Fondamenti di Analisi Statistica dei Segnali

Letizia Lo Presti e Fabrizio Sellone

Dipartimento di Elettronica
Politecnico di Torino

Indice

1 Processi casuali a tempo discreto **1**

1.1 Definizione 1

1.2 Descrizione statistica 2

1.2.1 Medie di insieme 2

1.2.2 Indipendenza, correlazione e ortogonalità 5

1.3 Stazionarietà 6

1.3.1 Proprietà dell'autocorrelazione dei processi WSS 7

1.4 Medie temporali ed ergodicità 8

1.5 Classi di processi casuali a tempo discreto 9

1.5.1 Sequenza bianca o rumore bianco 9

1.5.2 Processo gaussiano 11

1.5.3 Processi di Markov del primo ordine 12

1.6 Descrizione di un processo nel dominio della frequenza 13

1.6.1 Definizione di spettro di potenza 13

1.6.2 Mutua correlazione e spettro mutuo 14

2 Trasformazioni lineari di processi casuali WSS **17**

2.1 Risposta di un sistema lineare ad un processo WSS 18

2.1.1 Medie di insieme del processo filtrato 18

2.1.2 Densità spettrale 19
2.1.3 Filtraggio di un rumore bianco 20
2.2 Rappresentazione di Wold 20
2.2.1 Sistema generatore nella scomposizione di Wold 23
2.2.2 Analisi e sintesi di processi 24
2.3 Stazionarietà, transitorio e stabilità nella rappresentazione di Wold 25
2.4 Il problema del transitorio per un processo AR del primo ordine . 26
2.4.1 Autocorrelazione e spettro asintotici 26
2.4.2 Autocorrelazione con ingresso causale 27
2.5 Lo spettro e il sistema generatore 29
2.5.1 Generatore di tipo FIR (MA) 29
2.5.2 Generatore con passa-tutto e FIR (MA) a fase minima . . 30
2.5.3 Generatore di tipo IIR (AR e ARMA) 31

3 Elementi di teoria della stima 33
3.1 Stima e stimatore 33
3.2 Qualità degli stimatori 34
3.2.1 L'errore quadratico medio (MSE) 36
3.3 Stima della media 36
3.4 Stima dell'autocorrelazione 38
3.4.1 Primo stimatore 39
3.4.2 Secondo stimatore 40

4 Stima di parametri 43
4.1 Descrizione vettoriale dei dati del problema 43
4.1.1 Descrizione in termini di densità di probabilità 44
4.2 Esempio di riferimento 45

4.3 Stima bayesiana 47
4.3.1 Stima MMSE 48
4.3.2 Stima MAP 48
4.4 Stima a massima verosimiglianza (ML) 51
4.5 Cramer-Rao Bound (CRB) 54

5 Stima spettrale classica 55
5.1 Il periodogramma 55
5.1.1 Il problema della stima spettrale 59
5.2 I metodi della stima spettrale classica 59
5.2.1 Stima spettrale per mezzo del periodogramma 60
5.3 Il periodogramma semplice 60
5.3.1 Risoluzione spettrale e polarizzazione locale. 65
5.4 Il periodogramma di Bartlett 67
5.5 Il periodogramma di Welch 69
5.5.1 Finestre di pesatura 71
5.5.2 Il Periodogramma di Welch 76

6 Stimatori lineari a Minimo Errore Quadratico Medio (MMSE) 79
6.1 Stimatore lineare 79
6.1.1 Ortogonalità dell'errore 81
6.1.2 Errore residuo 82
6.1.3 Stimatore lineare per sequenze infinite 82
6.2 Predittore lineare 83
6.2.1 Il filtro FIR per la predizione con i processi WSS 84
6.2.2 Predittore lineare di ordine p 85
6.2.3 Il filtro dell'errore di predizione 89

6.3 Il filtro di Wiener 90
6.3.1 Il filtro di Wiener per segnale e rumore scorrelati 91
6.3.2 Il filtro di Wiener per sequenze infinite 92
6.3.3 Il filtro di Wiener con sequenze di apprendimento 94
6.3.4 Generalizzazione del filtro di Wiener 95
6.3.5 Algoritmo LMS 95

7 Stima spettrale parametrica con modelli AR 99

7.1 Modello AR 100
7.2 Il predittore lineare e il modello AR 101
7.2.1 Dalle equazioni normali alle equazioni di Yule-Walker . . . 102
7.3 Metodo diretto per ottenere le equazioni di Yule-Walker 104
7.4 Stima spettrale basata sulle equazioni di Yule-Walker 105
7.5 Scelta dell'ordine p del modello AR 106

1

Processi casuali a tempo discreto

In questo capitolo verranno introdotti alcuni concetti di base relativi alla classe dei segnali a tempo discreto denominati *processi casuali a tempo discreto*.

1.1 Definizione

Si definisce *processo casuale a tempo discreto* o più semplicemente *processo casuale discreto* una sequenza di variabili casuali, indicata con la notazione $X[n]$, definita per tutti i valori interi della variabile n, chiamata *tempo discreto*. In alcuni contesti applicativi un processo casuale discreto $X[n]$ è denominato *serie temporale*.

Ad ogni istante discreto n si estrae dal processo una variabile casuale che può essere continua, discreta, mista, reale o complessa. Nel seguito, se non è specificato altrimenti, verranno considerati processi reali e continui in ampiezza.

In molti applicazioni un processo discreto $X[n]$ è ottenuto campionando un processo a tempo continuo $X(t)$, cioè

$$X[n] = X(nT_c)$$

dove T_c è l'intervallo di campionamento. In questi casi si dice che $X[n]$ rappresenta un *processo campionato* e $f_c = 1/T_c$ è la frequenza di campionamento.

Nel seguito la notazione $X[n]$ verrà utilizzata sia per indicare il processo casuale sia la singola variabile casuale estraibile dal processo all'istante discreto n. Si è preferito non introdurre due notazioni diverse, poiché la differenza è quasi sempre formale e due notazioni avrebbero appesantito la trattazione dell'argomento.

1.2 Descrizione statistica

Un processo reale a tempo discreto $X[n]$ viene descritto statisticamente in modo analogo ad un processo a tempo continuo. Ad ogni istante discreto n è associata una variabile casuale, descritta statisticamente dalla funzione di distribuzione cumulativa. Tale funzione varia con n ed è idonea a descrivere la statistica del primo ordine del processo. Si definisce pertanto la *funzione di distribuzione cumulativa del primo ordine* del processo discreto come

$$F_X(x;n) \triangleq P\{X[n] \leq x\}$$

e la *densità di probabilità del primo ordine*

$$f_X(x;n) = \frac{\partial}{\partial x} F_X(x;n)$$

In modo analogo si possono definire le funzioni di distribuzione cumulativa e le densità di probabilità di ordine superiore.

Si può introdurre anche il concetto di processo casuale complesso. Esso è definito come

$$X[n] = X_R[n] + jX_I[n]$$

dove $X_R[n]$ e $X_I[n]$ sono processi reali, descritti statisticamente per mezzo delle relative densità di probabilità.

1.2.1 Medie di insieme

Sebbene dal punto di vista teorico la densità di probabilità sia in grado di descrivere adeguatamente le caratteristiche statistiche di un processo, spesso tale funzione non viene utilizzata nelle applicazioni di ingegneria, poiché è molto complessa o difficile da valutare. Si preferisce in questi casi descrivere il processo in termini di *medie di insieme*. Di esse le più significative sono quelle riportate qui di seguito.

Valore atteso. Il valore atteso di un processo reale è definito come

$$\mu[n] = \mathrm{E}\{X[n]\} = \int_{-\infty}^{+\infty} x f_X(x;n)dx \tag{1.1}$$

ed è una funzione del tempo discreto n. Nel seguito sarà indicato più brevemente con μ_n. Comunemente μ_n viene chiamato anche *media* o *valore medio* del processo.

Per ciascun istante n, il valore atteso può essere messo in relazione con la media aritmetica dei valori assunti, indicati con $x^{(i)}[n]$, dalla variabile casuale $X[n]$ in N esperimenti casuali. In particolare si ha

$$\mu_n = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x^{(i)}[n]$$

dove il termine

$$\frac{1}{N} \sum_{i=1}^{N} x^{(i)}[n]$$

ha il significato di *frequenza relativa*.

Se il processo è complesso si ha

$$\mu_n = \mathrm{E}\{X[n]\} = \mathrm{E}\{X_R[n]\} + j\mathrm{E}\{X_I[n]\} \tag{1.2}$$

Autocorrelazione. Per un processo reale essa è definita come

$$R_X[m,n] = \mathrm{E}\{X[m]X[n]\} \tag{1.3}$$

e può essere calcolata a partire dalla densità di probabilità del secondo ordine.

In termini di frequenza relativa si può scrivere

$$R_X[m,n] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x^{(i)}[m]x^{(i)}[n]$$

dove $x^{(i)}[m]$ e $x^{(i)}[n]$ sono la coppia di valori assunti dalle due variabili casuali $X[m]$ e $X[n]$ all'i-esimo esperimento casuale.

Alcuni autori chiamano l'autocorrelazione semplicemente *correlazione*.

Se il processo è complesso l'autocorrelazione è definita come

$$R_X[m,n] = \mathrm{E}\{X^*[m]X[n]\} \tag{1.4}$$

Autocovarianza. Per un processo reale essa è definita come

$$C_X[m,n] = \mathrm{E}\{(X[m]-\mu_m)(X[n]-\mu_n)\} \tag{1.5}$$

Da essa si ricava la *varianza* del processo definita come

$$\sigma_n^2 = C_X[n,n] \tag{1.6}$$

Per un processo complesso si ha

$$C_X[m,n] = \mathrm{E}\{(X[m]-\mu_m)^*(X[n]-\mu_n)\} \tag{1.7}$$

Matrice di correlazione. Essa è una matrice quadrata che contiene un insieme limitato di valori dell'autocorrelazione. Si osservi che, per processi reali, tale matrice è simmetrica essendo $R_X[m,n] = R_X[n,m]$, come si ricava facilmente dalle (1.3). Pertanto la *matrice di autocorrelazione* o di *correlazione* di ordine L è una matrice simmetrica di dimensioni $L \times L$, definita come

$$\mathbf{R}_X = \begin{bmatrix} R_X[0,0] & R_X[0,1] & \dots & R_X[0,L-1] \\ R_X[0,1] & R_X[1,1] & \dots & R_X[1,L-1] \\ \vdots & \vdots & \ddots & \vdots \\ R_X[0,L-1] & R_X[1,L-1] & \dots & R_X[L-1,L-1] \end{bmatrix} \tag{1.8}$$

Matrice di covarianza. Essa è definita analogamente alla matrice di correlazione. Pertanto è una matrice quadrata che contiene un insieme limitato di valori dell'autocovarianza, simmetrica per processi reali essendo $C_X[m,n] = C_X[n,m]$, come si ricava facilmente dalle (1.5). La *matrice di autocovarianza* di ordine L di processi reali si definisce come

$$\mathbf{C}_X = \begin{bmatrix} C_X[0,0] & C_X[0,1] & \dots & C_X[0,L-1] \\ C_X[0,1] & C_X[1,1] & \dots & C_X[1,L-1] \\ \vdots & \vdots & \ddots & \vdots \\ C_X[0,L-1] & C_X[1,L-1] & \dots & C_X[L-1,L-1] \end{bmatrix} \tag{1.9}$$

Momenti di ordine superiore. E' possibile definire dei valori attesi legati alla statistica di ordine superiore del processo e le corrispondenti trasformate di Fourier, detti *spettri di ordine superiore*. Tra i momenti di ordine superiore più usati vi sono quelli di ordine tre, che dà origine al *bispettro* e quello di ordine quattro che dà origine al *trispettro*. Si può osservare che lo spettro di potenza comunemente usato è originato dalla statistica di ordine due.

Una trattazione introduttiva su questo argomento si può trovare in [4].

Le matrici di autocorrelazione e di covarianza si possono scrivere anche in forma compatta introducendo le notazioni vettoriali indicate nel seguito.

- Il vettore delle variabili casuali

$$\mathbf{X}_L = [\, X[0] \quad X[1] \quad \dots \quad X[L-1] \,]^T$$

- Il vettore dei valori medi

$$\mathbf{M}_L = [\, \mu_0 \quad \mu_1 \quad \dots \quad \mu_{L-1} \,]^T$$

Infatti si verifica facilmente che valgono le relazioni

$$\mathbf{R}_X = \mathrm{E}\{\mathbf{X}_L \mathbf{X}_L^T\}$$

e

$$\mathbf{C}_X = \mathbf{R}_X - \mathbf{M}_L \mathbf{M}_L^T$$

1.2.2 Indipendenza, correlazione e ortogonalità

Le consuete definizioni di *indipendenza statistica* e di *scorrelazione*, valide per le variabili casuali, possono essere utilizzate per introdurre gli stessi concetti per i processi a tempo discreto. Pertanto la definizione di indipendenza sarà legata al concetto di densità di probabilità congiunta.

Ragionando in termini di momenti, si può dire che se due campioni $X[m]$ e $X[p]$ di un processo discreto sono *indipendenti* si ha che

$$\mathrm{E}\{X[m]X[p]\} = \mathrm{E}\{X[m]\}\mathrm{E}\{X[p]\}$$

da cui segue

$$C_X[m,p] = 0 \tag{1.10}$$

per la definizione (1.5) di autocovarianza. Se vale l'equazione (1.10) i due campioni si dicono *scorrelati*. Pertanto campioni indipendenti di un processo sono anche scorrelati, ma in generale non è vero il viceversa. Un caso particolare per il quale la scorrelazione implica l'indipendenza statistica è quello in cui i due campioni siano variabili casuali gaussiane.

Inoltre i campioni $X[m]$ e $X[p]$ si dicono *ortogonali* se

$$\mathrm{E}\{X[m]X[p]\} = 0 \tag{1.11}$$

Segue dalla definizione (1.5) e dall'equazione (1.10) che $X[m]$ e $X[p]$ sono ortogonali se sono scorrelati e se μ_m o μ_p o entrambi sono nulli.

1.3 Stazionarietà

Un processo casuale a tempo discreto si dice *stazionario in senso stretto* se la sua funzione distribuzione cumulativa di ordine K

$$F_X(x[n_1], \ldots, x[n_K])$$

è invariante per traslazioni temporali, per ogni K (intero positivo) e per ogni insieme $(n_1, \ldots, n_K)$, ovvero

$$F_X(x[n_1], \ldots, x[n_K]) = F(x[n_1 + n_0], \ldots, x[n_K + n_0])$$

dove n_0 è un intervallo di durata arbitraria.

Tuttavia nella pratica si utilizza soltanto il concetto di *stazionarietà in senso lato* riferita soltanto alle caratteristiche di regolarità dei momenti del primo e del secondo ordine. Si osservi che sebbene la stazionarietà in senso stretto, per come è stata qui definita, sia una condizione molto più restrittiva della stazionarietà in senso lato, in generale non è vero che un processo stazionario in senso stretto sia necessariamente stazionario in senso lato, in quanto potrebbe non avere momenti finiti.

In particolare si dice che un processo è stazionario in senso lato se

$$\mathrm{E}\{X[n]\} = \mathrm{E}\{X[m]\} = \mu \tag{1.12}$$

e

$$R_X[m, n] = \mathrm{E}\{X[m]X[n]\} = R_X[m - n] = r_X[l] \tag{1.13}$$

dove $l = m - n$ rappresenta il ritardo tra due campioni del processo (in inglese *correlation lag*). Si noti che è stata introdotta in questo caso la notazione $r_X[l]$, che verrà adottata per distinguere in modo netto l'autocorrelazione dei processi stazionari da quella dei processi non stazionari. E' immediato verificare che vale la relazione

$$r_X[l] = \mathrm{E}\{X[n + l]X[n]\} \tag{1.14}$$

Come per i processi a tempo continuo, anche in questo caso i processi a tempo discreto stazionari in senso lato vengono indicati con la sigla inglese WSS. Essi sono anche chiamati processi *debolmente stazionari*.

Se il processo è complesso si ha

$$r_X[l] = \mathrm{E}\{X^*[n]X[n + l]\} \tag{1.15}$$

1.3.1 Proprietà dell'autocorrelazione dei processi WSS

L'autocorrelazione di un processo WSS gode di tre importanti proprietà. Si cominci a considerare il caso di processo relae.

- L'autocorrelazione $r_X[l]$ è una funzione pari. Come per i processi a tempo continuo (si veda [20]) si dimostra facilmente che

$$r_X[l] = r_X[-l]$$

- L'autocorrelazione $r_X[l]$ ha un massimo assoluto nell'origine, cioè

$$\mathrm{E}\{|X[n]|^2\} = r_X[0] \geq |r_X[l]|$$

  La dimostrazione è analoga a quella per processi a tempo continuo (si veda [20]).

- La matrice di autocorrelazione assume la forma

$$\mathbf{R}_X = \begin{bmatrix} r_X[0] & r_X[1] & \cdots & r_X[L-1] \\ r_X[1] & r_X[0] & \ddots & r_X[L-2] \\ \vdots & \ddots & \ddots & \vdots \\ r_X[L-1] & r_X[L-2] & \cdots & r_X[0] \end{bmatrix} \tag{1.16}$$

  ovvero la matrice di autocorrelazione $(L \times L)$ è una matrice Toeplitz simmetrica. In quanto tale essa ha in totale L elementi distinti. Si dimostra che essa è una matrice *semi-definita positiva* [3, pag.131][1].

Nel caso di processo complesso le proprietà precedenti si modificano come segue.

- L'autocorrelazione $r_X[l]$ di un processo complesso è una funzione hermitiana, ossia

$$r_X[l] = r_X^*[-l]$$

[1] Una matrice $\mathbf{A}$ si dice *semi-definita positiva* se per qualsiasi vettore $\mathbf{a}$ di dimensione $(L \times 1)$, vale la relazione $\mathbf{a}^T\mathbf{A}\mathbf{a} \geq 0$. Se $\mathbf{A}$ è una matrice di correlazione può essere conveniente riscrivere la condizione appena vista nella forma $\sum_{i=1}^{L}\sum_{j=1}^{L} a_i a_j r[|i-j|] \geq 0$; da questa espressione si ricava una dimostrazione della proprietà enunciata in termini di filtraggio, come verrà mostrato nel paragrafo 2.1.

- L'autocorrelazione $r_X[l]$ di un processo complesso è reale nell'origine, ed è tale che
$$\mathrm{E}\{|X[n]|^2\} = r_X[0] \geq |r_X[l]|$$

- La matrice di autocorrelazione assume la forma
$$\mathbf{R}_X = \begin{bmatrix} r_X[0] & r_X[1] & \cdots & r_X[L-1] \\ r_X^*[1] & r_X[0] & \ddots & r_X[L-2] \\ \vdots & \ddots & \ddots & \vdots \\ r_X^*[L-1] & r_X^*[L-2] & \cdots & r_X[0] \end{bmatrix} \tag{1.17}$$
ovvero la matrice di autocorrelazione $(L \times L)$ è una matrice Toeplitz hermitiana. Si dimostra che essa è una matrice *semi-definita positiva*.

1.4 Medie temporali ed ergodicità

Si consideri una realizzazione $x[n]$ di un processo $X[n]$. Si dice *media temporale* di una funzione $g(x[n])$ la quantità

$$< g(x[n]) >= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} g(x[n])$$

Ad esempio la media temporale di $x[n]$ è

$$< x[n] >= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} x[n]$$

Se $X[n]$ è WSS e $< x[n] >= \mathrm{E}\{X[n]\}$ si dice che la sequenza è ergodica per la media.

In modo analogo si può definire l'ergodicità per una generica funzione $g(x[n])$.

Energia e Potenza. Una media temporale molto significativa per descrivere un segnale è la *potenza*. Essa è legata all'*energia* definita come

$$\mathcal{E}_x = \sum_{n=-\infty}^{+\infty} |x[n]|^2 \tag{1.18}$$

Se il segnale non è ad energia finita, ossia $\mathcal{E}_x \to \infty$, si definisce la *potenza* come la media temporale

$$\mathcal{P}_x =< |x[n]|^2 >= \lim_{N\to\infty} \frac{1}{2N+1} \sum_{n=-N}^{N} |x[n]|^2$$

dove la quantità

$$\mathcal{E}_{x,\,N} = \sum_{n=-N}^{N} |x[n]|^2$$

è detta *energia parziale* del segnale $x[n]$.

1.5 Classi di processi casuali a tempo discreto

Nelle applicazioni ricorrono spesso alcuni particolari processi. Alcuni di essi sono indicati qui di seguito.

1.5.1 Sequenza bianca o rumore bianco

Si consideri un processo casuale $X[n]$ stazionario, a media nulla. La sua densità di probabilità del primo ordine $f_X(x;n)$ non dipende dal tempo, pertanto si può esprimere come $f_X(x)$. Se la sua densità di probabilità di ordine K,

$$f_X(x[n_1], x[n_2], \ldots, x[n_K])$$

soddisfa per qualsiasi K la relazione

$$f_X(x[n_1], x[n_2], \ldots, x[n_K]) = f_X(x[n_1]) f_X(x[n_2]) \ldots f_X(x[n_K]) \tag{1.19}$$

si dice che $X[n]$ è una *sequenza puramente casuale*. La (1.19) è la condizione di indipendenza statistica di tutte le variabili casuali estraibili dal processo. Per tale ragione si dice che i campioni di $X[n]$ sono *indipendenti e identicamente distribuiti (iid)* e i processi di questo tipo sono detti *processi iid*.

Si osservi che un processo iid è sempre stazionario in senso stretto, ma non è necessariamente stazionario in senso lato. Ciò accade, ad esempio, nel caso di una sequenza iid con distribuzione di Cauchy (tra i numerosi testi sull'argomento si veda ad esempio [22])

Se il processo iid è WSS l'indipendenza statistica implica che

$$r_X[n-m] = \mathrm{E}\{X[n]X[m]\} = \sigma_X^2 \delta[n-m]$$

dove σ_X^2 è la varianza del processo e $\delta[i]$ è la delta di Kronecker.

Questa caratteristica impulsiva dell'autocorrelazione è simile a quella del rumore bianco di tipo analogico. Per tale ragione un processo iid è detto anche *sequenza bianca* o *rumore bianco*. Si osservi che in questo caso il processo non è a varianza infinita come nel caso di un rumore bianco a tempo continuo, la cui autocorrelazione è una delta di Dirac nell'origine. Inoltre non è corretto vedere un rumore bianco a tempo discreto come il risultato del campionamento di un rumore bianco a tempo continuo. Infatti, come si vedrà nel seguito, lo spettro di una sequenza bianca è definito solo per frequenze numeriche nell'intervallo $(-1/2, 1/2)$. Inoltre una sequenza bianca è un processo a potenza media finita, mentre un processo bianco a tempo continuo non ha potenza media finita avendo spettro costante su tutto l'asse reale.

Dal momento che sequenze indipendenti sono necessariamente scorrelate, tutte le sequenze iid a varianza finita sono bianche, mentre non è vero in generale il viceversa. Infatti si è visto che sequenze scorrelate possono non essere indipendenti. Si ponga attenzione al fatto che alcuni autori usano il termine 'puramente casuale' per descrivere sequenze di variabili casuali scorrelate ma non necessariamente indipendenti. La definizione di sequenza puramente casuale qui adottata presuppone invece l'indipendenza statistica. Pertanto i termini sequenza bianca, sequenza iid e sequenza puramente casuale sono qui del tutto equivalenti.

Si osservino alcune importanti caratteristiche dei processi iid WSS.

- La matrice di autocorrelazione $\mathbf{R}_X$ di un processo iid è di tipo diagonale, cioè ha la forma

  $$\mathbf{R}_X = \sigma_X^2 \mathbf{I}$$

  dove $\mathbf{I}$ è la matrice identità $(L \times L)$.

- La densità di probabilità $f_X(x)$ non è fissata. Processi iid usati frequentemente nelle applicazioni hanno densità di probabilità gaussiana o uniforme. Si possono avere comunque altri tipi di densità di probabilità.

Nel seguito un processo bianco verrà indicato con la notazione $W[n]$.

1.5.2 Processo gaussiano

Si dice che un processo $X[n]$ è gaussiano se un insieme di L campioni estraibili dal processo ha una densità di probabilità congiunta di ordine L di tipo gaussiano, per ogni L intero positivo. Pertanto fissato un insieme $(n_1, \ldots, n_L)$ di istanti di tempo discreto e ponendo

$$\mathbf{X}_L = [\, X[n_1] \quad \ldots \quad X[n_L] \,]^T$$

si può scrivere la densità di probabilità di ordine L come

$$f(\mathbf{X}_L) = \frac{1}{(2\pi)^{(L/2)}\sqrt{\det(\mathbf{C}_X)}} \exp\left[-\frac{1}{2}(\mathbf{X}_L - \mathbf{M}_L)^T \mathbf{C}_X^{-1}(\mathbf{X}_L - \mathbf{M}_L)\right] \tag{1.20}$$

dove si è definito il vettore dei valori medi

$$\mathbf{M}_L = [\, \mu_{n_1} \quad \ldots \quad \mu_{n_L} \,]^T$$

e $\mathbf{C}_X$ è la matrice di covarianza definita dalla relazione (1.9), per gli elementi di $\mathbf{X}_L$.

In generale, ad ogni istante discreto si ha una variabile casuale $X[n]$ gaussiana con media μ_n e varianza σ_n^2.

Nel caso particolare di processo WSS a media nulla, $\mu_n = \mu = 0$ e $\sigma_n^2 = \sigma_X^2$ per ogni n, quindi alla matrice $\mathbf{C}_X^{-1}$ si può sostituire la matrice di autocorrelazione $\mathbf{R}_X^{-1}$, che si ha dalla definizione (1.8), nell'espressione (1.20).

Se oltre ad essere WSS e a media nulla il processo è anche iid si ha un *processo gaussiano bianco*. In questo caso si ha

$$\mathbf{R}_X^{-1} = \frac{1}{\sigma_X^2}\mathbf{I}$$

e $\det(\mathbf{R}_X) = \sigma_X^{2L}$. L'espressione (1.20) diventa quindi

$$f(\mathbf{X}_L) = \frac{1}{(2\pi)^{(L/2)}\sqrt{\sigma_X^{2L}}} \exp\left[-\frac{1}{2\sigma_X^2}|\mathbf{X}_L|^2\right] \tag{1.21}$$

I processi gaussiani godono delle seguenti proprietà.

- Un rumore bianco gaussiano è necessariamente un processo iid. Ciò deriva dal fatto che per variabili casuali gaussiane la scorrelazione è condizione sufficiente per l'indipendenza statistica.

- Qualsiasi trasformazione lineare di un processo gaussiano è ancora un processo gaussiano.

- Tutti i momenti di ordine superiore di una distribuzione gaussiana possono essere espressi facendo uso dei soli momenti del primo e del secondo ordine della stessa distribuzione.

1.5.3 Processi di Markov del primo ordine

Si è visto che la densità di probabilità di un processo $X[n]$ è in generale una funzione di due variabili: x e il tempo discreto n. Se si considerano due variabili casuali estratte dal processo ai due istanti n_1 e n_2, a tali variabili, X_1 e X_2, è associata una densità di probabilità del secondo ordine del tipo

$$f_X(x_1, x_2;\, n_1, n_2)$$

Tale notazione può essere allegerita scrivendo la densità di probabilità del secondo ordine come

$$f_X(x_{n_1}, x_{n_2})$$

Ad essa è associata una densità di probabilità condizionata del tipo

$$f_X(x_{n_1}|x_{n_2})$$

Passando da una coppia di variabili casuali ad un'n-upla di variabili casuali si hanno le densità di probalità congiunta e condizionata di ordine n. In particolare interessa la densità di probabilità condizionata

$$f_X(x_n|x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4},\, \ldots)$$

Se tale densità di probabilità è del tipo

$$f_X(x_n|x_{n-1}, x_{n-2}, x_{n-3}, x_{n-4},\, \ldots) = f_X(x_n|x_{n-1})$$

si dice che il processo è markoviano del primo ordine.

In modo analogo si possono definire processi markoviani di ordine superiore.

1.6 Descrizione di un processo nel dominio della frequenza

Una sequenza stazionaria $X[n]$ è necessariamente di durata infinita. Ogni realizzazione del processo, $x^{(k)}[n]$, è un segnale deterministico a tempo discreto, generalmente ad energia infinita, cioè tale per cui

$$\mathcal{E}_{x,\,k} = \sum_n |x^{(k)}[n]|^2 \to \infty$$

Segnali di questo tipo non ammettono trasformata di Fourier (DTFT) e pertanto si pone il problema di come descrivere tali segnali, e quindi tali processi, nel dominio della frequenza.

Il problema è analogo a quello che si ha quando si vuole introdurre il concetto di spettro per processi WSS a tempo continuo. Anche in questo caso si ricorre ad una definizione legata all'autocorrelazione.

1.6.1 Definizione di spettro di potenza

Si consideri un processo WSS a media nulla, con autocorrelazione $r_X[l]$ assolutamente sommabile. La trasformata z di $r_X[l]$

$$S_X(z) = \sum_{l=-\infty}^{+\infty} r_X[l] z^{-l}$$

converge in una corona circolare nel piano complesso z. Si noti che, poiché $r_X[l] = r_X[-l]$, allora

$$S_X(z) = S_X(z^{-1})$$

Sul cerchio di raggio unitario $z = \mathrm{e}^{j\omega} = \mathrm{e}^{j2\pi f}$, $S_X(z)$ diventa

$$S_X(\mathrm{e}^{j2\pi f}) = \sum_{l=-\infty}^{+\infty} r_X[l] \mathrm{e}^{-j2\pi f l} \tag{1.22}$$

La $S_X(\mathrm{e}^{j2\pi f})$ è chiamata *densità spettrale di potenza* o *spettro di potenza* del processo casuale $X[n]$. In altri termini $S_X(\mathrm{e}^{j2\pi f})$ è la DTFT di $r_X[l]$. Pertanto si ha

$$r_X[l] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} S_X(\mathrm{e}^{j\omega}) \mathrm{e}^{j\omega l} d\omega = \int_{-1/2}^{+1/2} S_X(\mathrm{e}^{j2\pi f}) \mathrm{e}^{j2\pi f l} df$$

E' interessante osservare alcune proprietà della densità spettrale.

- $S_X(\mathrm{e}^{j2\pi f})$ è una funzione reale e pari, essendo la trasformata di una funzione reale e pari. Inoltre si può dimostrare che è una funzione non negativa (la dimostrazione è rimandata al capitolo 2, in quanto necessita di argomenti qui non ancora trattati).

- Il valore quadratico medio del processo si può ottenere dalla densità spettrale per integrazione. Infatti

$$\mathrm{E}\{|X[n]|^2\} = r_X[0] = \frac{1}{2\pi}\int_{-\pi}^{+\pi} S_X(\mathrm{e}^{j\omega})d\omega = \int_{-1/2}^{+1/2} S_X(\mathrm{e}^{j2\pi f})df$$

- $S_X(\mathrm{e}^{j2\pi f})$ è una funzione periodica di periodo 1, essendo $S_X(\mathrm{e}^{j2\pi f}) = S_X(\mathrm{e}^{j2\pi(f+k)})$, per ogni k intero positivo o negativo.

- Se $X[n]$ è una sequenza bianca, allora $r_X[l] = \sigma_X^2\delta[l]$. Pertanto trasformando si ha $S_X(\mathrm{e}^{j2\pi f}) = \sigma_X^2$. Ne segue che lo spettro di potenza di una sequenza bianca è una funzione costante di f.

1.6.2 Mutua correlazione e spettro mutuo

Dati due processi reali $X[n]$ e $Y[n]$ si possono definire la funzione di *mutua correlazione* e la funzione di *densità spettrale di potenza mutua* o, più brevemente, *spettro mutuo*, in modo analogo a quanto fatto per la funzione di autocorrelazione di un processo. In particolare si definisce mutua correlazione la funzione

$$R_{XY}[m,n] = \mathrm{E}\{X[m]Y[n]\}$$

Se $X[n]$ e $Y[n]$ sono WSS e la loro mutua correlazione dipende solo dall'intervallo l che separa un campione di uno dei due processi da un campione dell'altro, si parla di *stazionarietà congiunta* dei processi $X[n]$, $Y[n]$ e la loro correlazione mutua diventa

$$r_{XY}[l] = \mathrm{E}\{X[n]Y[n+l]\}$$

Nell'ipotesi di stazionarietà congiunta valgono le seguenti proprietà:

- La mutua correlazione non è, in generale, una funzione pari. Tuttavia si ha

$$r_{XY}[l] = r_{YX}[-l]$$

  Infatti ponendo

$$r_{XY}[l] = \mathrm{E}\{X[n+k]Y[n+k+l]\}$$

per $k = -l$ si ottiene

$$r_{XY}[l] = \mathrm{E}\{X[n-l]Y[n]\} = r_{YX}[-l]$$

- Dalla diseguaglianza di Schwartz si ha

$$|r_{XY}[l]| \leq \sqrt{r_X[0]r_Y[0]}$$

Si definisce spettro mutuo dei processi $X[n]$, $Y[n]$ la funzione

$$S_{XY}(\mathrm{e}^{j2\pi f}) = \sum_{l=-\infty}^{+\infty} r_{XY}[l]\mathrm{e}^{-j2\pi fl} = \sum_{l=-\infty}^{+\infty} r_{YX}[-l]\mathrm{e}^{-j2\pi fl}$$

Si dimostra facilmente che

$$S_{XY}(\mathrm{e}^{j2\pi f}) = S^*_{YX}(\mathrm{e}^{j2\pi f})$$

Definizioni per processi complessi. Le definizioni di mutua correlazione e di spettro mutuo per i processi complessi sono analoghe a quelle per i processi reali appena date. Si definisce mutua correlazione di due processi complessi $X[n]$, $Y[n]$ la funzione

$$R_{XY}[m,n] = \mathrm{E}\{X[m]Y^*[n]\}$$

che nel caso di stazionarietà congiunta diventa

$$r_{XY}[l] = \mathrm{E}\{X[n]Y^*[n+l]\}$$

con le proprietà

- $r_{XY}[l] = r^*_{XY}[-l]$
- $|r_{XY}[l]| \leq \sqrt{r_X[0]r_Y[0]}$

Analogamente si può definire lo spettro mutuo di due processi complessi.

2

Trasformazioni lineari di processi casuali WSS

In questo capitolo verrano esaminati i metodi di analisi dei processi WSS trasformati dai sistemi lineari.

Per semplicità verranno considerati solo processi a media nulla. Si osservi che questa ipotesi non costituisce una limitazione all'analisi qui compiuta. Infatti si può immaginare un processo WSS $Z[n]$ con media $\mathrm{E}\{Z[n]\} = \mu_Z \neq 0$, scomposto nella forma $Z[n] = X[n] + \mu_Z$ con $\mathrm{E}\{X[n]\} = 0$.

Va osservato che uno degli scopi principali di questa trattazione sarà l'analisi spettrale di un processo. Per questa ragione si può così concentrare l'analisi solo sul processo $X[n]$, in quanto il contributo spettrale del valor medio μ_Z, è limitato all'origine dell'asse del frequenze, che, in generale, è di scarso interesse nell'analisi delle variazioni di un segnale. Gli effetti di un valor medio non nullo vanno di fatto esaminati opportunamente, in separata sede. Il valor medio rappresenta infatti la componente continua del segnale o del processo elaborato, ed è pertanto costante ed illimitata nel tempo. Il suo studio richiede quindi una notevole astrazione ed esula dagli scopi di questo testo.

2.1 Risposta di un sistema lineare ad un processo WSS

Si abbia un processo casuale reale $X[n]$ all'ingresso di un sistema lineare tempo-invariante (LTI), con risposta all'impulso $h[n]$ reale. Tale sistema produce in uscita un processo $Y[n]$ legato a $X[n]$ dalla relazione

$$Y[n] = \sum_{i=-\infty}^{+\infty} h[n-i]X[i]$$

2.1.1 Medie di insieme del processo filtrato

Si dimostra facilmente che un sistema LTI con un processo WSS a media nulla in ingresso produce in uscita un processo ancora a media nulla, ossia

$$\mathrm{E}\{Y[n]\} = \mathrm{E}\{X[n]\} = 0$$

Infatti si ha

$$\mathrm{E}\{Y[n]\} = \sum_{1=-\infty}^{+\infty} h[n-i]\mathrm{E}\{X[n]\} = 0$$

E' possibile dimostrare che l'autocorrelazione del processo in uscita è data da

$$r_Y[l] = r_X[l] * g_h[l] \tag{2.1}$$

dove

$$g_h[l] = \sum_{i=-\infty}^{+\infty} h[l+i]h[i] \tag{2.2}$$

è chiamata *autocorrelazione temporale* di $h[n]$.

A questo punto è facile dimostrare la proprietà enunciata nel paragrafo 1.3.1, dove si affermava che una matrice $r_X[l]$ è semidefinita positiva. Infatti dalle (2.1) e (2.2) possiamo scivere $r_Y[0] = \sum_{m=-\infty}^{+\infty} r_X[m]g_h[m] = \sum_{m=-\infty}^{+\infty}\sum_{i=-\infty}^{+\infty} r_X[m]h[m+i]h[i] \geq 0$. Ponendo $m+i=j$ e considerando $h[n]$ la risposta di un filtro FIR causale si ottiene la condizione indicata in nota nel paragrafo 1.3.1.

Dall (2.1) si deduce che un processo WSS all'ingresso di un sistema LTI produce in uscita un processo ancora WSS. Ciò si può anche dimostrare scrivendo

l'autocorrelazione del processo di uscita $R_Y[n, n+l] = \mathrm{E}\{Y[n]Y[n+l]\}$ nella forma

$$R_Y[n, n+l] = \mathrm{E}\left\{ \sum_{i_1=-\infty}^{+\infty} \sum_{i_2=-\infty}^{+\infty} h[i_1]h[i_2]X[n-i_1]X[n+l-i_2] \right\}$$

Portando l'operatore di media all'interno delle sommatorie

$$R_Y[n, n+l] = \sum_{i_1=-\infty}^{+\infty} \sum_{i_2=-\infty}^{+\infty} h[i_1]h[i_2]r_X[l+i_1-i_2]$$

si osserva che il secondo membro dell'equazione non dipende da n, da cui si deduce che

$$R_Y[n, n+l] = r_Y[l]$$

e quindi $Y[n]$ è WSS.

Analogamente si può trovare che la mutua correlazione tra ingresso e uscita è data da

$$r_{XY}[l] = r_X[l] * h[l] \tag{2.3}$$

2.1.2 Densità spettrale

Dalla (2.1) si ricava facilmente l'espressione dello spettro di potenza di $Y[n]$. Infatti è noto che la trasformata z di una convoluzione è il prodotto delle trasformate z dei termini della convoluzione, da cui

$$S_Y(z) = G_h(z)S_X(z)$$

Osservando che $g_h[l] = h[l] * h[-l]$, si ha che

$$G_h(z) = H(z)H(z^{-1})$$

e quindi

$$S_Y(z) = H(z)H(z^{-1})S_X(z)$$

Sulla circonferenza a raggio unitario del piano z si ha $z = \mathrm{e}^{j2\pi f}$ e $H(\mathrm{e}^{-\mathrm{j}2\pi f}) = H^*(\mathrm{e}^{j2\pi f})$. Da ciò segue

$$S_Y(\mathrm{e}^{j2\pi f}) = |H(\mathrm{e}^{j2\pi f})|^2 S_X(\mathrm{e}^{j2\pi f}) \tag{2.4}$$

Dalla (2.3) si ricava invece lo spettro mutuo

$$S_{XY}(\mathrm{e}^{j2\pi f}) = H(\mathrm{e}^{j2\pi f})S_X(\mathrm{e}^{j2\pi f}) \tag{2.5}$$

2.1.3 Filtraggio di un rumore bianco

Un processo bianco $W[n]$ con varianza σ_W^2 ha spettro costante

$$S_W(\mathrm{e}^{j2\pi f}) = \sigma_W^2 \tag{2.6}$$

come si ricava facilmente trasformando l'autocorrelazione $r_W[l] = \sigma_W^2 \delta[l]$.

Filtrando un processo bianco attraverso un sistema LTI si ottiene un processo $Y[n]$ con spettro

$$S_Y(\mathrm{e}^{j2\pi f}) = |H(\mathrm{e}^{j2\pi f})|^2 \sigma_W^2 \tag{2.7}$$

come si ricava facilmente dalla (2.4). Per lo spettro mutuo vale la relazione

$$S_{WY}(\mathrm{e}^{j2\pi f}) = H(\mathrm{e}^{j2\pi f}) \sigma_W^2 \tag{2.8}$$

E' evidente che questa relazione può essere utilizzata per valutare la funzione di trasferimento di un sistema LTI. Infatti alimentando il sistema con un rumore bianco si può ottenere la funzione di trasferimento del sistema come

$$H(\mathrm{e}^{j2\pi f}) = \frac{S_{WY}(\mathrm{e}^{j2\pi f})}{\sigma_W^2} \tag{2.9}$$

La tecnica di alimentare un sistema con una sequenza bianca per valutarne la funzione di trasferimento è una delle più semplici e diffuse nel campo della modellistica e dell'identificazione dei sistemi lineari.

2.2 Rappresentazione di Wold

Esiste un importante teorema, noto come *Teorema di Wold*, che consente di esprimere qualsiasi processo stazionario $X[n]$ come l'uscita da un sistema LTI alimentato da un rumore bianco.

L'enunciato del teorema è il seguente.

Teorema di Wold. Qualsiasi processo stazionario tempo-discreto $X[n]$, privo di componenti deterministiche, può essere rappresentato come l'uscita di un sistema a tempo discreto LTI, causale e stabile, alimentato da un rumore bianco tempo-discreto $W[n]$, cioè

$$X[n] = \sum_{i=0}^{+\infty} h[i] W[n-i] \tag{2.10}$$

Si dice che $X[n]$ è un "processo lineare.

In altre parole, questo teorema afferma che un processo a tempo discreto WSS può essere rappresentato come l'uscita di sistema lineare causale e *causalmente invertibile* (ovvero, con una funzione di trasferimento tale che se invertita descriva ancora un sistema causale) al cui ingresso sia posto un rumore bianco. La condizione che il sistema sia causalmente invertibile permette anche di definire la trasformazione inversa. In altri termini si può ottenere un sistema 'sbiancatore' che, alimentato da un generico processo $X[n]$, dà in uscita un processo bianco.

Dimostrazione. Si consideri un processo WSS $X[n]$ che abbia autocorrelazione $r_X[m]$ e densità spettrale di potenza $S_X(\mathrm{e}^{j2\pi f})$ (con $|f| \leq 0.5$). Si assuma $S_X(\mathrm{e}^{j2\pi f})$ reale e continua per ogni $|f| \leq 0.5$.

Dalla trasformata z dell'autocorrelazione $r_X[m]$ si ha

$$S_X(z) = \sum_{m=-\infty}^{\infty} r_X[m] z^{-m} \tag{2.11}$$

che valutata sulla circonferenza a raggio unitario $z = \mathrm{e}^{j2\pi f}$ dà la densità spettrale di potenza.

Si assuma ora che $S_X(z)$ sia una funzione analitica (sia derivabile a qualsiasi ordine) in una regione anulare del piano z (ossia $\rho_1 < |z| < \rho_2$ con $\rho_1 < 1$ e $\rho_2 > 1$). Con queste ipotesi $S_X(z)$ può essere espressa come serie di Laurent

$$\log S_X(z) = \sum_{m=-\infty}^{\infty} v[m] z^{-m} \tag{2.12}$$

dove $v[m]$ assume come valori i coefficienti dell'espansione in serie. Si può interpretare $v[m]$ come una sequenza la cui trasformata z è $V(z) = \log S_X(z)$. Valutando la precedente espressione sulla circonferenza a raggio unitario si ha

$$\log S_X(\mathrm{e}^{j2\pi f}) = \sum_{m=-\infty}^{\infty} v[m] \mathrm{e}^{-\mathrm{j}2\pi f m} \tag{2.13}$$

In questo modo i coefficienti $v[m]$ sono i coefficienti dell'espansione in serie di Fourier della funzione $\log S_X(\mathrm{e}^{j2\pi f})$

$$v[m] = \int_{-1/2}^{1/2} [\log S_X(\mathrm{e}^{j2\pi f})] \mathrm{e}^{-\mathrm{j}2\pi f m} \mathrm{d}\, f$$

Dall'equazione (2.12) si ha

$$S_X(z) = \exp\left[\sum_{m=-\infty}^{\infty} v[m]z^{-m}\right]$$

che può essere scritta come

$$S_X(z) = \sigma_W^2 H(z)H(z^{-1}) \tag{2.14}$$

dove per definizione $\sigma_W^2 = \mathrm{e}^{v[0]}$ e

$$H(z) = \mathrm{e}^{\sum_{m=1}^{\infty} v[m]z^{-m}} \qquad |z| > \rho_1 \tag{2.15}$$

come si vede ricordando che la sequenza $v[m]$ è pari. Valutando l'espressione (2.14) sulla circonferenza a raggio unitario si ha

$$S_X(\mathrm{e}^{j2\pi f}) = \sigma_W^2 |H(\mathrm{e}^{j2\pi f})|^2 \tag{2.16}$$

Dalla definizione di $H(z)$ data dalla (2.15) è evidente che la parte causale di $v[m]$ è associata ad $H(z)$ e la parte anticausale a $H(z^{-1})$.

La funzione $H(z)$ è analitica nella regione $\rho_1 < |z|$. Pertanto in questa regione essa ammette un'espansione in serie di Taylor della forma

$$H(z) = \sum_{m=0}^{\infty} h[n]z^{-n} \tag{2.17}$$

ovvero $H(z)$ rappresenta la funzione di trasferimento di un filtro causale. Alimentando questo filtro con un rumore bianco avente densità spettrale pari a σ_W^2 si ha alla sua uscita un processo con densità spettrale $S_X(\mathrm{e}^{j2\pi f}) = \sigma_W^2 |H(\mathrm{e}^{j2\pi f})|^2$. Al contrario ponendo all'ingresso di un sistema con funzione di trasferimento $1/H(z)$ un processo con densità spettrale $S_X(\mathrm{e}^{j2\pi f})$ si ha all'uscita un processo bianco. Si parla in questo caso di filtro "sbiancante". □

Questo teorema ha delle vastissime applicazioni pratiche e rappresenta uno dei risultati fondamentali della teoria dei processi casuali a tempo discreto. Di fatto l'enunciato precedente permette di ricondurre l'analisi di un processo casuale privo di componenti deterministiche all'analisi di un sistema LTI. In termini spettrali ciò equivale a ricondurre lo studio del processo casuale all'equazione (2.7).

2.2.1 Sistema generatore nella scomposizione di Wold

Il sistema della scomposizione di Wold è fisicamente realizzabile e stabile. Pertanto $H(z)$ deve essere di tipo razionale e con un numero finito di parametri, cioè del tipo

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_m z^{-m}}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_l z^{-l}} \tag{2.18}$$

Affinché tale sistema sia anche stabile i poli di $H(z)$ devono essere contenuti all'interno del cerchio unitario.

I modelli associati alla (2.18) vengono denominati come indicato nel seguito.

Modelli AR. Un modello AR (dall'inglese *autoregressive*) ha una funzione di trasferimento a tutti poli del tipo

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_l z^{-l}} \tag{2.19}$$

dove si è normalizzato a 1 il numeratore per semplificare la scrittura. Si osservi che questo non comporta una perdita di generalità nella trattazione poiché il numeratore dell'equazione(2.19) rappresenta solo un fattore di scala che può essere controllato attraverso la varianza della sequenza bianca in ingresso.

Un modello AR è in grado di produrre un processo WSS $X[n]$ del tipo

$$X[n] = -a_1 X[n-1] - a_2 X[n-2] - \ldots - a_l X[n-l] + W[n] \tag{2.20}$$

Modelli MA. Un modello MA (dall'inglese *moving average*) ha una funzione di trasferimento a tutti zeri del tipo

$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots + b_m z^{-m} \tag{2.21}$$

Un modello MA è in grado di produrre un processo WSS $X[n]$ del tipo

$$X[n] = b_0 W[n] + b_1 W[n-1] + b_2 W[n-2] + \ldots + b_m W[n-m] \tag{2.22}$$

Modelli ARMA. Un modello ARMA (dall'inglese *autoregressive moving average*) ha una funzione di trasferimento contenente sia poli che zeri del tipo (2.18).

Nel seguito i modelli AR, MA e ARMA verrano indicati come *modelli* o *sistemi generatori*. In generale i modelli ARMA sono più complessi da ottenere rispetto ai modelli AR e MA, avendo sia poli che zeri, ma hanno il vantaggio di richiedere normalmente un numero minore di coefficienti per rappresentare un processo casuale con lo stesso grado di fedeltà. Questa caratteristica può essere assai rilevante in applicazioni che, per ragioni computazionali, richiedano $H(z)$ con un basso numero di coefficienti.

2.2.2 Analisi e sintesi di processi

I modelli AR, MA e ARMA vengono utilizzati in due classi di problemi ben distinti, che presentano spesso problematiche e metodologie anche molto diverse tra loro. In particolare si possono individuare due ruoli distinti dei modelli in questione, a seconda che si utilizzino per generare un processo $X[n]$ o per "rendere bianco" il processo stesso, cioè per riottenere $W[n]$ da $X[n]$. Le due classi di problemi sono indicate qui di seguito.

Sintesi di processi. Dati $W[n]$ e $H(z)$ si costruisce $X[n]$. Infatti, generando un certo numero di sequenze iid differenti per eccitare il sistema, si possono sintetizzare altrettante realizzazioni del processo $X[n]$ utilizzando gli stessi parametri, ovvero lo stesso filtro $H(z)$. Questo implica che tutte le sequenze sintetizzate dal filtro avranno la stessa statistica.

Si osservi ancora che vi è una profonda differenza tra i modelli AR ed i modelli MA nel comportamento in regime transitorio. Infatti la risposta all'impulso di un modello MA si esaurisce dopo un numero finito di campioni come appare dall'equazione (2.22), mentre quella di un modello AR non si esaurisce mai. Di questo si discuterà nel paragrafo 2.3.

Analisi di processi. Dati $X[n]$ e $H(z)$ si vuole riottenere $W[n]$, attraverso un filtraggio inverso. In altre parole si fa passare $X[n]$ attraverso un filtro

$$G(z) = \frac{1}{H(z)}$$

in modo da rendere bianco il segnale. $G(z)$ è detto *filtro di analisi*.

Si osservi che il fatto che $H(z)$ sia stabile e causale non implica che $G(z)$ sia stabile e causale. Come si vedrà nel seguito è comunque sempre possibile, con opportuni artifici, trovare un filtro di analisi causale e stabile che renda bianco il processo.

In generale, comunque, il problema più serio non è il filtraggio diretto o inverso di un processo quanto la stima dei parametri del modello del filtro.

2.3 Stazionarietà, transitorio e stabilità nella rappresentazione di Wold

E' noto che un processo WSS è costituito da realizzazioni di durata infinita. Ciò pone dei seri problemi realizzativi nei procedimenti di analisi e sintesi basati sul teorema di Wold. Infatti le realizzazioni di $X[n]$ e di $W[n]$ avranno sempre degli istanti iniziali e quindi, in generale, si avranno dei fenomeni transitori che alterano le caratteristiche di stazionarietà dei processi in esame.

Un risultato generale sui processi generati come uscita di un filtro lineare al cui ingresso sia posto un rumore bianco è il seguente.

Teorema. Un processo $X[n]$ generato usando un modello riconducibile all'equazione (2.10) è stazionario se e solo se il sistema generatore è stabile e soltanto dopo l'estinzione dei transitori.

Questo risultato potrebbe apparire per qualche verso in contraddizione con l'espressione (2.7), dalla quale la stazionarietà del processo di uscita sembra garantita per qualunque $H(z)$. Va però osservato che i risultati del paragrafo 2.1.2 seguono da un'analisi su tutto l'asse temporale, da $-\infty$ a $+\infty$, nella quale quindi non vi sono gli effetti di eventuali condizioni iniziali del sistema. Limitando lo studio ad una porzione dell'intero asse temporale il problema delle condizioni iniziali emerge e con esso quello del comportamento in regime transitorio. Il comportamento in regime transitorio di un sistema è strettamente connesso con la sua stabilità. Si osservi che questo problema è molto più sentito per i modelli AR ed ARMA piuttosto che per quelli MA. Infatti per questi ultimi la stabilità è garantita dall'ipotesi che i coefficienti $b_0, \ldots, b_m$ siano finiti. Per gli altri due è necessario che tutti i poli si trovino all'interno del cerchio a raggio unitario.

Per chiarire il concetto è meglio considerare un caso particolare, come illustrato nel paragrafo 2.4.

2.4 Il problema del transitorio per un processo AR del primo ordine

Si consideri un modello AR del primo ordine, caratterizzato da una funzione di trasferimento

$$H(z) = \frac{1}{1 - d_1 z^{-1}}$$

ottenuta ponendo $a_1 = -d_1$ nella (2.19), e col vincolo $|d_1| < 1$, per ragioni di stabilità. Si dimostra facilmente che la risposta all'impulso del modello è

$$h[n] = d_1^n u[n]$$

dove $u[n]$ è un gradino tempo-discreto causale. Indicando con $X[n]$ il processo in uscita, quando all'ingresso è posto il rumore bianco $W[n]$, si può esprimere $X[n]$ nella forma

$$X[n] = W[n] + d_1 X[n-1] \tag{2.23}$$

oppure

$$X[n] = \sum_{i=0}^{+\infty} h[i] W[n-i] = \sum_{i=0}^{+\infty} d_1^i W[n-i] \tag{2.24}$$

Quindi per generare $X[n]$ occorrono infiniti campioni di $W[n]$. In pratica ciò non è possibile e quindi si ricorre a sequenze di rumore 'causali' come illustrato nel paragrafo 2.4.2.

2.4.1 Autocorrelazione e spettro asintotici

Si consideri l'autocorrelazione del processo $X[n]$. Essa può essere valutata utilizzando le equazioni (2.1) e (2.2). Dall'equazione (2.1) si può scrivere

$$r_X[l] = r_W[l] * g_h[l] = \sigma_W^2 \delta[l] * g_h[l] = \sigma_W^2 g_h[l]$$

Dalla (2.2), e ricordando che il modello generatore è causale, si può scrivere l'espressione di $g_h[l]$ valida per $l > 0$. Si ha

$$g_h[l] = \sum_{i=0}^{+\infty} h[l+i] h[i] = \sum_{i=0}^{+\infty} d_1^{l+i} d_1^i = \frac{d_1^l}{1 - d_1^2}$$

Da quest'ultima espressione e ricordando che $r_X[l]$ è una funzione pari si ottiene infine

$$r_X[l] = \sigma_W^2 \frac{d_1^{|l|}}{1 - d_1^2} \tag{2.25}$$

dove si ricorda che $l = m - n$.

Si osservi che deve valere l'ipotesi $|d_1| < 1$, altrimenti la serie non converge.

Lo spettro si ricava utilizzando la (1.22), da cui si ottiene

$$\begin{aligned} S_X(\mathrm{e}^{j2\pi f}) &= \frac{\sigma_W^2}{1-d_1^2} \sum_{i=-\infty}^{+\infty} d_1^{|i|} \mathrm{e}^{-j2\pi f i} \\ &= \frac{\sigma_W^2}{1-d_1^2} \left(\sum_{i=0}^{+\infty} d_1^i \mathrm{e}^{-j2\pi f i} + \sum_{i=-\infty}^{-1} d_1^{-i} \mathrm{e}^{-j2\pi f i} \right) \\ &= \frac{\sigma_W^2}{1-d_1^2} \left(\frac{1}{1-d_1 \mathrm{e}^{-j2\pi f}} + \frac{1}{1-d_1 \mathrm{e}^{j2\pi f i}} - 1 \right) \end{aligned}$$

Con qualche semplificazione si ottiene infine

$$S_X(\mathrm{e}^{j2\pi f}) = \frac{\sigma_W^2}{1 - 2d_1 \cos(2\pi f) + d_1^2} \tag{2.26}$$

2.4.2 Autocorrelazione con ingresso causale

Come si è detto in precedenza, nelle applicazioni pratiche relazioni come la (2.23) o la (2.24) non possono essere utilizzate, poiché richiedono l'elaborazione di infiniti termini. Si utilizza, quindi, un processo bianco "causale, nel senso che si considera un processo $W'[n]$, definito come

$$W'[n] = \begin{cases} 0 & \text{per } n < 0 \\ W[n] & \text{per } n \geq 0 \end{cases}$$

e quindi si può scrivere

$$X'[n] = \sum_{i=0}^{n} d_1^i W[n-i]$$

cioè

$$X'[n] = W[n] + d_1 W[n-1] + \ldots + d_1^n W[0]$$

dove $X'[n]$ è il processo che si ottiene quando all'ingresso è posto $W'[n]$.

Si osservi che la sequenza $X'[n]$ è un processo di Markov del primo ordine. Questo si può vedere scrivendo la relazione ingresso-uscita in forma ricorsiva; trattandosi di un sistema a un polo si otterrà un sistema del primo ordine, per il quale ciascun campione $X[n]$ in uscita dipende direttamente solo dal campione immediatamente precedente $X[n-1]$.

L'autocorrelazione $R_{X'}[n,m]$ si può esprimere come

$$R_{X'}[n,m] = \sum_{i=0}^{n}\sum_{k=0}^{m} d_1^i d_1^k \mathrm{E}\{W[n-i]W[m-k]\} \tag{2.27}$$

dove $\mathrm{E}\{W[n-i]W[m-k]\} = \sigma_W^2 \delta[n-m-i+k] \neq 0$ solo se $n-i = m-k$. Si cominci a considerare il caso $m \geq n$. Introducendo la sequenza *porta*

$$p[k] = \begin{cases} 1 & \text{per } k \in [0,m] \\ 0 & \text{altrove} \end{cases}$$

la relazione (2.27) diventa

$$R_{X'}[n,m] = \sum_{i=0}^{n} p[m+i-n] d_1^i d_1^{m+i-n} \sigma_W^2 = \sigma_W^2 \sum_{i=0}^{n} d_1^{m-n} d_1^{2i} \tag{2.28}$$

Sommando la progressione, si ottiene

$$R_{X'}[n,m] = \sigma_W^2 d_1^{m-n} \frac{1-d_1^{2(n+1)}}{1-d_1^2} \tag{2.29}$$

Questa espressione dipende sia da m che da n, pertanto il processo $X'[n]$ non è WSS. Scrivendo $m = n+l$, la $R_{X'}[n,m]$ si può riscrivere come

$$R_{X'}[n,n+l] = \sigma_W^2 d_1^l \frac{1-d_1^{2(n+1)}}{1-d_1^2} \tag{2.30}$$

E' interessante confrontare questa espressione con la (2.25). Infatti si osserva che

$$\lim_{n\to\infty} R_{X'}[n,n+l] = r_X[l] = \sigma_W^2 \frac{d_1^{|l|}}{1-d_1^2} \tag{2.31}$$

Pertanto il processo $X'[n]$ è asintoticamente stazionario ed equivalente ad $X[n]$. In pratica ciò vuol dire che, per n sufficientemente grande, $X'[n]$ si comporta[1] come $X[n]$. In altre parole si può dire che $X'[n]$ equivale a $X[n]$ una volta esuarito il transitorio.

[1] nel senso dell'autocorrelazione

2.5 Lo spettro e il sistema generatore

Dato un processo $X[n]$ e la sua scomposizione di Wold, si può esprimere lo spettro di $X[n]$ in funzione della funzione di trasferimento del sistema generatore. Dalla (2.4) si ricava

$$S_X(\mathrm{e}^{j2\pi f}) = |H(\mathrm{e}^{j2\pi f})|^2 \sigma_W^2 \tag{2.32}$$

E' evidente che lo spettro non dipende dalla fase di $H(\mathrm{e}^{j2\pi f})$, ma soltanto dal modulo. Quindi esistono diversi sistemi generatori in grado di produrre processi $X[n]$ con identiche caratteristiche spettrali. Le sequenze ottenute da tali sistemi si chiamano sequenze *spettralmente equivalenti*.

Per meglio esaminare le sequenze spettralmente equivalenti, si scriva la (2.32) nel dominio z:

$$S_X(z) = H(z)H(z^{-1})\sigma_W^2 \tag{2.33}$$

e si definisca la funzione $S_h(z) = H(z)H(z^{-1})$, detta *funzione generatrice dello spettro*.

2.5.1 Generatore di tipo FIR (MA)

Se $H(z)$ è di tipo FIR vuol dire che $H(z)$ è caratterizzato da L zeri z_i, $i = 1, \ldots, L$, dove L è la lunghezza del filtro FIR. La funzione generatrice dello spettro si può scrivere nella forma

$$S_h(z) = H(z)H(z^{-1}) = K \prod_{i=1}^{m} (z - z_i)(z^{-1} - z_i) \tag{2.34}$$

dove K è una costante. La funzione $S_h(z)$ ha zeri in z_i e in $1/z_i$; ad ogni zero $z_i = m_i \mathrm{e}^{j\phi_i}$ corrisponde uno zero $z_i' = 1/m_i \mathrm{e}^{-j\phi_i}$. Essendo $H(z)$ a coefficienti reali gli zeri sono a coppie complesse coniugate. Quindi ad ogni coppia di zeri complessi coniugati corrisponde una quaterna di zeri di $S_h(z)$ disposti come nella figura 2.1.

E' evidente che gli zeri di una funzione generatrice $S_h(z)$ possono essere riordinati in più modi, dando luogo a diversi sistemi generatori. In particolare è possibile individuare

- un sistema generatore, detto *a fase minima*, con zeri tutti compresi all'interno del cerchio di raggio unitario,

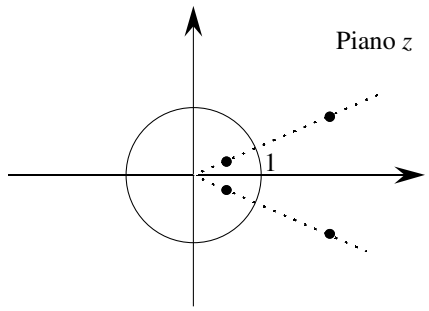

Figura 2.1. Quaterna di zeri di $S_h(z)$

- un sistema generatore, detto *a fase massima*, con zeri tutti esterni al cerchio unitario,
- diversi sistemi generatori *a fase mista*.

La ragione del termine *fase minima* è chiaramente illustrata in [1, pag. 434]. E' interessante osservare che il sistema generatore a fase minima è quello che presenta la massima energia parziale, [1, pag.436], [3, pag.45].

2.5.2 Generatore con passa-tutto e FIR (MA) a fase minima

E' possibile costruire un sistema generatore costituito da un filtro passatutto seguito da un filtro a fase minima.

Per ottenere tale sistema si procede suddividendo gli zeri di $H(z)$ in N_1 zeri d_i e N_2 zeri g_i ripettivamente interni ed esterni al cerchio di raggio unitario. Si definiscono poi le due funzioni di trasferimento: $H_d(z)$ contenente gli zeri d_i, e $H_g(z)$ contenente gli zeri g_i.

A questo punto si introducono delle coppie di zeri e poli operando come segue. Gli zeri g_i vengono "ribaltati" dentro il cerchio a raggio unitario dando luogo a N_2 nuovi zeri g_i' che conservano le fasi dei corripondenti g_i ma tali che $|g_i'| = 1/|g_i|$, come mostrato nella figura 2.2.

L'effetto di questi zeri viene compensato da altrettanti poli nelle stesse posizioni.

Utilizzando questo insieme di poli e zeri si scrive $H(z)$ nella forma

$$H(z) = AH_d(z)H'_g(z)H_g(z)\frac{1}{H'_g(z)} = AH_{min}(z)H_{ap}(z)$$

dove $H'_g(z)$ contiene gli zeri g_i,

$$H_{min}(z) = H_d(z)H'_g(z)$$

e

$$H_{ap}(z) = \frac{H_g(z)}{H'_g(z)}$$

E' immediato verificare che $H_{min}(z)$ è un filtro a fase minima. Inoltre si può dimostrare [1, pag.372] che $H_{ap}(z)$ è un filtro a modulo costante, chiamato filtro passa-tutto (*all-pass* in inglese).

E' evidente che la sagomatura dello spettro è realizzata soltanto dal filtro a fase minima.

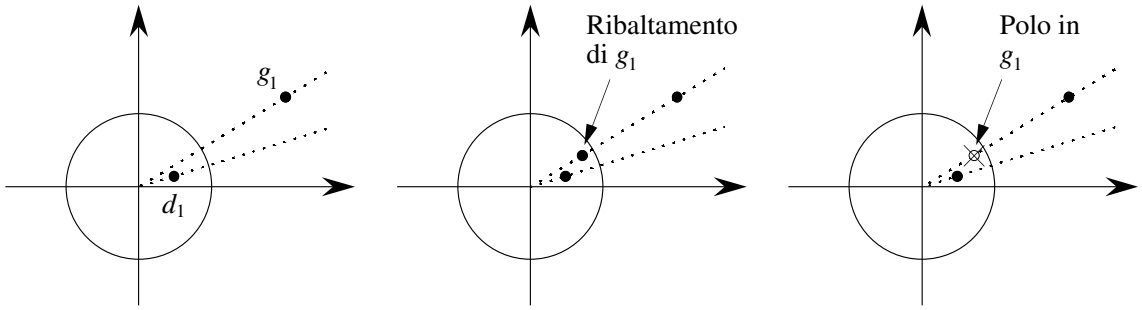

Figura 2.2. Zeri g_i di $H_g(z)$ e zeri g'_i di $H'_g(z)$

2.5.3 Generatore di tipo IIR (AR e ARMA)

Analoghe fattorizzazioni possono essere fatte con generatori di tipo IIR, operando sugli zeri. I poli infatti devono essere sempre all'interno del cerchio di raggio unitario.

Una proprietà interessante dei filtri IIR a fase minima è legata alla definizione di filtro inverso. Infatti se

$$H(z) = \frac{A(z)}{B(z)}$$

è un filtro IIR stabile a fase minima, il filtro inverso, definito come

$$H_i(z) = \frac{B(z)}{A(z)}$$

è ancora un filtro stabile a fase minima.

3

Elementi di teoria della stima

Nei capitoli precedenti si è visto come arrivare ad una descrizione dei processi nel dominio della frequenza, definendo lo spettro di potenza come trasformata di un valore atteso, cioè dell'autocorrelazione. Altri valori attesi (la media, la mutua correlazione, ecc...) sono stati introdotti e in parte interpretati. Nelle applicazioni pratiche è quasi sempre impossibile calcolare dei valori attesi, in quanto si dovrebbe conoscere il modello probabilistico dei processi in esame. Spesso i valori attesi devono essere *stimati* a partire da un certo numero di campioni di una realizzazione del processo.

Si pone quindi il problema dell'attendibilità della stima effettuata. Questo problema è affrontato dalla *teoria della stima*, di cui verranno dati alcuni cenni in questo capitolo.

Si osservi che la teoria della stima non si applica soltanto ai valori attesi, ma in generale tratta del problema di attribuire dei valori (reali o complessi) ad un insieme di dati, sulla base di un insieme di misure.

3.1 Stima e stimatore

Si supponga di volere stimare un parametro y_v a partire da un insieme di dati misurati $x_N[n]$ $(n = 0, 1, \ldots, N-1)$, nell'ipotesi che $x_N[n]$ rappresenti un segmento temporale di una singola realizzazione, $x[n]$, di un processo casuale a tempo discreto $X[n]$.

Il parametro y_v può essere di tipo casuale oppure di tipo deterministico, come illustrato nell'esempio che segue.

Si supponga di avere una sequenza di N campioni, $x_N[n]$, di un segnale vocale descritto dal processo casuale tempo continuo $X(t)$. Se campioni sono stati estratti con passo di campionamento T_s in un intervallo temporale $[0, (N-1)T_s]$, si ha $x_N[n] = x(nT_s)$, per ogni $n \in [0, N-1]$, dove con $x(t)$ si intende una specifica realizzazione del processo $X(t)$. Si supponga che il parametro da stimare y_v sia il valore assunto da un campione successivo agli N campioni prelevati. Esso potrebbe essere, ad esempio, il campione $(N+1)$-esimo, per cui si avrebbe $y_v = X(NT_s)$.

Si osservi che in questo caso y_v è una variabile casuale.

Se il parametro da stimare y_v fosse invece il valore medio del processo $X(t)$, esso sarebbe di tipo deterministico.

In generale, per stimare y_v si costruisce la funzione

$$y = h(x[n]) \tag{3.1}$$

dove $h(.)$ rappresenta una funzione dei dati misurati e y è una *stima* del valore "vero y_v. Nella (3.1) si preferisce omettere il pedice N ed utilizzare la realizzazione $x[n]$ invece di $x_N[n]$, in quanto è possibile inglobare nella definizione della funzione $h(.)$ il troncamento della $x[n]$ nella finestra $[0, N-1]$. E' evidente che y può essere visto come uno dei possibili valori assunti della variabile casuale

$$Y = h(X[n])$$

chiamata *stimatore*.

Una notazione usata molto di frequente per indicare una stima di un parametro y_v è $\hat{y}_v$; lo stimatore in questo caso si può indicare col simbolo $\hat{Y}_v$[1].

3.2 Qualità degli stimatori

Per formulare un giudizio sulla "qualità di uno stimatore vengono introdotti alcuni parametri idonei a descrivere sostanzialmente le seguenti caratteristiche sulle stime effettuate.

[1] Questa regola semplice che associa una lettera minuscola alla stima e la lettera maiuscola allo stimatore non potrà essere sempre seguita, poiché, a volte, alcuni parametri sono tipicamente indicati con lettere greche oppure con simboli speciali. In questi casi si provvederà caso per caso ad introdurre una notazione non ambigua.

- La capacità di approssimare il valore vero y_v che interessa valutare.
- La variabilità della stima y al variare della realizzazione considerata.
- Il modo in cui lo stimatore converge ad un valore limite quando N tende all'infinito (in questo caso si parla di comportamento asintotico).

I parametri che vengono utilizzati nella teoria della stima per valutare da questi punti di vista la qualità di uno stimatore, dato un certo modello statistico di riferimento, sono i seguenti.

La polarizzazione (in inglese *bias*) descrive la prima caratteristica. Si dice che uno stimatore Y di uno scalare y_v è *non polarizzato* (o *centrato*) se

$$\mathrm{E}\{Y\} = y_v \qquad \text{per qualsiasi } N$$

Se invece questa condizione è vera solo quando N diventa molto grande, il che ovviamente è molto meno restrittivo, allora si dice che Y è uno stimatore di y_v *asintoticamente non polarizzato* (o *centrato*). Si ha allora solo

$$\lim_{N\to\infty} \mathrm{E}\{Y\} = y_v$$

Si chiama *polarizzazione* di una stimatore Y il parametro

$$B(Y) = \mathrm{E}\{Y\} - y_v \tag{3.2}$$

La varianza σ_Y^2 della variabile casuale Y, o la sua radice, è il parametro che si usa per descrivere la seconda caratteristica. Si ha

$$\sigma_Y^2 = \mathrm{E}\{(Y - \mathrm{E}\{Y\})^2\}$$

Anche qui si parla di *varianza asintotica dello stimatore* per indicare il limite a cui tende σ_Y^2 quando N tende all'infinito.

La varianza dello stimatore verrà indicata anche col simbolo $\mathrm{Var}\{Y\}$.

La consistenza è una qualità che caratterizza il comportamento di alcuni stimatori per N tendente all'infinito. Si dice che uno stimatore è consistente se è asintoticamente centrato e la sua varianza asintotica è nulla[2].

Si osservi che uno stimatore centrato non è necessariamente consistente.

[2] Altri autori (si veda ad esempio [11, pag.271]) definiscono la consistenza utilizzando il concetto di *convergenza in probabilità*

L'efficienza descrive invece le caratteristiche di convergenza degli stimatori centrati per N finito. Si dice che uno stimatore centrato Y_1 di un parametro scalare y_v è più *efficiente* di un altro stimatore centrato Y_2 se

$$\sigma^2_{Y_1} \leq \sigma^2_{Y_2} \qquad \text{per qualsiasi } N \tag{3.3}$$

Per valutare l'efficienza di uno stimatore centrato si usa generalmente la *disuguaglianza di Cramèr-Rao* [8, pag. 47] che fornisce un limite inferiore per la varianza di un qualsiasi stimatore centrato.

3.2.1 L'errore quadratico medio (MSE)

Un parametro importante sul quale sono basate molte stime è l'*errore quadratico medio* (MSE dall'inglese *mean squared error*), definito come

$$\mathrm{MSE}(Y) = \mathrm{E}\{|Y - y_v|^2\} \tag{3.4}$$

E' facile dimostrare che

$$\mathrm{MSE}(Y) = \sigma^2_Y + B^2(Y)$$

da cui appare evidente che MSE(Y) è un parametro che tiene conto sia della varianza che della polarizzazione dello stimatore. Per tale ragione viene spesso utilizzato come punto di partenza per ottenere delle stime adeguate di y_v.

3.3 Stima della media

Si ricordi la definizione di media data nel paragrafo 1.2.1. Si è visto che μ_n, definito dall'espressione (1.1), può essere messo in relazione con la media aritmetica dei valori assunti, indicati con $x^{(i)}[n]$, dalla variabile casuale $X[n]$ in N esperimenti casuali successivi. In particolare si era visto che μ_n è il limite per $N \to \infty$ della frequenza relativa

$$\mu_{n,\,N} = \frac{1}{N} \sum_{i=0}^{N-1} x^{(i)}[n] \tag{3.5}$$

Questa espressione non può essere utilizzata nelle condizioni ipotizzate in questo capitolo, cioè nell'ipotesi di conoscere un segmento temporale di un'*unica* realizzazione $x[n]$ del processo $X[n]$. Infatti la frequenza relativa è definita su diverse

realizzazioni di uno stesso campione del processo casuale in esame, mentre qui si hanno alcuni campioni di una realizzazione dello stesso processo. Si può dire che l'uso dell'espressione (3.5) presuppone un'analisi per "verticali" del processo, ovvero con n costante ma diverse realizzazioni, mentre in questo caso l'unico tipo di analisi effettuabile non può che essere per "orizzontali", ovvero con un'unica realizzazione ed n variabile in un dato intervallo.

Tuttavia uno stimatore, con un'espressione simile alla (3.5), può essere introdotto nel caso di processo *stazionario*. In queste condizioni è noto che $\mu_n = \mu$. Si costruisca quindi la stima

$$\hat{m}_N = \frac{1}{N} \sum_{n=0}^{N-1} x[n] \tag{3.6}$$

E' evidente che la (3.6) ha la struttura delle *medie temporali* viste nel paragrafo 1.4. A tale stima è associato lo stimatore $\hat{\mu}_N$

$$\hat{\mu}_N = \frac{1}{N} \sum_{n=0}^{N-1} X[n] \tag{3.7}$$

Per valutare la qualità di tale stimatore si possono valutarne la polarizzazione e la varianza.

Polarizzazione. E' molto semplice verificare che

$$\mathrm{E}\{\hat{\mu}_N\} = \frac{1}{N} \sum_{n=0}^{N-1} \mathrm{E}\{X[n]\} = \mu \tag{3.8}$$

da cui segue che lo stimatore (3.7) è uno stimatore centrato.

Varianza. La varianza della stima è data da

$$\sigma^2_{\hat{\mu}_N} = E\left\{ \left[\frac{1}{N} \sum_{n=0}^{N-1} X[n] \right]^2 \right\} - \mu^2$$

Da semplici passaggi si ricava

$$\sigma^2_{\hat{\mu}_N} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} C_X[i,k] \tag{3.9}$$

che diventa, per processi WSS,

$$\sigma^2_{\hat{\mu}_N} = \frac{1}{N^2} \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} C_X[i-k] \tag{3.10}$$

Si può dimostrare [22] che

$$\lim_{l \to \infty} C_X[l] = 0 \tag{3.11}$$

è una condizione sufficiente affinché $\sigma^2_{\hat{\mu}_N} \to 0$ per $N \to \infty$.

Si può concludere dicendo che la (3.7) è uno stimatore consistente del valore atteso di un processo $X[n]$ WSS se vale la (3.11). In questo caso si può scrivere che

$$\lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} x[n] = \mu \tag{3.12}$$

Un processo WSS per il quale vale la (3.12) è detto *ergodico per la media*.

Molti processi fisici stazionari sono anche ergodici. Per tale ragione, nelle applicazioni ingegneristiche, viene comunemente assunta l'ergodicità per la media. Un ragionamento analogo vale anche per l'ergodicità dei momenti del secondo ordine.

Si ponga attenzione sul fatto che comunque la stazionarietà di un processo non ne assicura l'ergodicità. Peraltro in questa sede si è discusso solo dell'ergodicità di processi stazionari mentre la teoria dell'ergodicità è assai più generale e riferibile a processi casuali non necessariamente stazionari. La teoria dei processi ergodici si presenta in modo assai diverso se non si considera l'ipotesi di stazionarietà. Alcuni autori [10], per questa ragione, parlano di *processi strettamente ergodici* quando si riferiscono a processi sia ergodici sia stazionari. La complessità di questa materia non permette di addentrersi in questa sede in ulteriori dettagli per i quali si rimanda il lettore interessato a [10].

3.4 Stima dell'autocorrelazione

In modo analogo a quanto fatto per la stima della media, l'autocorrelazione $r_X[l] = \mathrm{E}\{X[n]X[n+l]\}$ di un processo WSS può essere messa in relazione con

la frequenza relativa

$$r_X[l] = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} x^{(i)}[n]x^{(i)}[n+l]$$

dove $x^{(i)}[n]$ e $x^{(i)}[n+l]$ sono la coppia di valori assunti dalle due variabili casuali $X[n]$ e $X[n+l]$ all'i-esimo esperimento casuale.

Questa espressione per le stesse ragioni addotte in precedenza, non può essere utilizzata nelle condizioni ipotizzate in questo capitolo, cioè nell'ipotesi di conoscere un segmento temporale di un'*unica* realizzazione $x[n]$ del processo $X[n]$.

In pratica si devono usare degli stimatori che abbiano la struttura di medie temporali.

3.4.1 Primo stimatore

Una stima molto usata è la seguente

$$\hat{r}_N[l] = \frac{1}{N-|l|} \sum_{n=0}^{N-|l|-1} x[n]x[n+|l|] \tag{3.13}$$

a cui è associato lo stimatore

$$\hat{R}_N[l] = \frac{1}{N-|l|} \sum_{n=0}^{N-|l|-1} X[n]X[n+|l|] \tag{3.14}$$

Per valutare la qualità di tale stimatore se ne possono valutare la polarizzazione e la varianza.

Polarizzazione. E' molto semplice verificare che

$$\mathrm{E}\{\hat{R}_N[l]\} = \frac{1}{N-|l|} \sum_{n=0}^{N-|l|-1} \mathrm{E}\{X[n]X[n+|l|]\} = r_X[l]$$

da cui segue che la (3.14) è uno stimatore centrato.

Varianza. Prima di valutare la varianza si possono cominciare a fare alcune considerazioni qualitative sullo stimatore introdotto. Osservando lo stimatore (3.13) si vede che, al crescere di $|l|$, diminuisce il numero dei campioni

di $x[n]$ che determinano la stima dell'autocorrelazione. Pertanto la stima diventa sempre meno affidabile al crescere di $|l|$. Al limite per $|l| = N$ la stima è calcolata con un singolo addendo della sommatoria (3.13). Questo effetto può essere quantificato per mezzo della varianza dello stimatore. Si può dimostrare [23] che per processi gaussiani ed N sufficientemente grande essa vale

$$\mathrm{Var}\{\hat{R}_N[l]\} \approx \frac{1}{N-|l|} \sum_{i=-\infty}^{\infty} \{r_X^2[l] + r_X[i-l]r_X[i+l]\}$$

Per $N \to \infty$, $\mathrm{Var}\{\hat{R}_N[l]\} \to 0$, quindi lo stimatore è consistente. D'altra parte, per N finito, al crescere di $|l|$ la varianza della stima aumenta. In generale per contenere la varianza dello stimatore ed avere stime affidabili di $r_x[l]$ è opportuno assicurare che $|l| \ll N$. Una regola empirica è ritenere la stima dell'autocorrelazione adeguata per per valori di l nell'intervallo $|l| < N/4$.

Si noti infine che tale stimatore non garantisce più che la matrice di autocorrelazione $\hat{\mathbf{R}}_X$, costruita con le autocorrelazioni stimate attraverso l'espressione (3.13), sia semi-definita positiva. Questo effetto va trattato con cautela quando è necessario invertire, specie per via numerica, la matrice delle autocorrelazioni stimate $\hat{\mathbf{R}}_X$, come capita in numerose applicazioni.

3.4.2 Secondo stimatore

Un'altra stima molto usata è

$$\hat{r}'_N[l] = \frac{1}{N} \sum_{n=0}^{N-|l|-1} x[n]x[n+|l|] \tag{3.15}$$

a cui è associato lo stimatore

$$\hat{R}'_N[l] = \frac{1}{N} \sum_{n=0}^{N-|l|-1} X[n]X[n+|l|] \tag{3.16}$$

Si osservi che con questo stimatore la matrice delle autocorrelazioni stimate $\hat{\mathbf{R}}'_X$ è semi-definita positiva, quindi meno critica da trattare per via numerica.

Per valutare la qualità di tale stimatore se ne possono valutare la polarizzazione e la varianza.

Polarizzazione. E' molto semplice verificare che

$$\mathrm{E}\{\hat{R}'_N[l]\} = \frac{N-|l|}{N} r_X[l]$$

da cui segue che l'espressione (3.16) è propria di uno stimatore non centrato. Per $N \to \infty$ lo stimatore diventa centrato essendo

$$\lim_{N\to\infty} \mathrm{E}\{\hat{R}'_N[l]\} = r_X[l]$$

In pratica per $|l| \ll N$ la polarizzazione è trascurabile.

Varianza. Come nel caso del primo stimatore, anche ora al crescere di $|l|$, diminuisce il numero dei campioni di $x[n]$ che determinano la stima dell'autocorrelazione. Pertanto la stima diventa sempre meno significativa al crescere di $|l|$. Al limite per $|l| = N$ la stima è costituita da un singolo addendo della sommatoria (3.15).

Analogamente al caso del primo stimatore, si può dimostrare che per processi gaussiani e per N sufficientemente grande la varianza dello stimatore è data da

$$\mathrm{Var}\{\hat{R}'_N[l]\} \approx \frac{1}{N} \sum_{i=-\infty}^{\infty} \{r_X^2[l] + r_X[i-l] r_X[i+l]\}$$

Si osservi che per $N \to \infty$, $\mathrm{Var}\{\hat{R}_N[l]\} \to 0$, quindi lo stimatore è consistente. Confrontando con lo stimatore precedente si vede immediatamente che in questo caso la varianza è minore.

E' interessante confrontare i due stimatori. Il primo è centrato, ma il secondo ha un errore quadratico medio (MSE) inferiore. Pertanto in molti casi si preferisce il secondo stimatore anche se è polarizzato.

Come per il primo stimatore, il ritardo $|l|$ deve essere piccolo rispetto a N per avere una bassa varianza.

4

Stima di parametri

Nei paragrafi 3.3 e 3.4 del capitolo 3 sono stati introdotti degli stimatori le cui caratteristiche di consistenza sono state verficate a posteriori. La teoria della stima insegna a ricavare gli stimatori sulla base di criteri stabiliti a priori. Si può dire brevemente che i criteri di stima sono basati sulle distribuzioni di probabilità delle variabili in gioco, oppure, più semplicemente, su alcuni valori attesi, in particolare sull'autocorrelazione. In questo capitolo verranno dati alcuni cenni sui metodi di stima basati sulla conoscenza della statistica delle variabili in gioco. Per avere ulteriori dettagli su questo complesso tema il lettore interessato può consultare [3, 5, 6, 8].

4.1 Descrizione vettoriale dei dati del problema

Si ritorni di nuovo nelle condizioni descritte all'inizio del capitolo 3, supponendo di avere un vettore di dati misurati

$$\mathbf{x}_m = (x[0],x[1],\cdots,x[N-1])$$

Tale insieme di dati può essere visto come un'istanza di un processo casuale

$$\mathbf{X} = (X[0],X[1],\cdots,X[N-1])$$

a cui è associato uno spazio vettoriale $\mathbf{x}$, che rappresenta l'insieme dei possibili valori assunti da $\mathbf{X}$.

A partire dai dati misurati si vuole stimare un vettore di M parametri

$$\mathbf{a}_v = (a_{v,1},a_{v,2},\cdots,a_{v,M})$$

per mezzo di un vettore di stime

$$\hat{\mathbf{a}} = (\hat{a}_1, \hat{a}_2, \cdots, \hat{a}_M)$$

a cui è associato uno stimatore $\hat{\mathbf{A}}$. In modo più esplicito, il vettore $\hat{\mathbf{a}}$ si può indicare con la notazione $\hat{\mathbf{a}}(\mathbf{x}_m)$, così da mettere in evidenza la sua dipendenza dai dati misurati. Nei metodi illustrati nel seguito il vettore $\hat{\mathbf{a}}$ viene visto come l'incognita di un problema di ottimizzazione rispetto ad un criterio assegnato. In questi casi si preferisce adottare la notazione più snella $\hat{\mathbf{a}}$.

I parametri da stimare (cioeè il vettore $\mathbf{a}_v$) possono essere di tipo deterministico o casuale. In generale si associa al vettore $\mathbf{a}_v$ un vettore di variabili casuali

$$\mathbf{A} = (A_1, A_2, \cdots, A_M)$$

e si indica con $\mathbf{a} = (a_1, a_2, \cdots, a_M)$ lo spazio dei possibili valori assunti da $\mathbf{A}$. Se i parametri non sono di tipo casuale possono essere modellati comunque come variabili casuali con densità di probabilità uniforme.

4.1.1 Descrizione in termini di densità di probabilità

La descrizione in termini di densità di probabilità dei dati in esame si può ottenere dalla conoscenza della densità di probabilità congiunta dei due vettori casuali, $\mathbf{X}$ e $\mathbf{A}$, cioè

$$f_{\mathbf{X},\mathbf{A}}(\mathbf{x},\mathbf{a})$$

E' noto infatti che da tale distribuzione si possono ottenere sia le densità marginali, sia le densità di probabilità condizionate. In particolare giocano un ruolo importante le due densità:

1. *a priori*

$$f_{\mathbf{X}|\mathbf{A}=\mathbf{a}}(\mathbf{x}|\mathbf{A} = \mathbf{a})$$

2. e *a posteriori*

$$f_{\mathbf{A}|\mathbf{X}=\mathbf{x}}(\mathbf{a}|\mathbf{X} = \mathbf{x})$$

   Quest'ultima viene valutata generalmente in corrispondenza dei dati misurati, cioè

$$f_{\mathbf{A}|\mathbf{X}=\mathbf{x}_m}(\mathbf{a}|\mathbf{X} = \mathbf{x}_m)$$

Come è noto, la densità di probabilità congiunta si può esprimere a partire dalle densità di probabilità condizionate nella forma

$$f_{\mathbf{X},\mathbf{A}}(\mathbf{x},\mathbf{a}) = \begin{cases} f_{\mathbf{X}|\mathbf{A}=\mathbf{a}}(\mathbf{x}|\mathbf{A}=\mathbf{a}) f_{\mathbf{A}}(\mathbf{a}) \\ f_{\mathbf{A}|\mathbf{X}=\mathbf{x}}(\mathbf{a}|\mathbf{X}=\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) \end{cases}$$

A questo punto ci chiediamo come utilizzare tutte le informazioni statistiche sui vettori $\mathbf{X}$ e $\mathbf{A}$, per arrivare ad una ragionevole stima di $\mathbf{a}_v$.

Riassumendo abbiamo alcuni elementi del mondo reale (i dati misurati e i parametri da stimare) e modelli probabilistici degli elemenbti del mondo reale, come illustrato in sintesi nella tabella 4.2. Come vedremo, i metodi della teoria della stima portano a mescolare i due mondi in modo da formulare una stima $\hat{\mathbf{a}}$ basata sui modelli probabilistici.

Mondo reale	Mondo dei modelli
$\mathbf{a}_v$	$\mathbf{A}$ $f_{\mathbf{A}}(\mathbf{a})$
$\mathbf{x}_m$	$\mathbf{X}$ $f_{\mathbf{X},\mathbf{A}}(\mathbf{x},\mathbf{a})$ $f_{\mathbf{X}\|\mathbf{A}=\mathbf{a}}(\mathbf{x}\|\mathbf{A}=\mathbf{a})$ $f_{\mathbf{A}\|\mathbf{X}=\mathbf{x}}(\mathbf{a}\|\mathbf{X}=\mathbf{x})$
$\hat{\mathbf{a}}$ (stima)	$\hat{\mathbf{A}}$ (stimatore)
	oppure
$\hat{\mathbf{a}}(\mathbf{x}_m)$	$\hat{\mathbf{a}}(\mathbf{X})$ (stimatore)

Tabella 4.1. Il mondo reale e il suo modello

4.2 Esempio di riferimento

Nel seguito faremo sempre riferimento ad un esempio semplice che ci consenta di interpretare i metodi di stima in modo intuitivo. Consideriamo in particolare un processo casuale del tipo

$$X[n] = a_v + W[n] \tag{4.1}$$

dove a_v è il parametro da stimare a partire da un insieme di dati misurati $\mathbf{x}_m$ e $W[n]$ è una sequenza bianca a media nulla.

Supponiamo inoltre che il parametro sconosciuto a_v sia in realtà un'istanza di una variabile casuale A con una ddp nota. Nel seguito faremo degli esempi con

ddp di A di tipo bimodale per mettere in evidenza alcuni aspetti dei metodi di stima descritti.

Nella figura 4.1 è illustrato un esempio molto semplice, che ci consente di fare alcune considerazioni intuitive, per introdurre l'argomento. La ddp del processo di rumore è illustrata nella figura 4.1(c). Il vettore dei dati misurati $\mathbf{x}_m$ è illustrato nella figura 4.1(a). Come si vede tutti i dati sono compresi in una fascia ristretta intorno al valore 1.2, pertanto ci aspettiamo che il valore del parametro sia vicino al valore 1.2. Il parametro sconosciuto a_v è in realtà un'istanza di una variabile casuale A con la ddp asimmetrica mostrata nella figura 4.1(b). Dalla figura vediamo che il valore $A \sim 1.2$ non è molto verosimile. Ci chiediamo a questo punto quanto peso vogliamo che abbia nella nostra stima questo fatto.

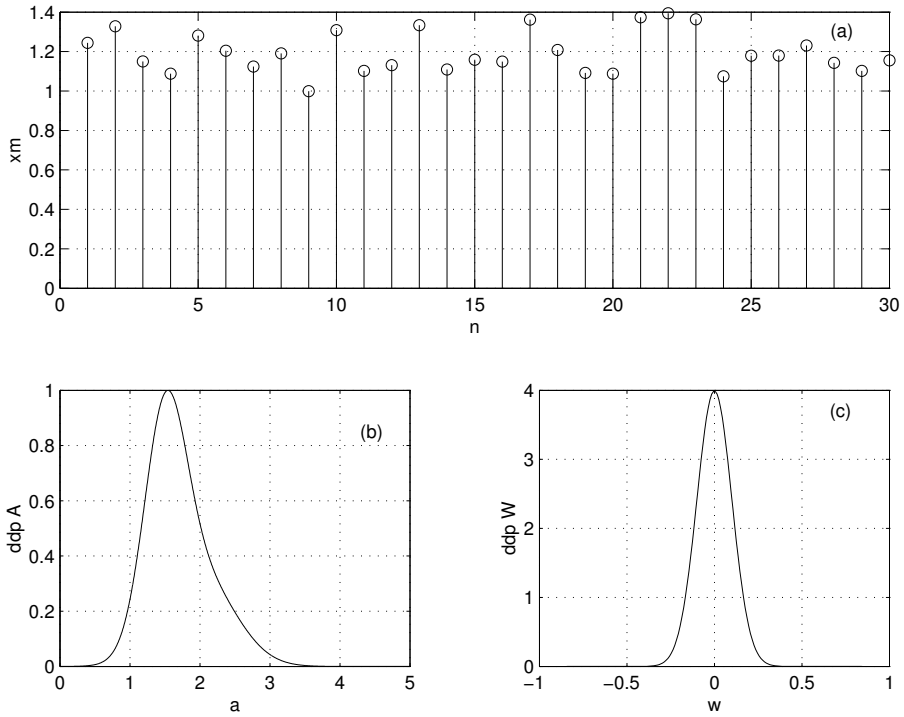

Figura 4.1. (a): vettore $\mathbf{x}_m$. (b): ddp (asimmetrica) di A. (c): ddp del rumore

Per dare una risposta a questo quesito dobbiamo prima introdurre i metodi classici della stima di parametri. Nel paragrafo 4.3 verranno descritti i metodi della stima di *Bayes* e nel paragrafo 4.4 il metodo della stima a massima verosimiglianza. Alla fine di questa descrizione potremo fare qualche commento sull'esempio presentato in questo paragrafo.

4.3 Stima bayesiana

La stima bayesiana è basata sul concetto di attribuire una funzione di costo a tutti i possibili valori del vettore stimato $\hat{\mathbf{a}}$ e a tutti i possibili valori del vettore dei parametri A. Si definisce una funzione del tipo

$$C(\hat{\mathbf{a}},\mathbf{A})$$

In tale funzione la stima $\hat{\mathbf{a}}$ compare come una variabile incognita che si vuole valutare, mentre il vettore dei parametri interviene con tutti i suoi possibili valori e con la sua statistica. Per tale motivo è modellato con la variabile casuale $\mathbf{A}$ e con la sua ddp (in linea di principio estratta dalla ddp congiunta $f_{\mathbf{X},\mathbf{A}}(\mathbf{x},\mathbf{a})$). L'idea è quella di assegnare all'incognita $\hat{\mathbf{a}}$ il valore che minimizza il costo medio (chiamato *funzione di rischio*), condizionato all'evento $\mathbf{X} = \mathbf{x}_m$.

Ad esempio si potrebbe porre

$$C(\hat{\mathbf{a}},\mathbf{A}) = (\mathbf{A} - \hat{\mathbf{a}})^2$$

In questo caso il valore di $\hat{\mathbf{a}}$ che minimizza il costo medio corrisponde al valore che minimizza lo scarto quadratico medio.

Nota la ddp congiunta, la funzione di rischio si scrive come

$$\begin{aligned}
\mathcal{R}(\hat{\mathbf{a}}) &= \mathrm{E}\{C(\hat{\mathbf{a}},\mathbf{A})\} \\
&= \int_{\mathbf{x}}\int_{\mathbf{a}} C(\hat{\mathbf{a}},\mathbf{a}) f_{\mathbf{X},\mathbf{A}}(\mathbf{x},\mathbf{a}) d\mathbf{x} d\mathbf{a} \\
&= \int_{\mathbf{x}}\int_{\mathbf{a}} C(\hat{\mathbf{a}},\mathbf{a}) f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} d\mathbf{a}
\end{aligned}$$

La strategia consiste nell'attribuire un costo elevato ai valori indesiderati della v.c. (vettoriale) $\mathbf{A}$. Dato il vettore $\mathbf{x}_m$ dei dati misurati, il rischio condizionato diventa

$$\mathcal{R}(\hat{\mathbf{a}}|\mathbf{x}_m) = \int_{\mathbf{a}} C(\hat{\mathbf{a}},\mathbf{a}) f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a}$$

e la stima bayesiana

$$\hat{\mathbf{a}}_b = \arg\min_{\hat{\mathbf{a}}} \left(\int_{\mathbf{a}} C(\hat{\mathbf{a}},\mathbf{a}) f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a} \right)$$

4.3.1 Stima MMSE

Si dice che una stima $\hat{\mathbf{a}}$ di $\mathbf{a}_v$ è di tipo MMSE (*Minimum Mean–Squared Error*) se la funzione di costo è

$$C(\hat{\mathbf{a}},\mathbf{A}) = |\mathbf{A} - \hat{\mathbf{a}}|^2$$

e la stima è di tipo bayesiano. La funzione di rischio si scrive come

$$\begin{aligned} \mathcal{R}_{\mathrm{MMSE}}(\hat{\mathbf{a}}|\mathbf{x}_m) &= \mathrm{E}[|\mathbf{A} - \hat{\mathbf{a}}|^2|\mathbf{x}_m] \\ &= \int_{\mathbf{a}} |\mathbf{a} - \hat{\mathbf{a}}|^2 f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a} \end{aligned} \tag{4.2}$$

e la stima MMSE è il vettore $\hat{\mathbf{a}}_{\mathrm{MMSE}}$ che minimizza $\mathcal{R}_{\mathrm{MMSE}}(\hat{\mathbf{a}}|\mathbf{x}_m)$. Si può dimostrare che $\hat{\mathbf{a}}_{\mathrm{MMSE}}$ è la media della ddp a posteriori $f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m)$.

Dimostrazione. Dalla (4.2) si ottiene

$$\mathcal{R}_{\mathrm{MMSE}}(\hat{\mathbf{a}}|\mathbf{x}_m) = \int_{\mathbf{a}} \mathbf{a}^2 f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a} + \int_{\mathbf{a}} \hat{\mathbf{a}}^2 f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a} - 2\int_{\mathbf{a}} \mathbf{a}\hat{\mathbf{a}} f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a}$$

Derivando rispetto a $\hat{\mathbf{a}}$ si ha

$$\frac{d}{d\hat{\mathbf{a}}}\mathcal{R}_{\mathrm{MMSE}}(\hat{\mathbf{a}}|\mathbf{x}_m) = 2\hat{\mathbf{a}} - 2\int_{\mathbf{a}} \mathbf{a} f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a}$$

La derivata si annulla per

$$\hat{\mathbf{a}} = \int_{\mathbf{a}} \mathbf{a} f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a}$$

Nel capitolo 6 il tema degli stimatori MMSE verrà approfondito, con particolare attenzione ad una classe importante di tali stimatori, detti *stimatori lineari*.

4.3.2 Stima MAP

Utilizzando una funzione di costo del tipo

$$C(\hat{\mathbf{a}},\mathbf{A}) = 1 - \delta(\hat{\mathbf{a}},\mathbf{A})$$

dove $\delta(\hat{\mathbf{a}},\mathbf{A})$ è la delta di Kronecker (vettoriale), si ottiene il rischio condizionato

$$\mathcal{R}_{\mathrm{MAP}}(\hat{\mathbf{a}}|\mathbf{x}_m) = \int_{\mathbf{a}} [1 - \delta(\hat{\mathbf{a}},\mathbf{a})] f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) d\mathbf{a} = 1 - f_{\mathbf{A}|\mathbf{X}}(\hat{\mathbf{a}}|\mathbf{x}_m)$$

a cui corrisponde la stima

$$\hat{\mathbf{a}}_{\mathrm{MAP}} = \arg\max_{\mathbf{a}} f_{\mathbf{A}|\mathbf{X}}(\mathbf{a}|\mathbf{x}_m) = \arg\max_{\mathbf{a}}[f_{\mathbf{X}|\mathbf{A}}(\mathbf{x}_m|\mathbf{a})f_{\mathbf{A}}(\mathbf{a})] \tag{4.3}$$

nota come stima MAP (massima a posteriori). Questo nome deriva dal significato che possiamo facilmente attribuire a questo tipo di stima. Dall'equazione (4.3) vediamo che tale stima corrisponde al valore più probabile di $\mathbf{A}$, quando $\mathbf{X} = \mathbf{x}_m$.

Un esempio è illustrato nella figura 4.2, dove il problema descritto nel paragrafo 4.2 è esaminato nel caso di ddp di A di tipo bimodale (curva nel riquadro alto). Ricordiamo che si tratta di stimare il parametro a_v di un processo del tipo

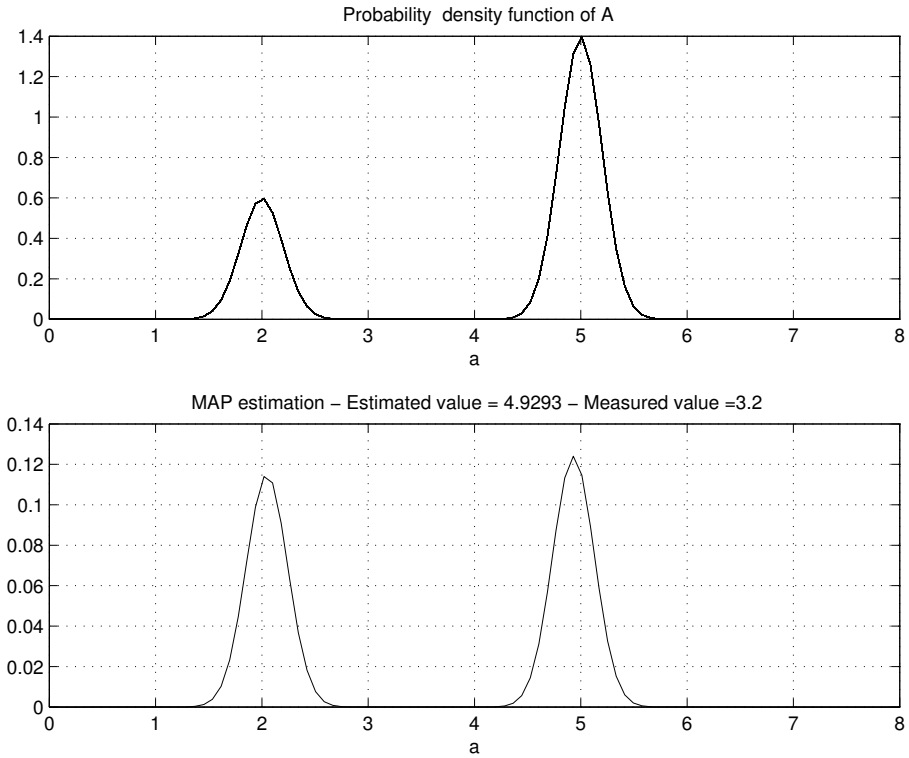

Figura 4.2. Densità di probabilità di A e densità di probabilità congiunta per $x = x_m$

$$X[n] = a_v + W[n]$$

dove $W[n]$ è un rumore a media nulla. Allo scopo di rendere leggibili i risultati della stima attraverso i grafici delle densità di probabilità, consideriamo il

caso molto semplice di avere un solo valore misurato. Pertanto il vettore $\mathbf{x}_m$ si riduce ad uno scalare x_m. Nel riquadro in basso della figura 4.2 è tracciata la $f_{X|A}(x_m|a)f_A(a)$, ricavata dalla ddp congiunta mostrata nella figura 4.3, nel caso $x_m = 3.2$. Osservando la figura 4.2 (riquadro in basso) si ottiene facilmente il valore

$$\hat{\mathbf{a}}_{\mathrm{MAP}} = 4.93 \tag{4.4}$$

ottenuto applicando la (4.3). Osserviamo che il valore stimato è molto diverso dal valore misurato. Ciò è dovuto al fatto che la ddp di A ha un peso rilevante nel metodo di stima e quindi il massimo della $f_{X|A}(x_m|a)f_A(a)$ dipende fortemente dalla distribuzione della A. Poiché nel nostro modello (dove $X[n] = a_v + W[n]$) a_v è il valore medio di $X[n]$ potremmo essere tentati di scegliere $\hat{a} = x_m = 3.2$, rinunciando al criterio MAP. Come vedremo questa scelta arbitraria corrisponde al criterio ML descritto nel paragrafo 4.4.

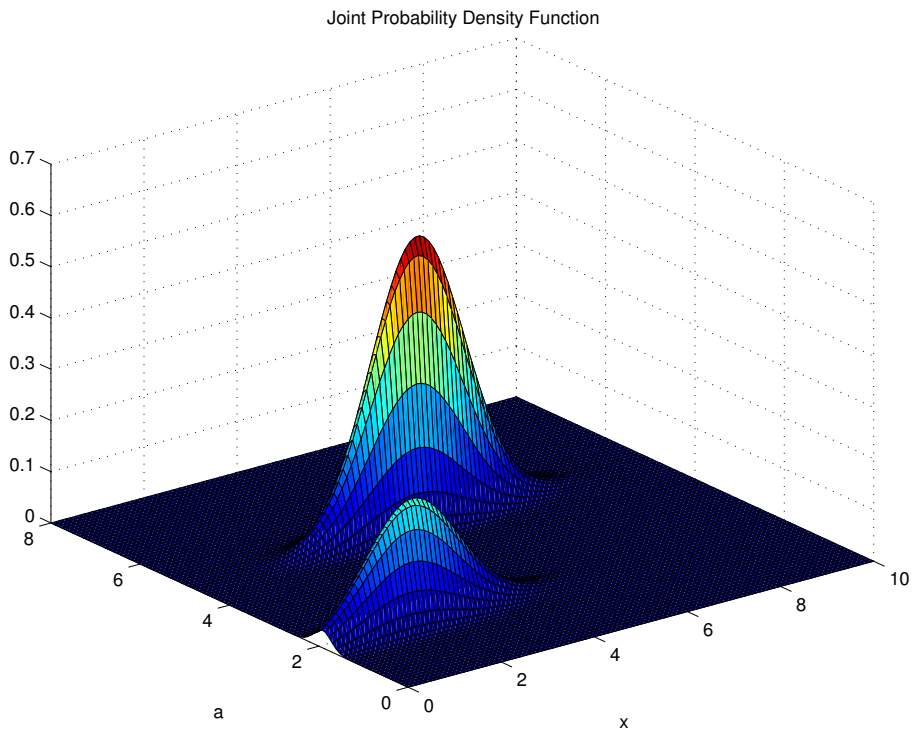

Figura 4.3. Densità di probabilità congiunta (per stima MAP)

4.4 Stima a massima verosimiglianza (ML)

Ritorniamo al metodo MAP e consideriamo il caso in cui si abbia massima incertezza su A, cioè il caso di ddp uniforme per tutti i valori significativi di $f_{\mathbf{X}|\mathbf{A}}(\mathbf{x}_m|\mathbf{a})$ In questo caso la stima MAP diventa

$$\hat{\mathbf{a}}_{\mathrm{MAP}} \Rightarrow \arg\max_{\mathbf{a}}[f_{\mathbf{X}|\mathbf{A}}(\mathbf{x}_m|\mathbf{a})] \tag{4.5}$$

La stima che si ottiene non è influenzata dalla statistica di $\mathbf{A}$ ed è usata in molte applicazioni.

Formalmente si definisce una funzione di verosimiglianza come

$$l(\mathbf{a},\mathbf{x}_m) = f_{\mathbf{X}|\mathbf{A}}(\mathbf{x}_m|\mathbf{a})$$

e si introduce un vettore $\hat{\mathbf{a}}_{\mathrm{ML}}$ cha massimizza tale funzione, cioè

$$\hat{\mathbf{a}}_{\mathrm{ML}} = \arg\max_{\mathbf{a}} f_{\mathbf{X}|\mathbf{A}}(\mathbf{x}_m|\mathbf{a}) \tag{4.6}$$

Tale stima è detta a massima verosimiglianza (ML *Maximum Likelihood*). In genere si preferisce lavorare con la *funzione di verosimiglianza logaritmica*

$$L(\mathbf{a},\mathbf{x}_m) = \ln f_{\mathbf{X}|\mathbf{A}}(\mathbf{x}_m|\mathbf{a})$$

per semplificare gli aspetti computazionali.

Per chiarire le differenze tra stima MAP e ML si consideri di nuovo l'esempio del paragrafo 4.2, cioè il processo

$$X[n] = a_v + W[n]$$

di cui si vuole stimare il parametro a_v. Come nel caso visto nel paragrafo 4.3.2 si dispone soltanto di un dato misurato $x_m = 3.2$. La ddp di A è di nuovo la bimodale dell'esempio precedente, riportata, per comodità, nel riquadro in alto della figura 4.4. Nel riquadro in basso della figura 4.4 è tracciata la $f_{X|A}(x_m|a)$, ricavata dalla ddp condizionata mostrata nella figura 4.5, nel caso di $x_m = 3.2$. E' evidente che in questo caso la curva non è influenzata dalla ddp di A. Osservando la figura 4.4 (riquadro in basso) si ottiene facilmente il valore

$$\hat{a}_{\mathrm{ML}} = 3.2 \tag{4.7}$$

ottenuto applicando la (4.6). Osserviamo che il valore stimato è uguale al valore misurato. Ciò è dovuto al fatto che la ddp di A non ha alcun peso nel metodo di

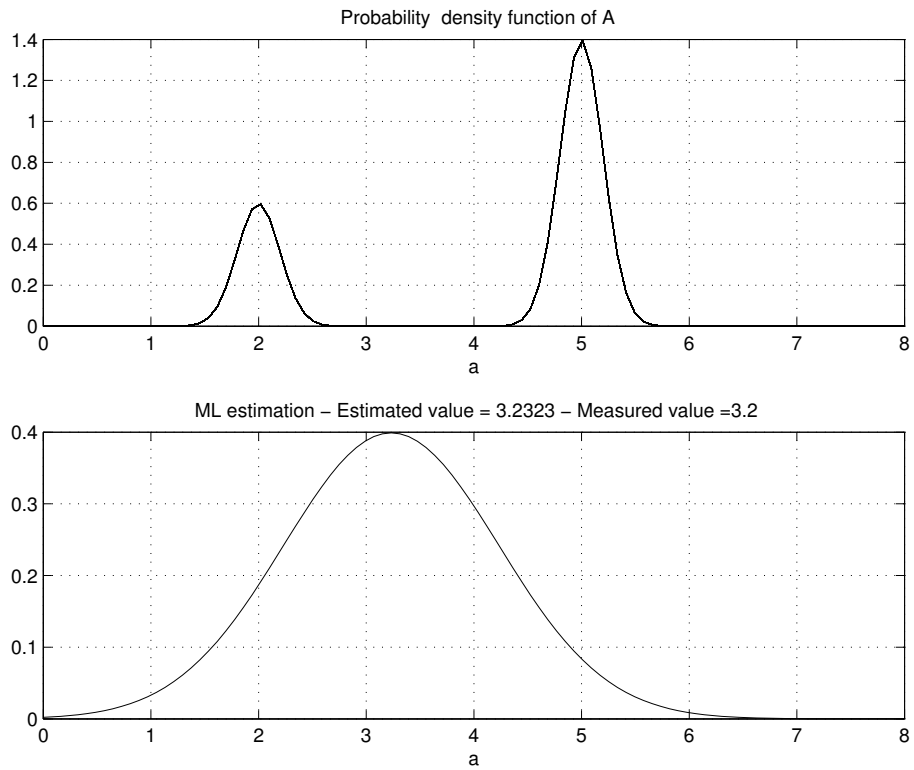

Figura 4.4. Densità di probabilità di A e funzione di verosimiglianza

stima. Pertanto la scelta intuitiva $\hat{a} = x_m = 3.2$ corrisponde al valore della stima ML.

La stima ML si applica nei casi in cui il valore vero da stimare è un parametro deterministico sconosciuto, ma sempre uguale in tutti i dati misurati. Nell'esempio del paragrafo 4.2 ci troviamo in questa situazione: in ogni dato misurato è presente sempre lo stesso valore di a_v, mentre il rumore (che è l'elemento di disturbo) varia sempre.

Tale metodo di stima si usa anche quando il parametro è casuale, ma la sua ddp è sconosciuta.

A questo punto possiamo aggiornare la tabella 4.2, come indicato nella tabella **??**

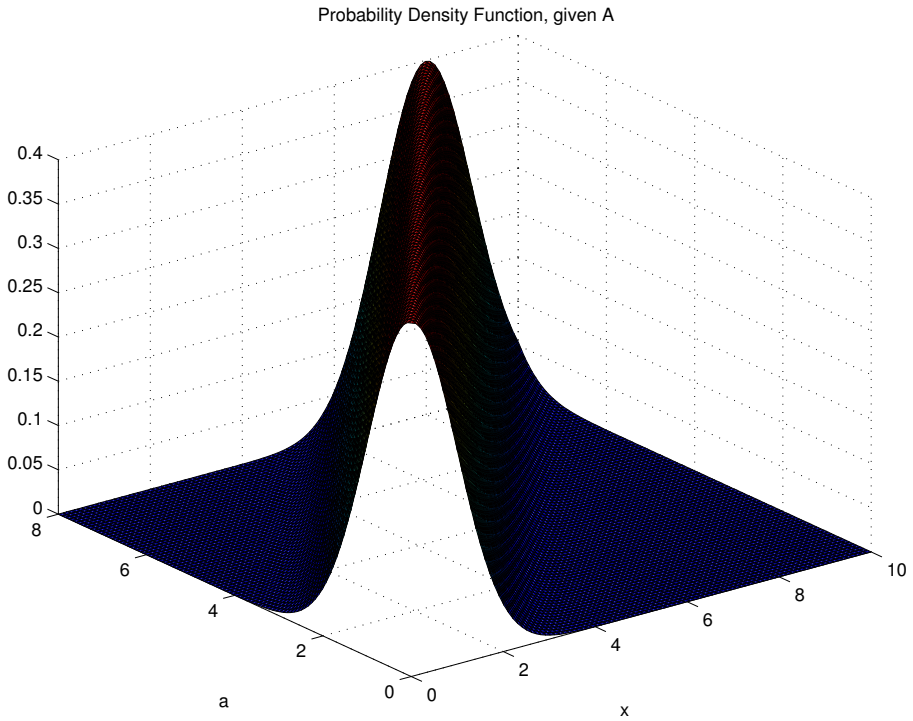

Figura 4.5. Densità di probabilità condizionata (per stima ML)

Mondo reale	Mondo dei modelli
$\mathbf{a}_v$	$\mathbf{A}$ $f_{\mathbf{A}}(\mathbf{a})$
$\mathbf{x}_m$	$\mathbf{X}$ $f_{\mathbf{X},\mathbf{A}}(\mathbf{x},\mathbf{a})$ $f_{\mathbf{X}\|\mathbf{A}=\mathbf{a}}(\mathbf{x}\|\mathbf{A}=\mathbf{a})$ $f_{\mathbf{A}\|\mathbf{X}=\mathbf{x}}(\mathbf{a}\|\mathbf{X}=\mathbf{x})$
$\hat{\mathbf{a}}$ (stima)	$\hat{\mathbf{A}}$ (stimatore)
Mondo della stima	
$\hat{\mathbf{a}}_b = \arg\min_{\hat{\mathbf{a}}} \left(\int_{\mathbf{a}} C(\hat{\mathbf{a}},\mathbf{a}) f_{\mathbf{A}\|\mathbf{X}}(\mathbf{a}\|\mathbf{x}_m) d\mathbf{a}\right)$ $\hat{\mathbf{a}}_{\mathrm{MAP}} = \arg\max_{\mathbf{a}}[f_{\mathbf{X}\|\mathbf{A}}(\mathbf{x}_m\|\mathbf{a}) f_{\mathbf{A}}(\mathbf{a})]$ $\hat{\mathbf{a}}_{\mathrm{ML}} = \arg\max_{\mathbf{a}} f_{\mathbf{X}\|\mathbf{A}}(\mathbf{x}_m\|\mathbf{a})$	

Tabella 4.2. Il mondo reale, il suo modello e le sue stime

4.5 Cramer-Rao Bound (CRB)

La varianza della stima di uno stimatore non polarizzato non può superare un certo valore limite che dipende dalla distribuzione dei dati misurati e dei parametri. Tale limite, noto come *Cramer-Rao Bound*, si esprime come

$$\mathrm{Var}[\hat{\mathbf{A}}] \geq \mathbf{J}^{-1}(\mathbf{a})$$

dove $\mathbf{J}$ è una matrice $M \times M$, detta *Matrice di informazione di Fisher*. Nel caso di stima di parametri deterministici si ha

$$\mathbf{J}_{i,j} = -\mathrm{E}[(\hat{\mathbf{A}} - \mathbf{a}_v)^2 | \mathbf{A} = \mathbf{a}_v)] = -\mathrm{E}\left(\frac{\partial^2 \ln f_{\mathbf{X},\mathbf{A}}(\mathbf{X}|\mathbf{A})}{\partial a_i \partial a_j} | \mathbf{A} = \mathbf{a}_v\right)$$

Nel caso di stima di un vettore di variabili casuali si ha

$$\mathbf{J}_{i,j} = -\mathrm{E}[(\hat{\mathbf{A}} - \mathbf{A})^2] = -\mathrm{E}\left(\frac{\partial^2 \ln f_{\mathbf{X},\mathbf{A}}(\mathbf{X},\mathbf{A})}{\partial a_i \partial a_j}\right)$$

La dimostrazione si può trovare, ad esempio, in [16, par. 10.6].

5

Stima spettrale classica

In questo capitolo verranno descritti i metodi di *stima spettrale classica*, i più importanti dei quali sono i *metodi a periodogramma*. La caratteristica principale che distingue queste tecniche da altri metodi di valutazione dello spettro sta nel fatto che le prime non assumono alcuna particolare struttura del modello matematico del processo in esame. In ragione di questo, si parla spesso di *stima spettrale non parametrica*, per distinguerla dalla *stima spettrale parametrica* che, invece, si basa su di un ben preciso modello matematico. Va notato che le tecniche di stima spettrale non parametrica sono generalmente più robuste di quelle parametriche e sono da preferirsi a queste ultime qualora non si possieda una sufficiente quantità di informazione per poter costruire un modello fedele del processo in esame. Infatti, come è noto anche da altri contesti, i metodi che sfruttano la struttura del modello matematico del sistema considerato forniscono, in genere, risultati migliori di altre tecniche solamente quando il modello su cui si basano è sufficientemente accurato.

In Figura 5.1 è riportata una sommaria classificazione dei principali metodi di stima spettrale. In grigio sono evidenziati i metodi che verranno descritti in dettaglio in questo capitolo, mentre altre tecniche di stima spettrale verranno prese in considerazione nel corso dei prossimi capitoli.

5.1 Il periodogramma

Definizione 5.1.1 (Periodogrammi) *Si consideri un processo casuale $X[n]$, WSS, a media nulla e con spettro di potenza $S_X\left(\mathrm{e}^{j2\pi f}\right)$. Si definisce* periodo-

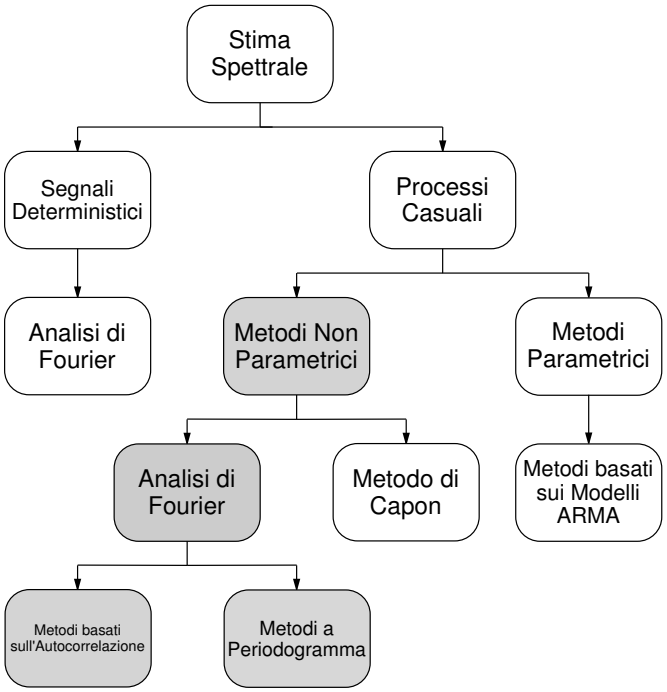

Figura 5.1. Classificazione dei principali metodi di stima spettrale.

gramma *di* $X[n]$ *la quantità*

$$P_{X,\,M}\left(\mathrm{e}^{j2\pi f}\right) \triangleq \frac{1}{2M+1}\left|\sum_{n=-M}^{+M} X[n]\mathrm{e}^{-j2\pi fn}\right|^2 \tag{5.1}$$

Si definisce periodogramma medio *il valore atteso del periodogramma*

$$\bar{P}_{X,\,M}\left(\mathrm{e}^{j2\pi f}\right) \triangleq \mathbb{E}\left\{P_{X,\,M}\left(\mathrm{e}^{j2\pi f}\right)\right\} = \mathbb{E}\left\{\frac{1}{2M+1}\left|\sum_{n=-M}^{+M} X[n]\mathrm{e}^{-j2\pi fn}\right|^2\right\} \tag{5.2}$$

Si definisce periodogramma asintotico *il limite del periodogramma medio*

$$P_X\left(\mathrm{e}^{j2\pi f}\right) \triangleq \lim_{M\to\infty} \mathbb{E}\left\{\frac{1}{2M+1}\left|\sum_{n=-M}^{+M} X[n]\mathrm{e}^{-j2\pi fn}\right|^2\right\} \tag{5.3}$$

□

Un risultato molto importante è il seguente

Teorema 5.1.1 *Si consideri un processo casuale $X[n]$, WSS, a media nulla, con autocorrelazione $r_X[l]$ e spettro di potenza $S_X\left(\mathrm{e}^{j2\pi f}\right)$. Se*

$$\sum_{l=-\infty}^{+\infty} |l| r_X[l] < \infty \tag{5.4}$$

allora lo spettro di potenza e il periodogramma asintotico di $X[n]$ coincidono

$$S_X\left(\mathrm{e}^{j2\pi f}\right) = P_X\left(\mathrm{e}^{j2\pi f}\right) \tag{5.5}$$

□

Nel seguito, per alleggerire le notazioni, lo spettro di potenza ed il periodogramma asintotico verranno indicati con $S_X(f)$ e $P_X(f)$, rispettivamente.

Dimostrazione: Si consideri il limite

$$\lim_{M\to\infty} \mathbb{E}\left\{P_{X,M}(f)\right\}$$

che viene scritto, utilizzando la (5.2), nella forma

$$\begin{aligned}
\lim_{M\to\infty} \mathbb{E}\left\{P_{X,M}(f)\right\} &= \lim_{M\to\infty} \mathbb{E}\left\{\frac{1}{2M+1}\left|\sum_{n=-M}^{+M} X[n]\mathrm{e}^{-j2\pi fn}\right|^2\right\} \\
&= \lim_{M\to\infty} \mathbb{E}\left\{\frac{1}{2M+1}\sum_{n=-M}^{+M}\sum_{m=-M}^{+M} X[n]X^*[m]\mathrm{e}^{-j2\pi f(n-m)}\right\} \\
&= \lim_{M\to\infty} \frac{1}{2M+1}\sum_{n=-M}^{+M}\sum_{m=-M}^{+M} r_X[n-m]\mathrm{e}^{-j2\pi f(n-m)} \qquad (5.6)
\end{aligned}$$

Quest'ultima espressione è stata ottenuta portando l'operatore di media all'interno della sommatoria doppia, e sfruttando il fatto che $X[n]$ è WSS.

La (5.6) si può scrivere in una forma più semplice, osservando che, in generale, vale la relazione

$$\sum_{n=-M}^{+M}\sum_{m=-M}^{+M} a[n-m] = \sum_{i=-2M}^{+2M} a[i](2M+1-|i|) \tag{5.7}$$

come si dimostra facilmente osservando che la sommatoria doppia si può vedere come la somma di tutte le diagonali di una matrice Toeplitz di dimensioni $(2M+1)\times(2M+1)$ che ha la forma seguente

$$\begin{bmatrix} a[0] & a[1] & \dots & & a[2M] \\ a[-1] & a[0] & a[1] & \dots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \\ a[-2M] & \dots & & & a[0] \end{bmatrix}$$

Pertanto

$$\begin{aligned} \lim_{M\to\infty} \mathbb{E}\left\{P_{X,M}(f)\right\} &= \lim_{M\to\infty} \frac{1}{2M+1} \sum_{i=-2M}^{+2M} (2M+1-|i|) r_X[i] \mathrm{e}^{-j2\pi fi} \\ &= \lim_{M\to\infty} \sum_{i=-2M}^{+2M} r_X[i] \mathrm{e}^{-j2\pi fi} \\ &\quad - \lim_{M\to\infty} \frac{1}{2M+1} \sum_{i=-2M}^{+2M} |i| r_X[i] \mathrm{e}^{-j2\pi fi} \end{aligned} \tag{5.8}$$

Quest'ultima espressione consente di dimostrare la (5.5) e di trovarne le condizioni di validità. Infatti, se è verificata la condizione

$$\sum_{i=-\infty}^{+\infty} |i| r_X[i] < \infty \tag{5.9}$$

si può dire che il secondo limite nella (5.8) scompare, essendo

$$\sum_{i=-\infty}^{+\infty} |i| r_X[i] \mathrm{e}^{-j2\pi fi} < \sum_{i=-\infty}^{+\infty} |i| r_X[i] < \infty$$

La (5.8) diventa pertanto

$$\lim_{M\to\infty} \mathbb{E}\left\{P_{X,M}(f)\right\} = \sum_{i=-\infty}^{+\infty} r_X[i] \mathrm{e}^{-j2\pi fi} \tag{5.10}$$

da cui, ricordando la definizione di spettro di potenza, si ottiene

$$S_X(f) = \lim_{M\to\infty} \mathbb{E}\left\{P_{X,M}(f)\right\} \tag{5.11}$$

□

5.1.1 Il problema della stima spettrale

Il periodogramma è il punto di partenza dei metodi più diffusi di stima spettrale. Tali metodi sono basati sulla determinazione di uno stimatore consistente del periodogramma asintotico $P_X(f)$. Nel caso in cui la condizione (5.4) sia verificata, tale stimatore può essere considerato uno stimatore dello spettro di potenza $S_X(f)$, in virtù del Teorema 5.1.1.

5.2 I metodi della stima spettrale classica

Volendo utilizzare metodi numerici per valutare $P_X(f)$, è necessario considerare solo un insieme discreto e finito di valori di f. Siano f_k i valori di f che compongono questo insieme, dove $k = 0, \ldots, K-1$ e K è una costante intera positiva. Il calcolo di $P_X(f_k)$ attraverso la relazione (5.3) richiede i seguenti tre passi concettuali distinti:

1. calcolo di
$$P_{X,\,M}(f_k) = \frac{1}{2M+1} \left| \sum_{n=-M}^{+M} X[n] \mathrm{e}^{-j2\pi f_k n} \right|^2 \tag{5.12}$$

2. stima del valore atteso della (5.12)
$$\mathbb{E}\left\{P_{X,\,M}(f_k)\right\} \tag{5.13}$$

3. passaggio al limite
$$\lim_{M\to\infty} \mathbb{E}\left\{P_{X,\,M}(f_k)\right\} \tag{5.14}$$

Di queste operazioni soltanto la (5.12) può essere calcolata per via numerica, a partire da un segmento di durata finita $x_N[n]$ di una singola realizzazione $x[n]$ del processo $X[n]$, mentre la (5.13) e la (5.14) possono, in linea di principio, essere calcolate soltanto per via analitica. Inoltre la (5.13) richiede la conoscenza delle caratteristiche statistiche del processo. Nel seguito verranno presi in considerazione i tre metodi di stima spettrale basati sul periodogramma utilizzati in pratica. Essi possono essere considerati come successivi affinamenti di una tecnica elementare basata sulla valutazione della (5.12) e consentono di approssimare i passi concettuali descritti dai punti 2. e 3.

5.2.1 Stima spettrale per mezzo del periodogramma

Si consideri il problema di stimare il valore atteso di $P_{X,M}(f_k)$ per mezzo di uno stimatore non polarizzato, a bassa varianza, e basato su di un insieme finito di dati misurati, ottenuti da un segmento $x_N[n]$ di una singola realizzazione $x[n]$ del processo $X[n]$. I tre metodi di stima spettrale basati sul periodogramma possono essere brevemente riassunti come segue.

- **Periodogramma semplice.** Non è altro che il calcolo della (5.12), senza tentare di stimare il valore atteso espresso dalla (5.13), né tantomeno il passaggio al limite descritto dalla (5.14). È l'implementazione del solo primo passo concettuale e permette di ottenere uno stimatore molto grezzo, generalmente ad elevata varianza e fortemente polarizzato. Il metodo è descritto in maggior dettaglio nel paragrafo 5.3.

- **Periodogramma di Bartlett.** Viene introdotto per ridurre la varianza del periodogramma semplice. Esso si ottiene dividendo l'intervallo di osservazione in N_s segmenti. Si calcola, quindi, il periodogramma semplice di ciascun segmento, e si mediano gli N_s periodogrammi così ottenuti. In questo modo si ottiene uno stimatore a varianza ridotta, ma ancora fortemente polarizzato (nella quasi totalità delle applicazioni). Il metodo è descritto in maggior dettaglio nel paragrafo 5.4.

- **Periodogramma di Welch.** Viene utilizzato per ridurre la polarizzazione dei due precedenti stimatori. Esso si ottiene introducendo una finestra di pesatura sui segmenti del periodogramma di Bartlett. Ciò consente di ridurre gli effetti di polarizzazione a larga banda che si hanno in assenza di pesatura o, per meglio dire, in presenza di pesatura rettangolare. Il metodo è descritto in maggior dettaglio nel paragrafo 5.5.

5.3 Il periodogramma semplice

Ignorando sia l'operazione di media che il passaggio al limite, si può incominciare a stimare $P_X(f)$ attraverso il

Definizione 5.3.1 (Periodogramma semplice) *Si consideri un processo casuale $X[n]$, WSS, a media nulla. Sia $x[n]$ una singola realizzazione del processo*

$X[n]$. *Si chiama* Periodogramma semplice[1], *la funzione*

$$ {}_s\hat{p}_{X,N}(f) \triangleq \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\mathrm{e}^{-j2\pi fn}\right|^2 \tag{5.15}$$

a cui è associato lo stimatore

$$ {}_s\hat{P}_{X,N}(f) \triangleq \frac{1}{N}\left|\sum_{n=0}^{N-1} X[n]\mathrm{e}^{-j2\pi fn}\right|^2 \tag{5.16}$$

□

Si osservi che questa espressione è del tutto simile a quella del periodogramma (5.1) eccetto per il fatto che, in quel caso, la sommatoria era estesa da $-M$ ad M in modo da ricondurre il passaggio al limite per $M \to \infty$, alla definizione di spettro di potenza $S_X(f)$. Nel caso della (5.16) si vuole, invece, fare esplicito riferimento ad un segmento causale di N campioni $x_N[n]$, di una realizzazione $x[n]$ del processo causale $X[n]$. Poichè la funzione che ci interessa stimare è uno spettro di potenza, il quale non è influenzato da traslazioni temporali del processo in esame, i valori finiti assunti dagli estremi delle sommatorie che compaiono nei periodogrammi sono del tutto ininfluenti. Questo spiega le differenze tra le due espressioni.

Si valutano ora la polarizzazione e la varianza dello stimatore (5.16).

Polarizzazione dello stimatore. La polarizzazione dello stimatore si ottiene in modo analogo a quanto fatto nella dimostrazione della relazione (5.5). Il valore atteso si ha quindi dall'espressione (5.16), riscritta nella forma

$$ {}_s\hat{P}_{X,N}(f) = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} X[n]X^*[m]\mathrm{e}^{-j2\pi f(n-m)} \tag{5.17}$$

Mediando si ottiene

$$ \mu_s(f) \triangleq \mathbb{E}\left\{{}_s\hat{P}_{X,N}(f)\right\} = \frac{1}{N}\sum_{n=0}^{N-1}\sum_{m=0}^{N-1} r_X[n-m]\mathrm{e}^{-j2\pi f(n-m)} \tag{5.18}$$

Usando la relazione

$$ \sum_{n=0}^{M-1}\sum_{m=0}^{M-1} a[n-m] = \sum_{i=-(M-1)}^{M-1} (M-|i|)a[i] \tag{5.19}$$

[1] In inglese *sample spectrum*.

che si dimostra, come prima, osservando che si tratta di sommare le diagonali di una matrice Toeplitz, in questo caso di dimensioni $M \times M$, si ricava

$$\mu_s(f) = \frac{1}{N} \sum_{i=-(N-1)}^{N-1} (N - |i|) r_X[i] \mathrm{e}^{-j2\pi f i} \tag{5.20}$$

o, in altri termini,

$$\mu_s(f) = \sum_{i=-\infty}^{+\infty} w_N[i] r_X[i] \mathrm{e}^{-j2\pi f i} \tag{5.21}$$

dove

$$w_N[i] \triangleq \begin{cases} 1 - |i|/N & \text{per } |i| \leq N-1 \\ 0 & \text{per } |i| > N-1 \end{cases} \tag{5.22}$$

rappresenta una finestra triangolare, il cui andamento è mostrato nella figura 5.2.

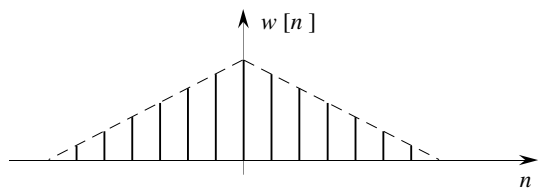

Figura 5.2. Sequenza $w_N[i]$

Dalla relazione (5.21) si vede che $\mu_s(f)$ rappresenta la DTFT della sequenza $w_N[i] r_X[i]$. L'esistenza di questa DTFT è assicurata se è verificata la condizione (5.4). Infatti, come si è visto, essa è condizione sufficiente per garantire la convergenza della serie (5.21). Dalle proprietà della DTFT è noto che al prodotto nel dominio del tempo discreto corrisponde una convoluzione nel dominio della frequenza. Pertanto si può scrivere

$$\mu_s(f) = S_X(f) * W_N(f) \tag{5.23}$$

dove $W_N(f)$ è la DTFT della sequenza $w_N[i]$

$$W_N(f) \triangleq \mathrm{DTFT}\left\{w_N[i]\right\} = \frac{1}{N} \left[\frac{\sin(\pi f N)}{\sin(\pi f)}\right]^2 \tag{5.24}$$

Nella figura 5.3 è mostrato l'andamento in dB di $W_N(f)$ nel caso $N = 20$. Si nota che $W_N(f)$ ha un lobo principale intorno all'origine e dei lobi

secondari decrescenti ai lati. Gli zeri della funzione si trovano nei punti $f_{z_i} = \pm i/N$, con i intero. Il lobo principale è largo[2] $2/N$. Alla larghezza del lobo principale è legata la definizione di *risoluzione* della stima spettrale, come si vedrà tra breve.

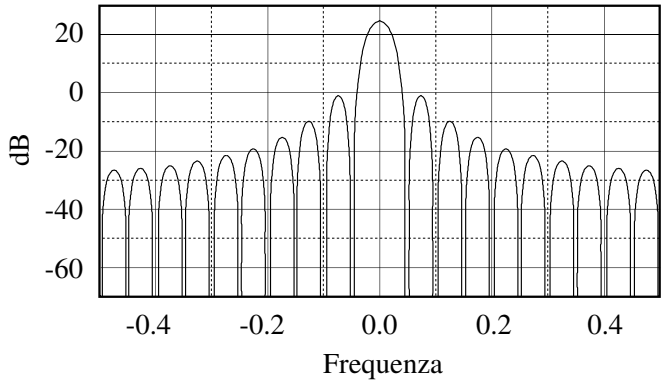

Figura 5.3. DTFT di $w_N[i]$

L'espressione (5.23) rivela che ${}_s\hat{P}_{X,N}(f)$ non è uno stimatore centrato di $S_X(f)$, perché in generale

$$\mu_s(f) \neq S_X(f) \tag{5.25}$$

Nel caso in cui $N \to \infty$ si ottiene

$$\lim_{N\to\infty} W_N(f) = \delta(f) \tag{5.26}$$

da cui segue

$$\lim_{N\to\infty} \mu_s(f) = P_X(f) \tag{5.27}$$

che rappresenta una stima centrata di $P_X(f)$. Se, inoltre, è verificata la condizione (5.4), allora il passaggio al limite (5.27) dà anche una stima centrata di $S_X(f)$. Da questo segue che il periodogramma semplice è uno stimatore asintoticamente centrato di $S_X(f)$.

Varianza dello stimatore. La valutazione della varianza richiede la conoscenza dei momenti del quarto ordine del processo. In generale si tratta di

[2] In questo caso per larghezza del lobo principale si intende la distanza che separa i due zeri più prossimi all'origine.

un calcolo molto complesso, che porta ad un'espressione in forma chiusa soltanto in casi particolari.

Uno di questi casi si ha quando $X[n]$ è un processo gaussiano bianco. In questo caso i momenti del quarto ordine si ottengono dai momenti del secondo ordine e, dopo alcuni passaggi (si veda [13]), si ottiene

$$\sigma_s^2(f) = P_X^2(f)\left[1 + \left(\frac{\sin(2\pi f N)}{N\sin(2\pi f)}\right)^2\right] \tag{5.28}$$

E' interessante osservare che $\sigma_s^2(f)$ non tende mai a zero, nemmeno per $N \to \infty$. Ne deriva che il periodogramma semplice è uno stimatore non consistente di $P_X(f)$, e pertanto non è adatto ad essere utilizzato come metodo di stima spettrale. In particolare, si noti che la varianza è dello stesso ordine di $P_X^2(f)$ per qualsiasi N, e quindi la deviazione standard è dello stesso ordine di grandezza della quantità da stimare. L'effetto che si osserva nello spettro stimato è la comparsa di rapide ed ampie fluttuazioni al posto di un andamento dolce.

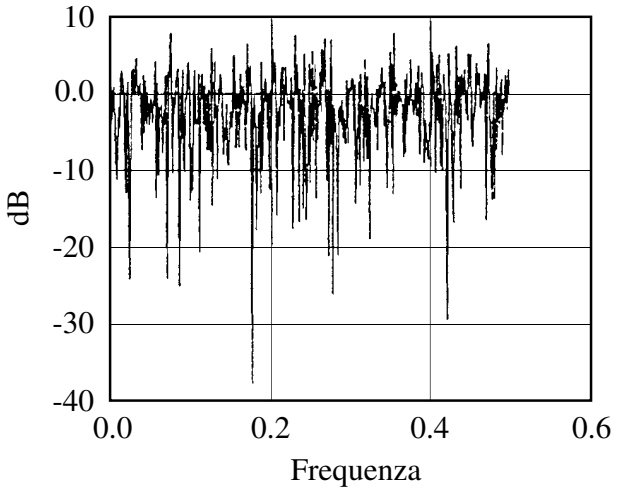

Figura 5.4. Periodogramma semplice di una sequenza bianca

Nel caso di sequenza non bianca il calcolo della varianza è molto più complicato. Si può ritenere, come si afferma in [13], che la varianza non tenda a zero neppure in questo caso.

Nella figura 5.4 è illustrato il periodogramma semplice di una sequenza bianca. Il segmento considerato per valutare ${}_s\hat{P}_{X,N}(f)$ contiene $N = 800$

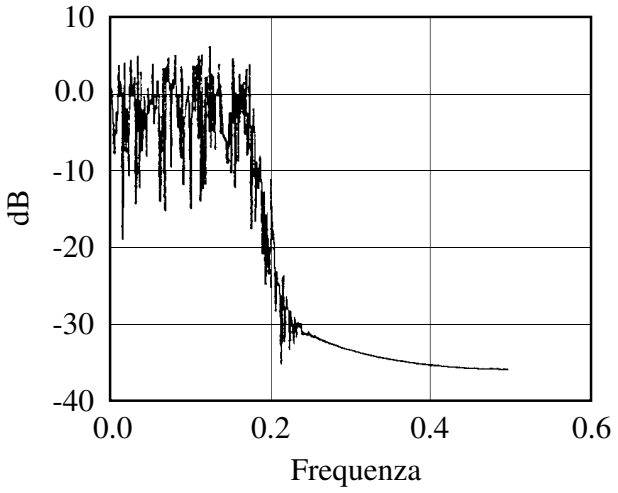

Figura 5.5. Periodogramma semplice di una sequenza bianca filtrata

campioni; anche i punti f_k in cui è stato valutato lo spettro sono 800. Le fluttuazioni che mantengono elevata la varianza sono molto evidenti.

Nella figura 5.5 è invece riportato lo spettro di un rumore bianco filtrato con un filtro passabasso, valutato anch'esso su un segmento di 800 punti. Si osserva lo stesso tipo di fluttuazioni riscontrate nel caso del rumore bianco non filtrato.

5.3.1 Risoluzione spettrale e polarizzazione locale.

Anche se il periodogramma semplice non è adatto alla stima spettrale, è conveniente usarlo qui per fare alcuni commenti relativi all'effetto della finestra $w_N[i]$, per motivi che appariranno chiari nel paragrafo seguente.

Dalla (5.23) si vede che il valore medio dello spettro stimato è legato allo spettro vero da un'operazione di convoluzione con $W_N(f)$ la cui espressione è data dalla (5.24).

Si considerino ora una singola frequenza f_0, ed i valori che vi assumono lo spettro vero $P_X(f_0)$ ed il valore atteso dello spettro stimato $\mathbb{E}\left\{{}_s\hat{P}_{X,N}(f_0)\right\}$. L'operazione di convoluzione della (5.23), da cui si ottiene $\mathbb{E}\left\{{}_s\hat{P}_{X,N}(f_0)\right\}$, può essere

riscritta nella forma

$$\mathbb{E}\left\{{}_s\hat{P}_{X,N}(f_0)\right\} = \int_{l_p} P_X(f)W_N(f_0 - f)df + \int_{l_s} P_X(f)W_N(f_0 - f)df \qquad (5.29)$$

dove l_p rappresenta l'intervallo di frequenze intorno a f_0 relativo al lobo principale e l_s rappresenta l'intervallo complementare, cioè quello relativo ai lobi secondari.

Considerando separatamente i due contributi all'integrale di convoluzione e supponendo per il momento trascurabile il contributo dovuto al secondo integrale, si può osservare che nel calcolo di $\mathbb{E}\left\{{}_s\hat{P}_{X,N}(f_0)\right\}$ intervengono tutti i valori di $P_X(f)$ in un intorno di f_0 di ampiezza pari all'ampiezza del lobo principale di $W_N(f)$. In altre parole, a causa della finestra di pesatura, l'analisi spettrale non *risolve* contributi diversi all'interno di quell'intervallo. Questo porta direttamente alla definizione di risoluzione spettrale del periodogramma semplice come $R_s = 1/N$. E' evidente che la risoluzione migliora, cioè R_s diminuisce, se aumenta la lunghezza del segmento $x_N[n]$.

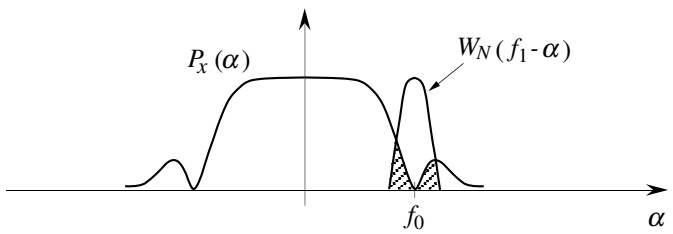

Figura 5.6. Risoluzione spettrale

Allo stesso modo, e per lo stesso motivo, si può avere una deformazione nel profilo dello spettro. Infatti la convoluzione attribuisce alla frequenza f_0 tutta l'energia contenuta nell'intervallo di risoluzione del lobo principale. Nel caso di spettro bianco ciò non provoca problemi, ma se lo spettro intorno a f_0 non è piatto, come nella figura 5.6, la stima ne risulta fortemente polarizzata. Questo tipo di polarizzazione viene detta *polarizzazione locale* [12].

Un diverso tipo di polarizzazione, detta *polarizzazione a larga banda*, è, invece, dovuta ai lobi secondari. Quest'ultima si manifesta in modo apprezzabile quando il secondo integrale della (5.29) non si può considerare trascurabile. Ciò accade, ad esempio, quando $P_X(f)$ intorno a f_0 è piccolo confrontato con i valori che assume nell'intervallo l_s. Questo aspetto del problema è analizzato più in dettaglio nel paragrafo 5.5.

Per ora si noti che la risoluzione spettrale e la polarizzazione locale sono in qualche modo legate, nel senso che quando migliora la prima, in quanto, ad esempio, si considera un segmento contenente più campioni, migliora di conseguenza anche la seconda.

5.4 Il periodogramma di Bartlett

Il periodogramma di Bartlett è un metodo che consente di approssimare l'operazione di media (5.13) ignorata nel periodogramma semplice.

Per comprendere il ragionamento su cui il metodo è basato occorre ricordare che il processo WSS è un'astrazione matematica che viene utilizzata come modello del fenomeno in osservazione. Essa è ottenuta estendendo a tutto l'asse temporale certe caratteristiche di regolarità presentate dal fenomeno nell'intervallo di tempo in cui esso viene osservato. Si può dunque immaginare di suddividere l'intervallo di osservazione $(0, N-1)$ in N_s intervalli più brevi, ciascuno di durata D, e considerare i segmenti di segnale così ottenuti come membri diversi del processo casuale complessivo. Questa operazione è schematizzata nella figura 5.7, dove i vari segmenti di $x[n]$ sono denominati $x^{(p)}[n]$, con $p = 1, \ldots, N_s$.

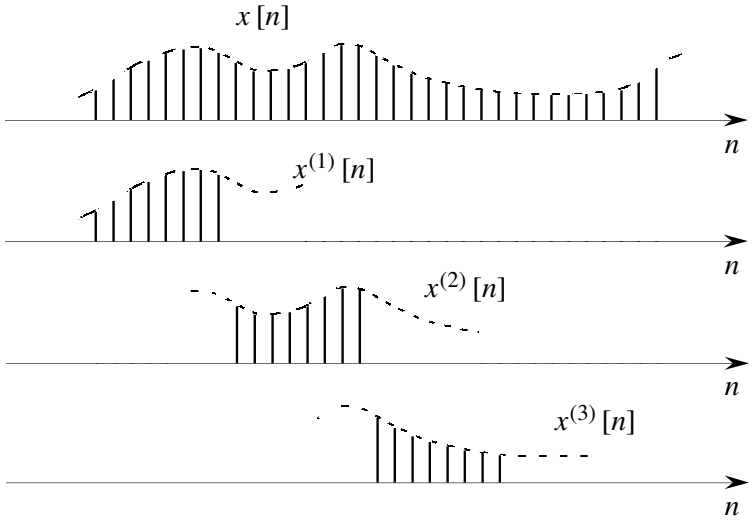

Figura 5.7. Segmentazione del processo $X[n]$

Definizione 5.4.1 (Periodogramma di Bartlett) *Si consideri un processo casuale $X[n]$, WSS, a media nulla. Sia $x_N[n]$ un segmento di N campioni di una singola realizzazione $x[n]$ del processo $X[n]$. Siano $x^{(p)}[n]$, $p = 1, 2, \ldots N_s$ dei segmenti di D campioni ottenuti a partire da $x[n]$. Il* periodogramma di Bartlett *${}_b\hat{P}_{X,N}(f)$ è definito come la media degli N_s periodogrammi semplici ${}_s\hat{P}^{(p)}_{X,D}(f)$ di ciascun segmento*

$$ {}_b\hat{P}_{X,N}(f) \triangleq \frac{1}{N_s} \sum_{p=1}^{N_s} {}_s\hat{P}^{(p)}_{X,D}(f) \tag{5.30} $$

□

L'espressione (5.30) consente di approssimare l'operazione di media che era stata trascurata nel periodogramma semplice. É evidente che si tratta ancora di uno stimatore non centrato, poichè non scompare nella media di ${}_b\hat{P}_{X,N}(f)$ la convoluzione con la DTFT della finestra triangolare. Infatti, dopo brevi passaggi, si può scrivere

$$ \mu_b(f) \triangleq \mathbb{E}\left\{ {}_b\hat{P}_{X,N}(f) \right\} = S_X(f) * W_D(f) \tag{5.31} $$

dove $W_D(f)$ si ottiene ponendo $N = D$ nella (5.24). La risoluzione spettrale dipenderà dalla durata di ciascun segmento. Pertanto, per il periodogramma di Bartlett, la risoluzione spettrale vale $R_s = 1/D$.

Il pregio del metodo di Bartlett è quello di ridurre la varianza dello stimatore ottenibile col periodogramma semplice. Se i segmenti sono statisticamente indipendenti, la varianza di ${}_b\hat{P}_{X,N}(f)$ si ottiene sommando le varianze di ${}_s\hat{P}^{(p)}_{X,D}(f)$ e dividendo per N_s. Nel caso di processo gaussiano bianco, utilizzando la (5.28) ed utilizzando D in luogo di N, si ottiene

$$ \sigma_b^2(f) = \frac{1}{N_s} P_X^2(f) \left[1 + \left(\frac{\sin(2\pi f D)}{D \sin(2\pi f)} \right)^2 \right] $$

La varianza è, in questo caso, inversamente proporzionale al numero di segmenti N_s. Se i segmenti non sono indipendenti il calcolo della varianza di ${}_b\hat{P}_{X,N}(f)$ è più complicato, ma le considerazioni di base restano valide.

In conclusione, si può dire che per un intervallo di osservazione fisso, cioè per $N = N_s D$ fisso, si dovrà cercare un compromesso tra il desiderio di ottenere una elevata risoluzione, che spingerebbe a scegliere D il più grande possibile e la necessità di avere una bassa varianza della stima, che vincola a mantenere grande N_s.

Nella figura 5.8 è mostrato lo spettro di una sequenza bianca stimato con il periodogramma di Bartlett con $N = 800$ e $N_s = 8$. È evidente il miglioramento in termini di varianza rispetto al caso del periodogramma semplice, riportato in figura 5.5.

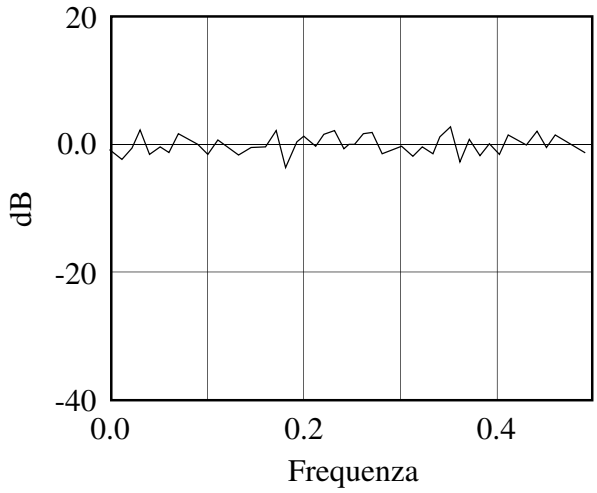

Figura 5.8. Periodogramma di Bartlett di una sequenza bianca ($N = 800$, $D = 100$, $N_s = 8$)

Analoghi risultati si ottengono nel caso di sequenza filtrata. Nella figura 5.9 è illustrato lo spettro di Bartlett di un rumore bianco filtrato con un filtro Butterworth passa basso con banda a 3 dB $B = 0{,}18$ e banda di transizione molto ripida (16 poli); sono stati usati $N_s = 10$ segmenti, ciascuno di durata $D = 800$ campioni. Si confronti, nella figura 5.10, il risultato ottenuto con lo spettro stimato per mezzo di un periodogramma semplice di $N = 8000$ campioni, cioè utilizzando la stessa sequenza di dati. É evidente il miglioramento in termini di varianza ottenuto col metodo di Bartlett.

5.5 Il periodogramma di Welch

Come già detto nel paragrafo precedente, lo spettro della figura 5.9 è stato ottenuto filtrando un rumore bianco con varianza σ_η^2 attraverso un filtro di Butterworth a 16 poli con banda a 3 dB $B = 0{,}18$. Il modulo al quadrato della funzione di trasferimento del filtro è mostrato nella figura 5.11, dove si vede che $|H(f)|^2$ of-

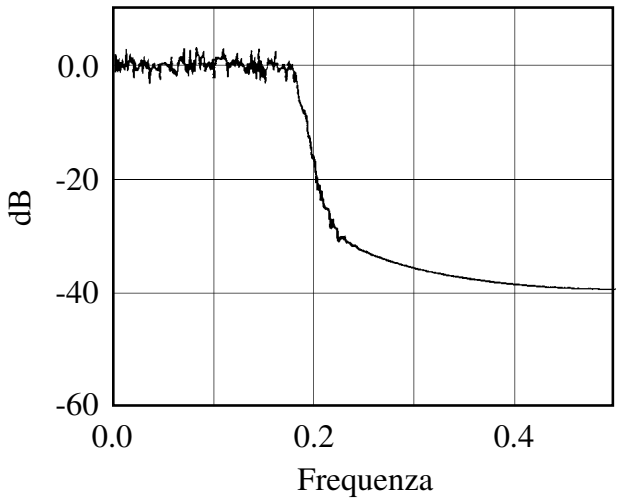

Figura 5.9. Periodogramma di Bartlett di una sequenza bianca filtrata ($D = 800$, $N = 8000$, $N_s = 10$)

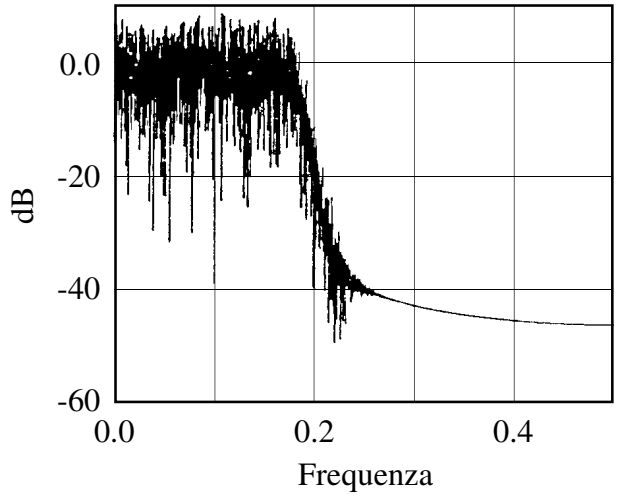

Figura 5.10. Periodogramma semplice di una sequenza bianca filtrata ($N = 8000$)

fre un'attenuazione fuori banda molto elevata, ed ha un andamento molto ripido nelle banda di transizione.

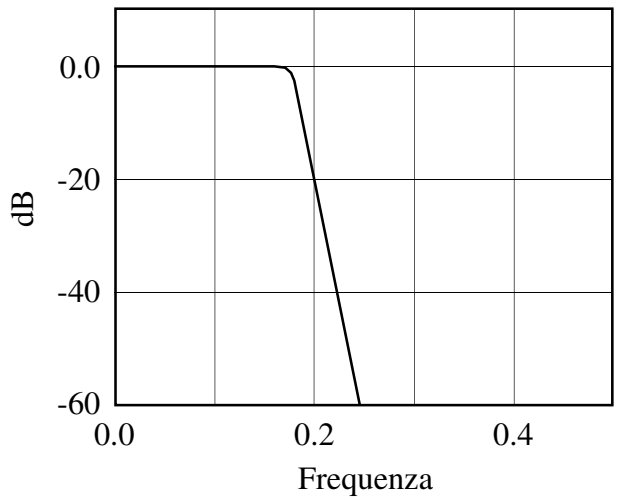

Figura 5.11. Modulo al quadrato della funzione di trasferimento di un filtro Butterworth a 16 poli

Lo spettro vero vale pertanto

$$S_X(f) = \sigma_\eta^2 |H(f)|^2$$

Nella figura 5.9 si scopre, invece, che lo spettro stimato non scende mai al di sotto di –40 dB. Questo effetto è dovuto alla convoluzione tra lo spettro vero e lo spettro della finestra di pesatura, indicato nella (5.31), che produce valori non trascurabili nell'intervallo l_s definito nella (5.29). Il fenomeno è illustrato qualitativamente nella figura 5.12, dove si vede che la polarizzazione dello stimatore dello spettro nella banda attenuata è dovuta quasi unicamente al contributo del secondo integrale della (5.29), legato ai lobi secondari di $W_D(f)$. Per questo motivo questo tipo di polarizzazione viene chiamato *polarizzazione a larga banda*[3], [12].

5.5.1 Finestre di pesatura

Un modo per combattere la polarizzazione a larga banda consiste nel sostituire la finestra rettangolare implicitamente contenuta nel segmento $x^{(p)}[n]$ con una finestra di pesatura la cui DTFT presenti lobi secondari più bassi. Questa riduzione dei lobi secondari si tradurrà in un minor contributo del secondo integrale

[3]In inglese si usa spesso anche il termine *leakage*.

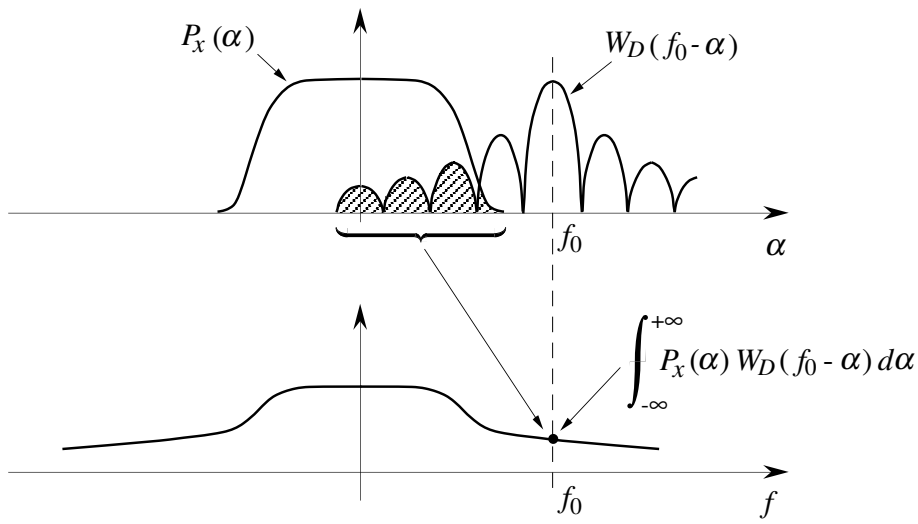

Figura 5.12. Polarizzazione a larga banda

della (5.29). La finestratura viene applicata semplicemente moltiplicando ogni campione di ogni segmento $x^{(p)}[n]$ per il corrispondente campione della finestra scelta $w_g[n]$ ottenendo la sequenza finestrata $x_w^{(p)}[n] = x^{(p)}[n] \odot w_g[n]$. Il prodotto nel dominio del tempo si traduce in una convoluzione nel dominio della frequenza. Pertanto, nell'espressione (5.31) la convoluzione verrà effettuata con lo spettro della nuova finestra $W_g(f)$, anzichè con quello della finestra rettangolare.

Utilizzando finestre di pesatura appositamente studiate si può ridurre di decine di dB la polarizzazione a larga banda. Si noti tuttavia che, a parità di lunghezza D delle varie finestre, la riduzione dei lobi secondari si ottiene sempre a spese di un aumento della larghezza del lobo principale. Ne segue che la scelta della finestra da usare comporta un compromesso tra risoluzione, e quindi anche polarizzazione locale, e polarizzazione a larga banda.

Numerose soluzioni a questo problema sono state proposte dai ricercatori. In questo paragrafo verranno considerate alcune delle più importanti finestre utilizzate nella pratica, e ne saranno descritte brevemente le caratteristiche, senza approfondire la teoria che sta alla base del progetto di tali finestre. Il lettore interessato all'argomento può fare riferimento a [21], dove sono descritte alcune delle finestre più comuni. Per una discussione più approfondita ci si riferisca invece a [13] o [24].

Innanzitutto è utile dare una sommaria classificazione di queste finestre. Si chiamano *finestre fisse* quelle che, a parità di lunghezza D, danno un ammontare fisso di polarizzazione a larga banda. Appartengono a questa categoria la finestra *rettangolare*, quella di *Hamming* e quella di *Hann*, erroneamente conosciuta come finestra di Hanning. Invece, le finestre che non appartengono alla categoria delle finestre fisse, sono definite in funzione di parametri che consentono, a parità di D, di scegliere il giusto compromesso tra polarizzazione a larga banda e risoluzione. Appartengono a quest'ultima classe la finestra di *Kaiser* e la finestra di *Dolph-Chebyshev*. In figura 5.13 sono riportati gli andamenti delle finestre qui menzionate nel dominio del tempo (a sinistra) ed in quello della frequenza (a destra). Per quest'ultimo dominio sono riportati gli andamenti in dB del modulo dello spettro. La righe di questa figura riportano nell'ordine, gli andamenti della finestra rettangolare, di quella di Hamming, di quella di Hann, di quella di Kaiser e di quella di Dolph-Chebyshev, per una lunghezza $D = 32$. Gli spettri sono stati valutati utilizzando una FFT a 512 punti. Le ultime due finestre sono state progettate per fornire un livello dei lobi secondari rispetto al lobo principale di circa -40 dB. Come si può osservare, i profili delle varie finestre sono molto simili nel dominio del tempo, ma molto diversi nel dominio della frequenza. In particolare risulta evidente l'abbassamento del livello dei lobi secondari rispetto al caso di finestra rettangolare ed il conseguente allargamento del lobo principale.

In tabella 5.5.1 sono riassunte le principali proprietà delle finestre considerate, in termini di estensione del lobo principale ed altezza massima dei lobi secondari, in dB, rispetto al lobo principale.

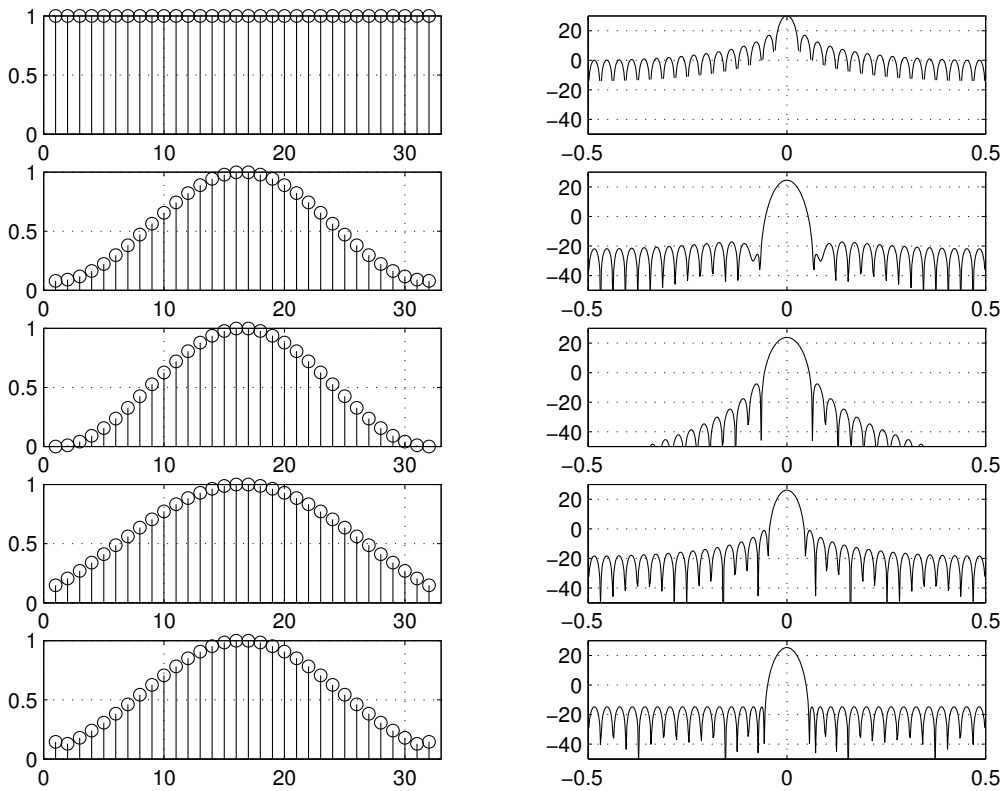

Figura 5.13. Andamenti delle principali finestre di pesatura. Dominio del tempo (sinistra), dominio della frequenza (destra).

Tipo di finestra	Altezza lobi secondari	Estensione lobo principale
Rettangolare	-13	$\frac{0.905}{D-1}$
Hamming	-43	$\frac{3.135}{D-1}$
Hann	-32	$\frac{2.505}{D-1}$
Kaiser	-A	$\frac{A-8}{14.357(D-1)}$
Dolph-Chebyshev	-A	$\frac{1}{2\pi}\cos^{-1}\left[\left(\cosh\frac{\cosh^{-1}10^{A/20}}{D-1}\right)^{-1}\right]$

Consideriamo più in dettaglio le caratteristiche delle finestre mostrate in figura 5.13.

Finestra di Hamming. L'espressione nel dominio del tempo di questa finestra è

$$w_{Hm}[n] = 0{,}54 - 0{,}46 \cos\left(\frac{2\pi n}{D-1}\right) \quad \text{per } n = 0, 1, \ldots D-1 \tag{5.32}$$

Finestra di Hann. L'espressione nel dominio del tempo di questa finestra è

$$w_{Hn}[n] = 0{,}5 - 0{,}5 \cos\left(\frac{2\pi n}{D-1}\right) \quad \text{per } n = 0, 1, \ldots D-1 \tag{5.33}$$

Si tratta di una funzione a coseno rialzato.

Finestra di Kaiser. L'espressione nel dominio del tempo di questa finestra è

$$w_{Ka}[n] = \frac{I_0\left(\beta\sqrt{1-[1-2n/(D-1)]^2}\right)}{I_0(\beta)} \quad \text{per } n = 0, 1, \ldots D-1 \tag{5.34}$$

dove $I_0(\cdot)$ è la funzione di Bessel modificata di prima specie, di ordine zero, mentre β è un parametro che consente di ottenere opportuni livelli dei lobi secondari, e corrispondenti estensioni del lobo principale. Si osservi che ponendo $\beta = 0$ si ottiene la finestra rettangolare. Valori di $\beta > 0$ decrementano il livello dei lobi secondari, a fronte di un incremento dell'estensione del lobo principale. Il parametro β può essere scelto in modo da ottenere approssimativamente un livello di lobi secondari A sulla base della seguente relazione

$$\beta \backsimeq \begin{cases} 0 & A \leq 21 \\ 0.5842(A-21)^{0.4} + 0.07886(A-21) & 21 < A \leq 50 \\ 0.1102(A-8.7) & A > 50 \end{cases} \tag{5.35}$$

Inoltre, per ottenere un livello di lobi secondari A ed un estensione del lobo principale Δf, la lunghezza della finestra D deve soddisfare l'equazione

$$\Delta f = \frac{A-8}{14.357(D-1)} \tag{5.36}$$

Finestra di Dolph-Chebyshev. Questa finestra è ottenuta come IDFT dei polinomi di Chebyshev, valutata in D equispaziati valori di frequenza sulla

circonferenza di raggio unitario del piano $\mathcal{Z}$. Essa è caratterizzata dall'avere un livello di lobi secondari costante. Si dice che questo tipo di finestra ha un comportamento *equiripple*. L'estensione del lobo principale Δf, il livello di lobi secondari A e la lunghezza della finestra D sono legati dalla seguente relazione

$$\Delta f = \frac{1}{2\pi} \cos^{-1} \left[\left(\cosh \frac{\cosh^{-1} 10^{A/20}}{D-1} \right)^{-1} \right] \tag{5.37}$$

5.5.2 Il Periodogramma di Welch

Un modo per migliorare la risoluzione spettrale, senza rinunciare alla finestratura, è stato proposto da Welch. Egli propone di modificare l'approccio del periodogramma di Bartlett, consentendo una parziale sovrapposizione degli N_s segmenti considerati, estratti dalla sequenza di dati sperimentali. Ciò consente di aumentare la lunghezza D di ogni segmento, senza variare né la durata N dell'intervallo di osservazione, né il numero N_s di segmenti.

In [13] si possono trovare le caratteristiche (media e varianza) del periodogramma di Welch. Qui si conclude mostrando con un esempio il miglioramento che si può ottenere col metodo proposto. Nella figura 5.14.d è mostrato il periodogramma di Welch relativo allo stesso rumore filtrato che aveva prodotto il periodogramma di Bartlett della figura 5.9. Anche i parametri per la stima spettrale sono gli stessi usati nella figura 5.9, cioè $D = 800$ e $N = 8000$; la sovrapposizione è di 200 campioni. Il miglioramento ottenuto nella banda attenuata si commenta da solo. Nella figura 5.14 sono mostrate per confronto alcune curve già illustrate precedentemente. In particolare si hanno le curve del modulo quadro della risposta in frequenza del filtro utilizzato per filtrare il rumore bianco, il periodogramma semplice ed il periodogramma di Bartlett.

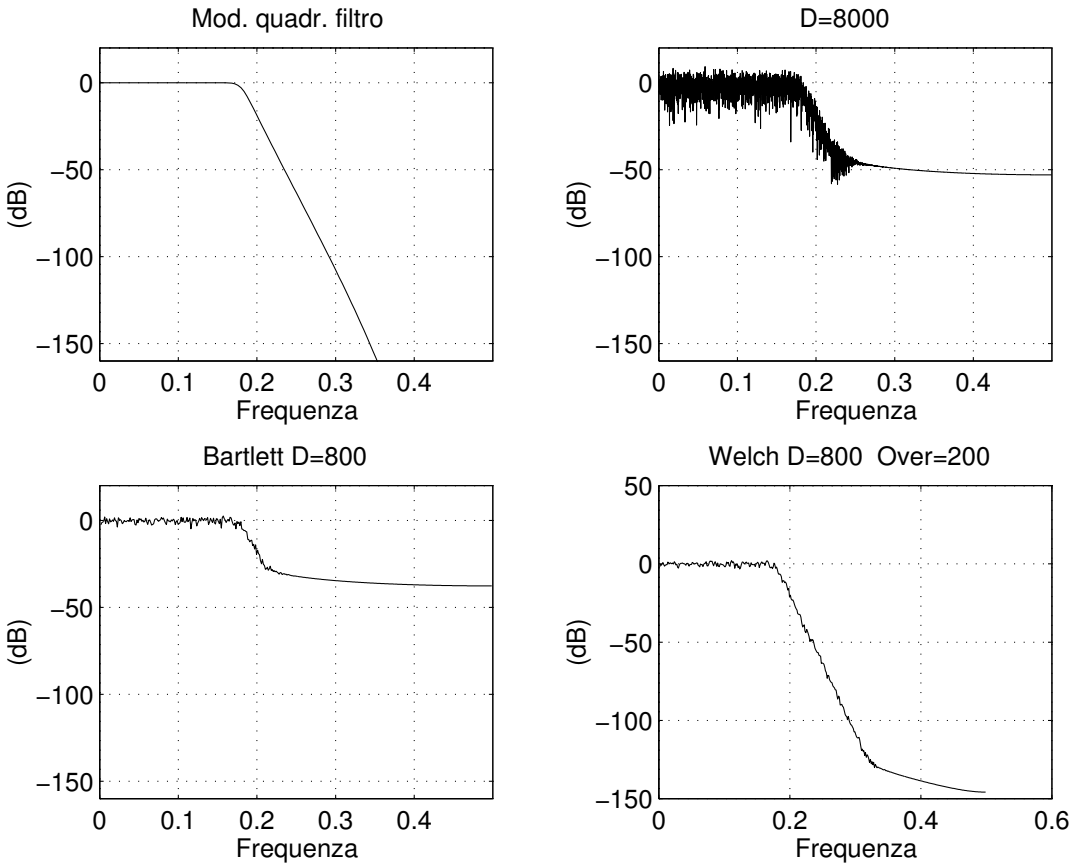

Figura 5.14. Periodogramma di Welch (con sovrapposizione del 30%) di un rumore bianco filtrato ($N = 8000$)

6

Stimatori lineari a Minimo Errore Quadratico Medio (MMSE)

In questo capitolo verrano esaminati alcuni problemi di stima di parametri casuali basati sulla minimizzazione dell'errore quadratico medio. Tale criterio di stima, noto come metodo MMSE (dall'inglese *Minimum Mean Square Error*), è già stato presentato nel paragrafo 4.3.1.

Il problema è ancora quello di volere stimare un parametro y_v a partire da un insieme di dati misurati $x_N[n]$, con $n = 0, 1, \ldots, N-1$, nell'ipotesi che $x_N[n]$ rappresenti un segmento temporale di una singola realizzazione, $x[n]$, di un processo casuale a tempo discreto $X[n]$. Nel capitolo 3 era stata introdotta la *stima* di y_v come funzione dei dati misurati $y = h(x[n])$. In questo capitolo si pone un vincolo sul tipo di funzione $h(\cdot)$. In particolare si impone che tale funzione sia di tipo lineare.

Come si vedrà nel capitolo 7 i metodi parametrici di stima spettrale sono collegati con i metodi di stima lineare.

6.1 Stimatore lineare

Gli stimatori lineari basati sul criterio MMSE hanno la proprietà fondamentale di poter essere rappresentati da un'espressione in forma chiusa dipendente solo dalla

statistica del secondo ordine del processo, cosa che in generale non è possibile con altri stimatori di tipo MMSE. Nella pratica quesi stimatori, noti comunemente come *filtri di Wiener*, sono molto utilizzati in quanto poco onerosi dal punto di vista realizzativo.

Si dice che uno stimatore è di tipo lineare se può essere espresso come

$$Y = h(X[n]) = \sum_{n=0}^{N-1} h_n X[n] \tag{6.1}$$

Utilizzando come criterio di stima il metodo MMSE, introdotto nel paragrafo 4.3.1, il punto di partenza per la valutazione dei coefficienti dello stimatore è il funzionale J

$$J = \mathrm{E}\{|y_v - Y|^2\} = \mathrm{E}\left\{\left|y_v - \sum_{n=0}^{N-1} h_n X[n]\right|^2\right\} \tag{6.2}$$

Supponendo che $X[n]$ sia un processo reale, sviluppando i quadrati si ha

$$J = \mathrm{E}\{y_v^2\} + \sum_{n=0}^{N-1}\sum_{i=0}^{N-1} h_n h_i \mathrm{E}\{X[n]X[i]\} - 2\sum_{n=0}^{N-1} h_n \mathrm{E}\{y_v X[n]\} \tag{6.3}$$

Derivando l'espressione (6.3) rispetto ad una generica variabile h_j, con $j = 0, 1, \ldots, N-1$, si ottiene

$$\frac{\partial J}{\partial h_j} = 2h_j \mathrm{E}\{|X[j]|^2\} + 2\sum_{n\neq j} h_n \mathrm{E}\{X[n]X[j]\} - 2\mathrm{E}\{y_v X[j]\}$$

o, riscrivendo in un unico termine i primi due addendi a secondo membro

$$\frac{\partial J}{\partial h_j} = 2\sum_{n} h_n \mathrm{E}\{X[n]X[j]\} - 2\mathrm{E}\{y_v X[j]\} \tag{6.4}$$

Uguagliando a zero la derivata (6.4) si ha

$$\sum_{n} h_n \mathrm{E}\{X[n]X[j]\} = \mathrm{E}\{y_v X[j]\} \qquad j = 0, 1, \ldots, N-1 \tag{6.5}$$

Questo risultato costituisce un insieme di equazioni *lineari* nelle N incognite h_j, note come *equazioni normali*.

Se $X[n]$ è un processo WSS, oltre che reale, l'equazione (6.5) diventa

$$\sum_{n} h_n r_X[n-j] = \mathrm{E}\{y_v X[j]\} \qquad j = 0, 1, \ldots, N-1 \tag{6.6}$$

a cui corrisponde il sistema

$$\begin{aligned}
h_0 r_X[0] + h_1 r_X[1] + h_2 r_X[2] + \ldots + h_{N-1} r_X[N-1] &= \mathrm{E}\{y_v X[0]\} \\
h_0 r_X[1] + h_1 r_X[0] + h_2 r_X[1] + \ldots + h_{N-1} r_X[N-2] &= \mathrm{E}\{y_v X[1]\} \\
\vdots \quad \vdots \quad \vdots \\
h_0 r_X[N-1] + h_1 r_X[N-2] + \ldots + h_{N-1} r_X[0] &= \mathrm{E}\{y_v X[N-1]\}
\end{aligned}$$

che può essere riscritto in forma matriciale utilizzando la matrice di autocorrelazione $\mathbf{R}_X$, che, come è noto, nel caso di processi WSS reali è una matrice Toeplitz simmetrica. Si ottiene

$$\mathbf{R}_X \, \mathbf{h}_N = \mathrm{E}\{y_v \mathbf{X}_N\} \tag{6.7}$$

dove

$$\begin{aligned}
\mathbf{R}_X &= \begin{bmatrix} r_X[0] & r_X[1] & \ldots & r_X[N-1] \\ r_X[1] & r_X[0] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \\ r_X[N-1] & \ldots & & r_X[0] \end{bmatrix} \\
\mathbf{h}_N &= [\, h_0 \quad h_1 \quad \ldots \quad h_{N-1} \,]^T \\
\mathbf{X}_N &= [\, X[0] \quad X[1] \quad \ldots \quad X[N-1] \,]^T
\end{aligned}$$

Si noti che, per mantenere valida in generale l'espressione (6.7), il parametro y_v è stato mantenuto all'interno dell'operatore di media, dal momento che y_v può essere sia deterministico che casuale.

Si osservi che il termine $(y_v - Y)$ è una variabile casuale di cui non è necessario conoscere la densità di probabilità. Infatti le equazioni normali consentono di ricavare i pesi h_n dello stimatore soltanto a partire dalla conoscenza di momenti del secondo ordine.

6.1.1 Ortogonalità dell'errore

Si richiama l'espressione dell'errore di stima

$$e_y = y_v - \sum_{n=0}^{N-1} h_n x[n] \tag{6.8}$$

a cui è associata la variabile casuale

$$E_y = y_v - \sum_{n=0}^{N-1} h_n X[n] \tag{6.9}$$

Utilizzando l'equazione (6.5) si dimostra facilmente che

$$\mathrm{E}\{E_y X[j]\} = 0$$

Ricordando la definizione (1.11) si può affermare che l'errore risulta ortogonale, ovvero scorrelato, ai campioni disponibili del processo. Per tale ragione le equazioni (6.5) e (6.6) sono dette *equazioni normali*, e il sistema (6.7) è detto sistema delle equazioni normali.

L'*ortogonalità* dell'errore ai dati acquisiti è una propietà fondamentale degli stimatori lineari MMSE.

6.1.2 Errore residuo

Le equazioni normali consentono di stimare y_v a meno dell'errore E_y espresso dalla (6.9). Il funzionale J rappresenta il valore quadratico medio di tale errore. Quando sono soddisfatte le equazioni normali J raggiunge il suo valore minimo $J_{\min}$.

Per calcolare $J_{\min}$ basta sostituire l'equazione (6.5) nell'equazione (6.3). In pratica nella (6.3) si pone al posto di $\sum_n h_n \mathrm{E}\{X[n]X[j]\}$ il valore atteso $\mathrm{E}\{y_v X[j]\}$, come si ricava dalla (6.5), e si ottiene

$$J_{\min} = \mathrm{E}\{y_v^2\} + \sum_{n=0}^{N-1} h_n \mathrm{E}\{y_v X[n]\} - 2 \sum_{n=0}^{N-1} h_n \mathrm{E}\{y_v X[n]\}$$

da cui infine

$$J_{\min} = \mathrm{E}\{y_v^2\} - \sum_{n=0}^{N-1} h_n \mathrm{E}\{y_v X[n]\} \tag{6.10}$$

6.1.3 Stimatore lineare per sequenze infinite

Si osservi che l'espressione (6.5) può essere ottenuta formalmente anche nel caso in cui la sequenza $x[n]$ sia di durata infinita. Seguendo il metodo illustrato nel paragrafo 6.1 si vede che è possibile scrivere l'espressione dello *stimatore* LMMSE *asintotico*

$$Y = h(X[n]) = \sum_{n=-\infty}^{+\infty} h_n X[n] \tag{6.11}$$

a cui per un processo WSS corrisponde l'insieme di equazioni normali

$$\sum_{i=-\infty}^{+\infty} h_i r_X[i-j] = \mathrm{E}\{y_v X[j]\} \qquad \text{con } j \text{ intero} \tag{6.12}$$

Il sistema che si ottiene ha infinite incognite ed infinite equazioni, quindi, in generale, non è possibile esprimerne la soluzione in modo elementare.

6.2 Predittore lineare

Si consideri il caso, illustrato nella figura 6.1, in cui l'insieme dei dati misurati $x_N[n]$, con $n = 0, \ldots N-1$, rappresenti campioni di una realizzazione $x[n]$ di un processo casuale $X[n]$. Si supponga di volere stimare il campione successivo

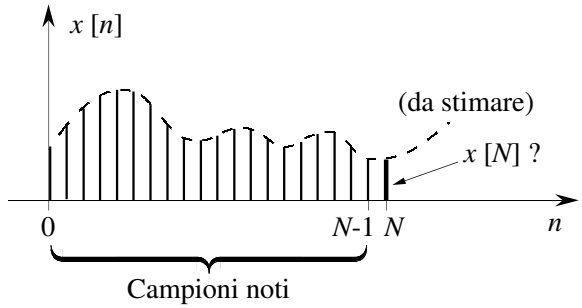

Figura 6.1. Segmento di segnale con campione da predire

$X[N]$. Se si utilizza uno stimatore con un'espressione del tipo della (6.1), al campione successivo è attribuito un valore stimato (cioè una stima) $\hat{x}[N]$ che è ottenuto come combinazione lineare dei dati misurati $x[n]$. Si ha infatti

$$\hat{x}[N] = \sum_{n=0}^{N-1} h_n x[n] \tag{6.13}$$

Si osservino alcuni aspetti del problema.

- Alla stima (6.13) è associato uno stimatore $\hat{X}[N]$. Da un confronto con l'espressione (6.1) si ha che, in questo caso, $Y = \hat{X}[N]$. Possiamo pertanto utilizzare il metodo esposto precedentemente, ponendo $y_v = X[N]$. Infatti in questo caso il valore "vero da stimare è in generale una variabile casuale.

- L'espressione (6.13) rappresenta una previsione del valore assunto dal campione futuro $x[N]$ sulla base dei campioni passati. Per questa ragione lo stimatore $\hat{X}[N]$, espresso dalla relazione (6.13) viene chiamato *predittore lineare* del processo $X[n]$.

- Usando il criterio MMSE, nell'ipotesi di processo reale WSS, il migliore predittore lineare si ottiene risolvendo il sistema espresso dall'equazione (6.6), che può essere riscritto come

$$\sum_n h_n r_X[n-j] = \mathrm{E}\{X[N]X[j]\} = r_X[N-j] \tag{6.14}$$

con $j = 0, 1, \ldots, N-1$, ovvero come

$$\begin{aligned}
h_0 r_X[0] + h_1 r_X[1] + h_2 r_X[2] + \ldots + h_{N-1} r_X[N-1] &= r_X[N] \\
h_0 r_X[1] + h_1 r_X[0] + h_2 r_X[1] + \ldots + h_{N-1} r_X[N-2] &= r_X[N-1] \\
\vdots \qquad & \quad \vdots \\
h_0 r_X[N-1] + h_1 r_X[N-2] + \ldots + h_{N-1} r_X[0] &= r_X[1]
\end{aligned}$$

Si noti che nulla è stato detto sul problema della stima delle autocorrelazioni $r_X[N-j]$.

6.2.1 Il filtro FIR per la predizione con i processi WSS

Il predittore lineare può essere visto anche come un filtro FIR di lunghezza N e con coefficienti b_i che opera sui dati misurati $x[n]$ come indicato nella 6.2. Scrivendo l'uscita del filtro al passo N si ottiene

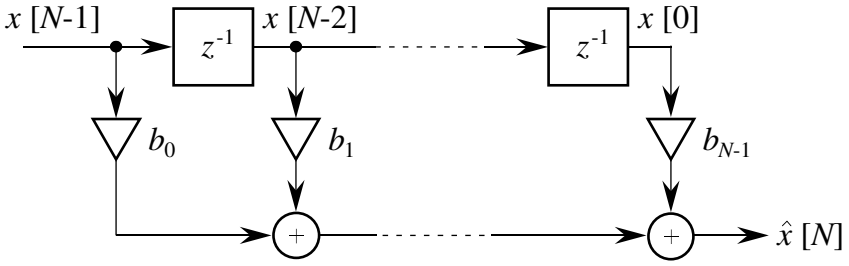

Figura 6.2. Filtro FIR per la predizione lineare del campione $X[N]$

$$\hat{x}[N] = \sum_{i=0}^{N-1} b_{N-1-i} x[i] = \sum_{i=0}^{N-1} b_i x[N-1-i]$$

Questa espressione deve essere confrontata con la (6.13), che viene riscritta cambiando l'indice della sommatoria come

$$\hat{x}[N] = \sum_{i=0}^{N-1} h_i x[i]$$

Confrontando le due espressioni di $\hat{x}[N]$ si vede che, progettando i coefficienti del filtro secondo il criterio MMSE, si arriva all'espressione (6.14) dove $h_n = b_{N-1-n}$, cioè

$$\sum_n b_{N-1-n} r_X[n-j] = r_X[N-j] \qquad \text{con } j = 0, 1, \ldots, N-1$$

Quest'ultima espressione può essere riscritta operando i due cambiamenti di variabile

$$k = N-1-n \qquad\qquad m = N-1-j$$

Si ottiene

$$\sum_k b_k r_X[k-m] = r_X[m+1] \qquad \text{con } m = 0, 1, \ldots, N-1 \tag{6.15}$$

a cui corrisponde il sistema

$$\begin{aligned}
b_0 r_X[0] + b_1 r_X[1] + b_2 r_X[2] + \ldots + b_{N-1} r_X[N-1] &= r_X[1] \\
b_0 r_X[1] + b_1 r_X[0] + b_2 r_X[1] + \ldots + b_{N-1} r_X[N-2] &= r_X[2] \\
\vdots \quad \vdots \quad \vdots \\
b_0 r_X[N-1] + b_1 r_X[N-2] + \ldots + b_{N-1} r_X[0] &= r_X[N]
\end{aligned}$$

Un metodo veloce per invertire la matrice $\mathbf{R}_X$ e risolvere quindi il sistema (6.15) è utilizzare l'algoritmo recursivo di Levinson [3, pag.131].

6.2.2 Predittore lineare di ordine p

Si è visto che un predittore lineare può essere rappresentato come un filtro FIR di lunghezza N, dove N sono i dati misurati. A questo punto è possibile riformulare il problema del progetto del predittore lineare, svincolando la lunghezza del filtro FIR dalla lunghezza del vettore dei dati.

Si consideri un segmento temporale $x_N[n]$ di una realizzazione di un processo casuale $X[n]$ WSS, contenente p campioni, e un istante discreto n generico, esterno e successivo al segmento temporale considerato. Il campione $x[n]$ può essere stimato utilizzando un predittore così definito

$$\hat{x}[n] = \sum_{i=1}^{p} c_i x[n - n_0 - (i-1)] \tag{6.16}$$

chiamato *predittore lineare di ordine p con distanza di predizione* n_0. Nella figura 6.3 è illustrato un esempio nel caso $n_0 = 5$. E' evidente che la (6.16) può essere

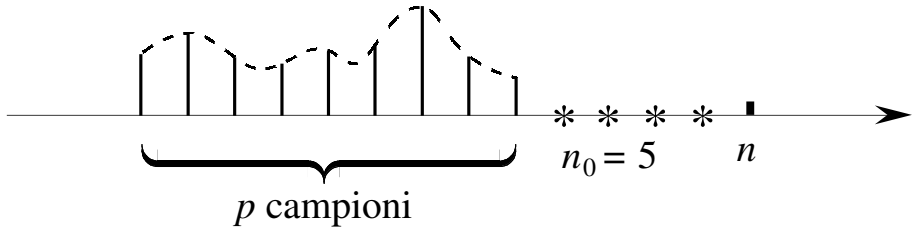

Figura 6.3. Predittore con distanza di predizione $n_0 = 5$

utilizzata per predire $X[n]$ in qualsiasi istante discreto a patto di conoscere p campioni passati di una realizzazione del processo in esame.

Predittore lineare ad un passo Se $n_0 = 1$ la distanza di predizione si riduce ad uno. In questo caso la predizione al passo n viene fatta sulla base dei p campioni precedenti, come illustrato nella figura 6.4, e l'equazione (6.16) diventa

$$\hat{x}[n] = \sum_{i=1}^{p} c_i x[n - i] \tag{6.17}$$

Il predittore che si ottiene viene detto *predittore lineare di ordine p ad un passo.* Come è evidente, tale predittore può essere rappresentato come un filtro FIR. Lo schema di tale filtro è mostrato nella figura 6.5 e si ricava facilmente dallo schema del precedente filtro FIR indicato nella figura 6.2. La lunghezza p del filtro non dipende dal numero di dati misurati, ma è un parametro intrinseco del predittore.

Operando come nel paragrafo 6.2.1, i coefficienti c_i possono essere valutati risolvendo il sistema

$$c_1 r_X[0] + c_2 r_X[1] + c_3 r_X[2] + \ldots + c_p r_X[p-1] \quad = \quad r_X[1]$$

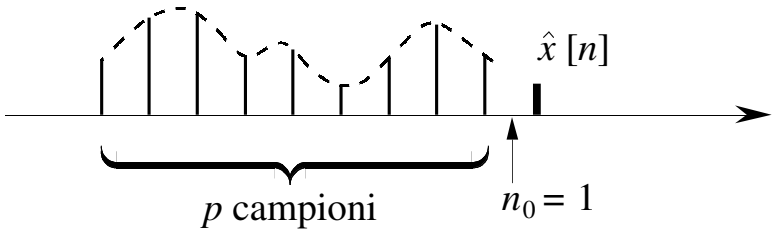

Figura 6.4. Predittore con distanza di predizione $n_0 = 1$

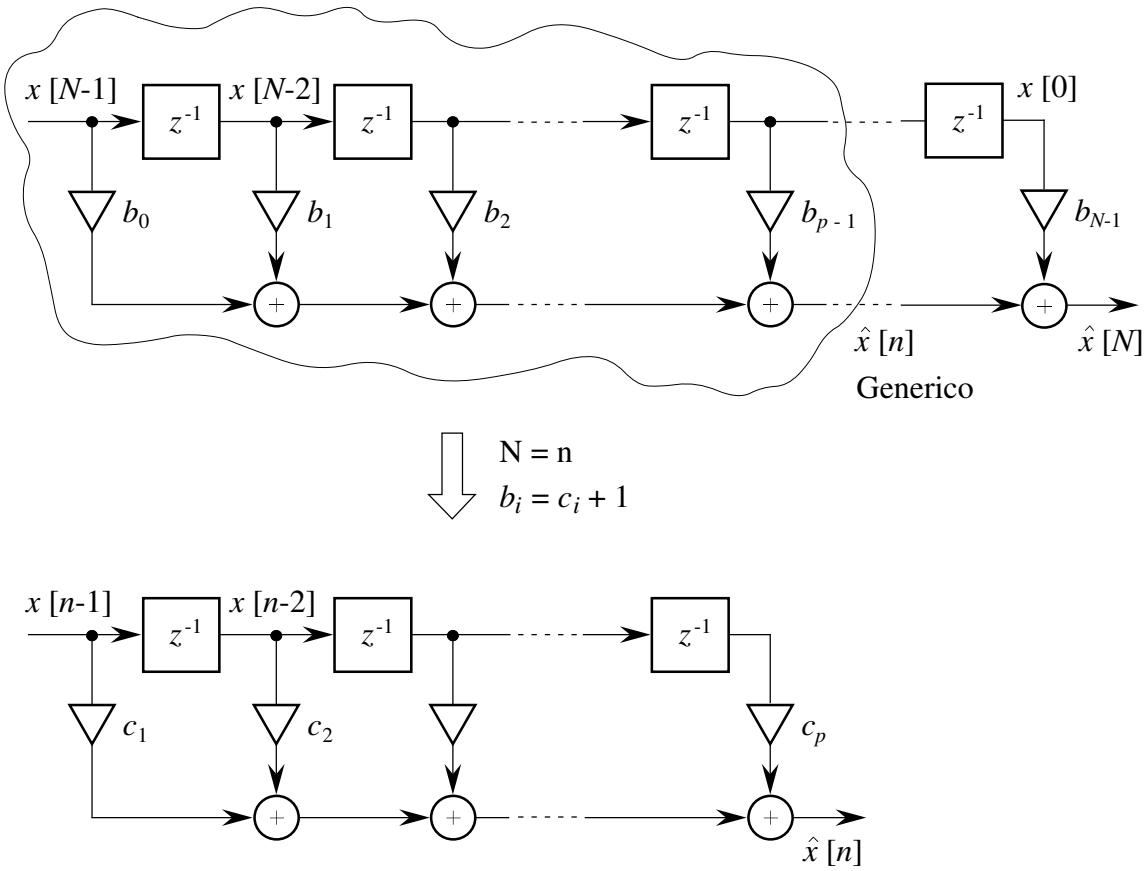

Figura 6.5. Predittore ad un passo di ordine p

$$c_1 r_X[1] + c_2 r_X[0] + c_3 r_X[1] + \ldots + c_p r_X[p-2] \quad = \quad r_X[2]$$

$$\begin{aligned} & \vdots \quad \vdots \quad \vdots \\ c_1 r_X[p-1] + c_2 r_X[p-2] + \ldots + c_p r_X[0] &= r_X[p] \end{aligned} \tag{6.18}$$

Predittore ad un passo: Relazioni tra N, p, b_i, c_i. Per scrivere in maniera corretta le relazioni tra N, p, b_i, c_i è necessario ottenere l'equazione (6.17) a partire dall'espressione generale del predittore lineare, cioè dall'equazione

$$\hat{x}[N] = \sum_{i=0}^{N-1} h_i x[i] \tag{6.19}$$

Ponendo $h_i = b_{N-1-i}$ in questa espressione si ha

$$\hat{x}[N] = \sum_{i=0}^{N-1} b_{N-1-i} x[i]$$

Poiché nel predittore di ordine p i campioni più vecchi vengono eliminati dalla predizione, introducendo

$$b'_i = \begin{cases} b_i & \text{per } 0 \leq i \leq p-1 \\ \\ 0 & \text{altrove} \end{cases}$$

tale predittore si può scrivere nella forma

$$\hat{x}[N] = \sum_{i=0}^{N-1} b'_{N-1-i} x[i]$$

dove si può porre $N = n$, quindi

$$\hat{x}[n] = \sum_{i=0}^{n-1} b'_{n-1-i} x[i] = \sum_{i=0}^{n-1} b'_i x[n-1-i]$$

Richiamando la relazione tra b'_i e b_i pocanzi definita si ha

$$\hat{x}[n] = \sum_{i=0}^{p-1} b_i x[n-1-i]$$

oppure, con un cambiamento di variabile,

$$\hat{x}[n] = \sum_{i=1}^{p} b_{i-1} x[n-i]$$

Questa espressione coincide con l'espressione (6.17) se si pone

$$c_i = b_{i-1} \qquad\qquad i = 1, \ldots, p$$

Nella figura 6.6 le relazioni tra i coefficienti h_i, b_i e c_i sono indicate sotto forma di tabelle. E' indicata inoltre la relazione tra N e n.

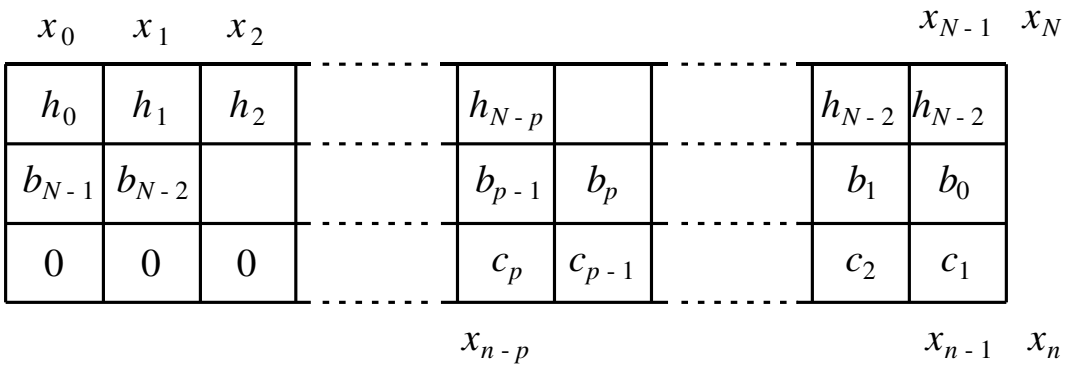

Figura 6.6. Relazioni tra N, h_i, b_i, c_i

6.2.3 Il filtro dell'errore di predizione

Si consideri il predittore ad un passo espresso dall'equazione (6.17) e l'errore di predizione $e_x[n] = x[n] - \hat{x}[n]$. Si ha

$$e_x[n] = x[n] - \sum_{i=1}^{p} c_i x[n-i] = \sum_{i=0}^{p} d_i x[n-i] \tag{6.20}$$

dove i coefficienti d_i possono essere scritti in forma vettoriale come

$$\mathbf{d} = [\,1 \quad -c_1 \quad -c_2 \quad \ldots \quad -c_p\,]^T$$

Il vettore $\mathbf{d}$ individua un filtro FIR, detto *filtro dell'errore di predizione* (PEF dall'inglese *Prediction Error Filter*). Il filtro PEF si utilizza per riscrivere il sistema di equazioni normali in funzione del vettore $\mathbf{d}$, aumentato di un'equazione nell'incognita $J_{\min}$. Dall'equazione (6.10) si ha che

$$J_{\min} = \mathrm{E}\{y_v^2\} - \sum_{i=0}^{N-1} h_i \mathrm{E}\{y_v X[i]\}$$

Utilizzando la figura 6.6 è possibile riscrivere questa espressione nel caso in cui $y_v = X[n]$. Se $X[n]$ è stimato con un predittore ad un passo di p elementi, l'espressione di $J_{\min}$ si ottiene sostituendo alla coppia $(h_i, X[i])$ la coppia $(c_i, X[n-i])$, da cui

$$J_{\min} = \mathrm{E}\{X[n]^2\} - \sum_{i=1}^{p} c_i \mathrm{E}\{X[n]X[n-i]\} = \sum_{i=0}^{p} d_i r_X[i]$$

Questa equazione insieme al sistema (6.18) dà luogo ad un nuovo sistema di $(p+1)$ equazioni in $(p+1)$ incognite rappresentate dai parametri c_i e $J_{\min}$. Tale sistema si scrive normalmente nella forma

$$\begin{aligned}
r_X[0] - c_1 r_X[1] - c_2 r_X[2] - \ldots - c_p r_X[p] &= J_{\min} \\
r_X[1] - (c_1 r_X[0] + c_2 r_X[1] + c_3 r_X[2] + \ldots + c_p r_X[p-1]) &= 0 \\
r_X[2] - (c_1 r_X[1] + c_2 r_X[0] + c_3 r_X[1] + \ldots + c_p r_X[p-2]) &= 0 \\
\vdots \quad \vdots \quad \vdots \\
r_X[p] - (c_1 r_X[p-1] + c_2 r_X[p-2] + \ldots + c_p r_X[0]) &= 0
\end{aligned} \tag{6.21}$$

Per risolvere il sistema si può usare opportunamente l'algoritmo di Levinson, oppure una sua variante, nota come *algoritmo di Levinson-Durbin*, che si avvantaggia della struttura più semplice del secondo membro e inoltre fornisce anche $J_{\min}$.

Si osservi ancora che sulla base delle precedenti considerazioni è possibile dedurre il predittore opportuno per stimare i campioni antecedenti a quelli acquisiti o i campioni intermedi a due sequenze di campioni acquisiti. In queste sede non ci si occuperà ulteriormente di queste categorie di predittori.

6.3 Il filtro di Wiener

Si consideri il caso in cui l'insieme dei dati misurati $x[n]$ rappresenti i campioni di una realizzazione di un processo casuale a tempo discreto $X[n]$ del tipo

$$X[n] = S[n] + W[n] \tag{6.22}$$

dove $S[n]$ è un processo di segnale e $W[n]$ è un processo di rumore. Si supponga di volere stimare il campione della sola componente di segnale al passo $N-1$, cioè

$s[N-1]$. Utilizzando uno stimatore lineare del tipo descritto dalla relazione (6.1), si può definire un valore stimato $\hat{s}[N-1]$ come segue

$$y = \hat{s}[N-1] = \sum_{n=0}^{N-1} h_n x[n] \tag{6.23}$$

Applicando un criterio di ottimizzazione ai minimi quadrati, come visto nel paragrafo 6.1, si ottiene il sistema di equazioni normali espresso dalla relazione (6.5), che può essere riscritto come

$$\sum_i h_i \mathrm{E}\{X[l]X[i]\} = \mathrm{E}\{S[N-1]X[l]\} \qquad l = 0, 1, \ldots, N-1 \tag{6.24}$$

E' possibile, utilizzando la stessa tecnica vista nel paragrafo 6.2.1, interpretare la relazione (6.23) come un'operazione di filtraggio e scrivere

$$\hat{s}[N-1] = \sum_{i=0}^{N-1} \gamma_{N-1-i} x[i] \tag{6.25}$$

Il filtro risultante è un filtro FIR che prende il nome di *filtro di Wiener* di lunghezza finita.

L'equazione (6.24) può essere scritta in termini di coefficienti γ_i. Ponendo $k = N-1-i$ e $m = N-1-l$, si ottiene

$$\sum_{k=0}^{N-1} \gamma_k \mathrm{E}\{X[N-1-m]X[N-1-k]\} = \mathrm{E}\{S[N-1]X[N-1-m]\} \qquad m = 0, 1, \ldots, N-1 \tag{6.26}$$

Se $X[n]$ e $W[n]$ sono processi WSS l'equazione (6.26) diventa

$$\sum_{k=0}^{N-1} \gamma_k r_X[m-k] = R_{SX}[m] \qquad m = 0, 1, \ldots, N-1 \tag{6.27}$$

6.3.1 Il filtro di Wiener per segnale e rumore scorrelati

Una situazione molto frequente si ha quando $W[n]$ e $S[n]$ sono scorrelati e a media nulla. Con queste ipotesi si dimostra facilmente che $r_X[m] = r_S[m] + r_W[m]$ e $r_{SX}[m] = r_S[m]$. In questo caso il sistema di equazioni normali diventa

$$\sum_{k=0}^{N-1} \gamma_k r_X[m-k] = r_S[m] \qquad m = 0, 1, \ldots, N-1 \tag{6.28}$$

Le equazioni normali di questa forma prendono il nome di *equazioni di Wiener–Hopf*.

E' possibile generalizzare i risultati appena ottenuti riformulando il problema nel modo seguente.

Si consideri un insieme illimitato di dati $x[n] = s[n] + w[n]$, con $n = 0, 1, \ldots$, e si supponga di volere stimare la componente di segnale al generico passo ν utilizzando il campione attuale e quelli passati di $x[n]$, cioè i campioni corrispondenti a $n \leq \nu$. Si definisce la stima

$$\hat{s}[\nu] = \sum_{i=0}^{\nu} \gamma_{\nu-i} x[i] \tag{6.29}$$

E' evidente che il problema si riconduce facilmente a quello precedente. Basta porre $N-1 = \nu$ e supporre ν variabile. Le equazioni normali, nel caso di processi WSS, rimangono le (6.28). Il filtro che si ottiene è chiaramente non invariante; infatti per ogni valore di ν si avrà un filtro di lunghezza crescente e con valori dipendenti da ν. Per valori sufficientemente grandi di ν il filtro tende a diventare invariante. Il risultato asintotico è il *filtro di Wiener di lunghezza infinita*, con equazioni normali del tipo

$$\sum_{k=0}^{N-1} \gamma_k r_X[m-k] = r_S[m] \qquad \text{con } m = 0, 1, \ldots \tag{6.30}$$

In [5, pag.404] si osserva che la soluzione asintotica coincide con quella che si otterrebbe, ad ogni passo n, con uno stimatore "causale del tipo

$$\hat{S}[n] = \sum_{i=0}^{+\infty} \gamma_i X[n-i] \tag{6.31}$$

La funzione di trasferimento del filtro asintotico può essere espressa nel dominio z come descritto in [5, pag.413].

6.3.2 Il filtro di Wiener per sequenze infinite

Il problema del filtro di Wiener può essere ulteriormente riformulato, considerando una sequenza $x[n]$ di durata infinita ($-\infty < n < +\infty$). Di nuovo si consideri soltanto il caso in cui $x[n]$ sia una realizzazione del processo espresso dall'equazione (6.22).

Si vuole stimare il valore del segnale ad un generico istante ν per mezzo di una stima lineare del tipo

$$\hat{s}[\nu] = \sum_{i=-\infty}^{+\infty} \gamma_{\nu-i} x[i]$$

Si può dimostrare che le equazioni normali associate a tale stimatore si ottengono dall'equazione (6.26) e si possono scrivere come

$$\sum_{k=-\infty}^{+\infty} \gamma_k \mathrm{E}\{X[\nu-m]X[\nu-k]\} = \mathrm{E}\{S[\nu]X[\nu-m]\} \qquad (6.32)$$
$$\text{con } -\infty < m < +\infty$$

Da questa espressione si vede che il filtro di Wiener, in questo caso, è definito da infiniti coefficienti γ_k; pertanto ad ogni istante ν si devono valutare infiniti coefficienti, i cui valori dipendono da tutti i valori passati e futuri del processo $X[n]$. Si può osservare che come conseguenza si ha un filtro non causale e non invariante nel tempo.

Filtro invariante. Il filtro diventa invariante nel caso in cui $X[n]$ e $W[n]$ siano processi WSS. Infatti, in questo caso, l'equazione (6.33) diventa

$$\sum_{k=-\infty}^{+\infty} \gamma_k r_X[m-k] = r_{SX}[m] \qquad \text{con } -\infty < m < +\infty \qquad (6.33)$$

quindi indipendente da ν. L'equazione (6.33) esprime una convoluzione tra sequenze e quindi può essere espressa nel dominio z, come

$$\Gamma(z) = \sum_{k=-\infty}^{+\infty} \gamma_k z^{-k} = \frac{S_{SX}(z)}{S_X(z)}$$

dove $S_X(z)$ è la trasformata z dell'autocorrelazione $r_X[n]$ e $S_{SX}(z)$ è la trasformata z della mutuacorrelazione $r_{SX}[n]$.

Se i due processi $S[n]$ e $W[n]$ sono scorrelati e a media nulla si ottiene il filtro di Wiener

$$\Gamma(z) = \sum_{k=-\infty}^{+\infty} \gamma_k z^{-k} = \frac{S_S(z)}{S_X(z)}$$

dove $S_S(z)$ è la trasformata z dell'autocorrelazione $r_S[n]$.

6.3.3 Il filtro di Wiener con sequenze di apprendimento

Si consideri il caso in cui l'insieme dei dati misurati $x[n]$ rappresenti i campioni di una realizzazione di un processo casuale a tempo discreto $X[n]$ del tipo

$$X[n] = r[n] + W[n] \tag{6.34}$$

dove $r[n]$ non è più un processo casuale (come nel caso espresso dalla (6.22)), ma un segnale di riferimento noto, mentre $W[n]$ è di nuovo un processo di rumore. Il campione del segnale di riferimento al passo $N-1$, cioè $r[N-1]$, pur essendo noto, può essere stimato, utilizzando esattamente lo stesso tipo di stimatore lineare descritto dalla relazione (6.23), cioè

$$y = \hat{r}[N-1] = \sum_{n=0}^{N-1} h_n x[n] \tag{6.35}$$

Lo scopo di questa stima, apparentemente inutile, si vedrà tra breve. Per il momento si può dire che, applicando un criterio di ottimizzazione ai minimi quadrati, come visto nel paragrafo 6.1, si ottiene il sistema di equazioni normali espresso dalla relazione (6.24), che può essere riscritto come

$$\sum_i h_i \mathrm{E}\{X[l]X[i]\} = \mathrm{E}\{r[N-1]X[l]\} \qquad l = 0, 1, \ldots, N-1 \tag{6.36}$$

oppure in forma matriciale come

$$\mathbf{R}_X \, \mathbf{h}_N = \mathrm{E}\{r[N-1]\mathbf{X}_N\} \tag{6.37}$$

da cui si ottiene

$$\mathbf{h}_N = \mathbf{R}_X^{-1} \mathrm{E}\{r[N-1]\mathbf{X}_N\} \tag{6.38}$$

E' evidente che il risultato ottenuto non è nient'altro che il vettore dei coefficienti del filtro di Wiener (letto nel verso opportuno). Tale filtro può essere utilizzato per ripulire al meglio dal rumore (nel senso MMSE) un segnale $r[n]$, per $n > N-1$, quando sono noti i primi N campioni del segnale stesso. In altri termini, in alcune applicazioni è possibile addestrare un sistema di *denoising* (cioè un sistema che riduce il rumore) con un segnale di riferimento noto. Una volta addestrato, il sistema continuerà a filtrare il rumore in modo efficiente (ottimo nel senso MMSE), fino a quando le quantità espresse nella relazione (6.38) saranno in accordo con le caratteristiche del segnale entrante.

6.3.4 Generalizzazione del filtro di Wiener

L'espressione (6.38) può essere generalizzata, svincolandosi dalla sua interpretazione in termini di filtro LTI. Si consideri a questo scopo:

- Un segnale di riferimento $r[n]$ ad un generico istante n,
- Un vettore $\mathbf{x}[n] = (\, x_1[n] \quad x_2[n] \quad \cdots \quad x_N[n] \,)$ di N dati misurati $x_i[n]$, a cui è associato un vettore $\mathbf{X}[n] = (\, X_1[n] \quad X_2[n] \quad \cdots \quad X_N[n] \,)$ di N processi $X_i[n]$ con statistica congiunta nota (o parzialmente nota). Questo caso interessa quando i dati vengono misurati contemporaneamente in diversi punti dello spazio (ad esempio all'uscita di una schiera di antenne).

Ad ogni istante discreto n è possibile costruire un vettore di pesi, generalizzando l'espressione (6.38) come segue

$$\mathbf{h}[n] = \mathbf{R}_X^{-1} \mathrm{E}\{r[n]\mathbf{X}[n]\} \tag{6.39}$$

dove $\mathbf{R}_X$ rappresenta, a questo punto, una sorta di "autocorrelazione istantanea dei dati misurati. Applicando tali pesi ai dati misurati $X_i[n]$ si estrae da essi, istante per istante, la migliore stima (nel senso MMSE) del segnale $r[n]$.

Nelle applicazioni pratiche le due medie congiunte $\mathbf{R}_X$ e $\mathrm{E}\{r[n]\mathbf{X}[n]\}$ non sono disponibili e quindi si deve ricorrere a loro stime.

Un altro metodo, relativamente semplice, per ottenere $\mathbf{h}[n]$ è noto come algoritmo LMS (*Least Mean Square*). Tale metodo, descritto nel paragrafo seguente, usa un meccanismo iterativo per arrivare alla minimizzazione dell'MSE.

6.3.5 Algoritmo LMS

Il metodo LMS è basato sul fatto che la (6.39) deriva dalla minimizzazione del valore quadratico medio del segnale errore

$$\varepsilon[n] = r[n] - \mathbf{h}^H[n]\mathbf{X}[n]$$

cioè dell'espressione

$$\begin{aligned} \mathrm{MSE}_n &= E\{|\varepsilon[n]|^2\} \\ &= \mathrm{E}\{|r[n]|^2\} + \mathbf{h}^H[n]\mathbf{R}_X\mathbf{h}[n] - 2\mathbf{h}^H[n]\mathbf{z} \end{aligned} \tag{6.40}$$

dove $\mathbf{z} = \mathrm{E}\{r[n]\mathbf{X}[n]\}$ e il segnale $r[n]$ può essere in generale sia deterministico che casuale. La superfice MSE è una funzione quadratica di $\mathbf{h}[n]$ ed è minimizzata imponendo che il gradiente rispetto a $\mathbf{h}[n]$ sia zero. La (6.39) si ottiene appunto da questo processo di minimizzazione, da cui si ottiene che il minimo MSE, cioè l'MMSE è dato da

$$\mathrm{MMSE}_n = \mathrm{E}\{|r[n]|^2\} - \mathbf{z}^H \mathbf{R}_X^{-1} \mathbf{z}$$

L'idea alla base dell'algoritmo LMS è quella di calcolare i coefficienti (o pesi) $\mathbf{h}[n]$ in modo iterativo, a partire da valori iniziali (anche del tutto arbitrari), fino a raggiungere il minimo della superfice quadratica. Si procede come segue.

- Si osserva che, in certo istante n, il gradiente dell'MSE dipende dai pesi $\mathbf{h}[n]$. Infatti si ha

$$\nabla_{\mathbf{h}} \mathrm{MSE}_n|_{\mathbf{h}=\mathbf{h}[n]} = 2\mathbf{R}_X \mathbf{h}[n] - 2\mathbf{z}$$

- Si cerca una stima $g(\mathbf{h}[n])$ di tale gradiente. Nel caso più semplice tale stima si ottiene sostituendo $\mathbf{R}_X$ e $\mathbf{z}$ con le loro stime rumorose, ottenute direttamente dai dati misurati, cioè

$$g(\mathbf{h}[n]) = 2\mathbf{x}[n]\mathbf{x}^H[n]\mathbf{h}[n] - 2\mathbf{x}[n]r[n]$$

- Al passo $n+1$ i pesi vengono aggiornati in modo da spostarsi verso il punto della superfice MSE a gradiente nullo. Tale aggiornamento è governato da un parametro μ secondo la formula

$$\mathbf{h}[n+1] = \mathbf{h}[n] - \mu g(\mathbf{h}[n])$$

  Lo scalare positivo μ controlla le caratteristiche di convergenza dell'algoritmo.

La scelta del parametro μ è uno dei punti critici dell'algoritmo. Dettagli si possono trovare in [25].

Spiegazione semplice dell'algoritmo LMS Un modo semplice per capire l'algoritmo LMS è considerare il caso monodimensionale. Si consideri una funzione del tipo $y = (x-a)^2$, con derivata $dy/dx = 2x - 2a$. Il valore di x che minimizza y si può trovare in forma iterativa, scrivendo

$$x_m(n+1) = x_m(n) - \mu \left.\frac{dy}{dx}\right|_{x=x_m(n)}$$

dove x_m è il punto di minimo che si sta cercando ed n il passo di iterazione. Disegnando la parabola e la sua derivata il meccanismo iterativo appare evidente. In formule si ha

$$x_m(n+1) = x_m(n) - \mu(2x_m(n) - 2a)$$

Se il meccanismo converge e $x_m(n+1) \to x_m(n)$, allora, sostituendo nell'espressione precedente, si ha

$$x_m(n) \to a$$

che è il valore cercato. Se ne conclude che il minimo si può ottenere calcolando una serie di derivate, invece che azzerando la derivata stessa.

7

Stima spettrale parametrica con modelli AR

Le tecniche spettrali basate sul periodogramma consentono di ottenere una stima spettrale soltanto dall'elaborazione dei dati di misura. Le tecniche di tipo parametrico utilizzano invece un modello di generazione del processo e hanno come obbiettivo la stima dei parametri del modello.

La qualità dei risultati ottenuti con le tecniche parameriche dipende da quanto fedelmente il modello descriva il fenomeno che ha prodotto il processo, dai criteri di stima e dagli algoritmi di calcolo. Essendo l'argomento molto vasto, per una trattazione approfondita si rimanda il lettore ai due testi classici di Marple [13] e di Kay [14]. Gli stessi argomenti sono trattati anche in [15] e [12] ma con un approccio più orientato ai problemi della metrologia.

In questo capitolo verranno dati i concetti introduttivi della stima spettrale parametrica. Verranno esaminati soltanto i modelli di tipo AR, fino ad arrivare alla scrittura delle equazioni di Yule-Walker, che costiuiscono la base dei metodi parametrici.

La stima spettrale parametrica, a differenza della stima a periodogramma, è basata sull'ipotesi che esista un modello di generazione del processo in esame e che la struttura di tale modello sia in parte nota. La conoscenza, anche parziale, del modello costituisce un'informazione importante, in grado di migliorare la qualità della stima stessa, anche avendo a disposizione pochi campioni del processo. In un certo senso si può dire che sapere che il processo sia riconducibile ad un modello con una struttura ben precisa compensa il fatto di operare una stima spettrale

a partire da un insieme limitato di dati. E' evidente che il miglioramento in termini di stima è legato al fatto di operare con un modello 'corretto'; se il modello ipotizzato non è adeguato, invece di un miglioramento si può ottenere una stima spettrale decisamente errata.

Un'altra proprietà interessante dei metodi di stima spettrale parametrica riguarda il problema del troncamento dei dati. E' noto che con le tecniche non parametriche gli spettri vengono stimati a partire da sequenze che sono implicitamente o esplicitamente moltiplicate nel tempo per delle funzioni a finestra. In altre parole gli spettri vengono stimati osservando solo dei segmenti temporali di una data sequenza, e non tenendo conto di tutta la sequenza. Ciò si traduce nell'assumere che le sequenze di dati o i coefficienti di correlazione al di fuori di queste finestre siano identicamente nulli. Ovviamente questa ipotesi non è realisticamente plausibile. Inoltre è proprio la presenza delle finestre che, come è stato visto, limita la risoluzione dello stimatore spettrale. Quando si fa riferimento ad un modello per la generazione di un processo casuale, si suppone che tale modello generi tutti i campioni della sequenza e non solo quelli osservati in un dato intervallo di tempo. In questo modo si effettua un'estrapolazione e non vi è più l'effetto di troncamento delle finestre. Pertanto si incrementa la risoluzione dello stimatore.

Il problema del troncamento si rivela importante in tutti quei casi in cui la sequenza di dati disponibile per la stima è particolarmente breve, sia perchè essa costituisce tutta l'informazione disponibile o perchè il processo può essere considerato WSS solo per brevi intevalli di tempo.

7.1 Modello AR

E' noto dal teorema di Wold che un processo $X[n]$ WSS generato da un modello AR può essere espresso come

$$X[n] = -a_1 X[n-1] - a_2 X[n-2] - \ldots - a_M X[n-M] + W[n] \tag{7.1}$$

dove $W[n]$ è un processo bianco a media nulla e varianza σ_W^2 e a_i sono i parametri del modello. Tale modello ha una funzione di trasferimento a tutti poli del tipo

$$H(z) = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \ldots + a_M z^{-M}}$$

quindi lo spettro di $X[n]$ ha la forma

$$S_X(\mathrm{e}^{j2\pi f}) = |H(\mathrm{e}^{j2\pi f})|^2 \sigma_W^2 = \frac{\sigma_W^2}{|1 + a_1 \mathrm{e}^{-j2\pi f} + \ldots + a_M \mathrm{e}^{-j2\pi f M}|^2} \tag{7.2}$$

Nel seguito un processo con uno spettro di questo tipo verrà brevemente indicato come *processo AR*.

Si vedrà nel paragrafo 7.2 che i coefficienti a_i possono essere ottenuti utilizzando un predittore lineare di ordine $p > M$, il cui vettore dei coeffcienti $\mathbf{c}$ risulta essere $\mathbf{c} = [\,-a_1 \quad \ldots \quad -a_M \quad 0 \quad \ldots \quad 0\,]^T$.

Si possono usare anche altri criteri di stima ottenendo modelli diversi, come quelli trattati in [3, paragrafo 6.5] e [3, pag. 432].

Il vero problema di tutti questi metodi è che l'autocorrelazione $r_X[l]$ non è mai nota in forma analitica, ma deve essere stimata. Questo problema è trattato nel paragrafo 7.4.

7.2 Il predittore lineare e il modello AR

I modelli AR possono essere messi in relazione con i filtri per la predizione lineare visti nel capitolo 6. Si consideri in particolare un predittore del tipo descritto dall'equazione (6.17), cioè definito come

$$\hat{x}[n] = \sum_{i=1}^{p} c_i x[n-i]$$

e si applichi tale predittore ai dati estratti da una singola realizzazione $x[n]$ di un processo $X[n]$ generato da un modello AR. La realizzazione $x[n]$ si può scrivere nella forma

$$x[n] = -a_1 x[n-1] - a_2 x[n-2] - \ldots - a_M x[n-M] + w[n] \tag{7.3}$$

L'errore di predizione $e_x[n] = x[n] - \hat{x}[n]$, nel caso $p = M$, assume la forma

$$\begin{aligned} e_x[n] &= -[(a_1 + c_1)x[n-1] + (a_2 + c_2)x[n-2] + \ldots + (a_p + c_p)x[n-p]] + w[n] \\ &= \sum_{i=1}^{p} \alpha_i x[n-i] + w[n] \end{aligned}$$

dove $\alpha_i = -(a_i + c_i)$. Tale errore è di tipo casuale e quindi può essere visto come una realizzazione di un processo $E_X[n]$ del tipo

$$E_X[n] = \sum_{i=1}^{p} \alpha_i X[n-i] + W[n] \tag{7.4}$$

Se $X[n]$ è WSS anche $E_X[n]$ è WSS.

Come è noto dalla teoria illustrata nel capitolo 6, i coefficienti c_i del predittore lineare ad un passo sono quelli che minimizzano il funzionale di costo $J = \mathrm{E}\{|X[n] - \hat{X}[n]|^2\}$, cioè $J = \mathrm{E}\{|E_X[n]|^2\}$. Utilizzando la (7.4) J assume la forma

$$\begin{aligned} J &= \mathrm{E}\{\left|\sum_{i=1}^{p} \alpha_i X[n-i] + W[n]\right|^2\} \\ &= \sum_{i=1}^{p}\sum_{m=1}^{p} \alpha_i \alpha_m \mathrm{E}\{X[n-i]X^*[n-m]\} + \sigma_W^2 + 2\sum_{i=1}^{p} \alpha_i \mathrm{E}\{X[n-i]W[n]\} \\ &= \sum_{i=1}^{p}\sum_{m=1}^{p} \alpha_i \alpha_m \mathrm{E}\{X[n-i]X^*[n-m]\} + \sigma_W^2 \end{aligned}$$

essendo $\mathrm{E}\{X[n-i]W[n]\} = 0$, in quanto $W[n]$ è indipendente dai campioni di $X[n]$ negli istanti $n-i$, con $i > 1$; ciò è dovuto al fatto che il sistema generatore è causale e quindi $W[n]$ sarà correlato soltanto con i valori futuri di $X[n]$ e non con quelli passati.

Si consideri ora la minimizzazione del funzionale J. Derivando rispetto ad una generica α_i e uguagliando a zero si ottiene

$$\sum_{i=1}^{p} \alpha_i \mathrm{E}\{X[n-i]X^*[n-m]\} = 0$$

da cui si ha che deve essere $\alpha_i = 0$, e quindi

$$c_i = -a_i \qquad i = 1, 2, \ldots, p \tag{7.5}$$

Tali valori sostituiti nell'espressione di J consentono di trovarne il valore minimo, che risulta essere

$$J_{\min} = \sigma_W^2 \tag{7.6}$$

7.2.1 Dalle equazioni normali alle equazioni di Yule-Walker

I risultati ottenuti nel paragrafo precedente portano a concludere che i coeffcienti del predittore di ordine $p = M$ di un processo AR di ordine M coincidono, a meno del segno, con i coefficienti del modello AR. Pertanto le equazioni normali

associate al predittore possono essere utilizzate per valutare il modello AR. Utilizzando le equazioni (7.5) e (7.6) le equazioni normali (6.21) si possono riscrivere come segue

$$\begin{aligned}
r_X[0] + a_1 r_X[1] + a_2 r_X[2] + \ldots + a_p r_X[p] &= \sigma_W^2 \\
r_X[1] + a_1 r_X[0] + a_2 r_X[1] + a_3 r_X[2] + \ldots + a_p r_X[p-1] &= 0 \\
r_X[2] + a_1 r_X[1] + a_2 r_X[0] + a_3 r_X[1] + \ldots + a_p r_X[p-2] &= 0 \\
&\vdots \\
r_X[p] + a_1 r_X[p-1] + a_2 r_X[p-2] + \ldots + a_p r_X[0] &= 0 \qquad (7.7)
\end{aligned}$$

da cui si deduce che i parametri del modello AR possono essere ricavati dalla conoscenza della matrice (con dimensione $p \times p$) di autocorrelazione. Si noti che lo spettro, calcolato come trasformata dell'autocorrelazione $r_X[l]$, richiede la conoscenza di $r_X[l]$ per tutti i valori di l. Le equazioni (7.7) consentono invece di calcolare i parametri a_i e σ_W^2 e quindi lo spettro del processo soltanto a partire dalla matrice di autocorrelazione ($p \times p$).

Le equazioni (7.7) sono note come *equazioni di Yule-Walker*. Esse hanno esattamente la stessa struttura delle equazioni normali (6.21) e pertanto possono essere risolte utilizzando l'*algoritmo di Levinson-Durbin*. La differenza tra i due sistemi di equazioni è soltanto nel significato delle incognite.

In particolare, un'interpretazione significativa può essere data del risultato $J_{\min} = \sigma_W^2$. Infatti, osservando che un predittore che opera su un processo con una struttura come quella espressa dalla (7.1) non può predire la componente $W[n]$, che è per sua natura non prevedibile, $J_{\min}$ esprime la potenza dell'errore non prevedibile.

Si osservi che l'errore residuo di un predittore lineare non è necessariamente bianco. Invece nei modelli AR, ottenuti come appena descritto, l'errore residuo è bianco, essendo il modello AR ottenuto dalla rappresentazione di Wold.

Se un processo $X[n]$ è di tipo AR, ma non sono noti i coefficienti a_i e nemmeno M, si pone il problema della scelta dell'ordine p del predittore. Se il valore scelto per p è tale per cui $p \geq M$ il predittore lineare dà come risultato un vettore dei coefficienti $\mathbf{c}$ (di dimensioni $p \times 1$) del tipo

$$\mathbf{c} = [\, -a_1 \quad \ldots \quad -a_M \quad 0 \quad \ldots \quad 0 \,]^T$$

Ciò significa che scegliendo p sufficientemente elevato, sarà sempre possibile trovare il modello AR. Nelle applicazioni pratiche ciò non accade in maniera così

precisa perché le equazioni di Yule-Walker vengono risolte usando delle stime dell'autocorrelazione e non il valore vero.

La stretta analogia tra il modello AR e il predittore lineare può essere estesa alla funzione di trasferimento $H_p(z)$ del filtro PEF introdotto nel paragrafo 6.2.3. Una proprietà notevole del PEF è data dal fatto che l'operatore $F(z) = 1/H_p(z)$ (filtro di analisi) è un filtro FIR i cui zeri giacciano all'interno del cerchio di raggio unitario. In altri termini $F(z)$ è un filtro a fase minima e quindi il suo inverso $H_p(z)$ è sempre stabile. Una dimostrazione di questa importante proprietà è riportata nell'appendice 3D di [3].

7.3 Metodo diretto per ottenere le equazioni di Yule-Walker

Le equazioni di Yule-Walker si possono ottenere anche operando direttamente sul processo, senza passare dal predittore a un passo.

A questo scopo si consideri un processo AR di ordine M, ossia un processo descritto dalla (7.1) e se ne valuti l'autocorrelazione. Si ottiene

$$\begin{aligned} r_X[l] &= \mathrm{E}\{X[n]X^*[n-l]\} \\ &= -\mathrm{E}\{\sum_{i=1}^{M} a_i X[n-i]X^*[n-l]\} + \mathrm{E}\{W[n]X^*[n-l]\} \\ &= -\sum_{i=1}^{M} a_i r_X[l-i] + \mathrm{E}\{W[n]X^*[n-l]\} \end{aligned}$$

da cui

$$r_X[l] + \sum_{i=1}^{M} a_i r_X[l-i] = \mathrm{E}\{W[n]X^*[n-l]\} \tag{7.8}$$

Si consideri ora il secondo membro $\mathrm{E}\{W[n]X^*[n-l]\}$. Esso esprime la mutua correlazione tra ingresso e uscita del filtro AR. Si consideri un generico istante n_0. Poiché il filtro è causale il campione $X[n_0 - l]$, supposto $l > 0$, non dipende dal campione $W[n_0]$. Inoltre $W[n]$ è un processo bianco, pertanto $W[n_0]$ non è correlato con $W[n]$ per qualsiasi $n \neq n_0$. Ne segue che il termine $\mathrm{E}\{W[n]X^*[n-l]\}$ risulta nullo per ogni $l > 0$.

Lo stesso risultato può essere ottenuto esprimendo $X[n]$ come convoluzione della risposta all'impulso del filtro $h[n]$ e del processo bianco $W[n]$ come segue

$$X[n] = \sum_{i=0}^{\infty} h[i]W[n-i] \tag{7.9}$$

da cui si ha

$$\begin{aligned}
\mathrm{E}\{W[n]X^*[n-l]\} &= \mathrm{E}\{W[n]\sum_{i=0}^{\infty} h[i]W^*[n-l-i]\} \\
&= \sum_{i=0}^{\infty} h[i]r_W[l+i] \\
&= \sum_{i=0}^{\infty} h[i]\sigma_W^2\delta[l+i] \\
&= \begin{cases} 0 & \text{se } l > 0 \\ h[0]\sigma_W^2 & \text{se } l = 0 \\ h[l]\sigma_W^2 & \text{se } l < 0 \end{cases}
\end{aligned} \tag{7.10}$$

Utilizzando il teorema del valore iniziale si dimostra facilmente che

$$h[0] = \lim_{z\to\infty} \sum_{i=0}^{\infty} h[i]z^{-i} = \lim_{z\to\infty} H(z) = 1$$

Pertanto dalla (7.8) si può ottenere il sistema di equazioni

$$\begin{aligned}
r_X[0] + a_1 r_X[1] + a_2 r_X[2] + \ldots + a_M r_X[M] &= \sigma_W^2 \\
r_X[1] + a_1 r_X[0] + a_2 r_X[1] + a_3 r_X[2] + \ldots + a_M r_X[M-1] &= 0 \\
r_X[2] + a_1 r_X[1] + a_2 r_X[0] + a_3 r_X[1] + \ldots + a_M r_X[M-2] &= 0 \\
&\vdots \\
r_X[M] + a_1 r_X[M-1] + a_2 r_X[M-2] + \ldots + a_M r_X[0] &= 0
\end{aligned} \tag{7.11}$$

che coincide con il sistema di equazioni (7.7), ovvero con le *equazioni di Yule-Walker*.

7.4 Stima spettrale basata sulle equazioni di Yule-Walker

Un punto di partenza abbastanza ovvio per stimare lo spettro di un processo $X[n]$ di tipo AR è quello di utilizzare le equazioni di Yule-Walker sostituendo

all'autocorrelazione $r_X[l]$ una sua stima $\hat{r}_X[l]$. Ad esempio si può usare la stima espressa dalla (3.15), cioè

$$\hat{r}'_L[l] = \frac{1}{L} \sum_{i=0}^{L-|l|-1} x[n]x[n+|l|] \tag{7.12}$$

Si osservi che questa è una stima polarizzata. Come suggerito nel paragrafo 3.4.1 i valori attendibili di $\hat{r}'_L[l]$ sono in prima approssimazione quelli contenuti nell'intervallo $|l| < L/4$, da cui la regola empirica $L > 4p$. Utilizzando tale stima (o un'altra equivalente) si possono stimare i coefficienti a_i del modello AR e la varianza del rumore σ^2_W, risolvendo le equazioni di Yule-Walker (7.7).

Indicando con $\hat{a}_i$ i coefficienti stimati e con $\hat{\sigma}^2_W$ la varianza stimata, si può introdurre una stima spettrale del tipo

$$S_{Y,W}(\mathrm{e}^{j2\pi f}) = \frac{\hat{\sigma}^2_W}{|1 + \hat{a}_1 \mathrm{e}^{-j2\pi f} - \ldots + \hat{a}_p \mathrm{e}^{-j2\pi fp}|^2}$$

dove il pedice Y,W indica che si tratta di uno spettro stimato per mezzo delle equazioni di Yule Walker. Il metodo usato è noto come *Metodo di stima spettrale AR alla Yule-Walker*.

L'uso di uno stimatore polarizzato come quello definito dall'equazione (7.12) è molto frequente poiché permette di avere matrici di correlazione stimate semidefinite positive. Questa condizione è particolarmente importante per evitare instabilità numeriche negli algoritmi di inversione di matrice, utilizzati per cercare la soluzione di un sistema di equazioni lineari.

Questo metodo è affidabile quando si ha a disposizione un numero sufficientemente elevato di campioni. Quando invece si ha solo una sequenza molto breve a disposizione è opportuno utilizzare altri metodi di stima AR per ottenere una risoluzione più elevata. Tra questi si può citare il *metodo di Burg*, [13].

7.5 Scelta dell'ordine p del modello AR

Una volta che si è determinato che una stima attraverso un modello AR è appropriata, prima di effetture la stima dello spettro di un dato processo è necessario definire l'ordine p del modello. Ovviamente si sarebbe portati ad acquisire il minor numero M di campioni, purché consistente con la struttura dei dati, specie per limitare il più possibile la complessità computazionale. D'altra parte se si

ha un basso numero di campioni a disposizione la scelta di un modello AR di ordine p troppo elevato può portare ad avere una stima dello spettro poco affidabile, per la presenza di effetti spuri, dovuti, ad esempio, ad una stima inadeguata dell'autocorrelazione.

Nel caso poi in cui il processo sia poi di tipo MA o ARMA o sia semplicemente disturbato dalla presenza di un rumore addittivo, allora l'ordine del modello AR dovrebbe essere infinito, poiché qualsiasi modello di ordine finito darebbe luogo a stime polarizzate.

Nella pratica si è soliti calcolare la stima dello spettro per diversi ordini p entro un intervallo ragionevole, per valutare in un secondo tempo quale è il valore di p più idoneo sulla base di un determinato criterio. Ad esempio si potrebbe dedurre un criterio per la scelta di p, dall'osservazione dei valori minimi $J_{\min} = \hat{\sigma}^2_W$ assunti dal funzionale di costo al variare di p. Una scelta possibile, in questi casi, consiste nel prendere il più basso valore di p che garantisca di mantenere il valore minimo del funzionale al di sotto di una soglia prefissata. Resta da chiarire come definire quasta soglia, ovvero quando si può considerare accettabile l'errore residuo a seguito della minimizzazione del funzionale.

Un altro criterio potrebbe essere dato dalla scelta dell'ordine minimo tale che l'errore residuo sia bianco, dal momento che l'operazione di filtraggio del modello AR dovrebbe lasciare un errore bianco. Tuttavia non è chiaro come si debba stimare lo spettro dell'errore residuo.

Evidentemente la scelta dell'ordine del modello AR non è banale. In letteratura sono stati proposti diversi criteri dei quali menzioniamo i principali.

Criterio dell'errore di predizione finale. E' basato sulla funzione, indicata come FPE (dall'inglese *Final Prediction Error*),

$$\mathrm{FPE}[p] = \hat{\sigma}^2_{W\,p} \left[\frac{L + (p-1)}{L - (p+1)} \right] \tag{7.13}$$

dove L è la lunghezza della sequenza acquisita e $\hat{\sigma}^2_{W\,p}$ rappresenta la varianza dell'errore residuo associato al modello di ordine p.

La funzione FPE$[p]$ è data dal prodotto di due fattori: $\hat{\sigma}^2_{W\,p}$ che diminuisce al crescere di p e la frazione che segue che invece cresce con p. L'ordine del modello scelto è quello che rende minima la funzione FPE$[p]$.

Si dimostra che questo criterio tende a sottostimare il vero ordine del modello AR.

Criterio dell'informazione di Akaike. E' basato sulla funzione

$$\mathrm{AIC}[p] = L \ln(\hat{\sigma}^2_{W\,p}) + 2p \tag{7.14}$$

Anche in questo caso si hanno due termini di cui il primo diminuisce con p mentre il secondo cresce. L'ordine del modello viene scelto in modo da rendere minima la funzione AIC[p].

Questo criterio tende a sovrastimare l'ordine vero del modello AR.

Criterio della funzione di trasferimento autoregressiva. Si introduce la funzione

$$\mathrm{CAT}[p] = \left[\frac{1}{L}\sum_{i=1}^{p} \frac{1}{\hat{\sigma}^2_{W\,p}\sqrt{L/(L-i)}}\right] - \frac{1}{\hat{\sigma}_{W\,p}[L/(L-p)]} \tag{7.15}$$

Ancora una volta l'ordine del modello viene scelto in modo da rendere minima la funzione CAT[p].

Come conclusione si noti che nessun criterio è in generale preferibile agli altri.

Bibliografia

[1] **J. G. Proakis, D. G. Manolakis,** *Introduction to Digital Signal Processing,* Prentice-Hall, Englewood Cliffs, NJ, 1993

[2] **A. Papoulis,** *Probability, Random Variables, and Stochastic Processes,* McGraw-Hill Book Company, New York, 1965

[3] **P. M. Clarkson,** *Optimal and Adaptive Signal Processing*, CRC Press, Boca Raton, Florida, 1993

[4] **J. G. Proakis, C. M. Rader, F. Ling, C. L. Nikias** *Advanced digital signal Processing,* Mcmillman Publishing Company, New York, 1992.

[5] **S. M. Kay,** *Fundamentals of Statistical Signal Processing: Estimation Theory,* Prentice-Hall, Englewood Cliffs, NJ, 1993

[6] **L. L. Scharf,** *Statistical Signal Processing. Detection, Estimation, and Time Series Analysis*, Addison-Wesley, July 1991

[7] **B. Porat,** *Digital Processing of Random Signals. Theory and Methods*, Prentice-Hall, Englewood Cliffs, NJ, 1994

[8] **J. M. Mendel,** *Lessons in Digital Estimation Theory,* Prentice-Hall, Englewood Cliffs, NJ, 1987

[9] **W. Gardner,** *Introduction to Random Processes with Applications to Signals and Systems,* McGraw-Hill Publishing Company, 1990

[10] **R. M. Gray, L. D. Davisson,** *Random Processes: A Mathematical Approach for Engineers,* Prentice-Hall, Englewood Cliffs, NJ, 1986

[11] **H. Stark, J. W. Woods,** *Probability, Random processes, and Estimation Theory for Engineers,* Seconda edizione, Prentice-Hall, Englewood Cliffs, NJ, 1994

[12] **D.B. Percival, A. T. Walden,** *Spectral Analysis for Physical Applications,* Cambridge University Press, New York, 1993

[13] **S. L. Marple,** *Digital Spectral Analysis with Applications,* Prentice-Hall, Englewood Cliffs, NJ, 1987

[14] **.... Kay,** *.....,* Prentice-Hall, Englewood Cliffs, NJ, ...

[15] **M. B. Priestley,** *Spectral Analysis and Time Series,* Academic Press, London,1981

[16] **Robert N. McDonough, Anthony D. Whalen** *Detection of Signals in Noises* Academic Press, Second edition, London,1971

[17] **Steven J. Leon** *Linear Algebra with Applications* MacMillan, Fourth Edition, 1994

[18] **D.J. Thomson,** *Spectrum estimation and harmonic analysis,* Proc. IEEE, vol. 70, pp. 1055-1096, 1982

[19] **L. Lo Presti, F. Neri,** *L'analisi dei segnali,* Seconda edizione, CLUT, Torino, settembre 1992

[20] **L. Lo Presti, F. Neri,** *Introduzione ai processi casuali,* CLUT, Torino, novembre 1992

[21] **A. De Marchi, L. Lo Presti,** *Incertezze di misura: come quantificare l'incertezza nella conoscenza della realtà,* CLUT, Torino, novembre 1993

[22] **W. W. S. Wei,** *Time series analysis: Univariate and Multivariate Methods,* Addison-Wesley, 1990.

[23] M. S. Bartlett, "On the theoretical specification of sampling properties of autocorrelated time series, *Journal of Royal Statistic Soc. Suppl.,* vol. 8, pp. 27–41, 1946.

[24] F. Harris, "On the use of windows for harmonic analysis with a discrete Fourier transform, *Proceedings of the IEEE,* Vol. 1, No. 66, pp. 51–83, January, 1978.

[25] Lac C. Godara, "Application of Antenna Array to Mobile Communications, Part II: Beam-Forming and Direction-of-Arrival Considerations, *Proceedings of the IEEE,* Vol. 85, No. 8, August 1997.