informazione. Non ha infatti senso confrontare le energie spese da due sistemi che trasmettono un numero diverso di bit[9]. Conviene quindi far riferimento all'energia spesa per bit d'informazione, data da $E_b = E_s / \log_2 M$.

[9] sarebbe come dire che un sistema di trasmissione binario è più efficiente se utilizzato per trasmettere un solo bit in tutta la sua vita, perché si spende poca energia mentre continuando a trasmettere se ne consuma tanta

2.4 Probabilità d'errore: trasmissione non binaria

In qualche caso non binario le regioni di decisione hanno una forma abbastanza regolare da consentire il calcolo esatto della probabilità d'errore. Ad esempio per la costellazione 4PSK di fig. 2.1 non si ha errore solo se entrambe le componenti del rumore non superano (nel verso pericoloso) $d/2$, dove d è la distanza minima tra segnali, e quindi

$$1 - P(E) = \left(1 - Q(\frac{d}{\sqrt{2N_0}})\right)^2 \tag{2.27}$$

ovvero, trascurando $Q^2(\cdot)$ rispetto a $Q(\cdot)$,

$$P(E) \approx 2Q(\frac{d}{\sqrt{2N_0}}) \tag{2.28}$$

Se la sorgente emette cifre binarie e queste sono trasmesse a coppie mediante un simbolo 4PSK, più che la probabilità $P(E)$ che il *simbolo* deciso sia errato interessa la probabilità che i *bit* decisi siano errati. Quest'ultima probabilità verrà indicata con $P_b(E)$, e detta probabilità d'errore *sui bit*[10] per distinguerla da quella sui simboli. Dato il simbolo trasmesso $\mathbf{s}_i$ sono possibili decisioni errate $\mathbf{s}_j$ diverse (tre, nel caso di quattro segnali), che producono un diverso numero di bit errati. Quindi $P_b(E)$ non dipende solo dalla geometria dei segnali ma anche da come questi sono associati alle coppie di bit, o più in generale agli $m = \log_2 M$ bit. Per questo motivo nella costellazione di fig. 2.1 sono indicate anche le coppie di bit corrispondenti ai segnali.

In generale se $\mathbf{s}_i$ e $\mathbf{s}_j$ differiscono per n_{ij} bit si ha

$$P_b(E) = \frac{1}{\log_2 M} \sum_{i=1}^{M} P(\mathbf{s}_i) \sum_{j \neq i} n_{ij} P(\mathbf{s}_j/\mathbf{s}_i) \tag{2.29}$$

dove $n_{ij}/\log_2 M$ è la frazione di bit errati rispetto a quelli trasmessi[11]. In ogni caso $1 \leq n_{ij} \leq \log_2 M$, e quindi si ottiene

$$\frac{1}{\log_2 M} P(E) \leq P_b(E) \leq P(E) \tag{2.30}$$

[10] modo di dire ben poco elegante, ma sintetico

[11] chi non sia del tutto convinto pensi al seguente caso: $M = 4$, $P(E) = 10^{-5}$ e $n_{ij} = 1$ (ritenendo ininfluente, perché poco probabile, il caso di due errori); si trasmettano 10^{10} simboli, e quindi $2 \cdot 10^{10}$ bit; si hanno circa 10^5 simboli errati e 10^5 bit errati; ne risulta $P_b(E) = 5 \cdot 10^{-6} = P(E)/2$

per cui in un primo progetto di massima può essere sufficiente determinare $P(E)$.

Per la costellazione 4PSK, detta E_s l'energia di un simbolo la distanza minima è $d = \sqrt{2E_s}$ e poiché $E_s = 2E_b$ si ottiene con semplici calcoli lasciati al lettore (e senza dover fare alcuna approssimazione)

$$P_b(E) = Q\Big(\sqrt{\frac{2E_b}{N_0}}\Big) \tag{2.31}$$

esattamente come nel caso binario antipodale. Il risultato non deve sorprendere. Basta ricordare, come discusso nel Cap. 1, che la trasmissione *successiva* di due bit con segnali antipodali può essere interpretata come la trasmissione di uno tra quattro segnali aventi la stessa geometria del caso 4PSK, e lo stesso *mapping*; oppure basta pensare alla modulazione 4PSK come la trasmissione indipendente di due bit sugli assi in fase e quadratura, con modulazione binaria antipodale.

Un altro caso in cui la determinazione della probabilità d'errore è agevole è la modulazione d'ampiezza multilivello (fig. 1.2, nel caso di 8 livelli). Detta d la distanza tra livelli adiacenti, si ha $P(E/\mathbf{s}_i) = Q(d/\sqrt{2N_0})$ per i due livelli più esterni, e una probabilità d'errore doppia per quelli interni. Non resta che determinare l'energia media per simbolo, e quindi per bit, per ottenere, nel caso di segnali equiprobabili, il risultato (es. 2.6):

$$P(E) = \frac{2(M-1)}{M} Q\Big(\sqrt{\frac{6E_b \log_2 M}{(M^2-1)N_0}}\Big) \tag{2.32}$$

Si può poi notare che in caso di errore è quasi certo che si sbagli a favore di un livello immediatamente adiacente. Conviene quindi assegnare i bit ai livelli in modo che livelli adiacenti differiscano per un solo bit, e quindi sia $n_{ij} = 1$. La cosa è possibile con il cosiddetto *mapping di Gray* (si vedano le costellazioni M-PSK di fig. 2.1 e 2.4; la costruzione è facilmente generalizzabile a 16 o più livelli). Si ottiene quindi $P_b(E) = P(E)/\log_2 M$.

Ovviamente lo stesso valore di $P_b(E)$ si ottiene nel caso di costellazioni QAM con M livelli su ciascun asse, e quindi M^2 punti, se si utilizza per entrambi gli assi il *mapping di Gray*. Per quanto riguarda $P(E)$, se dovesse interessare, si può notare che i punti interni della costellazione hanno quattro concorrenti a distanza minima, quelli lungo i bordi tre, ed i quattro vertici due.

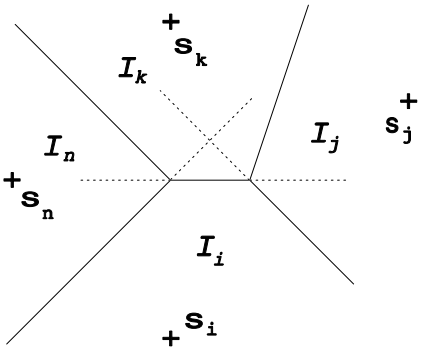

Fig. 2.3 - Regione di decisione I_i, delimitata dagli assi delle congiungenti il segnale $\mathbf{s}_i$ con i segnali $\mathbf{s}_j$, $\mathbf{s}_k$, $\mathbf{s}_n$, ecc.

Per un confronto rapido tra sistemi di trasmissione si usa trascurare i fattori moltiplicativi della funzione $Q(\cdot)$ e considerare invece dominante l'argomento. Dalla (2.32) si vede che passando dalla trasmissione binaria ($M = 2$) a quella quaternaria ($M = 4$) occorre aumentare E_b/N_0 di $10\log_{10}(5/2) = 4$ dB. Occorrono altri 4.5 dB per passare a otto livelli, e ancora 4.8 dB per sedici. Non si deve concludere affrettatamente che la trasmissione multilivello è inefficiente. Poiché si trasmettono 2, 3 e 4 bit per dimensione la banda viene ridotta di altrettante volte. Sarà chiaro nel seguito che banda ed energia possono essere barattate[12].

2.5 Calcolo approssimato della probabilità d'errore

In pochi altri casi il calcolo esatto della probabilità d'errore ha complessità accettabile. Un metodo semplice ed utile per approssimare per eccesso $P(E)$ e $P_b(E)$ è lo *union bound*, maggiorazione sostanzialmente basata sul fatto che la probabilità dell'unione di più eventi è minore o uguale alla somma delle relative probabilità.

La regione di decisione I_j relativa a $\mathbf{s}_j$ è delimitata dall'insieme dei punti,

[12] ed anzi c'è un terzo termine del baratto, la complessità; tutti i sistemi semplici, anche il binario antipodale, sono inefficienti

rette, piani, ecc. secondo il numero N di dimensioni, assi della congiungente l'estremo del vettore $\mathbf{s}_j$ con tutti gli altri concorrenti, come mostrato in due dimensioni in fig. 2.3, e in particolare con quello trasmesso $\mathbf{s}_i$. Ne deriva che la regione di decisione I_j è inclusa nel semispazio delimitato dall'asse della congiungente $\mathbf{s}_i$ con $\mathbf{s}_j$. La probabilità $P(\mathbf{s}_j/\mathbf{s}_i)$ è quindi minore o uguale alla probabilità che avendo trasmesso $\mathbf{s}_i$ il vettore ricevuto cada nel semipiano più vicino a $\mathbf{s}_j$ che ad $\mathbf{s}_i$, e questa è semplicemente la probabilità d'errore nella *trasmissione binaria* con i due soli segnali $\mathbf{s}_i$ ed $\mathbf{s}_j$ data da $Q(d_{ij}/\sqrt{2N_0})$, dove $d_{ij} = |\mathbf{s}_i - \mathbf{s}_j|$. Si ha quindi, per $P(E)$ e $P_b(E)$ rispettivamente,

$$P(E) = \sum_{i=1}^{M} P(\mathbf{s}_i) \sum_{j \neq i} P(\mathbf{s}_j/\mathbf{s}_i) \leq \sum_{i=1}^{M} P(\mathbf{s}_i) \sum_{j \neq i} Q\Big(\frac{d_{ij}}{\sqrt{2N_0}}\Big) \tag{2.33}$$

$$\begin{aligned} P_b(E) &= \frac{1}{\log_2 M} \sum_{i=1}^{M} P(\mathbf{s}_i) \sum_{j \neq i} n_{ij} P(\mathbf{s}_j/\mathbf{s}_i) \leq \\ &\leq \frac{1}{\log_2 M} \sum_{i=1}^{M} P(\mathbf{s}_i) \sum_{j \neq i} n_{ij} Q\Big(\frac{d_{ij}}{\sqrt{2N_0}}\Big) \end{aligned} \tag{2.34}$$

Si osservi che $P(\mathbf{s}_j/\mathbf{s}_i)$ e $P(\mathbf{s}_i/\mathbf{s}_j)$ possono essere diverse. Entrambe vengono maggiorate dallo *union bound* dalla stessa probabilità $Q(d_{ij}/\sqrt{2N_0})$.

Un modo alternativo di presentare lo *union bound*, che meglio rende conto del nome, è osservare che $P(E/\mathbf{s}_i)$ è la probabilità dell'unione degli eventi, non disgiunti, $|\mathbf{r} - \mathbf{s}_j| < |\mathbf{r} - \mathbf{s}_i|$ $(j \neq i)$, quindi è maggiorata dalla somma delle probabilità dei singoli eventi. Da ciò si ottiene la (2.33), ma non la (2.34). Si è quindi preferita una diversa derivazione.

Per una valutazione approssimata di $P(E)$ basta dunque conoscere l'insieme delle distanze d_{ij} tra i segnali, presi a coppie. In genere poi la funzione $Q(\cdot)$ varia così rapidamente con l'argomento che basta considerare un insieme ridotto di distanze, o addirittura se N_0 è piccolo, quindi ad alto rapporto segnale-rumore, solo la *distanza minima* tra i segnali.

La (2.33) suggerisce, per quanto possibile, di disporre i segnali in modo che l'insieme delle distanze d_{ij} tra il segnale trasmesso $\mathbf{s}_i$ e i concorrenti, o perlomeno la distanza minima, sia indipendente dal segnale trasmesso. Se

così non fosse, la probabilità d'errore $P(E)$ sarebbe largamente dominata dal caso peggiore[13].

Per una prima grossolana analisi si considera quindi solo la distanza minima tra segnali[14]. L'analisi può poi essere raffinata contando i concorrenti a distanza minima, ed eventualmente considerando concorrenti a distanza un po' maggiore della minima. Infine si può valutare n_{ij}, perlomeno per i concorrenti più vicini, e quindi $P_b(E)$.

Per usare criticamente lo *union bound* occorre aver acquisito una certa esperienza. Per E_b/N_0 molto piccolo il risultato potrebbe addirittura essere $P(E) < 10$ (certamente vero, ma poco utile). Spesso in questi casi il valore di $P(E)$ è comunque così elevato che non interessa determinarlo. Solo per valori ragionevolmente utili di $P(E)$ si ottengono risultati significativi.

Per sistemi non troppo complessi lo *union bound* risulta utilizzabile per tutte le probabilità d'errore di interesse pratico, ad esempio $P(E) < 10^{-3}$. Se il numero di segnali e di dimensioni è molto grande occorre un po' più di cautela. In ogni caso si tratta di una valutazione per eccesso, che porta ad un progetto conservativo. In qualche caso estremo, di cui si vedrà un esempio tra poco, lo *union bound* è veramente troppo pessimista.

Comunque quando si arriva al progetto vero e proprio di un sistema di trasmissione numerica si può ricorrere ad altre verifiche come il confronto con $P(E) > Q(d_{min}/\sqrt{2N_0})$ e soprattutto la *simulazione*, estremamente efficace per le alte probabilità d'errore dove lo *union bound* fallisce.

In qualche caso si maggiora la probabilità d'errore anche senza includere tutti i termini dello *union bound*. Un esempio è la modulazione a otto fasi (8PSK) con segnali disposti come in fig. 2.4, dove per semplicità non sono mostrati gli assi. Supponendo di aver trasmesso il segnale $\mathbf{s}_i$, ad esempio quello corrispondente a 000, e detti **s'** ed **s"** i due concorrenti adiacenti, corrispondenti a 001 e 100, si vede che l'unione dei due semipiani costituiti dai punti per cui $|\mathbf{r} - \mathbf{s'}| < |\mathbf{r} - \mathbf{s}_i|$ e $|\mathbf{r} - \mathbf{s"}| < |\mathbf{r} - \mathbf{s}_i|$ ricopre completamente la regione in cui si ha errore, ed anzi la regione di decisione corrispondente a 110 è contata due volte. La valutazione di $P(E/\mathbf{s}_i)$ mediante lo *union bound* non richiede quindi sette termini, ma bastano questi due. Detta E_s l'energia per simbolo si ottiene $d_{min} = 2\sqrt{E_s}\sin(\pi/8) = 0.765\sqrt{E_s} = \sqrt{0.586E_s}$ e quindi,

[13] spesso non ci si rende conto che la media tra 10^{-3} e 10^{-7} non è 10^{-5} ma $0.5 \cdot 10^{-3}$

[14] se ci fosse un solo concorrente del segnale trasmesso $\mathbf{s}_i$ alla distanza minima d_{min} si avrebbe $P(E/\mathbf{s}_i) = Q(d_{min}/\sqrt{2N_0})$; la presenza di altri segnali non può che aumentare la probabilità d'errore, sottraendo spazio alla regione di decisione corretta I_i; quindi la probabilità d'errore è compresa tra $Q(d_{min}/\sqrt{2N_0})$ e lo *union bound*

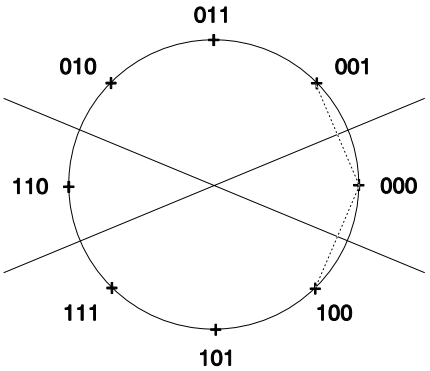

Fig. 2.4 - Costellazione 8PSK con *mapping di Gray*

poiché $E_s = 3E_b$,

$$P(E) = 2Q\Big(\sqrt{\frac{0.88E_b}{N_0}}\Big) \tag{2.35}$$

Sono assolutamente improbabili errori a favore di concorrenti non adiacenti, molto distanti, e quindi è opportuno utilizzare il *mapping di Gray* di fig. 2.4, in modo da sbagliare un solo bit su tre. Si ha quindi $P_b(E) = P(E)/3$.

Dalla (2.35) si vede che per passare dalla modulazione 4PSK alla 8PSK, con una riduzione della banda del fattore 2/3, occorre aumentare E_b/N_0 di $10\log_{10}(2/0.88) = 3.6$ dB. È un risultato piuttosto deludente.

Costellazioni 16PSK non sono praticamente mai usate. Comunque è immediato ottenere una espressione analoga alla (2.35).

2.6 Segnali ortogonali

Insiemi di segnali ortogonali sono raramente utilizzati in pratica, per motivi che saranno presto chiari, ma comunque interessanti teoricamente sia perché è possibile una valutazione esatta delle prestazioni, che consente il confronto con lo *union bound*, sia per il particolare comportamento asintotico all'aumentare del numero dei segnali.

Segnali ortogonali possono essere ottenuti in molti modi (si rivedano gli esercizi del Cap. 1). Detta E_s l'energia, gli $M = N$ segnali $\mathbf{s}_i$ sono disposti

lungo gli assi a distanza $\sqrt{E_s}$ dall'origine. Ogni segnale ha $M-1$ concorrenti, tutti alla stessa distanza $\sqrt{2E_s}$, e lo *union bound* fornisce immediatamente

$$P(E) \leq (M-1)Q\Big(\sqrt{\frac{E_s}{N_0}}\Big) = (M-1)Q\Big(\sqrt{\frac{E_b \log_2 M}{N_0}}\Big) \tag{2.36}$$

Quanto alla $P_b(E)$ si osservi che tutti gli errori possibili hanno la stessa probabilità. È dunque inutile perder tempo a studiare il *mapping*[15]. Si ottiene (es. 2.8)

$$P_b(E) \leq \frac{M}{2} Q\Big(\sqrt{\frac{E_b \log_2 M}{N_0}}\Big) \tag{2.37}$$

Per una valutazione esatta di $P(E)$, che non dipende dal segnale trasmesso, si supponga che sia stato trasmesso ad esempio $\mathbf{s}_1$. La componente r_1 ha valor medio $\sqrt{E_s}$ e tutte le altre hanno valor medio nullo. Tutte le componenti hanno varianza $N_0/2$. Si ha decisione corretta se e solo se[16] $r_j < r_1$ $(j = 2, \ldots, M)$. Quindi

$$\begin{aligned} 1 - P(E) &= 1 - P(E/\mathbf{s}_1) = P(r_2 < r_1, \ldots, r_M < r_1/\mathbf{s}_1) = \\ &= \int P(r_2 < r_1, \ldots, r_M < r_1/\mathbf{s}_1, r_1) f(r_1/\mathbf{s}_1) dr_1 = \\ &= \int \Big(1 - Q\big(\frac{r_1}{\sqrt{N_0/2}}\big)\Big)^{M-1} \frac{1}{\sqrt{\pi N_0}} \exp\Big(-\frac{(r_1 - \sqrt{E_s})^2}{N_0}\Big) dr_1 \end{aligned} \tag{2.38}$$

espressione da calcolare numericamente, ma comunque trattabile anche per M molto grande. I risultati sono mostrati in fig. 2.5, insieme allo *union bound*, per $M = 2$, 2^5, 2^{10}, 2^{20}, 2^{40} e 2^{80}.

Si osservi che le prestazioni migliorano all'aumentare di M. Lo *union bound* fornisce un'ottima approssimazione per ogni caso di possibile interesse pratico, ma si discosta sempre più dal risultato esatto all'aumentare di M. Inoltre le curve diventano sempre più ripide. Viene voglia di capire cosa succede per

[15] anche se il principiante è portato a credere, nel caso di ortogonalità ottenuta mediante separazione temporale o spettrale, che il segnale trasmesso sia meglio protetto dai segnali più "distanti" nel tempo o in frequenza che da quelli "adiacenti"

[16] sono le condizioni che definiscono la regione di decisione I_1; basta anche pensare che il ricevitore calcola le M componenti e sceglie il massimo

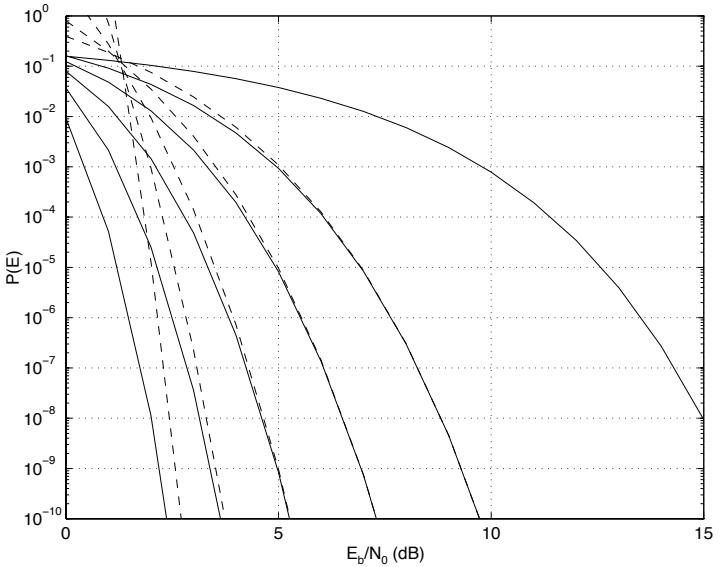

Fig. 2.5 - Probabilità d'errore per $M = 2$, 2^5, 2^{10}, 2^{20}, 2^{40} e 2^{80} segnali ortogonali (da destra a sinistra): confronto tra risultato esatto (curve continue) e *union bound* (tratteggio)

$M \to \infty$. Si può dimostrare (es. 2.2) che per $y \geq 0$ si ha $Q(y) \leq \frac{1}{2}\exp(-y^2/2)$, e quindi ponendo $M = \exp(\log M) = \exp(\log 2 \log_2 M)$ si ottiene

$$P(E) < M \exp(-\frac{E_b \log_2 M}{2N_0}) = \exp\big(-\log_2 M(\frac{E_b}{2N_0} - \log 2)\big) \qquad (2.39)$$

che tende a zero per $M \to \infty$ se $E_b/N_0 > 2\log 2$ (1.41 dB).

Per $M \to \infty$ lo *union bound* è addirittura troppo pessimista. Con argomenti più complessi si può dimostrare che perché $P(E)$ tenda a zero basta che sia $E_b/N_0 > \log 2$ (-1.59 dB). Se si confronta con $E_b/N_0 = 9.59$ dB del binario antipodale per ottenere un misero $P(E) = 10^{-5}$ il guadagno appare enorme.

Perché dunque non si usano gli ortogonali? Il numero di bit per dimensione, da cui dipende il rapporto tra ritmo di trasmissione e banda, è sconfortante: $\log_2 M/M$ tende a zero per $M \to \infty$. Gli ortogonali sono divoratori di banda, e quindi sono rarissime le occasioni per usarli.

La domanda che ci si deve porre, e a cui si risponderà nel seguito, è se lasciando crescere la complessità sia possibile trovare sistemi efficienti ma che non espandano a dismisura la banda.

I segnali ortogonali, pur nella loro inutilità pratica, hanno comunque un altro insegnamento da offrire. Per un valore prefissato di $E_b/N_0 > \log 2$ si consideri da un lato la distanza al quadrato tra i segnali pari a $d^2 = 2E_b \log_2 M$, e dall'altro il valor medio del quadrato della lunghezza del vettore rumore nelle M dimensioni, dato da $MN_0/2$. Si vede, all'aumentare di M, che $P(E)$ tende a zero mentre il rapporto tra la lunghezza del vettore rumore e la distanza tra segnali cresce all'infinito. Il rumore diventa enorme, ma fa sbagliare la decisione con probabilità decrescente! Sono i misteri della geometria in un gran numero di dimensioni, in particolare per insiemi di segnali estremamente rarefatti come gli ortogonali. Perché si commetta errore non basta affatto che la lunghezza del vettore rumore superi $d_{min}/2$; è importante anche la direzione del vettore. Solo se lo spazio dei segnali è denso, cioè se ci sono concorrenti ovunque il vettore rumore si diriga, la lunghezza del vettore rumore acquista importanza (si veda anche l'es. 2.16).

Appena un po' più interessanti degli ortogonali, nella pratica, sono i segnali *biortogonali* (N ortogonali, più gli N segnali opposti) già citati nel Cap. 1. A parità di numero M di segnali sono richieste $N = M/2$ dimensioni, e quindi l'efficienza spettrale è doppia. Il ricevitore effettua solo $M/2$ correlazioni e la probabilità d'errore è praticamente invariata (es. 2.9).

2.7 Insiemi di segnali più complessi

Dagli insiemi di segnali finora considerati si può intuire che i sistemi con maggiore efficienza spettrale richiedono valori maggiori di E_b/N_0. Per quanto riguarda gli ortogonali non si hanno elementi per dire se il miglioramento delle prestazioni all'aumentare del numero di dimensioni sia attribuibile alla sola riduzione di efficienza spettrale o non anche all'aumento di complessità. Si deve esaminare qualche insieme di segnali in cui l'aumento di complessità non si traduca in riduzione significativa del numero di bit per dimensione. L'ideale è quindi individuare, in un numero N sufficientemente grande di dimensioni, insiemi di 2^N, 2^{2N} o 2^{3N} segnali (in modo da trasmettere uno, due o tre bit per dimensione). Si vedranno alcuni semplici esempi in questa sezione.

Uno degli esempi più semplici che si possono costruire, che però non dà un

numero intero di bit per dimensione, è l'insieme di segnali in banda base[17]

$$s_i(t) = \sum_{k=1}^{9} a_k g(t - kT) \tag{2.40}$$

dove le $g(t - kT)$ sono ortogonali e $a_k = \pm 1$. Perché questa non sia la banale trasmissione successiva di nove bit con segnalazione binaria antipodale si imponga la regola che gli a_k con segno negativo devono essere in numero *pari*. La metà delle combinazioni di bit risultano vietate da questa regola (o *codice*). Ad esempio si può pensare che i primi otto a_k siano scelti secondo i bit d'informazione da trasmettere, ed il nono in modo da rispettare la parità. I segnali possibili sono $M = 2^8 = 256$, in uno spazio a 9 dimensioni. Il numero di bit per dimensione è $8/9$ e quindi l'efficienza spettrale è un po' ridotta rispetto alla trasmissione binaria antipodale.

Le coordinate del segnale $\mathbf{s}_i$ sono $\sqrt{E_g}a_{ik}$, dove E_g è l'energia di $g(t)$ e a_{ik} è il k-esimo bit dell'i-esimo segnale ($k = 1, \ldots, 9$; $i = 1, \ldots, 256$). Il quadrato della distanza tra $\mathbf{s}_i$ e $\mathbf{s}_j$ è dato da

$$d_{ij}^2 = \sum_{k=1}^{9} (s_{ik} - s_{jk})^2 = E_g \sum_{k=1}^{9} (a_{ik} - a_{jk})^2 \tag{2.41}$$

ed è evidente che solo le coordinate *diverse* danno un contributo, pari a $4E_g$. Infine il codice impone che $\mathbf{s}_i$ ed $\mathbf{s}_j$ differiscano per almeno due coordinate, e quindi la distanza minima tra i segnali è $d_{min}^2 = 8E_g$. I concorrenti a distanza minima sono $\binom{9}{2} = 36$, e analogamente si possono contare quelli a distanza maggiore. L'energia di ciascun segnale è $9E_g$. L'energia per bit d'informazione è $E_b = 9E_g/8$ e quindi si ha $d_{min}^2 = 64E_g/9$. Infine lo *union bound* dà

$$P(E) \leq \binom{9}{2} Q\Big(\sqrt{\frac{32E_b}{9N_0}}\Big) + \binom{9}{4} Q\Big(\sqrt{\frac{64E_b}{9N_0}}\Big) + \ldots \tag{2.42}$$

In modo analogo si può maggiorare $P_b(E)$. La fig. 2.6 mostra il risultato, insieme ad alcuni punti ottenuti mediante simulazione, a confronto con la segnalazione binaria antipodale. Si noti che per bassi valori di E_b/N_0 lo *union bound* è completamente inutile. Il primo termine dello *union bound* è il più importante, e per rapporto segnale-rumore molto alto il coefficiente

[17] il numero di simboli, nove per fare un esempio specifico, potrebbe essere qualsiasi

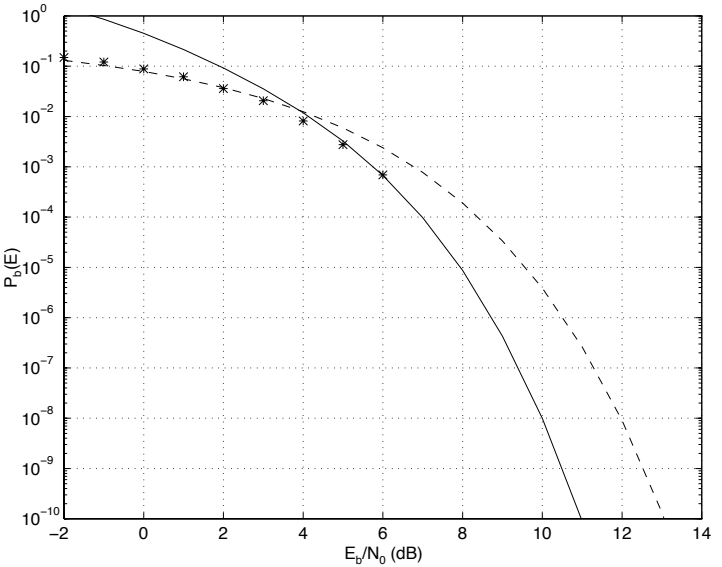

Fig. 2.6 - Probabilità d'errore per i 256 segnali in 9 dimensioni: *union bound* (curva continua) e simulazione (asterischi) a confronto con il binario antipodale (tratteggio)

moltiplicativo influisce poco, e si ha un *guadagno asintotico* rispetto al binario antipodale di $10\log_{10}(16/9) = 2.5$ dB[18]. Si noti che il guadagno reale, ad esempio a $P_b(E) = 10^{-5}$, è inferiore essendo circa 1.6 dB.

È interessante anche il ricevitore ML per questo insieme di segnali. Poiché l'energia $|\mathbf{s}_i|^2$ non dipende da i, è sufficiente determinare il segnale che rende massima la correlazione $\mathbf{r} \cdot \mathbf{s}_i$. Ignorando costanti moltiplicative inessenziali, si deve trovare il massimo di

$$\sum_{k=1}^{9} r_k a_{ik} \tag{2.43}$$

Si supponga che tutti gli r_k siano positivi. È evidente, senza calcolare tutte le 256 correlazioni, che il massimo si ottiene con tutti gli $a_k = 1$. Analogamente

[18] il lettore attento noterà che il fattore 16/9 deriva dal prodotto di due termini: 2 (pari al numero minimo di componenti diverse) e 8/9 (numero di bit per simbolo); questo risultato sarà ritrovato in generale nel Cap. 5

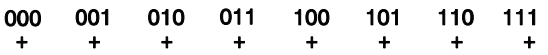

Fig. 2.7 - Costellazione 8PAM con *mapping* binario naturale

se un numero pari di r_k è negativo basta prendere i corrispondenti $a_k = -1$. Resta da esaminare il caso di un numero dispari di r_k negativi, perché non è lecito assegnare a tutti gli a_k il segno di r_k; sarebbe come dire che è stato trasmesso un segnale che non esiste. Essendo costretti ad assegnare ad almeno un a_k un segno diverso dal corrispondente r_k è evidente che lo si farà per uno solo, e solo per quello tra gli r_k minore in modulo.

Benché il ricevitore sia piuttosto semplice e si abbia effettivamente un guadagno rispetto alla segnalazione binaria antipodale, questo insieme di segnali non è usato frequentemente. Il motivo è che si può trovare di meglio, con complessità ancora accettabile.

Un altro esempio interessante è la trasmissione in banda base di otto simboli a otto livelli, con la costellazione di fig. 2.7, dove per semplicità non è mostrato l'asse; il *mapping* è *binario naturale* anziché di Gray, ma i bit indicati non sono i bit d'informazione. La forma d'onda trasmessa è

$$s_i(t) = \sum_{k=1}^{8} a_k g(t - kT) \tag{2.44}$$

dove $a_k = \pm 1, \pm 3, \pm 5, \pm 7$ e le repliche di $g(t)$ sono ortogonali. Il primo bit della rappresentazione binaria naturale di ciascun a_k sia scelto liberamente; il secondo bit rispettando una regola di parità (numero pari di uni, o di zeri); il terzo bit uguale per gli otto a_k (tutti zeri, oppure tutti uni).

Le dimensioni sono otto; e quanti sono i segnali? Si possono scegliere liberamente tutti i primi otto bit, sette dei secondi e infine uno solo dei terzi, per un totale di 16 bit d'informazione corrispondenti a $M = 2^{16}$ segnali. Si trasmettono 16/8=2 bit per dimensione, come in una modulazione non codificata a quattro livelli.

Ora la ricerca di coppie di segnali a distanza minima è un po' più difficile, soprattutto se fatta per tentativi; occorre essere sicuri di non aver trascurato qualche caso nascosto. Sia E_g l'energia della forma d'onda $g(t)$, e si considerino due segnali che differiscono solo in uno dei primi bit. Una solo delle otto coordinate è differente, e si vede che in tutti i casi la distanza al quadrato è $d^2 = 64E_g$. Si consideri poi il caso di due dei secondi bit differenti. Ora due

delle otto coordinate contribuiscono alla distanza al quadrato, ciascuna per $16E_g$, e quindi $d^2 = 32E_g$. Infine si considerino segnali che differiscono in otto componenti, a causa dell'ultimo bit diverso; si ha $d^2 = 8 \cdot 4E_g = 32E_g$. Per tutte le altre coppie di segnali la distanza è maggiore, e quindi $d^2_{min} = 32E_g$.

Si può poi verificare che gli otto livelli sono utilizzati da questo codice con pari probabilità, e quindi calcolare l'energia media per simbolo che risulta pari a $E_s = 21E_g$. Questa viene spesa per due bit d'informazione[19] e quindi $E_b = 21E_g/2$. Esprimendo infine d^2_{min} in funzione di E_b si ha $d^2_{min} = 64E_b/21$ e infine, lasciando al lettore il compito di determinare il numero di concorrenti a distanza minima,

$$P(E) \leq \cdots Q\Big(\sqrt{\frac{32E_b}{21N_0}}\Big) + \ldots \tag{2.45}$$

Non si deve confrontare questo risultato con il binario antipodale, che ha diversa efficienza spettrale, ma con la trasmissione a quattro livelli, che secondo la (2.32) ha come argomento, sotto radice, $4E_b/5N_0$. Dunque si ha un guadagno asintotico pari a $10\log_{10}\frac{32/21}{4/5} = 2.8$ dB.

Il ricevitore non è eccessivamente complicato (es. 2.10).

L'es. 2.11 propone di confrontare un analogo sistema a 16 livelli, che trasmette 3 bit per simbolo, con la trasmissione ad 8 livelli non codificata. Il guadagno asintotico è di 3 dB.

Un ultimo esempio, in banda passante. Si trasmette un segnale costituito da otto forme d'onda 8PSK

$$\begin{aligned} s_i(t) &= \mathrm{Re}\{\sum_{k=1}^{8} d_k g(t-kT)\exp(j2\pi f_0 t)\} = \\ &= \sum_{k=1}^{8} a_k g(t-kT)\cos 2\pi f_0 t - \sum_{k=1}^{8} b_k g(t-kT)\sin 2\pi f_0 t \end{aligned} \tag{2.46}$$

dove, al solito, le repliche di $g(t)$ sono ortogonali e i dati complessi d_k sono tratti dalla costellazione di fig. 2.8. Dette 0,1,...,7 le fasi possibili per d_k, le otto fasi trasmesse sono *tutte pari* oppure *tutte dispari*; inoltre la somma delle otto fasi è un *multiplo di* 4. Ad esempio le otto fasi possono essere 00000000,

[19] naturalmente si potrebbe calcolare l'energia di ciascun segnale, pari a $8E_s$ e dividere per il numero complessivo di bit, pari a 16, ottenendo lo stesso risultato

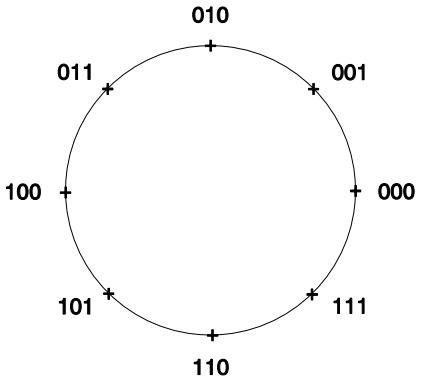

Fig. 2.8 - Costellazione 8PSK con *mapping* binario naturale

40000000, 22000000, 62000000, 11111111, 51111111, 33111111, 73111111 ma non 11110000 (alcune fasi pari, altre dispari) o 20000000 (somma non multipla di 4).

Quanti sono i segnali in questo spazio a 16 dimensioni? Un modo per calcolarlo è osservare che si può scegliere innanzitutto se usare fasi pari o dispari; poi in quattro modi ciascuna delle prime sette fasi; infine l'ultima fase deve rendere la somma pari ad un multiplo di 4, e ciò può essere ottenuto in due modi. In totale si hanno $2 \cdot 4^7 \cdot 2 = 2^{16}$ configurazioni, fra le 2^{24} che sarebbero possibili senza codice, e quindi 16/16=1 bit per dimensione, pari a 2 bit per simbolo. Il riferimento con cui confrontare è quindi la modulazione non codificata 4PSK, che ha la stessa efficienza spettrale.

Il lettore calcoli ora, ad esempio, la distanza tra segnali corrispondenti alle fasi 00000000 e 40000000. Si ottiene, se E_s è l'energia per simbolo, $d^2 = 4E_s$. Analogamente se si considerano le fasi 00000000 e 22000000 (oppure 62000000) si ottiene $d^2 = 4E_s$. Se infine si considerano 00000000 e 11111111 (oppure 77111111) si ottiene $d^2 = 4.69E_s$. L'energia spesa per bit è $E_b = E_s/2$, da cui $d^2_{min} = 8E_b$, ed infine (resta da vedere quanti sono i concorrenti, perlomeno a distanza minima)

$$P(E) \leq \cdots Q\Big(\sqrt{\frac{4E_b}{N_0}}\Big) + \ldots \tag{2.47}$$

con un guadagno asintotico di 3 dB rispetto al 4PSK. L'espressione per $P_b(E)$ è analoga. La fig. 2.9 mostra la probabilità d'errore $P_b(E)$ valutata con lo

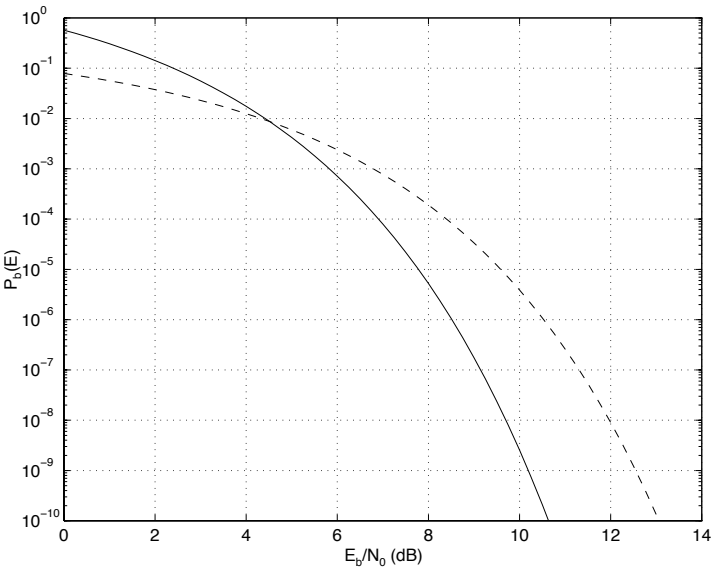

Fig. 2.9 - Probabilità d'errore per la trasmissione 8PSK codificata (*union bound* troncato; curva continua) e confronto con il 4PSK non codificato (tratteggio)

union bound (troncato; sono inclusi solo i termini con $d^2 = 4E_s$ e $4.69E_s$), e per confronto quella del 4PSK non codificato. Anche in questo caso il guadagno reale è minore di quello asintotico.

Anche in questo caso il ricevitore non è molto complicato (es. 2.13).

2.8 Considerazioni finali

Volendo riassumere il contenuto di questo capitolo, dove si sono presentati i fondamenti della trasmissione numerica, si può dire che tutta l'informazione rilevante contenuta nella forma d'onda ricevuta $r(t)$ è concentrata, nel caso di rumore gaussiano bianco nella banda dei segnali, nelle N componenti r_k che giacciono nello spazio dei segnali. Le verosimiglianze dei segnali $\mathbf{s}_i$ sono semplici funzioni del quadrato $|\mathbf{r} - \mathbf{s}_i|^2$ della distanza tra vettore ricevuto e segnali. È poi particolarmente comodo, e normalmente poco costoso, ignorare le probabilità a priori $P(\mathbf{s}_i)$ dei messaggi (ricevitore ML anziché MAP). Da

ciò deriva la struttura del ricevitore.

La probabilità d'errore è calcolabile esattamente in qualche caso semplice. La maggiorazione mediante lo *union bound* è utile nella maggior parte dei casi pratici.

I segnali ortogonali, purtroppo raramente utilizzabili, mostrano un comportamento singolare: al crescere del numero di segnali, e di dimensioni, l'andamento di $P(E)$ in funzione di E_b/N_0 diventa infinitamente ripido; si può ottenere una probabilità d'errore piccola a piacere con un valore finito, ed anche piuttosto piccolo, di E_b/N_0. Non può non nascere la curiosità di sapere se ciò sia dovuto alla banda che tende all'infinito, oppure alla complessità tendente all'infinito (o eventualmente a entrambi i motivi).

Non è poi difficile costruire esempi di segnali che consentono prestazioni migliori, a parità di efficienza spettrale, dei più semplici sistemi di trasmissione numerica in banda base o in banda passante. Ciò invita a rivolgere l'indagine, in modo più sistematico, in due direzioni: quali siano i limiti teorici insuperabili qualunque sia la complessità dell'insieme di segnali, e presumibilmente del ricevitore, e come si possano costruire sistemi pratici di complessità accettabile, avendo ben presente che questa cresce con il progredire della tecnologia. Nel seguito si darà risposta ad entrambe le questioni, subito dopo aver presentato nel prossimo capitolo qualche utile complemento alla teoria.

2.9 Esercizi

2.1 - Nel caso di trasmissione binaria antipodale con segnali di energia $E_s = E_b$ si mostri che il rapporto tra il quadrato dell'ampiezza del segnale utile all'uscita del correlatore, o del filtro adattato, e la varianza della componente del rumore, cioè il rapporto segnale-rumore all'ingresso del decisore, è $2E_b/N_0$.

2.2 - Siano x_1 e x_2 variabili casuali gaussiane indipendenti, con valor medio nullo e varianza unitaria. L'evento $A=(x_1 > y, x_2 > y)$, con $y \geq 0$, ha probabilità $Q^2(y)$ ed è incluso nell'evento $B=(x_1 > 0, x_2 > 0, x_1^2 + x_2^2 > 2y^2)$. Si calcoli, in coordinate polari, $P(B)$ e si mostri che $Q(y) \leq \frac{1}{2}\exp(-y^2/2)$.

2.3 - Dato l'insieme di segnali $\mathbf{s}_i$, aventi probabilità a priori $P(\mathbf{s}_i)$, l'energia

media trasmessa è

$$E_m = \sum_{i=1}^{M} |\mathbf{s}_i|^2 P(\mathbf{s}_i)$$

Si sottragga a tutti i segnali un vettore $\mathbf{m}$, e si mostri che l'energia media risulta minima se

$$\mathbf{m} = \sum_{i=1}^{M} \mathbf{s}_i P(\mathbf{s}_i) \quad \text{ovvero} \quad m(t) = \sum_{i=1}^{M} s_i(t) P(\mathbf{s}_i)$$

dove $\mathbf{m}$ può essere considerato il baricentro di M masse pari alle probabilità. Si mostri anche che il risparmio di energia media è pari a $|\mathbf{m}|^2$. *Suggerimento*: la derivazione rispetto al vettore $\mathbf{m}$ segue le regole formali della derivazione rispetto ad una variabile scalare.

2.4 - Si considerino, nell'intervallo di tempo $(0, T)$, le due forme d'onda equiprobabili a priori $s_i(t) = A\cos 2\pi f_i t$, con $f_2 - f_1$ multiplo di $1/2T$, e si mostri che sono ortogonali. Il baricentro dei due segnali è $\mathbf{m} = (\mathbf{s}_1 + \mathbf{s}_2)/2$, cui corrisponde la forma d'onda $m(t) = s_1(t)/2 + s_2(t)/2$. Si mostri che sottraendo tale baricentro ad entrambe le forme d'onda si ottengono segnali antipodali, con energia media dimezzata. Si mostri però che la potenza *di picco* delle forme d'onda non è variata, e si è solo ottenuto come risultato di avere forme d'onda più difficili da generare perché con inviluppo non costante.

2.5 - Si determini la soglia di decisione s del ricevitore MAP per segnali antipodali $\mathbf{s}_1$ e $\mathbf{s}_2 = -\mathbf{s}_1$ aventi probabilità P_1 e $P_2 = 1 - P_1$. Si valutino poi le probabilità d'errore $P(E/\mathbf{s}_1)$, $P(E/\mathbf{s}_2)$ e $P(E)$. *Suggerimento*: quando $r_1 = s$ le due probabilità a posteriori sono uguali. *Commento*: la probabilità d'errore *media* si riduce, rispetto al ricevitore ML, ma quella del caso peggiore *aumenta*.

2.6 - Si verifichi l'espressione (2.32) per la probabilità d'errore nella trasmissione multilivello, con M segnali equiprobabili a priori. *Suggerimento*: si ricordi che

$$\sum_{n=1}^{M/2} (2n+1)^2 = \frac{M(M^2-1)}{6}$$

2.7 - Si consideri un sistema di trasmissione in cui sette segnali sono ottenuti come nella modulazione M-PSK, con fasi equispaziate, e l'ottava forma d'onda è nulla. Si confrontino le prestazioni con la modulazione 8PSK, sia a parità di energia media sia di energia massima.

2.8 - Si mostri che in caso di errore nella trasmissione con M segnali ortogonali il numero medio di bit errati è $M \log_2 M/2(M-1)$. *Suggerimento*: ci sono $M-1$ errori possibili, equiprobabili; inoltre la somma per $j \neq i$ degli n_{ij} è data da $M \log_2 M/2$ (perché? conviene assumere che sia stato trasmesso il segnale rappresentato con $\log_2 M$ zeri).

2.9 - Per un insieme di M segnali biortogonali si determinino il ricevitore ML, la probabilità d'errore maggiorata con lo *union bound* e quella esatta, data da una espressione analoga alla (2.38).

2.10 - Si determini il ricevitore ML per il sistema di trasmissione codificato, descritto nel testo, che utilizza blocchi di otto simboli con la costellazione PAM di fig. 2.7. *Suggerimento*: si può suddividere l'insieme dei 2^{16} segnali in due gruppi di 2^{15}, con ultimi bit rispettivamente 0 e 1, e determinare *separatamente* il segnale più verosimile all'interno di ciascun gruppo. Fatto questo, resterà da eseguire un agevole confronto fra *due* sole ipotesi. I segnali di ciascun gruppo utilizzano una costellazione 4PAM (traslata). Si consideri dapprima il segnale a minima distanza *ignorando* la regola di parità sul secondo bit. Se la parità risulta rispettata non occorre altro; in caso contrario si deve cambiare il (secondo) bit corrispondente alla decisione meno affidabile.

2.11 - Si consideri un sistema di trasmissione analogo a quello dell'esercizio precedente, ma utilizzante una costellazione 16PAM con *mapping* binario naturale, in cui primo e secondo bit sono liberi, il terzo è soggetto ad una regola di parità e infine i quarti bit sono tutti uguali. Si determini il numero di segnali, il numero di bit per dimensione ed il guadagno asintotico rispetto ad una costellazione PAM non codificata di pari efficienza spettrale. *Commento*: il ricevitore ML è del tutto analogo al precedente.

2.12 - Si consideri un sistema di trasmissione analogo ai due precedenti, ma con costellazione 4PAM. Si mostri che anche in questo caso si ha guadagno asintotico rispetto al binario antipodale di pari efficienza spettrale, in termini

di energia *media* per bit d'informazione. Si mostri tuttavia che se si confrontano i *picchi* del segnale, anziché le energie medie, questo sistema perde di interesse. *Suggerimento*: il picco del segnale trasmesso dipende dalla forma d'onda in banda base $g(t)$ e dall'istante di tempo, e il massimo si ha a metà tra i simboli; per semplicità si confrontino i valori agli istanti kT.

2.13 - Si determini il ricevitore ML per il sistema di trasmissione codificato, descritto nel testo, che utilizza otto simboli 8PSK, con il *mapping* di fig. 2.8. *Suggerimento*: analogamente a precedenti esercizi si suddivida l'insieme dei segnali in due gruppi, con fasi rispettivamente pari e dispari.

2.14 - Si consideri la trasmissione di quattro simboli 4PSK, con fasi tutte pari o tutte dispari. Inoltre, numerate le fasi da 0 a 3, la somma delle quattro fasi è multipla di 4. Si determini quanti sono i segnali, eventualmente enumerando tutti i casi, e quanti bit si trasmettono per dimensione. Quali sono le prestazioni asintotiche?

2.15 - Si considerino i segnali

$$s_i(t) = \begin{cases} \displaystyle\sum_{k=1}^{4} a_k g(t-kT)\cos 2\pi f_0 t \quad \text{oppure} \\ \displaystyle\sum_{k=1}^{4} a_k g(t-kT)\sin 2\pi f_0 t \end{cases}$$

dove $a_k = \pm 1$ ed è consentito un numero pari di segni negativi. Si determini quanti sono i segnali, eventualmente enumerando tutti i casi, e quanti bit si trasmettono per dimensione. Quali sono le prestazioni asintotiche? A conti fatti si mostri che si tratta dello stesso insieme di segnali dell'esercizio precedente, descritto in modo diverso.

2.16 - (*Scherzi della geometria*) Se d è la distanza minima tra il segnale trasmesso $\mathbf{s}_i$ ed i concorrenti, la decisione è certamente corretta se la lunghezza $|\mathbf{n}|$ del vettore rumore non supera $d/2$, qualunque ne sia la direzione. Ciò corrisponde, in fig. 2.3, ad approssimare la regione di decisione I_i con la sfera ad N dimensioni di raggio $d/2$ e suggerirebbe di maggiorare la probabilità d'errore con $1 - P(|\mathbf{n}| < d/2)$. Si calcoli tale probabilità e si mostri, ad esempio con segnali ortogonali, che tale risultato è estremamente pessimista e

quindi *del tutto inutile.*
Suggerimento: sia $K_N R^N$ il volume della sfera di raggio R in N dimensioni, e quindi $NK_N R^{N-1}$ la superficie. Da ciò e dalla *ddp* congiunta di $n_1, \cdots, n_N$ si mostri che la *ddp* della variabile casuale

$$x = \sum_{k=1}^{N} n_k^2$$

è

$$f(x) = \frac{NK_N}{(\pi N_0)^{N/2}} x^{N-1} \exp(-x^2/N_0)$$

Infine si può ottenere $P(x < (d/2)^2)$ analiticamente integrando per parti, con l'aiuto di programmi di analisi simbolica come Maple e Mathematica, oppure con strumenti numerici come Matlab. È poi noto che $K_N = \pi^{N/2}/(N/2)!$ per N pari, e che $K_N = 2^N \pi^{(N-1)/2} (\frac{N-1}{2})!/N!$ per N dispari. *Commento*: in un gran numero di dimensioni quasi tutto il volume di una sfera è vicino alla superficie; contrariamente a quanto suggerisce la fig. 2.3, per forza di cose bidimensionale, le parti trascurate approssimando la regione di decisione I_i con una sfera hanno un volume *enorme*.

Capitolo 3

Trasmissione numerica: complementi

3.1 Introduzione

I principi generali per la determinazione del ricevitore ottimo nella trasmissione numerica in presenza di rumore additivo gaussiano, bianco nella banda dei segnali, sono stati esposti nel capitolo precedente. Per completezza si vedranno in questo capitolo alcuni utili complementi teorici, relativi a situazioni che si presentano con una certa frequenza.

Innanzitutto si è finora fatto conto di conoscere perfettamente funzioni base e segnali, e quindi di poter calcolare le correlazioni tra il segnale ricevuto $r(t)$ e le funzioni base $\Phi_k(t)$ oppure i segnali $s_i(t)$. Può sembrare strano, a prima vista, ritenere possibile che le correlazioni non vengano effettuate correttamente e quindi preoccuparsi di questa evenienza. Ma con un po' di riflessione non è difficile rendersi conto di dove sia nascosto almeno un modo per sbagliare una correlazione come

$$\int r(t)\Phi_k(t)dt \tag{3.1}$$

oppure

$$\int r(t)s_i(t)dt \tag{3.2}$$

Basta infatti che non sia perfettamente nota la *temporizzazione* delle forme d'onda, cioè che non vi sia perfetto accordo tra gli orologi relativi alla forma

d'onda ricevuta $r(t)$ e la funzione con cui correlare, *generata localmente*, perché si calcoli

$$\int r(t)\Phi_k(t-\tau)dt \tag{3.3}$$

oppure

$$\int r(t)s_i(t-\tau)dt \tag{3.4}$$

anziché la correlazione desiderata.

Il problema della sincronizzazione dei due orologi può essere ignorato in un primo momento nel presentare i concetti fondamentali, ma si deve poi riconoscere che la concordanza perfetta è in pratica impossibile.

Analogamente, in banda passante, non può essere perfettamente nota la fase della portante ed è quindi inevitabile una piccola rotazione degli assi.

Le correlazioni possono essere non del tutto corrette anche per altri motivi. Ad esempio si è già osservato che nel calcolo numerico di correlazioni le forme d'onda $\Phi_k(t)$ o $s_i(t)$ vengono necessariamente troncate per limitare il numero di moltiplicazioni, e che i relativi campioni sono rappresentati numericamente con precisione finita, cioè quantizzati. In qualche altro caso sono semplificazioni circuitali a suggerire di approssimare la funzione "giusta" con un'altra più semplice. Sono tutti errori che si dovranno tenere sotto controllo, ma di cui si deve essere in grado di valutare l'effetto, e tutti riconducibili al correlare con funzioni "sbagliate".

Nel valutare le conseguenze di tali errori si dovrà anche tener presente se sono commessi consapevolmente oppure no. Ad esempio è chiaro che se si calcolano le coordinate del vettore ricevuto con un sistema di assi non corretto, ma *noto*, l'errore può essere almeno in parte compensato con successive elaborazioni. Se invece l'errore non è percepito come tale, non si provvede ad alcuna correzione e quindi l'effetto generalmente è più grave.

In qualche caso, sia pur non frequente, risulta comodo utilizzare un insieme di funzioni base non ortogonali. È interessante, al di là dell'utilità pratica, vedere brevemente come si possa procedere.

Si è anche già citata nel Cap. 1 la possibilità che ad uno stesso messaggio corrispondano forme d'onda diverse, o perché non è possibile controllare perfettamente alcuni *parametri indeterminati* dell'apparecchiatura trasmittente, come frequenza, fase, ecc. o perché le deformazioni introdotte dal canale non sono prevedibili a priori. In generale il calcolo delle verosimiglianze, o delle

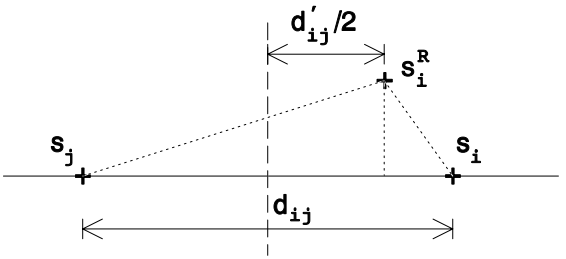

Fig. 3.1 - Geometria dei segnali in caso di segnale ricevuto (senza rumore) $\mathbf{s}_i^R$ distorto

probabilità a posteriori, diventa più complicato pur essendo basato sugli stessi concetti geometrici.

In questo capitolo si vedranno alcuni semplici esempi di ricezione in presenza di parametri indeterminati, e in particolare il caso, di un qualche interesse pratico, di sistemi di trasmissione in banda passante con fase della portante non nota (*ricezione non coerente*).

Infine verrà esaminato il caso di rumore gaussiano non bianco.

3.2 Ricezione basata su segnali approssimati

Si consideri la seguente situazione, che può presentarsi per i motivi più vari: i segnali $\mathbf{s}_i$ per i quali è progettato il ricevitore sono corretti, ma per qualche causa non sotto controllo, come distorsioni lineari o non lineari, assi ruotati a causa di un errore di fase, errori nel guadagno degli amplificatori di ricezione, ecc. il segnale ricevuto (escludendo il rumore che ad esso si somma) è $\mathbf{s}_i^R$ anziché $\mathbf{s}_i$.

Le regioni di decisione sono quindi quelle corrette, relative ai segnali $\mathbf{s}_i$, ma le distanze dal segnale effettivamente ricevuto $\mathbf{s}_i^R$ ai confini delle regioni di decisione sono diverse da quelle del caso ideale, come mostrato in fig. 3.1 in cui $\mathbf{s}_j$ è un generico concorrente del segnale trasmesso $\mathbf{s}_i$. Si approssimi, analogamente allo *union bound*, la probabilità $P(\mathbf{s}_j/\mathbf{s}_i)$ di decidere a favore di $\mathbf{s}_j$ avendo trasmesso $\mathbf{s}_i$ con la probabilità che il vettore ricevuto, somma del segnale $\mathbf{s}_i^R$ e del rumore, cada nel semipiano costituito dai punti più vicini a $\mathbf{s}_j$ che ad $\mathbf{s}_i$. Detta $d'_{ij}/2$ la distanza tra $\mathbf{s}_i^R$ e il confine tra le due regioni, si

ha

$$P(\mathbf{s}_j/\mathbf{s}_i) \le Q\Big(\frac{d'_{ij}}{\sqrt{2N_0}}\Big) \tag{3.5}$$

Resta da calcolare la distanza modificata d'_{ij}. Detta d_{ij} la distanza tra $\mathbf{s}_i$ ed $\mathbf{s}_j$, dalla fig. 3.1 si può vedere che

$$|\mathbf{s}_i^R - \mathbf{s}_j|^2 - |\mathbf{s}_i^R - \mathbf{s}_i|^2 = \Big(\frac{d_{ij}}{2} + \frac{d'_{ij}}{2}\Big)^2 - \Big(\frac{d_{ij}}{2} - \frac{d'_{ij}}{2}\Big)^2 = d_{ij}d'_{ij} \tag{3.6}$$

da cui si ottiene la formula di uso generale

$$d'_{ij} = \frac{|\mathbf{s}_i^R - \mathbf{s}_j|^2 - |\mathbf{s}_i^R - \mathbf{s}_i|^2}{d_{ij}} \tag{3.7}$$

Naturalmente in casi semplici si riesce a calcolare d'_{ij} anche senza ricorrere alla (3.7). Si noti anche che in generale $d'_{ij} \neq d'_{ji}$, perché quest'ultima distanza dipende dalla posizione di $\mathbf{s}_j^R$. La (3.7) insieme al solito *union bound* consente di maggiorare le probabilità $P(\mathbf{s}_j/\mathbf{s}_i)$ e quindi $P(E)$ o $P_b(E)$.

Una situazione analoga si ha se si ricevono, ignorando il rumore, i segnali $\mathbf{s}_i$ indistorti mentre il ricevitore è progettato per un insieme di segnali $\mathbf{s}'_i$ non corretto; è ciò che avviene ad esempio se si correla con funzioni sbagliate. Naturalmente in questo caso nell'applicare la (3.7) il vettore $\mathbf{s}_i^R$ va sostituito con $\mathbf{s}_i$, ed i vettori $\mathbf{s}_i$ ed $\mathbf{s}_j$ con $\mathbf{s}'_i$ ed $\mathbf{s}'_j$.

Come semplice esempio si consideri una modulazione 4PSK ed un errore di fase ε negli assi: il segnale trasmesso è

$$a\ g(t)\cos 2\pi f_0 t - b\ g(t)\sin 2\pi f_0 t \tag{3.8}$$

ma in ricezione si correla con $g(t)\cos(2\pi f_0 t + \varepsilon)$ e con $-g(t)\sin(2\pi f_0 t + \varepsilon)$. È quasi immediato verificare, a dire il vero anche senza la (3.7), che le distanze $\sqrt{2E_s} = 2\sqrt{E_b}$ del caso ideale vengono moltiplicate per $\cos\varepsilon \pm \sin\varepsilon$, che può essere approssimato con $1 \pm \varepsilon$ se $\varepsilon \ll 1$. La probabilità d'errore passa da $Q(\sqrt{2E_b/N_0})$ a (es. 3.2)

$$P(E) = \frac{1}{2}Q\Big(\sqrt{\frac{2E_b(1-\varepsilon)}{N_0}}\Big) + \frac{1}{2}Q\Big(\sqrt{\frac{2E_b(1+\varepsilon)}{N_0}}\Big) \tag{3.9}$$

Si noti che se l'errore di fase ε fosse noto sarebbe banale correggerlo moltiplicando il vettore ricevuto, rappresentato con un numero complesso, per

$\exp(j\varepsilon)$. Ciò otterrebbe l'effetto di ruotare gli assi riportandoli nella posizione corretta.

Come esempio un po' più complesso si consideri la trasmissione di una coppia di bit mediante la forma d'onda in banda base

$$s_i(t) = a_1 g(t-T) + a_2 g(t-2T) \tag{3.10}$$

dove $a_1, a_2 = \pm 1$, e si supponga che in ricezione a causa di un errore τ di temporizzazione si calcolino le correlazioni con $g(t-T-\tau)$ e $g(t-2T-\tau)$. Il calcolo delle distanze modificate è lasciato al lettore (es. 3.4). È tuttavia interessante osservare che, contrariamente al caso precedente, il piano dove giacciono i quattro segnali $\mathbf{s}_i$ non coincide con il piano formato dai due assi utilizzati in ricezione. Quindi anche se l'errore di temporizzazione τ fosse noto non sarebbe possibile calcolare *esattamente* le coordinate corrette da quelle sbagliate, cioè compensare l'errore τ *dopo* averlo commesso. Naturalmente se τ è piccolo la degradazione è comunque modesta.

In pratica trasmettere solo due bit, come nella (3.10), non ha senso. Il lettore può provare a rifare il calcolo nel caso di trasmissione di una lunga sequenza di bit.

3.3 Assi non ortogonali

In qualche raro caso risulta più comodo calcolare le componenti del vettore ricevuto $\mathbf{r}$ e le correlazioni $\mathbf{r} \cdot \mathbf{s}_i$ facendo uso di un insieme di funzioni base non ortogonali.

È istruttivo vedere, sia pur rapidamente, come si deve operare. Siano $v_k(t)$ le N funzioni base normalizzate ma non ortogonali che si vogliono utilizzare, e $\mathbf{v}_k$ i vettori corrispondenti. Sia poi A la matrice di dimensione $N \cdot N$ che ha come colonne le coordinate dei versori $\mathbf{v}_k$ $(k = 1, \ldots, N)$ rispetto ad un riferimento *ortogonale* prefissato (ma che non occorre specificare, come si vedrà tra breve). Siano r ed s_i vettori colonna di N elementi formati dalle coordinate del vettore ricevuto e dell'i-esimo segnale rispetto agli stessi assi ortogonali. Per non dover calcolare esplicitamente queste coordinate ortogonali si può osservare che il prodotto scalare $\mathbf{r} \cdot \mathbf{s}_i$ è dato dal prodotto $r^T s_i$ dove r^T è il vettore riga trasposto di r, e che può essere espresso anche come

$$\mathbf{r} \cdot \mathbf{s}_i = r^T s_i = r^T A A^{-1} s_i = (r^T A)(A^{-1} s_i) \tag{3.11}$$

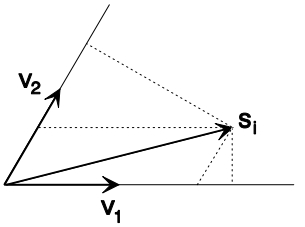

Fig. 3.2 - Componenti controvarianti (proiezioni ortogonali agli assi) e covarianti (proiezioni parallele agli assi) del segnale $\mathbf{s}_i$, in riferimento ad assi non ortogonali

dove A^{-1} è la matrice inversa di A. Ciò implicitamente richiede che la matrice A non sia singolare cioè che le colonne di A, e quindi gli N versori non ortogonali, siano linearmente indipendenti.

Ora è facile dare un semplice significato agli N elementi di $r^T A$: si tratta infatti dei prodotti scalari, calcolati mediante le componenti ortogonali, tra il vettore ricevuto $\mathbf{r}$ ed i versori non ortogonali $\mathbf{v}_k$. Questi prodotti scalari possono anche essere calcolati *non geometricamente* mediante gli integrali $\int r(t)v_k(t)dt$ esattamente come si farebbe per una base ortogonale.

Le componenti di $A^{-1}s_i$ possono essere *precalcolate* per ogni segnale e memorizzate. Il procedimento, un po' involuto, consiste nello scegliere una base ortogonale e calcolare rispetto ad essa le componenti dei segnali $\mathbf{s}_i$ e dei versori $\mathbf{v}_k$, e quindi la matrice A, determinare poi la matrice inversa A^{-1} e infine calcolare gli elementi dei vettori $A^{-1}s_i$. È però interessante osservare che una base ortogonale non è strettamente necessaria. Infatti si ha

$$A^{-1}s_i = A^{-1}(A^T)^{-1}A^T s_i = (A^T A)^{-1}A^T s_i = (A^T A)^{-1}(s_i^T A)^T \qquad (3.12)$$

Ora si osservi che $A^T A$ ha come elementi di indice k, n i prodotti scalari tra i versori $\mathbf{v}_k$ e $\mathbf{v}_n$, calcolabili come integrali nel tempo senza ricorrere ad una esplicita base ortogonale. Occorre poi calcolare, una volta per tutte, la matrice inversa $(A^T A)^{-1}$. I vettori $s_i^T A$, del tutto analoghi a $r^T A$, hanno come elementi i prodotti scalari tra $\mathbf{s}_i$ e i versori $\mathbf{v}_k$, anch'essi esprimibili come integrali nel tempo. In conclusione non occorre aver scelto una base ortogonale di servizio.

In riferimento agli assi $\mathbf{v}_k$ non ortogonali, le componenti di $r^T A$ e $s_i^T A$ sono dette *controvarianti*. Si tratta dei prodotti scalari $\mathbf{r} \cdot \mathbf{v}_k$ (o $\mathbf{s}_i \cdot \mathbf{v}_k$), e quindi

delle proiezioni *ortogonali* del vettore sul k-esimo asse (fig. 3.2).

Le componenti $A^{-1}s_i$ sono invece dette *covarianti*. Il lettore che avesse la curiosità di dare anche a queste un significato geometrico potrebbe osservare innanzitutto che $AA^{-1}s_i = s_i$. Inoltre $A(A^{-1}s_i)$ è un vettore colonna, dato dalla somma delle colonne di A con pesi pari alle componenti covarianti $A^{-1}s_i$. Poiché le colonne di A corrispondono, nel sistema di assi ortogonali, ai versori $\mathbf{v}_k$ si tratta della scomposizione del vettore $\mathbf{s}_i$ in somma dei versori $\mathbf{v}_k$ con pesi pari alle componenti covarianti. Le componenti covarianti quindi non sono altro che le componenti cartesiane in coordinate non ortogonali, cioè le proiezioni *parallele* agli assi (fig. 3.2).

Per riassumere, i prodotti scalari in uno spazio con versori non ortogonali si calcolano come somme di prodotti tra componenti controvarianti di uno dei due vettori e componenti covarianti dell'altro. Le componenti controvarianti sono date dagli usuali prodotti scalari tra vettori e versori, e quindi si calcolano con correlazioni tra funzioni del tempo. Le componenti covarianti, pur avendo un evidente significato geometrico, non si calcolano altrettanto facilmente. Spesso è preferibile determinare le componenti controvarianti e moltiplicarle per $(A^TA)^{-1}$. Infine in coordinate ortogonali componenti controvarianti e covarianti coincidono, e A^TA è la matrice identità $N \cdot N$. Le coordinate ortogonali sono dunque ben più comode di quelle oblique, a cui si ricorre solo in casi speciali.

3.4 Parametri indeterminati

Si supponga che all'i-esimo messaggio non sia associata un'unica forma d'onda $s_i(t)$ ma l'insieme di forme d'onda $s_i(t, \vartheta)$, dove ϑ è un parametro non noto a priori. Ad esempio potrebbero non essere ben note l'ampiezza o la posizione temporale della forma d'onda, o anche, in banda passante, la frequenza o la fase della portante.

Potrebbero, in generale, essere indeterminati più parametri $\vartheta_1, \vartheta_2, \ldots$ ma l'estensione a questo caso è immediata, per cui nel seguito si considererà un solo parametro.

La difficoltà nasce dal fatto che il vettore $\mathbf{s}_i$ non è noto a priori. Se ha senso attribuire una densità di probabilità a priori $f(\vartheta)$ al parametro indeterminato ϑ, si può dapprima calcolare la verosimiglianza condizionata ad un valore del parametro ϑ. Detto $\mathbf{s}_i(\vartheta)$ il vettore che corrisponde alla forma d'onda $s_i(t, \vartheta)$,

si ha

$$f(\mathbf{r}/\mathbf{s}_i(\vartheta)) \equiv \exp\Big(-\frac{1}{N_0}|\mathbf{r}-\mathbf{s}_i(\vartheta)|^2\Big) \equiv$$
$$\equiv \exp\Big(\frac{2}{N_0}\mathbf{r}\cdot\mathbf{s}_i(\vartheta)\Big)\exp\Big(-\frac{1}{N_0}|\mathbf{s}_i(\vartheta)|^2\Big) \tag{3.13}$$

da cui si ottiene che la verosimiglianza, non condizionata, è proporzionale a

$$\int \exp\Big(-\frac{1}{N_0}|\mathbf{r}-\mathbf{s}_i(\vartheta)|^2\Big)f(\vartheta)d\vartheta \equiv$$
$$\equiv \int \exp\Big(\frac{2}{N_0}\mathbf{r}\cdot\mathbf{s}_i(\vartheta)\Big)\exp\Big(-\frac{1}{N_0}|\mathbf{s}_i(\vartheta)|^2\Big)f(\vartheta)d\vartheta \tag{3.14}$$

Si deve poi trovare il massimo rispetto all'indice i di tale verosimiglianza[1]. Chiaramente non si tratta più di determinare fra i possibili segnali $\mathbf{s}_i$ quello più vicino al vettore ricevuto $\mathbf{r}$, operazione priva di senso perché non c'è un unico vettore $\mathbf{s}_i$ ma piuttosto un insieme di vettori a cui sono state assegnate delle probabilità a priori in accordo con la *ddp* del parametro.

In generale la ricerca del massimo è faticosa, ed in certi casi è anche imbarazzante assegnare una *ddp* a priori al parametro[2]. Spesso si risolve il dilemma con una procedura *ad hoc*, consistente nel determinare la *coppia* i, ϑ che dà il massimo della verosimiglianza condizionata $f(\mathbf{r}/\mathbf{s}_i(\vartheta))$ oppure della funzione integranda nella (3.14), e quindi nel prendere come decisione il valore di i così ottenuto (scartando ϑ, che non interessa). La procedura ha il merito di essere assai più semplice; se ad esempio si ignora la *ddp* a priori $f(\vartheta)$ basta infatti individuare tra tutti i possibili vettori $\mathbf{s}_i(\vartheta)$ quello più vicino ad $\mathbf{r}$, e scegliere il valore di i corrispondente. È poi confortante scoprire che in alcuni casi particolari la decisione così ottenuta coincide con quella fornita dalla (3.14). Si osservi infine che nel caso di segnali di uguale energia basta cercare il massimo delle correlazioni $\mathbf{r}\cdot\mathbf{s}_i(\vartheta)$.

[1] eventualmente moltiplicata per la probabilità a priori dell'i-esimo messaggio se si preferisce il ricevitore MAP

[2] se non si sa assolutamente nulla della fase della portante è ragionevole supporla uniformemente distribuita tra 0 e 2π; ma se può esservi un errore imprecisato di frequenza lo supporremo uniformemente distribuito o no? e fra quali valori estremi?

3.5 Ricezione non coerente

Si consideri la trasmissione di uno tra i due segnali passa banda

$$s_i(t) = \mathrm{Re}\{z_i(t)\exp(j2\pi f_0 t + j\vartheta)\} \qquad (i = 1, 2) \tag{3.15}$$

con equivalenti passa basso $z_i(t)\exp(j\vartheta)$, dove la fase ϑ è sconosciuta per cui si può supporre uniformemente distribuita tra 0 e 2π. Le verosimiglianze condizionate $f(\mathbf{r}/\mathbf{s}_i(\vartheta))$ sono date dalla (3.13), e quindi si deve cercare il massimo di

$$2\mathbf{r}\cdot\mathbf{s}_i(\vartheta) - |\mathbf{s}_i(\vartheta)|^2 = \mathrm{Re}\{\int z(t)z_i^*(t)dt\exp(-j\vartheta)\} - \frac{1}{2}\int |z_i(t)|^2 dt \tag{3.16}$$

Lo spazio ha quattro dimensioni, anziché le due che basterebbero se ϑ fosse noto, e le due regioni di decisione non sono facilmente visualizzabili. Anche il calcolo della probabilità d'errore non è affatto banale. L'unico caso relativamente semplice è quello in cui i segnali sono ortogonali ed hanno la stessa energia. In tal caso si possono prendere come funzioni base i segnali stessi, normalizzati, e i corrispondenti in quadratura; inoltre nella (3.16) si può ignorare il termine $|\mathbf{s}_i(\vartheta)|^2$.

È poi evidente che la probabilità d'errore non dipende dalla fase effettiva del segnale trasmesso[3] ϑ, per cui si può assumere $\vartheta = 0$. Sia E_s l'energia di ciascun segnale. Indicando con $x + jy$ le componenti rispetto alle prime due funzioni base, e con $u + jv$ le altre due, e supponendo di aver trasmesso il segnale $s_1(t)$ si ha $E[x] = \sqrt{E_s}$ ed $E[y] = E[u] = E[v] = 0$. Le componenti sono variabili casuali indipendenti, ed hanno varianza $N_0/2$. Per risparmiare un po' di fatica si ponga $N_0/2 = 1$.

Il massimo rispetto a ϑ di $2\mathbf{r}\cdot\mathbf{s}_i(\vartheta)$ è evidentemente pari al *modulo* di $\int z(t)z_i^*(t)dt$. Quindi la decisione è basata sul confronto tra i moduli di $x+jy$ e $u+jv$, e si ha errore se $u^2+v^2 > x^2+y^2$. Pur con *due* soli segnali si noti quanto è complesso il confine tra le regioni di decisione. Fissati x ed y, si ha errore con probabilità pari all'integrale della *ddp* $f(u)f(v)$ nella regione $u^2+v^2 > x^2+y^2$. Il calcolo è immediato se eseguito in coordinate polari, ed il risultato è

$$P(E/x,y) = \exp(-\frac{x^2+y^2}{2}) \tag{3.17}$$

[3] chi non ci credesse ripeta il calcolo per un valore generico di ϑ

Si ottiene infine, con qualche calcolo[4],

$$P(E) = \int\int P(E/x,y)f(x)f(y)dxdy = \int\int \exp(-\frac{x^2}{2})\exp(-\frac{y^2}{2}) \\ \frac{1}{2\pi}\exp(-\frac{(x-\sqrt{E_s})^2}{2})\exp(-\frac{y^2}{2})dxdy = \frac{1}{2}\exp\big(-\frac{E_s}{4}\big) \tag{3.18}$$

Infine, ricordando che si era posto $N_0 = 2$,

$$P(E) = \frac{1}{2}\exp\big(-\frac{E_s}{2N_0}\big) \tag{3.19}$$

Può essere interessante osservare che se la fase ϑ fosse nota, e quindi il ricevitore coerente, la probabilità d'errore sarebbe data da

$$P(E) = Q\big(\sqrt{\frac{E_s}{N_0}}\big) \approx \frac{1}{\sqrt{2\pi E_s/N_0}}\exp\big(-\frac{E_s}{2N_0}\big) \tag{3.20}$$

Le prestazioni del ricevitore non coerente sono quindi *in questo caso* non lontane da quelle del ricevitore coerente. Ad esempio se si vuol ottenere $P(E) = 10^{-5}$ i valori di E_s/N_0 sono rispettivamente 13.35 e 12.6 dB, ed asintoticamente la differenza si annulla. Tuttavia si è già osservato che le semplici modulazioni binarie sono troppo inefficienti, in particolare per basse probabilità d'errore.

In generale la probabilità d'errore nel caso di ricezione non coerente di due segnali non ortogonali, o anche solo con energie diverse, è data da un'espressione piuttosto complicata che qui non viene riportata. Conviene tuttavia osservare che le prestazioni asintotiche, per alto rapporto segnale-rumore, dipendono dalla distanza minima tra segnali concorrenti anche se le regioni di decisione non sono delimitate da piani. Supponendo ad esempio di aver trasmesso il segnale con fase nulla $\mathbf{s}_1(0)$ si devono considerare come concorrenti tutti i vettori $\mathbf{s}_2(\vartheta)$, essendo lecita ogni fase. La minima di queste

[4] si ricordi che $\int \exp(-(z-a)^2/2\sigma^2)dz = \sqrt{2\pi}\sigma$; basta invocare tale risultato dopo aver spezzato nel prodotto di due integrali e aver completato i quadrati moltiplicando e dividendo per opportuni esponenziali

distanze determina le prestazioni asintotiche. Si ha

$$d^2(\vartheta) = |\mathbf{s}_1(0) - \mathbf{s}_2(\vartheta)|^2 = |\mathbf{s}_1|^2 + |\mathbf{s}_2|^2 - 2\mathbf{s}_1(0)\cdot\mathbf{s}_2(\vartheta) = \\ = \frac{1}{2}\int |z_1(t)|^2 dt + \frac{1}{2}\int |z_2(t)|^2 dt - \mathrm{Re}\{\int z_1(t) z_2^*(t)\exp(-j\vartheta)dt\} \tag{3.21}$$

Il massimo rispetto a ϑ della parte reale dell'ultimo termine coincide con il modulo, e quindi la distanza minima è data da

$$d^2_{min} = \frac{1}{2}\int |z_1(t)|^2 dt + \frac{1}{2}\int |z_2(t)|^2 dt - |\int z_1(t) z_2^*(t) dt\,| \tag{3.22}$$

Si noterà che nel caso di segnali ortogonali tale distanza minima è la stessa del caso coerente, essendo nullo il terzo termine. È però minore in tutti gli altri casi.

3.6 Demodulazione differenziale

Si consideri, come esempio di ricezione non coerente, la trasmissione di una sequenza di simboli M-PSK

$$\mathrm{Re}\{\sum d_k g(t - kT)\exp(j2\pi f_0 t + j\vartheta)\} \tag{3.23}$$

dove i valori possibili per i dati d_k sono $\exp(j2\pi/M)$ e la fase ϑ è sconosciuta, per cui si suppone $f(\vartheta)$ uniforme nell'intervallo da 0 a 2π. Poiché non si conosce la fase assoluta, si associa l'informazione da trasmettere alle *differenze* di fase tra simboli successivi (*codifica differenziale*[5]). In genere si rinuncia a determinare la fase ϑ, e quindi la si considera sconosciuta, quando essa fluttua troppo rapidamente e rende difficoltosa la sincronizzazione di portante. Ciò può accadere a causa di forti instabilità degli oscillatori, o più spesso in caso di canali variabili nel tempo. Normalmente in tali situazioni si assume che la fase incognita ϑ sia costante nell'intervallo di tempo occupato dalle forme d'onda usate per trasmettere due dati consecutivi d_{k-1} e d_k, e si impone a

[5]la codifica differenziale dei dati può essere usata, per altri motivi, anche in caso di ricezione coerente; infatti ad esempio nelle modulazioni 4PSK e QAM il recupero della portante soffre di una ambiguità di multipli di $\pi/2$; questa può essere risolta con provvedimenti *ad hoc* oppure codificando i dati in modo differenziale; unico inconveniente, sopportabile, è che quando si sbaglia una decisione questa ha effetto su due coppie consecutive di simboli e quindi la probabilità d'errore raddoppia

priori di utilizzare per la decisione solo le corrispondenti coordinate del vettore ricevuto[6]. Ciò è come dire che si opera come se si fosse trasmesso il segnale

$$\mathrm{Re}\{(d_{k-1}g(t-(k-1)T)+d_k g(t-kT))\exp(j2\pi f_0 t+j\vartheta)\} \tag{3.24}$$

Avendo i segnali pari energia, la verosimiglianza è data da

$$f(\mathbf{r}/\mathbf{s}_i(\vartheta)) \equiv \exp\Big(\frac{2}{N_0}\mathbf{r}\cdot\mathbf{s}_i(\vartheta)\Big) \tag{3.25}$$

e quindi basta cercare il massimo di

$$\mathbf{r}\cdot\mathbf{s}_i(\vartheta) = \mathrm{Re}\{(r_{k-1}d_{k-1}^* + r_k d_k^*)\exp(-j\vartheta)\} \tag{3.26}$$

Il massimo rispetto a ϑ è ovviamente dato dal modulo del numero complesso, e quindi si deve cercare il massimo di $|r_{k-1}d_{k-1}^* + r_k d_k^*|$. Il corrispondente valore di d_k/d_{k-1}, cioè la più verosimile differenza tra le fasi trasmesse da cui poi si determinano i bit d'informazione decisi, corrisponde al miglior allineamento possibile tra i vettori $r_{k-1}d_{k-1}^*$ e $r_k d_k^*$.

Non è difficile mostrare (es. 3.11) che la regola di decisione può essere espressa anche nel seguente modo: si determina in quale regione di decisione della usuale modulazione M-PSK *coerente* cade il numero complesso $r_k r_{k-1}^*$. L'esempio più semplice è quello della modulazione binaria 2PSK, per cui la decisione è semplicemente basata sul segno di $\mathrm{Re}\{r_k r_{k-1}^*\}$.

Se le fluttuazioni di ϑ non sono troppo rapide si può ritenere la fase costante su più di due simboli. Ad esempio per tre simboli si deve cercare il massimo (ma non è altrettanto facile, in pratica) di $|r_{k-2}d_{k-2}^* + r_{k-1}d_{k-1}^* + r_k d_k^*|$. Anche senza far calcoli si intuisce che le prestazioni sono migliori, per cui si potrebbe pensare di aumentare ulteriormente il numero di campioni r_k considerati (es. 3.13). Ma ovviamente il semplice modello di una fase *sconosciuta ma costante* diventa sempre meno accurato, ed anzi a rigore non è mai corretto. Tornando al caso di due soli campioni, se si suppone che la fase sia ϑ nel primo simbolo ed abbia un valore leggermente diverso $\vartheta+\varphi$ nel secondo, si deve cercare il massimo rispetto ai dati e alle due fasi ϑ e φ di

$$f(\mathbf{r}/\mathbf{s}_i(\vartheta,\varphi))f(\vartheta,\varphi) \equiv \exp\Big(\frac{2}{N_0}\mathbf{r}\cdot\mathbf{s}_i(\vartheta,\varphi)\Big)f(\varphi) \tag{3.27}$$

[6] strategia criticabile perché se la fase è costante su due intervalli di simbolo lo sarà presumibilmente anche in alcuni altri adiacenti, che sarebbe meglio utilizzare; se invece varia così rapidamente da non essere costante al di fuori, non lo sarà neppure in due consecutivi e quindi il modello non è corretto

dove si è assunto ϑ uniforme e φ indipendente da ϑ. Se $f(\varphi)$ è funzione decrescente del modulo dell'argomento si trova, con un qualche ragionamento, che la regola di decisione non cambia rispetto al caso $\varphi = 0$. Peggiorano però le prestazioni, come si discuterà nel seguito.

Analogamente nel caso di più di due campioni si dovrebbero considerare le variazioni di fase, ovviamente correlate, tra primo e secondo simbolo, primo e terzo, e così via. La complessità aumenta notevolmente, ma d'altra parte questo sembra l'unico modo corretto per ottenere il massimo delle prestazioni.

Il calcolo della probabilità d'errore nella demodulazione differenziale è semplice solo nel caso di modulazione 2PSK, quando si osservino solo due campioni e si possa ritenere costante la fase ϑ. Infatti è immediato riconoscere che i due segnali (3.24) ottenuti con $d_{k-1} = 1; d_k = 1$ e $d_{k-1} = 1; d_k = -1$ sono ortogonali ed hanno energia $2E_b$ se E_b è l'energia per simbolo, e quindi anche per bit d'informazione. Dalla (3.19) si ha quindi immediatamente

$$P(E) = \frac{1}{2} \exp\Big(- \frac{E_b}{N_0}\Big) \tag{3.28}$$

Può essere interessante osservare che se la fase ϑ fosse nota, e quindi il ricevitore coerente, la probabilità d'errore sarebbe data da

$$P(E) = Q\Big(\sqrt{\frac{2E_b}{N_0}}\Big) \approx \frac{1}{\sqrt{4\pi E_b/N_0}} \exp\Big(- \frac{E_b}{N_0}\Big) \tag{3.29}$$

Le prestazioni asintotiche del ricevitore non coerente sono quindi prossime a quelle del ricevitore coerente. In pratica tuttavia le occasioni per utilizzare la modulazione 2PSK sono piuttosto rare a causa dello spreco di banda.

Se si considera la ricezione differenziale nel caso 4PSK i segnali sono ancora dati dalla (3.24), ma i dati d_{k-1} e d_k hanno quattro valori: $\pm 1, \pm j$, e ad esempio i segnali corrispondenti alle coppie $d_{k-1} = 1, d_k = 1$ e $d_{k-1} = 1, d_k = j$ non sono ortogonali. Le regioni di decisione sono piuttosto complesse. Supponendo $d_{k-1} = d_k = 1$ e $\vartheta = 0$ il segnale trasmesso ha componenti $\sqrt{E_s}, 0, \sqrt{E_s}, 0$. Un concorrente ha componenti $\sqrt{E_s}\cos\vartheta, -\sqrt{E_s}\sin\vartheta, \sqrt{E_s}\sin\vartheta, \sqrt{E_s}\cos\vartheta$. La distanza dipende da ϑ, e le prestazioni asintotiche dipendono dalla minima distanza possibile. Questa risulta pari a $d^2 = 2(2-\sqrt{2})E_s = 4(2-\sqrt{2})E_b$, come il lettore può verificare, mentre nel caso coerente è $4E_b$. La perdita asintotica è quindi tutt'altro che trascurabile, essendo data da $10\log_{10} 1/(2-\sqrt{2}) = 2.3$ dB.

Se invece la fase non si mantiene costante dal primo al secondo simbolo la distanza minima tra concorrenti si riduce ulteriormente, ed è persino possibile

che diventi nulla. In tal caso la degradazione asintotica è *infinita*. Se infatti si calcola, o si simula, la probabilità di errore in funzione di E_b/N_0 si scopre che essa non scende al di sotto di un livello minimo (*floor*), cioè che c'è una probabilità non nulla di errore anche in assenza di rumore. Si tratta appunto della probabilità che segnali corrispondenti a *messaggi diversi* coincidano.

Un ultimo caso molto interessante, ma che non si analizzerà, è quello in cui il guadagno (complesso) di un canale a radiofrequenza varia nel tempo. Assumendo che esso resti praticamente costante nell'intervallo occupato da un simbolo, le componenti del vettore ricevuto sono date da[7]

$$r_k = c_k s_{ik} + n_k \tag{3.30}$$

dove n_k sono i campioni complessi del rumore e c_k è una successione di variabili casuali complesse di cui è nota la funzione di autocorrelazione. In particolare se il segnale viene ricevuto come somma di molti contributi provenienti da cammini diversi (canale *multipath*) con ampiezze e fasi casuali, come è tipico ad esempio della propagazione nel canale radiomobile terrestre, c_k è una sequenza di variabili casuali gaussiane complesse, a valor medio nullo, con funzione di autocorrelazione determinabile in base alle caratteristiche fisiche della propagazione. Se esiste anche un percorso dominante in visibilità si può attribuire alla parte reale di c_k un valor medio non nullo.

Il compito del ricevitore è determinare la sequenza di dati più verosimile, tenendo conto della densità di probabilità congiunta dei guadagni c_k. Poiché questa è gaussiana si comprende come vi sia qualche speranza di poter affrontare questo interessante problema, comunque tutt'altro che semplice.

3.7 Ricezione in diversità

Si supponga di ricevere due repliche indipendenti $r_1(t) = A_1 s_i(t) + n_1(t)$ ed $r_2(t) = A_2 s_i(t) + n_2(t)$ del segnale trasmesso $s_i(t)$, con ampiezze A_1 ed A_2 note, sovrapposte a realizzazioni indipendenti $n_1(t)$ ed $n_2(t)$ del rumore. Il rumore sia gaussiano e bianco nella banda dei segnali, ed abbia densità spettrale bilatera $N_0/2$. Il calcolo di $P(\mathbf{s}_i/r_1(t), r_2(t))$ procede al solito modo con la regola di Bayes, e ad esempio nel caso di segnali con pari energia la

[7] se le caratteristiche del canale variano rapidamente rispetto all'intervallo tra i simboli il problema, già non semplice, si complica notevolmente

verosimiglianza $f(\mathbf{r}_1, \mathbf{r}_2/\mathbf{s}_i)$ è data da

$$f(\mathbf{r}_1, \mathbf{r}_2/\mathbf{s}_i) \equiv \exp\Big(-\frac{1}{N_0}|\mathbf{r}_1 - A_1\mathbf{s}_i|^2\Big)\exp\Big(-\frac{1}{N_0}|\mathbf{r}_2 - A_2\mathbf{s}_i|^2\Big) \equiv$$
$$\equiv \exp\Big(\frac{2}{N_0}(\mathbf{r}_1 \cdot A_1\mathbf{s}_i + \mathbf{r}_2 \cdot A_2\mathbf{s}_i)\Big) \equiv \exp\Big(\frac{2}{N_0}(A_1\mathbf{r}_1 + A_2\mathbf{r}_2)\cdot\mathbf{s}_i\Big) \tag{3.31}$$

Dunque basta combinare i segnali $r_1(t)$ ed $r_2(t)$ con pesi proporzionali alle ampiezze A_1 ed A_2, e procedere come al solito. Si può poi mostrare che ai fini della probabilità d'errore le energia delle due repliche del segnale contenute in $r_1(t)$ ed $r_2(t)$ si sommano. Questa tecnica presuppone la conoscenza di A_1 ed A_2, che sono numeri complessi nel caso in banda passante, e viene detta *maximal ratio combining*.

In un ricevitore non coerente si dovrebbe analogamente cercare il massimo di $|\mathbf{r}\cdot A_1\mathbf{s}_1(\vartheta_1) + \mathbf{r}\cdot A_2\mathbf{s}_2(\vartheta_2)|$. Spesso, per semplicità, si considera invece la somma dei quadrati dei due moduli, cioè si sommano in modo non coerente le uscite di due ricevitori. Non di rado poi le ampiezze non sono note, e quindi si somma con pesi uguali.

Per comprendere il motivo per cui si utilizzano ricevitori in diversità basta pensare al caso in cui il guadagno del canale è una variabile casuale complessa con componenti in fase e quadratura gaussiane. In tal caso la *ddp* dell'ampiezza del segnale ricevuto è di *Rayleigh* e la *ddp* dell'energia ricevuta è *esponenziale* (es. 3.8).

Si consideri ad esempio la trasmissione 2PSK differenziale, e sia E_b il valor *medio* dell'energia ricevuta per simbolo. Il valore di E_b/N_0 fluttua casualmente essendo moltiplicato per un guadagno in potenza G con *ddp* esponenziale, e con esso la probabilità d'errore. La probabilità d'errore *media* è data da

$$P(E) = \int_0^\infty \frac{1}{2}\exp\Big(-\frac{GE_b}{N_0}\Big)\exp(-G)dG = \frac{1}{2(1+E_b/N_0)} \tag{3.32}$$

risultato veramente orribile perché $P(E)$ non decresce esponenzialmente, ma linearmente, con E_b/N_0. Il motivo è che per una frazione non trascurabile del tempo il guadagno del canale è molto minore di uno e quindi la probabilità d'errore è prossima a 1/2. Un risultato quasi altrettanto disperante si ottiene con il ricevitore coerente (es. 3.14), pur con l'ardita ipotesi di poter recuperare perfettamente la fase della portante e il sincronismo di simbolo anche quando

il segnale è molto affievolito[8].

Potendo disporre di due repliche *indipendenti* del segnale, la probabilità che *entrambe* siano fortemente affievolite si riduce notevolmente. Il calcolo non è semplice, ma si può mostrare che la probabilità d'errore è asintoticamente proporzionale a $(E_b/N_0)^{-2}$, e che con diversità di ordine L è proporzionale a $(E_b/N_0)^{-L}$.

Risultati analoghi si ottengono per la modulazione 4PSK differenziale.

Per concludere questi rapidi cenni sulla ricezione in diversità, argomento estremamente vario e complesso, si può osservare che la diversità può essere ottenuta in un gran numero di modi, quali ad esempio: diversità di *frequenza* (trasmissione con portanti separate da una frequenza tale che i relativi contributi al ricevitore siano indipendenti; l'ordine di grandezza della separazione in frequenza è l'inverso della dispersione dei tempi d'arrivo delle diverse repliche del segnale); di *spazio* (antenne riceventi poste a distanza tale da produrre segnali affievoliti in modo indipendente; ciò si ottiene con una distanza pari ad alcune volte la lunghezza d'onda); di *tempo* (trasmissione in momenti diversi, sufficientemente distanti; ciò dipende dalla coerenza temporale del canale, cioè dall'intervallo di tempo oltre il quale si può ritenere che i cammini multipli si ricombinino in modo indipendente).

3.8 Rumore gaussiano non bianco

In presenza di rumore gaussiano, ma non bianco nella banda dei segnali, la rappresentazione geometrica dei segnali e del rumore è un po' complicata dal fatto che le funzioni base che verrebbero comode per rappresentare i segnali non producono componenti di rumore incorrelate. Detto altrimenti, possono essere necessarie infinite funzioni base anche per un insieme finito di segnali. Inoltre le componenti n_k del rumore, pur incorrelate, hanno varianze σ_k^2 diverse.

La teoria risulta faticosa se ci si pone il vincolo, normalmente non giustificato, di osservare il segnale ricevuto $r(t)$ solo in un intervallo di tempo pari alla durata dei segnali, come si farebbe nel caso di rumore bianco. Essendo il rumore correlato può risultare utile dare un'occhiata anche al di fuori di tale

[8] un modo per avvicinarsi a tale condizione è trasmettere, di tanto in tanto, simboli noti; dai campioni ricevuti si può stimare il guadagno complesso del canale negli istanti corrispondenti e, interpolando, anche negli altri istanti; per la trasmissione di tali *simboli pilota* occorre naturalmente spendere dell'energia, che non reca *direttamente* informazione

intervallo.

L'analisi è invece piuttosto semplice se si consente l'utilizzo di un filtro che renda bianco il rumore, perlomeno nella banda dei segnali, per poi applicare le tecniche usuali. Se la densità spettrale del rumore è nota e se il filtro sbiancatore esiste ed è invertibile, come sempre accade in pratica, non si fa alcun danno irreversibile sbiancando. Infatti si può sempre tornare al punto di partenza e quindi non si rinuncia alla ottimalità del ricevitore.

Resta solo da osservare che il ricevitore, a valle del filtro sbiancatore, deve essere progettato per i segnali *uscenti* dal filtro, diversi da quelli *trasmessi* e con energia diversa, in generale. Dunque anche nel semplice caso binario antipodale la probabilità d'errore non dipende direttamente dall'energia dei segnali (es. 3.15). Occorre anche notare che se i segnali sono ottenuti in trasmissione come somma di funzioni ortogonali, tale ortogonalità viene persa nel passaggio attraverso il filtro sbiancatore. Di ciò comunque si sa come tenere conto.

Nel caso di densità spettrali di potenza esprimibili come funzioni razionali fratte la funzione di trasferimento del filtro sbiancatore è immediata: basta fattorizzare lo spettro del rumore e prendere poli e zeri che corrispondono ad un filtro causale stabile e con inverso causale stabile. Ad esempio se $S_n(f) = (1+A^2f^2)/(1+B^2f^2)$ il filtro sbiancatore ha funzione di trasferimento $H(f) = (1+jBf)/(1+jAf)$. La risposta impulsiva, causale, ha una qualche durata[9] t_0. Supponendo i segnali non nulli nell'intervallo da 0 a T_0, il filtro da un lato sente l'effetto (del solo rumore) anche nell'intervallo da $-t_0$ a 0, dall'altro allunga i segnali fino all'istante T_0+t_0. Le correlazioni eseguite a valle del filtro sbiancatore nell'intervallo da 0 a $T_0 + t_0$ sono dunque combinazioni lineari dei valori di $r(t)$ nell'intervallo da $-t_0$ a $T_0 + t_0$. Al di fuori di tale intervallo è inutile guardare perché la correlazione del rumore non si estende oltre.

Se invece si imponesse artificialmente il vincolo di osservare $r(t)$ solo nell'intervallo da 0 a T_0 il progetto del ricevitore sarebbe più complicato, e le prestazioni lievemente peggiori avendo buttato dell'informazione. In pratica imporre tale vincolo potrebbe sembrare giustificato solo se si volesse trasmettere una successione di segnali in intervalli di durata T_0 temporalmente separati, per non avere problemi di interferenza intersimbolica. Ma l'interferenza intersimbolica viene curata con altri mezzi, che si vedranno in seguito.

[9] non si obietti che t_0 è infinito, in teoria

3.9 Esercizi

3.1 - Si sia trasmessa una forma d'onda triangolare simmetrica di durata T con ampiezza $\pm A$, in presenza di rumore bianco. In ricezione si preferisce eseguire la correlazione con un *rettangolo* (simmetrico) di durata T_0 e ampiezza unitaria, per evitare di dover eseguire prodotti. Si calcoli la degradazione delle prestazioni rispetto al ricevitore ML e si mostri che esiste un valore ottimo di T_0, minore di T. *Suggerimento*: basta calcolare l'angolo formato dai vettori corrispondenti alle funzioni triangolare e rettangolare.

3.2 - Si verifichino le espressioni date nel testo per la probabilità d'errore nella modulazione 4PSK in presenza di un errore ε di fase. Si approssimi la probabilità d'errore sviluppando in serie di potenze la funzione $Q(y)$ fino al secondo ordine. Infine si traduca l'aumento di $P(E)$ in una equivalente perdita di E_b/N_0 (espressa in dB), e si mostri che tale perdita è approssimabile, per piccoli valori di ε, con $4.34(2E_b/N_0)^2\varepsilon^2$ dB.
Suggerimenti: ponendo $\rho = \sqrt{2E_b/N_0}$ si ha

$$Q(\rho(1 \pm \varepsilon)) \approx Q(\rho)(1 \mp \rho^2\varepsilon + \rho^4\varepsilon^2/2)$$

$$Q(\rho(1-\delta)) \approx Q(\rho)(1+\rho^2\delta)$$

$$20\log_{10}(1-\delta) = 20\log(1-\delta)/\log 10 \approx -8.68\delta$$

3.3 - Si ripeta il calcolo dell'esercizio precedente per la modulazione 8PSK.

3.4 - Si consideri il caso, citato nel testo, di trasmissione di due bit successivi con forme d'onda a radice di Nyquist e correlazioni eseguite con un errore di temporizzazione τ. Si calcolino le distanze modificate d'_{ij}. Si estenda poi al caso di trasmissione di una lunga sequenza di bit.

3.5 - Nella trasmissione 4PSK si correli il segnale ricevuto con $g(t)\cos(2\pi f_0 t)$ e $g(t)\cos(2\pi f_0 t + \varphi)$. Si mostri come dalle due componenti *non ortogonali* è possibile ottenere le coordinate ortogonali *senza errore*, purché le due fasi non siano coincidenti od opposte; in pratica fasi quasi uguali danno problemi di precisione numerica. *Suggerimento*: data la semplicità della geometria non occorre una teoria generale su assi non ortogonali; si verifichi comunque che essa fornisce i risultati corretti.

3.6 - Il ricevitore ML per segnali *on-off* $\mathbf{s}_1$ di energia E ed $\mathbf{s}_2 = 0$ prevede che la decisione sia basata su una soglia posta a $\sqrt{E}/2$. Si consideri il caso in cui si vuole rivelare l'eventuale presenza del segnale $\mathbf{s}_1$, ma non ne sia nota l'ampiezza. Ad esempio si voglia rivelare la presenza di eco radar proveniente da un eventuale bersaglio, con ampiezza incognita perché dipendente sia dalla distanza sia dalla sezione efficace dell'oggetto. Non sapendo come fissare la soglia, un criterio ragionevole è basato sul fatto che il segnale è noto *se è assente*. Si può fissare la soglia per una prefissata probabilità $P(\mathbf{s}_1/\mathbf{s}_2)$ di *falso allarme*. Se questa è ad esempio 10^{-3}, si determini la soglia e la probabilità $P(\mathbf{s}_2/\mathbf{s}_1)$ di *mancata rivelazione* in funzione di E. *Commento*: si noti l'estrema semplificazione del supporre che si sia trasmesso *un solo* impulso radar; in realtà la ricerca del bersaglio viene continuamente ripetuta.

3.7 - Si considerino i segnali $s_i(t) = A\cos 2\pi f_i t$, nell'intervallo da 0 a T. Quale è la struttura del ricevitore non coerente? Quale deve essere la spaziatura in frequenza perché i segnali siano ortogonali? Quale è la probabilità d'errore nel caso binario ortogonale?

3.8 - Si mostri che l'ampiezza A di un vettore in due dimensioni con componenti u e v gaussiane indipendenti a valor medio nullo e con varianza $N_0/2$ ha *ddp* di *Rayleigh*

$$f(A) = \frac{2A}{N_0}\exp(-\frac{A^2}{N_0})$$

Si verifichi che $E[A^2] = 2N_0/2 = N_0$. Si mostri che la *ddp* del quadrato dell'ampiezza è esponenziale con valor medio N_0.

3.9 - Si mostri che l'ampiezza A di un vettore in due dimensioni con componenti x e y gaussiane indipendenti, con $E[x] = \sqrt{E_s}$ e $E[y] = 0$, e con varianza $N_0/2$ ha *ddp* di *Rice*

$$f(A) = \frac{2A}{N_0}\exp(-\frac{A^2 + E_s}{N_0})I_0(\frac{2A\sqrt{E_s}}{N_0})$$

dove

$$I_0(z) = \frac{1}{2\pi}\int_0^{2\pi}\exp(z\cos\varphi)d\varphi$$

è la funzione di Bessel modificata di prima specie e di ordine zero.

Si mostri che lo stesso risultato vale con $E[x] = \sqrt{E_s}\cos\vartheta$ e $E[y] = \sqrt{E_s}\sin\vartheta$. *Commento*: per $E_s = 0$ la *ddp* di Rice coincide con quella di Rayleigh.

3.10 - Si mostri che se $E_s \gg N_0/2$ la *ddp* di Rice tende ad una gaussiana con valor medio $\sqrt{E_s}$ e varianza $N_0/2$. *Suggerimento*: $\sqrt{x^2+y^2} \approx x$.

3.11 - Si mostri che nella ricezione differenziale M-PSK (su due simboli) la decisione è basata sul confronto tra il numero complesso $r_k r_{k-1}^*$ e le regioni di decisione della demodulazione coerente. *Suggerimento*: si consideri la ricerca del massimo di $|r_{k-1}d_{k-1}^* + r_k d_k^*|^2$; si espanda il quadrato e si ignorino i termini indipendenti dai dati.

3.12 - Si consideri la seguente realizzazione del ricevitore differenziale per la modulazione 2PSK: il segnale ricevuto viene filtrato con un passa banda con risposta impulsiva $g(t_0-t)\cos(2\pi f_0 t + \psi)$, con ψ arbitrario; l'uscita $u(t)$ viene moltiplicata per la replica ritardata $u(t-T)$; vengono eliminate le componenti intorno alla frequenza $2f_0$ e si campiona all'istante t_0; infine si decide in base al segno. Si mostri che il ricevitore equivale a quello descritto nel testo se $f_0 T$ è un numero *intero*, e si spieghi che modifica occorrerebbe introdurre altrimenti. *Suggerimento*: si usino gli equivalenti passa basso. *Commento*: in genere $f_0 T \gg 1$ ed è difficile controllare il ritardo con precisione sufficiente, per cui questo ricevitore è più teorico che pratico.

3.13 - Si consideri la trasmissione di simboli 4PSK con ricezione differenziale basata sull'osservazione di n campioni ($n > 2$), supponendo che la fase ϑ sia costante nel corrispondente intervallo di tempo. Si calcoli, in funzione di n, la degradazione asintotica (in dB) rispetto alla demodulazione coerente, e si mostri che tende a zero per $n \to \infty$.

3.14 - Si consideri la trasmissione binaria antipodale in banda passante su un canale di Rayleigh. Detta E_b l'energia media ricevuta, e supponendo possibile il perfetto recupero della portante e dei sincronismi di simbolo anche in presenza di forti affievolimenti, la probabilità d'errore condizionata all'ampiezza A, normalizzata, del segnale ricevuto è

$$P(E/A) = Q\Big(\sqrt{\frac{2E_b}{N_0}}A\Big)$$

dove A ha *ddp* di Rayleigh $f(A) = 2A\exp(-A^2)$. Si calcoli la probabilità d'errore *media*, e la si confronti con quella del ricevitore non coerente. *Commento*: il risultato è quasi altrettanto tragico.

3.15 - I segnali $s_1(t) = A\frac{\sin \pi t/T}{\pi t/T}$ ed $s_2(t) = -s_1(t)$ con energia E_b sono utilizzati su un canale con rumore gaussiano con densità spettrale di potenza *unilatera* $N_0/(1 + 4\pi^2 f^2 T^2)$. Si determini il ricevitore a massima verosimiglianza e la probabilità d'errore. *Suggerimento*: può essere utile osservare che $\int s_i(t)s_i'(t)dt = 0$. *Commento*: forme d'onda con *roll-off* nullo vanno bene solo negli esercizi, per semplificare i calcoli.

3.16 - Il filtro sbiancatore causale ed invertibile per un rumore gaussiano con densità spettrale di potenza proporzionale a $1/(1 + A^2 f^2)$, come nell'esercizio precedente, ha risposta impulsiva istantanea e quindi il ricevitore ottimo per segnali di durata T_0 non utilizza il segnale ricevuto $r(t)$ al di fuori di tale intervallo. Si dia una spiegazione intuitiva di tale fatto. *Suggerimento*: il rumore considerato è un processo di Markov.

Capitolo 4

Capacità di canale

4.1 Introduzione

Nei capitoli precedenti si sono introdotti, in modo non sistematico, alcuni sistemi di trasmissione *codificata* in cui per semplificare la generazione del segnale da inviare sul canale e la realizzazione del ricevitore si utilizza una sequenza di forme d'onda elementari corrispondenti a simboli monodimensionali o bidimensionali, rispettivamente in banda base e in banda passante.

Già da questi primi esempi si trae l'indicazione che è possibile ottenere una maggior efficienza nell'uso dell'energia se i punti della costellazione, cioè le coordinate del vettore trasmesso, non sono scelti indipendentemente simbolo per simbolo.

Combinando opportunamente la dimensione della costellazione elementare e le regole del codice si può controllare il numero di bit d'informazione per dimensione, cioè l'efficienza spettrale. Si supponga ad esempio di voler trasmettere un bit per dimensione. Tra i sistemi non codificati non c'è praticamente scelta: binario antipodale in banda base; quattro fasi in banda passante. Se invece si utilizza uno spazio dei segnali ad N dimensioni basta in linea di principio selezionare un insieme di $M = 2^N$ punti, opportunamente disposti in modo da ottenere un buon insieme di distanze reciproche tra i segnali (ponendo particolare attenzione alla distanza minima). Il valore del numero di dimensioni N è del tutto libero, ed anche fissato questo ci sono moltissimi modi di disporre i segnali. L'unico timore può essere di non saper usufruire di questi troppo numerosi gradi di libertà. In altre parole, la pigrizia mentale suggerirebbe di non volgere lo sguardo oltre una o due dimensioni.

Se più in generale si vogliono trasmettere R bit per dimensione basta selezionare $M = 2^{NR}$ punti, in modo da trasmettere NR bit d'informazione in N dimensioni.

Si apre quindi una vastissima scelta. Naturalmente si deve essere disposti ad aumentare la complessità dell'insieme di segnali, e quindi la complessità del ricevitore ed il ritardo con cui vengono generati i segnali e prese le decisioni in ricezione.

Apertisi questi orizzonti, si può affrontare il problema da due punti di vista, complementari. Un primo consiste nel ricercare sempre nuove soluzioni pratiche, che consentano le prestazioni desiderate con il minimo di complessità; il secondo, più speculativo, nel domandarsi se vi sia un limite invalicabile alle prestazioni ottenibili, da poter confrontare con i sistemi pratici già individuati. Poiché la maggior libertà possibile nella scelta dei segnali si ha con un grande numero di dimensioni, ci si chiederà quali possano essere le prestazioni non ponendo alcun limite al numero N di dimensioni, e quindi alla complessità e al ritardo[1].

In questo capitolo si affronterà il problema dei limiti teorici; solo successivamente ci si rivolgerà alle soluzioni pratiche.

4.2 Il "cutoff rate"

La prima difficoltà che si incontra è la scelta dell'insieme dei segnali, non avendo validi criteri che aiutino a disporre i punti nello spazio ad N dimensioni. La seconda difficoltà è la valutazione delle prestazioni. L'unico strumento semplice è lo *union bound*. Tuttavia, ai livelli di probabilità d'errore di interesse pratico, gli esempi elementari visti nei capitoli precedenti sembrano indicare che il *bound* fornisce indicazioni sempre meno attendibili all'aumentare del numero dei segnali e delle dimensioni. Si tratta comunque di una maggiorazione, e non ci si lamenterà troppo se le prestazioni risulteranno migliori, anche di molto, di quanto predetto dallo *union bound*.

Per quanto questi strumenti sembrino modesti, se ne possono ottenere interessanti risultati. Ad esempio utilizzando lo *union bound* e la ulteriore

[1] nella gran parte dei casi pratici il ritardo prodotto da un migliaio di dimensioni è del tutto trascurabile; il vero vincolo è la complessità del ricevitore, dovuta alla ricerca del massimo tra $M = 2^{NR}$ verosimiglianze

maggiorazione[2] $Q(y) < \exp(-y^2/2)$, valida per $y \geq 0$, si ha

$$P(E) \leq \sum_i P(\mathbf{s}_i) \sum_{j \neq i} Q\Big(\frac{|\mathbf{s}_i - \mathbf{s}_j|}{\sqrt{2N_0}}\Big) < \sum_i P(\mathbf{s}_i) \sum_{j \neq i} \exp\Big(-\frac{|\mathbf{s}_i - \mathbf{s}_j|^2}{4N_0}\Big) \quad (4.1)$$

Mentre è un compito arduo valutare tale espressione per uno specifico insieme di moltissimi segnali, che occorre naturalmente aver definito, risulta molto più agevole calcolarne la media rispetto agli insiemi di segnali che si ottengono scegliendo *casualmente* e *indipendentemente* i punti $\mathbf{s}_i$ $(i = 1, \ldots, M)$ con una *ddp* prefissata. Si noti che la scelta *indipendente* dei segnali $\mathbf{s}_i$ implica che lo stesso punto potrebbe essere scelto due volte $(\mathbf{s}_j = \mathbf{s}_i)$ o comunque potrebbero essere selezionati punti molto vicini tra loro; in uno spazio con un gran numero di dimensioni tale infelice evenienza risulta però poco probabile.

Certamente esiste almeno un codice con probabilità d'errore non maggiore di quella *media* su tutti i codici. Se dunque questa risulta accettabile, l'esistenza di almeno un codice con prestazioni soddisfacenti è garantita, ed anzi non è difficile dedurre il confortante risultato che una frazione considerevole dei codici ha prestazioni soddisfacenti. Si deve tuttavia osservare che non si sono trovati esplicitamente i codici, né si è avuta alcuna indicazione sulla complessità del ricevitore, che dipende fortemente dalle caratteristiche dell'insieme dei segnali.

Se i segnali $\mathbf{s}_i$ ed $\mathbf{s}_j$ sono scelti indipendentemente e con una stessa *ddp*, il valor medio di $\exp(-|\mathbf{s}_i - \mathbf{s}_j|^2/4N_0)$ non dipende dagli indici i e j. Osservando poi che $\sum P(\mathbf{s}_i) = 1$ e che nella (4.1) la somma interna ha $M - 1$ termini si ottiene

$$P(E) < ME[\exp(-\frac{|\mathbf{s}_i - \mathbf{s}_j|^2}{4N_0})] \quad (4.2)$$

Come ulteriore semplificazione si supponga infine che i vettori $\mathbf{s}_i$ siano ottenuti accostando n simboli $\mathbf{x}_{ik}$ $(k = 1, \ldots, n)$ scelti *casualmente* e *indipendentemente* da una qualche costellazione prefissata, con m punti, tipicamente (ma non necessariamente) ad una dimensione in banda base e due in banda passante. Il generico punto della costellazione sia scelto con probabilità prefissata p_l

[2] si è già mostrato che $Q(y) \leq \frac{1}{2}\exp(-y^2/2)$ (es. 2.2); per gli scopi di questo capitolo il fattore $1/2$ può essere ignorato

$(l = 1, \ldots, m)$. Si ha

$$E[\exp(-\frac{|\mathbf{s}_i - \mathbf{s}_j|^2}{4N_0})] = E[\prod_{k=1}^{n} \exp(-\frac{|\mathbf{x}_{ik} - \mathbf{x}_{jk}|^2}{4N_0})] = \\ = \left(E[\exp(-\frac{|\mathbf{x}_{ik} - \mathbf{x}_{jk}|^2}{4N_0})]\right)^n = \left(\sum_{i=1}^{m}\sum_{j=1}^{m} \exp(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4N_0}) p_i p_j\right)^n \tag{4.3}$$

Si osservi che nell'ultima espressione si sono riutilizzati gli indici i e j per enumerare gli m segnali della costellazione elementare, anziché gli M segnali dello spazio ad N dimensioni.

Infine ponendo $M = 2^{nR}$, dove ora R è espresso in bit per simbolo, e definendo il *cutoff rate* (anch'esso espresso in bit per simbolo)

$$R_0 = -\log_2 \sum_{i=1}^{m}\sum_{j=1}^{m} \exp(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{4N_0}) p_i p_j \tag{4.4}$$

si ottiene

$$P(E) < 2^{-n(R_0 - R)} \tag{4.5}$$

Ovviamente il ritmo di trasmissione R ed il *cutoff rate* R_0 possono essere espressi, se lo si preferisce, in bit per dimensione anziché per simbolo. In tal caso il numero di segnali è dato da $M = 2^{NR}$ e si ha

$$P(E) < 2^{-N(R_0 - R)} \tag{4.6}$$

Il punto sorprendente è che si è dimostrata l'esistenza di sistemi di trasmissione con probabilità d'errore piccola a piacere purché il ritmo di trasmissione sia minore di un valore *finito* dato dal *cutoff rate* R_0. Basta infatti scegliere n, e quindi il numero N di dimensioni, sufficientemente grande. Ed anzi queste semplici considerazioni danno anche una idea, pur grossolana, del valore di n che potrà essere richiesto in pratica.

Con argomenti decisamente più complessi si può dimostrare che una relazione analoga alla (4.6), ma con un valore di R_0 maggiore[3], vale anche se i punti $\mathbf{s}_i$ vengono scelti con *ddp* uniforme *sulla superficie* di una sfera ad N dimensioni di raggio prefissato, e dunque con un valore prefissato dell'energia,

[3] l'espressione di R_0 è piuttosto involuta, e tutto sommato poco interessante

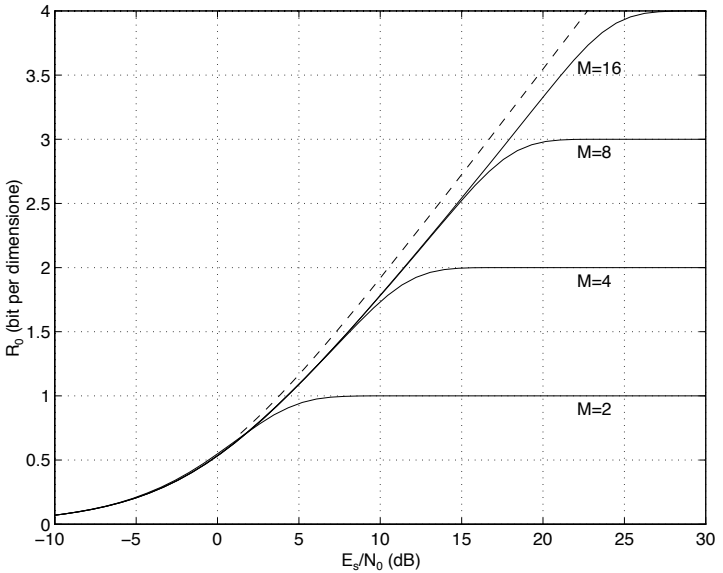

Fig. 4.1 - R_0 in funzione di E_s/N_0 per costellazioni monodimensionali ad M livelli equispaziati ed equiprobabili e, per confronto, per punti con *ddp* uniforme sulla sfera ad N dimensioni (curva tratteggiata)

oppure *entro* la stessa sfera, e quindi con un vincolo sull'energia massima. Si noti che in questi ultimi casi le singole coordinate del vettore trasmesso *non* vengono scelte indipendentemente.

La fig. 4.1 mostra, nel caso particolare di segnali elementari $\mathbf{x}_i$ *equiprobabili* tratti da costellazioni M-PAM il valore di R_0 (in bit per simbolo, ovvero bit per dimensione) in funzione del rapporto E_s/N_0 tra energia per simbolo e densità spettrale unilatera del rumore. Nelle modulazioni multilivello i livelli estremi sono lievemente favoriti rispetto a quelli interni, e dovrebbero quindi essere utilizzati con frequenza un po' maggiore. Ciò complica notevolmente il calcolo di R_0 ed anche l'insieme dei segnali $\mathbf{s}_i$, dando in compenso un vantaggio molto modesto. Questa possibilità viene quindi di norma ignorata.

In fig. 4.1 è anche mostrato per confronto (per N grande) il valore di R_0, espresso in bit per dimensione, che si ottiene scegliendo punti uniformi sulla sfera ad N dimensioni. Il miglioramento rispetto alla scelta indipendente simbolo per simbolo può essere attribuito al modo più raffinato di selezione

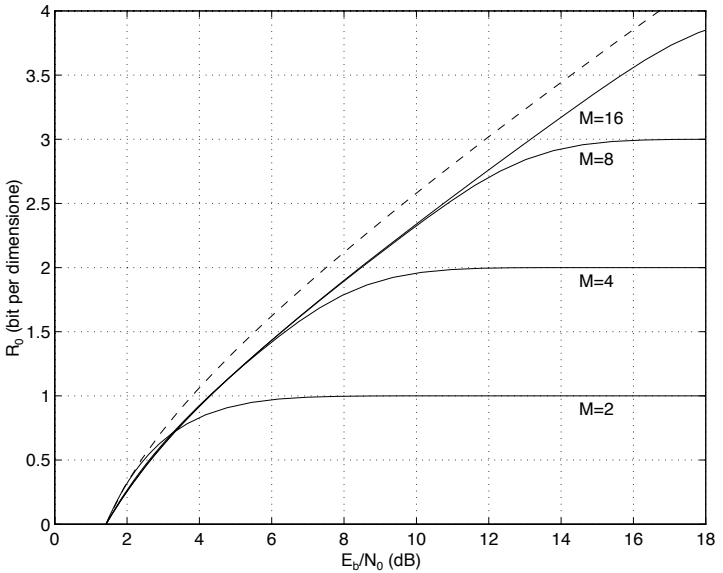

Fig. 4.2 - R_0 in funzione di E_b/N_0 per costellazioni monodimensionali ad M livelli equispaziati ed equiprobabili e, per confronto, per punti con *ddp* uniforme sulla sfera ad N dimensioni (curva tratteggiata)

dei punti $\mathbf{s}_i$.

Se si vuol determinare il valore di E_b/N_0 richiesto per ottenere un valore prefissato di R_0 basta osservare che $E_s = E_b R$ e porre $R = R_0$. La fig. 4.2 mostra gli stessi grafici di fig. 4.1 in funzione di E_b/N_0.

La fig. 4.3 mostra grafici analoghi per alcune costellazioni bidimensionali, sempre con punti equiprobabili[4]. Si osservi che l'utilizzo di una costellazione QAM con punti equiprobabili è del tutto equivalente alla trasmissione di due simboli PAM indipendenti. Quindi R_0, se espresso in bit per dimensione, ha lo stesso valore (come mostrano le figure). Si può infine osservare che il valore di E_b/N_0 richiesto per R_0 tendente a zero non dipende dalla costellazione (es. 4.1) e coincide con quello trovato nel Cap. 2 per segnali *ortogonali* o anche *biortogonali*.

[4]la costellazione *a croce* 32CR (*Cross* con 32 punti) è una costellazione QAM 6x6 da cui sono tolti i quattro punti con energia maggiore; in modo analogo si può ottenere, ad esempio, la costellazione 128CR togliendo 4 punti da ogni angolo di una costellazione QAM 12x12

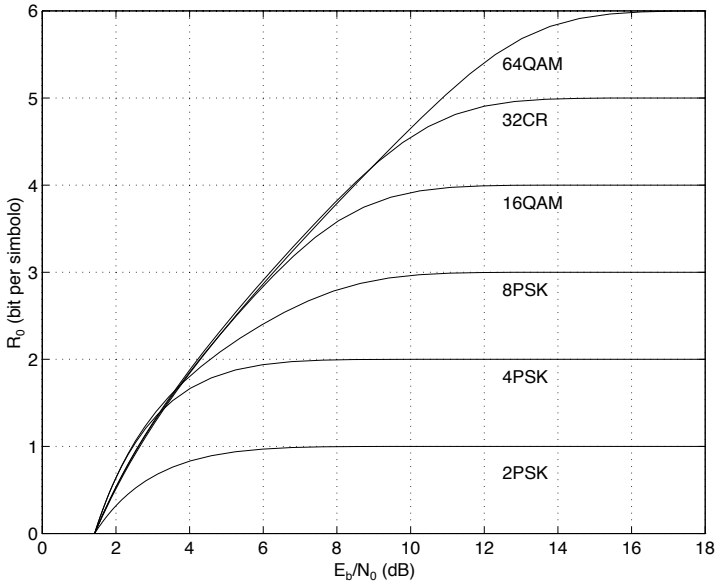

Fig. 4.3 - R_0 in funzione di E_b/N_0 per costellazioni bidimensionali (punti equiprobabili)

La maggiorazione della probabilità d'errore ottenuta con il *cutoff rate*, molto significativa in teoria, potrebbe esserlo meno in pratica; infatti la complessità di un insieme di segnali scelti casualmente, e del relativo ricevitore ML, è certamente eccessiva. Può essere interessante offrire un esempio di confronto tra le prestazioni ottenibili con un particolare sistema di modulazione multidimensionale e quelle previste per la media dei codici. Si può considerare un sistema di trasmissione che utilizzi segnali ben disposti sulla superficie di una sfera, quali possono essere i segnali ortogonali o biortogonali. L'insieme è purtroppo poco denso e ha applicazioni pratiche solo in casi speciali, a causa della modesta efficienza spettrale. In compenso è relativamente facile calcolare la probabilità d'errore. La fig. 4.4 mostra, al variare di N, la probabilità d'errore $P(E)$ per segnali ortogonali mantenendo E_b/N_0 pari a $4\log 2$, e per confronto quella maggiorata mediante la (4.6), che vale per la media dei codici. Come si vede c'è un discreto accordo, a parte un fattore moltiplicativo. Risultati analoghi si otterrebbero per segnali biortogonali.

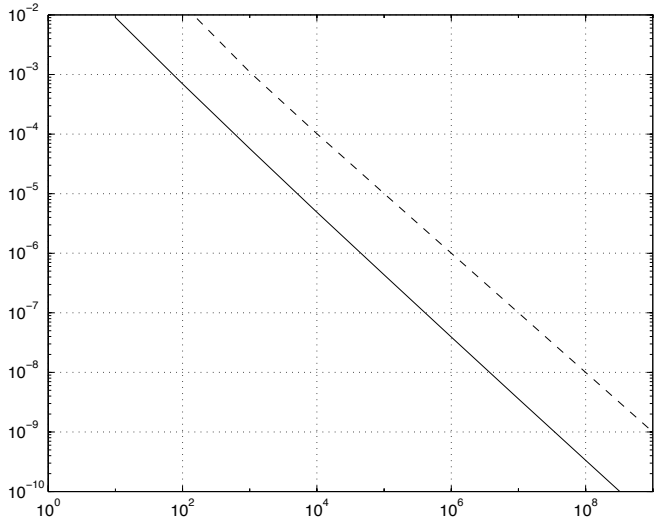

Fig. 4.4 - Probabilità d'errore per N segnali ortogonali ($E_b/N_0 = 4\log 2$; curva continua) e maggiorazione per la media dei codici mediante la (4.6) (tratteggio)

4.3 Capacità di canale

Dopo aver mostrato che è possibile trasmettere l'informazione ad un ritmo finito con probabilità d'errore piccola a piacere, risulterà meno sconvolgente venire a sapere che il ritmo può essere persino maggiore di R_0, qualunque sia la probabilità d'errore desiderata. È comunque interessante determinare quale sia il vero limite teorico. Per ottenere questo risultato occorrono tecniche di maggiorazione assai più raffinate, con le quali si può dimostrare che

$$P(E) \leq 2^{-NE(R)} \tag{4.7}$$

dove la funzione $E(R)$ è detta *esponente d'errore*. Nel primo tratto si ha $E(R) = R_0 - R$, in accordo con quanto già trovato con mezzi elementari. Tuttavia l'esponente d'errore $E(R)$ risulta positivo anche oltre R_0, fino alla cosiddetta *capacità di canale* C. Si può infine mostrare che la capacità di canale è davvero il limite invalicabile, nel senso che se $R > C$ la probabilità d'errore $P(E)$ tende ad uno per N tendente all'infinito, *qualunque sia il codice*.

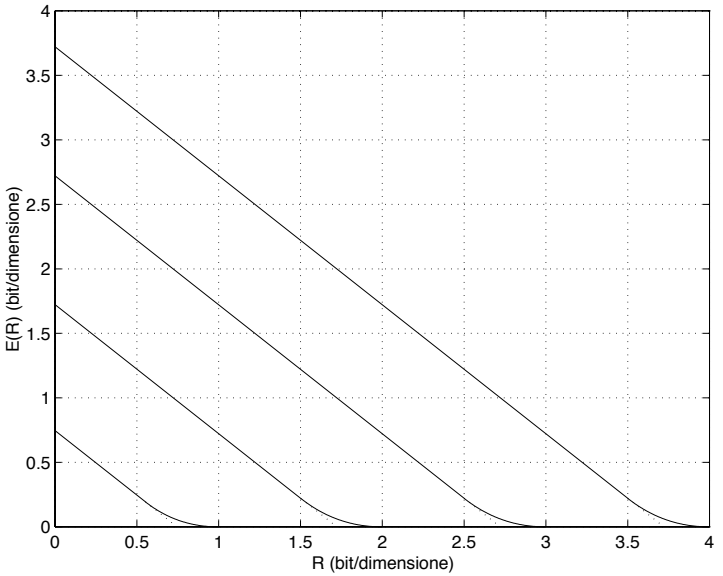

Fig. 4.5 - Andamenti tipici dell'esponente d'errore $E(R)$ (punti con *ddp* uniforme sulla sfera ad N dimensioni)

La fig. 4.5 mostra alcuni andamenti tipici dell'esponente d'errore, nel caso di punti distribuiti uniformemente sulla superficie della sfera ad N dimensioni. I valori di E_s/N_0 per i quattro grafici sono scelti in modo che risulti $C = 1$, 2, 3 e 4 bit per dimensione rispettivamente. Si può osservare, ed è un fatto generale, che la curva dell'esponente d'errore raggiunge la capacità con pendenza nulla. Ci si deve aspettare quindi una estrema difficoltà nell'ottenere basse probabilità d'errore per valori di R prossimi a C. Non di rado si ritiene che il *cutoff rate* R_0 dia una buona indicazione del ritmo di trasmissione *praticamente* raggiungibile, lasciando alla capacità un valore quasi solo teorico. Peraltro per elevate capacità R_0 non si discosta molto da C.

Nel caso di punti uniformi *sulla* sfera o *entro* la sfera ad N dimensioni la capacità C ha una espressione molto semplice, e giustamente famosa. In bit per dimensione essa è data da

$$C = \frac{1}{2}\log_2(1 + \frac{E_d}{N_0/2}) \tag{4.8}$$

dove E_d è l'energia per dimensione. Se poi si pone $N = 2BT_0$, cioè si assume che non vi sia alcun eccesso di banda e dunque siano disponibili $2B$ dimensioni al secondo, e quindi $P = 2BE_d$ è la potenza media dei segnali, si ottiene facilmente il valore di C in bit/s:

$$C = B \log_2(1 + \frac{P}{BN_0}) \tag{4.9}$$

Tale espressione è particolarmente conveniente se sono fissati a priori i valori della banda e della potenza disponibile (ovviamente il ritmo di trasmissione realizzabile in pratica sarà minore della capacità). Si può osservare che se $P/BN_0 \gg 1$ la capacità cresce piuttosto lentamente con la potenza, e risulta invece pressoché proporzionale alla banda. Viceversa, se $P/BN_0 \ll 1$ la banda ha un effetto trascurabile sulla capacità, che risulta proporzionale alla potenza. Infatti poiché $\log_2(1 + x) \approx x / \log 2$ per $x \ll 1$, si ottiene

$$C \approx \frac{P}{N_0 \log 2} \tag{4.10}$$

Nei due casi estremi si usa dire che il sistema di trasmissione è rispettivamente *limitato in banda* o *limitato in potenza*.

Se si impone $R \leq C$ e nella (4.8) si sostituisce $E_d = RE_b$, con R in bit per dimensione, e infine si risolve rispetto ad E_b/N_0 si ottiene la semplicissima espressione

$$\frac{E_b}{N_0} \geq \frac{2^{2R} - 1}{2R} \tag{4.11}$$

che mostra che il valore minimo teorico di E_b/N_0 dipende solo dal numero di bit per dimensione che si vuole trasmettere. Si noti che il termine $2R$ che compare nella (4.11) non è altro che l'efficienza spettrale teorica, in bit/s/Hz, nel caso in cui non vi sia alcun eccesso di banda. La (4.11) è quindi particolarmente comoda quando sono fissati a priori il ritmo di trasmissione e la banda, cioè l'efficienza spettrale. Il minimo teorico della potenza richiesta si ottiene poi moltiplicando l'energia minima per bit E_b per il ritmo di trasmissione in bit al secondo.

La fig. 4.6 confronta la capacità di canale C ed il *cutoff rate* R_0, già mostrato in fig. 4.2, in funzione di E_b/N_0 nel caso di punti uniformi sulla sfera ad N dimensioni. Per C ed R_0 prossimi a zero la differenza tende a 3 dB; per valori molto elevati si può mostrare che tende a 1.68 dB.

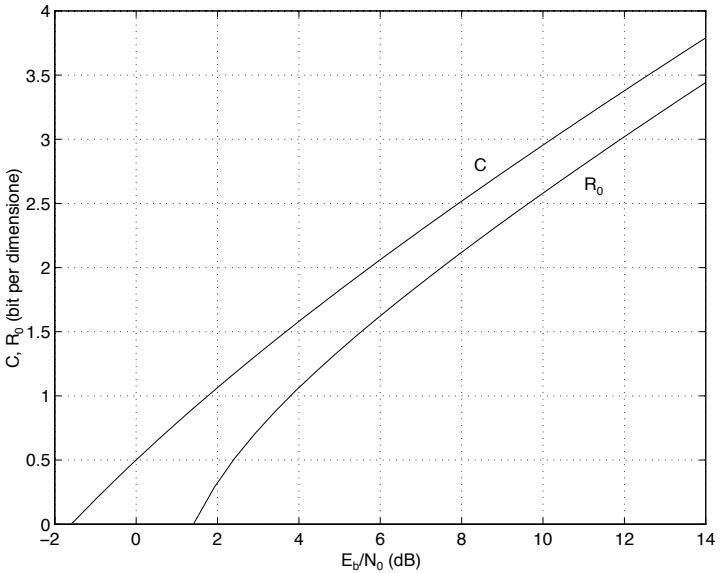

Fig. 4.6 - Confronto tra capacità di canale C e *cutoff rate* R_0, per punti con *ddp* uniforme sulla sfera ad N dimensioni

Per R tendente a zero, cioè efficienza spettrale nulla, il minimo valore teorico per E_b/N_0 è $\log 2$, pari a -1.59 dB; per $R = 0.5$ bit per dimensione è 0 dB; per $R = 1, 2$ e 3 bit per dimensione (come nella segnalazione non codificata a due, quattro e otto livelli per asse) 1.76, 5.74 e 10.21 dB rispettivamente[5].

Se si ritenesse praticamente invalicabile il *cutoff rate* R_0 i valori minimi di E_b/N_0 sarebbero rispettivamente 1.4, 2.4, 3.8, 7.5 e 11.9 dB. Resterebbe comunque una grande differenza, ad esempio per $R = 1$, tra il poter ottenere di una probabilità d'errore comunque piccola con $E_b/N_0 \geq 3.8$ dB e il valore di E_b/N_0 pari a 9.6 dB richiesto per ottenere $P(E) = 10^{-5}$ con la segnalazione binaria antipodale, e addirittura $E_b/N_0 = 13.1$ dB se si volesse $P(E) = 10^{-10}$.

Normalmente, per i motivi di semplicità già discussi, si fissa a priori la costellazione elementare. Ad esempio la fig. 4.7 mostra i valori della capacità C, in bit per simbolo, in funzione del rapporto E_b/N_0 per le costellazioni M-PAM da 2 a 16 livelli equiprobabili, e per confronto nel caso di punti

[5] è quindi giusto "pagare" 4 dB per passare da 1 a 2 bit per dimensione, ed altri 4.5 dB per trasmettere 3 bit per dimensione

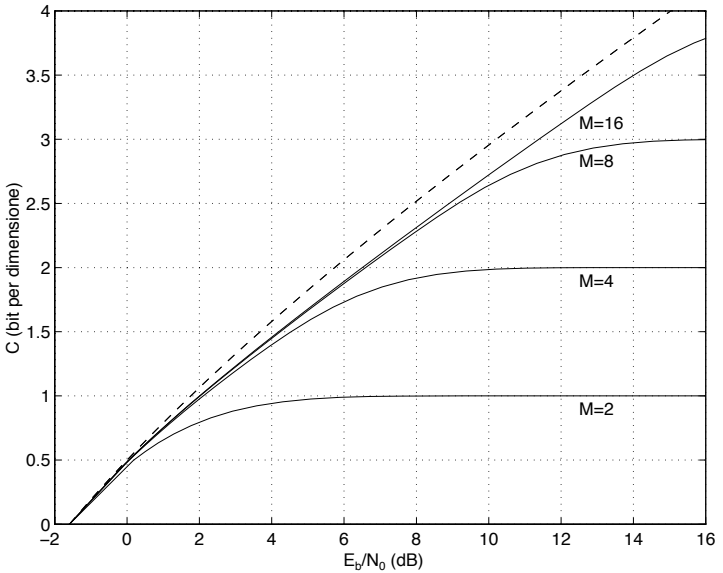

Fig. 4.7 - Capacità C per costellazioni monodimensionali ad M livelli equispaziati ed equiprobabili e, per confronto, per punti con *ddp* uniforme sulla sfera ad N dimensioni (curva tratteggiata)

uniformi sulla sfera ad N dimensioni. L'andamento qualitativo delle curve è molto simile a quello del *cutoff rate* R_0 di fig. 4.2. Quantitativamente le curve sono traslate, per un ampio tratto, di circa 2 dB. Anche i grafici della capacità, come quelli del *cutoff rate*, mostrano che imporre a priori una costellazione comporta una qualche penalizzazione. Questa è tuttavia modesta se il numero di livelli è sufficiente, ed è largamente compensata dalla maggior semplicità.

4.4 Capacità nel caso di rumore non bianco

Se il rumore che si somma ai segnali $s_i(t)$ è gaussiano ma non bianco, la capacità può essere calcolata immaginando di suddividere la banda disponibile in piccoli intervalli disgiunti, in cui si possa ritenere costante la densità spettrale del rumore. Questo potrebbe addirittura essere preso come un suggerimento pratico; infatti stanno diffondendosi sistemi di trasmissione

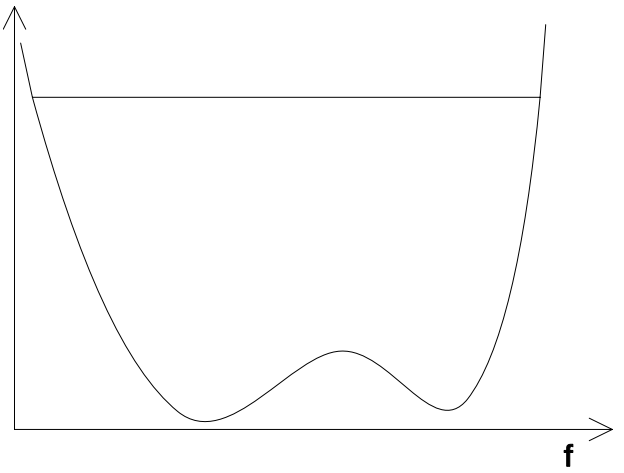

Fig. 4.8 - "Water pouring"

multifrequenza per la radiodiffusione di segnali audio e video, in cui la banda viene suddivisa in centinaia o addirittura migliaia di portanti. Supponendo, per maggior generalità, che il guadagno del canale possa variare con la frequenza, la capacità C in bit/s è data da una espressione del tutto analoga alla (4.9)

$$C = \int_0^\infty \log_2 \left(1 + \frac{S_s(f)C(f)}{N_0(f)}\right) df \tag{4.12}$$

dove $S_s(f)$ è la densità spettrale di potenza (unilatera) del segnale trasmesso, $C(f)$ è il guadagno in potenza del canale, $S_s(f)C(f)$ è quindi la potenza ricevuta per unità di banda, mentre $N_0(f)$ è la densità spettrale (unilatera) del rumore. La potenza media trasmessa è

$$P = \int_0^\infty S_s(f) df \tag{4.13}$$

È naturale chiedersi quale debba essere $S_s(f)$ per ottenere la massima capacità a parità di potenza trasmessa. Un semplice calcolo porta a

$$S_s(f) = \lambda - \frac{N_0(f)}{C(f)} \tag{4.14}$$

dove la costante λ è determinata dalla potenza disponibile. Si usa presentare visivamente tale risultato come in fig. 4.8: si deve immaginare di "versare" potenza, finché ce n'è, in un recipiente il cui fondo abbia livello pari a $N_0(f)/C(f)$ come se fosse acqua (*water pouring*); λ è il livello raggiunto. Implicitamente resta determinata anche la banda da utilizzare. Se si ha poca potenza conviene usare solo le frequenze con il miglior rapporto segnale-rumore. La capacità, dapprima proporzionale alla potenza, cresce però sempre più lentamente per cui conviene cominciare ad utilizzare anche altre frequenze (sempre che siano disponibili, e non già assegnate ad altri utenti o servizi). Se la potenza è sovrabbondante, in primissima approssimazione $S_s(f)$ è costante, in una banda che però dipende dalla potenza disponibile.

Dalla (4.12) si può infine calcolare il valore della capacità, che in bit/s risulta data da

$$C = \int \log_2 \frac{\lambda C(f)}{N_0(f)} df \tag{4.15}$$

dove l'integrale è esteso alla banda effettivamente utilizzata. In un sistema multifrequenza quest'ultima espressione suggerirebbe anche come suddividere i flussi di bit d'informazione fra le varie portanti. Naturalmente la capacità è solo un limite teorico; i ritmi di trasmissione in ciascuna banda saranno opportunamente ridotti.

4.5 Considerazioni finali

La capacità di canale, qui introdotta con specifico riferimento al caso di rumore additivo gaussiano, è generalizzabile ad altri contesti. È stata infatti definita da *Shannon* per ogni tipo di canale, con ingressi e uscite sia discreti sia continui, ed è calcolabile dalla descrizione statistica del legame ingresso-uscita. Per quanto la definizione generale sia interessante e suggestiva, il calcolo del valore della capacità non è però sempre agevole. Inoltre in un contesto di trasmissione numerica interessa soprattutto la probabilità d'errore in funzione del numero N di dimensioni, cioè l'esponente d'errore $E(R)$. Una indicazione utile si ottiene facilmente calcolando il *cutoff rate* R_0, ed eventualmente confrontandolo con la capacità C.

Da un punto di vista pratico i risultati relativi al *cutoff rate* e alla capacità di canale sono stati importanti perché hanno fortemente stimolato la ricerca di sistemi codificati che mantenessero almeno in parte le straordinarie promesse teoriche.

4.6 Esercizi

4.1 - La (4.4) è l'espressione di R_0 in bit per simbolo, per una costellazione prefissata. Per $R = R_0$ tendente a zero l'energia per simbolo $E_s = E_b R_0$ tende a zero. Usando le approssimazioni $\exp(-x) \approx 1 - x$ e $\log_2(1-x) \approx -x/\log 2$, valide per $x \ll 1$, si mostri che la (4.4) fornisce

$$R_0 = \frac{E_b R_0}{2 N_0 \log 2}$$

cioè $E_b/N_0 = 2 \log 2$ per *qualunque costellazione* con baricentro nell'origine.

4.2 - Si può dimostrare che per N grande un punto scelto uniformemente sulla superficie della sfera in N dimensioni ha componenti lungo ciascun asse con *ddp marginale* che tende alla gaussiana. Le variabili casuali sono però, sia pur debolmente, correlate. Si consideri invece il caso in cui le componenti sono scelte *indipendentemente* con *ddp* gaussiana. Si calcoli il valore di R_0, generalizzando la (4.4) al caso continuo, e si mostri che è minore di quello mostrato in fig. 4.6. *Suggerimento*: $x_i - x_j$ è una v.c. gaussiana con varianza $2E$, dove E è l'energia per dimensione; inoltre $\int \exp(-z^2/2\sigma^2) dz = \sqrt{2\pi}\sigma$. Si possono fare i seguenti commenti al risultato:

- a parità di *ddp marginale* la presenza di una correlazione, anche debole, tra le componenti x_{ik} del segnale $\mathbf{s}_i$ può migliorare le prestazioni; ciò vale probabilmente anche per le costellazioni monodimensionali e bidimensionali considerate in fig. 4.2 e 4.3; i codici pratici, in cui le componenti non sono ovviamente scelte a caso, dovrebbero quindi avere prestazioni migliori di quanto predetto dalla (4.5)

- la *ddp* marginale delle componenti x_{ik} gaussiana sembra essere una condizione necessaria per ottenere le massime prestazioni teoriche

- si può dimostrare (es. 4.4) che anche con componenti x_{ik} gaussiane *indipendenti* si raggiunge la capacità C data dalla (4.8), però l'esponente d'errore $E(R)$ è minore di quello ottenibile con punti uniformi sulla sfera (ciò è confermato, per il primo tratto, dalla riduzione del valore di R_0); inoltre viene a mancare la garanzia sull'energia massima dei segnali (questo non è però molto preoccupante: l'energia per dimensione $\frac{1}{N}\sum s_{ik}^2$ è ben poco casuale se N è grande)

4.3 - La formula per il calcolo della capacità di canale, per costellazioni con un numero finito m di punti $\mathbf{x}_i$ scelti *indipendentemente* con probabilità p_i $(i = 1, \ldots, m)$, è

$$C = \sum p_i \int f(\mathbf{r}/\mathbf{x}_i) \log_2 \Big(\frac{f(\mathbf{r}/\mathbf{x}_i)}{\sum p_i f(\mathbf{r}/\mathbf{x}_i)} \Big) d\mathbf{r} \tag{4.16}$$

dove l'integrale è in tante dimensioni quante ne occupano i simboli $\mathbf{x}_i$ (tipicamente una o due). Osservando che $f(\mathbf{r}/\mathbf{x}_i) = f_{\mathbf{n}}(\mathbf{r} - \mathbf{x}_i)$, dove $f_{\mathbf{n}}(\cdot)$ è la *ddp* del rumore, e che $\sum p_i f(\mathbf{r}/\mathbf{x}_i) = f(\mathbf{r})$ si mostri che occorre valutare due integrali del tipo $\int f(\mathbf{y}) \log_2 f(\mathbf{y}) d\mathbf{y}$, con $\mathbf{y} = \mathbf{n}$ e $\mathbf{y} = \mathbf{r}$ rispettivamente. *Commento*: i grafici di fig. 4.7 sono ottenuti in tal modo.

4.4 - Si generalizzi il risultato dell'esercizio precedente ad un insieme $\mathbf{x}$ continuo. Si valuti poi la capacità nel caso di vettori $\mathbf{x}$ monodimensionali con *ddp gaussiana* e si mostri che coincide con quello che si ottiene con punti uniformi sulla sfera ad N dimensioni, dato dalla (4.8). *Suggerimento*: $f(\mathbf{r})$ è gaussiana, con varianza pari alla somma delle varianze di $\mathbf{x}$ ed $\mathbf{n}$.

4.5 - Secondo la (4.11), che dà il valore minimo di E_b/N_0 in funzione dell'efficienza spettrale, quanto sarebbe giusto "pagare" in E_b/N_0 per passare da 1 a 1.5 bit per dimensione? *Commento*: si ricorderà che la modulazione 8PSK (non codificata) richiede 3.6 dB in più rispetto al 4PSK; è quindi più inefficiente della (già inefficiente) modulazione 4PSK.

Capitolo 5

Sistemi di trasmissione codificati

5.1 Introduzione

Nei tempi andati si distingueva tra codifica e modulazione, considerandole operazioni ben distinte. In un certo senso si chiamavano modulazione e codifica rispettivamente ciò che aveva a che fare con le forme d'onda e con le cifre d'informazione, solitamente binarie.

L'approccio tradizionale alla codifica di canale è il seguente. Si immagini di dover trasmettere un flusso di dati binari e di scegliere il sistema di modulazione, eventualmente multilivello, sulla base di ritmo di trasmissione, banda disponibile e probabilità d'errore. Il tutto può essere considerato un canale binario: da un lato entrano bit e dall'altro ne escono. Nel caso di dati particolarmente delicati se si vuole garantire una probabilità d'errore molto bassa, ad esempio 10^{-13}, non risulta conveniente aumentare il rapporto segnale-rumore: con la segnalazione binaria antipodale per passare da $P(E) = 10^{-5}$ a 10^{-13} occorrono circa 4.7 dB. Piuttosto si ricorre ad un *codice* per *proteggere* i dati dagli errori, con una delle due modalità seguenti: trasmissione a blocchi e rivelazione della presenza di errori, nel qual caso si scarta il blocco e se ne chiede la ritrasmissione; oppure identificazione e correzione degli errori (purché non troppo numerosi). Per ottenere questa protezione occorre aggiungere ai bit d'informazione dei *simboli ridondanti* ottenuti da quelli d'informazione secondo le regole di un codice. Si paga in riduzione del flusso utile di bit d'informazione, oppure a parità di questo in

aumento del ritmo sul canale e quindi della banda. "I codici espandono la banda" è un ben noto teorema popolare, che può tuttavia essere ambiguo come si vedrà in questo capitolo.

Nello schema a blocchi di tali sistemi di trasmissione sono quindi ben riconoscibili un *codificatore*, un *modulatore*, un *ricevitore* per forme d'onda e un *decodificatore*.

Si riconsideri però uno tra i più semplici esempi presentati nel Cap. 2: la forma d'onda trasmessa è

$$s_i(t) = \sum_{k=1}^{N} a_k g(t - kT) \tag{5.1}$$

dove le forme d'onda $g(t-kT)$ sono ortogonali e $a_k = \pm 1$; i primi $N-1$ simboli a_k possono essere scelti liberamente, mentre l'ultimo è vincolato dal *codice* che impone una regola di parità sui segni. Si può immaginare un codificatore in cui entrano blocchi di $N-1$ bit ed escono blocchi codificati di N bit, seguito da un modulatore che genera la forma d'onda (5.1). In ricezione però non si potrebbero mettere in cascata un ricevitore per la segnalazione binaria antipodale, che fornisce blocchi di N bit decisi, seguito da un decodificatore, che fa del suo meglio per individuare e correggere gli errori[1]. Il ricevitore ML, già visto nel Cap. 2, semplicemente considera gli $s_i(t)$ come un insieme di 2^{N-1} segnali in uno spazio a N dimensioni, e sceglie quello a distanza minima dal segnale ricevuto; ciò è tra l'altro molto facile, come si ricorderà.

Se anche si vogliono tenere ben distinte le due operazioni di codifica e modulazione, il ricevitore va invece considerato indivisibile. Tuttavia è preferibile pensare anche ai due blocchi in trasmissione come ad un modo semplice per generare uno tra 2^{N-1} segnali convenientemente disposti in uno spazio ad N dimensioni. Il codice è quindi semplicemente una regola per scegliere le coordinate del vettore trasmesso. Tutti i sistemi del Cap. 2 con prestazioni di un qualche interesse sono interpretabili in questo modo, come risulterà chiaro in questo capitolo.

[1] se il danno di prendere *decisioni indipendenti bit per bit* è già stato fatto, cioè se si sono prese decisioni *hard*, non resta di meglio che cercare tra le parole di codice quella che differisce nel minor numero possibile di posizioni dalla N-pla decisa; nel nostro semplice esempio basta però un solo errore per trovarsi imbarazzati a metà strada tra due parole di codice

5.2 Distanza geometrica e distanza di Hamming

Nell'esempio precedente si può anche osservare che, essendo le forme d'onda $g(t-kT)$ ortogonali, la distanza al quadrato tra due generici segnali $s_i(t)$ e $s_j(t)$ è data da

$$d_{ij}^2 = \sum_{k=1}^{N} (a_{ik} - a_{jk})^2 E_s = 4E_s d(i,j) \tag{5.2}$$

dove E_s è l'energia per simbolo, a_{ik} il k-esimo bit dell'i-esimo segnale e $d(i,j)$ il numero di componenti di segno opposto nelle due N-ple $\{a_{ik}\}$ e $\{a_{jk}\}$ $(k = 1, \ldots, N)$. Dunque la distanza geometrica (detta anche *euclidea*) d_{ij}^2 è proporzionale alla cosiddetta *distanza di Hamming* tra blocchi di bit. Il calcolo delle prestazioni mediante lo *union bound* richiede quindi solo la conoscenza delle distanze di Hamming tra le N-ple *parole* del codice, e le prestazioni asintotiche sono determinate dalla minima distanza di Hamming[2].

Inoltre il numero di dimensioni N dello spazio dei segnali è dato dalla lunghezza del blocco, e il numero M di segnali è pari al numero di parole di codice. In breve, lo studio del codice binario è praticamente equivalente all'indagine sulla geometria dei segnali.

È immediato verificare che le stesse considerazioni valgono per blocchi di simboli con modulazione 4PSK e con *mapping* di Gray (fig. 2.1): di nuovo basta contare i bit diversi per ottenere la distanza geometrica. Si noti invece che non è altrettanto conveniente la numerazione binaria naturale delle quattro possibili fasi (00, 01, 10 e 11, nell'ordine), perché la distanza geometrica non risulta funzione di quella di Hamming.

Per le costellazioni QAM, o comunque disposte su una griglia quadrata, vi è un semplice *mapping* che talvolta può risultare utile. Questo è ottenuto, in analogia con il caso 4PSK, accostando gruppi di 4 punti codificati rispettivamente con ..00, ..01, ..11, ..10 come mostrato in fig. 5.1 per il 16QAM.

Indicando con $2\sqrt{E}$ la distanza minima tra punti della costellazione, un codice che agisca solo sugli ultimi due bit ottiene $d_{ij}^2 = 4Ed(i,j)$, cioè il codice può essere scelto sulla base della sola distanza di Hamming.

Naturalmente punti come 0000 e 0100 non sono protetti in alcun modo dal codice. La loro distanza è però già abbastanza elevata, perlomeno per ottenere

[2] in tutto il capitolo d^2 indica una distanza geometrica al quadrato e d, o eventualmente d^H, una distanza di Hamming

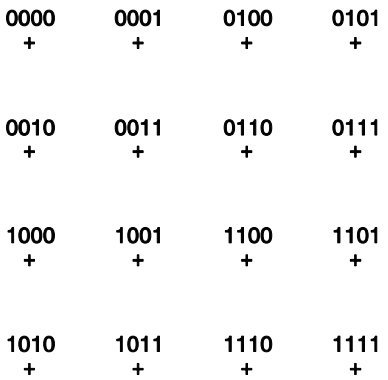

Fig. 5.1 - Costellazione 16QAM con *mapping* basato sulla distanza di Hamming

semplici sistemi codificati con prestazioni decorose[3] (es. 5.22,5.37).

Nelle costellazioni QAM non è possibile spingere oltre la proporzionalità tra distanza geometrica al quadrato e distanza di Hamming. Infatti una distanza di Hamming pari a 3 (tre bit diversi) non può corrispondere a coppie di punti della costellazione con $d^2 = 12E$, che non esistono.

Se si considera una costellazione M-PAM, detta ancora $2\sqrt{E}$ la distanza minima tra i punti, i valori possibili per d^2 sono $4E, 16E, \ldots$ per cui non c'è modo di avere d^2 proporzionale ad una qualche distanza di Hamming.

Anche nelle costellazioni 8PSK e 16PSK è impossibile avere d^2 proporzionale ad una distanza di Hamming. Ad esempio in una costellazione 8PSK (fig. 2.4 e 2.8) i valori possibili per d^2 sono $(2\sin\pi/8)^2 E_s = 0.586E_s$, $2E_s$, $3.414E_s$ e $4E_s$. È ben vero che con il *mapping* di Gray la distanza geometrica minima garantita risulta funzione della distanza di Hamming. Infatti è facile verificare che con 1, 2 e 3 bit di differenza d^2 vale rispettivamente almeno $0.586E_s$, $2E_s$ e $3.414E_s$. Questa non è però una proporzionalità diretta, e non è facile ricavarne codici efficienti.

Per tutte queste costellazioni esiste un diverso *mapping*, basato sulla cosiddetta tecnica del *set partitioning* che verrà presentata più avanti.

Nel seguito si considereranno dapprima i sistemi codificati in cui d^2 è proporzionale alla distanza di Hamming. Questi risultano più semplici perché

[3] l'osservazione che nelle costellazioni multilivello i punti sufficientemente lontani non richiedono protezione da parte di un codice è del tutto generale

in pratica si riducono al solo studio del codice binario. I sistemi codificati basati sul *set partitioning* verranno presentati successivamente.

5.3 Codici a blocco binari lineari

Si considererà dapprima il caso, concettualmente più semplice, di codifica binaria *a blocchi*: ad un blocco di K bit d'informazione viene associata una N-pla codificata ($N > K$). Ci si limiterà a trattare codici lineari, cioè codici in cui le cifre trasmesse sono ottenute come combinazioni lineari (in algebra binaria) di quelle d'informazione. Questi codici sono assai più semplici, ed offrono prestazioni di poco inferiori ai codici non lineari.

Se non vi sono motivi per fare diversamente si fanno coincidere le prime K cifre del blocco con quelle d'informazione. In tal caso il codice viene detto *sistematico*. Sui canali più comuni le prestazioni dei codici sistematici e non sistematici sono del tutto equivalenti.

Linearità significa che la somma di due N-ple parole di codice è essa stessa parola di codice. La somma è intesa bit a bit, ed ovviamente con algebra modulo 2 (OR esclusivo: 1+1=0).

Sommando due parole uguali si ottiene la N-pla di tutti zeri, che è quindi sempre una parola di codice. La distanza di Hamming tra due parole di codice, definita come il numero di componenti diverse, non è altro che il numero di uni nella loro somma, che è a sua volta una parola di codice. Ne deriva che la distanza minima di Hamming di un codice è pari al minimo numero di uni nelle parole del codice, esclusa naturalmente quella di tutti zeri.

Un'altra conseguenza molto importante, a cui si arriva facilmente, è che risulta indipendente dalla parola trasmessa non solo l'insieme delle distanze di Hamming tra questa e tutte le altre, ma anche la disposizione relativa, nello spazio dei segnali, del vettore trasmesso e dei $2^K - 1$ vettori concorrenti[4]. Per valutare la probabilità d'errore, perlomeno per canali simmetrici, si può quindi assumere che sia trasmessa la N-pla di tutti zeri, e vedere le parole non nulle come concorrenti di questa.

Se un codice a blocco è usato su un canale reso binario da decisioni *hard* simbolo per simbolo[5], si possono definire il *potere rivelatore* e *correttore* del codice. Un codice con distanza minima d rivela ogni configurazione di $d - 1$, o meno, bit errati nel blocco; infatti questi non possono produrre una parola

[4] ciò vale a meno di cambiamenti di verso degli assi, cioè di simmetrie speculari

[5] le decisioni non *hard* sono dette *soft*

di codice. Lo stesso codice garantisce la correzione di un numero di bit errati fino a $t = (d-1)/2$ se d è dispari e $t = d/2 - 1$ se d è pari; infatti con un tale numero di errori la parola ricevuta è ancora più vicina a quella trasmessa che alle concorrenti.

La costruzione di codici a blocco efficienti è tutt'altro che banale, e solo dei più semplici si riesce a dare una giustificazione intuitiva. Di norma occorrono approfondite nozioni di algebra, in particolare dei *campi di Galois*, cioè dei campi con un numero finito di elementi. Ci si limiterà pertanto ad una descrizione che consenta di usare questi codici, senza alcuna velleità di capire come nascano le loro proprietà.

Dei codici lineari a blocco è possibile dare una descrizione elementare, ma di scarsa utilità pratica, mediante le cosiddette matrici *generatrice* e *di parità* per le quali si rimanda eventualmente a libri sui codici.

Per rappresentare la N-pla di bit della parola di codice risulta invece particolarmente conveniente utilizzare polinomi di grado $N-1$ in una variabile che viene indicata con x, z, D o altro, secondo i gusti[6]. Gli N coefficienti corrispondono ai bit, e possono valere solo zero o uno: la presenza o l'assenza, nel polinomio, di un certo monomio indica che il bit corrispondente è uno o è zero. Resta da intendersi sulla corrispondenza tra elementi della N-pla e coefficienti del polinomio. La convenzione più diffusa, e che verrà utilizzata nel seguito, è $c_{N-1}D^{N-1} + \ldots + c_2D^2 + c_1D + c_0$; altri però preferiscono numerare i coefficienti da 1 a N, e quindi trovano comodo indicare il polinomio con $c_1D^{N-1} + \ldots + c_{N-2}D^2 + c_{N-1}D + c_N$. In entrambi i casi si assume che il termine di grado più elevato corrisponda al primo bit trasmesso in ordine di tempo, cioè che i bit siano trasmessi rispettivamente nell'ordine $c_{N-1}, \ldots, c_0$ oppure $c_1, \ldots, c_N$.

Con questa notazione, che potrebbe apparire artificiale e faticosa, un codice è individuato da un *polinomio generatore* $g(D)$, di grado $N-K$: sono parole di codice, per definizione, tutte e sole quelle corrispondenti ai polinomi divisibili per $g(D)$. Le 2^K parole di codice possono essere generate moltiplicando per $g(D)$ un generico polinomio di grado minore o uguale a $K-1$ (ce ne sono effettivamente 2^K), ottenendo polinomi di grado minore o uguale a $N-1$. Detto altrimenti, $g(D)$ e le sue repliche traslate sono prototipi di parole di codice, e tutte le parole di codice sono combinazioni lineari di queste.

La circuiteria richiesta è molto semplice. Poiché moltiplicare due trasfor-

[6] il polinomio non è altro che la trasformata zeta della N-pla; x è la prima variabile che viene in mente, z ricorda la trasformata zeta, D sta per *Delay*

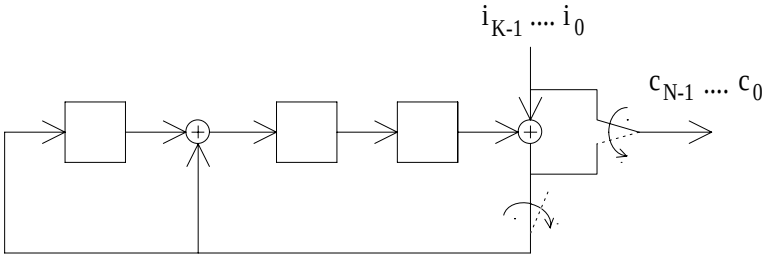

Fig. 5.2 - Codificatore per il codice con $g(D) = D^3 + D + 1$

mate zeta equivale a convolvere due sequenze, il codificatore è un filtro FIR con $N - K + 1$ coefficienti binari, che esegue la convoluzione, naturalmente con algebra binaria, delle cifre d'informazione $i_{K-1}, \ldots, i_0$ con $g_{N-K}, \ldots, g_0$. Il codificatore richiede quindi solo sommatori modulo 2, in numero pari ai coefficienti non nulli di $g(D)$. Il codice così ottenuto non è sistematico.

Se invece si vuole che il codice sia sistematico occorre trasmettere dapprima le cifre d'informazione $i_{K-1}, \ldots, i_0$, cioè il polinomio $i(D)D^{N-K}$, dove la moltiplicazione per D^{N-K} ha il solo compito di disporre i bit d'informazione al tempo giusto. Il polinomio $i(D)D^{N-K}$ ha $N-K$ zeri nelle ultime posizioni, dove vengono poste le rimanenti cifre, dette *di parità*. Queste sono scelte in modo che il polinomio trasmesso sia una parola di codice valida, cioè sia divisibile per $g(D)$. Si ottiene questo risultato sottraendo da $i(D)D^{N-K}$ il resto della divisione di questo stesso polinomio per $g(D)$; se infatti una divisione dà resto basta sottrarlo dal dividendo, e non vi sarà più resto:

$$c(D) = i(D)D^{N-K} - R_{g(D)}[i(D)D^{N-K}] \tag{5.3}$$

dove $R_{g(D)}[\cdot]$ indica il resto della divisione per $g(D)$, che al massimo ha grado $N - K - 1$ come desiderato. In algebra binaria la differenza è poi uguale alla somma. Si noti, per inciso, che il quoziente della divisione per $g(D)$ non è di alcun interesse.

In fig. 5.2 è mostrato, ad esempio, il codificatore per $g(D) = D^3 + D + 1$. Inizialmente i registri sono azzerati. Seguendo passo passo, con un po' di pazienza, le operazioni che si farebbero manualmente per dividere due polinomi e quelle effettuate dal circuito di fig. 5.2 non è difficile vedere che ogni colpo di *clock* equivale a un passo della divisione e che dopo K *clock*

negli $N - K$ registri è contenuto il resto[7]. Basta quindi aprire l'anello di retroazione (la divisione è finita!) e inviare sul canale il contenuto dei registri. La complessità del codificatore è praticamente la stessa della versione non sistematica. Ovviamente anche le 2^K parole codificate sono le stesse; cambia semplicemente il modo di associarle ai K bit d'informazione.

Se poi per qualche motivo i valori di N e K di un certo codice non risultassero convenienti, il codice può essere *accorciato*. L'operazione, banale, consiste nel porre a zero i bit d'informazione nelle prime b posizioni. Naturalmente non si trasmettono questi dati, e si ottiene quindi un codice con $K' = K - b$ ed $N' = N - b$.

Nella maggior parte dei casi i codici (non accorciati) di un qualche interesse sono *ciclici*, o semplici modificazioni di codici ciclici. Il nome deriva dal fatto che una parola di codice ruotata ciclicamente (ad esempio $c_0, c_{N-1}, \ldots, c_2, c_1$) è anch'essa parola di codice. Si dimostra facilmente che in tal caso $g(D)$ è un divisore di $D^N - 1$.

Alcune classi di codici a blocco (anche banali) sono:

- *codice universo*: una N-pla non codificata! $K = N$; $d = 1$
- *codice a parità semplice*: $K = N - 1$ cifre d'informazione, con N qualsiasi, e una cifra di parità somma di quelle d'informazione di modo che il numero complessivo di uni sia pari. Il polinomio generatore è $g(D) = D + 1$, e $d = 2$ come è facile verificare.
- *codice a ripetizione*: una sola cifra d'informazione, ripetuta N volte; $K = 1$; $d = N$
- *codici di Hamming*: classe infinita di codici, con coppie di valori di N e K che soddisfano la condizione $N = 2^{N-K} - 1$: (7,4), (15,11), (31,26), (63,57), (127,120), e così via. I corrispondenti polinomi generatori possono essere (ne esiste più d'uno) $D^3 + D + 1$, $D^4 + D + 1$, $D^5 + D^2 + 1$, $D^6 + D + 1$, $D^7 + D^3 + 1$, ... La distanza minima d è però sempre pari a 3, per cui i codici con N grande hanno scarso interesse, se non su canali poco rumorosi. Anche quelli con N piccolo non sono molto interessanti perché troppo semplici; infatti occupano un piccolo numero di dimensioni. I codici di Hamming possono comunque avere un ruolo

[7] chi volesse *dimostrare* questo risultato consideri dapprima il caso di un solo bit d'informazione i_{K-m} ($m = 1, \ldots, K$) diverso da zero; invochi poi, per il contenuto finale di ciascuno dei registri, la sovrapposizione degli effetti

nella costruzione di codici più elaborati. Del principio su cui sono basati questi codici esiste una spiegazione elementare, che si può trovare su tutti i libri di codici.

- *codici di Hamming estesi*: aggiungendo ad un qualsiasi codice lineare un bit di parità complessiva, senza cambiare K, si ottiene un codice (non ciclico) con distanza d pari, e ovviamente non diminuita (es. 5.1), e quindi 4 nel caso dei codici di Hamming. I valori di N e K sono (8,4), (16,11), (32,26), ecc.

- *codici BCH* (*Bose, Ray-Chaudhuri, Hocquenghem*): la classe più famosa di codici ciclici, che include come caso particolare i codici di Hamming. Incomprensibili senza algebra dei campi di Galois. La tab. 1 mostra alcune combinazioni di $N = 2^m - 1$, K, d e $g(D)$. Per risparmiare spazio si usa dare i coefficienti del polinomio generatore in ottale, o talvolta in esadecimale, anziché in binario. In prima approssimazione, ma con diverse eccezioni, ogni aumento di 2 della distanza, cioè ogni aumento di una unità del potere correttore, costa m cifre di parità.

- *codici BCH non primitivi*: codici ciclici, con N divisore di $2^m - 1$; solitamente le tabelle includono solo quelli che risultano migliori dei normali BCH, accorciati alla stessa lunghezza N. Quello di gran lunga più famoso è il codice di *Golay*, con $N = (2^{11} - 1)/89 = 23$, $K = 12$, $d = 7$ e generatore (ottale) 5343 oppure 6165, che ha la particolarità di essere l'unico codice correttore di tutti e soli gli errori semplici, doppi e tripli.

- *codice di Golay esteso*: appendendo una cifra di parità complessiva si ottiene un codice (non ciclico) con $N = 24$, $K = 12$ e $d = 8$.

- *codici Reed-Muller*: classe infinita, estensione di codici ciclici. Fissato $N = 2^m$, il codice di *ordine* r ha $K = 1 + \binom{m}{1} + \binom{m}{2} + \cdots + \binom{m}{r}$ e distanza minima $d = 2^{m-r}$. Il caso $m - r = 2$ risulta equivalente ai codici di Hamming estesi. Il caso $r = 1$ corrisponde a segnali biortogonali. I codici con N non molto grande (qualche decina) sono molto buoni, ed anche decodificabili in modo efficiente. Con valori elevati di N sono meno interessanti. Per i polinomi generatori si rimanda ai testi sui codici.

N	K	d	$g(D)$ (ottale)
15	11	3	23
	7	5	721
	5	7	2467
	1	15	77777
31	26	3	45
	21	5	3551
	16	7	107657
	11	11	5423325
	...	...	...
63	57	3	103
	51	5	12471
	45	7	1701317
	39	9	...
	36	11	...
	...	...	...

N	K	d	$g(D)$ (ottale)
127	120	3	211
	113	5	41567
	106	7	11554743
	99	9	...
	92	11	...
	...	...	...
255	247	3	435
	239	5	267543
	231	7	156720655
	223	9	...
	215	11	...
	207	13	...
	199	15	...
	191	17	...
	...	...	...

Tab. 1 - Alcuni codici BCH

5.4 Prestazioni dei codici a blocco lineari

Nel caso di ricevitore ML le prestazioni sono valutabili, con la solita approssimazione data dallo *union bound*, conoscendo la struttura del codice e in particolare la distanza minima. Ad alto rapporto segnale-rumore, se d è la distanza minima di Hamming il termine dominante è dato da $Q(\sqrt{\frac{2E_s d}{N_0}})$, ovvero $Q(\sqrt{\frac{2E_b}{N_0}\frac{K}{N}d})$, moltiplicato per il numero di concorrenti a distanza minima. Se si vuole $P_b(E)$ occorre moltiplicare per il rapporto tra il numero d di bit errati e il numero N di bit nel blocco.

In non pochi casi si fa l'operazione, pur concettualmente criticabile, di prendere decisioni binarie simbolo per simbolo, per poi cercare di correggere gli errori individuandoli sulla base del resto della divisione della N-pla decisa per $g(D)$, detto *sindrome*. Naturalmente la sindrome, essendo un polinomio di grado $N-K-1$, può assumere solo 2^{N-K} configurazioni mentre le configurazioni d'errore possibili sono 2^N, includendo anche il caso di assenza di errori. Ogni sindrome corrisponde a 2^K ipotesi (es. 5.6), fra le quali si dovrà

scegliere quella a distanza minima. Salvo casi elementari, di poco interesse, di norma per comprendere la decodifica *algebrica* occorrono solide nozioni di algebra.

Se comunque t è il numero massimo di errori correggibili, il termine dominante della probabilità d'errore è la probabilità che nel blocco vi siano $t+1$ errori, data da $\binom{N}{t+1} p^{t+1}(1-p)^{N-t-1}$, dove $p = Q(\sqrt{\frac{2E_b}{N_0}\frac{K}{N}})$ è la probabilità d'errore sul canale binario. Con qualche grossolana approssimazione (es. 5.8) si ottiene una probabilità d'errore proporzionale a $Q\big(\sqrt{\frac{2E_b}{N_0}\frac{K}{N}(t+1)}\big)$. In pratica si è osservato che la perdita rispetto al ricevitore ML risulta intorno ai 2 dB. Naturalmente si usano codici molto potenti, e con valori di K fino ad alcune migliaia, per i quali il ricevitore ML non è assolutamente realizzabile mentre la decodifica algebrica è ancora possibile[8]. La maggior complessità del codice recupera almeno in parte la perdita dovuta al ricevitore non ottimo.

In taluni casi il codice viene utilizzato solo come *rivelatore* di errori. Dopo le decisioni bit per bit si verifica mediante la semplice divisione per $g(D)$ se il blocco deciso corrisponde ad una parola di codice, nel qual caso viene accettato, oppure no (e viene scartato, chiedendo la ritrasmissione). Il pericolo è che gli errori sul canale siano almeno d ed in posizioni opportune (inopportune, a dire il vero) in modo da trasformare una parola di codice in una diversa parola di codice. In primissima approssimazione si può mostrare che ciò avviene con probabilità proporzionale a $Q(\sqrt{\frac{2E_b}{N_0}\frac{K}{N}d})$.

5.5 Codici convoluzionali

Come insegna la teoria della capacità di canale, i codici efficienti richiedono un numero elevato di dimensioni dello spazio dei segnali. In un certo senso le difficoltà nel decodificare in modo ottimo i codici a blocco derivano dal fatto che parole di codice a piccola distanza da quella trasmessa possono trovarsi ovunque nello spazio. Si considerino ad esempio le parole di codice con tre uni, dirette concorrenti della parola nulla in un semplice codice di Hamming. Fissata la posizione di uno dei tre uni si verifica che un altro può trovarsi in tutte le altre posizioni, e che anche il terzo è distribuito uniformemente nel blocco, apparentemente in modo indipendente dai primi due (ovviamente

[8] è buffo pensare che un codificatore con $K = 1000$ può generare $2^{1000} \approx 10^{300}$ parole diverse, ma in tutta la sua vita ne produrrà una frazione infinitesima

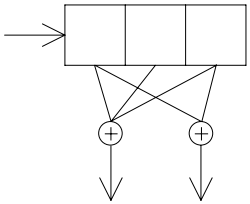

Fig. 5.3 - Semplice codificatore convoluzionale

invece c'è una regola). Occorre trovare un qualche modo per classificare le parole di codice in modo da limitare l'intorno da esplorare. Una regola generale esiste, ma è utilizzabile in pratica solo per codici con un piccolo numero di cifre di parità[9]. Il fatto straordinario è che non è invece limitante la lunghezza del blocco, come risulterà chiaro tra breve per i codici convoluzionali per i quali tale lunghezza non ha limiti.

La struttura di un semplice codificatore convoluzionale è mostrata in fig. 5.3. La lunghezza K del registro a scorrimento ($K = 3$ nell'esempio) è detta *constraint length*. All'inizio i registri sono azzerati. Entra un bit d'informazione per volta e ne escono due, combinazioni lineari del bit attuale e di alcuni precedenti (due, in figura). I bit codificati sono convoluzioni della sequenza in ingresso con due diverse risposte impulsive; da ciò deriva il nome di questa classe di codici, detti talvolta anche *ricorrenti*. Per quanto lunga sia la sequenza in ingresso, un generico bit d'informazione ha effetto sui bit codificati solo in un intervallo molto limitato, contrariamente ai codici a blocco. A parità di distanza, ciò produce un insieme di segnali un po' meno denso (es. 5.13,5.14) ma consente la decodifica ML.

Conviene distinguere, nel contenuto del registro a scorrimento di lunghezza $K = 3$, una parte riservata al bit attuale ed una contenente i $K - 1 = 2$ bit del passato ancora memorizzati, che costituiscono lo *stato* del codificatore. Sono possibili 2^{K-1} stati. Le uscite attuali dipendono sia dallo stato sia dal dato attuale. Lo stato al passo successivo, corrispondente all'ingresso di un nuovo bit d'informazione, è anch'esso funzione dello stato e del dato attuale.

Le *transizioni di stato* possibili per i primi passi sono indicate in fig. 5.4, dove gli stati sono enumerati secondo il contenuto della coppia di registri

[9] verrà presentata più avanti, perché più facilmente comprensibile dopo i codici convoluzionali; fu scoperta infatti per analogia con questi

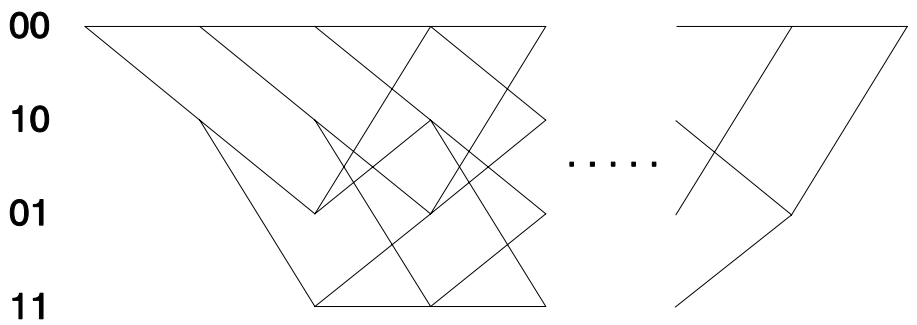

Fig. 5.4 - Traliccio delle transizioni di stato del codice convoluzionale di fig. 5.3

di memoria, nell'ordine in cui appaiono in fig. 5.3. Si osservi che da ogni stato se ne possono raggiungere solo due, corrispondenti ai due valori del bit d'informazione che si sposta dalla prima alla seconda cella, così come ogni stato è raggiungibile solo da due, poiché viene scartato un solo bit dal registro a scorrimento. In 2 passi ($K-1$, in generale) è possibile raggiungere qualsiasi stato, dopo di che il diagramma di fig. 5.4, detto *traliccio* (*trellis*), si ripete invariato. Quando eventualmente la sequenza di bit d'informazione termina, risulta conveniente per motivi che si vedranno in seguito forzare lo stato finale ad uno noto, ad esempio quello nullo, terminando la sequenza con $K-1$ zeri.

Si noti che le transizioni di stato, mostrate in fig. 5.4, dipendono solo dalla particolare struttura a registro a scorrimento, e non dalle connessioni tra celle del registro e sommatori modulo 2. Inoltre ciascuna transizione di stato individua non solo il contenuto attuale delle celle di memoria (stato iniziale), ma anche il bit d'informazione attuale (che è dato dal primo bit dello stato finale). Ogni sequenza di bit d'informazione è in corrispondenza biunivoca con una successione di transizioni di stato, cioè con un percorso nel traliccio. Visivamente risulta comodo osservare che ad ogni biforcazione il percorso superiore corrisponde all'ingresso di uno zero e quello inferiore ad un uno.

I bit codificati, e di conseguenza la forma d'onda trasmessa, dipendono sia dal contenuto dei registri sia dalle connessioni con i sommatori modulo 2. Per tutti i valori di K di interesse pratico la ricerca di buoni codici è stata effettuata in modo esaustivo provando tutte le possibili connessioni, con

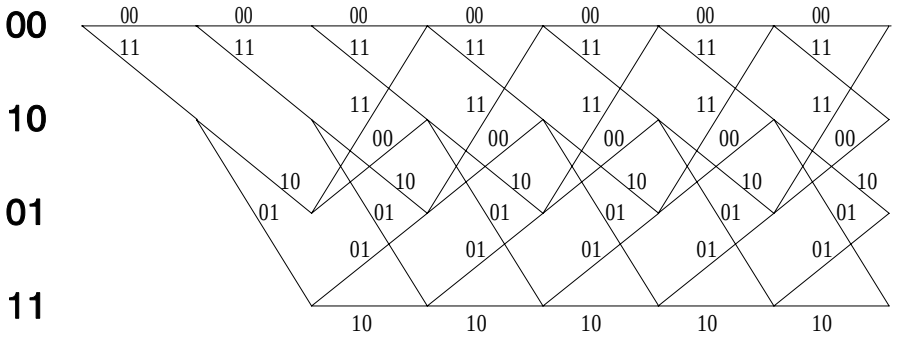

Fig. 5.5 - Traliccio del codificatore di fig. 5.3; sono mostrati anche i bit codificati

qualche regola per scartare al più presto le cattive soluzioni. I risultati sono tabulati sui libri di codici.

Il codice di fig. 5.3 è, ad esempio, il migliore fra quelli a quattro stati[10]. In fig. 5.5 è riprodotto il traliccio, con l'aggiunta delle coppie di bit codificati corrispondenti a ciascuna transizione di stato (già per otto stati la figura risulterebbe quasi illeggibile, e converrebbe sostituirla con una tabella). Il traliccio mostra tutti i possibili segnali, che sono 2^L se L è il numero dei passi, e quindi le coordinate del vettore trasmesso: ad esempio con segnalazione binaria antipodale 1 corrisponde a $\sqrt{E_s}$ e 0 a $-\sqrt{E_s}$. Si noti che il codice non è sistematico. Contrariamente ai codici a blocco, i codici convoluzionali non sistematici sono generalmente migliori dei sistematici.

Il ricevitore a massima verosimiglianza richiede il calcolo delle correlazioni del vettore ricevuto con tutti i possibili segnali. Una generica correlazione può essere calcolata sommando via via le coppie di termini corrispondenti a ciascuna transizione di stato. Nel primo intervallo occorre calcolare le due correlazioni $-r_1 - r_2$ e $r_1 + r_2$; nel secondo occorre aggiungere a queste, rispettivamente, $-r_3 - r_4$, $r_3 + r_4$, $r_3 - r_4$ e $-r_3 + r_4$; nel terzo intervallo le quattro ipotesi finora considerate si biforcano ancora, ed alle correlazioni già calcolate si deve sommare una delle quattro combinazioni $\pm r_5 \pm r_6$, secondo le indicazioni del traliccio. La novità è che le otto strade convergono a due a due in uno stesso stato, ed è quindi possibile scartare definitivamente la peggiore

[10] una variante banale è scambiare i due bit codificati

delle due mantenendo memoria solo di quella *sopravvissuta* al confronto (una per ciascuno stato; si memorizza il percorso ed il valore della correlazione).

Al passo successivo le quattro strade sopravvissute si biforcano nuovamente, ma di nuovo si incontrano a due a due, per cui si potrà scartare allo stesso modo un'altra metà delle ipotesi. In definitiva occorrerà effettuare ad ogni passo quattro aggiornamenti e confronti, e conservare le informazioni relative ai quattro percorsi sopravvissuti.

È questo il famosissimo *algoritmo di Viterbi*, proposto verso la fine degli anni '60 per i codici convoluzionali ma che risulta utile anche in altri contesti, come si vedrà in seguito. La complessità è proporzionale al numero 2^{K-1} degli stati, e cresce solo linearmente con la lunghezza L della sequenza. In altri termini nel funzionamento in tempo reale basta che nel tempo di un bit d'informazione vengano effettuate tutte le operazioni di somma, confronto e memorizzazione per tutti gli stati[11].

Il codice a quattro stati risulta fin troppo semplice in pratica; molto usato è quello a 64 stati, per il quale sono disponibili circuiti integrati estremamente economici che effettuano tutte le operazioni richieste, fino a velocità di decine di Mb/s. Naturalmente i livelli ricevuti r_k sono rappresentati con precisione finita, cioè sono quantizzati. Si scopre che un bit per il segno e due o tre per la parte frazionaria sono largamente sufficienti.

Unico inconveniente, se L non è piccolo, è la dimensione della memoria per i cammini sopravvissuti. Peraltro risulta difficile credere che si debba attendere la fine della trasmissione per decidere sui primi bit d'informazione. Ed infatti avviene il seguente fenomeno: ad un generico istante, con elevata probabilità i 2^{K-1} cammini sopravvissuti coincidono fino a $4 \div 5K$ passi precedenti. Il decodificatore può quindi già annunciare le relative decisioni, dopo di che è inutile mantenere in memoria la parte già decisa. Ad ogni passo sono memorizzati 2^{K-1} cammini sopravvissuti per una lunghezza pari ad esempio a $5K$ transizioni di stato, e viene emessa la decisione relativa al bit che sta uscendo dalla memoria disponibile[12].

Quando la trasmissione ha termine si è costretti a prendere una decisione anche sugli ultimi bit d'informazione, che risulterebbero quindi meno affidabili. È questo il motivo per cui si preferisce avere uno stato finale certo, ottenuto

[11] la struttura a farfalle del traliccio consente una forte parallelizzazione, con unità elementari che trattano coppie di stati

[12] occasionalmente i 2^{K-1} percorsi sopravvissuti non coincidono, ma si è forzati comunque a prendere una decisione; ad esempio si potrà scegliere il bit d'informazione corrispondente alla sequenza sopravvissuta con la miglior correlazione al momento attuale

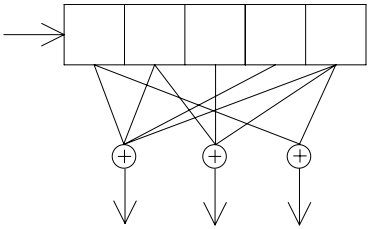

Fig. 5.6 - Codificatore convoluzionale con *rate* $R = 2/3$ a otto stati; i bit entrano a coppie

forzando a zero il contenuto finale dei registri, e quindi un solo cammino sopravvissuto da considerare.

Una volta capito l'algoritmo di Viterbi risulta facile vedere le modifiche richieste con altri valori dei parametri del codice. Una modifica banale è variare il numero di bit di codice per bit d'informazione, portandolo ad esempio a tre: il codificatore ha tre sommatori, anziché due; ogni ramo nel traliccio corrisponde ad una terna di bit, e le correlazioni vengono aggiornate sommando tre contributi. Il rapporto tra bit d'informazione e bit codificati (pari a 1/3, nel caso in esame) viene detto il *rate* del codice, e solitamente indicato con R[13]. *Rate* è un termine difficilmente traducibile (ritmo? velocità?), e nel seguito non si tenterà di farlo.

Più interessante è il caso di *rate* con numeratore $b \neq 1$, ad esempio $R = 2/3$. La soluzione più intuitiva è far entrare i bit d'informazione nel registro a scorrimento a due per volta, ed avere tre sommatori, come in fig. 5.6 che rappresenta un codificatore a $2^{K-b} = 8$ stati[14]. La novità è che ad ogni transizione di stato si hanno due bit freschi, e quindi da ogni stato se ne possono raggiungere quattro ed in ogni stato se ne ricongiungono quattro. L'algoritmo di Viterbi esegue dunque confronti tra quattro ipotesi, e mantiene solo la migliore. Tanto meglio, verrebbe da dire, si sfoltisce maggiormente; però i confronti tra quattro concorrenti sono un po' più fastidiosi di quelli tra due. Ed infatti è stata anche proposta una diversa soluzione, basata sulla

[13] usare, come è consuetudine, lo stesso simbolo R per indicare sia il ritmo di trasmissione (in bit/s) sia il *rate* di un codice non è l'ideale; generalmente il contesto elimina l'ambiguità

[14] osservando che i bit d'informazione *pari* e *dispari* vanno in celle distinte del registro a scorrimento, il codificatore può essere rappresentato in modo equivalente con due registri di tre e due celle, alimentati in parallelo

perforazione di codici con *rate* 1/2 (codici *punctured*[15]): una volta si prelevano entrambi i bit codificati, ed una volta uno solo, periodicamente; in totale si sono trasmessi tre bit ogni due d'informazione. Per ottenere $R = 3/4$ basta considerare prima due bit, poi uno, poi ancora uno. In ricezione basta porre $r_k = 0$ in corrispondenza dei bit mancanti, esattamente a metà strada tra i livelli nominali, ed utilizzare il normale algoritmo di Viterbi. Particolarmente apprezzati sono i codici *punctured* universali, cioè che offrono buone prestazioni per molti valori di R ad esempio da 1/2 a 7/8, modificando solo le regole di perforazione[16].

E quali bit è meglio eliminare? L'analisi esaustiva di tutte le possibili perforazioni conduce alle tabelle dei migliori tra tali codici.

Le prestazioni dei codici convoluzionali *punctured* sono quasi equivalenti a quelle dei migliori codici, a parità di numero di stati.

5.6 Prestazioni dei codici convoluzionali

Una valutazione approssimata delle prestazioni dei codici convoluzionali può essere ottenuta mediante il solito *union bound*. Se non si limita a priori la durata della trasmissione è evidente che prima o poi qualche errore verrà commesso, e quindi la probabilità d'errore è pari a uno (risultato assolutamente inespressivo sulla qualità del collegamento). Occorre definire in modo conveniente *eventi errore* di cui sia facile valutare la probabilità. In particolare si trova conveniente calcolare la probabilità che ad un generico istante il percorso scelto dal ricevitore *inizi a divergere* da quello corretto, cioè la frequenza con cui il ricevitore devia dalla retta via. Un evento errore è concluso, per definizione, quando si ricongiunge *per la prima volta* al percorso corretto. Si noti che se si volesse invece valutare la probabilità che ad un generico istante la transizione di stato decisa sia errata si dovrebbe tener conto anche degli eventi errore già in corso.

Occorre enumerare i possibili eventi errore, e la distanza geometrica (proporzionale a quella di Hamming) tra il segnale corretto e ciascun concorrente. La linearità del codice consente di assumere che sia stata trasmessa una

[15] l'autore di queste note li chiama anche (liberamente) codici *azzoppati*

[16] però gli eventi errore dei codici *punctured* sono decisamente più lunghi, perché una distanza consistente viene accumulata solo su un gran numero di simboli; nel decodificatore occorre una memoria per i percorsi sopravvissuti pari ad alcune volte quella richiesta dal codice con *rate* 1/2

particolare sequenza, ad esempio quella di tutti zeri. Un evento errore deve terminare nello stato $0 \ldots 0$ e quindi non può avere lunghezza minore di K, mentre non c'è limite alla lunghezza massima. Per codici semplici, come il quattro stati di cui si è visto il traliccio, l'ispezione si può fare manualmente. Ad esempio si individuano in fig. 5.5 un percorso errato di lunghezza 3 con distanza di Hamming pari a 5, due con distanza pari a 6 e lunghezza rispettivamente 4 e 5, quattro con distanza 7 e lunghezze comprese tra 5 e 7, e così via. Ricordando che $E_s = E_b R$, si ha che la probabilità dell'inizio di un evento errore è maggiorata dallo *union bound*

$$\begin{aligned} P(E) &\leq Q\Big(\sqrt{\frac{2E_b}{N_0}5R}\Big) + 2Q\Big(\sqrt{\frac{2E_b}{N_0}6R}\Big) + 4Q\Big(\sqrt{\frac{2E_b}{N_0}7R}\Big) + \ldots = \\ &= \sum_d a(d) Q\Big(\sqrt{\frac{2E_b}{N_0}dR}\Big) \end{aligned} \tag{5.4}$$

dove $a(d)$ è il numero di eventi errore a distanza d. Ad alto rapporto segnale-rumore è particolarmente importante la distanza minima del codice. Questa, in generale, non corrisponde all'evento errore di lunghezza minore e va quindi ricercata senza porre limiti a tale lunghezza. Tale distanza è detta *distanza libera (free distance)* ed indicata con d_f.

Il guadagno asintotico del codice rispetto alla trasmissione binaria antipodale non codificata, ignorando come al solito i coefficienti moltiplicativi all'esterno della funzione $Q(\cdot)$, è evidentemente dato da

$$G = d_f R \tag{5.5}$$

ed è pari a $5/2$ (4 dB) nel caso in esame. Naturalmente solo una parte di questo guadagno è da attribuire al codice; la parte restante all'espansione di banda. Infatti secondo quanto trovato nel capitolo precedente la sola espansione di banda corrispondente alla trasmissione di $1/2$ bit per dimensione anziché 1 bit "vale" circa 1.8 dB.

Se si vuol valutare la probabilità che i bit d'informazione siano errati occorre pesare la probabilità di ogni evento errore con il numero di bit d'informazione errati. Ad esempio si vede che con l'evento errore a distanza minima si sbaglia un solo bit d'informazione (i bit d'informazione decodificati sono 100 anziché 000); con ciascuno dei due a distanza 6 si hanno due bit

errati, e così via. La probabilità d'errore sui bit è maggiorata da

$$\begin{aligned} P_b(E) &\le Q\Big(\sqrt{\frac{2E_b}{N_0}5R}\Big) + 2Q\Big(\sqrt{\frac{2E_b}{N_0}6R}\Big) + 2Q\Big(\sqrt{\frac{2E_b}{N_0}6R}\Big) + \ldots = \\ &= \sum_d \sum_i i\, n(d,i) Q\Big(\sqrt{\frac{2E_b}{N_0}dR}\Big) = \sum_d w(d) Q\Big(\sqrt{\frac{2E_b}{N_0}dR}\Big) \end{aligned} \tag{5.6}$$

dove $n(d,i)$ è il numero di eventi errore a distanza d e con i bit d'informazione errati, e $w(d) = \sum i\, n(d,i)$ è il numero complessivo di bit d'informazione errati negli eventi errori aventi distanza d. La formula tiene implicitamente conto non solo degli eventi errore che iniziano ad un generico istante, ma anche di quelli già in corso. Infatti la frequenza con cui *inizia* un evento errore viene moltiplicata per il numero *complessivo* di bit d'informazione errati che esso produce, e non solo per quello prodotto nella prima transizione di stato. Se il numeratore b del *rate* del codice è diverso da uno occorre anche tener conto del numero di bit trasmessi per ciascuna transizione di stato e si ha

$$P_b(E) \le \frac{1}{b} \sum_d w(d) Q\Big(\sqrt{\frac{2E_b}{N_0}dR}\Big) \tag{5.7}$$

Il lettore non del tutto convinto da tali formule (che effettivamente la prima volta danno da pensare) immagini un evento errore con probabilità P, che quindi abbia inizio mediamente ogni $1/P$ passi nel traliccio, e che provochi w bit d'informazione errati. In L passi, con L molto grande, si trasmettono Lb bit d'informazione e se ne sbagliano $w\,LP$, in media. Il contributo dell'evento errore a $P_b(E)$ è quindi wP/b. La maggiorazione (5.7) non è altro che la somma estesa a tutti i possibili eventi errore.

Una formula analoga vale per i codici *punctured*, con l'avvertenza di mediare la probabilità degli eventi errore su più transizioni di stato (es. 5.12).

Le tab. 2 e 3 danno un'idea dei valori di distanza d_f ottenibili con codici convoluzionali con *rate* $R = 1/2$ e $1/3$. Sono indicati anche il numero complessivo $w(d_f)$ di bit d'informazione errati negli eventi errore a distanza minima, per valutare almeno il termine dominante della (5.6), e la configurazione dei sommatori, in ottale. Si noti che il codice con $R = 1/2$ e 64 stati, utilizzato molto spesso, ha un guadagno asintotico di ben 7 dB.

Tra i codici *punctured*, limitandosi per semplicità a quelli derivati dall'ottimo codice con *rate* $R = 1/2$ a 64 stati, si trovano ad esempio quelli con *rate*

K	numero di stati	d_f	$w(d_f)$	generatori
3	4	5	1	7,5
4	8	6	2	17,15
5	16	7	4	35,23
6	32	8	2	75,53
7	64	10	36	171,133
8	128	10	2	371,247

Tab. 2 - Codici convoluzionali con *rate* $R = 1/2$

K	numero di stati	d_f	$w(d_f)$	generatori
3	4	8	3	7,7,5
4	8	10	6	17,15,13
5	16	12	12	37,33,25
6	32	13	1	75,53,47
7	64	15	7	171,165,133
8	128	16	1	367,331,225

Tab. 3 - Codici convoluzionali con *rate* $R = 1/3$

$R = 2/3$, $3/4$ e $7/8$ che hanno $d_f = 6$, 5 e 3, rispettivamente. Anche l'ultimo dà un guadagno asintotico degno di nota (4.2 dB).

L'enumerazione degli eventi errore, facile per l'esempio a quattro stati, diventa faticosa nei casi veramente interessanti e deve essere meccanizzata. Ad esempio esplorando in modo esaustivo il traliccio del codice con $R = 1/2$ e 64 stati[17] si trovano 11 eventi errore a distanza 10 con lunghezza compresa tra 7 e 16, che contribuiscono un totale di 36 bit d'informazione errati, 38 a distanza 12 (lunghezza tra 9 e 24; 211 bit errati), 193 a distanza 14 (lunghezza tra 10 e 28; 1404 bit errati) e così via.

[17] gli eventi errore sono infiniti; ovviamente si scartano quelli che già ad una certa profondità nel traliccio superano la massima distanza che interessa

5.7 Funzione di trasferimento

Un modo elegante per enumerare gli eventi errore è la *funzione di trasferimento*. Per individuare i percorsi nel traliccio che divergono inizialmente dalla sequenza di tutti zeri, e vi si ricongiungono *per la prima volta* dopo un numero imprecisato di passi, basta cancellare nel primo passo del traliccio la connessione tra $0\ldots0$ e $0\ldots0$, che consentirebbe di *non* dare inizio all'evento errore, e nei passi successivi tra $0\ldots0$ e uno stato qualsiasi, che consentirebbe di prolungare un evento errore già terminato. Basta poi associare ad ogni transizione di stato un termine che rappresenti distanza, bit d'informazione errati ed ogni altra informazione che interessi (ad esempio la lunghezza dell'evento errore). Per il codice con quattro stati, al primo passo si ha il vettore

$$x_0(D,I) = \begin{bmatrix} 0 \\ D^2 I \\ 0 \\ 0 \end{bmatrix} \tag{5.8}$$

dove il termine $D^2 I$ corrisponde alla transizione dallo stato 00 a 10, con distanza 2 e con un bit d'informazione errato (se non si fosse interessati al numero di bit errati basterebbe porre $I = 1$). Un generico passo nel traliccio è descritto dalla matrice

$$A(D,I) = \begin{bmatrix} 0 & 0 & D^2 & 0 \\ 0 & 0 & I & 0 \\ 0 & D & 0 & D \\ 0 & DI & 0 & DI \end{bmatrix} \tag{5.9}$$

dove le colonne corrispondono allo stato iniziale e le righe a quello finale.

Il prodotto $A(D,I)x_0(D,I)$ dà un vettore che enumera tutti i percorsi esistenti nei primi due passi del traliccio a partire dallo stato nullo e verso tutti gli stati. Si ottiene

$$A(D,I)x_0(D,I) = \begin{bmatrix} 0 \\ 0 \\ D^3 I \\ D^3 I^2 \end{bmatrix} \tag{5.10}$$

dove i termini nulli indicano che non vi sono percorsi. Ripetendo n volte la moltiplicazione per $A(D,I)$ si ottiene l'enumerazione di tutti i percorsi

di lunghezza $n+1$. Un generico monomio nel polinomio primo elemento del vettore corrisponde ad un evento errore di lunghezza $n+1$; gli esponenti di D e I indicano distanza e bit errati dell'evento errore. Ad esempio è facile verificare che il primo elemento di $A^4(D,I)x_0(D,I)$ è $D^6I^2 + D^7I^3$, che corrisponde ai due eventi errore di lunghezza 5 facilmente individuabili nel traliccio di fig. 5.5. Volendo considerare tutti gli eventi errore, di qualsiasi lunghezza, si sommerà per tutti gli n:

$$T(D,I) = \sum_{n=0}^{\infty} A^n(D,I)x_0(D,I) \tag{5.11}$$

È inteso che del risultato si prenda solo il primo polinomio, corrispondente allo stato finale nullo[18]. Osservando che anche per le matrici vale, sempre che la somma converga, la relazione $\sum_{n=0}^{\infty} A^n = (\mathbf{1}-A)^{-1}$, dove $\mathbf{1}$ è la matrice identità, si ottiene

$$T(D,I) = (\mathbf{1} - A(D,I))^{-1}x_0(D,I) \tag{5.12}$$

Con l'aiuto di un programma di analisi simbolica si può poi sviluppare in serie di potenze, ottenendo $T(D,I) = \sum_d \sum_i n(d,i)D^dI^i$ e quindi l'enumerazione degli eventi errore. Se invece i bit errati non dovessero interessare basterebbe porre $I=1$ e considerare $T(D) = \sum_d a(d)D^d$. Infine si possono utilizzare la (5.4) e la (5.6) o (5.7), per maggiorare $P(E)$ o $P_b(E)$ rispettivamente.

Se si accetta un'ulteriore approssimazione si può usare direttamente in modo numerico la funzione di trasferimento. Per esempio si può sfruttare la maggiorazione (es. 5.20)[19]

$$Q(\sqrt{x+y}) \leq Q(\sqrt{x})\exp(-y/2) \tag{5.13}$$

[18] poiché non esistono eventi errore di lunghezza minore di K si potrebbe sommare a partire dall'indice $K-1$

[19] un risultato più semplice, ma meno accurato, si ottiene dalla maggiorazione, già vista nel Cap. 2, $Q(x) \leq \frac{1}{2}\exp(-x^2/2)$, valida per $x \geq 0$

valida per $x, y \geq 0$, per ottenere facilmente ponendo $d = d_f + d'$

$$\begin{aligned} P(E) &\leq \sum_d a(d) Q\Big(\sqrt{\frac{2E_b}{N_0} dR}\Big) \leq \\ &\leq Q\Big(\sqrt{\frac{2E_b}{N_0} d_f R}\Big) \exp\Big(\frac{E_b}{N_0} d_f R\Big)\, T(D)|_{D=\exp(-RE_b/N_0)} \end{aligned} \tag{5.14}$$

Si osservi che non occorre valutare simbolicamente $T(D) = (\mathbf{1} - A(D))^{-1} x_0(D)$ e poi sostituire $D = \exp(-RE_b/N_0)$; si può subito sostituire il valore di D e procedere numericamente.

Un po' più elaborato è il calcolo di $P_b(E)$. Occorre osservare che

$$\sum_d \sum_i i\, n(d,i) D^d = \frac{\partial T(D,I)}{\partial I}|_{I=1} \tag{5.15}$$

e quindi

$$\begin{aligned} P_b(E) &\leq \frac{1}{b} \sum_d \sum_i i\, n(d,i) Q\Big(\sqrt{\frac{2E_b}{N_0} dR}\Big) \leq \\ &\leq Q\Big(\sqrt{\frac{2E_b}{N_0} d_f R}\Big) \exp\Big(\frac{E_b}{N_0} d_f R\Big) \frac{\partial T(D,I)}{\partial I}|_{D=\exp(-RE_b/N_0);\ I=1} \end{aligned} \tag{5.16}$$

Per la derivata dell'inversa di una matrice $B(x)$ vale la relazione, forse non notissima ma analoga alla derivata dell'inverso di una funzione scalare,

$$\frac{dB^{-1}(x)}{dx} = -B^{-1}(x) \frac{dB(x)}{dx} B^{-1}(x) \tag{5.17}$$

e quindi si ottiene

$$\begin{aligned} \frac{\partial T}{\partial I} &= \frac{\partial (\mathbf{1} - A)^{-1}}{\partial I} x_0 + (\mathbf{1} - A)^{-1} \frac{\partial x_0}{\partial I} = \\ &= (\mathbf{1} - A)^{-1} \frac{\partial A}{\partial I} (\mathbf{1} - A)^{-1} x_0 + (\mathbf{1} - A)^{-1} \frac{\partial x_0}{\partial I} \end{aligned} \tag{5.18}$$

Le derivate di A e x_0 rispetto ad I non offrono difficoltà: basta moltiplicare ogni termine per l'esponente di I. Poi si sostituisce $I = 1$, e si procede

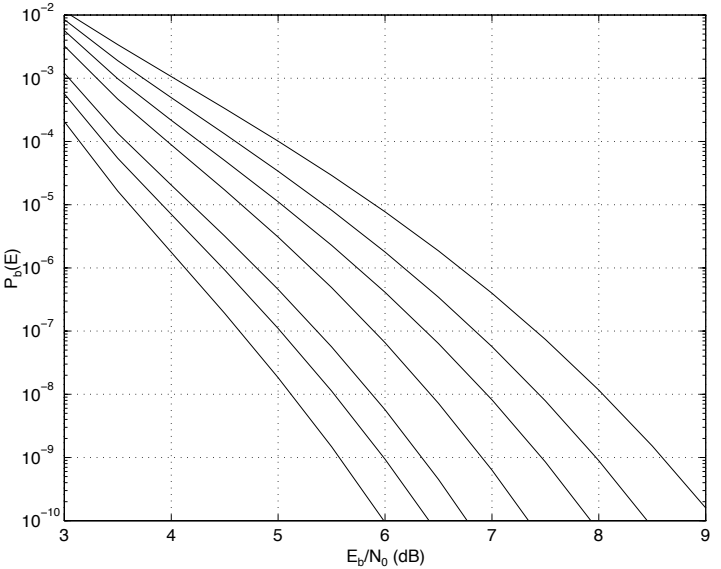

Fig. 5.7 - Maggiorazione della probabilità d'errore dei codici convoluzionali con *rate* 1/2 da 4 a 256 stati (da destra a sinistra)

numericamente. Nel prodotto delle tre matrici e del vettore x_0 conviene partire dal fondo, in modo da ottenere sempre un vettore[20]. L'unica parte costosa è quindi valutare numericamente la matrice inversa $(\mathbf{1} - A)^{-1}$, come nel calcolo di $P(E)$.

Le fig. 5.7 e 5.8 presentano i risultati così ottenuti per codici con *rate* $R = 1/2$ e $2/3$. Per basso rapporto segnale-rumore lo *union bound* è molto largo, e quindi inutilizzabile, e la serie può addirittura divergere. Inutile dire che in tal caso si ottengono numeri privi di senso[21].

Probabilità d'errore così elevate sono, per la gran parte dei sistemi pratici, inaccettabili e quindi di nessun interesse se non in situazioni particolari. Il caso più tipico è quello della codifica *concatenata*, in cui un codice *interno* opera a rapporto segnale-rumore molto basso e quindi produce una probabilità

[20] il risultato è lo stesso se si parte moltiplicando le matrici, ma il calcolo è più pesante

[21] come se si tentasse di utilizzare per $x > 1$ la ben nota somma della serie $\sum_{i=0}^{\infty} x^i = \frac{1}{1-x}$

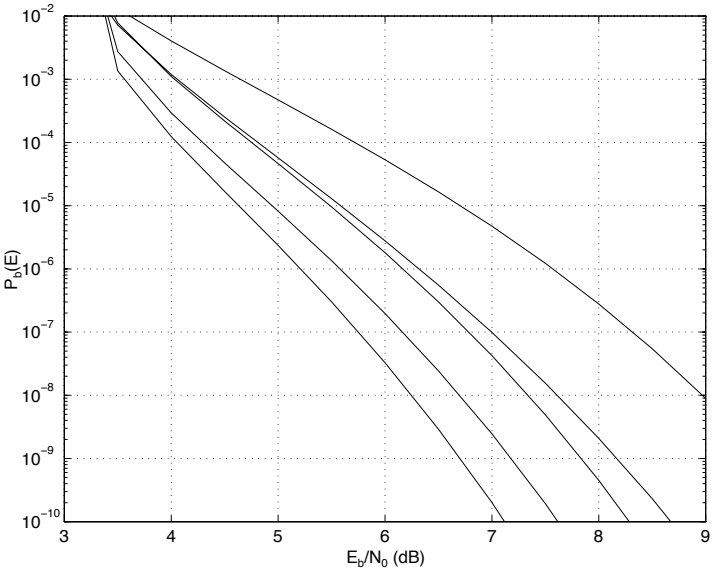

Fig. 5.8 - Maggiorazione della probabilità d'errore dei codici convoluzionali con *rate* 2/3 da 4 a 64 stati (da destra a sinistra)

d'errore molto elevata; un codice *esterno* riduce poi la probabilità d'errore ai valori desiderati. Solitamente il codice esterno è un codice a blocco correttore di molti errori. Poiché l'andamento della probabilità d'errore all'uscita del decodificatore esterno è una funzione rapidamente variabile della probabilità d'errore in ingresso, occorre valutare quest'ultima con precisione, e ciò è possibile solo mediante simulazione. Fortunatamente gli errori si presentano con frequenza elevata, e quindi non sono richiesti (solitamente) tempi di calcolo eccessivi.

Per il codice a 64 stati con *rate* $R = 1/2$ e ad $E_b/N_0 = 2.85$ dB, la fig. 5.9 mostra le frequenze degli eventi errore (raggruppati per lunghezza) ottenute sia mediante simulazione sia con lo *union bound*. Quest'ultimo è effettivamente un po' pessimista (è comunque interessante vedere la somiglianza tra i due istogrammi). A questo valore di E_b/N_0 si ha $P(E) = 2 \cdot 10^{-4}$, e quindi in media ha inizio un evento errore ogni 5000 passi nel traliccio; la simulazione

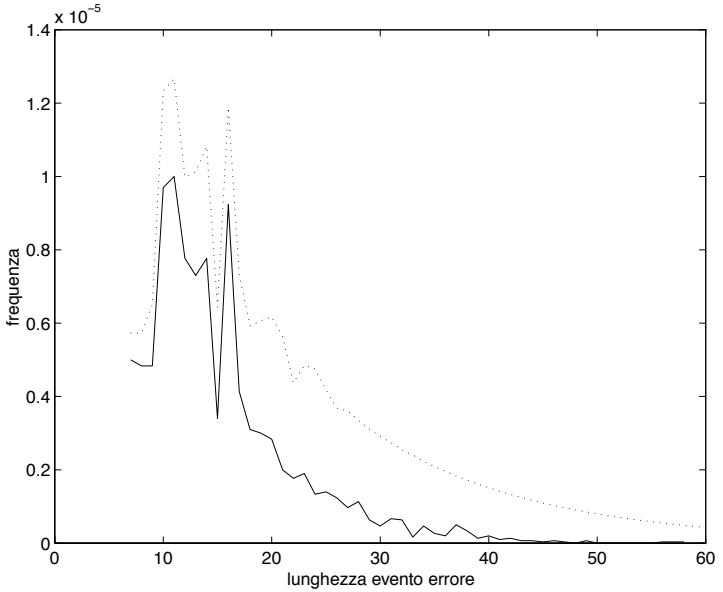

Fig. 5.9 - Frequenze degli eventi errore (raggruppati per lunghezza) nel caso di codice convoluzionale a 64 stati con *rate* 1/2 a $E_b/N_0 = 2.85$ dB (curva continua), confrontate con lo *union bound* (curva punteggiata)

è quindi ancora possibile[22]; diventa via via più pesante all'aumentare del rapporto segnale-rumore perché gli errori sono sempre più rari.

5.8 Codifica concatenata

I codici convoluzionali forniscono prestazioni particolarmente interessanti per probabilità d'errore intorno a $10^{-4} \div 10^{-7}$, peraltro adeguate per molte applicazioni. L'andamento di $P(E)$ o $P_b(E)$ in funzione di E_b/N_0 non è però particolarmente ripido, e diventa ben presto quasi parallelo a quello della trasmissione binaria antipodale. Quando si vogliono ottenere probabilità d'errore molto basse i codici convoluzionali perdono almeno in parte la loro attrattiva. I più potenti codici a blocco, al contrario, dopo aver pagato

[22] in realtà volendo mettere in evidenza gli eventi errore fino a lunghezze intorno a 40-50, poco probabili, occorre un bel po' di tempo di simulazione

la considerevole (e quasi inevitabile) perdita di un paio di dB dovuta alla decodifica *hard* recuperano il terreno perduto grazie alla maggior distanza di Hamming.

Uno schema di codifica concatenato consiste nella cascata di due codici. Quello *interno*, quasi sempre convoluzionale, opera con un rapporto segnale-rumore molto basso, e produce all'uscita del decodificatore una probabilità d'errore relativamente elevata. Al canale binario così ottenuto si applica poi un codice *esterno*, quasi sempre a blocco, che riduce la probabilità d'errore al livello desiderato.

Molto spesso il codice esterno è un *Reed-Solomon*. Si tratta di una classe di codici *non binari* particolarmente efficienti. Se $N = 2^m - 1$ è la lunghezza del blocco, l'alfabeto utilizzato dal codice ha 2^m elementi[23]. Gli N simboli c_i che costituiscono una parola di codice possono corrispondere, ad esempio, a 2^m segnali ortogonali[24]. Più spesso il simbolo viene rappresentato mediante un *byte* di m bit, e ogni parola di codice equivale a N byte oppure a $N \cdot m$ bit. Perché un codice Reed-Solomon possa correggere t errori occorre che abbia $N - K = 2t$ simboli di parità. È inteso che il codice corregge t *byte*, indipendentemente dal numero di bit errati in un byte. Ad esempio un codice di lunghezza $N = 255$ e con $K = 223$ simboli d'informazione ha byte di 8 bit e corregge fino a 16 byte errati (pari a un massimo di 128 bit, nel caso più favorevole).

Quando il decodificatore del codice interno convoluzionale produce un evento errore molti bit consecutivi possono risultare errati, soprattutto a basso rapporto segnale-rumore (si riveda la fig. 5.9). Si noti il fatto favorevole che più bit errati possono cadere nello stesso byte, e quindi contare come un solo errore per il decodificatore del codice esterno. Tuttavia un evento errore molto lungo fa sbagliare certamente molti byte e ciò può portare troppo vicino al potere correttore del codice. Al limite, se ogni evento errore producesse più di t byte errati il codice esterno sarebbe del tutto inefficace. La soluzione comunemente adottata consiste nel disperdere, più o meno casualmente, i byte errati in modo che vadano a cadere in parole di codice diverse. Ciò richiede di riordinare il flusso di byte tra i due codificatori mediante una permutazione

[23] è possibile portare la lunghezza del blocco a $N = 2^m$ o $2^m + 1$, se occorre, ma non oltre; naturalmente è sempre possibile *accorciare* il codice

[24] utilizzando segnali ortogonali ogni simbolo può essere sbagliato, con pari probabilità, in $2^m - 1$ modi e il codice Reed-Solomon si trova completamente a suo agio perché non ha alcuna preferenza per errori particolari; è un vero peccato che i segnali ortogonali siano quasi da bandire per problemi di banda

detta *interleaving* e di compiere l'operazione opposta di *deinterleaving* tra i due decodificatori. In questo modo gli eventi errore prodotti dal decodificatore convoluzionale, dopo essere stati frammentati in byte, vengono dispersi su parole di codice diverse.

Si consideri, ad esempio, lo schema concatenato dove il codice interno è l'ottimo convoluzionale a 64 stati con *rate* $R = 1/2$ della tab. 2 ed il codice esterno è il Reed-Solomon (255,239) correttore di $t = 8$ errori. Si può mostrare che per ottenere $P_b(E) = 10^{-10}$ basta $E_b/N_0 = 3.1$ dB se si utilizza *interleaving* infinito; in pratica è sufficiente disperdere gli eventi errore su quattro o cinque parole di codice. Senza *interleaving*, invece, le prestazioni sono molto scadenti: allo stesso rapporto segnale-rumore si ha $P_b \approx 10^{-4}$. Si noti che le operazioni di *interleaving* e *deinterleaving* introducono un ritardo pari a diverse volte la durata di una parola del codice Reed-Solomon.

Le prestazioni migliorano di circa 0.3 dB se si utilizza il codice Reed-Solomon (255,223) correttore di 16 errori. Oltre tale ridondanza si ha invece un peggioramento: il maggior potere correttore del codice non compensa più la riduzione dell'energia per bit di canale, data da $E_s = R\frac{K}{N}E_b$.

5.9 Modulazione e codifica integrate

Prima di introdurre i sistemi codificati in cui la distanza geometrica non è proporzionale ad una qualche distanza di Hamming, è opportuna una breve digressione teorica sulla trasmissione di informazione con bassa probabilità d'errore, ad un ritmo corrispondente a un numero prefissato di bit per dimensione.

Se tale rapporto è 1/2 la modulazione binaria antipodale (o 4PSK) insieme ad un codice a blocco o convoluzionale con *rate* 1/2 è la risposta adeguata. Resta solo da scegliere il codice, e in particolare la sua complessità.

Volendo trasmettere un bit per dimensione la modulazione binaria non lascia ridondanza per un codice, e quindi non è la soluzione. Certamente si può non codificare per nulla, ma è ben noto che si resta lontanissimi dalle prestazioni limite previste dalla capacità di canale.

Piuttosto si potrebbe pensare ad una modulazione d'ampiezza a 4 livelli, combinata con un qualche codice con *rate* 1/2. Per ogni bit d'informazione ne escono due dal codificatore, e questi vengono trasmessi con un simbolo a quattro livelli; si trasmette quindi un bit d'informazione per simbolo, cioè un bit per dimensione. La scelta del codice non è immediata per le costellazioni

a 4 livelli, ma questo è un problema che si affronterà tra poco e che per ora si può ignorare.

Si noti però che esistono altre soluzioni. Ad esempio una modulazione a 8 livelli combinata con un codice con *rate* 1/3. Oppure, in banda passante, un codice con *rate* 2/3 e modulazione 8PSK: entrano nel codificatore due bit d'informazione e ne escono tre, che individuano il simbolo trasmesso; ogni due bit d'informazione si trasmette un simbolo bidimensionale, quindi un bit per dimensione. Le soluzioni sono dunque moltissime, e bisogna trovare qualche criterio per orientarsi nella scelta.

Si noterà anche che in tutti questi casi a parità di banda si trasmette l'informazione allo stesso ritmo del binario antipodale. L'affermazione "i codici espandono la banda" è quindi vera solo nel senso che *a parità di costellazione* l'introduzione di un codice riduce il ritmo di trasmissione, ma è tuttavia poco significativa perché nulla vieta di cambiare la costellazione.

Allo stesso modo è facile immaginare esempi in cui si trasmette un diverso numero, anche molto maggiore di 1, di bit per dimensione.

La notissima formula per la capacità di canale $C = B \log_2(1 + P/N_0 B)$, e l'equivalente per E_b/N_0 in funzione dell'efficienza spettrale, valgono quando non si pongano vincoli sull'ampiezza delle singole componenti del vettore trasmesso. Se invece si impone l'uso di una particolare costellazione le singole componenti sono quantizzate, e ciò riduce la capacità, a valori che è comunque possibile calcolare (si veda il Cap. 4).

Anziché la capacità, spesso si calcola il *cutoff rate* R_0, che alcuni ritengono vicino al limite *pratico* raggiungibile dai sistemi codificati. In una modulazione multilivello in teoria si potrebbe sfruttare il fatto che i livelli estremi sono lievemente favoriti rispetto a quelli interni, e dovrebbero quindi essere utilizzati con frequenza un po' maggiore. Ciò complica il codice, e dà un vantaggio molto modesto. Questa possibilità viene quindi ignorata. La fig. 5.10, già presentata nel Cap. 4, mostra i valori di R_0 in funzione di E_b/N_0 per le costellazioni monodimensionali a M livelli equispaziati ed equiprobabili.

Come si vede non c'è motivo per usare una costellazione multilivello se si vuol trasmettere 1/2 bit o 3/4 di bit per dimensione[25]. Ma se ci si avvicina a un bit per dimensione, per non parlare del caso non codificato, la segnalazione binaria diventa inefficiente. Occorre passare a 4 livelli, ma non vi è necessità di andare oltre. Quattro livelli, insieme ad un opportuno codice,

[25] ed anzi le costellazioni multilivello sono lievemente penalizzate dalla scelta di usare i livelli con pari probabilità

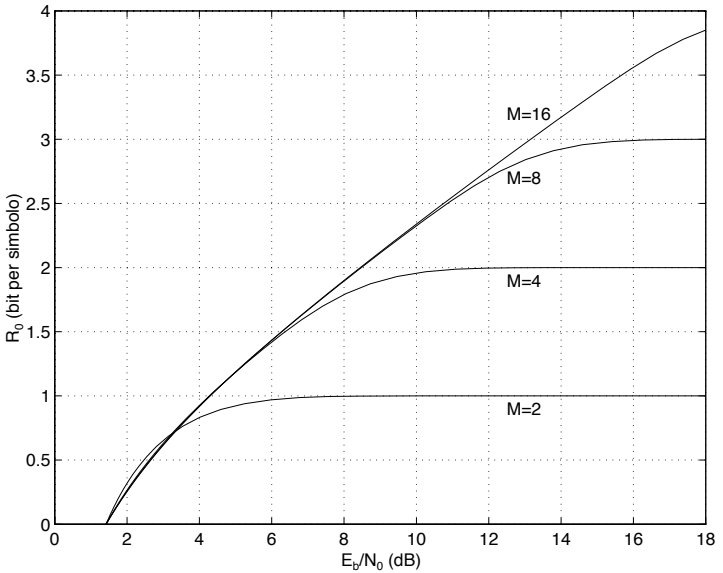

Fig. 5.10 - R_0 per costellazioni monodimensionali ad M livelli equispaziati ed equiprobabili

sono raccomandabili fino a 1.5 bit per dimensione, o poco più. Oltre occorrono 8 livelli, e così via.

La fig. 5.11 riproduce i risultati del Cap. 4 per costellazioni bidimensionali. Praticamente in tutti i casi è sufficiente usare una costellazione con un numero di punti doppio di quello che sarebbe richiesto nella trasmissione non codificata (la costellazione 2PSK è utile solo per frazioni di bit per simbolo). Sia nel caso monodimensionale sia nel bidimensionale l'andamento della capacità C è simile a quello di R_0, e porta a conclusioni analoghe.

Avendo un'indicazione sulle costellazioni da usare e sul *rate* del codice non resta che trovare i codici, e valutarne le prestazioni.

5.10 "Set partitioning"

Si consideri la semplice costellazione 4PAM in fig. 5.12, con numerazione binaria naturale dei quattro punti. Detta 2 la distanza tra punti adiacenti, è immediato verificare che una differenza nel secondo bit garantisce $d^2 \geq 4$; se

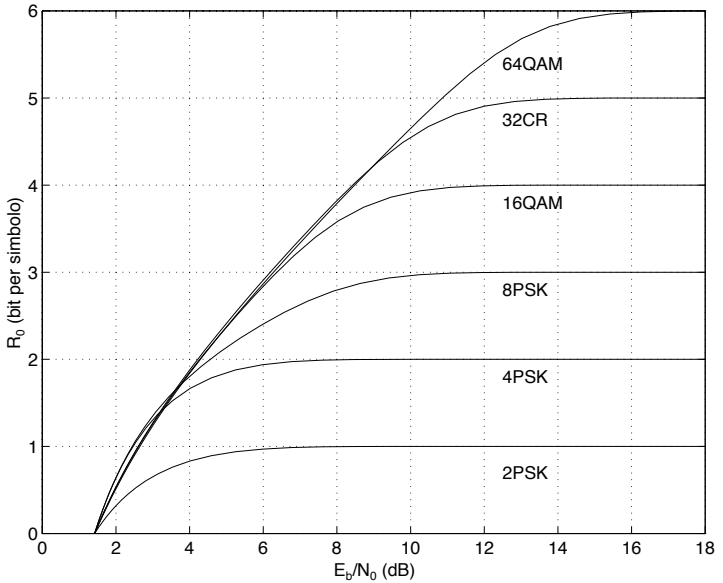

Fig. 5.11 - R_0 per costellazioni bidimensionali (punti equiprobabili)

Fig. 5.12 - Costellazione monodimensionale a 4 livelli con *mapping* naturale

il secondo bit coincide, una differenza nel primo dà $d^2 = 16$.

Analogamente per la costellazione 8PAM in fig. 5.13 una differenza nel terzo bit garantisce $d^2 \geq 4$; nel caso il terzo bit coincida, una differenza nel secondo dà $d^2 \geq 16$; se infine coincidono sia il secondo sia il terzo bit, il primo fornisce una distanza $d^2 = 64$.

La costellazione 8PSK di fig. 5.14 con *mapping* naturale ha proprietà analoghe: detta 1 l'energia per simbolo, una differenza nel terzo bit garantisce $d^2 \geq 0.586$; se il terzo bit coincide, una differenza nel secondo dà $d^2 = 2$; se coincidono sia il secondo sia il terzo bit, il primo fornisce una distanza $d^2 = 4$. Analoghe considerazioni valgono per la costellazione 16PSK, peraltro raramente utilizzata.

Per le costellazioni QAM si consideri il *mapping* di fig. 5.15. Normaliz-

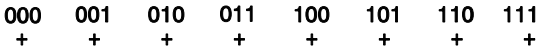

Fig. 5.13 - Costellazione monodimensionale a 8 livelli con *mapping* naturale

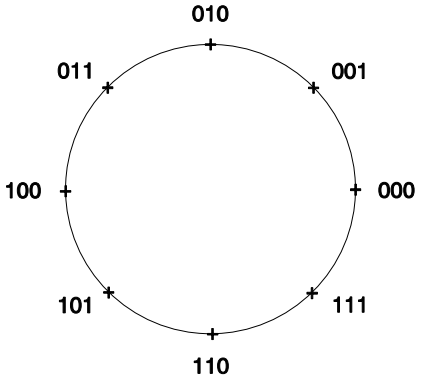

Fig. 5.14 - Costellazione 8PSK con *mapping* naturale

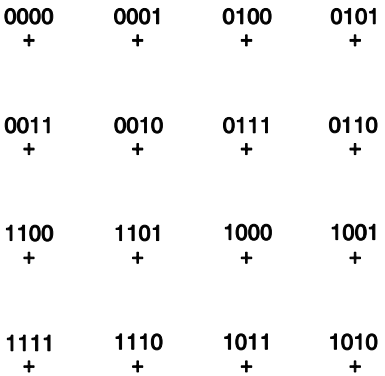

Fig. 5.15 - Costellazione 16QAM con *set partitioning*

zando a 2 la distanza tra punti adiacenti, il quarto bit differente garantisce $d^2 \geq 4$; nel caso coincida, una differenza nel terzo bit garantisce $d^2 \geq 8$; se coincidono terzo e quarto bit, il secondo bit produce una distanza $d^2 \geq 16$; e se ancora occorresse (ma non avviene, di norma) il primo bit potrebbe garantire $d^2 \geq 32$.

In tutti questi casi il valore dell'ultimo bit suddivide l'insieme di punti della costellazione in due sottoinsiemi (*subset*) tali che la distanza tra un elemento qualsiasi del primo ed uno del secondo è $d^2 \geq d_1^2$. La distanza tra punti di uno stesso *subset* è invece $d^2 \geq d_2^2$, con $d_2^2 > d_1^2$. Entrambi i *subset* possono essere a loro volta partizionati secondo il valore del penultimo bit. Si ottengono quattro *subset*, individuati dagli ultimi due bit. La distanza tra punti di *subset* che differiscono nell'ultimo bit è $d^2 \geq d_1^2$; tra *subset* con l'ultimo bit uguale ma che differiscono nel penultimo è $d^2 \geq d_2^2$; tra punti dello stesso *subset* $d^2 \geq d_3^2$, con $d_3^2 > d_2^2$.

Se la costellazione contiene almeno otto punti, l'operazione può essere ripetuta ottenendo otto *subset*, e di norma non occorre andare oltre. La distanza tra punti generici ...0 e ...1 è $d^2 \geq d_1^2$; tra $..0x$ e $..1x$ ($x = 0, 1$) è $d^2 \geq d_2^2$; tra $.0xy$ e $.1xy$ è $d^2 \geq d_3^2$; tra punti di uno stesso *subset* $.xyz$ e $.xyz$ è $d^2 \geq d_4^2$ (con $d_4^2 > d_3^2 > d_2^2 > d_1^2$; se i *subset* hanno un solo punto si suppone $d_4^2 = \infty$).

Come si vedrà, queste proprietà consentono di trovare buoni sistemi codificati. Si esamineranno dapprima i sistemi basati su codici convoluzionali, proposti alla fine degli anni '70 e rapidamente affermatisi, poi quelli con codifica a blocco, proposti negli stessi anni ma che non hanno ancora larga diffusione.

5.11 Modulazione codificata a traliccio

Si scelgono i *subset* in modo tale che i punti all'interno di ciascuno siano già sufficientemente distanti da non richiedere protezione da parte di un codice. Punti appartenenti a *subset* diversi possono invece essere troppo vicini. Si seleziona la sequenza dei *subset* trasmessi mediante un codice convoluzionale, in modo da poter accumulare su più simboli una distanza sufficiente anche in questo caso. Questi codici, introdotti da Ungerboeck alla fine degli anni '70, sono spesso indicati con la sigla TCM (*Trellis Coded Modulation*).

Un semplicissimo esempio con modulazione 8PSK, uno dei primi proposti, aiuterà a chiarire. I punti dei quattro *subset* sono $x00$, $x01$, $x10$ e $x11$ ($x = 0$

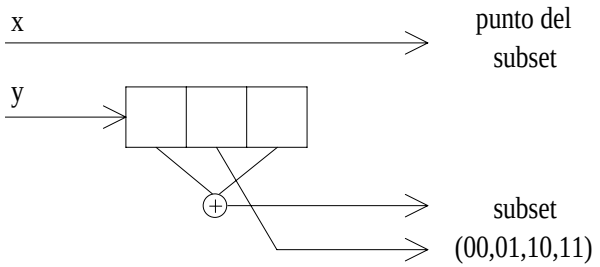

Fig. 5.16 - Codificatore TCM 8PSK a quattro stati

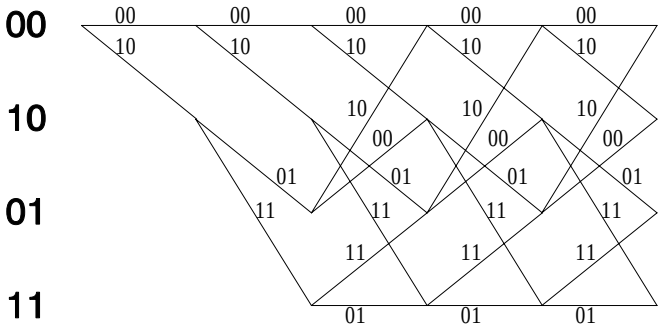

Fig. 5.17 - Traliccio del codificatore TCM 8PSK a quattro stati

o 1). Ogni *subset* contiene due punti, e la distanza al quadrato tra questi è $4E_s$. Il terzo bit diverso garantisce $d^2 \geq 0.586E_s$; se il terzo coincide, una differenza nel secondo dà $d^2 = 2E_s$. La sequenza dei *subset* è selezionata dal codice convoluzionale in fig. 5.16, alimentato dalla sequenza $\{y\}$. Si noti che il codice è diverso dal miglior convoluzionale con *rate* 1/2 a quattro stati della tab. 2, perché non deve rendere massima la distanza di Hamming, ma piuttosto quella geometrica nello spazio dei segnali.

In fig. 5.17 è mostrato il traliccio, dove esistono (anche se non sono indicate esplicitamente) transizioni di stato *parallele*: se ad esempio il codice convoluzionale decreta che la transizione di stato sia da 00 a 00 e quindi che si debba trasmettere il *subset* 00, resta la possibilità di trasmettere la fase 000 oppure 100 a seconda del valore x del bit non codificato. Ogni ramo del traliccio corrisponde quindi ad una coppia di segnali possibili.

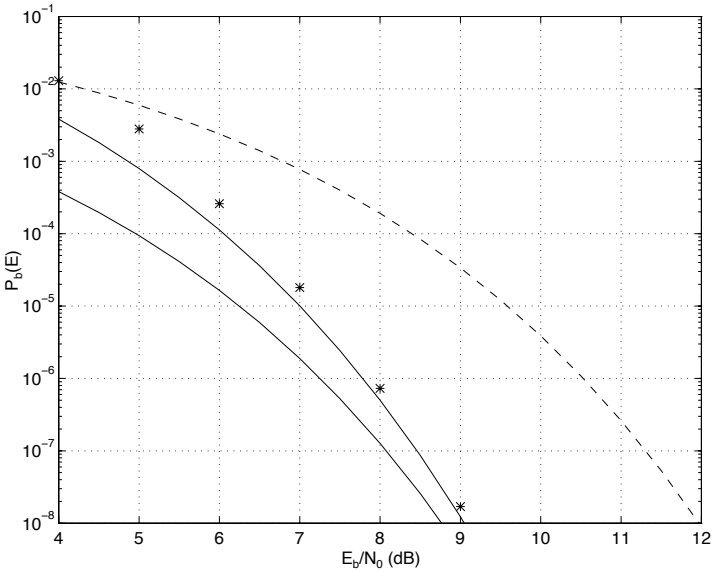

Fig. 5.18 - Probabilità d'errore per il codice TCM 8PSK a quattro stati; simulazione (asterischi) e *union bound* (curve continue: eventi errore a distanza minima, o fino a $d^2 = 4.586E_s$) a confronto con il 4PSK non codificato (tratteggio)

L'ispezione del traliccio, con l'aiuto delle proprietà del *set partitioning* mostra che la minima distanza per transizioni non parallele è $d^2 = 4.586E_s$, e si ha con la sequenza di *subset* 10,01,10. A causa delle transizioni parallele vi sono otto di tali eventi errore, che possono produrre fino a quattro bit d'informazione errati.

La distanza corrispondente alle transizioni parallele, eventi errore che si esauriscono in un passo e producono un solo bit errato, è invece $d^2 = 4E_s$, e domina ad alto rapporto segnale-rumore. Sostituendo $E_s = 2E_b$ si vede che il codice ha un guadagno asintotico di 3 dB rispetto al 4PSK non codificato, *a parità di ritmo di trasmissione e banda.*

Le prestazioni del codice sono mostrate in fig. 5.18: i punti indicati con un asterisco sono ottenuti mediante simulazione[26]; sono mostrate per confronto le

[26] per $P_b(E)$ molto bassa occorrono tecniche particolari, cosiddette di *importance sampling*

prestazioni del 4PSK non codificato, e quelle valutate mediante lo *union bound* tenendo conto dell'unico evento errore a distanza minima, oppure anche degli otto con distanza $d^2 = 4.586E_s$. Questi ultimi danno un contributo non del tutto trascurabile perché sono più numerosi, hanno distanza poco superiore alla minima e producono un maggior numero di bit errati.

Con lo stesso codificatore di fig. 5.16 il lettore è invitato ad esaminare il caso di costellazione monodimensionale a 8 livelli, che trasmette due bit per dimensione. I quattro *subset* hanno due punti ciascuno, e quindi vi è un bit non codificato e vi sono coppie di transizioni parallele. Con semplici calcoli si trova un guadagno asintotico, rispetto ai 4 livelli non codificati, di 3.31 dB (es. 5.23). A conti fatti si osserverà anche che questa volta la distanza minima non viene dalle transizioni parallele, ma dal codice convoluzionale. Si possono quindi migliorare le prestazioni aumentando la complessità, cioè il numero degli stati del codice. Infatti si potrebbe mostrare che con 4 *subset* e 8 stati si può arrivare a 3.77 dB, con 16 stati a 4.18 dB, e così via fino a quasi 6 dB con 256 stati. Non si dimentichi però che sono guadagni *asintotici*, di norma raggiunti abbastanza rapidamente nei codici semplici e sempre più lentamente nei casi più complessi.

Tornando al primo esempio, con costellazione 8PSK, risulta invece limitante la distanza delle transizioni parallele. Quindi almeno asintoticamente non serve a nulla migliorare il codice. Occorre piuttosto *eliminare le transizioni parallele*, cioè partizionare in 8 *subset* di un solo punto ciascuno. Non vi sono quindi bit non codificati (o *liberi*, come talvolta si dice) e il codice convoluzionale ha *rate* effettivo 2/3. Con 8 stati si riesce a ottenere un guadagno asintotico di 3.60 dB, e con 256 stati di 5.75 dB.

Per le costellazioni QAM valgono considerazioni analoghe. I codici più semplici, a 4 stati, richiedono 4 *subset* e quindi codificatori con *rate* 1/2, e danno guadagni asintotici intorno a 3 dB. A partire da 8 stati occorre partizionare in 8 *subset*. Se quindi ad esempio si vogliono trasmettere 5 bit per simbolo (2.5 bit per dimensione) si usa una costellazione di $2 \cdot 2^5 = 64$ punti, partizionata in 8 *subset* di 8 punti ciascuno; ci sono quindi 8 transizioni parallele. Due bit d'informazione entrano nel codificatore convoluzionale con *rate* $R = 2/3$. Escono 3 bit che selezionano il *subset*. I 3 bit non codificati selezionano il punto all'interno del *subset*.

Si noti quanto poco cambierebbe se si volessero trasmettere 7 bit per simbolo (3.5 bit per dimensione). Si partizionerebbe un 256QAM in 8 *subset* di 32 punti, e si avrebbero 32 transizioni parallele. I 3 bit uscenti dal codificatore

sceglierebbero, come prima, il *subset* e i 5 non codificati il punto nel *subset*. In un certo senso si può dire che per le alte capacità il codice è *universale*, cioè praticamente indipendente dalla dimensione della costellazione.

Anche per le costellazioni QAM i guadagni asintotici con 256 stati risultano intorno ai 6 dB. In nessun caso si è trovato utile partizionare in 16 *subset*.

Il ricevitore può essere realizzato come per i codici convoluzionali, con l'unica novità delle transizioni parallele. Risulta conveniente determinare come primo passo per ogni gruppo di transizioni parallele il punto del corrispondente *subset* che risulta più vicino al vettore ricevuto. Dopo aver così sfoltito il grafo, lasciando un solo candidato per ciascuna transizione di stato, si applica l'usuale algoritmo di Viterbi.

Per la valutazione delle prestazioni valgono considerazioni analoghe ai codici convoluzionali, con l'unica avvertenza di contare con opportuna molteplicità nello *union bound* le transizioni parallele. La distanza minima può dipendere dalla sequenza trasmessa e può talvolta risultare maggiore di quella garantita dal *set partitioning* (es. 5.28); tuttavia quasi sempre si trovano sequenze per cui vale l'uguaglianza, e quindi si deve fare conto solo sulla distanza minima garantita. Analogamente il numero di eventi errore a distanza minima può dipendere dalla sequenza trasmessa. Con qualche cautela, il metodo della funzione di trasferimento può essere utile anche per i codici TCM.

Infine per la ricerca di buoni codici Ungerboeck ha proposto alcune regole empiriche che si sono rivelate molto valide. Infatti con l'indagine esaustiva si trova poco di meglio. Per le tabelle dei migliori codici si rimanda ai testi specializzati. Occorre però osservare che la notazione di Ungerboeck è diversa da quella qui utilizzata. A parte la numerazione degli indici, la differenza principale è che i codificatori sono dati in forma *sistematica* (ciò è possibile solo facendo uso di *strutture retroazionate*). Due esempi sono presentati negli es. 5.26-5.27.

5.12 Codici TCM multidimensionali

La fig. 5.11 mostra che le costellazioni 16QAM, 32CR e 64QAM potrebbero essere utilizzate, senza penalità significativa, per trasmettere rispettivamente 3.5, 4.5 e 5.5 bit per simbolo anziché 3, 4 e 5 come avviene per i codici TCM descritti nella sezione precedente. Analoghe conclusioni si trarrebbero dai grafici della capacità di canale.

Per ottenere un numero frazionario di bit per simbolo si possono raggrup-

pare due simboli bidimensionali in un unico simbolo a quattro dimensioni che trasmette, nei tre casi presi ad esempio, rispettivamente 7, 9 e 11 bit d'informazione.

Per fare un esempio specifico si consideri la costellazione 64QAM. Se si usa un codice con *rate* $R = 1/2$, degli 11 bit d'informazione trasmessi ogni due tempi di simbolo uno va all'ingresso del codice convoluzionale, e altri 10 non vengono codificati; dei 12 bit che definiscono il punto della coppia di simboli 64QAM, due sono quindi codificati e 10 liberi. Analogamente con un codice con *rate* $R = 2/3$ due bit entrano nel codificatore, e ne producono tre in uscita, e nove sono liberi.

Naturalmente occorre aver opportunamente partizionato rispettivamente in quattro oppure otto *subset* i 2^{12} punti della costellazione in quattro dimensioni, con criteri analoghi a quelli discussi per le costellazioni bidimensionali. Si noti il gran numero di transizioni parallele, pari a 1024 o 512 rispettivamente.

Un codice TCM multidimensionale può essere utilizzato anche in altro modo. Si supponga di voler usare coppie di simboli QAM per trasmettere un numero intero di bit d'informazione per simbolo, ad esempio 5, cioè 10 bit d'informazione ogni due simboli QAM. Utilizzando codificatori convoluzionali con *rate* $R = 1/2$ o $2/3$, la differenza tra bit codificati e bit d'informazione è pari ad uno, e basta un insieme di 2^{11} punti in quattro dimensioni. Senza voler entrare in dettagli, si possono selezionare fra le 2^{12} combinazioni corrispondenti ad una coppia di costellazioni 64QAM le 2^{11} più convenienti, ad esempio dal punto di vista dell'energia media o dell'energia di picco, ed utilizzare solo queste.

Quanto alle prestazioni, basti dire che solitamente i codici TCM multidimensionali offrono guadagni asintotici maggiori di quelli bidimensionali, ma con un numero di concorrenti a distanza minima più elevato. Il guadagno asintotico viene quindi ottenuto a livelli più bassi di probabilità d'errore.

Infine si può organizzare qualche bel giochino anche con i bit liberi, introducendo un po' di ridondanza nella costellazione. La tecnica, detta *trellis shaping*, porta a guadagnare circa 1 dB in termini di energia *media*.

5.13 Modulazione codificata a blocchi

Si descriverà solo la classe più semplice di codici, che già offre molte soluzioni.

Con il *set partitioning* si ottengono *subset* tali che la distanza garantita tra ...0 e ...1 è almeno d_1^2; tra ..0x e ..1x almeno d_2^2; tra .0xy e .1xy almeno

d_3^2; infine tra punti di uno stesso *subset* $.xyz$ e $.xyz$ la distanza è almeno d_4^2 (con $d_4^2 > d_3^2 > d_2^2 > d_1^2$). Si supponga di codificare l'ultimo, il penultimo e il terzultimo bit con tre codici a blocco *diversi*, con uguale lunghezza N del blocco e con K_1, K_2 e K_3 bit d'informazione e distanza di Hamming d_1^H, d_2^H e d_3^H. Quindi si trasmettono $K = K_1 + K_2 + K_3$ bit con N simboli.

Se due generici vettori concorrenti contengono parole diverse del primo codice, cioè quello relativo all'ultimo bit, la loro distanza geometrica deriva da almeno d_1^H simboli diversi, ciascuno dei quali garantisce distanza non inferiore a d_1^2. Se sono uguali le parole del primo codice, ma diverse quelle del secondo (relativo al penultimo bit), quest'ultimo garantisce almeno d_2^H simboli diversi, ciascuno con distanza almeno d_2^2. Analogamente se sono uguali le parole dei primi due codici, ma diverse quelle del terzo, vi sono almeno d_3^H simboli diversi con distanza almeno d_3^2. Se infine le parole dei tre codici coincidono ci deve essere almeno un bit non codificato diverso, e quindi almeno un simbolo con distanza non minore di d_4^2. In definitiva la distanza minima dell'insieme di 2^K segnali è

$$d^2 = \min(d_1^H d_1^2, d_2^H d_2^2, d_3^H d_3^2, d_4^2) \tag{5.19}$$

Noti i valori $d_1^2, \ldots, d_4^2$, che dipendono dalla costellazione, è sufficiente scegliere tra i codici a blocco quelli con distanze di Hamming appropriate.

Ad esempio per una costellazione 8PSK (con $E_s = 1$) si ha $d_1^2 = 0.586$, $d_2^2 = 2$, $d_3^2 = 4$ e $d_4^2 = \infty$ (i *subset* contengono un solo punto). Ponendo $N = 8$ si può scegliere come primo codice quello *a ripetizione*, che contiene solo le due parole $0 \ldots 0$ e $1 \ldots 1$, con $K_1 = 1$ e distanza di Hamming $d_1^H = 8$; come secondo il codice *a parità singola*, con $K_2 = 7$ e $d_2^H = 2$; come terzo il codice *universo* con $K_3 = 8$ e $d_3^H = 1$. Ne risulta $K = 16$, cioè si trasmettono 2 bit per simbolo come in un 4PSK non codificato, e $d^2 = 4$, con un guadagno asintotico di 3 dB[27].

Analogamente si possono trovare molti altri sistemi codificati, anche per costellazioni PAM e QAM, spesso con valori stravaganti del numero di bit per dimensione; alcuni sono proposti tra gli esercizi. Il lettore è anche invitato a rivedere alla luce di quanto detto i sistemi codificati a blocco del Cap. 2.

Per i codici più semplici è facile realizzare il ricevitore ML, come mostrato negli esercizi del Cap. 2. Un ricevitore alternativo, solitamente non lontano dall'ottimo, si ottiene con una struttura *multistadio*. Questa consiste nel

[27] il lettore attento avrà notato che si tratta dello stesso sistema codificato presentato nel Cap. 2

cominciare a decodificare il primo codice ignorando gli altri, supponendo che non vi sia alcun vincolo sui bit corrispondenti. Nel semplice esempio 8PSK si tratta quindi di decidere a massima verosimiglianza tra due ipotesi: l'ultimo bit è 0 o 1, ovvero le fasi sono tutte pari o tutte dispari[28]. Dato il vettore ricevuto si calcola il logaritmo della verosimiglianza (es. 5.38) e si prende una decisione definitiva tra le (due) parole del primo codice. Assumendo poi che tale decisione sia corretta, ad esempio che le fasi siano tutte pari, si prende una decisione a massima verosimiglianza tra le parole del secondo codice. Trattandosi di un codice a parità semplice, basterà prendere decisioni *hard* e verificare il rispetto della parità. Se questa non è soddisfatta si cambia il bit meno affidabile. Prese poi per buone anche queste decisioni, non resta che decodificare il terzo codice. Poiché questo è un codice universo basta prendere decisioni indipendenti simbolo per simbolo.

Da tale descrizione si vede come sia delicato soprattutto il primo passo, perché tutto il resto si basa sulla sua correttezza. È quindi opportuno che il primo codice garantisca una distanza geometrica decisamente maggiore di quella minima. Pare anche che sia opportuno che il terzo codice sia un codice universo; ciò è come dire che in realtà si partiziona in 4 *subset* anziché 8.

Non è ancora del tutto chiaro quali prestazioni possano offrire in pratica i sistemi codificati a blocco. Sembrerebbero interessanti soprattutto i sistemi relativamente semplici, con guadagni asintotici dell'ordine dei 3 dB o poco più.

5.14 Decodifica ML dei codici a blocco

Il calcolo esaustivo delle 2^K correlazioni con tutti i segnali è proponibile solo per codici molto semplici. Un metodo generale, almeno in linea di principio, per decodificare in modo ottimale i codici a blocco è di far corrispondere le parole di codice ai rami di un traliccio, come per i codici convoluzionali.

Per semplicità di esposizione ci si limiterà ai codici ciclici, eventualmente accorciati, il cui codificatore ha una struttura come quella di fig. 5.2, corrispondente al codice di Hamming con $N = 7$ e $K = 4$. Definendo *stato* il contenuto degli $N - K$ registri è facile tracciare il traliccio. Il numero degli stati è $2^{N-K} = 8$ e la fig. 5.19 mostra il diagramma a traliccio. Ovviamente ad ogni transizione di stato corrisponde un solo bit trasmesso. Si noti che il ramo superiore corrisponde alla trasmissione di uno zero per le transizioni uscenti

[28] si è visto invece nel Cap. 2 che il ricevitore ML rimanda tale decisione come ultima; ma ciò è possibile solo perché le ipotesi da considerare sono poche

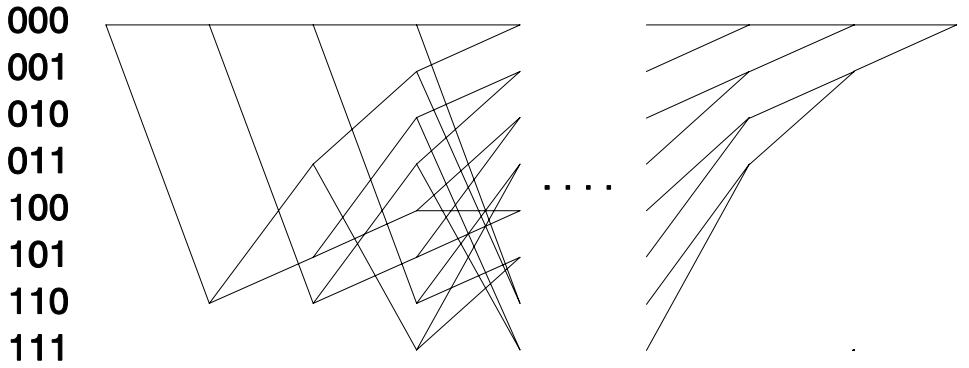

Fig. 5.19 - Traliccio del codice a blocco di fig. 5.2

dagli stati ..0 e di un uno per gli stati ..1. Gli ultimi passi corrispondono all'emissione delle cifre di parità. Con N passi dell'algoritmo di Viterbi si determina il blocco con la massima verosimiglianza. È evidente che ci si deve limitare a codici con un piccolo numero di cifre di parità, perché il numero di stati sia trattabile.

La fig. 5.20 mostra le prestazioni del codice di Hamming con $N = 31$ e $K = 26$ valutate simulando un numero sufficiente di volte il ricevitore ML. Per confronto è mostrato anche lo *union bound* calcolato includendo tutti i concorrenti a distanza di Hamming da 3 fino a 5.

5.15 Considerazioni conclusive

Il panorama dei sistemi di trasmissione codificati è molto vasto, e sempre in evoluzione. In questo capitolo si sono presentati i sistemi di codifica ormai classici (molti codici sono stati standardizzati per le più varie applicazioni).

I codici furono pensati, in origine, per un canale binario per sua natura o comunque reso binario da decisioni *hard*. Tuttavia ci si rese presto conto che l'informazione *soft* è preziosa, se disponibile, e non dovrebbe essere sprecata. Il metodo classico di decodifica dei codici convoluzionali, l'algoritmo di Viterbi, utilizza l'informazione *soft* senza alcuna fatica. Ciò ha portato alla grande diffusione di questa classe di codici.

I metodi tradizionali di decodifica dei codici a blocco sono invece basati sull'algebra dei campi finiti, e sono intrinsecamente *hard*. Qualche tentativo

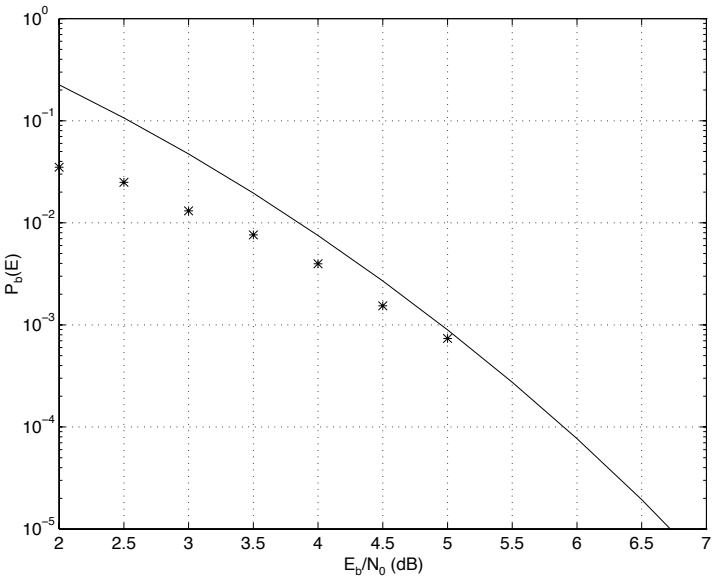

Fig. 5.20 - Probabilità d'errore per il codice di Hamming (31,26) decodificato a massima verosimiglianza (asterischi), e *union bound* (concorrenti fino a distanza 5)

di decodifica *soft* è stato fatto, ma sempre con un incremento non trascurabile della complessità. Una tecnica abbastanza intuitiva consiste nell'individuare, anziché la parola di codice alla minima distanza di Hamming dal blocco ricevuto, un insieme non troppo numeroso di possibili candidati fra i quali poi scegliere valutando le correlazioni, cioè le effettive verosimiglianze (metodo di *Chase*).

Quanto alle prestazioni dei codici binari, in estrema sintesi si può dire che i convoluzionali danno il meglio per probabilità d'errore intorno a $10^{-5} \div 10^{-7}$ e che quelli a blocco sono invece interessanti per valori molto più bassi.

Metodi di codifica concatenata consentono di ottenere prestazioni rispettabili con valori di E_b/N_0 intorno a 2 dB. Si può osservare che se fosse disponibile un metodo *soft* per la decodifica del codice esterno si dovrebbe richiedere al decodificatore interno di produrre come uscita non semplicemente i bit decodificati ma anche una misura della probabilità a posteriori degli stessi bit. Si vorrebbe cioè un decodificatore di Viterbi con uscita *soft*. Ciò è possibile, e si

stanno studiando molti sistemi codificati basati su queste tecniche.

Un'ultima semplice osservazione sui codici binari. Benché non detto esplicitamente, è chiaro che il ricevitore deve individuare in un sistema codificato a blocchi l'inizio di ogni blocco, cioè recuperare il *sincronismo di trama* tra gli N possibili. Nel caso di un codice convoluzionale con *rate* $R = 1/2$ il problema è più semplice, ma sempre presente: si deve individuare il corretto sincronismo in una trama fatta di due bit codificati.

Infine, volendo ottenere efficienze spettrali elevate con segnalazioni multilivello, i codici binari basati sulla ottimizzazione della distanza di Hamming non sono adatti perché non producono la massima distanza geometrica nello spazio dei segnali. I concetti di *set partitioning* portano ad individuare i codici più convenienti.

In questa introduzione ai sistemi codificati non si è fatto cenno a canali particolari, come il radiomobile, in cui l'ampiezza (complessa) del segnale ricevuto fluttua casualmente. Quasi inutile dire che i codici qui presentati non sono necessariamente i più adatti a tali canali.

5.16 Esercizi

5.1 - Si mostri che se $D + 1$ è un fattore del polinomio generatore $g(D)$ di un codice a blocco binario, ogni parola di codice ha un numero pari di uni e quindi la distanza minima del codice è pari. *Suggerimento*: basta mostrare che le parole di codice sono divisibili per $D + 1$ e usare il teorema del resto; il resto della divisione di una parola di codice $c(D)$ per $D + 1$ è pari a $c(1)$. In particolare si può poi osservare che il polinomio generatore $g(D)$ ha un numero pari di coefficienti non nulli.

5.2 - Si può ottenere un codice a blocco con $(N, K) = (15, 10)$ e distanza minima $d = 4$ estendendo il codice di Hamming (15,11), ottenendo quindi un codice (16,11), e poi accorciandolo a (15,10). Il codice così ottenuto non è ciclico. Con il risultato dell'esercizio precedente si mostri che si ottiene più semplicemente un codice ciclico con gli stessi parametri utilizzando come polinomio generatore $g(D)$ il prodotto di $D + 1$ e del polinomio $D^4 + D + 1$, generatore del codice di Hamming (15,11).

5.3 - Si ottenga un codice a blocco di lunghezza $N = 8$ accostando ad una parola del codice a parità semplice di lunghezza 4 la stessa parola ripetuta,

oppure con tutti i bit negati. Ad esempio da 0101 si ottengono 01010101 e 01011010. Si mostri che $K = 4$ e $d = 4$. *Commento*: si tratta del codice Reed-Muller (8,4).

5.4 - Per decodificare in modo efficiente il codice dell'esercizio precedente si ricerchi dapprima la più verosimile tra le otto parole ottenute per ripetizione. Si mostri che basta considerare $r_1+r_5, \ldots, r_4+r_8$, e poiché si tratta di un codice a parità semplice ... Analogamente basta considerare $r_1 - r_5, \ldots, r_4 - r_8$ per determinare la più verosimile tra le altre otto parole. Infine basta confrontare le due correlazioni vincenti.

5.5 - Si considerino i codici Reed-Muller con $N = 32$, per tutti i valori di r da 0 a 5. Si mostri che solo per $r = 2$ si ottiene un nuovo codice; negli altri casi si ottengono rispettivamente i seguenti codici: a ripetizione, biortogonale, Hamming esteso, a parità semplice e universo.

5.6 - Si mostri che per un codice a blocco (N, K) esistono 2^K configurazioni d'errore (inclusa quella nulla, cioè senza errori) che producono la stessa sindrome. *Suggerimento*: ogni parola di codice dà sindrome nulla.

5.7 - Il guadagno asintotico, con decodifica ML, di un codice a blocco (N, K) con distanza minima d è dato da $G = \frac{K}{N}d$ (espressione del tutto analoga a quella per i codici convoluzionali). Analizzando le tabelle dei codici si vede che con codici complessi sono possibili guadagni asintotici molto elevati. Esiste ad esempio un codice BCH con $N = 1023$, $K = 543$ e $d \geq 107$ (il valore esatto non è noto). Si calcoli il guadagno asintotico, in dB. Anche tralasciando il problema della complessità del decodificatore, si mostri che guadagni così elevati non hanno alcun significato pratico. *Suggerimento*: il numero di concorrenti a distanza minima ...

5.8 - Se $p = Q(\sqrt{\frac{2E_b}{N_0}\frac{K}{N}})$ è la probabilità d'errore sul singolo bit di codice con decisione *hard*, si mostri che in primissima approssimazione la probabilità d'errore sul blocco per un codice correttore di t errori è $Q(\sqrt{\frac{2E_b}{N_0}\frac{K}{N}(t+1)})$. *Suggerimento*: $Q(x) \approx \exp(-x^2/2)$.

5.9 - In un sistema di trasmissione *binario antipodale* si utilizza come forma d'onda elementare una sequenza *pseudocasuale* $\{a_k\}$ di lunghezza N molto

grande. L'energia complessiva trasmessa è E e la densità spettrale del rumore, gaussiano, è $N_0/2$. Si determini il ricevitore ottimo, e la corrispondente probabilità di errore. Si spieghi, eventualmente solo in modo qualitativo, perché questo sistema di trasmissione è molto tollerante ai disturbi a banda stretta (ad esempio, una sinusoide ad una qualsiasi frequenza). Si consideri poi il seguente ricevitore, non ottimo: si prendono N decisioni binarie indipendenti e si contano le coincidenze di segno con la sequenza $\{a_k\}$. Si decide a favore dell'uno se le coincidenze superano $N/2$, a favore dello zero altrimenti. Si calcolino (approssimativamente) la probabilità di errore e la degradazione, in dB, rispetto al ricevitore ottimo. *Suggerimento*: si osservi che la probabilità d'errore sul singolo bit è poco minore di 1/2 (quanto?); si considerino poi il valor medio e la varianza del numero di coincidenze; infine si utilizzi l'approssimazione gaussiana della distribuzione binomiale.

5.10 - Con segnalazione *binaria antipodale* il miglior codice convoluzionale con *rate* $R = 1/2$ e due stati ha generatori 2,3. Si tracci il traliccio del codice e si determinino distanza minima e guadagno asintotico (in dB). Secondo la formula della capacità di canale, quale parte del guadagno può essere attribuita all'espansione di banda? *Commento*: il codice ha prestazioni così poco interessanti da non meritare di essere inserito nelle tabelle.

5.11 - Si determini la *free distance* d_f del codice convoluzionale a quattro stati con *rate* $R = 1/3$ e generatori (ottali) 7, 7 e 5. Quale è il guadagno asintotico del codice, con segnalazione binaria antipodale? *Suggerimento*: si tracci solo la parte strettamente necessaria del diagramma a traliccio.

5.12 - Si consideri il codice convoluzionale con *rate* $R = 1/2$ di fig. 5.3 con segnalazione binaria antipodale e si supponga che vengano cancellati, cioè non vengano trasmessi, i bit nelle posizioni $4, 8, 12, \ldots$ ottenendo quindi un codice *punctured* con *rate* $R = 2/3$. Si disegni il traliccio, e si osservi che l'insieme dei possibili eventi errore varia periodicamente. Si indichi quindi come occorre modificare le formule (5.4) e (5.7) per il calcolo della probabilità d'errore, e si determini il guadagno asintotico del codice.

5.13 - Un codice convoluzionale, se terminato facendo seguire ai bit d'informazione K zeri, è interpretabile anche come codice a blocco. Ad esempio dal codice a 64 stati con *rate* $R = 1/2$ di tab. 2, con 11 bit d'informazione e 6

zeri si ottiene un blocco di lunghezza $N = 34$, con distanza minima 10. Si noti che il codice BCH (31,11) pur con N minore ha distanza maggiore. Si provino altre combinazioni, per arrivare a conclusioni analoghe. Perché allora i codici convoluzionali sono così diffusi? *Suggerimento*: quale è la complessità del decodificatore *ottimo* per il codice BCH? e quali sarebbero le prestazioni del decodificatore *algebrico*?

5.14 - I migliori codici convoluzionali con *rate* $R = 2/3$ e 4,8,16,32,64 stati hanno distanza $d_f = 3, 4, 5, 6, 7$ rispettivamente. Si propongano esempi di codici a blocco ottenuti terminando con zeri come nell'esercizio precedente, e se ne valutino i parametri. *Suggerimento*: si faccia attenzione a quanti zeri occorrono.

5.15 - Si consideri un "codice" convoluzionale a due stati ($K = 2$) con *rate* $R = 1$ (!), con l'unico sommatore connesso a entrambe le celle del registro. La segnalazione sia binaria antipodale, su un canale con rumore additivo gaussiano bianco. Si tracci il diagramma a traliccio e si determini la distanza minima tra segnali corrispondenti a sequenze "codificate" diverse. Quale è, o sembra essere, il guadagno asintotico del codice (in dB)? Si osservi ora che *non c'è codice*, ma solo una precodifica differenziale dei dati $x_k = i_k + i_{k-1}$ (modulo 2), seguita da modulazione non codificata. Dunque *non può esserci guadagno.* Perché in questo caso la distanza minima non è un parametro significativo? *Suggerimento*: si tenti di usare lo *union bound* (5.4).

5.16 - Un sistema di trasmissione usa la segnalazione binaria antipodale ed il codice convoluzionale a quattro stati di fig. 5.3. Il rumore è additivo gaussiano bianco. Si consideri il seguente ricevitore: a partire da uno stato che si suppone *noto*, si correla il segnale ricevuto con quelli corrispondenti a *tutte* le ipotesi possibili per una durata pari a N bit d'informazione, e si sceglie il massimo; si assume poi che la *prima* transizione di stato così ottenuta sia corretta, cioè si decide *solo* per il *primo bit* d'informazione, e per lo *stato* così raggiunto; si prosegue allo stesso modo, a partire da questo stato. Trascurando la propagazione degli errori dovuta a decisioni errate, e quindi a stati di partenza errati, si calcolino le prestazioni asintotiche del ricevitore per diversi valori di N, e le si confrontino con quelle della stima della sequenza a massima verosimiglianza ottenuta con l'algoritmo di Viterbi. *Suggerimento*: per la linearità del codice è lecito assumere, nel calcolo delle prestazioni, che

sia stata trasmessa la sequenza di tutti zeri.

5.17 - Si consideri un sistema di trasmissione in cui le coppie di bit uscenti da un codice convoluzionale con *rate* $R = 1/2$ vengono associate a quattro segnali ortogonali, ad esempio mediante modulazione di frequenza o di posizione. La distanza tra le forme d'onda *non* è proporzionale alla distanza di Hamming: è infatti la stessa per coppie di bit che differiscono sia in una sola sia in due posizioni. Considerando il traliccio del banale codice convoluzionale dell'es. 10, e quello dell'ancora più semplice codice con generatori 2,1 che ha *free distance* $d_f = 2$ anziché 3, si mostri che le prestazioni sono *uguali*. Passando poi a codici con quattro stati, si determinino le prestazioni del codice con generatori 7,5 (ottimo per la segnalazione binaria antipodale) e si cerchi un codice più semplice che abbia le stesse prestazioni. *Commento*: la (5.5) vale *solo* in caso di segnalazione binaria antipodale.

5.18 - (*Codice di linea bipolare*) Si trasmette la forma d'onda

$$s(t) = \sum c_k g(t - kT)$$

dove $g(t)$ è una forma d'onda a radice di Nyquist. Se si indicano con a_k i simboli binari (equiprobabili) da trasmettere, il codice bipolare è descritto dalle transizioni di stato

$$s_k = s_{k-1} + a_k \qquad a_k = 0, 1;\ s_k = 0, 1$$

dove la somma è modulo due. I livelli trasmessi sono dati da

$$c_k = s_k - s_{k-1}$$

Si determini il diagramma a traliccio del codice, ponendo attenzione alla corrispondenza fra transizioni di stato e dati a_k; si calcoli la distanza minima tra le sequenze lecite; si confrontino le prestazioni asintotiche con quelle della segnalazione binaria antipodale, a pari energia media per bit. Si mostri che la probabilità d'errore *dipende* dalla sequenza trasmessa. Se questa è $x00..001..$, con m zeri consecutivi ($m = 0, 1, \ldots$) quale è il numero di sequenze a distanza minima? Utilizzando l'usuale maggiorazione della probabilità d'errore, e *tenendo conto* delle probabilità a priori delle sequenze di dati, si calcoli la probabilità degli eventi errore e la probabilità d'errore sui bit. *Suggerimento*: $\sum_{n=1}^{\infty} n2^{-n} = 2$.

5.19 - Per il codice convoluzionale con *rate* $R = 1/2$ e quattro stati si determini la funzione di trasferimento, data dal primo elemento di

$$(\mathbf{1} - A(D, I))^{-1} x_0(D, I)$$

Poi si sviluppi in serie e si confronti il risultato con gli eventi errore riconoscibili nel traliccio di fig. 5.5.

5.20 - Si dimostri la diseguaglianza (5.13). *Suggerimento*: in

$$Q(\sqrt{x+y}) = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{x+y}}^{\infty} \exp(-z^2/2) dz$$

si ponga $z^2 = t^2 + y$ e si osservi che $dz \leq dt$.

5.21 - Si supponga di avere un canale con rapporto segnale-rumore variabile nel tempo. Si mostri che occorre essere molto cauti nella scelta di un codice. In particolare con codici molto potenti, e quindi con curve di probabilità d'errore in funzione di E_b/N_0 ripidissime, si mostri che le prestazioni sono sostanzialmente determinate dal rapporto segnale-rumore *peggiore*. *Suggerimento*: se per metà del tempo la probabilità d'errore è 10^{-10} e per l'altra metà è 10^{-1}, la probabilità d'errore è ... *Commento*: si ottiene un netto miglioramento, a costo di un maggior ritardo, mediante l'*interleaving*.

5.22 - Volendo utilizzare un codice convoluzionale tradizionale, come quelli di tab. 2, per una costellazione QAM con il *mapping* di fig. 5.1 si mostri che, almeno ad alto rapporto segnale-rumore, risulta poco utile scegliere codici con un numero elevato di stati; si noti infatti che il guadagno asintotico è limitato dalla presenza di bit non codificati. Si confrontino le prestazioni con quelle di costellazioni bidimensionali non codificate che trasmettono lo stesso numero di bit per simbolo.

5.23 - Si determinino le prestazioni del codice TCM di fig. 5.16, utilizzato con la costellazione 8PAM con *mapping* naturale di fig. 5.13.

5.24 - Si provi a disegnare il traliccio per una modulazione 8PSK codificata a quattro stati con otto *subset*. Si mostri che si riesce ad ottenere lo stesso guadagno asintotico del codice con quattro *subset* e con transizioni parallele,

ma si ha un numero maggiore di concorrenti a distanza minima e quindi un codice dalle prestazioni inferiori.

5.25 - Si consideri un codice TCM a otto stati con generatori (ottali) 15,23,04. Si calcoli il guadagno asintotico nel caso di costellazione 8PSK e di costellazione QAM. *Suggerimento*: per semplificare il calcolo si consideri un numero M di punti molto elevato.

5.26 - In fig. 5.16, indicando con $i(D)$ la sequenza di bit d'informazione e con $c_1(D)$ e $c_2(D)$ le sequenze di bit codificati si ha $c_1(D) = i(D)(1 + D^2)$ e $c_2(D) = i(D)D$. Se poi si pone $i'(D) = i(D)(1 + D^2)$ si ha $c_1(D) = i'(D)$ e $c_2(D) = i'(D)D/(1 + D^2)$. Il codificatore può quindi essere ottenuto con una precodifica differenziale $1 + D^2$ seguita da un circuito lineare retroazionato che realizza la funzione di trasferimento $D/(1 + D^2)$. Quale è la struttura di tale circuito? *Commento*: l'insieme delle sequenze codificate, e quindi l'insieme delle distanze tra i segnali, non varia se si elimina la precodifica differenziale. Si ottiene quindi un codificatore *sistematico* con *retroazione*. È questa la struttura dei codificatori che si trovano nelle tabelle di Ungerboeck.

5.27 - Si consideri il codice TCM a otto stati dell'es. 5.25, con *rate* $R = 2/3$ e generatori (ottali) 15,23,04. Dette $i_1(D)$ e $i_2(D)$ le sequenze corrispondenti alle coppie di bit d'informazione, si mostri che $c_1(D) = i_1(D)(D + D^2) + i_2(D)$, $c_2(D) = i_1(D)(1 + D^2) + i_2(D)D$ e $c_3(D) = i_1(D)D$. Allo scopo di ottenere un codificatore sistematico si ponga $i'_1(D) = c_1(D)$ e $i'_2(D) = c_2(D)$ e si mostri che $c_3(D) = (i'_1(D)D^2 + i'_2(D)D)/(1 + D^3)$. Quale è quindi la struttura del codificatore *sistematico* con *retroazione* così ottenuto? *Commento*: nelle tabelle di Ungerboeck il codificatore appare in questa forma.

5.28 - In presenza di rumore additivo gaussiano bianco e su un canale ideale si trasmette il segnale

$$s(t) = \sum a_k g(t - kT)$$

dove le $g(t - kT)$ sono forme d'onda ortogonali con banda $0.7/T$. I livelli a_k sono ottenuti con un codice convoluzionale con *rate* $R = 1/2$ e generatori 7,2 (ottale), e *mapping* naturale: 00,01,10 e 11 corrispondono rispettivamente ad $a_k = -3, -1, 1$ e 3.

- Quanti bit d'informazione per simbolo si trasmettono, e quale è la banda richiesta per trasmettere 10 Mb/s?
- Quale è (approssimativamente) il valore di E_b/N_0 richiesto per ottenere $P(E) = 10^{-5}$?
- Come è fatto il ricevitore?
- Rispetto ad un sistema non codificato di pari efficienza spettrale, il codice dà un guadagno anche sulla potenza di picco (anziché media)?
- Si mostri che la sequenza a distanza minima da quella di tutti zeri ha distanza *maggiore* della minima garantita.
- Cosa cambierebbe se si volessero trasmettere due bit per dimensione, con lo stesso codice convoluzionale? Si avrebbe guadagno in potenza di picco rispetto ad un sistema non codificato di pari efficienza spettrale?
- Dalle tabelle dei codici TCM si troverebbero i generatori 5,2 (ottale) anziché 7,2. Confrontando i codificatori e i corrispondenti tralicci si mostri che i due codici si equivalgono. *Suggerimento*: se il secondo bit codificato è uno zero ...; se è un uno ...

5.29 - Dalle tabelle dei codici TCM per modulazione d'ampiezza a otto livelli si ricava un codice con generatori (ottali) 15,02. Quale è la struttura del codificatore?
Nelle stesse tabelle si trova anche che il guadagno asintotico rispetto alla modulazione d'ampiezza non codificata di pari efficienza spettrale è 3.77 dB. Indicando i livelli trasmessi nel caso codificato con $\pm\sqrt{E}$, $\pm 3\sqrt{E}$, $\pm 5\sqrt{E}$ e $\pm 7\sqrt{E}$, quale è quindi la distanza minima tra sequenze concorrenti? *Suggerimento*: nel caso non codificato $d^2 = \ldots E_b$; quindi nel caso codificato $d^2 = \ldots E_b$; inoltre $E_b = \ldots E$, e quindi $d^2 = \ldots E$.
Si cerchi nel traliccio l'evento errore corrispondente (si osservi che l'ultimo passo nel traliccio è necessariamente dallo stato 001 allo stato 000, e fornisce $d^2 = \ldots$; quindi basta cercare percorsi fino allo stato 001 con $d^2 = \ldots$). Tenendo conto delle transizioni parallele, quanti possono essere gli eventi errore a distanza minima?

5.30 - Siano disponibili codificatore e decodificatore per un codice TCM a otto stati per la modulazione 64QAM. Come si possono riutilizzare questi

componenti, con piccole modifiche, per un sistema di trasmissione in banda base con modulazione d'ampiezza a otto livelli?
Secondo le tabelle, il codice TCM per la modulazione 64QAM ha un guadagno asintotico di 3.77 dB rispetto alla trasmissione non codificata con costellazione 32CR (una costellazione 6x6 da cui sono tolti i quattro punti con energia maggiore). Da questi dati si determini (approssimativamente) il valore di E_b/N_0 richiesto dal sistema in banda base per ottenere $P(E) = 10^{-5}$.
I generatori del codice sono (in ottale) 15,23,04. Quale è quindi la struttura del codificatore?

5.31 - Si trasmette una successione di simboli a_k a quattro livelli, associati a coppie di bit (00 = -3; 01 = -1; 10 = 1; 11 = 3). I simboli vengono codificati a blocchi di otto, con le seguenti regole: i primi bit degli otto simboli non sono codificati, cioè sono del tutto liberi; gli otto secondi bit sono parole del codice di Hamming esteso $(8, 4)$. Quali prestazioni ha il ricevitore ottimo? e quale potrebbe esserne la struttura?

5.32 - In presenza di rumore additivo gaussiano bianco, si trasmette una successione di simboli d_k con modulazione 4PSK e con il seguente codice a blocco di lunghezza pari a quattro simboli: dette 0, 1, 2 e 3 le quattro fasi 0, $\pi/2$, π e $3\pi/2$, i quattro simboli hanno fasi tutte pari oppure tutte dispari; inoltre la somma delle fasi è divisibile per 4. Quanti bit per simbolo si trasmettono? Quale è la struttura del ricevitore ottimo? Quali le prestazioni? Quale codice convoluzionale darebbe banda occupata e prestazioni comparabili?

5.33 - Si consideri una modulazione 4PSK codificata a blocchi di quattro simboli, con la seguente regola: numerate le fasi di ciascun simbolo da 0 a 3, la somma delle fasi dei due simboli è pari. Quanti bit per simbolo si trasmettono? Quali sono le prestazioni? Come può essere fatto il ricevitore?

5.34 - Si consideri una modulazione 8PSK codificata a blocchi di N simboli, in cui la somma delle fasi è pari. Quanti bit per simbolo si trasmettono? Quali sono le prestazioni? È conveniente un valore di N molto grande? Come può essere fatto il ricevitore?

5.35 - Si trasmette una successione di simboli d_k tratti dalla costellazione 16QAM con il *mapping* di fig. 5.15. I simboli vengono codificati a blocchi

di quattro, con le seguenti regole: i quarti bit dei quattro simboli sono tutti zeri o tutti uni; il terzo bit del quarto simbolo è la somma modulo 2 di quelli dei primi tre simboli (e quindi in terza posizione c'è un numero pari di uni); i secondi e i primi bit non sono codificati, cioè sono del tutto liberi. Quanti bit per simbolo si trasmettono, e quali sono le prestazioni del codice?

5.36 - Come nell'esercizio precedente, ma con blocchi di otto simboli e con i seguenti codici: i quarti bit sono tutti zeri o tutti uni; i terzi bit danno parole del codice di Hamming esteso (8,4); i secondi bit danno il codice a parità semplice (8,7); i primi bit non sono codificati, cioè sono del tutto liberi.

5.37 - Si trasmette una successione di simboli d_k tratti dalla costellazione 16QAM di fig. 5.1 utilizzando una forma d'onda equivalente in banda base $g(t)$ e le sue repliche ortogonali $g(t - kT)$.
I simboli vengono codificati a blocchi di quattro, con le seguenti regole: le quattro coppie formate da terzo e quarto bit di ogni simbolo, prese nell'ordine, sono parole del codice di Hamming esteso con $N = 8$ e $K = 4$; i secondi e i primi bit non sono codificati, cioè sono del tutto liberi. Quali sono le prestazioni del codice?

5.38 - Si consideri il decodificatore multistadio per la modulazione 8PSK codificata a blocchi di $N = 8$ simboli, citata nel testo, con codici rispettivamente a ripetizione, a parità singola e universo. In particolare si calcoli la verosimiglianza delle due ipotesi relative al primo codice (fasi tutte pari o tutte dispari) *ignorando* il secondo codice, cioè considerando possibili (ed equiprobabili a priori) tutte le N-ple di terne di bit ..0 o ..1, rispettivamente. Come si può eventualmente semplificare il calcolo tenendo conto solo dei termini più importanti?

Capitolo 6

Equalizzazione

6.1 Introduzione

Nei capitoli precedenti si è implicitamente fatto conto di poter trasmettere sequenze di forme d'onda senza interferenza reciproca. In un numero considerevole di casi ciò è effettivamente possibile. Infatti se si riceve il segnale $r(t) = \sum a_k g(t - kT) + n(t)$, o nel caso complesso l'equivalente passa basso $z(t) = \sum d_k g(t - kT) + n(t)$, e se le repliche traslate della forma d'onda elementare $g(t)$ sono ortogonali, i simboli successivi vanno ad occupare via via nuove dimensioni dello spazio dei segnali e le correlazioni con le funzioni base $g(t - kT)$ forniscono via via le coordinate del segnale ricevuto senza *interferenza intersimbolica* (ISI: *InterSymbol Interference*). Se poi i dati sono indipendenti, ovvero se non c'è un codice che vieti alcune sequenze, le decisioni possono essere prese indipendentemente simbolo per simbolo, ed il ricevitore ottimo è semplice; nel caso codificato l'elaborazione delle coordinate del vettore ricevuto, più complessa, segue le linee indicate nel Cap. 5.

Se invece le forme d'onda $g(t - kT)$ non sono ortogonali il ricevitore è tutt'altro che banale anche senza codice, come si vedrà. L'ortogonalità, facile da ottenere in trasmissione, è bensì richiesta in ricezione. In altre parole, occorre fare i conti con il canale. Se questo non è troppo distorto, ma soprattutto se le sue caratteristiche sono note e non variano nel tempo, si può provvedere modificando opportunamente la forma d'onda trasmessa (si veda però l'es. 6.1). Se il canale varia nel tempo in modo impredicibile bisogna rassegnarsi alla mancata ortogonalità, e si può sperare di riottenere l'informazione trasmessa solo se, in modo esplicito o implicito, si riesce a

misurare la risposta impulsiva del canale, cioè a determinare la forma d'onda elementare ricevuta $g(t)$.

Non di rado i canali radio, in particolare i ponti radio e i sistemi radiomobili, hanno questa sfavorevole caratteristica a causa dei percorsi multipli (*multipath*) da antenna trasmittente a ricevente, dovuti a riflessione e diffrazione dei raggi. Poiché le lunghezze dei percorsi sono diverse, i contributi si sommano con fasi e ritardi differenti. Il divario tra i ritardi può essere ignorato se è piccolo rispetto al tempo di simbolo (grossolanamente, fino a $0.2T$). Comunque differenze di percorso anche piccole, pari ad alcune lunghezze d'onda, rendono pressoché casuali le fasi dei vari contributi. Il numero dei raggi ricevuti, e l'eventuale presenza di un termine dominante (*raggio diretto*) determina la distribuzione dell'ampiezza ricevuta. Se questa è nota, il progetto può essere fatto abbastanza semplicemente calcolando la probabilità d'errore condizionata all'ampiezza ricevuta, e poi mediando rispetto alla distribuzione di questa.

Se la dispersione dei ritardi è forte, la risposta impulsiva del canale radio non è rappresentabile solo con un guadagno complesso. La funzione di trasferimento non è piatta nella banda del segnale, le forme d'onda trasmesse vengono deformate e allungate e l'ortogonalità è persa. Nel caso di mezzi mobili, poi, la risposta impulsiva del canale varia molto rapidamente; infatti basta che il mezzo mobile si sposti di frazioni di lunghezza d'onda perché i raggi si ricompongano con fasi totalmente diverse.

Anche altri mezzi trasmissivi, ad esempio linee e cavi, introducono forti distorsioni, anche se di norma lentamente variabili nel tempo. Nella trasmissione numerica mediante *modem* su linee telefoniche commutate si incontrano, da volta a volta, canali molto diversi.

Per introdurre i concetti fondamentali conviene per il momento supporre nota la forma d'onda ricevuta $g(t)$, cosa del resto vera in alcuni casi.

6.2 Stima della sequenza a massima verosimiglianza

Prima di analizzare le numerose e semplici soluzioni *ad hoc* al problema dell'equalizzazione di canale, alcune proposte già una quarantina di anni addietro, è opportuno chiedersi se e quali indicazioni darebbe la teoria svolta nel Cap. 2. Si supponga di aver trasmesso N simboli. Il ricevitore ML cerca il massimo, rispetto alla sequenza di dati ad M livelli $\{d_k\}$, di $2\mathbf{r} \cdot \mathbf{s}_i - |\mathbf{s}_i|^2$

ovvero di[1]

$$\mathrm{Re}\{\int z(t)\sum_{k=1}^{N} d_k^* g^*(t-kT)dt\} - \frac{1}{2}\int \sum_{n=1}^{N}\sum_{k=1}^{N} d_n d_k^* g(t-nT)g^*(t-kT)dt = \tag{6.1}$$
$$= \sum_{k=1}^{N} \mathrm{Re}\{y_k d_k^*\} - \frac{1}{2}\sum_{n=1}^{N}\sum_{k=1}^{N} d_n d_k^* R_{k-n}$$

dove $g(t)$ è l'equivalente passa basso della forma d'onda elementare *ricevuta*; inoltre

$$y_k = \int z(t) g^*(t-kT)dt \tag{6.2}$$

è la correlazione del segnale ricevuto $z(t)$ con $g^*(t-kT)$, ovvero, se si preferisce, l'uscita del filtro adattato all'istante kT, e

$$R_{k-n} = \int g(t-nT)g^*(t-kT)dt \tag{6.3}$$

è l'autocorrelazione di $g(t)$ valutata in $t=(k-n)T$. È importante osservare che i valori dell'autocorrelazione R_l possono essere precalcolati e che solo alcuni sono diversi da zero, ad esempio quelli per $|l| \leq L$, essendo limitata la durata di $g(t)$. Inoltre si noti che $R(-l) = R^*(l)$.

Basta quindi calcolare le componenti (non ortogonali) y_k del segnale ricevuto, poi tutte le verosimiglianze (che sono, ad esempio, 2^N nella trasmissione binaria non codificata) e infine scegliere il massimo. Vale la pena di osservare che le correlazioni y_k, ovvero le uscite del filtro adattato, costituiscono una *statistica sufficiente*, cioè quanto occorre e basta per la stima a massima verosimiglianza della sequenza (MLSE: *Maximum Likelihood Sequence Estimation*) anche se non sono componenti ortogonali.

Quanto alla complessità, a prima vista sembra che la crescita esponenziale con N del numero di correlazioni renda il calcolo impraticabile, ma non è così in realtà. Infatti se si considera la somma parziale J_K che della (6.1) include tutti e soli i dati con indice minore o uguale a K, si può notare che essa può

[1] per maggior generalità si considera la trasmissione in banda passante; un'espressione analoga vale nel caso di segnali in banda base

essere calcolata recursivamente nel modo seguente

$$J_K(d_1,\ldots,d_{K-1},d_K) = J_{K-1}(d_1,\ldots,d_{K-1})+$$
$$+\mathrm{Re}\{y_K d_K^*\} - \frac{1}{2}|d_K|^2 R_0 - \mathrm{Re}\{d_K^* \sum_{l=1}^{L} d_{K-l} R_l\} \tag{6.4}$$

dove si è utilizzata la simmetria complessa coniugata dell'autocorrelazione, e dove per $K \leq L$ vanno ignorati i termini con indice $K-l \leq 0$. Se si definisce *stato* la L-pla $d_{K-L},\ldots,d_{K-1}$ si vede che J_K viene aggiornato sommando a J_{K-1} il contributo $\mathrm{Re}\{y_K d_K^*\}$ del campione ricevuto e una funzione, che può essere precalcolata e memorizzata, del dato attuale e dello stato, cioè una funzione della transizione di stato. Si vede infatti che, come per i codici convoluzionali, lo stato evolve come il contenuto di un registro a scorrimento. Quindi il calcolo può procedere secondo l'algoritmo di Viterbi, con M^L stati.

Le prestazioni sono valutabili con le solite tecniche, quali lo *union bound*. La distanza al quadrato tra due sequenze concorrenti è

$$|\mathbf{s}_i - \mathbf{s}_j|^2 = 4 \sum_{k=1}^{\infty} \left(|\varepsilon_k|^2 R_0 + 2\mathrm{Re}\{\sum_{l=1}^{L} \varepsilon_k \varepsilon_{k-l}^* R_l\} \right) \tag{6.5}$$

dove $\varepsilon_k = (d_k^i - d_k^j)/2$; ad esempio $\varepsilon_k = -1, 0, 1$ nel caso binario. Si ignorano naturalmente i termini con $k-l \leq 0$. Particolarmente importante sarà, al solito, la distanza minima.

La complessità dell'elaborazione MLSE non è piccola; soprattutto se la modulazione non è binaria il numero di stati M^L può essere molto elevato; si pensi per esempio ad una modulazione M-QAM. Una considerevole limitazione è poi che $g(t)$ deve essere nota, o stimata in qualche modo. Eppure in alcuni casi importanti, ad esempio nel sistema radiomobile numerico GSM, la via risulta praticabile. Molto più spesso, però, si ricorre alle soluzioni descritte nel seguito, più antiche e molto più semplici.

6.3 Equalizzazione "zero-forcing" e "decision-feedback"

In molti casi pratici non si ha né il tempo né la possibilità di stimare la forma d'onda elementare $g(t)$. Ciò pone un primo problema: quale filtro

di ricezione usare, cioè con quale forma d'onda correlare il segnale ricevuto? È chiaro che si ricorrerà ad un filtro prefissato, con ogni probabilità quello progettato per il canale indistorto. Non si deve infatti pensare che il canale sia sempre distorto: per una frazione considerevole del tempo i ponti radio non subiscono il *multipath*, anche se occorre predisporre contromisure per i momenti sfavorevoli.

Poiché in generale il filtro non è adattato, non si ottiene una statistica sufficiente e si ha una qualche degradazione delle prestazioni. Inoltre la risposta complessiva a valle del filtro di ricezione non è simmetrica.

Conviene supporre ancora per un momento nota la forma d'onda ricevuta $g(t)$. Considerando per semplicità la trasmissione in banda base, è immediato esprimere i campioni y_k come combinazioni lineari dei dati a_k più rumore:

$$y_k = \sum h_i a_{k-i} + n_k \tag{6.6}$$

dove la risposta impulsiva discreta $\{h_i\}$ è data dai campioni della convoluzione di $g(t)$ con la risposta impulsiva del filtro di ricezione. È anche facile calcolare l'autocorrelazione dei campioni del rumore n_k. Se, come per semplicità si supporrà nel seguito, il rumore è bianco e il filtro di ricezione è *radice di Nyquist* la sequenza $\{n_k\}$ è bianca.

Si consideri un caso semplicissimo: $h_0 = 1$, $h_1 = A$ e tutti gli altri nulli, ovvero $y_k = a_k + Aa_{k-1} + n_k$. Si ha interferenza solo da parte del simbolo che precede l'attuale, e la funzione di trasferimento è

$$H(z) = 1 + Az^{-1} \tag{6.7}$$

Il primo equalizzatore che viene in mente è, se esiste, il filtro inverso

$$\frac{1}{H(z)} = \frac{1}{1 + Az^{-1}} = 1 - Az^{-1} + A^2 z^{-2} - A^3 z^{-3} + \ldots \tag{6.8}$$

Lo sviluppo in serie vale, e il filtro inverso esiste, solo se $|A| < 1$.

L'equalizzatore potrebbe essere in teoria un filtro IIR con un solo polo, e dunque facile da realizzare. Gli equalizzatori IIR, pur studiati, trovano tuttavia rarissime applicazioni perché risulta sempre problematico garantirne la stabilità, oltre che l'aggiornamento adattativo dei coefficienti, e verranno ignorati nel seguito. Resta piuttosto la possibilità di realizzare un equalizzatore approssimato, mediante un filtro FIR con risposta impulsiva c_i data dalla (6.8) troncata ad un numero finito di termini. È quasi una regola che l'equalizzatore risulti in teoria infinito, ma in pratica sia realizzato con pochi coefficienti

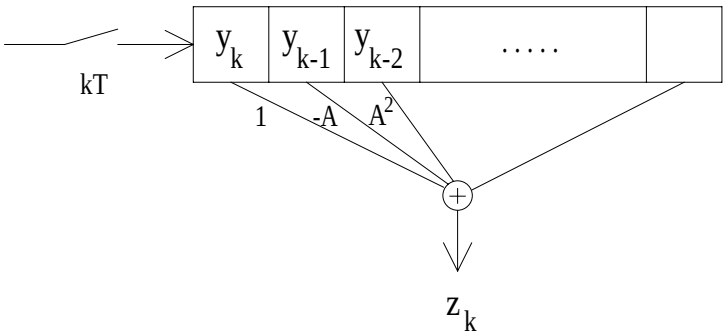

Fig. 6.1 - Struttura dell'equalizzatore per il canale $1 + Az^{-1}$

(*prese*, nel gergo tecnico) contando sul fatto che la risposta si esaurisce rapidamente.

La struttura dell'equalizzatore è quella rappresentata in modo schematico in fig. 6.1, dove ogni ramo include un moltiplicatore. I campioni del rumore, che supponiamo indipendenti, si trovano pesati e sommati all'uscita, con varianza data da

$$\sigma^2 = \sum c_i^2 \sigma_n^2 = \frac{\sigma_n^2}{1 - A^2} \tag{6.9}$$

dove per semplificare il calcolo si è considerato un numero infinito di prese. L'ISI viene quindi cancellata a costo di un incremento del rumore, modesto se il canale è poco distorto ($|A| \ll 1$) ma inaccettabile se $|A|$ è prossimo all'unità. Poiché viene azzerata la risposta impulsiva complessiva $\{h_i\} * \{c_i\}$ (per $i \neq 0$) la tecnica di equalizzazione viene detta *zero-forcing* e indicata con la sigla ZF.

Un'altra tecnica che facilmente viene in mente è basata sull'osservazione che quando si prende una decisione sul simbolo a_k si sono già prese le precedenti. È quindi disponibile $\hat{a}_{k-1}$, e si può pensare di moltiplicarlo per A e sottrarlo da y_k, ottenendo

$$z_k = y_k - A\hat{a}_{k-1} = a_k + A(a_{k-1} - \hat{a}_{k-1}) + n_k \tag{6.10}$$

e quindi un'equalizzazione perfetta senza incremento del rumore, perlomeno se il dato precedente deciso è corretto, cosa che accadrà con elevata probabilità nel funzionamento normale. La tecnica è detta *decision-feedback* (DF) e può cancellare altrettanto bene, se occorre, il contributo di altri dati precedenti.

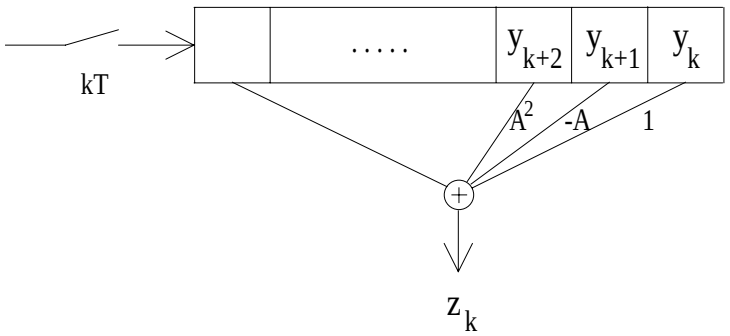

Fig. 6.2 - Struttura dell'equalizzatore per il canale anticausale $1 + Az$

Resta il dubbio di cosa può accadere se qualcuna delle decisioni precedenti è errata. L'ISI viene aumentata, anziché cancellata, e la probabilità che la decisione sul simbolo attuale sia anch'essa errata aumenta notevolmente. In generale si potrebbe temere un effetto a valanga della *propagazione degli errori*, molto difficile da studiare teoricamente ma in pratica quasi mai tragico.

Ora una piccola variante: $h_0 = 1$, $h_{-1} = A$ e tutti gli altri nulli, ovvero $y_k = a_k + Aa_{k+1} + n_k$. Si ha interferenza solo da parte del simbolo che segue l'attuale[2], e la funzione di trasferimento è

$$H(z) = 1 + Az \tag{6.11}$$

L'equalizzatore ZF è, se $|A| < 1$, il filtro inverso

$$\frac{1}{H(z)} = \frac{1}{1 + Az} = 1 - Az + A^2z^2 - A^3z^3 + \dots \tag{6.12}$$

La novità è che la risposta è anticausale. L'equalizzatore è realizzabile solo se, oltre a troncare la risposta, si introduce un ritardo pari alla lunghezza del filtro, cosa peraltro accettabile in tutti i casi pratici. La struttura è indicata in fig. 6.2.

La varianza del rumore all'uscita è ancora data dalla (6.9), e dunque non vi è alcuna differenza sostanziale tra il caso di risposta impulsiva causale (ed equalizzatore causale) o anticausale (ed equalizzatore anticausale).

[2]l'interferenza da parte di un simbolo futuro non ha nulla di strano; non si dimentichi che si ignorano sia il ritardo nella generazione del segnale trasmesso sia quello introdotto dal filtro di ricezione

La tecnica DF è invece disarmata, non essendo ancora disponibili nel funzionamento in tempo reale le decisioni future.

In pratica il canale non sarà né causale né anticausale, e si dovrà prevedere un equalizzatore che abbia una presa *centrale*, alcune causali ed altre anticausali. A parità di numero di prese, cioè di numero di moltiplicazioni e somme da eseguire per tempo di simbolo, la posizione della presa centrale è un grado di libertà a disposizione del progettista[3].

Vediamo ancora alcuni semplici esempi. Sia $h_0 = 1$, $h_1 = h_{-1} = A$, ovvero

$$\begin{aligned} H(z) &= 1 + Az^{-1} + Az = \frac{1}{1+B^2}(1+Bz^{-1})(1+Bz) \\ B &= \frac{2A}{1+\sqrt{1-4A^2}} \end{aligned} \tag{6.13}$$

Il calcolo della risposta dell'equalizzatore, che risulta simmetrica, e della varianza del rumore è lasciato al lettore, insieme alla verifica che la condizione $|B| < 1$ si traduce in $|A| < 0.5$ (es. 6.2). Quest'ultimo risultato è ovvio se si pensa che per $|A| = 0.5$ si hanno due zeri sul cerchio unitario, e quindi $H(z)$ non è invertibile.

Ci si può chiedere se la tecnica DF può intervenire almeno sulla parte causale. In effetti inviando i campioni y_k all'equalizzatore lineare anticausale $(1+B^2)/(1+Bz)$ si ottiene la risposta complessiva $1 + Bz^{-1}$, e si può poi completare l'opera con un equalizzatore DF.

Il calcolo della varianza del rumore, lasciato al lettore (es. 6.3), dà un risultato migliore dell'equalizzatore lineare ZF (a parte l'effetto della propagazione degli errori, difficile da valutare).

Un ultimo esempio: $h_0 = 1$, $h_1 = A$, $h_{-1} = -A$, ovvero

$$\begin{aligned} H(z) &= 1 + Az^{-1} - Az = \frac{1}{1-B^2}(1+Bz^{-1})(1-Bz) \\ B &= \frac{2A}{1+\sqrt{1+4A^2}} \end{aligned} \tag{6.14}$$

Il calcolo della risposta dell'equalizzatore e della varianza del rumore è lasciato

[3] come dire che si dovranno fare più progetti, tra cui poi scegliere; la posizione migliore per la presa *centrale* non risulta mai molto lontana dal centro

al lettore, insieme alla verifica che la condizione $|B| < 1$ è sempre soddisfatta (es. 6.4). L'esame della funzione di trasferimento sul cerchio unitario mostra la profonda differenza rispetto al caso precedente, non immediatamente percepibile dalla risposta impulsiva: si ha prevalentemente distorsione di fase, anziché d'ampiezza.

Naturalmente anche in quest'ultimo caso si possono combinare le tecniche ZF e DF (es. 6.5).

Tutte le considerazioni svolte in questa sezione si estendono senza difficoltà al caso complesso, ovvero alla trasmissione in banda passante. Basta osservare che non solo i dati e i campioni y_k, ma anche i coefficienti c_i dell'equalizzatore sono complessi. Infatti la distorsione del canale fa interferire reciprocamente i dati in fase e in quadratura, e solo un'operazione analoga può rimuovere l'ISI. Il lettore è invitato a ripercorrere i semplici esempi già visti, con A complesso (es. 6.6).

6.4 Equalizzazione a minimo errore quadratico medio

Per canali con forti distorsioni d'ampiezza la cancellazione completa dell'ISI è troppo onerosa. È preferibile una equalizzazione incompleta, che lasci un po' di ISI ma non esalti troppo il rumore. Come si vedrà alla fine del capitolo, non esiste un'espressione trattabile analiticamente per il contributo di interferenza e rumore alla probabilità d'errore. Un criterio non del tutto inadeguato, ma soprattutto trattabile, è la potenza complessiva del disturbo

$$E[(z_k - a_k)^2] = E[(\sum c_i y_{k-i} - a_k)^2] \tag{6.15}$$

dove non si sono indicati esplicitamente gli indici della somma perché dipendono dalla ripartizione delle prese tra *passato* e *futuro*. Fissato il numero N e la posizione delle prese, si possono scegliere i coefficienti c_i in modo da minimizzare l'errore quadratico medio. Il corrispondente equalizzatore viene indicato con la sigla MMSE (*Minimum Mean-Square Error*) oppure anche MSE (*Mean-Square Error*), o ancora con LMS (*Least-Mean-Square*).

I valori delle prese si trovano risolvendo un sistema di N equazioni lineari che ha come coefficienti la matrice di covarianza dei campioni y_k e come termini noti le correlazioni tra il dato desiderato a_k ed i campioni y_{k-i} (es. 6.8). Si potranno poi calcolare la risposta complessiva, non ideale,

e la varianza del rumore all'uscita dell'equalizzatore (es. 6.8), e infine (numericamente) la probabilità d'errore. Il calcolo, ripetuto per molti canali, rappresentativi di quelli che si potranno effettivamente incontrare, e per molte configurazioni dell'equalizzatore (numero e disposizione delle prese) porterà alle scelte di progetto. Normalmente le prestazioni risultano migliori di quelle dell'equalizzazione ZF, ma tendono a diventare simili per alto rapporto segnale-rumore[4].

L'equalizzazione MMSE può infine essere combinata con la tecnica DF: i coefficienti c_i per $i \leq 0$ sono determinati come appena visto; quelli in feedback di conseguenza, in modo da cancellare completamente l'ISI proveniente dai dati corrispondenti (es. 6.11).

È possibile tuttavia che le decisioni immediatamente *precedenti* non siano ancora disponibili. Ad esempio a velocità elevate se l'*hardware* dell'equalizzatore è in *pipeline* le decisioni precedenti, pur avendo un vantaggio temporale di $T, 2T, \ldots$ su quella attuale, sono *latenti*. Si può quindi usare solo un equalizzatore lineare. Anche nel caso di trasmissione codificata le decisioni vengono prese con troppo ritardo per poter essere retroazionate in tempo utile.

6.5 Equalizzazione adattativa MMSE

Vediamo ora come l'equalizzatore può determinare i valori appropriati dei coefficienti c_i e continuare ad adattarli per inseguire le variazioni della risposta del canale. Ci si concentrerà sulla cosiddetta tecnica del *gradiente stocastico* tralasciandone altre più veloci ma anche più complesse, richieste solo in casi particolari.

La funzione (6.15) è una *forma quadratica* nei coefficienti c_i, che si può dimostrare definita positiva (ovvero, ma non sorprende, con un unico minimo). La ricerca del minimo di questo *paraboloide* in N dimensioni può essere fatta iterativamente, a piccoli passi, seguendo la direzione di massima pendenza, che è quella opposta al gradiente. Se il paraboloide fosse *rotondo* si punterebbe costantemente verso il punto di minimo. Poiché non lo è il cammino sarà più tortuoso, ma continuando a scendere non si può che arrivare al fondo. Per quanto riguarda la lunghezza dei passi, è ragionevole fare passi più decisi dove la pendenza è forte, più piccoli dove si addolcisce (quindi ad esempio passi

[4] se c'è poco rumore l'equalizzatore MMSE cerca di azzerare, o quasi, l'ISI ed è quindi equivalente all'equalizzatore ZF

proporzionali al modulo del gradiente). Il gradiente ha come componenti

$$\frac{\partial}{\partial c_i} E[(z_k - a_k)^2] = 2E[(z_k - a_k)\frac{\partial}{\partial c_i}\sum c_i y_{k-i}] = 2E[(z_k - a_k)y_{k-i}] \quad (6.16)$$

Il valor medio può essere stimato accumulando e mediando somme di prodotti tra l'errore $z_k - a_k$ ed il contenuto y_{k-i} della i-esima cella dell'equalizzatore. I dati a_k, non disponibili, verranno sostituiti dai dati decisi $\hat{a}_k$, che sono praticamente sempre corretti almeno nel funzionamento a regime. L'operazione di media richiederebbe di lasciar passare un numero prefissato di campioni, e solo alla fine del conteggio aggiornare i valori dei c_i. Si trova più semplice utilizzare ogni prodotto $(z_k - \hat{a}_k)y_{k-i}$ non appena disponibile[5] e quindi l'aggiornamento dei coefficienti viene eseguito simbolo per simbolo:

$$c_i^{k+1} = c_i^k - \alpha(z_k - \hat{a}_k)y_{k-i} \quad (6.17)$$

La costante α determina il passo dell'algoritmo del gradiente stocastico. L'aggiornamento ha una componente casuale[6] e non si arresta, se non in media, neppure una volta raggiunto il minimo. È evidente che valori elevati di α sono più adatti durante la ricerca del minimo, mentre sono richiesti valori più piccoli per avere piccole fluttuazioni a regime. Senza voler entrare in dettagli, si provvederà a commutare tra due valori di α sulla base di una stima del valore attuale dell'errore quadratico medio.

Valori troppo elevati di α producono fluttuazioni così ampie che l'errore quadratico medio alimenta sé stesso e il controllo diventa instabile. È stato mostrato che il valore di α più conveniente durante la fase iniziale, cioè quello che dà la convergenza più rapida, è prossimo a $1/NE[y_k^2]$. A regime però le fluttuazioni casuali dell'errore quadratico medio non risultano trascurabili; mediamente il disturbo è circa doppio del minimo teorico, per cui si consiglia di ridurre α di cinque volte.

All'inizio della procedura i coefficienti dell'equalizzatore vengono tutti azzerati, tranne quello *centrale* posto uguale a 1. Se il canale è molto distorto è possibile, soprattutto nella trasmissione multilivello, che i dati decisi $\hat{a}_k$ siano così spesso errati da non poter ottenere una stima accettabile del gradiente, ovvero da non far convergere l'equalizzatore. La soluzione tradizionale, ove possibile, è di trasmettere una sequenza prefissata e quindi nota al ricevitore

[5] perché sprecare il tempo attendendo la fine del conteggio?

[6] si potrebbe dire algoritmo del *gradiente casuale*; l'aggettivo *stocastico*, pur sinonimo, è più elegante

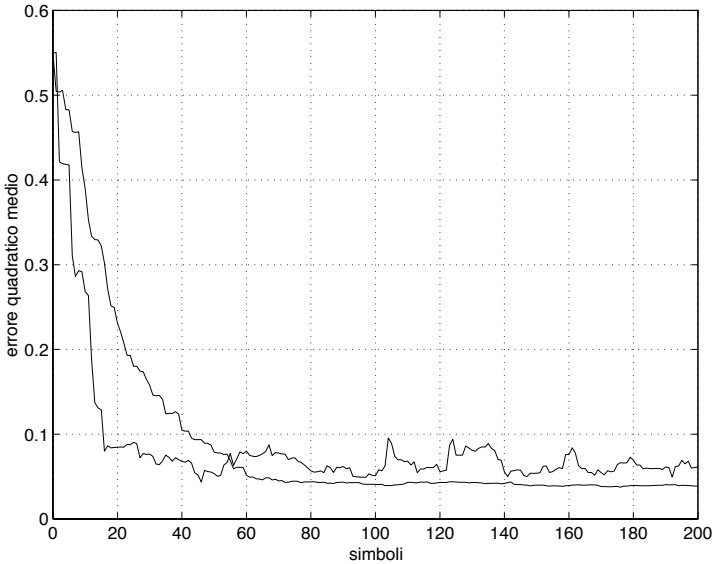

Fig. 6.3 - Esempi di convergenza di un equalizzatore a 11 prese durante la fase di apprendimento (risposta del canale: $-0.5, 1, 0.5$; $E_b/N_0 = 10$ dB; $\alpha = 1/NE[y_k^2]$ e $\alpha = 1/5NE[y_k^2]$)

(*sequenza di apprendimento*). Una volta raggiunta una configurazione accettabile delle prese, cosa solitamente molto rapida conoscendo i dati, le decisioni diventano equivalenti ad una sequenza nota. La fig. 6.3 mostra alcuni esempi di convergenza durante la fase di apprendimento.

In taluni casi non è possibile far precedere i dati da una sequenza nota. Ad esempio nei ponti radio ad alta capacità non si può proporre, ogni volta che il canale diventa così cattivo da non poter essere più equalizzato, di interrompere la trasmissione per dare inizio ad una nuova fase di apprendimento. Quando il canale torna ad essere equalizzabile, il ricevitore deve provvedere senza conoscere i dati trasmessi. Sono state proposte molte tecniche cosiddette di *equalizzazione cieca* in grado di risolvere questo problema, naturalmente non in tempi così rapidi come in fig. 6.3.

Per estendere al caso complesso l'algoritmo del gradiente stocastico occorre valutare le derivate rispetto a parte reale e immaginaria dei coefficienti c_i

dell'errore quadratico medio, dato da

$$E[(\mathrm{Re}\{z_k - d_k\})^2 + (\mathrm{Im}\{z_k - d_k\})^2] = E[|z_k - d_k|^2] = \\ = E[(z_k - d_k)(z_k^* - d_k^*)] \tag{6.18}$$

Il risultato di tale derivazione è spesso dato per scontato, ma non è così noto da non meritare una breve giustificazione. Sia $f(z)$ una funzione analitica della variabile complessa $z = u + jv$, e si vogliano calcolare le derivate, rispetto alle variabili reali u e v, della funzione $|f(z)|^2 = f(z)f^*(z)$. Questa non è una funzione analitica di z, cioè non è derivabile rispetto a z, mentre $f(z)$ lo è. Dando una variazione $dz = du + jdv$ all'argomento, la variazione della funzione è

$$d|f(z)|^2 = d(f(z)f^*(z)) = f'(z)dz\ f^*(z) + f'^*(z)dz^*\ f(z) = \\ = \mathrm{Re}\{2f(z)f'^*(z)dz^*\} = \mathrm{Re}\{2f(z)f'^*(z)\}du + \mathrm{Im}\{2f(z)f'^*(z)\}dv \tag{6.19}$$

dove forse il termine $f'^*(z)dz^*\ f(z)$ non risulterà ovvio; tuttavia basta pensare che la variazione $d|f(z)|^2$ è reale qualunque sia dz, e ciò richiede la somma di due termini complessi coniugati.

Dall'ultima espressione si vede che $2f(z)f'^*(z)$ contiene nella parte reale e immaginaria le derivate rispetto alla parte reale e immaginaria di z. Analogo calcolo si può ripetere per la funzione di N variabili complesse $|f(z_1, \ldots, z_N)|^2$. L'insieme degli N numeri complessi $2ff^*_{z_i}$ fornisce, a coppie, le $2N$ componenti del gradiente, e viene detto *gradiente complesso*.

Tornando all'equalizzatore, si trova che l'i-esima componente del gradiente complesso è data da $2E[(z_k - d_k)y^*_{k-i}]$, e quindi l'aggiornamento con il metodo del gradiente stocastico è

$$c_i^{k+1} = c_i^k - \alpha(z_k - \hat{d}_k)y^*_{k-i} \tag{6.20}$$

In analogia con quanto visto per il caso reale, il valore di α che dà la massima velocità di convergenza è prossimo a $1/NE[|y_k|^2]$.

6.6 Cancellazione d'eco

Si supponga che i campioni ricevuti y_k contengano un segnale *desiderato* indipendente dagli a_k, ma che a questo sia sovrapposta, oltre al rumore, una

replica *indesiderata* dei dati a_k, distorta da una qualche risposta impulsiva h_i *incognita*.

Tale situazione si presenta ad esempio nella trasmissione bidirezionale su linea bifilare o cavo coassiale. Anche senza prevedere per i due flussi di dati una separazione temporale, spettrale o mediante codici ortogonali (*divisione* di *tempo, frequenza* o *codice*, rispettivamente), la trasmissione bidirezionale risulta possibile se si disaccoppiano il ricevitore ed il trasmettitore locale mediante apposite strutture bilanciate. Il segnale trasmesso ha però potenza molto maggiore di quello ricevuto, e basta anche un piccolo inevitabile sbilanciamento perché al segnale utile si sovrapponga una *eco* dei dati trasmessi a_k.

Non è nota la risposta impulsiva h_i attraverso cui l'eco viene prodotta, ma sono disponibili i dati a_k. Si può sottrarre ai campioni ricevuti y_k una combinazione lineare dei dati a_{k-i}, con pesi c_i scelti in modo da avere la miglior cancellazione dei dati indesiderati (e si avrebbe una *cancellazione d'eco* di tipo ZF) oppure la minima varianza della somma di rumore e dati residui (cancellazione d'eco MMSE)[7]. Per quest'ultima si vuole ottenere il minimo di

$$E[(y_k - \sum c_i a_{k-i})^2] \tag{6.21}$$

dove, al solito, non sono indicati gli estremi della somma, che dipendono dal numero e dalla posizione temporale delle prese previste. La (6.21) è del tutto analoga alla (6.15), salvo il diverso ruolo delle variabili in gioco: i dati a_{k-i} sostituiscono i campioni ricevuti y_{k-i}, e il campione y_k da cui si vuole cancellare l'eco svolge il ruolo che nella (6.15) era del dato da decidere a_k. Il problema è però equivalente, dal punto di vista matematico. Ciò significa che l'aggiornamento dei coefficienti c_i può essere condotto in modo analogo a quello visto per l'equalizzazione, cioè secondo la (6.17), naturalmente con le opportune sostituzioni di variabili.

L'unico punto da sottolineare è che una buona cancellazione dell'eco richiede elevata precisione numerica, perché l'interferenza indesiderata da parte dei dati a_{k-i} è in genere più forte del segnale utile. Comunque, a parte queste particolarità dovute al diverso significato delle variabili in gioco, la struttura del cancellatore d'eco è concettualmente simile a quella di un equalizzatore.

[7] si osservi che poiché i dati sono indipendenti sia dal segnale ricevuto sia dal rumore la cancellazione opera solo sui dati; si noti anche che dal punto di vista del cancellatore d'eco il segnale *desiderato* è equivalente al rumore ricevuto, e disturba la convergenza dell'algoritmo di cancellazione

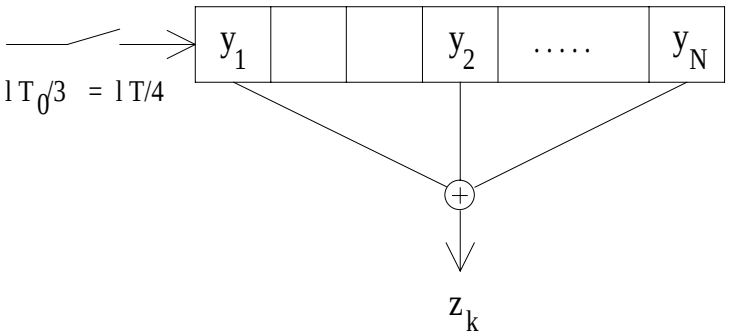

Fig. 6.4 - Equalizzatore a prese frazionarie con passo $T_0 = 3T/4$

6.7 Equalizzazione a prese frazionarie

Campionare l'uscita del filtro di ricezione al ritmo di un campione per simbolo sembra la cosa più naturale del mondo. Tuttavia lo è solo se produce una statistica sufficiente e se questa viene elaborata in modo ottimale. Poiché in tutti i casi pratici la banda del segnale numerico è maggiore di $1/2T$, il sottocampionamento con frequenza $1/T$ distrugge informazione, non più recuperabile perlomeno in modo lineare. Detto altrimenti, un equalizzatore lineare a passo T ha necessariamente una funzione di trasferimento con simmetria complessa coniugata intorno alla frequenza $1/2T$ mentre un filtro FIR *multirate* con spaziatura $T_0 < T$ all'ingresso e T all'uscita offre più gradi di libertà. Si potrà scegliere per semplicità $T_0 = T/2$, ma non di rado si preferisce $T_0 = mT/n = 2T/3$ o $3T/4$. L'equalizzatore viene detto *a prese frazionarie*.

Conviene modificare la numerazione dei campioni; chi infatti scriverebbe $y_{3l/4}$ per indicare i campioni in ingresso? Con riferimento alla fig. 6.4, i campioni sono prelevati con passo $T_0/m = T/n$ ed entrano nel registro a scorrimento n per volta ogni T secondi. Solo un campione ogni m di quelli contenuti nel registro contribuisce all'uscita dell'equalizzatore. Indicando con y_i e c_i i campioni utilizzati ed i relativi coefficienti si ha semplicemente

$$z = \sum_{i=1}^{N} c_i y_i \tag{6.22}$$

dove N è il numero dei moltiplicatori. La stessa notazione, che è svincolata

dall'istante attuale kT, potrebbe tra l'altro essere usata anche per gli equalizzatori *a prese intere*[8].

L'insieme dei c_i che minimizza l'errore quadratico medio si ottiene risolvendo il solito sistema di equazioni lineari, che il lettore può provare a scrivere, oppure nel solito modo adattativo (con qualche piccola anomalia, per cui si rimanda all'es. 6.13).

Normalmente, ma non è una regola, le prestazioni risultano migliori rispetto alla spaziatura intera, e ciò non solo a parità di durata della risposta dell'equalizzatore misurata in tempi di simbolo, il che appare ovvio, ma anche a parità di numero di coefficienti.

Un grande merito della spaziatura frazionaria è la scarsa sensibilità agli errori di temporizzazione. Il campionamento sufficientemente fitto consente un'efficace interpolazione (es. 6.14).

Infine anche l'equalizzazione MMSE a prese frazionarie può essere combinata con la tecnica DF (a prese intere, naturalmente).

6.8 Identificazione adattativa del canale

In alcuni contesti anziché determinare i coefficienti c_i dell'equalizzatore si preferisce stimare la risposta impulsiva del canale discreto. Se occorre è poi concettualmente facile calcolare la risposta dell'equalizzatore, ottimizzata secondo il criterio preferito.

In altre parole anziché individuare i coefficienti del filtro (eventualmente non lineare) che, alimentato dai campioni ricevuti y_k, riproduce al meglio i dati a_k si stima la risposta impulsiva $\hat{h}_i$ che, alimentata dai dati a_k, riproduce al meglio i campioni ricevuti y_k:

$$\hat{y}_k = \sum \hat{h}_i a_{k-i} \tag{6.23}$$

Poiché in pratica ogni risposta impulsiva ha durata finita si considera un numero finito di termini $\hat{h}_i$. Adottando ancora come criterio di bontà della stima di $\hat{y}_k$ l'errore quadratico medio, si cercano i coefficienti $\hat{h}_i$ che rendono minimo

$$E[(\hat{y}_k - y_k)^2] = E[(\sum \hat{h}_i a_{k-i} - y_k)^2] \tag{6.24}$$

[8] si può persino osservare che non vi è alcuna necessità di estrarre dal registro un campione ogni m; l'equalizzatore può avere *spaziatura variabile* (tanti altri gradi di libertà per complicare il progetto!)

Il problema è del tutto analogo a quello già visto per l'equalizzazione MMSE, a parte il diverso significato delle grandezze: l'uscita dell'equalizzatore z_k e il dato a_k sono sostituiti rispettivamente dai campioni stimati $\hat{y}_k$ e da quelli ricevuti y_k, e i campioni ricevuti y_{k-i} dai dati a_{k-i}. La ricerca del minimo errore quadratico medio con il metodo del gradiente stocastico è, analogamente a quanto già visto[9],

$$\hat{h}_i^{k+1} = \hat{h}_i^k - \alpha(\hat{y}_k - y_k)a_{k-i} = \hat{h}_i^k + \alpha(y_k - \hat{y}_k)a_{k-i} \tag{6.25}$$

Se i campioni ricevuti y_k sono dati dalla (6.6) e se, per semplicità, si suppongono i dati a_k incorrelati e il rumore incorrelato con i dati, è facile mostrare che il punto di equilibrio dell'algoritmo di stima è $\hat{h}_i = h_i$. Per quanto riguarda il valore più conveniente di α, le stesse considerazioni svolte per l'equalizzatore forniscono $1/NE[a_k^2]$ nella fase iniziale, da ridurre di circa cinque volte a regime.

In molti casi i dati trasmessi a_{k-i} non sono noti e nella (6.25) vengono sostituiti dai dati decisi $\hat{a}_{k-i}$, come nel caso dell'equalizzatore. Occorre naturalmente introdurre nello stimatore il ritardo necessario perché divengano disponibili i dati decisi *successivi* all'attuale, corrispondenti a valori negativi dell'indice i, che sono richiesti dall'algoritmo.

Nel caso complesso l'unica (ovvia) modifica nella (6.25) è che a_{k-i} è sostituito da d_{k-i}^*. Il valore di α nella fase iniziale è $1/NE[|d_k|^2]$.

6.9 Correlatore (o filtro adattato) "adattativo"

Un'applicazione interessante, anche se finora poco diffusa, della stima del canale è uno dei metodi che sono stati proposti per realizzare un correlatore (o filtro adattato) *adattativo*, per risolvere il problema a cui si è accennato della mancata conoscenza della forma d'onda elementare *ricevuta* $g(t)$. È noto che si possono calcolare le correlazioni di $r(t)$ e $g(t-kT)$ come somme di prodotti dei rispettivi campioni, purché si sia previsto un primo filtro di ricezione analogico che limiti la banda del segnale ricevuto. Questo può anche modificare, di solito in modo lieve, $g(t)$, peraltro senza conseguenze perché la stima di $g(t)$ viene effettuata a valle. Per semplicità di esposizione si supponga di operare con

[9] normalmente viene preferita l'ultima espressione, che differisce solo per un banale cambiamento di segno

due campioni per simbolo:

$$\int r(t)g(t-kT)dt = \int r(t+kT)g(t)dt \equiv$$
$$\equiv \sum r(nT/2+kT+\tau)g(nT/2+\tau) \tag{6.26}$$

Per maggior generalità si è supposto che i campioni di $r(t)$ siano prelevati con un *offset* temporale τ. Il risultato della correlazione non dipende da τ, purché si conoscano e utilizzino i campioni $g(nT/2+\tau)$ agli istanti di tempo corretti.

Basta quindi identificare i campioni della forma d'onda elementare $g(t+\tau)$, con passo $T/2$. Suddividendo questi campioni, per comodità, in *pari* e *dispari* ci si riduce a dover stimare due diverse risposte impulsive discrete, con passo pari ad un tempo di simbolo. D'altra parte si riconosce facilmente che i campioni *pari* di $r(t)$ sono dati da

$$r(kT+\tau) = \sum g(iT+\tau)a_{k-i} + n(kT+\tau) \tag{6.27}$$

e quelli *dispari* da

$$r(kT+T/2+\tau) = \sum g(iT+T/2+\tau)a_{k-i} + n(kT+T/2+\tau) \tag{6.28}$$

Quindi i campioni *pari* e *dispari* di $r(t)$, unitamente alla conoscenza dei dati a_k o delle corrispondenti decisioni $\hat{a}_k$, consentono di identificare *separatamente* le due risposte discrete[10] $g(iT+\tau)$ e $g(iT+T/2+\tau)$. Si può poi procedere al calcolo delle correlazioni (6.26) ed alle successive elaborazioni: equalizzazione, oppure stima MLSE; infatti dai campioni di $g(t)$ è possibile calcolare anche le correlazioni R_{k-n} definite dalla (6.3).

Vale la pena di sottolineare che l'errore di temporizzazione τ viene assorbito, cioè cancellato, da questa procedura: infatti nella fase di identificazione i campioni di $g(t)$ vengono automaticamente valutati con la giusta temporizzazione. Poiché si stima un numero *finito* di campioni di $g(t+\tau)$ occorre solo avere una idea grossolana del valore di τ, in modo da includere *tutti* i campioni significativi.

Nel caso complesso le varianti sono marginali: è richiesto il calcolo della correlazione tra gli equivalenti passa basso $z(t)$ e $g^*(t-kT)$, ed in generale la forma d'onda $g(t)$ è complessa. Anche un errore di fase $\exp(j\vartheta)$ nel segnale

[10] sarebbe meglio stimare le due risposte *congiuntamente*, ma prevale il desiderio di semplificare

ricevuto $z(t)$ viene automaticamente assorbito; infatti lo si ritrova nella stima di $g(t)$, e viene quindi cancellato quando nella correlazione si moltiplica per il complesso coniugato.

Con tutti questi meriti la tecnica meriterebbe maggior diffusione.

6.10 Equalizzazione adattativa ZF e DF

Con un numero N finito di prese non è possibile in generale forzare a zero l'ISI. Si indica comunque con la sigla ZF un equalizzatore che utilizza gli N gradi di libertà per imporre una risposta impulsiva pari a uno nell'origine e zero in $N-1$ posizioni prefissate. La versione adattativa è basata sulla misura della risposta impulsiva b_i a valle dell'equalizzatore. Se si suppongono i dati incorrelati, si verifica facilmente che il valor medio di $(z_k - a_k)a_{k-i}$ è proporzionale a b_i per $i \neq 0$ e a $b_0 - 1$ per $i = 0$. Se, e solo se, questi valori medi sono diversi da zero occorre intervenire. Non è semplice agire su un solo b_i perché ogni coefficiente dell'equalizzatore li modifica tutti:

$$b_i = \sum h_n c_{i-n} \tag{6.29}$$

Tuttavia se il canale non è troppo distorto il termine dominante è quello per $n = 0$, e c_i influenza b_i molto più che gli altri. Se dunque si sta misurando un valore scorretto di b_i si interviene sulla presa corrispondente c_i fino al raggiungimento dell'equilibrio[11]:

$$c_i^{k+1} = c_i^k - \alpha(z_k - \hat{a}_k)\hat{a}_{k-i} \tag{6.30}$$

Di questo algoritmo era molto apprezzato in passato il fatto che nel caso binario il prodotto per $\hat{a}_{k-i}$ richiede al più un cambiamento di segno. Ora il vantaggio è trascurabile, e sempre più spesso si preferisce l'equalizzazione MMSE.

Se si vuol utilizzare anche la tecnica DF si controllano i c_i, e quindi i b_i, mediante la (6.30) solo per $i \leq 0$. Occorre poi stimare la risposta impulsiva residua b_i a valle dell'equalizzatore (che è in parte lineare, in parte in feedback) per $i > 0$. L'aggiornamento del coefficiente A_i per cui viene moltiplicato, per poi essere sottratto, il dato $\hat{a}_{k-i}$ è

$$A_i^{k+1} = A_i^k + \alpha(z_k - \hat{a}_k)\hat{a}_{k-i} \qquad i > 0 \tag{6.31}$$

Si raggiunge l'equilibrio quando $b_i = 0$ cioè quando A_i ha il valore che cancella il contributo di a_{k-i}.

[11] quindi la risposta impulsiva b_i viene azzerata nelle $N-1$ posizioni per cui sono disponibili prese dell'equalizzatore

6.11 Probabilità d'errore in presenza di ISI

A valle dell'equalizzatore resta comunque un po' di ISI. Infatti la tecnica MMSE non si pone come obbiettivo la cancellazione completa, e quella ZF, che pure dichiara di averlo, annulla solo un numero limitato di contributi.

Occorre quindi saper calcolare la probabilità d'errore in presenza di ISI e di rumore additivo gaussiano. Considerando solo il caso di dati indipendenti, non codificati, si può semplificare la notazione indicando con a_0 il dato attuale e con $b_i a_i$ gli interferenti. La posizione temporale di questi non ha alcuna importanza, per cui si assumerà $i = 1, \ldots, N$, dove N è ora il numero di interferenti. Scalando, se occorre, il segnale e il rumore si può anche supporre $b_0 = 1$, e quindi

$$z = a_0 + \sum_{i=1}^{N} b_i a_i + n \tag{6.32}$$

dove $a_i = \pm 1, \pm 3, \ldots, \pm(M-1)$ ed n è una variabile casuale gaussiana di varianza σ^2. La probabilità $P(\sum b_i a_i + n > 1)$ che si superi la soglia di decisione è facilmente calcolabile condizionando a tutte le possibili N-ple di dati interferenti:

$$P(\sum_{i=1}^{N} b_i a_i + n > 1) = \sum_{1}^{M^N} P(a_1, \ldots, a_N) P(\sum_{i=1}^{N} b_i a_i + n > 1 / a_1, \ldots, a_N) = \tag{6.33}$$

$$= \frac{1}{M^N} \sum_{1}^{M^N} Q(\frac{1 - b_1 a_1 - \ldots - b_N a_N}{\sigma})$$

dove si sono supposti i dati equiprobabili, e le somme da 1 a M^N vanno intese estese a tutte le M^N possibili configurazioni interferenti. Se il numero di interferenti è piccolo il calcolo è rapido. Per una primissima valutazione ad alto rapporto segnale-rumore si può considerare il termine dominante, cioè quello con l'interferenza peggiore[12].

Per un grande numero di interferenti si può osservare che la *ddp* del disturbo è la convoluzione delle N degli interferenti e di quella del rumore, ed

[12] un altro metodo grossolano è valutare la varianza dell'ISI e sommarla a quella del rumore, come se la *ddp* dell'ISI fosse gaussiana (cosa piuttosto lontana dal vero, in particolare per le code)

utilizzare per il calcolo le trasformate di Fourier delle *ddp*, ovvero le funzioni caratteristiche. Detto $x = \sum b_i a_i + n$ il disturbo, si ha

$$P(x > 1) = \frac{1}{2} - \frac{1}{2}P(|x| < 1) = \frac{1}{2} - \frac{1}{2}\int_{-1}^{1} f(x)dx = \\ = \frac{1}{2} - \frac{1}{2}\int r(x)f(x)dx \tag{6.34}$$

dove la funzione rettangolare $r(x)$ vale 1 nell'intervallo $(-1, 1)$ ed è nulla altrove. Ora si può invocare il teorema di Parseval. Indicando con $M(u)$ la funzione caratteristica di x e con $M_i(u)$ quella di $b_i a_i$, si ottiene

$$P(x > 1) = \frac{1}{2} - \frac{1}{2}\frac{1}{2\pi}\int \frac{2\sin u}{u} M(u)du = \\ = \frac{1}{2} - \frac{1}{2\pi}\int \frac{\sin u}{u}\prod_{i=1}^{N} M_i(u)\exp(\frac{-u^2\sigma^2}{2})du \tag{6.35}$$

E poi facile verificare (es. 6.16) che la funzione caratteristica $M_i(u)$ è data da

$$M_i(u) = \begin{cases} \cos(b_i u) & M = 2 \\ \dfrac{\sin(Mb_i u)}{M\sin(b_i u)} & M = 4, 8, \ldots \end{cases} \tag{6.36}$$

Dal punto di vista numerico non è conveniente ottenere una probabilità piccola dalla differenza di due termini molto più grandi. Per rimediare, si immagini per un momento che non vi sia ISI ($b_i = 0$). Si ha

$$P(n > 1) = Q(\frac{1}{\sigma}) = \frac{1}{2} - \frac{1}{2\pi}\int \frac{\sin u}{u}\exp(\frac{-u^2\sigma^2}{2})du \tag{6.37}$$

che potrebbe essere un modo stravagante per calcolare la funzione Q! Sottraendo la (6.37) dalla (6.35) si cancella il termine 1/2 e si ottiene

$$P(x > 1) = Q(\frac{1}{\sigma}) + \frac{1}{2\pi}\int \frac{\sin u}{u}(1 - \prod_{i=1}^{N} M_i(u))\exp(\frac{-u^2\sigma^2}{2})du \tag{6.38}$$

dove, se fa piacere, si possono interpretare i due termini come i contributi alla probabilità d'errore del rumore e dell'ISI. L'antitrasformata della funzione

integranda è la convoluzione di tre termini: un rettangolo, la differenza tra un impulso e la *ddp* dell'ISI, e una gaussiana. L'antitrasformata ha supporto praticamente limitato, e ciò consente di discretizzare l'integrale (6.38) senza errore in base al risultato duale di quello discusso nel Cap. 1 per integrali nel dominio del tempo. Il rettangolo e l'ISI non superano rispettivamente 1 e $(M-1)\sum|b_i|$. Il rumore può essere considerato limitato a $K\sigma$, con $K = 3 \div 5$ secondo il livello di probabilità d'errore considerato. Nell'integrale discretizzato il passo deve quindi essere minore di $2\pi/(1+(M-1)\sum|b_i|+K)$ (non si dimentichi che u è una pulsazione, e non una frequenza). Infine la gaussiana tronca abbastanza rapidamente la somma (ai soli fini del calcolo il rumore è il benvenuto!) e la funzione integranda è pari, per cui spesso basta sommare ben pochi termini. È da notare anche che la complessità cresce lentamente con il numero N di interferenti.

6.12 Equalizzazione nel dominio delle frequenze

La parte lineare di un equalizzatore ZF o MMSE è un filtro, realizzabile perlomeno in linea di principio nel dominio delle frequenze. Si tratta di suddividere la sequenza di campioni ricevuti in blocchi, calcolarne la trasformata di Fourier discreta (FFT), moltiplicare per una funzione di trasferimento ed antitrasformare. Naturalmente si deve provvedere a giuntare nel modo corretto i blocchi, con le usuali e ben note tecniche di *overlap and add* oppure *overlap and save*. La funzione di trasferimento del filtro può essere aggiornata iterativamente, ad esempio con il gradiente stocastico.

Gli equalizzatori nel dominio delle frequenze non hanno avuto grande fortuna. L'elaborazione a blocchi complica la gestione dei dati e introduce un ritardo, svantaggi che non sembrano compensati da una consistente riduzione del numero di operazioni. Un filtro realizzato nel dominio delle frequenze può dare un vantaggio significativo solo nel caso di risposta impulsiva molto lunga. Solitamente gli equalizzatori per la trasmissione numerica richiedono invece poche prese (qualche decina, al massimo).

6.13 Sistemi OFDM

L'equalizzazione in frequenza è la soluzione più naturale quando si suddivide l'intera banda in un gran numero di canali adiacenti, utilizzati ciascuno per trasmettere una piccola frazione dei dati. Se la banda di ciascun canale è

molto piccola, e quindi il tempo di simbolo molto grande, si può ritenere che la funzione di trasferimento del canale, distorto, vari (lentamente) nel tempo ma sia quasi indipendente dalla frequenza all'interno del canale. Si tratta quindi di un semplice guadagno, generalmente complesso, variabile casualmente nel tempo. Una semplice tecnica di equalizzazione lineare consiste nello stimare il guadagno, indipendentemente per ciascun canale, e dividere i campioni ricevuti per il guadagno stimato.

Un'applicazione raffinata di tali concetti si può trovare nei sistemi OFDM (*Orthogonal Frequency Division Multiplexing*), che sono stati proposti e ormai standardizzati per la diffusione di segnali audio e televisivi numerici.

Si usa dire che non si possono utilizzare per la trasmissione numerica forme d'onda elementari con inviluppo rettangolare, perché ne soffre troppo la banda occupata. Ciò è vero finché si vuole imporre che i canali adiacenti siano separati spettralmente. Si ricorderà tuttavia, dal Cap. 1, che le forme d'onda $\cos 2\pi f_n t$ e $\sin 2\pi f_n t$ sono ortogonali, in un intervallo di durata T, se la separazione in frequenza Δf è pari a $1/T$, e inoltre che se il numero di tali funzioni è elevato è sostanzialmente rispettata la condizione $N = 2BT_0$, cioè non si ha alcuno spreco di banda. Nei sistemi OFDM si usano centinaia o migliaia di portanti.

Nell'intervallo $(kT, kT + T)$ l'equivalente passa basso del segnale da generare è

$$z(t) = \sum_n d_k^n g(t - kT) \exp(j2\pi f_n t) \tag{6.39}$$

dove $g(t)$ è il rettangolo di durata T e i dati d_k^n, eventualmente codificati, trasmessi con la n-esima portante nel k-esimo intervallo sono tratti da costellazioni PSK o QAM[13].

Generare un segnale OFDM sembra richiedere un grandissimo numero di operazioni. Si osservi però che gli N *campioni* di $z(t)$, presi con passo T/N all'interno del tempo di simbolo T, sono esprimibili come una trasformata discreta inversa di Fourier (IFFT) su N punti. Unica condizione perché la frequenza di campionamento N/T sia sufficiente è che il numero di portanti sia minore di N. Il modulatore OFDM è quindi il *chip* che calcola la IFFT.

Anche in ricezione, se non si considera la distorsione del canale, è facile vedere che le correlazioni sono realizzabili con una FFT.

[13] si possono utilizzare costellazioni diverse per i diversi canali; se il guadagno è molto variabile con la frequenza, ed è noto a priori, suddividere l'informazione da trasmettere in parti non uguali si accorda perfettamente con quanto osservato alla fine del Cap. 4

Volendo invece compensare la distorsione lineare del canale si potrebbe pensare di far *precedere* la FFT da un filtro inverso (equalizzatore ZF), realizzato nel dominio del tempo. Risulta normalmente più comodo far *seguire* la FFT da un equalizzatore nel dominio delle frequenze. Si osservi però che non c'è equivalenza perfetta tra le due strutture, perché la FFT realizza solo convoluzioni circolari. Un modo artificiale per ottenere l'equivalenza è fare in modo che anche il canale esegua una convoluzione circolare. Se T_0 è la durata massima possibile per la risposta impulsiva del canale, si premette a ciascun simbolo un *tempo di guardia* pari a T_0 durante il quale si trasmettono gli *ultimi* campioni del blocco. Questi sono quindi trasmessi due volte, con un incremento di energia, peraltro modesto se $T_0 \ll T$. Rispetto a quanto detto in precedenza occorre quindi un piccolo ritocco: il blocco su cui si eseguono le FFT ha durata $T + T_0$ anziché T. È anche possibile combinare le due tecniche: una equalizzazione lineare *incompleta* che riduce la durata della risposta del canale, e quindi il tempo di guardia richiesto, e una equalizzazione a valle della FFT.

La tecnica OFDM non è priva di inconvenienti. Il segnale trasmesso, somma di un gran numero di portanti modulate, ha componenti in fase e quadratura praticamente gaussiane ed ha quindi un fattore di picco piuttosto elevato. Inoltre il tempo di simbolo molto grande rende critico il problema della accuratezza della frequenza delle portanti ai fini della sincronizzazione di fase. Ciò che conta infatti è il prodotto tra errore di frequenza e tempo di simbolo, come si vedrà nel Cap. 9.

6.14 Considerazioni finali

Benché la teoria proponga per il ricevitore ottimo la tecnica MLSE, di solito si adottano altre soluzioni sia per motivi di complessità sia perché la forma d'onda elementare $g(t)$ non è nota o varia nel tempo. La soluzione, teoricamente apprezzabile, di identificare adattativamente il canale in modo da poter applicare la tecnica MLSE anche a canali varianti nel tempo non è molto diffusa.

Fra le soluzioni "semplici" è molto comune l'equalizzazione MMSE, eventualmente a prese frazionarie, che può essere combinata con la tecnica DF, sempre che il ritardo di elaborazione lo consenta e non si temano le conseguenze della propagazione degli errori. Il controllo della presa centrale non di rado è ZF (es. 6.7).

Si noti che si potrebbe avere interferenza intersimbolica persino in assenza di distorsione del canale. Un modo semplice, e al quale solitamente non si pensa, per cadere vittime dell'ISI è sbagliare la temporizzazione in ricezione, cioè correlare con le funzioni $g(t - kT - \tau)$, che non sono le funzioni base; le cure per l'ISI presentate in questo capitolo alleviano anche questo problema.

Tutte le tecniche di equalizzazione sono procedure *ad hoc* che non realizzano il ricevitore a massima verosimiglianza. Le prestazioni non possono quindi essere ricavate direttamente dalla geometria dei segnali, ma vanno valutate caso per caso tenendo conto anche dell'ISI residua.

La trasmissione in banda passante non differisce sostanzialmente da quella in banda base, e le tecniche ZF, MMSE e DF di equalizzazione sono analoghe. Per valutare le prestazioni dei vari tipi di equalizzatori è comodo, anche se non necessario, rappresentare il segnale numerico e il rumore con equivalenti passa basso, con i metodi descritti nel Cap. 1.

Per concludere si può fare un breve cenno al caso codificato, ad esempio con un codice convoluzionale. Non è difficile intuire che poiché sia il codificatore sia il canale discreto con ISI sono descritti dall'evoluzione di registri a scorrimento, cioè da macchine a stati finiti, i corrispondenti tralicci si possono combinare in uno solo (ma con un numero di stati che è il *prodotto* dei due componenti, e quindi spesso intrattabile). Sono state quindi studiate soluzioni che accorpano gli stati in un insieme ridotto di *superstati*, gruppi di stati scelti in modo tale da non ridurre troppo la distanza minima tra sequenze. La tecnica, indicata con la sigla RSMLSE (*Reduced State* MLSE), resta comunque piuttosto complessa. Molto spesso quindi si preferisce un normale equalizzatore *lineare*, dal momento che le decisioni vengono prese con troppo ritardo per poter essere retroazionate, che viene progettato indipendentemente dal codice.

6.15 Esercizi

6.1 - Si consideri (e critichi) il seguente progetto per un canale lineare fortemente distorto, come ad esempio una linea bifilare o un cavo coassiale (mezzi trasmissivi che hanno un'attenuazione fortemente variabile con la frequenza): fissata la segnalazione (ad esempio binaria antipodale) si fissa anche la forma d'onda elementare a valle del filtro di ricezione (ad esempio una forma d'onda di Nyquist con *roll-off* prefissato, per non avere ISI); poi si scelgono la forma d'onda trasmessa ed il filtro di ricezione in modo da avere la minima energia trasmessa a parità di varianza del rumore all'ingresso

del decisore (o, che è lo stesso, la minima varianza del rumore a parità di energia trasmessa). Si mostri, con un semplice calcolo variazionale, che il quadrato del modulo della trasformata di Fourier della forma d'onda trasmessa è proporzionale all'inverso del guadagno del canale (in ampiezza). Per un canale fortemente distorto si mostri poi che questa soluzione è lontanissima da quanto suggerito dalla teoria della capacità di canale (si ricordi il cosiddetto *water pouring*), e quindi insoddisfacente. *Commento*: la morale da trarre è che non è corretto cancellare l'ISI con mezzi elementari.

6.2 - Nel caso del canale con risposta $1+Az^{-1}+Az = \frac{1}{1+B^2}(1+Bz^{-1})(1+Bz)$ si verifichi il valore di B dato dalla (6.13) del testo, e si calcolino la risposta dell'equalizzatore ZF e la varianza del rumore supponendo indipendenti i campioni del rumore all'ingresso dell'equalizzatore. Si verifichi che la condizione $|B| < 1$ si traduce in $|A| < 0.5$.

6.3 - Per lo stesso canale dell'esercizio precedente si consideri l'equalizzatore lineare anticausale $(1 + B^2)/(1 + Bz)$ seguito da un equalizzatore DF che cancella l'ISI residua, causale. Si calcoli la varianza del rumore, e si mostri (ignorando l'effetto della propagazione degli errori) che è inferiore a quella del solo equalizzatore lineare ZF dell'esercizio precedente.

6.4 - Nel caso del canale con risposta $1+Az^{-1}-Az = \frac{1}{1-B^2}(1+Bz^{-1})(1-Bz)$ si verifichi il valore di B dato dalla (6.14) del testo, e si calcolino la risposta dell'equalizzatore ZF e la varianza del rumore supponendo indipendenti i campioni del rumore all'ingresso dell'equalizzatore. Si verifichi che la condizione $|B| < 1$ è sempre soddisfatta.

6.5 - Per lo stesso canale dell'esercizio precedente si consideri l'equalizzatore lineare anticausale $(1 - B^2)/(1 - Bz)$ seguito da un equalizzatore DF che cancella l'ISI residua, causale. Si calcoli la varianza del rumore, e si mostri (ignorando l'effetto della propagazione degli errori) che è inferiore a quella del solo equalizzatore lineare ZF dell'esercizio precedente.

6.6 - Si riconsiderino gli esempi della Sez. 3 con risposta del canale complessa $1 + Az^{-1}$, $1 + Az$, $1 + Az^{-1} + A^*z$ e $1 + Az^{-1} - A^*z$, rispettivamente.

6.7 - Un controllo automatico di guadagno per un canale non distorto può

essere considerato un equalizzatore con *una sola presa*. Si mostri che se l'algoritmo di controllo è ZF il guadagno complessivo a regime è pari ad uno, mentre se il controllo è MMSE il guadagno è lievemente inferiore a causa del rumore additivo. Nel caso MMSE si mostri poi che la varianza della somma della distorsione, dovuta al guadagno non unitario, e del rumore è effettivamente inferiore alla varianza del solo rumore nel caso ZF. Questa riduzione della varianza appare però piuttosto artificiale, e di norma si preferisce il controllo ZF. Per lo stesso motivo, non di rado si preferisce controllare la presa *centrale* di un equalizzatore MMSE mediante l'algoritmo ZF. Si scrivano le equazioni per il controllo adattativo degli N coefficienti di questo equalizzatore.

6.8 - Imponendo che siano nulle le derivate, rispetto ai coefficienti c_i, dell'errore quadratico medio dato dalla (6.15) si mostri che i valori degli N coefficienti per il minimo errore quadratico medio si ottengono risolvendo un sistema di equazioni *lineari*, la cui matrice dei coefficienti contiene le covarianze degli N campioni contenuti nel registro a scorrimento, e il vettore dei termini noti le correlazioni tra gli stessi e il dato desiderato all'uscita dell'equalizzatore (in un canale non distorto quello corrispondente alla *presa centrale*). Si mostri che la matrice dei coefficienti e il vettore dei termini noti sono facilmente calcolabili se è nota la forma d'onda $g(t)$ a valle del filtro di ricezione e la (eventuale) correlazione tra i campioni del rumore e tra i dati. Infine si indichi come si può determinare il valore minimo dell'errore quadratico medio.

6.9 - Si dimostri la seguente affermazione, a prima vista paradossale: in assenza di rumore, un equalizzatore MMSE di lunghezza finita non può lasciare interferenza intersimbolica (ISI) con varianza maggiore di quella del corrispondente equalizzatore ZF. Come è possibile che l'equalizzatore *che forza a zero l'ISI* ne lasci più dell'altro?

6.10 - Si consideri la trasmissione binaria su un canale con ISI che dà, all'uscita del filtro adattato,

$$y_k = a_k + Aa_{k-1} + n_k$$

dove i campioni del rumore sono incorrelati e hanno varianza σ^2. Si consideri un semplicissimo equalizzatore lineare, con *presa centrale* fissata ad uno e con

un solo grado di libertà

$$z_k = y_k + B y_{k-1}$$

Si calcoli il valore di B corrispondente alla equalizzazione ZF e MMSE, rispettivamente. Si valuti poi l'errore quadratico medio residuo e si mostri che esso è effettivamente minore nel secondo caso.

6.11 - Per un equalizzatore che combina le tecniche MMSE e DF si scriva il sistema di equazioni che fornisce i valori dei coefficienti sia della parte lineare sia di quella in retroazione.

6.12 - In un equalizzatore MMSE siano state predisposte prese sovrabbondanti, e quindi *inutili*. Anche ignorando la maggior complessità, si mostri che peggiorano le prestazioni dinamiche dell'equalizzatore, cioè la velocità di convergenza. *Suggerimento*: si consideri l'effetto del numero di prese N sul passo α di aggiornamento.

6.13 - Si consideri un sistema di trasmissione PAM con *roll-off* del 40%. Si mostri che per un equalizzatore MMSE a prese frazionarie con spaziatura $T/2$ e di lunghezza infinita, o comunque sovrabbondante, *in assenza di rumore* esistono *infinite* configurazioni del coefficienti che danno lo stesso minimo errore quadratico medio. *Suggerimento*: basta mostrare che la funzione di trasferimento è arbitraria in una qualche banda. *Commento*: in pratica, anche in presenza di rumore il minimo è molto piatto (cioè debole), e quindi la convergenza dell'equalizzatore può essere piuttosto lenta.

6.14 - Si consideri una trasmissione PAM su canale non distorto. Si mostri che, almeno in teoria, un equalizzatore a prese frazionarie sufficientemente lungo è in grado di compensare un errore di temporizzazione pari anche a mezzo tempo di simbolo, interpolando i campioni ricevuti. Si spieghi perché invece un equalizzatore a prese intere non può ottenere lo stesso risultato.

6.15 - Un tipo particolarissimo di distorsione è un errore di fase nel recupero della portante, per cui l'equivalente passa basso $z(t)$ viene moltiplicato per $\exp(j\vartheta)$. Un equalizzatore potrebbe compensare la distorsione modificando i valori di *tutti* i coefficienti, da c_i a $c_i \exp(-j\vartheta)$. L'aggiornamento di N coefficienti complessi (ovvero $2N$ reali) richiede però tempo, ed è inefficace

in caso di fluttuazioni rapide di ϑ. Normalmente si preferisce stimare ϑ e moltiplicare l'ingresso y_k dell'equalizzatore, o l'uscita z_k, per $\exp(-j\vartheta)$ (si veda il capitolo sulla sincronizzazione). Quale indizio può mostrare che deve essere più agevole stimare ed aggiornare un *unico* parametro ϑ piuttosto che $2N$?

6.16 - Si verifichi la (6.36), cioè la funzione caratteristica di una variabile casuale che assume con pari probabilità gli M livelli $-(M-1)b_i \ldots (M-1)b_i$. *Suggerimento*: il caso $M = 2$ è banale; per $M > 2$ si tratta di valutare

$$\frac{1}{M} \sum_{n=0}^{M-1} \exp(jb_i u(2n - M + 1))$$

per cui si può utilizzare la ben nota relazione

$$\sum_{n=0}^{M-1} a^n = \frac{1 - a^M}{1 - a}$$

ponendo $a = \exp(j2b_i u)$.

6.17 - Si verifichi (numericamente) che la formula (6.37) valuta correttamente la funzione $Q(\cdot)$. *Commento*: naturalmente tale formula non ha alcun interesse pratico.

6.18 - Si consideri la trasmissione binaria antipodale su un canale con ISI. Si valutino numericamente e si confrontino, per varie risposte impulsive, scelte a piacere, la probabilità d'errore esatta e quelle approssimate ottenute considerando il caso peggiore oppure assimilando l'ISI a rumore gaussiano.

6.19 - La forma d'onda a coseno rialzato

$$g(t) * g(t) = \frac{\sin(\pi t/T)}{\pi t/T} \frac{\cos(\alpha \pi t/T)}{1 - 4\alpha^2 t^2/T^2}$$

con *roll-off* $\alpha = 0.4$ ha negli istanti T, $2T$ e $3T$ pendenza $-0.86/T$, $0.26/T$ e $-0.06/T$. In presenza di rumore gaussiano bianco si trasmette il segnale $\sum a_k g(t - kT)$. In ricezione si ha un piccolo errore di temporizzazione τ.
Ai fini del calcolo della probabilità d'errore l'interferenza intersimbolica (ISI), purché piccola, può essere sommata in potenza al rumore gaussiano. Quale

può essere la varianza dell'ISI, rispetto a quella del rumore, se si accetta una degradazione di 0.2 dB? Quale errore di temporizzazione si può tollerare nel caso di trasmissione binaria antipodale con $E_b/N_0 = 10$ dB? *Suggerimento*: conviene porre $E_s = E_b = 1$, e quindi $N_0/2 = \ldots$
Si calcoli, a parità di degradazione e di probabilità d'errore, quale errore di temporizzazione si può tollerare nei due casi di trasmissione a quattro livelli ($E_b/N_0 \approx \ldots$), e binaria antipodale con il codice convoluzionale a 64 stati e *rate* $R = 1/2$, che a queste probabilità d'errore guadagna circa 5.5 dB.

6.20 - Si consideri una modulazione 8PSK con ISI dovuta ad un piccolo errore di temporizzazione, come nell'esercizio precedente. Con le stesse approssimazioni, quale può essere la varianza dell'ISI rispetto a quella del rumore se si accetta una degradazione di 0.2 dB? Quale errore di temporizzazione si può tollerare nel caso di trasmissione con $E_b/N_0 = 14$ dB?

6.21 - Si consideri un canale distorto che a valle del filtro adattato produce un canale discreto con risposta $1 + Az^{-1}$, e campioni di rumore incorrelati. La segnalazione sia binaria antipodale. Per quanto possibile in modo quantitativo (numericamente, se occorre) si confrontino le prestazioni di: ricevitore senza alcuna equalizzazione; con equalizzatore ZF; con equalizzatore MMSE; con equalizzatore DF; ricevitore MLSE. *Commento*: nell'ultimo caso si troverà il risultato, a prima vista sorprendente, che le prestazioni sono *migliori* di quelle del canale con risposta 1, cioè senza ISI; ma osservando che il termine Az^{-1} porta comunque energia, non è del tutto strano che lo si possa sfruttare a dovere.

6.22 - Nell'esercizio precedente cosa cambierebbe se la risposta discreta fosse $1 + Az$ (anticausale)?

Capitolo 7

Modulazione numerica di frequenza a fase continua

7.1 Introduzione

La modulazione di frequenza, o fase, ha l'importante caratteristica di produrre un segnale a radiofrequenza con inviluppo costante. Lo stadio finale del trasmettitore può lavorare in saturazione fornendo una potenza maggiore rispetto ad un amplificatore lineare, a parità di dispositivo e di consumo. Ciò è particolarmente interessante sia nel sistema radiomobile, in cui i terminali mobili sono alimentati da batterie, sia alle frequenze elevatissime, dove non c'è abbondanza di dispositivi lineari a basso costo.

Nel valutare le prestazioni di questi sistemi non si guardi dunque solo al valore di E_b/N_0 richiesto per una prefissata probabilità d'errore, che potrà sembrare buono ma non entusiasmante.

La modulazione di frequenza perde molto del suo valore se non si impone la continuità di fase del segnale trasmesso. Infatti lo spettro sarebbe insoddisfacente, e del tutto inadeguato ad esempio per il sistema radiomobile. Il vincolo della fase continua introduce una correlazione tra i simboli adiacenti, interpretabile come una forma di codifica. Ciò complicherà il ricevitore (ma più nessuno si spaventa davanti ad un codice); in compenso offrirà prestazioni migliorate anche dal punto di vista dell'efficenza nell'uso della potenza. Per dirla in breve oggi come oggi, e a maggior ragione domani, la modulazione di frequenza a fase non continua non merita di essere considerata. Per indicare la modulazione di frequenza a fase continua è usatissima la sigla CPM

(*Continuous Phase Modulation*).

Ai fini dell'analisi delle prestazioni (struttura del ricevitore, probabilità d'errore, spettri, ecc.) è meglio dimenticare la modulazione angolare analogica. Si vedrà che, contrariamente alle attese, non c'è alcuna convenienza ad aumentare la deviazione al di sopra di valori relativamente piccoli. Quindi il calcolo degli spettri e della banda occupata non ha nulla a che fare con formule come quella di Carson. Inoltre la struttura del ricevitore sarà derivata sulla base della teoria della stima a massima verosimiglianza di sequenze (MLSE).

Anche se raramente viene realizzato in questo modo, il modulatore CPM può in linea di principio essere visto come un modulatore di frequenza, cioè un VCO, al cui ingresso sia applicato un segnale numerico $\sum a_k g(t-kT)$, con dati a_k binari o multilivello. Il legame tra i dati a_k e la frequenza istantanea, e quindi la fase istantanea, è lineare. Ad esempio per la fase si ha

$$\varphi(t) = 2\pi h \int_{-\infty}^{t} \sum a_k g(t-kT) = 2\pi h \sum a_k q(t-kT) \tag{7.1}$$

È ormai consolidata la consuetudine di normalizzare i valori di a_k a $\pm 1, \pm 3, \ldots$ e l'area della forma d'onda $g(t)$ a $1/2$, cioè porre[1] $q(\infty) = 1/2$. Il contributo finale di ciascun simbolo alla deviazione di fase è $\pi h a_k$, ed h viene detto *indice di modulazione*. Il segnale trasmesso è

$$s(t) = A\cos(2\pi f_0 t + \varphi(t) + \vartheta) = \mathrm{Re}\{A\exp(j\varphi(t) + j\vartheta)\exp(j2\pi f_0 t)\} \tag{7.2}$$

La fase iniziale della portante ϑ, da recuperare con un apposito sincronizzatore, verrà sottintesa nel seguito.

Se la durata della forma d'onda modulante in frequenza $g(t)$, e quindi il tempo di salita da 0 a $1/2$ della forma d'onda di fase $q(t)$, non supera un tempo di simbolo T la modulazione CPM viene detta *a risposta totale*; altrimenti *a risposta parziale*. La maggior parte dei sistemi di interesse pratico è a risposta parziale. Tuttavia conviene partire dai casi più semplici, a risposta totale. Per questi si vede, esaminando la (7.1), che la sequenza dei simboli precedenti determina solo la fase iniziale, e che a questa si somma la transizione di fase dovuta al simbolo attuale, che a sua volta si completa entro il tempo di simbolo.

Un caso veramente semplice è quello in cui i dati a_k sono binari, l'impulso modulante è un rettangolo di durata T e l'indice di modulazione è 2 (o un multiplo comunque grande di 2)[2]. Infatti la fase finale di un simbolo, che

[1] anche se sembrerebbe molto più naturale porre $q(\infty) = 1$ oppure π o 2π

[2] un altro caso semplice, ma non interessante, è quello in cui l'impulso modulante ha area nulla (es. 7.1)

è poi la fase iniziale per il successivo, risulta multipla di 2π, cioè nulla, indipendentemente dai simboli trasmessi. Non c'è quindi correlazione tra i simboli successivi e le decisioni possono essere prese indipendentemente simbolo per simbolo. Inoltre è immediato verificare che i due possibili segnali sono ortogonali, e quindi la probabilità d'errore è

$$P(E) = Q(\sqrt{\frac{E_b}{N_0}}) \tag{7.3}$$

A nulla vale aumentare la deviazione se non a peggiorare lo spettro, allargandolo.

Se si dimezza la deviazione ($h = 1$) la fase finale è π (oppure $-\pi$, che è lo stesso) alla fine del primo simbolo, 0 alla fine del secondo, di nuovo π dopo il terzo, e così via. La fase iniziale è comunque sempre nota, indipendentemente dalla sequenza trasmessa, e ciò è quanto basta per prendere ancora decisioni indipendenti. È immediato verificare che i due possibili segnali in ogni intervallo T sono ancora ortogonali, e quindi non è cambiato nulla (a parte lo spettro, che sarà migliorato).

7.2 La modulazione MSK

Se si dimezza ancora la deviazione, la fase alla fine del primo intervallo è $\pi/2$ oppure $-\pi/2$, e quindi la fase iniziale del secondo intervallo dipende dal primo simbolo trasmesso. Alla fine del secondo intervallo la fase potrà essere π, 0 o $-\pi$ (che equivale a π), e così via come nella fig. 7.1 che mostra il *diagramma ad albero* delle fasi.

Se però si immagina la fig. 7.1 avvolta su un cilindro, in modo da far coincidere i multipli di 2π, si vede che l'albero è in realtà un traliccio a due stati[3].

Poiché l'energia non dipende dalla sequenza dei dati, il ricevitore MLSE cercherà la sequenza di dati binari a_k che dà la massima correlazione tra il corrispondente segnale modulato e la forma d'onda ricevuta $r(t)$. La

[3] se si è disturbati dal fatto che le fasi possibili sono alternativamente $\pm\pi/2$ e 0 o π si considerino 0 e 2 anziché -1 e 1 come livelli possibili per a_k, cosa che equivale a prendere come frequenza di riferimento $f_0 - 1/4T$ anziché f_0, e si vedrà che le fasi possibili sono 0 e π per tutti gli intervalli

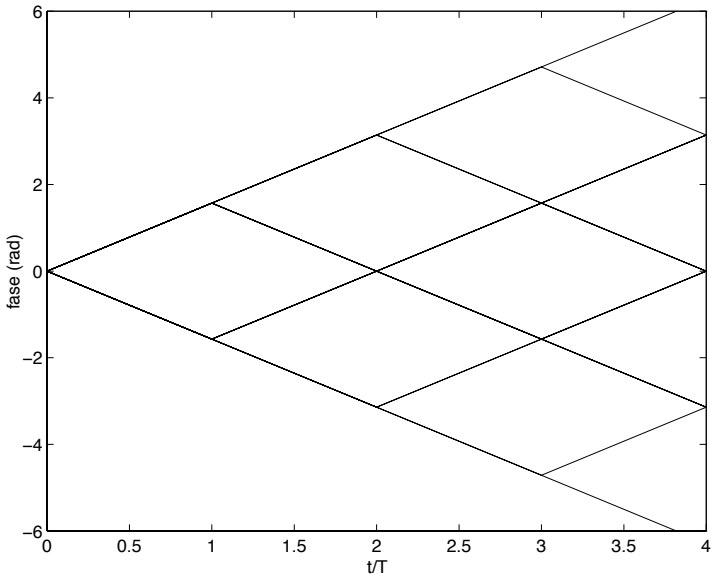

Fig. 7.1 - Albero delle fasi per la modulazione MSK

correlazione è calcolata evidentemente in banda base:

$$\begin{aligned} &\mathrm{Re}\{\int z(t)\exp(-j\pi\sum a_k q(t-kT)dt\} = \\ &= \sum_k \mathrm{Re}\{\exp(-j\frac{\pi}{2}\sum_{n=-\infty}^{k-1} a_n)\int_{kT}^{kT+T} z(t)\exp(-j\frac{\pi}{2}a_k(t-kT)/T)dt\} \end{aligned} \tag{7.4}$$

dove il termine $\exp\big(-j\frac{\pi}{2}\sum_{n=-\infty}^{k-1} a_n\big)$ ha come valori possibili alternativamente ± 1 e $\pm j$. Si vede dalla (7.4) che occorre calcolare due correlazioni per ogni simbolo, con $a_k = \pm 1$, e che queste vanno combinate con le precedenti tenendo conto dei due possibili valori della fase accumulata. È immediato vedere che ciò corrisponde alle quattro transizioni di stato possibili nel traliccio a due stati, e che quindi si potrà usare l'algoritmo di Viterbi per sfoltire via via le sequenze candidate.

Le prestazioni dipenderanno, come al solito, dalle distanze tra segnali concorrenti (non le distanze tra le fasi, ma tra i segnali modulati!). In particolare, per alto rapporto segnale-rumore basta considerare la distanza minima.

Per il calcolo della distanza in un intervallo di tempo NT tra due segnali CPM con fasi rispettivamente $\varphi_1(t)$ e $\varphi_2(t)$ si ottiene facilmente

$$\begin{aligned} d^2 &= |\mathbf{s}_1|^2 + |\mathbf{s}_2|^2 - 2\mathbf{s}_1 \cdot \mathbf{s}_2 = \\ &= NA^2T - 2A^2 \int_0^{NT} \cos(2\pi f_0 t + \varphi_1(t)) \cos(2\pi f_0 t + \varphi_2(t)) dt = \\ &= A^2 \int_0^{NT} (1 - \cos \Delta\varphi(t)) dt = \frac{2E_s}{T} \int_0^{NT} (1 - \cos \Delta\varphi(t)) dt \end{aligned} \tag{7.5}$$

dove $E_s = E_b \log_2 M$ è l'energia per simbolo e $\Delta\varphi(t) = \varphi_1(t) - \varphi_2(t)$. Nel caso in esame è semplice vedere che le sequenze a distanza minima differiscono in due bit consecutivi: $1, -1$ e $-1, 1$ oppure $1, 1$ e $-1, -1$. Nel secondo caso il ricongiungimento si ha con una differenza di fase (solo apparente!) di 2π. La distanza tra i segnali concorrenti è facilmente calcolabile, e risulta pari a

$$d^2 = 4E_s = 4E_b \tag{7.6}$$

e quindi la probabilità dell'evento errore[4] è

$$P(E) = Q\left(\sqrt{\frac{2E_b}{N_0}}\right) \tag{7.7}$$

È consuetudine ormai diffusa, e sarà quindi adottata nel seguito, quella di normalizzare la distanza minima d^2 dividendo per $2E_b$ in modo da ottenere direttamente il coefficiente di E_b/N_0 nell'argomento della funzione $Q(\cdot)$, che è pari a 2 nel caso in esame.

Come si vede, ponendo $h = 1/2$ si è dimezzata la deviazione, e presumibilmente migliorato lo spettro, e si sono pure guadagnati 3 dB. Intuitivamente ciò è dovuto al fatto che i due concorrenti sono ancora ortogonali, ma su un intervallo di tempo $2T$ doppio, quanto dura l'evento errore. Se invece si raddoppia la deviazione, tornando ad $h = 1$, si ha ricongiungimento delle traiettorie di fase nell'unico stato dopo un solo tempo di simbolo.

[4] per ogni evento errore si sbagliano due bit d'informazione a meno che si voglia prevedere una corrispondenza assoluta tra dati e fasi del segnale modulato (si veda l'es. 7.6)

Il sistema di modulazione appena descritto è famosissimo, anche se raramente usato, ed è indicato con la sigla MSK (*Minimum Shift Keying*) o FFSK (*Fast Frequency Shift Keying*). Una parte della fama deriva dal fatto che è interpretabile non solo come modulazione di frequenza, ma anche come modulazione d'ampiezza. Con un po' di pazienza (es. 7.4) si verifica che le componenti in fase e quadratura sono formate da due flussi di semionde di sinusoide di durata $2T$ modulate da dati binari legati da una semplice espressione ai dati a_k, con un tempo di simbolo $2T$ e con un *offset* relativo di mezzo tempo di simbolo (cioè T).

Il ricevitore dunque non richiede, come sembrava, l'algoritmo di Viterbi: poiché su ciascun asse le forme d'onda elementari hanno durata pari al tempo di simbolo ($2T$), bastano due filtri adattati in fase e quadratura, e due decisori a soglia.

Infine, trattandosi di modulazione d'ampiezza binaria antipodale il risultato (7.7) relativo alla probabilità d'errore non sorprende. Anche il calcolo dello spettro è banale; poiché l'impulso elementare è contenuto entro un tempo di simbolo ($2T$) il risultato è però poco più che discreto.

7.3 Altri schemi CPM a risposta totale

Di uno schema CPM a risposta totale si possono cambiare l'indice di modulazione, il numero dei livelli e la forma d'onda modulante.

Per quanto riguarda l'indice di modulazione si provi ad esempio a tracciare il traliccio delle fasi con $h = 1/3$ o $h = 2/3$, e si vedrà che il traliccio ha tre stati. Più in generale, il numero degli stati è pari al denominatore p dell'indice di modulazione espresso come rapporto di numeri primi: $h = k/p$. Un indice di modulazione irrazionale non corrisponde ad un numero finito di stati, e quindi è da evitare per poter realizzare il ricevitore[5].

Si può mostrare che per $h < 1$ il quadrato della distanza minima, normalizzata, per la modulazione binaria con impulso rettangolare è dato da (es. 7.7)

$$d^2 = 2(1 - \frac{\sin 2\pi h}{2\pi h}) \tag{7.8}$$

Non c'è quindi molto da guadagnare aumentando h. Il massimo è intorno ad $h = 0.7$, ma il miglioramento non sembra compensare il considerevole

[5] una conseguenza di ciò è che la deviazione di fase in trasmissione deve essere controllata con precisione

allargamento dello spettro. Si ricorderà poi la pessima prestazione dell'indice di modulazione $h = 1$. Indici che permettono eventi errore di durata pari a un simbolo sono detti *deboli*; spesso vanno evitati, o almeno considerati con attenzione. Si noti che $h = 1$ è un indice debole per qualunque modulazione CPM.

Il numero dei livelli non influenza il numero degli stati, ma solo il numero di correlazioni da calcolare per ciascun tempo di simbolo. Per $M = 4$ e impulso rettangolare, già $h = 1/3$ è un indice debole (es. 7.7). Non conviene superare valori di h intorno a 0.3. Per tale valore le prestazioni risultano assai simili all'MSK ($M = 2$, $h = 0.5$), ma lo spettro è un po' migliore.

Infine l'indagine sulla forma d'onda modulante porterebbe a vedere che essa ha un effetto tutto sommato modesto. Variazioni considerevoli si hanno solo rinunciando alla risposta totale.

7.4 Modulazioni CPM a risposta parziale

Se la forma d'onda modulante in frequenza $g(t)$ ha durata maggiore di un tempo di simbolo, ad esempio LT, cioè se il contributo πh di ciascun simbolo alla fase si distribuisce su diversi simboli, il traliccio delle fasi assume un aspetto assai più vario. Infatti, supponendo $g(t) \neq 0$ nell'intervallo $(0, LT)$, si possono individuare nella (7.1), valutata nell'intervallo $(kT, kT + T)$, tre termini:

$$\varphi(t) = \pi h \sum_{n=-\infty}^{k-L} a_n + 2\pi h \sum_{n=k-L+1}^{k-1} a_n q(t - nT) + 2\pi h a_k q(t - kT) \qquad (7.9)$$

Il primo rappresenta il contributo dei simboli precedenti $\ldots, a_{k-L-1}, a_{k-L}$ per i quali la forma d'onda $q(t - kT)$ ha già raggiunto il valore finale $1/2$; tale termine è rappresentabile con uno *stato di fase*. Il secondo termine, dovuto ai simboli $a_{k-L+1}, \ldots, a_{k-1}$, è esprimibile come una funzione del tempo t e dello *stato* di un registro a scorrimento contenente gli $L - 1$ simboli precedenti all'attuale. L'ultimo contributo viene dal simbolo attuale a_k.

Fissato un istante di tempo $kT \leq t \leq kT + T$, la fase istantanea e quindi il segnale modulato sono determinati univocamente dallo stato di fase, dallo stato del registro a scorrimento e dal simbolo attuale. Il simbolo attuale e gli stati del registro a scorrimento evolvono nel tempo con transizioni del tutto analoghe a quelle di un codice convoluzionale con registro di lunghezza L. In

più c'è lo stato di fase, il cui aggiornamento segue una semplice legge iterativa; infatti

$$\sum_{n=-\infty}^{k-L} a_n = \sum_{n=-\infty}^{k-L-1} a_n + a_{k-L} \tag{7.10}$$

Il modulatore CPM genererà un numero prefissato di campioni per tempo di simbolo, generalmente piccolo perché il segnale modulato ha banda stretta, inviati ad un convertitore digitale-analogico e ad un filtro interpolante. Il modulatore può essere costituito da una ROM in cui sono memorizzati i campioni delle traiettorie di fase, alimentata dal contenuto (a M livelli) del registro a scorrimento e dallo stato di fase (a p livelli), e naturalmente dall'informazione sulla posizione temporale del campione da generare.

Il numero complessivo di stati, per modulazione a M livelli, è pari al denominatore p dell'indice di modulazione, che determina il numero di stati di fase, moltiplicato per M^{L-1}, che è il numero di stati del registro a scorrimento. Poiché poi sono possibili M valori del simbolo attuale a_k il numero di traiettorie di fase possibili in un tempo di simbolo è pari a pM^L. Si possono prevedere in ricezione altrettanti correlatori. Se p è grande si risparmia qualcosa prevedendone $2M^L$, cioè uno in fase ed uno in quadratura per ogni configurazione dei dati nel registro a scorrimento. Tenendo conto che i sistemi più efficienti si ottengono con $M = 4$ e $L = 3$ o 4, risulta evidente la complessità del ricevitore. La ricerca quindi di ricevitori semplificati, non ottimi, a cui si accennerà in seguito è ben giustificata.

La fig. 7.2 mostra le traiettorie di fase per una modulazione binaria con $h = 1/2$ e forma d'onda modulante in frequenza a coseno rialzato (con *roll-off* del 100%) di durata $3T$, che viene indicata con la sigla 3RC.

Nei sistemi a risposta parziale l'enumerazione degli eventi errore ed il calcolo delle distanze, per una valutazione approssimata della probabilità d'errore tramite lo *union bound*, e in particolare il calcolo della distanza minima diventano più elaborati, non essendo facile vedere a colpo d'occhio le distanze tra segnali. Tuttavia come indicazione grossolana si può dire che nei casi di interesse pratico, almeno per valori di h fino a 0.5 la distanza minima corrisponde a sequenze che differiscono di 2 e -2 rispettivamente, in due simboli consecutivi. La distanza normalizzata risulta quindi

$$d^2 = \log_2 M \; \frac{1}{T} \int_0^{(L+1)T} \big(1 - \cos 4\pi h(q(t) - q(t-T))\big) dt \tag{7.11}$$

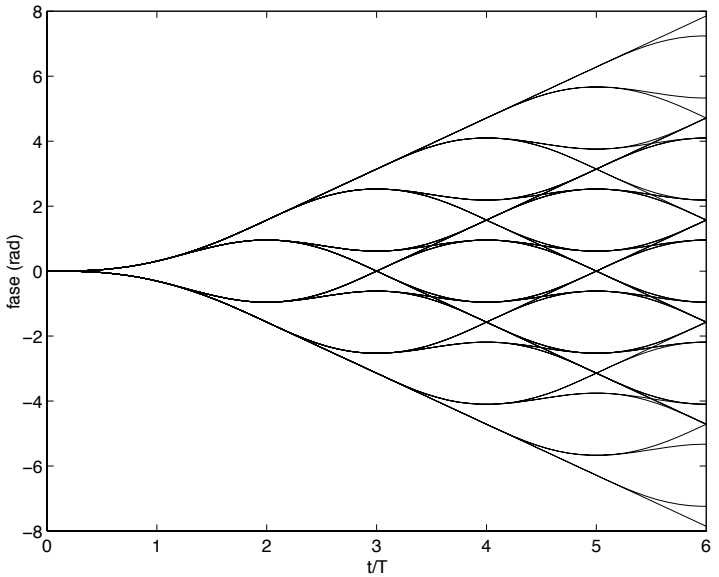

Fig. 7.2 - Traiettorie di fase per la modulazione binaria 3RC ($h = 0.5$)

Con impulso modulante 3RC non conviene nel caso binario superare $h = 0.5$, deviazione per cui $d^2 = 1.76$ come si può ottenere dalla (7.11). Per $M = 4$ la distanza d^2 raddoppia, e se si accetta un allargamento dello spettro si può guadagnare ancora qualcosa spingendosi fino ad $h = 0.6$ o 0.7.

7.5 Ricevitori semplificati

Un primo metodo che talvolta è utile per semplificare il ricevitore consiste, solo nel calcolo delle correlazioni in ricezione, nell'approssimare la forma d'onda modulante in fase $q(t)$ con una di durata minore. L'operazione non può essere fatta direttamente in trasmissione per non peggiorare, anche notevolmente, lo spettro. Lo scopo è quello di ridurre L e quindi sia il numero di correlazioni sia il numero di stati dell'algoritmo di Viterbi.

Un altro metodo, risolutivo in alcuni casi importanti, è quello di approssimare il segnale con una semplice modulazione d'ampiezza, traendo spunto dal caso MSK. Prima ancora che fosse disponibile una teoria in proposito, si

era scoperto in modo più o meno empirico che le modulazioni TFM e GMSK, descritte nel seguito, potevano essere demodulate con un ricevitore in fase e quadratura con *offset* simile a quello dell'MSK cambiando solo il filtro di ricezione, con una modesta degradazione delle prestazioni rispetto al ricevitore ottimo.

La teoria sviluppata successivamente, che risulta utile purtroppo solo per le modulazioni binarie, ha individuato uno sviluppo del segnale CPM in somma di 2^{L-1} segnali modulati in ampiezza, il primo dei quali è modulato dai dati a_k esattamente come nell'MSK, i restanti da combinazioni diverse dei dati. Nei due casi citati, quasi tutta la potenza del segnale è contenuta nel primo termine.

La stessa teoria indica poi come calcolare gli impulsi equivalenti in banda base e in particolare il primo, $h_0(t)$, e quindi consente di progettare il filtro di ricezione. Questo non sarà il semplice filtro adattato, perché in generale le repliche traslate di $h_0(t)$ non sono ortogonali, ed occorre tener conto dell'interferenza intersimbolica. Si può utilizzare come filtro la convoluzione di un filtro adattato e di un equalizzatore fisso. Un'altra soluzione, semplice ed efficace, è fare in modo che la cascata di $h_0(t)$ e del filtro di ricezione dia una risposta complessiva con spettro a coseno rialzato, con *roll-off* prefissato. Se si preferisce il ricevitore MLSE si possono includere altri termini dello sviluppo in somma di modulazioni d'ampiezza, se il loro contributo è significativo.

Un terzo metodo, utilizzabile anche per modulazioni multilivello, consiste nell'approssimare il segnale modulato durante ogni intervallo di simbolo con una sinusoide, di frequenza scelta in un insieme discreto di valori. Benché l'approssimazione possa sembrare brutale, si può mostrare che nella maggior parte dei casi pratici bastano solo sei diverse frequenze (o addirittura quattro), equispaziate con passo opportuno, per ottenere ottimi risultati. Basta quindi un numero limitatissimo di correlatori[6]. L'elaborazione viene fatta con l'algoritmo di Viterbi. Accenniamo, senza però approfondire, a due punti ai quali prestare attenzione: la base non è in generale ortogonale, e l'energia delle forme d'onda approssimate dipende, anche se non molto, dalla sequenza dei dati.

[6]dal punto di vista teorico ciò non è strano: il numero di dimensioni occupate da un segnale è circa $2BT$; i sistemi CPM, almeno quelli interessanti, hanno banda stretta; quindi, purché si trovi una base adeguata, bastano pochi correlatori

7.6 La modulazione TFM

Una modulazione CPM con spettro particolarmente compatto, proposta infatti originariamente per le comunicazioni radiomobili, è la *Tamed Frequency Modulation* (TFM), introdotta in modo euristico come variante dell'MSK. Ha traiettorie di fase molto addolcite, che alla fine di ogni simbolo portano la fase ad assumere valori sempre multipli $\pi/4$. La modulazione è binaria, con $h = 1/2$. La trasformata di Fourier della forma d'onda modulante in frequenza, centrata sull'asse dei tempi, è

$$G(f) = \frac{1}{2}\cos^2 \pi f T \frac{\pi f T}{\sin \pi f T} \tag{7.12}$$

per $|f| \leq 1/4T$ e nulla altrove. La forma d'onda modulante in fase $q(t)$ è mostrata in fig. 7.3. La durata è teoricamente infinita, ma in pratica si può ritenere[7] $L = 3$ o 4.

L'albero delle fasi è mostrato in fig. 7.4. Si noti l'andamento perfettamente costante della fase in certi tratti, corrispondenti alla trasmissione di dati positivi e negativi alternati.

Esaminando le traiettorie di fase si può mostrare che le sequenze di dati che producono i segnali a distanza minima sono $1, -1, \ldots$ e $-1, 1, \ldots$ e che la (7.11) fornisce come distanza normalizzata $d^2 = 1.59$, con un peggioramento quindi rispetto all'MSK (ovvero al binario antipodale) di 1 dB. È interessante osservare invece che le sequenze $1, 1, \ldots$ e $-1, -1, \ldots$ danno una distanza maggiore (es. 7.10). Questo tipo di errore sarà quindi poco frequente, perlomeno ad alto rapporto segnale-rumore.

Supponendo $L = 3 \div 4$, la teoria generale vorrebbe che nel demodulatore si usassero $16 \div 32$ correlatori e l'algoritmo di Viterbi con $8 \div 16$ stati. Ma se si usa l'approssimazione lineare discussa nella sezione precedente si può mostrare che il termine $h_0(t)$ contiene circa il 98% della potenza. Dunque la modulazione è con buona approssimazione lineare, e si può usare un semplice ricevitore con filtri in fase e quadratura e decisori binari, come per l'MSK. A conti fatti il costo del ricevitore semplificato è di pochi decimi di dB. La fig. 7.5 mostra l'andamento di $h_0(t)$; si osservi che la durata di $h_0(t)$ è decisamente maggiore di $2T$.

[7] ciò vale solo ai fini del progetto del ricevitore; nella generazione del segnale in trasmissione occorre maggior precisione, se si vuol ottenere l'ottimo spettro che la modulazione teoricamente consente

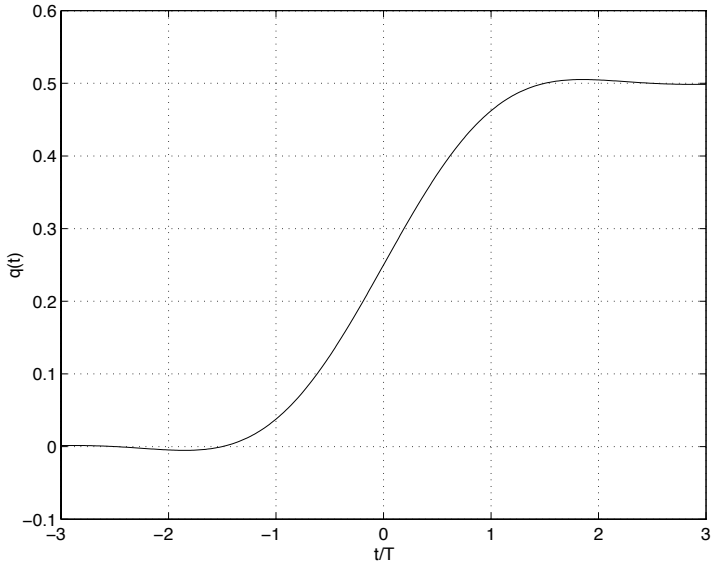

Fig. 7.3 - Forma d'onda modulante in fase (TFM)

7.7 La modulazione GMSK

La modulazione GMSK (Gaussian MSK) è una variante dell'MSK proposta per il radiomobile, che è stata poi effettivamente adottata per il sistema numerico europeo GSM.

L'indice di modulazione è $h = 1/2$, e gli impulsi rettangolari modulanti sono addolciti da un filtro gaussiano. La scelta della banda del filtro consente, entro certi limiti, di cercare un compromesso tra prestazioni e proprietà spettrali. Con una banda B a 3 dB pari a $0.2/T$ il segnale modulato risulta molto simile al TFM, mentre $B = \infty$ corrisponde all'MSK. Nel sistema GSM si è scelto $B = 0.3/T$. La risposta impulsiva del filtro è gaussiana, con $\sigma \approx 0.4T$ se $B = 0.3/T$. L'impulso modulante in frequenza $g(t)$ è la convoluzione del rettangolo di durata T con tale gaussiana.

Nel caso $B = 0.3/T$ l'impulso modulante in fase $q(t)$ e l'albero delle fasi risultano molto simili al 3RC di fig. 7.2. L'effetto del filtro gaussiano è lineare sul segnale modulante in frequenza e quindi anche sulla modulazione di fase $\varphi(t)$. Il filtro non ha alcun effetto quando si ha una sequenza di simboli uguali,

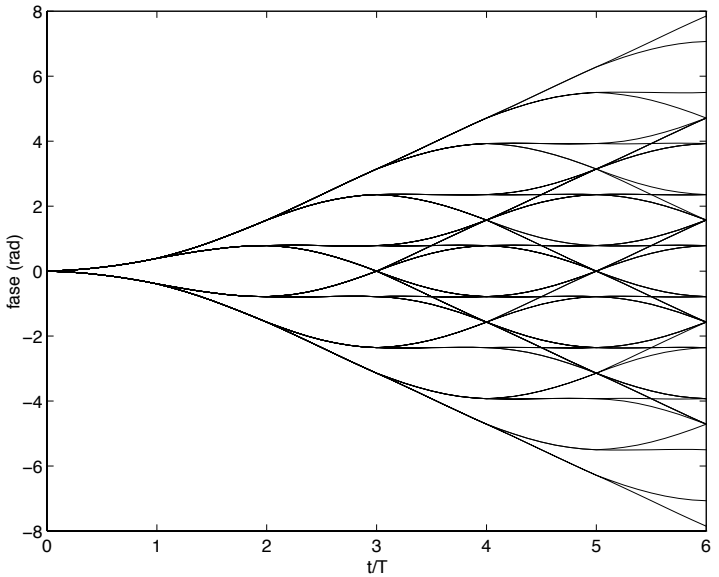

Fig. 7.4 - Albero delle fasi (TFM)

dal momento che il guadagno in continua è unitario. Con dati a segni alternati, l'andamento di $\varphi(t)$ a dente di sega con ampiezza $\pi/2$ dell'MSK è invece fortemente filtrato: ne resta solo la componente fondamentale a frequenza $1/2T$, e molto ridotta; l'escursione picco-picco è di circa 0.5 radianti.

Gli eventi errore sono gli stessi del TFM, e per $B = 0.3/T$ la distanza al quadrato normalizzata è $d^2 = 1.79$ per cui la perdita rispetto all'MSK è di circa 0.5 dB. Con tale valore di B il termine $h_0(t)$ contiene il 99.6% della potenza. La modulazione è quindi pressoché lineare, e si può utilizzare un ricevitore in fase e quadratura, con perdita modestissima rispetto all'ottimo. La fig. 7.6 mostra l'andamento di $h_0(t)$, che anche in questo caso ha durata maggiore di $2T$.

7.8 Spettro dei segnali CPM

Per il calcolo dello spettro di un segnale CPM basta determinare la funzione di autocorrelazione dell'equivalente passa basso $\exp(j\varphi(t))$. Tutti i segnali per

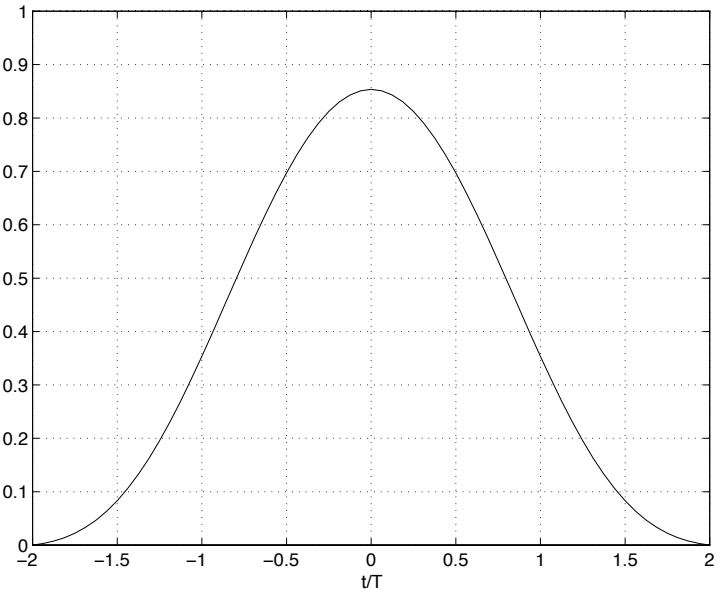

Fig. 7.5 - Forma d'onda elementare $h_0(t)$ (TFM)

la trasmissione numerica sono processi casuali non stazionari, ma piuttosto *ciclostazionari*, cioè con funzione di autocorrelazione $R(t, t+\tau)$ funzione periodica di t, con periodo pari al tempo di simbolo. Volendo definire una densità spettrale di potenza come se il processo fosse stazionario, cioè come trasformata di una funzione $R(\tau)$, occorre rimuovere la ciclostazionarietà mediando la funzione di autocorrelazione $R(t, t+\tau)$ su un periodo. Indicando la media temporale con la sopralineatura, si ha

$$R(\tau) = \overline{E[\exp(j\varphi(t+\tau))\exp(-j\varphi(t))]} \tag{7.13}$$

Si suppongano i dati a_k indipendenti ed equiprobabili. Sostituendo a $\varphi(t)$ e $\varphi(t+\tau)$ l'espressione (7.1), e utilizzando la funzione caratteristica dei dati, già incontrata nel Cap. 6,

$$C(f) = E[\exp(jua_k)] = \frac{\sin Mu}{M \sin u} \tag{7.14}$$

si ottiene

$$R(\tau) = \overline{\prod C\big(2\pi h(q(t+\tau-kT) - q(t-kT))\big)} \tag{7.15}$$

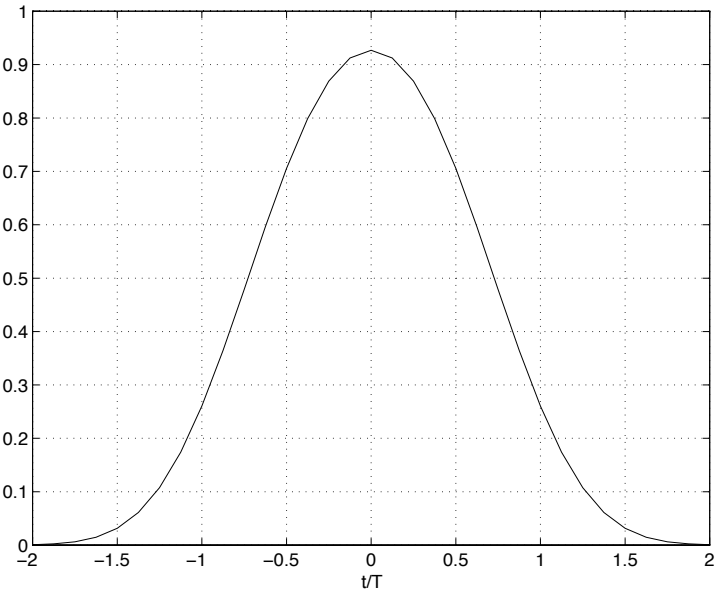

Fig. 7.6 - Forma d'onda elementare $h_0(t)$ (GMSK; $BT = 0.3$)

Per il calcolo si può anzitutto osservare che è sufficiente considerare solo $\tau \geq 0$, perché $R(-\tau) = R^*(\tau)$.

Per i valori di k per cui $t + \tau - kT$ e $t - kT$ sono entrambi negativi o entrambi maggiori di LT la funzione caratteristica vale 1, ed è inutile eseguire i corrispondenti prodotti. Basta quindi moltiplicare un piccolo numero di termini (che cresce però all'aumentare di τ).

Per $\tau \geq (L+1)T$ confrontando il calcolo di $R(\tau + nT)$, per $n \geq 1$, con quello di $R(\tau)$ si vede facilmente che $R(\tau + nT) = R(\tau)C^n(\pi h)$; non occorre quindi ripetere il calcolo dei prodotti. In definitiva basta applicare la (7.15) per $0 \leq \tau < (L+1)T$. I restanti campioni dell'autocorrelazione si ottengono di conseguenza.

Per la media temporale, se lo spettro è compatto basteranno pochi valori di t, ad esempio $0, T/4, T/2$ e $3T/4$, e altrettanti valori di τ per tempo di simbolo. Infine la densità spettrale di potenza è la trasformata di Fourier di $R(\tau)$, che ovviamente sarà calcolata con una FFT.

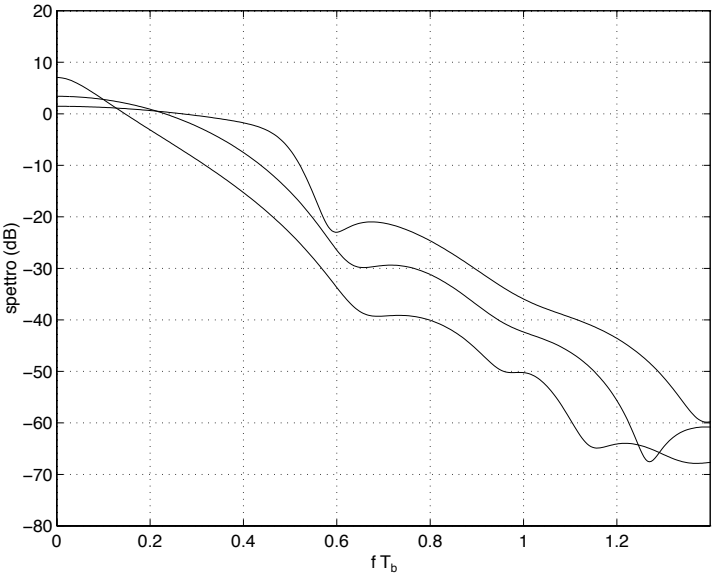

Fig. 7.7 - Spettro della modulazione 3RC ($M = 2$; $h = 0.3, 0.5, 0.7$)

Le fig. 7.7, 7.8 e 7.9 mostrano alcuni esempi di spettri nel caso 3RC binario e quaternario e per le modulazioni TFM e GMSK. Si noti che la frequenza è normalizzata al ritmo d'informazione in bit al secondo $1/T_b$, dove si indica con $T_b = T/\log_2 M$ l'intervallo tra bit d'informazione. Ad esempio nel caso quaternario $T_b = T/2$. Come mostrano le figure, lo spettro non cambia molto passando dalla modulazione binaria a quella quaternaria e mantenendo invariato h.

7.9 Considerazioni finali

La modulazione numerica di frequenza a fase continua (CPM) è particolarmente indicata quando si vogliono ottenere congiuntamente due proprietà che le modulazioni lineari non consentono: spettro compatto e inviluppo costante. L'applicazione dove ha incontrato maggior successo è il sistema radiomobile numerico, ma si sta diffondendo anche nei ponti radio a frequenze molto elevate.

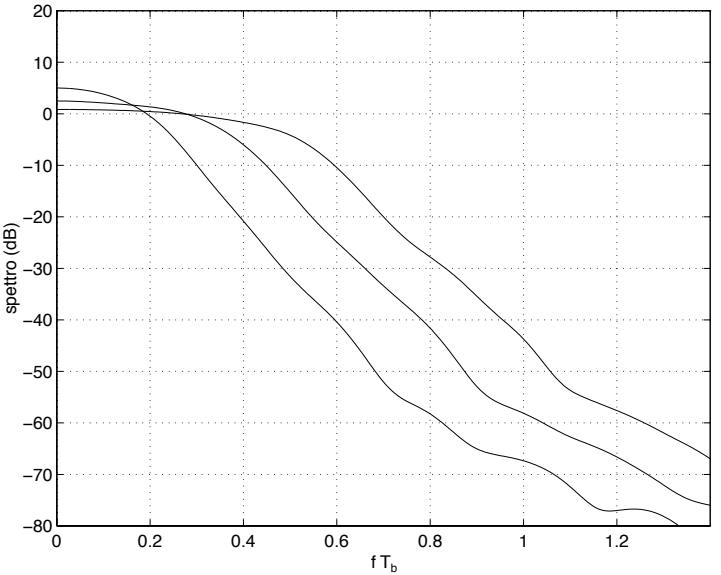

Fig. 7.8 - Spettro della modulazione 3RC ($M = 4$; $h = 0.3, 0.5, 0.7$)

La continuità di fase ha due effetti benefici: da un lato consente di ottenere, soprattutto con gli schemi a risposta parziale, spettri molto compatti del tutto impensabili senza questo vincolo; dall'altro introduce una, sia pur semplice, forma di dipendenza tra simboli adiacenti (equivalente ad una codifica) e quindi migliora anche l'efficienza nell'uso dell'energia.

Come indicazione di massima per la scelta tra i vari sistemi CPM, impulsi modulanti a risposta parziale di durata $3T$ o $4T$ possono offrire ottime prestazioni sia nel caso binario sia multilivello.

Dagli esempi presentati il lettore avrà potuto farsi un'idea delle prestazioni delle modulazioni binarie, che risultano particolarmente semplici per $h = 1/2$. Si può aggiungere, grossolanamente, che passando ad $M = 4$ ci si può attendere un guadagno asintotico rispetto al caso binario di 2 o 3 dB, a parità di banda occupata ma con un aumento non trascurabile della complessità del ricevitore.

I sistemi binari con indice di modulazione $h = 1/2$, come TFM e GMSK, sono invece demodulabili in modo molto semplice, e con piccola degradazione rispetto al ricevitore ML. Spesso in pratica i dati binari subiscono una precodifica differenziale prima di essere inviati al modulatore (es. 7.5).

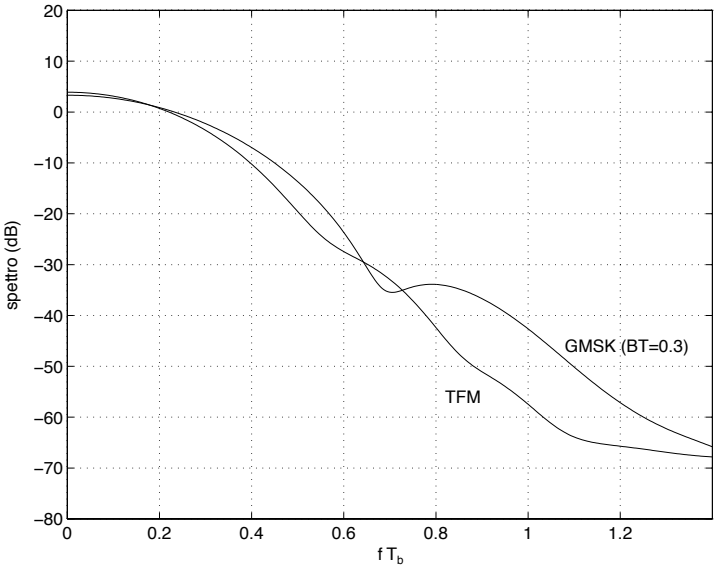

Fig. 7.9 - Spettro della modulazione TFM e GMSK ($BT = 0.3$)

Infine per le modulazioni CPM sono stati proposti anche molti tipi di ricevitori non coerenti, qui non analizzati per brevità.

7.10 Esercizi

7.1 - In un sistema CPM binario si consideri un impulso modulante in frequenza $g(t)$ con area nulla, costituito da due rettangoli di durata $T/2$ e con ampiezze opposte. L'impulso modulante in fase $\varphi(t)$ è quindi triangolare, con ampiezza Φ. Si mostri che il ricevitore può operare indipendentemente simbolo per simbolo, e si calcoli la distanza minima in funzione di Φ. Si trovi il valore di Φ che dà le migliori prestazioni e la corrispondente distanza (che comunque, come si vedrà, non risulta interessante).

7.2 - Si riconsideri l'esercizio precedente, con l'unica variante della segnalazione a quattro livelli con deviazione $\pm\Phi$, $\pm 3\Phi$.

7.3 - Si mostri che nei casi dei due esercizi precedenti lo spettro contiene una

riga, ovvero che una parte della potenza viene impiegata per trasmettere una componente sinusoidale alla frequenza della portante, senza recare informazione.

7.4 - Si mostri che si può scrivere l'equivalente passa basso del segnale MSK come

$$z(t) = \sum j^k c_k h_0(t - kT)$$

dove $h_0(t) = \sin \frac{\pi}{2} \frac{t-kT}{T}$ nell'intervallo $(0, 2T)$ e $c_k = \ldots a_{k-2} a_{k-1} a_k$.
Suggerimento: nell'intervallo $(kT, kT + T)$ si ha

$$z(t) = \exp(j\frac{\pi}{2} \ldots a_{k-2} a_{k-1}) \exp(j\frac{\pi}{2} a_k \frac{t - kT}{T}) =$$

$$= j^{k-1} \ldots a_{k-2} a_{k-1} \Big(\cos \frac{\pi}{2} \frac{t - kT}{T} + j a_k \sin \frac{\pi}{2} \frac{t - kT}{T} \Big)$$

e un'analoga espressione vale nell'intervallo $(kT + T, kT + 2T)$; si individui, nei *due* intervalli, il termine proporzionale a $j^k \ldots a_{k-1} a_k$. *Commento*: si tratta pertanto di una modulazione d'ampiezza con *offset* (in fase per k pari, in quadratura per k dispari); si osservi anche che j^k corrisponde ad un cambiamento di segno tra c_{k-2} e c_k, cioè per i dati trasmessi sullo stesso asse.

7.5 - Sia $\{b_k\}$ la sequenza di dati binari da trasmettere con modulazione MSK, e si utilizzi la precodifica differenziale $a_k = a_{k-1} b_k$. Si mostri che $b_k = c_k c_{k-2}$, e che quindi i dati b_k si ottengono dal confronto tra i segni di dati consecutivi su uno stesso asse (in fase, o in quadratura). *Commento*: in pratica occorre tener conto del cambiamento di segno introdotto da j^k; si osservi anche che se si sbaglia c_k si commettono *due* errori (b_k e b_{k+2}). *Commento*: considerazioni analoghe valgono per gli schemi di modulazione binari con indice di modulazione $h = 1/2$ rappresentabili (approssimativamente) come modulazioni in fase e quadratura (TFM, GMSK).

7.6 - Nelle modulazioni MSK, TFM e GMSK è possibile una precodifica dei dati b_k tale da ottenere $c_k = b_k$? e presenterebbe qualche inconveniente? *Suggerimento*: si consideri l'effetto di una rotazione di fase pari ad un multiplo di $\pi/2$ dovuta ad errata sincronizzazione di portante.

7.7 - Si consideri un sistema CPM con segnale modulante in frequenza $g(t)$ rettangolare di durata T (1REC) e segnalazione a M livelli. Si calcoli, almeno per piccoli valori dell'indice di modulazione h, la distanza minima. Si tracci il grafico per $M = 2$ e $M = 4$, e si spieghi perché il calcolo non è altrettanto semplice per valori elevati di h. Si mostri che per $M = 4$ i valori dell'indice di modulazione h=1/3, 1/2 e 2/3 sono deboli, e si calcolino le corrispondenti distanze minime, confrontandole con il grafico per i valori di h non deboli.

7.8 - Si consideri il sistema di modulazione binaria di frequenza a fase continua (CPM) in cui gli impulsi modulanti in frequenza sono rettangoli di durata $2T$, dove T è l'intervallo tra bit successivi. L'indice di modulazione h è pari a 1/2, cioè il contributo complessivo di ciascun bit alla fase è $\pm\pi/2$.
Quale è l'albero delle fasi, e quale la distanza minima tra i segnali (e quindi le prestazioni del ricevitore ottimo)? Come può essere fatto il ricevitore? Si può pensare ad un ricevitore con correlatori (o filtri adattati) in fase e quadratura, anziché con l'algoritmo di Viterbi? *Suggerimento*: si veda se la modulazione qui proposta ha qualche affinità con altre note.

7.9 - Si consideri il sistema di modulazione binaria di frequenza a fase continua (CPM) in cui gli impulsi modulanti in frequenza sono triangoli simmetrici di durata $2T$, dove T è l'intervallo tra bit successivi.
Quale è l'albero delle fasi, e quale (in funzione dell'indice di modulazione h) la distanza minima tra i segnali? Come può essere fatto il ricevitore, e quali ne sono le prestazioni?

7.10 - Utilizzando la (7.11) si verifichi il valore $d^2 = 1.59$ dato nel testo per la distanza minima nella modulazione TFM. Si calcoli anche la distanza tra i segnali corrispondenti alle ipotesi 1,1 e $-1,-1$, verificando che è maggiore della minima.

7.11 - Per valori interi dell'indice di modulazione h (valori peraltro deboli, e normalmente di nessun interesse pratico) risulta $|C(\pi h)| = 1$. Si mostri che ciò implica che $R(\tau)$ contenga una componente periodica, e quindi lo spettro contenga delle righe. In altre parole, una parte della potenza viene spesa per trasmettere inutili componenti sinusoidali.

7.12 - Si mostri che la ricerca della distanza minima tra segnali può essere

condotta inviando ad un ricevitore con algoritmo di Viterbi il segnale CPM (senza rumore) modulato da una sequenza casuale di dati e registrando la minima distanza (al quadrato) tra il percorso corretto ed i concorrenti. Si potrebbero usare anche le correlazioni, anziché le distanze al quadrato?

Capitolo 8

Stima di parametri continui

8.1 Introduzione

In un contesto di trasmissione numerica potrebbe sembrare fuori luogo occuparsi della stima di parametri continui. Tuttavia non lo è per almeno due buoni motivi. Innanzitutto quando si passa dalla teoria generale all'implementazione pratica dei sistemi di trasmissione numerica ci si scontra immediatamente con molti parametri continui che è necessario stimare: posizione temporale dei simboli, frequenza e fase della portante (in banda passante) sono gli esempi più tipici, che verranno ripresi in un apposito capitolo. Inoltre la rappresentazione geometrica dei segnali, rivelatasi finora così utile perlomeno nel caso di rumore gaussiano, è altrettanto indicata per determinare gli stimatori di parametri continui e le relative prestazioni.

Mentre nel caso discreto si utilizza una terminologia associata alla parola *decisione*, nel caso continuo si parla di *stima*. Occorre notare subito che concetti come la probabilità d'errore perdono ogni interesse nel caso continuo. La stima di una grandezza *reale* non può avere precisione infinita, tanto più in presenza di rumore; descrivere tale situazione affermando che la probabilità d'errore è uno, qualunque sia lo stimatore, non è evidentemente appropriato.

Ciò è analogo a quel che avviene quando nello studio della teoria delle probabilità si passa dagli eventi discreti alle variabili casuali continue. La probabilità che una variabile casuale abbia un preciso valore solitamente non dice nulla, perché tale probabilità è zero. Per una descrizione accurata si deve dare la *densità di probabilità* (*ddp*) della variabile casuale. Talvolta ci si accontenta, soprattutto se costretti dalla difficoltà del calcolo, di indicatori

sintetici come valor medio e varianza della variabile casuale.

Nel caso dello stimatore di un parametro continuo si è interessati all'errore commesso nella stima, cioè alla differenza tra valore stimato del parametro e valore vero. Il valor medio eventualmente non nullo dell'errore indica che lo stimatore è *polarizzato* cioè affetto da un errore sistematico (detto *bias*); la varianza oppure lo scarto quadratico medio danno una indicazione sintetica sull'errore che ci si può attendere. Polarizzazione e varianza possono essere funzioni del valore vero del parametro. Una descrizione più accurata del comportamento dell'errore è fornita dalla sua densità di probabilità, di cui però di solito non si sente un'esigenza stringente (basti pensare che tale *ddp* può essere funzione del valore del parametro, per cui si dovrebbero dare infinite *ddp*).

Prima ancora di affrontare la teoria conviene osservare che non esiste *lo* stimatore di un certo parametro, ma che se ne possono inventare a volontà; se ne possono individuare alcuni semplici e, soprattutto, sensati.

8.2 Criteri di stima

Si supponga di ricevere

$$r(t) = s(t, \alpha) + n(t) \tag{8.1}$$

dove $s(t, \alpha)$ è una forma d'onda che dipende dal parametro continuo α in modo noto al ricevitore[1] ed $n(t)$ è rumore additivo gaussiano, bianco nella banda dei segnali. Rappresentando segnali e rumore geometricamente si ha

$$\mathbf{r} = \mathbf{s}(\alpha) + \mathbf{n} \tag{8.2}$$

con l'unica avvertenza che ora il numero di possibili segnali $\mathbf{s}(\alpha)$ non è finito, e quindi non è escluso che, almeno in teoria, occorra uno spazio a infinite dimensioni.

Nel caso discreto *decidere* significa individuare un valore dell'indice i (il più probabile a posteriori, o il più verosimile). Ora *stimare* vuol dire attribuire un valore al parametro α. Se nel caso discreto si cercava il massimo delle probabilità a posteriori $P(\mathbf{s}_i/\mathbf{r})$ o delle verosimiglianze $f(\mathbf{r}/\mathbf{s}_i)$, ora si chiamano stimatori MAP e ML rispettivamente quelli che corrispondono al massimo della *ddp* a posteriori $f(\alpha/\mathbf{r})$ o della verosimiglianza $f(\mathbf{r}/\alpha)$.

[1] α potrebbe essere un'ampiezza, fase, frequenza, ecc.

Con la regola di Bayes si ottiene

$$f(\alpha/\mathbf{r}) = \frac{f(\mathbf{r}/\mathbf{s}(\alpha))f(\alpha)}{f(\mathbf{r})} \tag{8.3}$$

e, al solito, si può ignorare il denominatore nella ricerca del massimo. Analogamente al caso discreto si deve dare una *ddp a priori* del parametro α. Ciò può essere fastidioso o imbarazzante, e complica lo stimatore, per cui normalmente si preferisce considerare la verosimiglianza $f(\mathbf{r}/\mathbf{s}(\alpha))$.

Se $\hat{\alpha}$ è il valore stimato del parametro che corrisponde al massimo della *ddp* a posteriori o della verosimiglianza, quale è la probabilità che sia $\hat{\alpha} = \alpha$? Zero, ovviamente, essendo $\hat{\alpha}$ una variabile casuale. Dunque lo stimatore MAP, ed anche la sua versione semplificata ML, non sembrano imporsi senza discussione, contrariamente al caso discreto. Ciò naturalmente non significa che siano cattivi stimatori, ma solo che bisogna giustificarne meglio la validità.

A tale scopo occorre chiedersi cosa si potrebbe trovare di meglio. Una volta ricevuta la forma d'onda $r(t)$, tutta l'informazione disponibile su α è contenuta nella *ddp* a posteriori $f(\alpha/\mathbf{r})$. In presenza di rumore gaussiano bianco, questa è proporzionale a

$$f(\alpha/\mathbf{r}) \equiv \exp\Big(-\frac{1}{N_0}|\mathbf{r} - \mathbf{s}(\alpha)|^2\Big) f(\alpha) \tag{8.4}$$

Nel caso in esame la verosimiglianza è una funzione gaussiana di $\mathbf{r}$, ma non di α, in generale.

È possibile che la verosimiglianza e la *ddp* a posteriori abbiano due o più massimi di ampiezza quasi pari. In tal caso le stime ML e MAP, che ne privilegiano uno, non sono molto convincenti. Si potrebbe piuttosto cercare di rendere minima la varianza dell'errore (stimatore MMSE, a minimo errore quadratico medio). È ben noto che ciò si ottiene scegliendo come valore stimato $\hat{\alpha}$ il valor medio condizionato $E[\alpha/\mathbf{r}]$. Infatti per qualsiasi variabile casuale i momenti non centrali del secondo ordine sono maggiori di quello centrale. Il valor medio condizionato è (es. 8.1)

$$\hat{\alpha} = E[\alpha/\mathbf{r}] = \int \alpha f(\alpha/\mathbf{r}) d\alpha = \frac{\int \alpha \exp\Big(\frac{1}{N_0}|\mathbf{r} - \mathbf{s}(\alpha)|^2\Big) f(\alpha) d\alpha}{\int \exp\Big(\frac{1}{N_0}|\mathbf{r} - \mathbf{s}(\alpha)|^2\Big) f(\alpha) d\alpha} \tag{8.5}$$

espressione non del tutto rassicurante[2]. Fortunatamente in molti casi di interesse pratico ogni dubbio è risolto dal fatto che $f(\mathbf{r}/\mathbf{s}(\alpha))$ è approssimabile,

[2]in linea di principio si può precalcolare e memorizzare il valor medio condizionato per un insieme sufficientemente fitto di valori di $\mathbf{r}$; ma in pratica ciò risulta possibile solo in un numero molto limitato di dimensioni

come si vedrà fra breve, con una funzione di α gaussiana, e sufficientemente stretta, per cui le stime MAP, ML e MMSE sono vicine tra loro, soprattutto se la *ddp* a priori $f(\alpha)$ non varia troppo rapidamente. Spesso, quindi, la scelta dello stimatore è basata sulla semplicità di realizzazione più che sulle prestazioni[3].

8.3 Stima a massima verosimiglianza

Senza dubbio tra gli stimatori quello a massima verosimiglianza è concettualmente il più semplice. Richiede infatti di rendere minima la distanza $|\mathbf{r}-\mathbf{s}(\alpha)|$, esattamente come nel caso discreto con l'unica differenza che i vettori candidati sono infiniti. Al variare di α l'estremo del vettore $\mathbf{s}(\alpha)$ descrive una linea nello spazio, che si può immaginare punteggiata in α. Fra tutti i punti del luogo basta individuare quello a distanza minima dal vettore ricevuto $\mathbf{r}$, ed assegnare ad $\hat{\alpha}$ il valore di α corrispondente al punto.

Come nel caso discreto, il minimo della distanza equivale al massimo di

$$\mathbf{r} \cdot \mathbf{s}(\alpha) - \frac{1}{2}|\mathbf{s}(\alpha)|^2 \tag{8.6}$$

In particolare se l'energia $|\mathbf{s}(\alpha)|^2$ non dipende dal parametro basta cercare il massimo della correlazione $\mathbf{r} \cdot \mathbf{s}(\alpha)$.

Quanto poi la ricerca risulti semplice in pratica dipende dai casi, ovvero dalla geometria del luogo $\mathbf{s}(\alpha)$, come verrà chiarito da alcuni esempi presentati nel seguito.

Per il calcolo delle prestazioni dello stimatore ML, e in particolare dell'errore quadratico medio, si supponga di aver trasmesso il generico vettore $\mathbf{s}(\alpha_0)$ corrispondente al valore α_0 del parametro. Se il rumore è sufficientemente debole, $\mathbf{s}(\hat{\alpha})$ si discosta poco da $\mathbf{s}(\alpha_0)$. Si può quindi approssimare localmente il luogo geometrico con la tangente in $\mathbf{s}(\alpha_0)$, punteggiata uniformemente in α, ed $\mathbf{s}(\hat{\alpha})$ con la proiezione di $\mathbf{r}$ sulla tangente anziché sul luogo (fig. 8.1). Si ha quindi, per un generico α non troppo discosto da α_0,

$$\mathbf{s}(\alpha) \approx \mathbf{s}(\alpha_0) + (\alpha - \alpha_0)\frac{d\mathbf{s}(\alpha_0)}{d\alpha} \tag{8.7}$$

[3] non si vorrebbero mai incontrare *ddp* a posteriori multimodali, anche perché spesso la ricerca del massimo deve essere condotta con tecniche iterative che rendono assai facile finire in un massimo locale anziché globale; si deve comunque riconoscere che in questi casi lo stimatore MMSE sarebbe il più raccomandabile

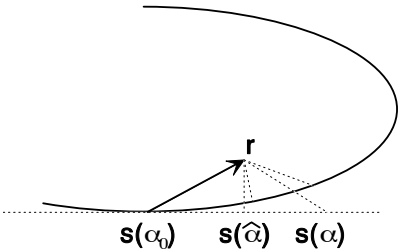

Fig. 8.1 - Rappresentazione geometrica di un tratto del luogo $\mathbf{s}(\alpha)$ e approssimazione con la tangente in un intorno del punto trasmesso

La fig. 8.1 mostra anche che la distanza al quadrato tra il vettore ricevuto $\mathbf{r}$ e il generico punto del luogo $\mathbf{s}(\alpha)$ è approssimabile con

$$\begin{aligned} &|\mathbf{r}-\mathbf{s}(\alpha)|^2 \approx |\mathbf{r}-\mathbf{s}(\hat{\alpha})|^2 + |\mathbf{s}(\hat{\alpha})-\mathbf{s}(\alpha)|^2 \approx \\ &\approx |\mathbf{r}-\mathbf{s}(\hat{\alpha})|^2 + (\hat{\alpha}-\alpha)^2 \Big|\frac{d\mathbf{s}(\alpha_0)}{d\alpha}\Big|^2 \end{aligned} \tag{8.8}$$

La verosimiglianza è proporzionale a $\exp(-|\mathbf{r}-\mathbf{s}(\alpha)|^2/N_0)$ ed è quindi in prima approssimazione una funzione gaussiana di α centrata in $\hat{\alpha}$. Come già osservato in tal caso gli stimatori ML, MAP ed MMSE sono quasi coincidenti. Dalla (8.7), valutata in $\alpha = \hat{\alpha}$, si ha poi che $\hat{\alpha} - \alpha_0$ è dato dal rapporto tra le ampiezze dei vettori $\mathbf{s}(\hat{\alpha}) - \mathbf{s}(\alpha_0)$ e $d\mathbf{s}(\alpha_0)/d\alpha$. Il primo dei due vettori è la componente del rumore lungo la tangente al luogo, quindi è una variabile casuale gaussiana con valor medio nullo e varianza $N_0/2$. Anche l'errore di stima è dunque approssimativamente gaussiano, ed ha varianza

$$E[(\hat{\alpha}-\alpha_0)^2] = \frac{N_0/2}{\Big|\frac{d\mathbf{s}(\alpha_0)}{d\alpha}\Big|^2} \tag{8.9}$$

Il fattore $|d\mathbf{s}(\alpha_0)/d\alpha|$ è la cosiddetta *dilatazione* del luogo[4]. La dilatazione (o *stretch*) è in generale una funzione del valore trasmesso α_0. In caso di dilatazione uniforme basta evidentemente calcolare il rapporto tra la lunghezza L del luogo e l'escursione $\Delta\alpha = \alpha_{max} - \alpha_{min}$ del parametro.

[4] si pensi ad un elastico punteggiato in α; si può variare la dilatazione tendendolo più o meno

Tuttavia, ancora una volta, non è strettamente indispensabile ricorrere alla rappresentazione geometrica per determinare la dilatazione del luogo. Infatti il vettore $d\mathbf{s}(\alpha_0)/d\alpha$ corrisponde alla forma d'onda $\partial s(t,\alpha_0)/\partial\alpha$ ed il suo modulo quadro non è altro che l'energia di tale forma d'onda:

$$\left|\frac{d\mathbf{s}(\alpha_0)}{d\alpha}\right|^2 = \int \left(\frac{\partial s(t,\alpha_0)}{\partial\alpha}\right)^2 dt \tag{8.10}$$

L'approssimazione del luogo con la tangente è lecita in caso di rumore *debole*; comincia a perdere validità quando diventano sensibili gli effetti della curvatura del luogo, e diventa del tutto inapplicabile quando non è trascurabile la probabilità che il vettore ricevuto $\mathbf{r}$ si avvicini troppo ad un altro ramo del luogo. Se ad esempio in fig. 8.1 si aumenta la componente del rumore nella direzione perpendicolare al luogo, la verosimiglianza $\exp(-|\mathbf{r}-\mathbf{s}(\alpha)|^2/N_0)$ dapprima diventa una funzione bimodale di α, poi ritorna ad avere un solo picco ma in corrispondenza di un valore gravemente sbagliato di α per cui l'approssimazione lineare perde ogni senso. L'effetto di tali errori, detti *anomali*, è difficile da calcolare e comunque peggiora nettamente le prestazioni rispetto a quanto previsto dalla (8.9). In primissima approssimazione la probabilità di fenomeni anomali è dell'ordine di $Q(d/\sqrt{2N_0})$, dove d è la distanza tra rami diversi del luogo. Occorre quindi che il luogo non ritorni ad una distanza inferiore a diverse volte $\sqrt{2N_0}$ da punti da cui è già passato.

Per concludere questi pochi cenni teorici si possono citare i casi di stima di un parametro in presenza di uno o più parametri indeterminati, oppure della stima congiunta di più parametri, quando la forma d'onda osservata è ad esempio $s(t,\alpha_1,\alpha_2)+n(t)$. In tal caso il luogo è una superficie punteggiata in α_1 e α_2 anziché una linea.

La stima ML di più parametri consiste nell'individuare il punto della superficie più vicino al vettore ricevuto. La teoria è abbastanza facilmente generalizzabile. Un punto da sottolineare è che in generale gli errori commessi nella stima *congiunta* di α_1 ed α_2 sono correlati. Infatti i vettori

$$\frac{\partial\mathbf{s}(\alpha_1,\alpha_2)}{\partial\alpha_i} \qquad i=1,2 \tag{8.11}$$

non sono, in generale, ortogonali. Quando dalle coordinate del vettore $\mathbf{s}(\hat{\alpha}_1,\hat{\alpha}_2)$ a distanza minima da $\mathbf{r}$ si risale ai valori dei parametri, i due errori risultano correlati. Il lettore non dovrebbe faticare a credere[5] che la matrice,

[5] se invece stenta a convincersi, occorrerebbe probabilmente troppo tempo per trattare l'argomento

di dimensione 2 per 2, delle covarianze degli errori di stima è pari a $N_0/2$ volte l'inversa della matrice che ha per elementi i quattro prodotti scalari

$$\frac{\partial \mathbf{s}(\alpha_1,\alpha_2)}{\partial \alpha_i}\frac{\partial \mathbf{s}(\alpha_1,\alpha_2)}{\partial \alpha_j} \qquad i,j=1,2 \tag{8.12}$$

Il caso di un parametro α_1 da stimare in presenza di un parametro indeterminato α_2 potrebbe ricondursi alla ricerca del massimo della media delle verosimiglianze condizionate. Spesso però la complessità è tale da preferire la stima congiunta di α_1 e α_2.

8.4 Esempi di stima ML

L'esempio più semplice è la *stima dell' ampiezza* dell'impulso $\alpha g(t)$. Occorre una sola funzione base, $\Phi_1(t) = g(t)/\sqrt{E_g}$ dove E_g è l'energia di $g(t)$. La rappresentazione geometrica del luogo $\mathbf{s}(\alpha)$ è molto semplice: si tratta di un segmento dell'asse Φ_1, ed il vettore $\mathbf{s}(\alpha)$ ha ascissa $\alpha\sqrt{E_g}$. La punteggiatura è lineare in α, ed è quindi evidente che una volta calcolata la componente

$$r_1 = \frac{\int r(t)g(t)dt}{\sqrt{E_g}} \tag{8.13}$$

del vettore ricevuto si ha $\hat{\alpha} = r_1/\sqrt{E_g}$, e quindi in definitiva

$$\hat{\alpha} = \frac{\int r(t)g(t)dt}{E_g} \tag{8.14}$$

Una semplice verifica, di norma consigliabile, è basata sull'osservare che se venisse a mancare il rumore si avrebbe perfetta coincidenza tra la stima ed il valore vero del parametro. È immediato verificare che se nella (8.14) si sostituisce $\alpha g(t)$ ad $r(t)$ si ottiene effettivamente $\hat{\alpha} = \alpha$.

Per quanto riguarda l'errore quadratico medio la dilatazione del luogo è facilmente calcolabile sia geometricamente sia mediante la (8.10), ed il risultato è $|d\mathbf{s}(\alpha)/d\alpha| = \sqrt{E_g}$. Si ha quindi

$$E[(\hat{\alpha}-\alpha_0)^2] = \frac{N_0/2}{E_g} \tag{8.15}$$

In una sola dimensione l'unico modo per ridurre l'errore è aumentare l'energia; un'osservazione analoga vale, come si ricorderà, per la trasmissione numerica

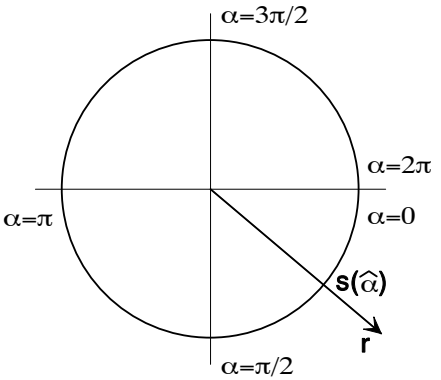

Fig. 8.2 - Stima della fase di un impulso: geometria del luogo $\mathbf{s}(\alpha)$ e punteggiatura in α

multilivello. Invadendo altre dimensioni si può incurvare il luogo e quindi ottenere una lunghezza considerevole senza aumentare l'energia, ma naturalmente a prezzo di una maggior banda occupata.

Non si può però esagerare nel ripiegare il luogo perché occorre proteggersi dagli errori anomali. Anche per la stima di parametri esiste una teoria, analoga a quella della capacità di canale, che determina le prestazioni limite senza vincoli sulla complessità dei segnali e dello stimatore.

Un semplice esempio di stima ML in due dimensioni è la *stima della fase* α di $s(t,\alpha) = g(t)\cos(2\pi f_0 t + \alpha)$. Il luogo occupa due dimensioni, come mostrato in fig. 8.2 dove si suppone che il parametro α abbia la massima escursione possibile, pari a 2π. Qualunque sia il vettore ricevuto $\mathbf{r}$, esterno o interno alla circonferenza, è facile riconoscere che il punto più vicino del luogo ha lo stesso argomento di $\mathbf{r}$, e quindi, osservando che la punteggiatura è antioraria,

$$\hat{\alpha} = -\arg \mathbf{r} = \arctan \frac{-r_2}{r_1} \tag{8.16}$$

dove l'arcotangente è da intendersi a quattro quadranti. Nel calcolo di r_1 ed r_2 può essere ignorata una stessa costante moltiplicativa, e quindi

$$r_1 = \int r(t) g(t) \cos 2\pi f_0 t dt \tag{8.17}$$

$$r_2 = \int r(t) g(t) \sin 2\pi f_0 t dt \tag{8.18}$$

Detta E_s l'energia del segnale, indipendente da α, la dilatazione del luogo è la lunghezza complessiva $2\pi\sqrt{E_s}$ divisa per l'escursione 2π del parametro, e quindi si ottiene

$$E[(\hat{\alpha} - \alpha_0)^2] = \frac{N_0/2}{E_s} \tag{8.19}$$

risultato che tornerà utile a proposito della sincronizzazione di portante.

La *stima della pulsazione* α del segnale $s(t, \alpha) = A\cos\alpha t$, osservato in presenza di rumore gaussiano bianco nell'intervallo di durata T che va da $-T/2$ a $T/2$ dà qualche problema in più. Se l'escursione del parametro è ampia il luogo[6] si contorce in modo difficile da visualizzare. Basta pensare che la distanza dall'origine $\sqrt{E_s}$ è costante e quindi il luogo giace su una sfera, e che per valori di α che differiscono per multipli di $2\pi/T$ i vettori $\mathbf{s}(\alpha)$ sono ortogonali. Poiché $|\mathbf{s}(\alpha)|^2$ non dipende dal parametro, la minima distanza equivale alla massima correlazione $\mathbf{r} \cdot \mathbf{s}(\alpha)$, e quindi lo stimatore ML deve trovare il valore di α che dà il massimo di

$$\int_{-T/2}^{T/2} r(t)\cos\alpha t dt \tag{8.20}$$

La ricerca, da eseguire in tutto l'intervallo di valori di pulsazione possibili, non è banale. In linea di principio occorrerebbe davvero calcolare infiniti integrali. Un modo di procedere può consistere in una prima localizzazione grossolana del massimo, campionando non troppo fittamente, seguita da un affinamento ricorsivo con qualche metodo numerico; se il punto iniziale è vicino al massimo la procedura converge. Per la prima fase lo strumento più raccomandabile è la trasformata discreta di Fourier (es. 8.5). In alcuni casi non occorre neppure procedere alla seconda fase, se il campionamento in pulsazione è già sufficientemente fitto. Si noti bene che una precisione elevata nella ricerca del valore di $\hat{\alpha}$ è illusoria: la precisione è comunque limitata dal rumore, ed è inutile utilizzare un campionamento con passo molto più fitto della deviazione standard attesa per l'errore (es. 8.5).

[6] il *serpente*, si dice anche in gergo colloquiale

Per quanto riguarda l'errore si ha, in presenza di rumore debole,

$$E[(\hat{\alpha}-\alpha_0)^2] = \frac{N_0/2}{\int_{-T/2}^{T/2} \left(\frac{\partial s(t,\alpha)}{\partial \alpha}\right)^2 dt} = \frac{N_0/2}{\int_{-T/2}^{T/2} t^2 \sin^2 \alpha t \, dt} = \tag{8.21}$$

$$= \frac{12N_0}{A^2T^3} = \frac{6N_0}{E_sT^2}$$

Si osservi la dipendenza non solo dall'energia del segnale ma anche dal quadrato della durata.

Se invece si ripetesse il calcolo con intervallo di osservazione $(0,T)$ anziché $(-T/2,T/2)$ si troverebbe un risultato quattro volte migliore. Ciò non deve sorprendere: l'espressione della dilatazione del luogo mostra chiaramente che un piccolo intervallo intorno a t ha peso t^2, e quindi in particolare un intorno dell'origine ha peso nullo. Ciò dipende dal fatto che qualunque sia α si è fissata a zero la fase all'istante $t=0$ e quindi solo allontanandosi dall'origine il segnale sente, e sempre più, il valore della pulsazione. Se si traslasse intorno all'origine il segnale $A\cos 2\pi\alpha t$ osservato nell'intervallo $(0,T)$ si riconoscerebbe una modulazione di fase in aggiunta a quella di frequenza, e cioè un segnale diverso da quello prima considerato[7].

Infine si consideri la *stima della posizione* dell'impulso $s(t,\alpha)=g(t-\alpha)$. Poiché l'energia non dipende dal parametro basta cercare il massimo della correlazione

$$\mathbf{r}\cdot\mathbf{s}(\alpha) = \int r(t)g(t-\alpha)dt \tag{8.22}$$

Se le correlazioni sono eseguite con circuiteria analogica, cioè con un filtro adattato, l'uscita del filtro, continua nel tempo, fornisce in tempo reale *tutte* le correlazioni desiderate: la correlazione con $s(t,\alpha)$ è l'uscita del filtro all'istante $\alpha+t_0$, dove t_0 è l'inevitabile ritardo da introdurre per rendere causale il filtro.

Se invece le correlazioni sono eseguite numericamente, occorre effettivamente calcolarne un buon numero. Al solito la discretizzazione dei valori di α

[7] non venga però la tentazione di utilizzare l'intervallo di osservazione $(t_0-T/2, t_0+T/2)$, con t_0 molto grande; il luogo verrebbe molto simile, localmente, ad una spirale lunghissima ma con passo molto piccolo; una piccola variazione $\Delta\alpha$ della pulsazione porta subito ad una distanza considerevole a causa della modulazione di fase, ma poi il luogo torna molto vicino ad $\mathbf{s}(\alpha_0)$ per valori di $\Delta\alpha t_0$ multipli di 2π. Solo con rumore debolissimo si eviterebbero errori anomali

sarà determinata sulla base della precisione consentita dal rumore.

L'errore quadratico medio è, con rumore debole,

$$E[(\hat{\alpha}-\alpha_0)^2] = \frac{N_0/2}{\int \left(\frac{\partial s(t,\alpha)}{\partial \alpha}\right)^2 dt} = \frac{N_0/2}{\int g'^2(t)dt} = \frac{N_0/2}{E_{g'}} \tag{8.23}$$

e dipende non direttamente dall'energia di $g(t)$ ma da quella della derivata $g'(t)$. Chiaramente lunghi tratti in cui $g(t)$ fosse costante non sarebbero di alcun aiuto nel localizzare temporalmente la forma d'onda; sono invece utili i fronti ripidi[8].

Una applicazione tipica della stima della posizione di un impulso si ha nel radar. La misura del ritardo dell'eco prodotta dall'eventuale bersaglio si traduce nell'informazione sulla distanza. Spesso interessa anche la velocità del bersaglio, misurabile attraverso lo spostamento Doppler della frequenza della portante. Si ha quindi un caso di stima congiunta di due parametri.

8.5 Stima non coerente

Per concludere con un breve cenno al caso di stima non coerente, si riprenda ad esempio il caso della modulazione di pulsazione $s(t,\alpha,\vartheta) = A\cos(\alpha t + \vartheta)$. Ad ogni valore della pulsazione α corrispondono infiniti segnali, che differiscono per la fase e quindi sono disposti a cerchio su un piano. Il luogo complessivo è quindi simile ad un tubo di sezione circolare, che si svolge su una sfera multidimensionale. Lo stimatore può essere basato sulla verosimiglianza

$$\int \exp\left(\frac{2}{N_0}\mathbf{r}\cdot\mathbf{s}(\alpha,\vartheta)\right) f(\vartheta)d\vartheta \tag{8.24}$$

oppure, più semplicemente, sulla correlazione

$$\mathbf{r}\cdot\mathbf{s}(\alpha,\vartheta) \tag{8.25}$$

In ogni caso occorre calcolare la correlazione di $r(t)$ sia con $\cos\alpha t$ sia con $\sin\alpha t$, e quindi un numero doppio di correlazioni rispetto al caso coerente[9].

[8] tuttavia la (8.9) non è applicabile a segnali con fronti infinitamente ripidi; si otterrebbe una dilatazione del luogo infinita e quindi errore nullo, risultato evidentemente non credibile

[9] se le correlazioni sono calcolate con la FFT le componenti in fase e quadratura sono già disponibili

Quanto alle prestazioni dello stimatore, ciò che conta è quanto rapidamente al variare di α il luogo si allontana dal punto trasmesso, intendendo però che la distanza debba essere valutata per tutte le fasi possibili e che si prenda il valore minimo. Dunque si deve valutare in α_0 il minimo del modulo quadro

$$\left|\frac{d\mathbf{s}(\alpha,\vartheta(\alpha))}{d\alpha}\right|^2 \tag{8.26}$$

per un qualsiasi andamento di $\vartheta(\alpha)$ nell'intorno del punto trasmesso. Poiché

$$\frac{d\mathbf{s}(\alpha,\vartheta(\alpha))}{d\alpha} = \frac{\partial\mathbf{s}(\alpha,\vartheta)}{\partial\alpha} + \frac{\partial\mathbf{s}(\alpha,\vartheta)}{\partial\vartheta}\frac{d\vartheta}{d\alpha} \tag{8.27}$$

si riconosce che il quadrato della dilatazione è il minimo rispetto a K (generico valore della derivata $d\vartheta/d\alpha$ in $\alpha=\alpha_0$) di

$$\left|\frac{\partial\mathbf{s}(\alpha,\vartheta)}{\partial\alpha} + K\frac{\partial\mathbf{s}(\alpha,\vartheta)}{\partial\vartheta}\right|^2 \tag{8.28}$$

Nel calcolo è poi lecito porre $\vartheta=0$ poiché il risultato non dipende dalla fase dell'impulso trasmesso.

8.6 Esercizi

8.1 - Si mostri che il valor medio condizionato $E[\alpha/\mathbf{r}]$ è dato dall'espressione (8.5) del testo. *Suggerimento*: basta mostrare che il denominatore normalizza l'area della *ddp* condizionata.

8.2 - Un parametro α ha densità di probabilità $f(\alpha) = \frac{1}{2}\exp(-|\alpha|)$. È disponibile il segnale $r(t) = \alpha + n(t)$ nell'intervallo di tempo da 0 a T. Il rumore $n(t)$ è gaussiano con densità spettrale $N_0/2$. Come si ottiene $\hat{\alpha}$:

- a massima verosimiglianza? e quale è l'errore quadratico medio?
- a massima probabilità a posteriori? (si faccia attenzione ai valori di r prossimi a zero)
- a minimo errore quadratico medio?

8.3 - Si consideri la forma d'onda, dipendente dal parametro α,

$$s(t,\alpha) = A\cos(\omega_0 t + \alpha t) \qquad -T/2 \le t \le T/2$$

Si mostri che

$$\mathbf{s}(\alpha) \cdot \frac{d}{d\alpha}\mathbf{s}(\alpha) = 0$$

Si determini l'angolo ϑ tra i vettori $\mathbf{s}(0)$ e $\mathbf{s}(\alpha)$. Per quale (o quali) α si ha $\vartheta = \pi/2$? Per quale (o quali) α si ha $\vartheta = 3\pi/4$?
Se $\alpha = 2\pi\,\Delta f$ e al segnale $s(t,\alpha)$ è sovrapposto un (debole) rumore gaussiano bianco, quale è la varianza della stima a massima verosimiglianza di Δf?

8.4 - Si determini lo stimatore ML della pulsazione dell'impulso

$$s(t,\alpha) = A\exp(-t^2/2T^2)\cos\alpha t$$

in presenza di rumore gaussiano bianco, e la varianza dell'errore di stima.

8.5 - Si mostri che il calcolo delle correlazioni (8.20) per la stima ML della pulsazione può essere effettuato mediante la trasformata discreta di Fourier (FFT). Come conviene operare nel caso di pulsazione centrale f_0 molto maggiore della pulsazione da stimare? Come conviene scegliere il numero di punti della FFT? *Suggerimento*: inutile cercare una precisione superiore, ad esempio, a metà o un quarto dell'inevitabile scarto quadratico medio dovuto al rumore.

8.6 - Si riceve, sovrapposta a rumore gaussiano bianco, una forma d'onda triangolare, di ampiezza A e durata T,

$$s(t,\tau) = \begin{cases} 2A(t-\tau)/T & \tau \le t \le \tau + T/2 \\ 2A(T+\tau-t)/T & \tau + T/2 \le t \le \tau + T \\ 0 & \text{altrove} \end{cases}$$

dove l'escursione possibile per il parametro è $0 \le \tau \le T$. Quale è lo stimatore a massima verosimiglianza del parametro τ? Quale è l'errore quadratico medio? Quale è la lunghezza del luogo geometrico descritto dal vettore $\mathbf{s}(\tau)$? Quale è la distanza tra gli estremi del luogo? Si risponda poi alle stesse domande supponendo che l'escursione del parametro sia $0 \le \tau \le 2T$.

8.7 - Si considerino le forme d'onda $s(t-\tau)\cos 2\pi f_0 t$ e $s(t-\tau)\cos 2\pi f_0(t-\tau)$, dove $s(t)$ è nota, sovrapposte a rumore gaussiano bianco (debole), ed in entrambi i casi si determini lo stimatore a massima verosimiglianza del parametro τ. Si discutano le differenze per quanto riguarda la struttura dello

stimatore e le prestazioni ottenibili. Si mostri che la maggior precisione del secondo caso è, con ogni probabilità, illusoria.

8.8 - Si determini lo stimatore a massima verosimiglianza della durata T_0 della forma d'onda

$$s(t,\alpha) = \begin{cases} A(T_0 - t) & 0 \le t \le T_0 \\ 0 & \text{altrove} \end{cases}$$

dove A è una costante nota, in presenza di (debole) rumore additivo gaussiano bianco e si calcoli la varianza della stima. Si noti che l'energia della forma d'onda dipende dal parametro.

8.9 - Quale è lo stimatore ML della "durata" T della forma d'onda gaussiana

$$s(t,\tau) = A\exp(-t^2/2T^2)$$

in presenza di rumore bianco? e quale la varianza della stima?

8.10 - Come si può stimare a massima verosimiglianza il parametro α dalla forma d'onda $s(t,\alpha) + n(t)$ dove $n(t)$ è rumore gaussiano bianco e

$$s(t,\alpha) = \begin{cases} A\sin^2\frac{\pi(t-\alpha)}{2T} & \alpha \le t \le \alpha + T \\ A & \alpha + T \le t \le T_0 \\ 0 & \text{altrove} \end{cases}$$

dove $T_0 > \alpha + T$, e quale è l'errore quadratico medio (con rumore debole)? *Suggerimento*: l'energia di $s(t,\alpha)$ dipende da α; si eviti il calcolo dell'energia di $s(t,\alpha)$ nell'intervallo $(\alpha, \alpha + T)$.

8.11 - Si consideri un luogo $\mathbf{s}(\alpha)$ costituito da un tratto circolare di raggio R, con dilatazione uniforme. Siano n_1 ed n_2 le componenti del rumore, rispettivamente tangente e ortogonale, nel piano in cui giace il luogo. Si mostri che l'errore di stima è proporzionale ad $R\arctan\frac{n_1}{R+n_2}$, anziché a n_1 come previsto dall'approssimazione lineare. Si consideri solo il primo termine non lineare dello sviluppo in serie dell'arcotangente, si mostri che è incorrelato con quello lineare, ed infine si determini la varianza dell'errore di stima. Si applichi il risultato alla stima della fase di una sinusoide. *Commento*: non occorre che il luogo sia circolare; basta che possa essere così approssimato localmente.

8.12 - Si consideri lo stimatore ML non coerente per la forma d'onda

$$s(t, \alpha, \vartheta) = A\cos(\alpha t + \vartheta)$$

osservata nell'intervallo (t_0, t_0+T). Si mostri che le prestazioni dello stimatore non dipendono da t_0, e si calcoli la varianza dell'errore di stima. Infine si confronti con il caso coerente.

Capitolo 9

Sincronizzazione

9.1 Introduzione

Finora si sono sempre considerate note la posizione temporale delle forme d'onda ricevute (sincronismo di simbolo) e la fase della portante (sincronismo di fase), oltre all'inizio del blocco codificato nel caso di codici a blocco o l'inizio della n-pla di simboli di canale in codici convoluzionali con *rate* k/n (sincronismi di trama), e così via. In realtà tutte queste informazioni devono essere ottenute, di norma, dallo stesso segnale ricevuto.

Per introdurre il problema della sincronizzazione di simbolo si consideri la semplice trasmissione binaria in banda base, non codificata, su canale ideale:

$$r(t) = \sum a_k g(t - kT - \tau) + n(t) \tag{9.1}$$

dove $\{a_k\}$ è una sequenza di simboli binari ($a_k = \pm 1$) equiprobabili ed indipendenti e τ corrisponde alla posizione temporale, ignota al ricevitore, delle forme d'onda elementari trasmesse. Per molti motivi, quali instabilità degli oscillatori, errori di temporizzazione accumulati in tratte precedenti, ecc., τ varia nel tempo e certamente non è pensabile determinarlo una volta per tutte. La stima del suo valore dovrà essere continuamente verificata ed aggiornata. Per il momento conviene però assumere che τ resti costante, anche se non noto, per un tempo di osservazione sufficiente.

In linea di principio sarebbe possibile affrontare il problema come stima della sequenza $\{a_k\}$ in presenza di un parametro indeterminato τ, oppure anche come stima congiunta della sequenza $\{a_k\}$ e del parametro τ. Tuttavia si preferisce normalmente separare i due aspetti: si procede dapprima a stimare

τ, e si utilizza poi tale stima come se fosse esatta. Tale approccio è molto più semplice, sia da realizzare sia da analizzare[1], e inoltre produce risultati comparabili nella maggior parte dei casi.

9.2 Sincronizzatore di simbolo a massima verosimiglianza

Concentrando l'attenzione sulla stima a massima verosimiglianza del parametro τ e volendo utilizzare la teoria svolta nel Cap. 8, si può considerare questo come un caso di stima in presenza di parametri indeterminati: i dati $\{a_k\}$ ai fini della determinazione del sincronismo di simbolo costituiscono un disturbo casuale. Fissato un intervallo di osservazione pari a N tempi di simbolo $(0, NT)$, la verosimiglianza condizionata ad una determinata sequenza di dati $\{a_k\}$ è

$$\Lambda(\tau, a_k) = \exp\Big[\frac{2}{N_0}\int_0^{NT} r(t)\sum a_k g(t-kT-\tau)dt\Big] \\ \exp\Big[-\frac{1}{N_0}\int_0^{NT}(\sum a_k g(t-kT-\tau))^2 dt\Big] \tag{9.2}$$

Per la maggior parte degli indici k la forma d'onda $g(t-kT-\tau)$ è tutta praticamente compresa nell'intervallo $(0, NT)$ oppure non dà contributo agli integrali. Ciò non vale però per i simboli intorno agli estremi dell'intervallo. Questi effetti di bordo rendono estremamente fastidioso il calcolo, e si preferisce approssimare la verosimiglianza troncando la somma a N termini e lasciando invece liberi gli estremi di integrazione:

$$\Lambda(\tau, a_k) \approx \exp\Big[\frac{2}{N_0}\int r(t)\sum_{k=1}^{N} a_k g(t-kT-\tau)dt\Big] \\ \exp\Big[-\frac{1}{N_0}\int(\sum_{k=1}^{N} a_k g(t-kT-\tau))^2 dt\Big] \tag{9.3}$$

[1] in particolare per i metodi di sincronizzazione molto raffinati resta spesso il dubbio della possibile esistenza di falsi punti di equilibrio

Con ciò la lieve dipendenza da τ dell'energia, presente nella (9.2), scompare. Se poi le forme d'onda $g(t-kT)$ sono ortogonali l'energia non dipende neppure dai dati, e quindi il secondo termine dell'esponenziale diventa inessenziale. La verosimiglianza condizionata può allora essere espressa come

$$\Lambda(\tau, a_k) = \exp(\frac{2}{N_0} \sum_{k=1}^{N} a_k y_k) \tag{9.4}$$

dove si è indicata con y_k la correlazione $\int r(t)g(t-kT-\tau)dt$ del segnale ricevuto $r(t)$ con la forma d'onda $g(t-kT-\tau)$. Si noti che anche se non esplicitamente indicato, per semplificare la notazione, y_k è funzione di τ. Il campione y_k può essere ottenuto dall'uscita $y(t)$ del filtro adattato all'istante $kT+\tau$ (ignorando il solito ritardo t_0 che rende realizzabile il filtro): $y_k = y(kT+\tau)$.

Resta da mediare la verosimiglianza rispetto ai dati binari. Si ottiene facilmente

$$\Lambda(\tau) = E[\exp(\frac{2}{N_0} \sum_{k=1}^{N} a_k y_k)] = \prod_{k=1}^{N} \mathrm{Ch}(\frac{2}{N_0} y_k) \tag{9.5}$$

Ai fini della ricerca del massimo si può considerare il logaritmo della verosimiglianza

$$\log \Lambda(\tau) = \sum_{k=1}^{N} \log \mathrm{Ch}(\frac{2}{N_0} y_k) \tag{9.6}$$

Si osserva subito che è scomodo cercare il massimo ripetendo il calcolo per molti valori di τ. Si dovrebbero infatti prelevare molti campioni per tempo di simbolo dell'uscita del filtro adattato, o addirittura calcolare numerose correlazioni per ogni simbolo, anziché quell'unica che è richiesta per prendere la decisione sul dato a_k. Si preferisce una ricerca iterativa del valore di τ con il metodo del gradiente stocastico, analogamente all'adattamento dei coefficienti di un equalizzatore. La ricerca del massimo viene guidata dalla derivata del logaritmo della verosimiglianza rispetto a τ, accumulata via via ad ogni passo dell'algoritmo. Poiché la derivata del logaritmo del coseno iperbolico è la tangente iperbolica, si ottiene

$$\tau_{k+1} = \tau_k + \alpha \mathrm{Th}(\frac{2}{N_0} y_k) \frac{dy_k}{d\tau} \tag{9.7}$$

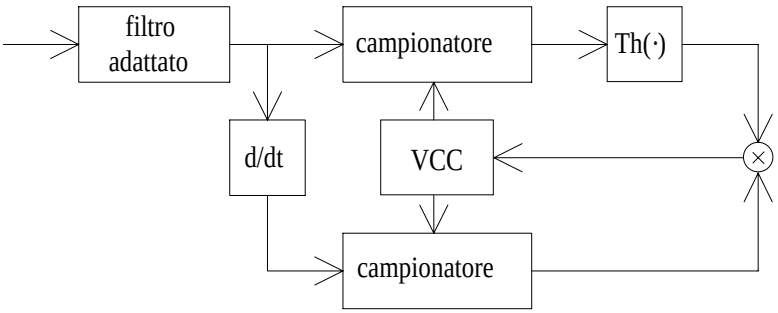

Fig. 9.1 - Sincronizzatore di simbolo a massima verosimiglianza

ad esempio con lo schema di fig. 9.1, dove VCC indica un generatore di *clock* controllato in tensione (*Voltage Controlled Clock*).

È lasciato al lettore verificare[2] che i campioni della derivata di y_k rispetto a τ sono pari ai campioni negli istanti $kT + \tau$ della derivata rispetto al tempo dell'uscita del filtro adattato.

Più che la tangente iperbolica, disturba in questo schema il dover conoscere il valore di N_0. Si può osservare che per piccoli valori dell'argomento la funzione è lineare, e il fattore $2/N_0$ può essere assorbito nella costante α, influenzando solo il guadagno dell'anello di regolazione; per grandi valori dell'argomento la tangente iperbolica è approssimabile con il segno dell'argomento, che a sua volta non è altro che la decisione binaria sul dato a_k. Ciò è in accordo con quanto si ottiene se si ripete il calcolo supponendo i dati a_k noti, anziché variabili casuali, e se si osserva che ad alto rapporto segnale-rumore i dati decisi $\hat{a}_k$ sono quasi sempre uguali ai dati trasmessi a_k (es. 9.1).

Lo schema di fig. 9.1 si presta bene ad una realizzazione analogica, in cui l'uscita del filtro adattato è continua nel tempo e quindi derivabile, ma non ad un ricevitore numerico. Infatti il correlatore fornisce un numero, e non una funzione del tempo. Il calcolo del campione della derivata richiederebbe un secondo correlatore numerico, che correlasse $r(t)$ con $g'(t - kT - \tau)$, con un raddoppio quindi della complessità del ricevitore.

Per alto rapporto segnale-rumore è possibile valutare il numero N di contributi da tenere in conto per avere una buona stima di τ, supponendo i dati a_k perfettamente noti (approssimazione favorevole, ma non lontana dal vero

[2] l'unico pericolo, modesto, è di sbagliare qualche segno

nel funzionamento normale a bassa probabilità d'errore, una volta acquisito il sincronismo) e quindi utilizzando i risultati del capitolo precedente ed in particolare la (8.9). Si ottiene[3]

$$\sigma^2_{\hat{\tau}} = \frac{N_0/2}{\int (\sum a_k g'(t-kT-\tau))^2 dt} \approx \frac{N_0/2}{N \int g'^2(t) dt} \tag{9.8}$$

Valori tipici di N richiesti in pratica sono dell'ordine delle decine o centinaia.

Naturalmente l'approssimazione perde ogni valore nel caso, che occorre evitare, in cui la sorgente trasmetta una lunga sequenza di simboli uguali. Infatti il segnale diventa una costante, da cui non si può pretendere di estrarre alcuna informazione sulla temporizzazione. Favorevole è invece il caso di simboli positivi e negativi alternati, sequenza solitamente trasmessa durante il preambolo nella trasmissione a pacchetti per favorire la sincronizzazione.

9.3 Quadratore e filtro

Fra le strutture analogiche merita di essere segnalata, perché molto diffusa, quella costituita da un quadratore[4] seguito da un filtro passa banda alla frequenza di cifra. Nel caso di dati binari indipendenti è facile vedere che il quadrato del segnale ricevuto ha un valor medio (cioè una componente deterministica) periodico, pari a

$$\begin{aligned} &E[a_k^2] \sum g(t-kT-\tau)^2 = \sum g(t-kT-\tau)^2 = \\ &= c_0 + 2c_1 \cos 2\pi (t-\tau)/T \end{aligned} \tag{9.9}$$

dove il coefficiente c_1 dello sviluppo in serie di Fourier è dato da

$$c_1 = \frac{1}{T} G(\frac{1}{T}) * G(\frac{1}{T}) \tag{9.10}$$

ed è diverso da zero in tutti i casi pratici; è infatti nullo solo nel caso di eccesso di banda nullo. Gli istanti di massimo[5] della sinusoide alla frequenza $1/T$ sono

[3] anche assumendo ortogonali le forme d'onda $g(t-kT-\tau)$, le derivate rispetto a τ non lo sono; si può però invocare l'incorrelazione tra i dati per concludere, perlomeno se N è grande, che l'energia della somma non si discosta molto dalla somma delle energie (es. 9.2)

[4] o altro dispositivo non lineare di più facile realizzazione pratica; l'analisi è però assai più difficile

[5] ovvero i passaggi per lo zero, con pendenza negativa, della derivata

gli istanti di campionamento desiderati. La sinusoide può essere estratta con un filtro centrato sulla frequenza di cifra e a banda stretta, per limitare il disturbo dovuto alle componenti casuali del quadrato del segnale ricevuto.

La realizzazione circuitale di tale filtro non è però priva di problemi. Ad esempio un filtro analogico passa banda introduce un errore di fase proporzionale all'errore nella frequenza centrale (*dissintonia*), e quindi dipendente da temperatura, vibrazioni meccaniche, invecchiamento dei componenti, ecc.

Il problema della dissintonia viene eliminato dai circuiti ad aggancio di fase (PLL: *Phase Locked Loop*). Questi introducono tuttavia il fastidioso fenomeno dell'*hang-up*: i transitori di aggancio possono essere molto lenti se quando si attiva il circuito l'errore di fase è prossimo a π. Ciò non ha alcuna importanza nella trasmissione continua, mentre nella trasmissione a pacchetti può far perdere un intero blocco di dati.

Nel caso di trasmissione in banda passante, una volta determinate le componenti in fase e quadratura cioè parte reale e immaginaria dell'equivalente passa basso, si ottengono segnali utili per la sincronizzazione dal quadrato di entrambe. Conviene evidentemente sommare i due termini. Detto $z(t) = x(t) + jy(t)$ l'equivalente passa basso del segnale ricevuto, nel caso del quadratore si ha $x^2(t) + y^2(t) = |z(t)|^2$; ciò mostra che il risultato è indipendente dalla fase della portante. Il sincronizzatore di simbolo è quindi in grado di agire anche prima che sia stato acquisito il sincronismo di fase.

Il sincronizzatore a massima verosimiglianza descritto in precedenza richiede invece che la portante sia già acquisita. Si può quindi temere una sfavorevole interazione tra i due sincronizzatori, se nessuno dei due è in grado di agire indipendentemente dall'altro[6]. Naturalmente se il sincronizzatore di simbolo richiede la conoscenza della fase della portante si può anche prevedere un sincronizzatore di portante che non richieda a sua volta la preventiva acquisizione del sincronismo di simbolo. Si otterrà prima il sincronismo di portante, e solo successivamente quello di simbolo.

[6] se non si hanno vincoli stringenti sul tempo di acquisizione ci si può affidare al semplice fatto che l'uno o l'altro dei sincronizzatori, mossi da segnali casuali, prima o poi si avvicina al sincronismo corretto; questo diventa il momento buono per agire anche per l'altro sincronizzatore

9.4 Strutture numeriche per la sincronizzazione di simbolo

Semplici strutture per la sincronizzazione di simbolo possono essere suggerite dai sincronizzatori analogici. Quello con quadratore e filtro passa banda mal si presta ad essere tradotto in una semplice versione numerica. Invece qualcosa di simile al sincronizzatore a massima verosimiglianza può essere ottenuto utilizzando il segnale $y(kT+\tau+T/2)-y(kT+\tau-T/2)$, che a meno di un fattore moltiplicativo è una discreta approssimazione della derivata nell'istante $kT+\tau$. Se è già prevista l'elaborazione a due campioni per tempo di simbolo, ad esempio perché si vuole utilizzare un equalizzatore a prese frazionarie spaziate di $T/2$, la derivata è ottenuta quasi senza costo[7].

La (9.7) suggerisce quindi, a basso ed alto rapporto segnale-rumore rispettivamente, l'aggiornamento

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha y_k (y_{k+1/2} - y_{k-1/2}) \tag{9.11}$$

oppure

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha \hat{a}_k (y_{k+1/2} - y_{k-1/2}) \qquad \hat{a}_k = \mathrm{sgn}(y_k) \tag{9.12}$$

dove $y_{k+1/2}$ indica $y(kT+T/2+\hat{\tau}_k)$, in modo un po' sbrigativo ma efficace. Se il termine con indice $k+1/2$ viene applicato con il ritardo di un tempo di simbolo, cioè se nell'aggiornamento di $\hat{\tau}_k$ viene sostituito con quello di indice $k-1/2$, cosa senza apprezzabili conseguenze data la lunga costante di tempo dell'anello di controllo, si ottengono rispettivamente il sincronizzatore di *Gardner*

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha y_{k-1} y_{k-1/2} - \alpha y_k y_{k-1/2} = \hat{\tau}_k + \alpha (y_{k-1} - y_k) y_{k-1/2} \tag{9.13}$$

e il cosiddetto *Data Transition Loop* (DTL)

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha \hat{a}_{k-1} y_{k-1/2} - \alpha \hat{a}_k y_{k-1/2} = \hat{\tau}_k + \alpha (\hat{a}_{k-1} - \hat{a}_k) y_{k-1/2} \tag{9.14}$$

Il DTL aggiorna il sincronismo di simbolo solo quando due dati successivi sono opposti. Intuitivamente è basato sul fatto che, con i dati opposti, se il sincronismo è corretto il valor medio di $y_{k-1/2}$ è nullo; altrimenti, almeno per piccoli errori, è proporzionale all'errore di temporizzazione ed alla differenza fra i dati, come si vede dal diagramma ad occhio di fig. 9.2.

[7] nel caso a prese intere $y(kT+\tau+T)-y(kT+\tau-T)$ può sembrare una approssimazione troppo brutale; tuttavia essa è ancora utilizzabile, come si vedrà

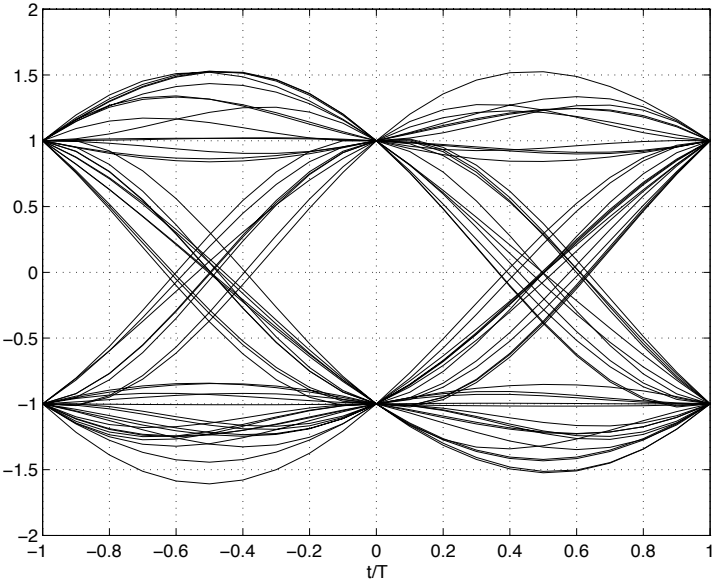

Fig. 9.2 - Diagramma ad occhio (*roll-off*=40%)

Nel caso complesso si possono sommare i contributi ottenuti dai flussi di dati in fase e in quadratura. È facile verificare ad esempio che il sincronizzatore di Gardner diventa

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha \mathrm{Re}\{(y_{k-1} - y_k) y^*_{k-1/2}\} \tag{9.15}$$

ed ha quindi il merito di essere indipendente dalla fase della portante, mentre il DTL non ha tale proprietà.

Molto interessanti sono infine i sincronizzatori di simbolo che richiedono un solo campione per tempo di simbolo, cioè lo stesso campione che è già richiesto per la decisione. L'approssimazione *a prese intere* della derivata dà il sincronizzatore di *Mueller e Müller*[8]

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha \hat{a}_k (y_{k+1} - y_{k-1}) \tag{9.16}$$

o anche, con la solita applicazione ritardata del primo termine,

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha (y_k \hat{a}_{k-1} - y_{k-1} \hat{a}_k) \tag{9.17}$$

[8]in questo caso la versione per basso rapporto segnale-rumore non funziona (es. 9.5)

Anche questo sincronizzatore ha una semplice interpretazione. Sia infatti $h(t) = g(t) * g(t)$ la risposta all'impulso elementare a valle del filtro adattato, che supponiamo di Nyquist: $h(iT) = 0$ $(i \neq 0)$. In presenza di un errore di temporizzazione $\Delta\tau$ si ha, includendo anche il rumore,

$$\begin{aligned} y_k &= y(kT + \tau + \Delta\tau) = \sum a_{k-i} h(iT + \Delta\tau) + n_k \approx \\ &\approx \Delta\tau \sum a_{k-i} h'(iT) + n_k \end{aligned} \tag{9.18}$$

e il valor medio di $y_k a_{k-1}$ è pari a $\Delta\tau h'(T)$. Analogo risultato si ottiene da $y_{k-1} a_k$. Dunque se i dati decisi sono corretti $(\hat{a}_k = a_k)$ il segnale di controllo ha valor medio $2\Delta\tau h'(T)$, e non è difficile calcolarne varianza e densità spettrale di potenza (es. 9.6).

Nel caso complesso sommando i contributi in fase e quadratura si ottiene il sincronizzatore, dipendente dalla fase della portante,

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha \mathrm{Re}\{y_k \hat{d}^*_{k-1} - y_{k-1} \hat{d}^*_k\} \tag{9.19}$$

Non è poi difficile mostrare che lo stesso sincronizzatore è efficace anche nel caso di modulazione 8PSK.

Un sincronizzatore che richiede un solo campione per tempo di simbolo, ed è invece indipendente dalla fase della portante, è

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha \mathrm{Re}\{y_k y^*_{k-1}\} \left(\frac{1}{|y_{k-1}|} - \frac{1}{|y_k|} \right) \tag{9.20}$$

9.5 Realizzazione numerica del sincronizzatore

È importante ricordare che in un ricevitore numerico, ad esempio con un solo campione y_k per simbolo, le correlazioni vengono calcolate discretizzando l'integrale con un passo $t_0 = T/n_c$ opportuno, come discusso nel Cap. 1:

$$\begin{aligned} &\int r(t) g(t - kT - \hat{\tau}) dt = \int r(t + kT) g(t - \hat{\tau}) dt = \\ &= t_0 \sum r(mt_0 + kT) g(mt_0 - \hat{\tau}) \end{aligned} \tag{9.21}$$

Per variare $\hat{\tau}$ basta quindi semplicemente usare un diverso insieme di coefficienti nel calcolo della correlazione discreta[9].

Anche per $\hat{\tau}$ sarà previsto un insieme discreto, sufficientemente fitto, di valori $\hat{\tau} = nT/n_\tau$. Una precisione infinita sarebbe illusoria, perché il valore stimato è comunque rumoroso. Se n_τ è multiplo di n_c, si deve calcolare

$$\sum r(mt_0 + kT)g(mt_0 - \hat{\tau}) = \sum r_{m+kn_c}\ g_{m(n_\tau/n_c)-n} \tag{9.22}$$

dove la spaziatura tra i campioni di $r(t)$ è T/n_c e quella tra i campioni di $g(t)$ è T/n_τ.

Si vede quindi che basta avere memorizzati i campioni di $g(t)$ con passo T/n_τ e selezionare dalla memoria, all'arrivo di ogni campione di $r(t)$, il coefficiente che corrisponde all'indice $m(n_\tau/n_c) - n$, funzione sia del tempo corrente sia del valore attuale di $\hat{\tau}$, rappresentato da n.

Può essere interessante svolgere qualche semplice considerazione sugli anelli di controllo numerici[10].

L'analisi generale è assai complessa, e richiede il supporto di simulazioni. Si cerca di individuare nel segnale di controllo due componenti, una deterministica (il valor medio) ed una casuale. Uno strumento assai utile per una prima valutazione è il grafico, in funzione dell'errore di temporizzazione, del valor medio; tale grafico, detto *curva ad S* a causa della forma tipica, è eventualmente accompagnato dal grafico della varianza. Un esempio è mostrato in fig. 9.3.

La curva ad S mostra quali sono le zone in cui si potrà utilizzare un'approssimazione lineare, e soprattutto dove le correzioni saranno quasi del tutto casuali[11].

A regime, con errore di temporizzazione ormai piccolo, l'anello è governato dall'equazione linearizzata

$$\hat{\tau}_{k+1} = \hat{\tau}_k + \alpha(-C(\hat{\tau}_k - \tau) + n_k) \tag{9.23}$$

[9] in linea di principio il *clock* usato per campionare il segnale ricevuto potrebbe non avere alcuna relazione con il tempo di simbolo T; tuttavia ritmi diversi tra dati trasmessi e dati demodulati richiedono un *buffer* che, alla lunga, si riempie o si svuota; quindi occorre che il segnale di sincronizzazione agisca, perlomeno a lungo termine, sulla frequenza con cui i dati escono dal ricevitore. In una rete di telecomunicazioni il problema di quali *clock* siano *master* e quali *slave* è assai complesso

[10] considerazioni analoghe valgono per i sincronizzatori di portante discussi in seguito

[11] come già accennato è un bene che vi siano vigorose fluttuazioni casuali, perché prima o poi l'anello si porta in una zona in cui può funzionare

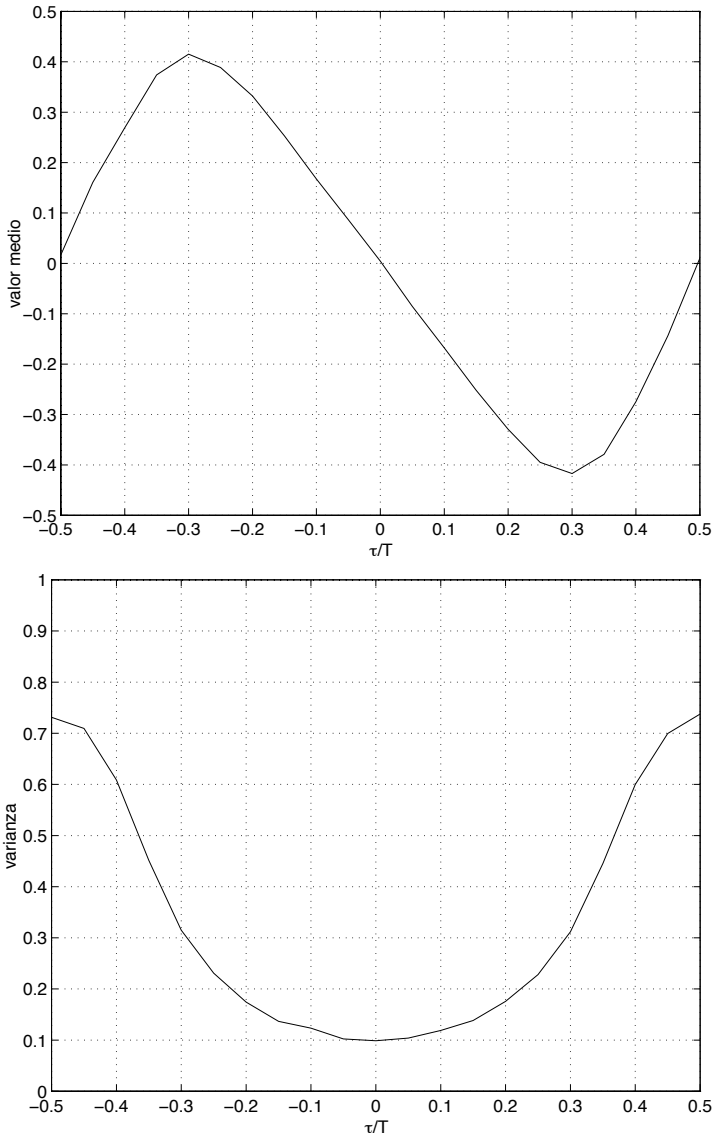

Fig. 9.3 - Valor medio e varianza del segnale di controllo $y_k \hat{a}_{k-1} - y_{k-1} \hat{a}_k$ della (9.17), con E_b/N_0=10 dB e *roll-off*=40%

dove C è il modulo della pendenza nell'origine della curva ad S e n_k è il termine di disturbo casuale (non bianco, di norma). In sostanza il segnale retroazionato è proporzionale ad una stima (rumorosa) dell'errore.

Il numero N di campioni necessari per avere una buona sincronizzazione non è mai piccolo, e quindi l'evoluzione temporale di τ_k è lenta. Ciò consente di ottenere risultati validi sia con la teoria dei sistemi di controllo a tempo discreto (trasformata zeta) sia a tempo continuo (trasformata di Laplace). L'autore di queste note ha incontrato prima il tempo continuo, e vi è ancora affezionato, per cui l'analisi con Laplace avrà la precedenza. È senz'altro conveniente normalizzare i tempi all'intervallo di simbolo T, cioè porre $T = 1$. Con questa scelta molte grandezze risultano adimensionali (ad esempio $\tau = 0.1$ corrisponde all'errore di un decimo di tempo di simbolo) ed anche le frequenze sono normalizzate alla frequenza di cifra $1/T = 1$.

Portando a primo membro $\hat{\tau}_k$ e considerando uguali la differenza $\hat{\tau}_{k+1} - \hat{\tau}_k$ e la derivata $d\hat{\tau}/dt$ si ottiene l'equazione a tempo continuo

$$\frac{d\hat{\tau}}{dt} = -\alpha C\hat{\tau} + \alpha C\tau + \alpha n(t) \tag{9.24}$$

Questo è un anello del primo ordine, con guadagno ad anello aperto $\alpha C/s$. La pulsazione di taglio è $\omega_0 = \alpha C$. Per quanto riguarda l'effetto su $\hat{\tau}$ del rumore $n(t)$, la funzione di trasferimento ad anello chiuso è

$$H_n(s) = \frac{1}{C}\frac{1}{1 + s/\alpha C} \tag{9.25}$$

e la banda unilatera di rumore, normalizzata alla frequenza di cifra, è

$$B_L = \int_0^\infty \frac{1}{1 + 4\pi^2 f^2/\alpha^2 C^2} df = \frac{\alpha C}{4} \tag{9.26}$$

Poiché la banda è normalmente molto minore della frequenza di cifra, per valutare la varianza dell'errore di temporizzazione basta moltiplicare la densità spettrale di potenza di $n(t)$ valutata a frequenza zero per il guadagno in potenza $1/C^2$ a frequenza zero e per la banda di rumore. Nel caso particolare di campioni n_k incorrelati, con varianza σ_n^2, la densità spettrale bilatera è $S(0) = \sigma_n^2$; basta infatti ricordare che l'integrale nella banda $(-1/2, 1/2)$ deve fornire la varianza σ_n^2. Quindi la varianza a regime di $\hat{\tau}_k$ è

$$\sigma_{\hat{\tau}_k}^2 = 2\sigma_n^2 \frac{1}{C^2}\frac{\alpha C}{4} = \frac{\alpha\sigma_n^2}{2C} = \frac{2B_L\sigma_n^2}{C^2} \tag{9.27}$$

L'anello ha una costante di tempo

$$t_0 = \frac{1}{\omega_0} = \frac{1}{\alpha C} = \frac{4}{B_L} \tag{9.28}$$

e i transitori sono esponenziali, con tale costante di tempo. La risposta ad un impulso di rumore è l'esponenziale $\alpha \exp(-k/t_0) = \alpha \exp(-k\alpha C)$. Questo consente una semplice verifica della varianza del *jitter* nel caso di rumore incorrelato, sommando le varianze dei vari contributi:

$$\sigma^2_{\tau_k} = \sum_{k=0}^{\infty} \alpha^2 \exp(-2k\alpha C)\sigma_n^2 = \frac{\alpha^2\sigma_n^2}{1-\exp(-2\alpha C)} \approx \frac{\alpha\sigma_n^2}{2C} \tag{9.29}$$

L'approssimazione finale è valida se $\alpha C \ll 1$, e dà un risultato in accordo con quello già trovato.

Se si vuol valutare la capacità dell'anello di inseguire le variazioni di τ, che come già detto non è in realtà costante, occorre la corrispondente funzione di trasferimento ad anello chiuso che, come si vede dalla (9.24), è pari alla (9.25) moltiplicata per C:

$$H_\tau(s) = \frac{1}{1 + s/\alpha C} \tag{9.30}$$

L'anello quindi insegue bene fluttuazioni fino alla pulsazione ω_0 e sempre più a fatica quelle più rapide.

Se si esamina la risposta impulsiva dell'anello si vede che il valore attuale di $\hat{\tau}_k$ anziché essere la media corrente degli ultimi N aggiornamenti, con pesi uguali, è una media pesata esponenzialmente di tutti i contributi precedenti. Tenendo conto che in realtà τ varia nel tempo, il fatto che i campioni più recenti abbiano peso maggiore è da vedere con favore.

Se infine si vuole un anello del secondo ordine, per poter assorbire non solo un errore di temporizzazione costante ma anche uno variabile linearmente nel tempo, basta introdurre un altro integratore ideale e, per avere un buon margine di stabilità, uno zero intorno alla pulsazione $b = 0.25 \div 0.5\omega_0$:

$$H(s) = \frac{1}{s}(s+b) = 1 + \frac{b}{s} \tag{9.31}$$

Per la realizzazione dell'integratore b/s si vede facilmente che occorre il filtro discreto, realizzabile con un moltiplicatore-accumulatore,

$$u_k = u_{k-1} + bi_k \tag{9.32}$$

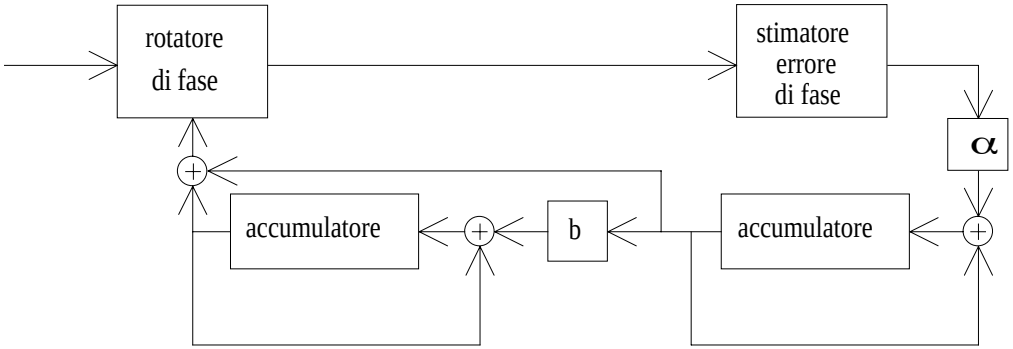

Fig. 9.4 - Struttura numerica del sincronizzatore di simbolo

dove i_k è l'ingresso e u_k è l'uscita. Quindi il sincronizzare numerico ha la struttura in fig. 9.4. Chiaramente è conveniente che α e b siano potenze (negative) di 2.

Vediamo ora come si ottengono gli stessi risultati con la trasformata zeta. L'equazione a tempo discreto (9.23) corrisponde alla funzione di trasferimento tra n_k e $\hat{\tau}_k$

$$H_n(z) = \frac{\alpha}{z - 1 + \alpha C} = \frac{\alpha z^{-1}}{1 - (1 - \alpha C) z^{-1}} \tag{9.33}$$

che mette in evidenza che l'anello introduce intrinsecamente un ritardo di un tempo di simbolo tra la misura dell'errore all'istante kT e l'applicazione della correzione all'istante $(k+1)T$. È poi immediato ottenere la risposta impulsiva, esponenziale,

$$h_n(k) = \alpha(1 - \alpha C)^{k-1} \qquad k = 1, 2, \ldots \tag{9.34}$$

Nel caso particolare di campioni n_k incorrelati si ottiene facilmente la varianza del *jitter*

$$\sigma_\tau^2 = \sum_{k=1}^{\infty} \alpha^2 (1 - \alpha C)^{2(k-1)} \sigma_n^2 = \frac{\alpha^2 \sigma_n^2}{1 - (1 - \alpha C)^2} = \frac{\alpha \sigma_n^2}{C(2 - \alpha C)} \tag{9.35}$$

Questa mostra che la stabilità impone che sia $\alpha C < 2$; tuttavia di norma $\alpha C \ll 1$ per avere un *jitter* tollerabile, e quindi (9.29) e (9.35) coincidono.

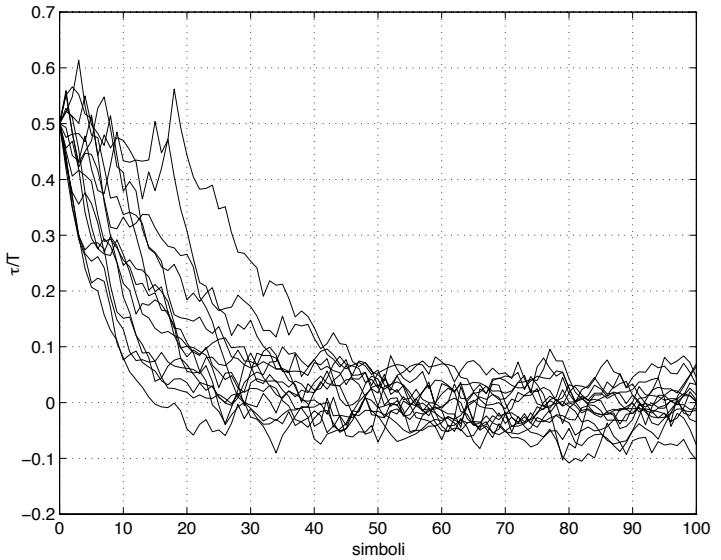

Fig. 9.5 - Errore di temporizzazione (E_b/N_0=10 dB; *roll-off*=40%; $\alpha = 0.05$)

Per valutare la capacità dell'anello di inseguire variazioni di τ si ottiene con semplici calcoli la funzione di trasferimento ad anello chiuso

$$H(z) = \frac{\alpha C z^{-1}}{1 - (1 - \alpha C) z^{-1}} \tag{9.36}$$

che evidenzia ancora una volta il ritardo di un tempo di simbolo. Infine si può anche includere nell'anello un secondo integratore, come nelle (9.31) e (9.32).

Le due trattazioni non sembrano avere molto in comune, a prima vista; ma concordano non appena si usa l'approssimazione $z^{-1} = \exp(-s) \approx 1 - s$, valida alle basse frequenze, le uniche che interessano se $\alpha C \ll 1$.

La fig. 9.5 mostra tipici andamenti dell'errore di temporizzazione, partendo da un errore iniziale di mezzo tempo di simbolo. Si osservino le possibili esitazioni iniziali. In metà dei casi, non mostrati in figura, il sincronizzatore si aggancia sul simbolo adiacente ($\tau/T = 1$), cosa del tutto equivalente.

Per concludere accenniamo solo di sfuggita al fatto che anche a regime a seguito di una combinazione sfavorevole di disturbi di tanto in tanto l'anello perde o guadagna un ciclo, cioè un intervallo di simbolo. Tale fenomeno è detto *cycle slip*.

9.6 Sincronizzazione di portante

Anche per la sincronizzazione di portante sono disponibili molti metodi, sia analogici sia numerici. Si preferirà concentrarsi su alcune tecniche numeriche. Su quelle analogiche, benché ancora abbastanza diffuse, si farà solo qualche cenno rimandando per approfondimenti alla letteratura, assai ampia.

Si consideri dapprima una modulazione 2PSK:

$$r(t) = \mathrm{Re}\{a(t)\exp(j\vartheta)\exp(j2\pi f_0 t)\} + n(t) \tag{9.37}$$

dove ϑ, che per il momento si suppone costante[12], è la fase della portante ricevuta rispetto al riferimento usato per la demodulazione e $a(t)$ è il segnale numerico modulante in banda base. Si vede che il valor medio del quadrato del segnale ricevuto contiene una componente $\frac{1}{2}E[a^2(t)]\cos(2\pi f_0 t + 2\vartheta)$ da cui si può estrarre una sinusoide alla frequenza $2f_0$ con un filtro passa banda (o un PLL). Infine con un divisore di frequenza si ottiene $\cos(2\pi f_0 t + \vartheta)$, oppure $\cos(2\pi f_0 t + \vartheta + \pi)$, cioè la portante desiderata a meno di un'inevitabile ambiguità di π.

Per il caso, più frequente, di modulazioni in fase e quadratura come 4PSK e QAM considerando $a(t)$ come somma di una parte reale ed una immaginaria indipendenti non è difficile verificare che con una quarta potenza si ottiene una sinusoide a frequenza $4f_0$. Un circuito divisore di frequenza per quattro dà la portante, a meno di un'ambiguità di $\pi/2$ o multipli.

Infine si può verificare che la modulazione 8PSK richiederebbe l'ottava potenza. Il contributo del rumore è via via più forte, ed anche più complicato da calcolare, all'aumentare dell'ordine della potenza.

Strutture analoghe, di cui si cita solo l'esistenza, sono i *Costas loop* e i cosiddetti *rimodulatori*, per il quali si rimanda alla letteratura.

9.7 Strutture numeriche per la sincronizzazione di portante

Si prenderanno in considerazione solo modulazioni lineari e sincronizzatori che richiedano un campione per simbolo, e si considererà noto il sincronismo di

[12] uno dei motivi principali per cui ϑ non è costante è il rumore di fase degli oscillatori in trasmissione e ricezione, che il sincronizzatore deve inseguire e compensare

simbolo[13].

Un'osservazione importante è che non occorre portare effettivamente la fase della portante del demodulatore a coincidere con ϑ agendo su un qualche VCO. Si ottiene lo stesso risultato, in banda base, moltiplicando per $\exp(-j\vartheta)$ il campione complesso y_k ottenuto dalla correlazione eseguita con assi ruotati. Ciò vale non solo se ϑ è costante ma anche se varia, purché non troppo rapidamente. Si pensi ad esempio a ϑ variabile linearmente nel tempo a causa di un errore fisso di frequenza tra i due oscillatori di trasmissione e ricezione (es. 9.16). Dunque l'oscillatore di ricezione può, in linea di principio, essere libero e il recupero della portante avvenire con una sola moltiplicazione complessa per simbolo, in banda base[14]. A questo scopo, il sincronizzatore deve solo fornire una stima aggiornata di ϑ.

In molti casi si può supporre che i dati (complessi) d_k siano noti; naturalmente nel funzionamento normale sono note le corrispondenti decisioni, e solo quelle passate[15]. Il logaritmo della verosimiglianza è dato da

$$\begin{aligned} \log \Lambda(\vartheta) &= \mathrm{Re}\{\int z(t)\big(\sum d_k g(t-kT)\exp(j\vartheta)\big)^* dt\} = \\ &= \mathrm{Re}\{\exp(-j\vartheta)\sum y_k d_k^*\} \end{aligned} \tag{9.38}$$

dove $z(t)$ è l'equivalente passa basso di $r(t)$ e $y_k = \int z(t)g^*(t-kT)dt$ è il campione complesso in banda base, che verrà moltiplicato per $\exp(-j\hat{\vartheta})$ per ottenere il campione corretto z_k.

Si utilizzi la solita ricerca iterativa del massimo guidata dalla derivata rispetto a ϑ. Poiché

$$\frac{d\log\Lambda(\vartheta)}{d\vartheta} = \mathrm{Im}\{\exp(-j\vartheta)\sum y_k d_k^*\} = \mathrm{Im}\{\sum z_k d_k^*\} \tag{9.39}$$

[13] ci si potrà poi eventualmente chiedere quanto degradano le prestazioni con un errore di temporizzazione; molti metodi funzionano ugualmente, sia pure con *jitter* di fase maggiore, che si ridurrà una volta acquisiti i sincronismi

[14] in realtà le fluttuazioni rapide e casuali di ϑ, che non tollerano il lungo ritardo di un anello che includa VCO e filtri in banda base (o correlatori), vengono corrette in questo modo; le fluttuazioni più prevedibili dovute ad un errore fisso di frequenza sono normalmente compensate agendo sul VCO

[15] nella trasmissione a pacchetti i dati sono preceduti da un *preambolo*, modulato da una sequenza *nota*

l'anello di controllo, applicando le correzioni campione per campione e utilizzando i dati decisi sarà

$$\hat{\vartheta}_{k+1} = \hat{\vartheta}_k + \alpha \mathrm{Im}\{z_k \hat{d}_k^*\} \tag{9.40}$$

Ad esempio, scomponendo z_k in parte reale e immaginaria ($z_k = u_k + jv_k$), nei casi 2PSK e 4PSK si ottiene rispettivamente

$$\hat{\vartheta}_{k+1} = \hat{\vartheta}_k + \alpha v_k \mathrm{sgn}(u_k) \tag{9.41}$$

$$\hat{\vartheta}_{k+1} = \hat{\vartheta}_k + \alpha (v_k \mathrm{sgn}(u_k) - u_k \mathrm{sgn}(v_k)) \tag{9.42}$$

dove $\mathrm{sgn}(u_k)$ e $\mathrm{sgn}(v_k)$ non sono altro che il dato deciso in fase e in quadratura.

Se non si volesse una ricerca iterativa, si può notare che fissato il numero N di termini precedenti da includere nella (9.38) il valore di ϑ che fornisce il massimo ha la semplice espressione

$$\hat{\vartheta} = \arg\{\sum_{i=1}^{N} z_{k-i} \hat{d}_{k-i}^*\} \tag{9.43}$$

dove ancora si sono usati i dati decisi. Dunque basta accumulare N parti reali ed immaginarie e calcolare un'arcotangente a quattro quadranti.

Quanto al valore di N richiesto, se il rumore è debole e si possono ritenere noti i dati ci si può affidare alla teoria del Cap. 8, da cui si ottiene facilmente (es. 9.13,9.14)

$$\sigma_{\hat{\vartheta}}^2 = \frac{N_0/2}{N E_s} \tag{9.44}$$

dove E_s è l'energia per simbolo. Tipicamente N va da qualche decina a qualche centinaia, secondo il tipo di modulazione.

L'analisi delle prestazioni dell'anello di recupero della portante è assai complessa. L'errore iniziale di fase è quasi sempre accompagnato da un errore di frequenza (rotazione della costellazione), che l'anello recupera solo dopo aver perso un gran numero di cicli[16].

L'analisi è invece relativamente semplice intorno al punto di equilibrio, e segue le stesse linee indicate per i sincronizzatori numerici di simbolo. Anche nella sincronizzazione di portante di tanto in tanto l'anello perde o guadagna un ciclo (ad esempio un quadrante). Tale fenomeno è detto *cycle slip*.

[16] è evidente che per avere errore di fase mediamente nullo a regime occorre un anello del secondo ordine

9.8 Metodo di Viterbi e Viterbi

Nel caso di modulazione M-PSK un metodo molto interessante è quello proposto da Viterbi e Viterbi. Data una N-pla di campioni y_k, il metodo consiste nei seguenti passi. Si rimuove la modulazione di fase per mezzo di una nonlinearità che moltiplica la fase per M

$$w_k = F(|y_k|)\exp(jM\arg\{y_k\}) \tag{9.45}$$

dove tipicamente $F(\rho) = \rho^K$ ($K = 0$ o 1, di solito). Lo scopo è quello di allineare i vettori w_k, tutti con fase $M\vartheta$ (ciò si otterrebbe esattamente in assenza di rumore). Poi si sommano gli N vettori w_k, allo scopo di migliorare di N volte il rapporto segnale-rumore. Infine, ottenuto un buon rapporto segnale-rumore, si valuta l'argomento della somma e lo si divide per M ottenendo la stima di ϑ, con la solita ambiguità ineliminabile di multipli di $2\pi/M$. Le prestazioni dello stimatore sono abbastanza vicine a quanto previsto dalla (9.44) e i valori pratici di N risultano intorno a 20-25 e 80-100 per la modulazione 4PSK e 8PSK, rispettivamente.

Si noti che non essendovi alcun riferimento ai dati decisi gli N campioni possono essere presi in posizione qualsiasi, anche tra quelli *futuri*. La scelta più conveniente è a cavallo dell'istante attuale, di modo che un piccolo errore di frequenza non abbia effetto: peggiora lievemente l'allineamento dei vettori, e quindi il rapporto segnale-rumore, ma non cambia il valor medio dell'argomento della somma. Non è difficile verificare che il massimo errore di frequenza tollerabile è dell'ordine di $1/2MNT$. Se il ritmo di trasmissione $1/T$ è elevato non ci sono problemi, mentre i casi difficili sono quelli a bassa velocità[17].

9.9 Considerazioni finali

L'argomento della sincronizzazione è assai vasto, o non è facile orientarvisi. Una trattazione un po' più completa avrebbe richiesto di considerare anche le modulazioni multilivello (PAM, QAM), l'effetto delle distorsioni del canale, le modulazioni non lineari (CPM), i sincronizzatori congiunti di simbolo e portante, i metodi per la stima di errori di frequenza, la ricerca dei sincronismi di trama, ecc. Si è voluto dare solo qualche idea di base.

[17] ovviamente la precisione assoluta degli oscillatori non dipende dal ritmo di trasmissione, ma solo dalla loro qualità e dalla frequenza della portante

Per concludere merita di essere segnalato il fatto che normalmente nei sistemi codificati si progettano i sincronizzatori come se il codice non ci fosse, con risultati soddisfacenti perché è stato mostrato che i buoni codici non modificano le statistiche rilevanti ai fini della sincronizzazione.

Si deve però osservare che un sistema codificato opera ad un valore di E_b/N_0 molto più basso, e spesso con E_s/N_0 ancora minore. Gli effetti del rumore diventano più difficili da valutare, e non tutti gli algoritmi di sincronizzazione risultano utilizzabili. Basti pensare che a causa del codice le decisioni sono disponibili con forte ritardo. Benché si possano prendere decisioni preliminari ignorando il codice, da utilizzare ai soli fini della sincronizzazione, queste sono assai poco affidabili a causa del basso rapporto segnale-rumore, e i sincronizzatori ne soffrono non poco. Unica consolazione è che i sistemi codificati, oltre a sopportare una maggior quantità di rumore, sono anche più tolleranti in fatto di errori di temporizzazione e di fase.

Un'ultima osservazione: nella decodifica di un codice a traliccio ogni percorso sopravvissuto corrisponde ad una diversa sequenza di dati. Se non si dovesse trovare un sincronizzatore di fase soddisfacente, si può pensare di utilizzarne uno diverso per ogni stato. Ciascuno di questi si basa sulla corrispondente sequenza di dati precedenti, utilizzando ad esempio la (9.43). Ogni stato quindi ha una sua diversa stima della fase della portante, o di altri parametri di interesse.

9.10 Esercizi

9.1 - Si determini lo stimatore a massima verosimiglianza dell'errore di temporizzazione τ per la modulazione PAM, supponendo *noti* i dati a_k. Si mostri che il risultato coincide con l'approssimazione per alto rapporto segnale-rumore dello stimatore descritto nella Sez. 2.

9.2 - Si ricavi l'espressione (9.8) della varianza dell'errore nella stima di τ. *Commento*: si noti che la varianza *dipende* dalla sequenza dei dati; il risultato vale solo per N grande, se è lecito invocare la legge dei grandi numeri.

9.3 - È noto che con *roll-off* del 100% il diagramma ad occhio di fig. 9.2 passa esattamente per lo zero negli istanti $\pm T/2$. Tuttavia tale eccesso di banda è poco frequente. In genere, quindi, il segnale di controllo del sincronizzatore di simbolo DTL ha fluttuazioni casuali, dipendenti dai dati, anche in assenza di

rumore additivo gaussiano. È però possibile sottrarre da $y_{k-1/2}$ una replica stimata dell'ISI nell'istante $kT - T/2$, eliminando quindi tale disturbo, detto *self noise* o *pattern noise*. Si descriva la struttura del sincronizzatore.
Commento: occorre conoscere le decisioni sia *precedenti* sia *seguenti*, perlomeno quelle che danno un contributo significativo.

9.4 - Si mostri che se i dati decisi sono corretti i successivi aggiornamenti nella (9.17) sono incorrelati, per cui basta conoscerne la varianza, mentre nella (9.16) non lo sono. Si mostri anche che la densità spettrale di potenza a frequenza zero ha (ovviamente) lo stesso valore nei due casi, e che quindi le prestazioni, almeno per anelli a banda stretta, sono uguali.

9.5 - Si mostri che la versione del sincronizzatore di simbolo con un solo campione per simbolo suggerita dal sincronizzatore a massima verosimiglianza per basso rapporto segnale-rumore, cioè quella con y_k anziché $\hat{a}_k = \mathrm{sgn}(y_k)$, non può funzionare; infatti $E[y_k y_{k-1}] = E[y_{k-1} y_k]$ a causa della stazionarietà, e quindi il segnale di controllo ha valor medio sempre nullo. Perché invece il sincronizzatore di Gardner con *due* campioni per simbolo funziona?
Suggerimento: si osservi che $E[y_{k-1} y_{k-1/2}] \neq E[y_{k-1/2} y_k]$ a causa della *ciclostazionarietà* del processo.

9.6 - In una modulazione binaria antipodale, con dati equiprobabili e indipendenti, le forme d'onda $g(t - kT)$ sono ortogonali ed hanno *roll-off* α. Posto per semplicità $T = 1$ e con un errore di temporizzazione τ, siano $y_k = \int r(t) g(t - k - \tau) dt$ i campioni all'uscita del filtro adattato.
Se $\tau \ll 1$ i mostri che il valor medio di $w_k = y_k a_{k-1} - y_{k-1} a_k$, cioè del segnale di controllo del sincronizzatore di Mueller e Müller, è proporzionale a τ, e ricordando l'espressione $\dfrac{\sin(\pi t)}{\pi t} \dfrac{\cos(\alpha \pi t)}{1 - 4\alpha^2 t^2}$ delle forme d'onda di Nyquist si determini il fattore di proporzionalità. *Suggerimento*: si valuti la pendenza in $t = 1$; si evitino calcoli inutili.
Si calcoli la varianza di w_k e si mostri che w_k e w_{k-i} sono incorrelati per $i \neq 0$ (per semplicità si supponga $\tau = 0$).
Si confrontino i risultati con quelli di fig. 9.3.

9.7 - Nei sincronizzatori di simbolo che utilizzano un solo campione per simbolo si potrebbe utilizzare anche la coppia di campioni y_{k+1} e y_{k-2}, ed eventualmente y_{k+2} e y_{k-3}, ecc. Si mostri che si trarrebbe vantaggio dal

fatto che in presenza di un errore di temporizzazione anche $h(\pm 2T + \Delta\tau)$, $h(\pm 3T + \Delta\tau)$, ecc. sono proporzionali all'errore $\Delta\tau$. Si mostri tuttavia che i vari contributi andrebbero pesati diversamente, per ottenere il minimo possibile di *jitter*. Quali dovrebbero essere i pesi dei vari contributi? Di quanto si ridurrebbe il *jitter* rispetto alla semplice soluzione (9.17)? Si mostri eseguendo il calcolo per vari valori di *roll-off* che il guadagno, a fronte di un notevole incremento di complessità, e di ritardo, sarebbe modesto.

9.8 - In una espressione come la (9.17) per l'aggiornamento del sincronismo di simbolo si supponga che $\hat{\tau}_k$ sia discretizzato con passo T/n_τ, cioè che sia $\hat{\tau}_k = n/n_\tau$ con n intero. Si mostri che basta moltiplicare la (9.17) per n_τ per ottenere l'espressione per l'aggiornamento dell'indice n. Quali potrebbero essere valori ragionevoli di n_τ per la segnalazione binaria antipodale? Gli stessi valori sarebbero adeguati per la segnalazione multilivello? La presenza di un codice, che porta ad operare con un rapporto E_b/N_0 inferiore, modifica sostanzialmente le conclusioni?

9.9 - Si discuta la seguente pignoleria: i campioni y_k e y_{k-1} utilizzati nella (9.17) sono calcolati in due istanti di tempo diversi, rispettivamente $kT + \tau_k$ e $kT - T + \tau_{k-1}$. Poiché $\tau_{k-1} \neq \tau_k$, la (9.33) non è corretta. Si indichi come andrebbe modificata. *Commento*: la maggior complessità non è compensata, in alcun caso pratico, da una maggior precisione; infatti la correzione sarebbe significativa solo per anelli con banda molto ampia, e quindi con *jitter* inaccettabile.

9.10 - Nella modulazione QAM l'ampiezza $|d_k|$ dei simboli varia casualmente. L'anello di controllo della portante (9.40) ha quindi un guadagno casuale, che per semplicità si può ritenere indipendente da simbolo a simbolo. Come cambiano l'acquisizione e il funzionamento a regime?

9.11 - Se nell'anello di controllo del *clock* o della fase si introduce un ritardo di L tempi di simbolo (ad esempio a causa di tempi richiesti per la elaborazione, o a causa di un codice) come variano le caratteristiche dell'anello? In particolare la stabilità diventa più critica, ed il *jitter* a regime aumenta. Si determini, eventualmente in modo numerico, di quanto aumenta il *jitter* in funzione del ritardo e della banda d'anello.

9.12 - L'introduzione di un secondo polo e di uno zero in un anello di controllo ha effetti benefici sull'errore a regime in presenza di *offset* di frequenza. Tuttavia ha anche un effetto negativo sul *jitter* a regime. Assumendo, come solitamente si consiglia, che lo zero sia posto alla pulsazione $b = 0.25\omega_0$ si determini, eventualmente in modo numerico, di quanto aumenta il *jitter*.

9.13 - Un sistema di trasmissione numerica a pacchetti utilizza la modulazione 2PSK ed un codice convoluzionale con $R = 1/4$. Il rumore è additivo gaussiano bianco ed E_b/N_0 è pari a 3 dB. Per la sincronizzazione del pacchetto si trasmette un *preambolo* di 30 simboli con fasi alternativamente uguali a 0 e π:

$$s(t, \vartheta) = \cos(2\pi f_0 t + \varphi_k + \vartheta) \qquad \varphi_k = 0, \pi \qquad 0 \le t \le 30T$$

Supponendo già acquisito il sincronismo di simbolo, quale è lo stimatore a massima verosimiglianza della fase ϑ della portante? e quale la varianza della stima?

9.14 - In presenza di rumore additivo gaussiano bianco, quale è lo stimatore a massima verosimiglianza della fase ϑ della forma d'onda

$$\sum_{k=1}^{N} a_k g(t - kT) \cos(\omega_0 t + \vartheta)$$

dove le $g(t - kT)$ sono ortogonali, con energia E_g, e la sequenza dei dati a_k è nota? Si considerino i casi $a_k = 1$ per tutti i k e $a_k = (-1)^k$. Quale è nei due casi la varianza della stima? Se $E_g/N_0 = 4$ dB, quanti simboli N occorrono per avere $\sigma_{\hat{\vartheta}} = 0.05$ rad?

9.15 - In una espressione come la (9.40) per l'aggiornamento del sincronismo di portante si supponga che $\hat{\vartheta}_k$ sia discretizzato con passo $2\pi/n_\vartheta$, cioè che sia $\hat{\vartheta}_k = 2\pi n/n_\vartheta$ con n intero. Si mostri che basta moltiplicare la (9.40) per $n_\vartheta/2\pi$ per ottenere l'espressione per l'aggiornamento dell'indice n. Quali potrebbero essere valori ragionevoli di n_ϑ per la segnalazione 4PSK? Gli stessi valori sarebbero adeguati per la segnalazione QAM? La presenza di un codice, che porta ad operare con un rapporto E_b/N_0 inferiore, modifica sostanzialmente le conclusioni?

9.16 - C'è qualche differenza tra la correzione di un errore di frequenza Δf ottenuta a monte della conversione in banda base oppure a valle moltiplicando

l'equivalente passa basso per $\exp(-j\hat{\vartheta}_k)$, dove $\hat{\vartheta}_k$ insegue la rotazione di fase dovuta all'errore di frequenza? *Suggerimento*: si supponga ΔfT molto grande; se invece ΔfT è piccolo ...