

# Elaborazione Numerica dei Segnali Parte II

**Fabio Rocca**

*Dipartimento di Ingegneria Elettronica  
e Informazione  
Politecnico di Milano*

Edizioni CUSL  
Milano

---

tutti i diritti riservati  
ISBN 88-8132-026-6

# Indice

<b>Prefazione</b>	<b>xiii</b>
<b>1 Analisi spettrale non parametrica</b>	<b>1</b>
1.1 Introduzione . . . . .	1
1.2 Analisi spettrale non parametrica . . . . .	3
1.2.1 Stima della potenza di segnali correlati gaussiani	4
1.3 Risoluzione spettrale e dispersione . . . . .	9
1.3.1 La dispersione della stima spettrale . . . . .	9
1.3.2 La polarizzazione della stima spettrale . . . . .	10
1.4 Il periodogramma . . . . .	11
1.5 La stima dell'autocorrelazione . . . . .	13
1.6 Errori da evitare . . . . .	17
1.7 Modifiche del periodogramma . . . . .	18
1.7.1 Smussamento dello spettro . . . . .	18
1.7.2 Media di periodogrammi di sottosequenze . . . . .	21
1.7.3 La tecnica WOSA . . . . .	22
1.8 Appendice 1 . . . . .	24
1.8.1 Processi casuali discreti . . . . .	24
1.8.2 Stima a massima verosimiglianza (da Wikipedia)	26
1.8.3 La stima lineare MMSE . . . . .	29
1.8.4 Sequenze correlate e loro spettri . . . . .	30
1.8.5 Matrici di Toeplitz: autovalori . . . . .	35
1.8.6 Un esempio numerico . . . . .	38
1.8.7 Stime degli autovalori di matrici di covarianza . .	41
1.9 Appendice 2 . . . . .	44
1.9.1 Autocorrelazione e spettro di potenza . . . . .	44
1.9.2 Dimensione delle variabili continue e discrete . . .	45
1.9.3 Teorema del limite centrale . . . . .	46

## INDICE

<b>2</b>	<b>Analisi Spettrale Parametrica</b>	<b>49</b>
2.1	Introduzione . . . . .	49
2.2	Analisi tutti zeri (MA) . . . . .	50
2.2.1	Troncamento della funzione di autocorrelazione . . . . .	52
2.3	Analisi tutti poli (AR) . . . . .	54
2.4	La stima spettrale AR . . . . .	57
<b>3</b>	<b>La predizione lineare</b>	<b>59</b>
3.1	L'errore di predizione . . . . .	61
3.2	Predizione lineare e analisi AR . . . . .	62
3.2.1	Allungamento del predittore . . . . .	63
3.3	La ricorsione di Levinson . . . . .	65
3.3.1	Vantaggi computazionali . . . . .	67
3.4	Il filtro a traliccio non ricorsivo . . . . .	67
3.5	Descrizioni alternative dell'autocorrelazione . . . . .	70
3.6	Ordine del predittore . . . . .	72
3.6.1	Risoluzione spettrale . . . . .	72
3.6.2	Lunghezza dei dati e criterio di Akaike . . . . .	73
3.7	Condizionamento di $\mathbf{R}$ . . . . .	75
3.8	Stima con rumore . . . . .	76
3.8.1	Il caso $\sigma_z^2 = 0$ . . . . .	80
3.8.2	Il caso $\sigma_z^2 \neq 0$ . . . . .	80
3.8.3	Predizione, stima e sbiancamento: un riassunto . . . . .	80
3.9	La deconvoluzione nel dominio delle frequenze . . . . .	84
3.9.1	Deconvoluzione predittiva nelle frequenze . . . . .	87
3.10	Le equazioni Y - W nel dominio degli autovettori . . . . .	88
<b>4</b>	<b>Filtraggio adattativo</b>	<b>91</b>
4.1	La tecnica di Widrow e Hoff . . . . .	93
4.2	Dispersione degli autovalori . . . . .	96
4.3	Rapidità della convergenza . . . . .	97
4.3.1	Dispersione del filtro errore $\mathbf{g}_i$ . . . . .	98
4.4	Ottimizzazione del passo di aggiornamento . . . . .	103
<b>5</b>	<b>Le direzioni di arrivo</b>	<b>105</b>
5.1	DOA con il periodogramma . . . . .	107
5.2	Stima AR: la risoluzione angolare . . . . .	108
5.2.1	Stima spettrale di Capon . . . . .	112

## INDICE

5.3	Reiezione direzionale del disturbo . . . . .	114
5.4	Appendice . . . . .	116
5.4.1	Stima lineare ottima di costanti o pendenze . . .	116
5.4.2	Esempio: Stima AR della frequenza di un processo casuale quasi monocromatico . . . . .	119
5.5	Stima ottima (U. Spagnolini) . . . . .	121
5.5.1	La regressione lineare . . . . .	121
5.5.2	Stima ottima di una frequenza . . . . .	123
5.5.3	Discussione sulla stima di una frequenza . . . . .	124
5.6	Stima di processi multidimensionali . . . . .	125
5.6.1	L'effetto cocktail e la separazione delle sorgenti .	127
<b>6</b>	<b>Il segnale vocale</b>	<b>131</b>
6.1	Tecniche di codifica del segnale vocale . . . . .	134
6.2	Codifica delle forme d'onda . . . . .	134
6.2.1	La Modulazione Delta . . . . .	135
6.2.2	L'ADPCM . . . . .	138
6.2.3	La codifica lineare predittiva . . . . .	140
6.3	La codifica del segnale musicale . . . . .	141
6.4	Cenni al riconoscimento del parlato . . . . .	142
<b>7</b>	<b>Segnali bidimensionali</b>	<b>143</b>
7.1	Le singolarità a 2D . . . . .	143
7.2	Sistemi lineari 2D . . . . .	147
7.2.1	Le funzioni cilindriche . . . . .	147
7.3	Le sinusoidi 2D . . . . .	149
7.4	L'antitrasformazione di Fourier 2D . . . . .	149
7.4.1	Funzioni a simmetria circolare . . . . .	151
7.5	Proprietà della trasformazione . . . . .	154
7.5.1	Funzioni cilindriche e separabili . . . . .	154
7.6	Filtri ideali 2D . . . . .	156
7.7	Filtri a ventaglio . . . . .	157
<b>8</b>	<b>Filtri 2D di natura fisica</b>	<b>159</b>
8.1	L'equazione dei potenziali . . . . .	159
8.1.1	Continuazione in basso e predizione lineare . . . .	161
8.2	La propagazione di onde . . . . .	163
8.2.1	Il problema del telerilevamento . . . . .	165

## INDICE

8.2.2	Le onde piane . . . . .	166
8.3	Il principio della fase stazionaria . . . . .	170
8.4	Il <i>chirp</i> , la lente sottile e l'ottica di Fourier . . . . .	171
8.5	L'equazione della diffusione . . . . .	176
<b>9</b>	<b>Sequenze 2D</b>	<b>179</b>
9.1	L'equivocazione a 2D . . . . .	180
9.1.1	Effetti di moiré . . . . .	181
9.1.2	Il campionamento esagonale . . . . .	183
9.2	Il campionamento del segnale televisivo . . . . .	186
9.2.1	Equivocazione spazio - temporale . . . . .	187
9.2.2	Lo spettro del segnale televisivo campionato . . . . .	189
9.3	Ricostruzione 2D . . . . .	190
9.4	Trasformata di Fourier discreta a 2D . . . . .	192
9.5	Filtraggi con prodotto di trasformate . . . . .	194
9.6	Filtri 2D . . . . .	194
9.6.1	La trasformazione di Mc Clellan . . . . .	196
9.6.2	Filtri per la continuazione di campo (wavefield continuation) . . . . .	197
9.7	Interpolazione mediante stima lineare . . . . .	199
9.7.1	Processi casuali a 2D . . . . .	200
9.7.2	Spettri di potenza a 2D . . . . .	200
9.7.3	Stima lineare a 2D . . . . .	201
9.7.4	Stima lineare ed interpolazione . . . . .	203
9.8	Poligoni di Voronoi e triangoli di Delaunay . . . . .	204
9.8.1	Il campionamento a 3D . . . . .	206
<b>10</b>	<b>La Tomografia</b>	<b>209</b>
10.1	Introduzione . . . . .	209
10.2	Trattamento nei numeri d'onda . . . . .	210
10.3	Il sinogramma . . . . .	215
10.4	La retroproiezione . . . . .	217
10.5	La retroproiezione filtrata . . . . .	219
10.5.1	Equivocazione angolare . . . . .	222
10.5.2	Retroproiezioni e numeri d'onda . . . . .	223
10.5.3	Proiezioni incomplete o mosse . . . . .	227
10.6	Tomografia 3D . . . . .	228
10.7	Tecniche iterative . . . . .	230

## INDICE

10.7.1	Retroproiezione e gradiente . . . . .	230
10.7.2	Retroproiezione e trasposizione . . . . .	231
10.7.3	Interpretazione stocastica della pseudo inversione	233
10.7.4	Interpretazione tomografica del gradiente . . . . .	234
<b>11</b>	<b>Tecniche di ottimizzazione</b>	<b>237</b>
11.1	Il gradiente . . . . .	237
11.2	Il gradiente coniugato . . . . .	239
11.2.1	Ortogonalità coniugata, incorrelazione e stima lineare	239
11.2.2	Connessione con la ricorsione di Levinson . . . . .	242
11.3	Gradiente e gradiente coniugato . . . . .	245

# Prefazione

Il corso di Elaborazione numerica dei segnali viene tenuto al Politecnico di Milano sin dall'A/A 1984 – 85. In generale, gli allievi hanno seguito al secondo anno di studi della laurea un corso che contiene elementi di *Teoria dei Segnali* dove vengono studiati i sistemi lineari, le trasformazioni di Fourier continua e discreta, i principi di teoria delle probabilità ed i processi stocastici. Inoltre, sono disponibili corsi di *Algoritmi e Circuiti di Comunicazione, Elaborazione e Trasmissione delle Immagini, Tecniche di Telerilevamento, “Applicazioni di Telerilevamento”* dove vengono sviluppate con maggiore respiro le possibili applicazioni. Il programma completo si è mantenuto pressoché eguale in tutti questi anni. La verifica viene fatta con un esame scritto.

Il testo è suddiviso in due volumi dedicati alle due parti del corso rispettivamente; in esso, ho cercato per quanto possibile di partire da considerazioni elementari dando immagini intuitive (spesso scritte in corsivo) di molti tra gli argomenti trattati. Certamente la trattazione analitica sarebbe stata guida più sicura, e la via intuitiva non è per questo più semplice; però, è più facile da memorizzare. Ho desiderato dare allo studente strumenti di tipo ingegneristico più che matematico per comprendere le relazioni intercorrenti tra ipotesi presunte ed risultati ottenuti per permettergli di valutare cosa può accadere nei frequenti casi in cui queste ipotesi non sono più attendibili. La frequentazione del corso di *Teoria dei Segnali* è comunque di grande aiuto per avere dimestichezza con il materiale trattato.

Il corso è di prevalente natura metodologica, anche se vengono sempre proposte applicazioni. La prima parte verte sui fondamenti dell'elabo-



## PREFAZIONE

razione lineare dei segnali e sui filtri numerici. La seconda affronta l'analisi spettrale monodimensionale con tecniche parametriche e non parametriche. Tra le applicazioni ho presentato sia il filtraggio adattativo che le schiere di sensori. Nell'ultima parte questi concetti sono estesi a segnali a due dimensioni; il corso si conclude con lo studio della tomografia e con la sua generalizzazione effettuata introducendo il problema inverso; quest'ultima parte non entra nel programma di insegnamento, ma viene comunque inclusa nel testo per introdurre un'interpretazione stocastica del gradiente coniugato e sottolinearne la stretta relazione con la ricorrenza di Levinson. I problemi inversi sono peraltro studiati in profondità nel corso "*Applicazioni di Telerilevamento*".

Ho cercato la massima concisione per fornire un ausilio alla didattica in aula, piuttosto che uno strumento per l'autodidattica. In questo caso, consiglio di cominciare con il libro di Mitra, inserito tra i riferimenti bibliografici, ricco di esempi *MATLAB<sup>TM</sup>*. Il testo di Proakis e Manolakis è più completo, ma può risultare più difficile ad una prima lettura.

In questo testo non si troveranno esercizi perchè questi, insieme a complementi e a numerosi testi d'esame sono contenuti nell'eserciziario Elaborazione Numerica dei Segnali: Temi d'esame '96/00 a cura di U. Spagnolini, F. Rocca, L. Zanzi, Edizioni CUSL, Milano.

Aggiungo una breve bibliografia commentata perchè lo studente possa trovare riferimenti completi per ogni argomento del corso.

Il corso ha un sito web

[http://www.elet.polimi.it/dsp/courses/ens\\_11/](http://www.elet.polimi.it/dsp/courses/ens_11/) dove gli studenti troveranno esercizi, data e luogo degli esami, esiti, soluzioni dei compiti, e materiale utile in generale.

Bibliografia.

■ Mitra, S., 1998, Digital Signal Processing, a computer - based approach, McGraw-Hill, pp 864. Libro elementare, ricco di esempi basati su *MATLAB<sup>TM</sup>*.

■ Proakis J., Manolakis D., 1995: Digital Signal Processing, Prentice Hall, 947 pp. e quattro appendici. Libro introduttivo didattico, semplice ma molto diffuso, che copre sia le tematiche del corso di Teoria dei Segnali che quelle dei primi 14 capitoli del testo.

■ Mitra S., Kaiser J., 1993: Handbook for Digital Signal Processing, Wiley Interscience, 1168 pp. Completo manuale che copre ampiamente i primi 14 capitoli del testo. Mancano la parte applicativa al segnale vocale, i filtri 2D, la tomografia.

## PREFAZIONE

■ Fliege, N., 1994, Multirate Digital Signal Processing, J. Wiley & Sons, pp340. Un buon libro sui filtri polifase e sui problemi di decimazione e interpolazione.

■ Haykin, S., 1996, Adaptive Filter Theory, Terza edizione, Prentice Hall Inc. Libro di testo fondamentale per il filtraggio adattativo.

■ Kondo, A. M., 1994, Digital Speech, J. Wiley & Sons. Un recente libro sulla codifica numerica del segnale vocale.

■ Bracewell, R. N., 1995, Two Dimensional Imaging, Prentice Hall International Editions. Ottimo libro, con buone qualità didattiche, su tutta la parte dei segnali 2D.



# Capitolo 1

## Analisi spettrale non parametrica

### 1.1 Introduzione

In questa seconda parte del testo, affronteremo il problema dell'analisi spettrale dei segnali e cioè la stima del loro spettro di potenza.

Il problema della stima spettrale é molto vasto ed ha applicazioni molto diverse tra loro. Per introdurre il problema, si può pensare allo studio dello spettro delle radiofrequenze in una grande città; l'uno accanto all'altro si incontrano i vari canali radio e televisivi, i segnali di telefonia cellulare, i segnali radar ecc. Gli intervalli di frequenza, se considerati con sufficiente larghezza di banda, sono occupati da segnali totalmente indipendenti tra loro, con diversa ampiezza e diverse caratteristiche statistiche. Per analizzare lo spettro, appare logico utilizzare un ricevitore selettivo che misura il contenuto energetico in ogni intervallo di frequenza. Cercheremo la stima più precisa possibile di queste energie nell'intervallo di tempo disponibile senza fare alcuna ipotesi ulteriore, non cercando modelli della generazione del segnale (**analisi spettrale non parametrica**).

Un caso ben diverso si ha ad esempio considerando lo spettro di un segnale acustico dovuto a vibrazioni o ad un segnale vocale. In questo caso, tutto il segnale nel suo complesso ha origini uniche e quindi vi saranno delle relazioni tra i contenuti delle varie bande spettrali. Un caso estremo poi di analisi spettrale é quello della ricerca della frequenza

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

di un oscillatore o delle direzioni di arrivo di onde piane su una schiera di sensori. In questi casi, si deve anzitutto scegliere un modello per la generazione del segnale e quindi determinare i parametri del modello stesso. In questo modo, si dá la descrizione dello spettro del segnale piú accurata possibile. Ad esempio, si cercheranno i parametri di un filtro lineare che, alimentato da un segnale a spettro uniforme (rumore bianco) produce uno spettro di potenza simile allo spettro sotto analisi. Questa tecnica (**analisi spettrale parametrica**) ha numerosissime applicazioni: studiare le vibrazioni di un sistema, studiare la stabilità di frequenza di un oscillatore, stimare posizione e numero delle sorgenti di segnale in un campo acustico, stimare i parametri del tratto vocale di un parlatore, e in generale stimare e predire segnali nel tempo o nello spazio.

Ovviamente, il successo della tecnica parametrica dipende dalla qualità e dalla correttezza del modello scelto; modelli validi portano ad una descrizione **parsimoniosa** del segnale, caratterizzata cioè dal minimo numero di parametri necessario; come vedremo, ciò porterá ad una migliore stima di questi parametri e quindi a risultati ottimali. Infine, e questo é poi uno dei risultati piú importanti, l'analisi spettrale parametrica porta all'identificazione del modello e questo apre una successiva fase di studio per comprendere e poi eventualmente controllare lo stato e l'evoluzione del sistema sotto osservazione.

L'analisi spettrale non parametrica é una questione concettualmente semplice se si fa uso del concetto di media di insieme: infatti, se si hanno abbastanza realizzazioni del segnale, basta calcolarne le trasformate discrete di Fourier e mediane le potenze, componente per componente. Tuttavia, raramente si hanno numerose repliche del segnale; spesso, si ha a disposizione una sola replica, per un intervallo di tempo assegnato la cui lunghezza indicheremo sistematicamente con  $T_{tot} = NT$  dove  $T$  é l'intervallo di campionamento. Per determinare lo spettro di potenza, si dovrá ricorrere ad ulteriori ipotesi come stazionarietà ed ergodicità che supporremo sempre valide, tranne quando verranno discussi i sistemi adattativi, in cui la non stazionarietà delle statistiche verrà considerata esplicitamente.

## 1.2 Analisi spettrale non parametrica

Volendo stimare lo spettro di una sequenza  $\{x_n\}$ ,  $n = 0, \dots, N - 1$  senza fare alcuna ipotesi addizionale, è necessario utilizzare un banco di  $K$  filtri passa banda con bande  $B_k$ ;  $k = 1, \dots, K$  e frequenze centrali  $f_k$ ;  $k = 1, \dots, K$  attraverso i quali fare transitare il segnale; si stimano poi le potenze medie delle uscite  $y_{n,k}$  dei vari filtri del banco (fig. 10.1). Emerge il problema di determinare la dispersione della stima spettrale e cioè la dispersione della stima  $\hat{S}_k$  della potenza (il valore quadratico medio)  $\sigma_k^2$  dei segnali  $y_{n,k}$  in uscita dei vari filtri.

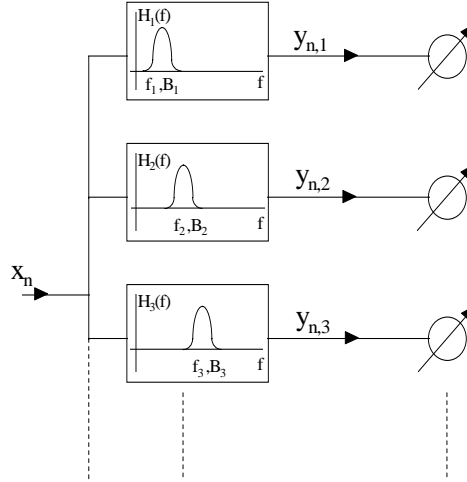


Figura 1.1: Stima dello spettro di un segnale

Questa stima è funzione sia del tempo totale di osservazione concesso  $T_{tot} = NT$  sia delle larghezze di banda  $B_k$  dei vari filtri (risoluzione spettrale), che non sono necessariamente eguali tra loro. Vedremo che dispersione della stima e risoluzione spettrale sono entrambe limitate dal valore di  $T_{tot}$  ed inoltre al migliorare della risoluzione spettrale peggiorerà (crescerà) la dispersione di  $\hat{S}_k$ ,  $\sigma_{\hat{S}_k}$ . Infatti, se la banda passante  $B_k$  del filtro  $k$ -esimo è molto stretta, i campioni (complessi) della sua uscita sono correlati tra loro per un intervallo di tempo lungo  $\Delta T_k \simeq 1/B_k$ . Quindi, il tempo totale necessario per misurare **all'uscita del filtro** un numero di campioni complessi indipendenti  $M$  abbastanza grande (per

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

avere una buona stima del loro valore quadratico medio e quindi della densità spettrale media in quella banda) sarà:

$$T_{tot} = NT = M\Delta T_k = M/B_k$$

e sarà tanto maggiore per quanto più ridotta è  $B_k$  e per quanto maggiore è  $M$ . A pari  $T_{tot}$  invece, dovremo scegliere se fare grande  $M$  (stima precisa) o  $\Delta T_k$  (elevata risoluzione spettrale).

### 1.2.1 Stima della potenza di segnali correlati gaussiani

#### Campioni indipendenti

Per iniziare lo studio della dispersione della stima del valore quadratico medio di un segnale di cui sono disponibili solo un numero limitato di campioni, eventualmente correlati tra loro, supponiamo di avere una sequenza di  $M$  campioni reali  $x_m$  che sappiamo essere **indipendenti e gaussiani**, a valor medio nullo e con eguale valore q.m.  $\sigma^2$ ; la migliore stima del loro v.q.m é:

$$\hat{S} = \frac{1}{M} \sum_{m=0}^{M-1} x_m^2$$

$\hat{S}$ , la stima della varianza della sequenza  $x_m$  é una variabile casuale, la cui distribuzione è nota come  $\chi^2$  con  $M$  gradi di libertà;  $\hat{S}$  è approssimativamente gaussiana, ha valor medio pari a  $\sigma^2$ ; la dispersione verrà indicata con  $\sigma_{\hat{S}}$ .  $\hat{S}$  non può essere gaussiana se non in modo approssimato, perché sempre positiva. Come verrà mostrato dall'equazione (1.4),  $\sigma_{\hat{S}}$  decresce con  $\sqrt{M}$ , e quindi lentamente.

Calcoleremo ora  $\sigma_{\hat{S}}$  nell'ipotesi che i campioni  $x_m$  siano incorrelati;

$$\begin{aligned} E[\hat{S}] &= \frac{1}{M} E\left[\sum_{m=0}^{M-1} x_m^2\right] = \frac{1}{M} \sum_{m=0}^{M-1} E[x_m^2] = \sigma^2 \\ \sigma_{\hat{S}}^2 &= E[\hat{S}^2] - E[\hat{S}]^2 = \frac{1}{M^2} \sum_{n,m=0}^{M-1} E[x_n^2 x_m^2] - \sigma^4 \\ &= \frac{ME[x_m^4] + (M^2 - M)E_{n \neq m}[x_n^2 x_m^2]}{M^2} - \sigma^4 = \end{aligned} \quad (1.1)$$

## 1.2 ANALISI SPETTRALE NON PARAMETRICA

$$\frac{3M\sigma^4 + (M^2 - M)\sigma^4}{M^2} - \sigma^4 = \frac{2\sigma^4}{M} \quad (1.2)$$

dove abbiamo utilizzato la relazione

$$E[x^4] = 3\sigma^4 \quad (1.3)$$

valida per variabili gaussiane a v.m. nullo, come sono le  $\{x_m\}$ . In conclusione abbiamo:

$$\sigma_{\hat{S}} = \sqrt{\frac{2}{M}}\sigma^2 = \sqrt{\frac{2}{M}}E[\hat{S}] \quad (1.4)$$

Supponiamo ora di voler calcolare il numero  $M$  di campioni indipendenti che permette una stima  $\hat{S}$  che, con una probabilità dell' 85%, sia differente dal valore vero  $\sigma^2$  di una quantità inferiore a  $\pm 0.05\sigma^2$ . Una variabile casuale gaussiana con varianza  $\sigma^2$  è contenuta entro un intervallo  $\pm k\sigma$  attorno al suo valor medio con una probabilità  $\text{erf}(k/\sqrt{2})$ ; poichè  $0.85 = \text{erf}(1.016)$ , si ha  $k = 1.016 \times \sqrt{2} = 1.43$ . Inoltre perchè sia

$$1.43\sigma_{\hat{S}} = 0.05\sigma^2$$

è necessario che:

$$\sigma_{\hat{S}} = \sqrt{\frac{2}{M}}\sigma^2 = \frac{\sigma^2}{28.6} \rightarrow M \simeq 1635$$

### Campioni correlati

Quando i campioni del segnale sono correlati, la dispersione della stima del loro v.q.m. cresce, perché decresce il numero equivalente di campioni indipendenti che contribuiscono alla stima: *se vogliamo stimare la distribuzione delle età dei cittadini, non è bene raccogliere dati in scuole elementari o case di riposo*. Fare valutazioni quantitative precise comporterebbe molti calcoli; cercheremo qui con un esempio di dare un'idea del problema.

Consideriamo due variabili casuali gaussiane a valor medio nullo ed eguale varianza che siano tra loro correlate, e le indichiamo con  $x, y$ . In generale possiamo supporre che sia:

$$y = \mu x + w; \quad E[x^2] = E[y^2] = \sigma^2; \quad (1.5)$$



## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

$$E[w^2] = \sigma^2(1 - \mu^2); \quad E[xw] = 0$$

dove  $w$  é una variabile casuale incorrelata con  $x$ . Risulta poi, utilizzando la solita relazione

$$\begin{aligned} E[x^4] &= 3\sigma^4 \\ E[xy] &= \mu\sigma^2; \quad E[x^2y^2] = (1 + 2\mu^2)\sigma^4 \end{aligned} \quad (1.6)$$

Consideriamo ora una sequenza  $\{x_m\}$  di campioni **correlati** tale che:

$$E[x_n x_m] = \mu\sigma^2; \quad n \neq m; \quad E[x_m^2] = \sigma^2$$

Ritornando ora alla stima del v.q.m. della sequenza:

$$\hat{S} = \frac{1}{M} \sum_{m=0}^{M-1} x_m^2$$

congruentemente con l'equazione (1.1) essa sarà caratterizzato dalla varianza

$$\sigma_{\hat{S}} = \sigma^2 \sqrt{\frac{2}{M}(1 - \mu^2) + 2\mu^2}$$

e la varianza **NON** decresce a zero anche se  $M$  tende ad  $\infty$ . Supponiamo ora che i campioni tendano a decorrelarsi tra loro al crescere della differenza di indice:

$$E[x_m x_{m+k}] = \rho^{|k|} \sigma^2; \quad \rho \leq 1 \quad (1.7)$$

La stima  $\hat{S}$  del v.q.m.  $\sigma^2$  della sequenza é sempre ottenuta come:

$$\hat{S} = \frac{1}{M} \sum_{m=0}^{M-1} x_m^2$$

Dopo qualche calcolo si trova che:

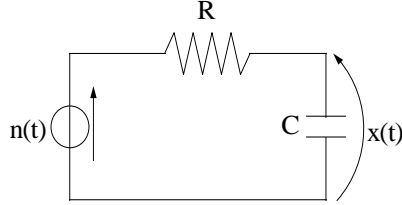
$$E[\hat{S}] = \sigma^2; \quad \sigma_{\hat{S}} \simeq \sigma^2 \sqrt{\frac{2}{M}} \times \sqrt{\frac{1 + \rho^2}{1 - \rho^2}} \quad (1.8)$$

Avendo quindi da stimare la varianza di una serie di  $M$  campioni non indipendenti, ma correlati con un decadimento esponenziale della correlazione di ragione  $\rho$ , é come avere a disposizione un numero equivalente inferiore di campioni indipendenti, pari a:

$$M_{eq} \simeq M \frac{1 - \rho^2}{1 + \rho^2} \quad (1.9)$$

e cioè decrescente con la correlazione tra i campioni.

## 1.2 ANALISI SPETTRALE NON PARAMETRICA



$$\tau = RC$$

Figura 1.2: Circuito RC alimentato da rumore bianco

### Segnali continui

Per proseguire l'esempio della stima della potenza di un segnale correlato, supponiamo di voler stimare la potenza in uscita di un circuito RC alimentato da rumore bianco (fig. 10.2) e quindi con uno spettro di potenza pari a:

$$\lim_{T_{tot} \rightarrow \infty} \frac{E[X(\omega)X(\omega)^*]}{T_{tot}} = S(\omega) = \frac{\lambda^2}{1 + (\omega\tau)^2} \quad (1.10)$$

Indichiamo con  $T_{tot}$  la durata dell'intervallo di osservazione, con  $\lambda^2$  la densità spettrale del rumore bianco in ingresso, l'intervallo di campionamento  $T = q\tau$ ,  $M = T_{tot}/q\tau$  il numero di campioni considerati; la correlazione tra i campioni successivi è:

$$\rho_q = e^{-q} \quad (1.11)$$

Il numero di campioni equivalente è:

$$M_{eq} \simeq \frac{T_{tot}}{T} \tanh q \quad (1.12)$$

La dispersione della stima è:

$$\sigma_{\hat{S}}^2 \simeq \sigma^4 \frac{2\tau}{T_{tot}} \frac{q}{\tanh q} \geq \sigma^4 \frac{2\tau}{T_{tot}} \quad (1.13)$$

Per  $q$  tendente a zero, e cioè campionando fittissimamente il segnale correlato uscente dal circuito  $RC$  (fig. 10.3),  $M$  tenderà all'infinito, ma

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

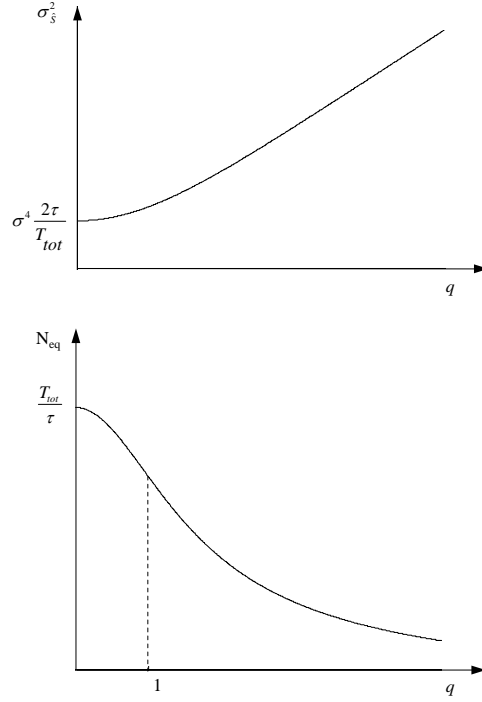


Figura 1.3: Varianza della misura in funzione della distanza  $qT$  tra i campioni (alta); Numero equivalente  $M_{eq}$  dei campioni (basso).

non  $M_{eq}$  perchè i campioni saranno sempre più correlati tra loro; risulta comunque:

$$\sigma_{\hat{s}} \geq \sigma^2 \sqrt{\frac{2\tau}{T_{tot}}} \quad (1.14)$$

e cioè anche se i campioni misurati sono in numero infinito, la dispersione **non tende a zero**. Se  $q$  cresce ma rimane inferiore circa ad 1, i campioni decrescono in numero e si decorrelano e la varianza della stima aumenta ma di poco. In compenso diminuisce drasticamente il costo computazionale. Se poi  $q$  cresce ancora, oltre il valore 1, i campioni diventano troppo pochi e la varianza cresce in proporzione a  $q$ . Da questo esempio si conclude che è inutile campionare in modo troppo fitto una serie correlata; la varianza della stima non scende perché i campi-

### 1.3 RISOLUZIONE SPETTRALE E DISPERSIONE

oni si assomigliano troppo. Peraltro, non bisogna essere troppo insistenti nell'imporre l'indipendenza dei campioni, perché se li si desidera troppo decorrelati, si perdono misure utili. E' ragionevole mantenere la correlazione intorno a 10 – 20%

Si osservi peraltro che nel caso di analisi spettrale di un segnale con un banco di filtri passa banda di banda  $B$  abbastanza selettivi, è ragionevole supporre che lo spettro del segnale sia uniforme in ogni banda (ma differente tra una banda e l'altra). Allora, se si riporta a frequenza zero ogni banda e si sottocampiona il segnale convertito alla frequenza di campionamento  $2B$ , si ottiene una sequenza di campioni reali indipendenti per la quale è semplice calcolare la dispersione della stima di potenza.

## 1.3 Risoluzione spettrale e dispersione

### 1.3.1 La dispersione della stima spettrale

Siamo ora in grado di valutare il necessario compromesso da raggiungere tra la risoluzione spettrale e la dispersione della stima se la sequenza di cui facciamo la stima spettrale è lunga  $N$  campioni e cioè è osservata in un intervallo di tempo  $T_{tot} = NT$ . Infatti sia  $B$  la risoluzione spettrale e quindi  $B$  sia la banda passante di ogni filtro da utilizzare per la stima dello spettro di potenza del segnale attorno ad una qualsiasi frequenza assegnata: in altre parole, la risoluzione spettrale non dipende dalla frequenza. Ripetendo in modo leggermente diverso il ragionamento fatto alla fine del paragrafo precedente, convertiamo in banda base il segnale passa banda uscita del filtro  $k$ -esimo. Centriamo quindi attorno alla frequenza 0 le componenti alle frequenze positive del segnale reale di partenza, eliminando del tutto quelle alle frequenze negative. Otteniamo così un segnale complesso con frequenza massima  $B/2$ ; se la banda  $B$  è piccola, sarà ragionevole ritenere uniforme lo spettro del segnale convertito in questa banda. Campionando la parte reale e la parte immaginaria del segnale complesso con l'intervallo di tempo  $\Delta T = 1/B$  e quindi alla frequenza  $B = 2 \times B/2$ , otteniamo due sequenze di campioni, tutti incorrelati tra loro (lo spettro è uniforme) e di pari v.q.m. Inoltre, poiché il filtro passa banda è a banda stretta e quindi combina linearmente tra loro numerosi campioni del segnale di ingresso campionati ad intervallo  $T \ll \Delta T$ , il segnale filtrato è gaussiano per il teorema del limite

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

centrale e pertanto l'incorrelazione tra i campioni implica anche la loro indipendenza. In ogni modo, **abbiamo per ogni banda**  $B$ ,  $M = 2BT_{tot}$  **campioni incorrelati.**

Quindi, utilizzando sia le parti reali che quelle immaginarie per la stima  $\hat{S}$  troviamo che:

$$\frac{\sigma_{\hat{S}}}{E(\hat{S})} = \frac{\sigma_{\hat{S}}}{\sigma^2} = \sqrt{\frac{2}{M}} = \sqrt{\frac{2}{2BT_{tot}}} = \frac{1}{\sqrt{BT_{tot}}} \quad (1.15)$$

Pertanto, se la risoluzione spettrale migliora ( $B$  decresce), aumenta anche la dispersione della stima; poiché però dobbiamo finestrare i dati nell'intervallo  $T_{tot}$  la risoluzione spettrale non può comunque essere maggiore di  $1/T_{tot}$  a causa della finestatura rettangolare che comporta la convoluzione con dei dati con  $W_r(\omega) = \text{sinc}(fT_{tot})$ .

### 1.3.2 La polarizzazione della stima spettrale

Ancora, la convoluzione con la trasformata della finestra comporta **la polarizzazione della stima.** Con questo termine si indica il fatto che:

$$E[\hat{S}_k] \neq \sigma_k^2 = S(\omega_k)B_k \quad (1.16)$$

e cioè che il valore medio dello stimatore è diverso dal valore medio della variabile casuale. Infatti, la presenza di componenti spettrali al di fuori della banda  $B_k$  (per esempio una sinusoide di grande ampiezza ad un'altra frequenza), a causa della finestra temporale adottata e quindi della convoluzione del segnale con  $W_r(\omega)$ , crea infiltrazioni spurie di energia (leaks) nella banda  $B_k$ : risulta infatti che la trasformata di Fourier del segnale finestrato è:

$$\begin{aligned} X_f(\omega) &= X(\omega) * W_r(\omega) \\ E[|X_f|^2] &= E\left[\int \int X(\psi) W_r(\omega - \psi) X^*(\eta) W_r^*(\eta - \omega) d\psi d\eta\right] = \\ &= \int \int W_r(\omega - \psi) W_r^*(\eta - \omega) E[X(\psi) X^*(\eta)] d\psi d\eta = \end{aligned}$$

e poichè, come vedremo in Appendice nel caso dei segnali campionati:

$$E[X(\psi) X^*(\eta)] = E[|X(\psi)|^2] \delta(\psi - \eta)$$

## 1.4 IL PERIODOGRAMMA

$$E[|X_f(\omega)|^2] = \int |W_r(\omega - \psi)|^2 E[|X(\psi)|^2] d\psi$$

e quindi lo spettro del segnale finestrato è la convoluzione dello spettro originale con il modulo quadro della trasformata della finestra

$$S_f(\omega) = S(\omega) * |W_r(\omega)|^2$$

e, nel caso discreto,

$$E[\hat{S}_k] = B_k \times (S(\omega) * |W_r(\omega)|^2)_{\omega=\omega_k}$$

La risoluzione spettrale peggiora se si usano finestre triangolari o cosinusoidali onde evitare che i lobi laterali della  $|W_r(\omega)|^2$ , a causa di componenti spettrali esterne all'intervallo di frequenza  $B_k$ , vi iniettino energia spuria e creino polarizzazione. Queste finestre rastremate diminuiscono la polarizzazione (i lobi laterali di ampiezza inferiore che con la finestra rettangolare) ma riducono la risoluzione spettrale. Il lobo centrale della trasformata della finestra avrà infatti larghezza maggiore che non quello corrispondente alla finestra rettangolare di eguale durata. Inoltre, la rastremazione della finestra porta a pesare poco dei dati e quindi ad accrescere la dispersione della stima.

Si osservi che la finestra che deve essere applicata ai dati per delimitarli temporalmente può essere qualsiasi, purchè abbia bassi lobi laterali. Vedremo invece più avanti che, quando finestreremo la funzione di autocorrelazione, avremo bisogno di finestre la cui trasformata di Fourier è sempre positiva.

## 1.4 Il periodogramma

Se abbiamo a disposizione solo un intervallo di osservazione dei dati di durata  $T_{tot}$ , abbiamo visto che il più piccolo intervallo di frequenza (che nel seguito indicheremo anche come cella) su cui si può avere una stima spettrale è:

$$B_{min} = \frac{1}{T_{tot}} = \frac{2}{N} f_N$$

Questa risoluzione spettrale (la massima possibile) si ottiene utilizzando una finestra rettangolare prima del filtraggio passa banda del segnale.

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

Dalla formula precedentemente vista, risulta che in questo caso:

$$\frac{\sigma_{\hat{S}}}{E(\hat{S})} = \sqrt{\frac{1}{B_{min}T_{tot}}} = 1 \quad (1.17)$$

e cioè la dispersione della stima non dipende dalla lunghezza dei dati; **allungando la serie dei dati e volendo sempre la massima risoluzione spettrale, la dispersione della stima non decresce perchè contemporaneamente cresce la risoluzione spettrale.** In questo caso, infatti, il numero  $M$  di campioni indipendenti usato per la stima dello spettro in ogni intervallo di frequenza è sempre 2.

Supponiamo ora che i filtri passa banda vengano realizzati con filtri a campionamento in frequenza, complessi, con frequenze centrali appunto

$$f_k = \frac{k}{T_{tot}}; \quad k = -\frac{N}{2} + 1, \dots, \frac{N}{2}$$

La risposta all'impulso del filtro  $k$ -esimo, complesso, è:

$$h_k(n) = e^{-j2\pi knT/T_{tot}} = e^{-j2\pi kn/N} = W_N^{-nk}; \quad n = 0, \dots, N-1 \quad (1.18)$$

$$W_N = e^{j2\pi/N}$$

L'ingresso del filtro è la sequenza  $\{x_n\}$ ,  $n = 0, \dots, N-1$ ; l'uscita del filtro, complessa, è valida solo all'istante  $n = N-1$  quando il transitorio iniziale di caricamento dei dati nel filtro è terminato, ma anche i dati sono terminati. In sostanza, con la sequenza di dati lunga  $N$ , per ogni filtro alla frequenza  $f_k$  disponiamo di solo un campione in uscita, complesso. Esso è:

$$X_k = \sum_{n=0}^{N-1} x_n W_N^{-nk} \quad (1.19)$$

Dimostreremo nell'appendice che gli  $X_k$ , uscite dei vari filtri passa banda che costituiscono il banco, sono sostanzialmente indipendenti tra loro, a garanzia che la risoluzione spettrale sia effettivamente  $1/NT$ . Determinare le  $N$  uscite  $X_k$  del banco degli  $N$  filtri passa banda corrisponde poi al calcolo della trasformata di Fourier discreta della sequenza.

## 1.5 LA STIMA DELL'AUTOCORRELAZIONE

L'unico campione a nostra disposizione dell'uscita del  $k$ -esimo filtro passa banda, il campione complesso  $X_k$ , ha potenza  $|X_k|^2$ ; questa, divisa per  $N$ , è la stima  $P_k$  della densità spettrale del segnale in quella banda; questa stima spettrale viene indicata con il nome di **periodogramma**.

Con il periodogramma abbiamo raggiunto la massima risoluzione possibile della stima spettrale e anche abbiamo rinunciato totalmente a migliorarne la dispersione: infatti, per fare la stima di energia usiamo solo  $M = 2$  dati indipendenti, e cioè le parti reale ed immaginaria di  $X(k)$ . Il periodogramma perciò ha massima risoluzione spettrale e massima dispersione della stima. Allungando  $T_{tot}$ , aumenta la risoluzione spettrale ma non migliora la dispersione. Peraltro, calcolare il periodogramma è molto poco costoso.

In figura 10.4 sono presentati gli spettri di potenza, calcolati in vari modi, corrispondenti ad un processo costituito da una senoide a frequenza 125 Hz ( $f_c = 1000\text{Hz}$ ) cui è aggiunta una sequenza di campioni indipendenti (rumore bianco) convoluti con la sequenza  $(-0.5, 0, 1, 0, -0.5)$  e quindi caratterizzata dallo spettro:

$$S(\omega) = [1 - \cos(2\omega T)]^2$$

In figura 10.4a è presentato il periodogramma calcolato su 2048 campioni. Si osservi il forte picco causato dalla senoide (tutta la potenza cade in una sola cella spettrale) e la forte dispersione della stima alle frequenze a cui lo spettro del segnale è diverso da zero. Dove invece l'ampiezza spettrale è piccola, anche la dispersione è piccola.

## 1.5 La stima dell'autocorrelazione

Stimare la densità spettrale equivale a stimare la funzione di autocorrelazione del processo: infatti l'una sarà la trasformata di Fourier dell'altra. Con autocorrelazione definiamo la seguente funzione,

$$r(\tau) = E[x_n^* x_{n+\tau}]; \quad r(-\tau) = r^*(\tau)$$

indipendente dall'indice  $n$ ; il processo causale  $\{x_n\}$  è a media nulla e può essere complesso. La trasformata di Fourier dell'autocorrelazione è quindi lo spettro di potenza

$$E[X(\omega)X^*(\omega)] \longleftrightarrow \sum_{n=0}^{N-1} E[x_n^* x_{n+\tau}]$$



## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

in quanto è la trasformata del valore atteso della convoluzione tra la sequenza e la stessa sequenza **ribaltata nel tempo e coniugata**.

E' quindi interessante individuare quale sia la stima dell'autocorrelazione corrispondente al periodogramma. La risposta e' semplice: per definizione il periodogramma è la sequenza di variabili casuali:

$$P_k = \frac{|X_k|^2}{N} = \frac{1}{N} \left( \sum_{n=0}^{N-1} x_n^* W_N^{nk} \right) \left( \sum_{n=0}^{N-1} x_n W_N^{-nk} \right) \quad (1.20)$$

esso quindi corrisponde, nel dominio della frequenza, alla convoluzione **circolare** della sequenza  $\{x_n\}$  con se stessa, ribaltata nei tempi e coniugata, se complessa (il coniugio della trasformata). Nel seguito, supporremo per semplicità che la sequenza  $\{x_n\}$  sia reale, tranne quando specificato. **Il periodogramma è la trasformata dell'autocorrelazione della sequenza periodicizzata**. Apparentemente, quindi la periodicizzazione ci porterebbe a gravi imprecisioni. Ma, stranamente, non è così: infatti si consideri la sequenza  $\{x_n\}$  nell'intervallo  $T_{tot}$  (la sequenza sia lunga  $N$  campioni) e il ragionevole stimatore **polarizzato** dell'autocorrelazione che vedremo corrispondere al periodogramma:

$$\hat{r}(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} x_n^* x_{n+\tau} \quad x_{n+\tau} = 0; \quad \forall n + \tau > N; \quad \forall n + \tau < 0;$$

Lo stimatore considerato è polarizzato perchè:

$$E[\hat{r}(\tau)] = \frac{N - |\tau|}{N} r(\tau) \neq r(\tau) \quad (1.21)$$

Ricordiamo qui che l'indice  $\tau$  dell'autocorrelazione viene anche chiamato in inglese *lag* (ritardo).

Questa stima dell'autocorrelazione si ottiene bordando con  $N$  zeri la sequenza  $\{x_n\}$  (fig. 10 .5) e poi facendone la correlazione con se stessa, eliminando quindi qualsiasi errore di periodicizzazione.

La Trasformata di Fourier  $X_k$  della sequenza  $x_n$  di lunghezza  $N$  bordata con  $N$  zeri è la Trasformata di Fourier della sequenza originale, interpolata ad intervallo di frequenza metà (e quindi pari a  $1/2NT$ ) con un sinc. Pertanto, i valori che assume alle frequenze  $f_k = k/NT$  sono uguali a quelli della Trasformata di Fourier della sequenza originale, **non bordata con zeri**. Quindi a parte un'irrelevante interpolazione su un

## 1.5 LA STIMA DELL'AUTOCORRELAZIONE

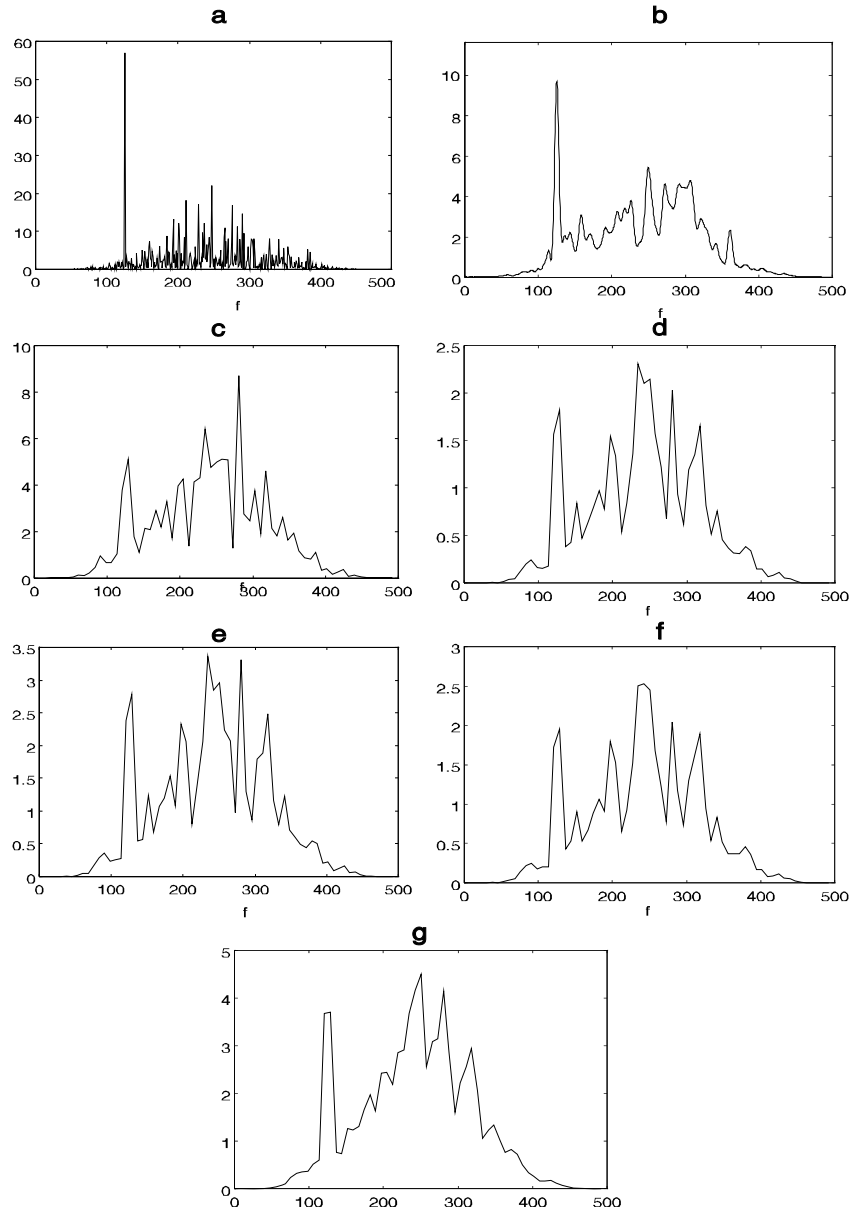


Figura 1.4: a) Spettro di potenza calcolato utilizzando un solo periodogramma su tutti i 2048 campioni; b) periodogramma precedente smussato c) Periodogramma ottenuto suddividendo i dati in 16 sottosequenze con finestre rettangolari e mediando le potenze; d) finestre triangolari e) finestre cosinusoidali f) finestre a coseno rialzato g) finestre a coseno rialzato sovrapposte (WOSA)

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

intervallo di frequenza metà la Trasformata di Fourier della  $\hat{r}_\tau$  coincide con il periodogramma, come si era detto; si deve ricordare tuttavia che l'interpolazione 2:1 deve essere fatta sulla trasformata  $X_k$  e non sul modulo quadrato  $|X_k^2|$ ; in altre parole, **prima si fa l'interpolazione e poi il modulo quadrato e non viceversa**. La periodicizzazione non ha influenza sulla stima spettrale anche se l'autocorrelazione periodicizzata é ben diversa da quella non periodicizzata. La polarizzazione nel do-

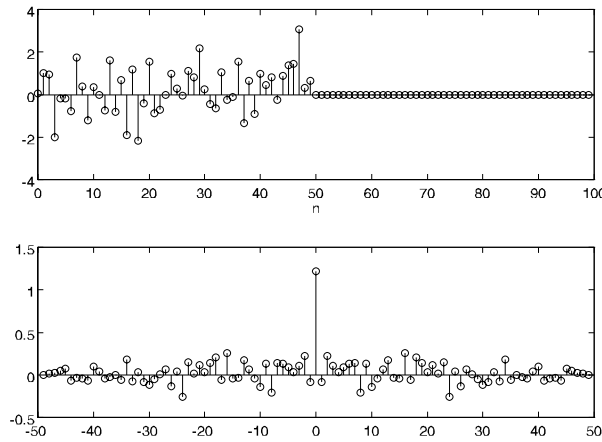


Figura 1.5: Sequenza di N campioni indipendenti bordata con N zeri e sua funzione di autocorrelazione

minio dei tempi che risulta dalla (10.23) corrisponde alla polarizzazione del periodogramma trovata nel dominio delle frequenze; sono entrambe dovute alla finestratura dei dati con una finestra rettangolare. Infatti, sia  $\{z_n\}$  una sequenza illimitata e  $\{x_n\}$  la sequenza ottenuta finestrando temporalmente la sequenza  $\{z_n\}$  con una finestra rettangolare  $w_{r,n}$  (fig. 10.6):

$$x_n = z_n \times w_{r,n}; \quad w_{r,n} = 1; \quad 0 \leq n \leq N-1 \quad (1.22)$$

$$E[\hat{r}(\tau)] = E\left[\frac{1}{N} \sum_{n=0}^{N-1} z_n w_{r,n} z_{n+\tau} w_{r,n+\tau}\right] = \frac{1}{N} \sum_{n=0}^{N-1} w_{r,n} w_{r,n+\tau} E[z_n z_{n+\tau}] = \quad (1.23)$$

## 1.6 ERRORI DA EVITARE

$$= r_{z,\tau} \frac{1}{N} \sum_{n=0}^{N-1} w_{r,n} w_{r,n+\tau} = r_{z,\tau} r_{w,\tau}$$

dove con  $r_{w,\tau}$  e  $r_{z,\tau}$  sono state indicate rispettivamente le autocorrelazioni della finestra e della sequenza di ingresso. Quindi, finestrare i dati comporta che il valore atteso della stima dell'autocorrelazione trovato sia il prodotto dell'autocorrelazione della sequenza originale per quella della finestra. Nel caso di finestra rettangolare, si ritrova il risultato che l'autocorrelazione "vera" è moltiplicata per una finestra triangolare che va a zero dopo  $N$  campioni.

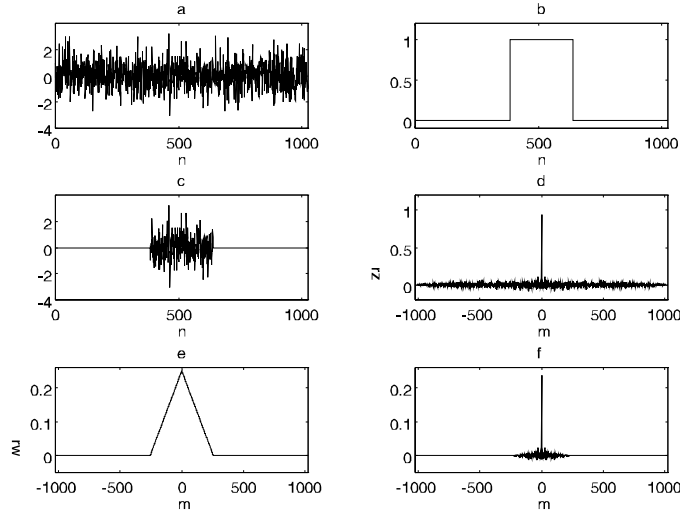


Figura 1.6: a) Sequenza illimitata  $z_n$  b) Finestra rettangolare  $w_n$  c) Sequenza finestrata  $x_n$  d) funzione di autocorrelazione della sequenza  $z_n$  e) funzione di autocorrelazione della finestra  $w_n$  f)  $r_z \times r_w$

## 1.6 Errori da evitare

Si potrebbe credere di potere fare meglio del periodogramma tentando di rimuoverne la polarizzazione ad esempio:

- a) depolarizzando la  $\hat{r}_{\tau}$  moltiplicandola per il fattore  $N/(N - |\tau|)$

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

b) utilizzando altri campioni del processo eventualmente disponibili e utilizzando come stimatore dell'autocorrelazione la funzione:

$$\hat{r}_{\cdot\tau} = \frac{1}{N} \sum_{n=0}^{N-1} x_n^* x_{n+\tau} \quad x_{n+\tau} \neq 0; \quad \forall n + \tau > N$$

Operare così porta a gravi errori anche se apparentemente si fa una cosa utile. Infatti, l'autocorrelazione polarizzata dell'equazione (1.21) ha spettro tutto positivo (il periodogramma appunto), come dovuto, ed il suo valore nell'origine, somma di tutti i termini dello spettro, tutti positivi, è massimo (le due questioni sono evidentemente collegate). Le autocorrelazioni non polarizzate ottenute con gli stimatori a, b non hanno affatto questa garanzia; sia nel caso a) che nel caso b) potrebbe benissimo accadere che  $\hat{r}_{\cdot\tau} > \hat{r}_0$  per qualche valore del ritardo  $\tau$ . Inoltre, nel caso a) la crescente amplificazione della stima dell'autocorrelazione per ritardi crescenti (dovuta alla divisione per  $N - |\tau|$  numero sempre più vicino allo zero) peraltro necessaria per evitare la polarizzazione, comporta un inaccettabile aumento della dispersione.

## 1.7 Modifiche del periodogramma

### 1.7.1 Smussamento dello spettro

Usando lo stimatore  $\hat{r}(\tau)$  viene fornita una stima dell'autocorrelazione fino a ritardi  $\tau \leq N$ ; è chiaro tuttavia che per ritardi grandi, l'autocorrelazione è stimata utilizzando ben pochi campioni (un solo campione se  $\tau = N - 1$ ) e quindi la stima ha grande dispersione; ciò risulta anche dal periodogramma dove la risoluzione spettrale è  $1/NT$  e quindi corrisponde ad un'autocorrelazione stimata fino al ritardo  $NT$ . Ciò porta ad un'elevata dispersione. Appare opportuno modificare il periodogramma in modo da ridurre sia la dispersione che la risoluzione spettrale.

Per ridurre la risoluzione spettrale basta p.e. mediare i valori assunti dalla stima spettrale fornita dal periodogramma in  $L = 2N/P$  intervalli di frequenza adiacenti; ciò corrisponde a smussare il periodogramma stesso convolvendolo con un operatore di filtro. Così facendo, si decresce anche la dispersione della stima in quanto si mediano tra loro stime basate su  $M = 2L$  variabili casuali indipendenti  $\hat{S}_k = |X_k|^2 = A_k^2 + B_k^2$  (vedi Appendice). Per effettuare poi questa riduzione della risoluzione spettrale

## 1.7 MODIFICHE DEL PERIODOGRAMMA

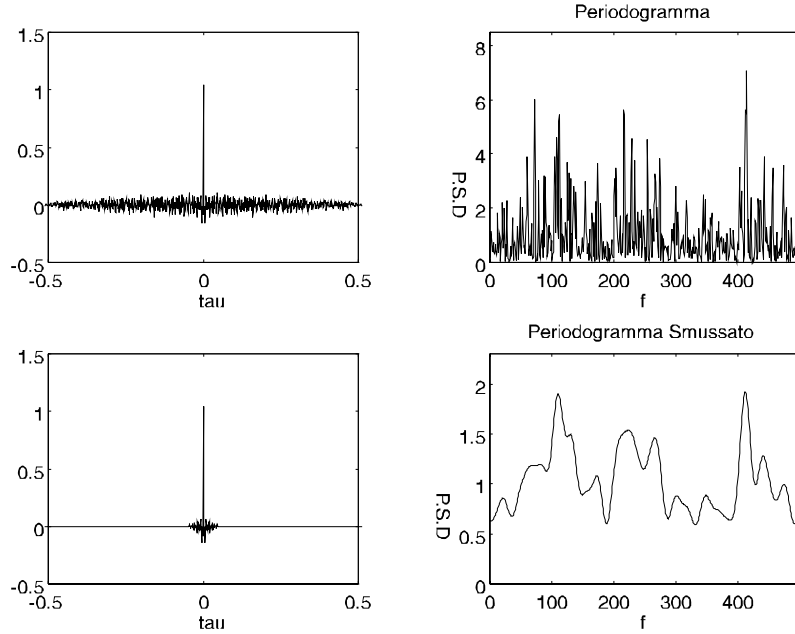


Figura 1.7: Troncamento della funzione di autocorrelazione e smussamento dello spettro

si può, nell'altro dominio, **moltiplicare** la funzione di autocorrelazione per una **finestra**, che è l'antitrasformata di Fourier del filtro che smussa il periodogramma (fig. 10.7). Tuttavia, non tutte le finestre sono valide: dovremo moltiplicare l'autocorrelazione per finestre con Trasformata di Fourier positiva per evitare di generare valori negativi dello spettro smussato, e quindi, in qualche situazione, ridurre pericolosamente  $\hat{r}_0$ . In genere, si userà una finestra triangolare nei tempi di durata corrispondente alla riduzione di risoluzione spettrale desiderata. Ad esempio, se si moltiplica l'autocorrelazione per una finestra triangolare che va a zero per  $\tau_{max} = PT \simeq NT/10$ , come viene spesso consigliato (metodo di Blackman e Tukey), il lobo principale della trasformata della finestra va a zero per  $f = 1/PT \simeq 10/NT$  e quindi combina insieme i contenuti  $\hat{S}_k$  di almeno 20 celle spettrali ad intervallo  $\Delta f = 1/NT$ . Facendo i conti più precisamente risulta però che, agli effetti della varianza della stima della locale densità spettrale, il numero equivalente di celle del periodogramma i cui contenuti sono combinati tra loro è  $\simeq 15$  invece che 20. Infatti,

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

indicando con  $\widehat{S}_B$  la stima spettrale secondo Blackman e Tukey, si ha, posto che:

$$E[\widehat{S}_k] = S; \quad E[\widehat{S}_k^2] - E^2[\widehat{S}_k] = \frac{2}{M} S^2 = S^2$$

$$\widehat{S}_B = \sum p_k \widehat{S}_k; \quad \sum p_k = 1; \quad E[\widehat{S}_B] = S$$

dove i pesi  $p_k$  sono dati dalla formula (trasformata di un triangolo campionato di lunghezza di base pari a  $2P$  periodicizzato su lunghezza  $N$ ;  $P$  dispari, altrimenti non torna la periodicità su  $N$  campioni):

$$p_k = \frac{1}{PN} \left( \frac{\sin(\pi k P/N)}{\sin(\pi k/N)} \right)^2 \quad (1.24)$$

$$p_k = \sum_{l=0}^{N-1} q_l W_N^{-kl}; \quad q_l = \frac{1}{N} \sum_{k=0}^{N-1} p_k W_N^{kl};$$

Si osservi che i pesi  $p_k$  hanno un lobo principale che comprende  $2N/P$  celle e quindi ci aspetteremmo una dispersione ridotta di un fattore  $\sqrt{P/2N}$ . Calcolando la dispersione come:

$$\sigma_{\widehat{S}_B}^2 = E \left[ \left( \sum p_k \widehat{S}_k \right)^2 \right] - \left( E \left[ \sum p_k \widehat{S}_k \right] \right)^2 = S^2 \sum p_k^2$$

risulta invece:

$$\frac{\sigma_{\widehat{S}_B}}{E[\widehat{S}_B]} = \sqrt{\sum p_k^2} \simeq \sqrt{\frac{2P}{3N}} = \sqrt{\frac{4}{3} \frac{P}{2N}} \quad (1.25)$$

Infatti i pesi nel dominio delle frequenze  $p_k$  sono trasformati di pesi nel dominio del tempo  $q_l$  decrescenti linearmente (a triangolo); si deve imporre che il valore al tempo 0 sia  $q_0 = 1/N$  perchè la somma dei pesi nel dominio trasformato  $p_k$  sia unitaria.

$$q_l = \frac{1}{N} \left( 1 - \frac{|l|}{P} \right); \quad \sum_{l=-P}^P q_l \simeq \frac{P}{N}; \quad \sum_{l=-P}^P q_l^2 \simeq \frac{2P}{3N^2};$$

e, usando la relazione di Parseval:

$$\sum p_k^2 = N \sum_{l=-P}^P q_l^2 \simeq \frac{2P}{3N} \quad (1.26)$$

## 1.7 MODIFICHE DEL PERIODOGRAMMA

In conclusione troviamo che la varianza cresce di un fattore  $4/3$  appunto perchè lo spettro della finestra (e quindi i pesi  $p_k$ ) è molto basso per  $f \simeq B$ ; nel caso considerato di una finestra triangolare che va a zero per  $\tau = NT/10$  si ha  $P = N/10$  e il numero equivalente di stime spettrali indipendenti  $\hat{S}_k$  combinate (15) è minore di quello totale (20) originalmente contenuto nelle celle spettrali combinate dalla media fatta con i pesi. Tenendo conto delle parti reali ed immaginarie della trasformata, i campioni di potenza indipendenti contenuti nelle 20 celle spettrali adiacenti da cui si stima lo spettro di potenza nell'intervallo cercato sono appunto  $M \simeq 30$ ; le stime indipendenti  $\hat{S}_k$  che sono state mediate si susseguono nel dominio delle frequenze. Smussamenti maggiori si otterrebbero con finestre sull'autocorrelazione che vanno a zero per tempi inferiori ecc. Si osservino gli spettri in figura 10.7.

Un altro risultato si sarebbe ottenuto con pesi tutti eguali, e cioè convolvendo il periodogramma con una finestra rettangolare. In questo caso, con maggiore efficienza, la riduzione di risoluzione spettrale di un fattore 20 avrebbe comportato la riduzione della dispersione di un fattore  $\sqrt{20}$ , invece che  $\sqrt{15}$ . Peraltro, sarebbe stato equivalente moltiplicare l'autocorrelazione per una finestra a seno cardinale, che non si annulla nemmeno a tempi grandi. In pratica, si usa spesso la tecnica di Blackman e Tukey, anche se è leggermente inefficiente.

### 1.7.2 Media di periodogrammi di sottosequenze

Un altro modo per diminuire la dispersione è quello di suddividere l'intervallo  $NT$  di osservazione dei dati in  $L$  sottointervalli disgiunti e quindi dividere la sequenza totale in  $L$  sottosequenze. Poi vengono mediate tra loro i periodogrammi delle sottosequenze ed il numero totale di campioni indipendenti combinati tra loro per ogni cella sarà  $M = 2L$ . Se p.e.  $L = 15$ ,  $M = 30$  e dal punto di vista della dispersione statistica i risultati sono equivalenti a quelli ottenibili con una finestratura triangolare dell'autocorrelazione che vada a zero per  $\tau_{max} \simeq NT/10$ . Quindi, con un periodogramma unico si ottiene la desiderata dispersione statistica con una risoluzione spettrale pari a circa  $B = 20/NT$ ; invece, segmentando in sottosequenze, per avere la stessa dispersione statistica, basta porre  $L = 15$ , la risoluzione spettrale del periodogramma di una sottosequenza diventa  $B' \simeq 15/NT$  e quindi un poco meglio rispetto al metodo di Blackman: abbiamo trattato i dati in modo diverso che sembra più



## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

efficiente e con risultati equivalenti a quelli che si avrebbero convolvendo lo spettro di un unico periodogramma con un rettangolo di larghezza  $B'$ .

In realtà succede di peggio; mediando gli  $L$  periodogrammi brevi ottenuti troncando i dati con finestre corte, la polarizzazione della stima peggiora a causa dei troncamenti e delle periodicizzazioni più fitte. Un disturbo monocromatico si ripercuote su tutto lo spettro con maggiore influenza. In presenza di disturbi fortemente colorati, è meglio smussare nel dominio delle frequenze il periodogramma della sequenza intera convolvendolo p.e. con un rettangolo di larghezza  $B$ , come visto alla fine del paragrafo precedente. In questo caso, gli effetti di un forte segnale monocromatico non si estendono al di là della banda  $B$  ed è come se avessimo rastremato a zero la funzione di autocorrelazione moltiplicandola per una finestra a *sinc*, come peraltro già osservato.

Per contro, mediando periodogrammi fatti su sottosequenze l'occupazione di memoria è minore ed anche i calcoli sono meno impegnativi. Nella figura 10 .4, si confrontino gli effetti dello smussamento del periodogramma (fig. 10 .4b) con quelli della media di periodogrammi fatti su sottointervalli (figg. 10 .4c - d - e - f - g); la qualità della stima in figura 10 .4b è ottenuta approssimativamente solo con la tecnica WOSA (fig. 10 .4g) di cui al prossimo paragrafo.

### 1.7.3 La tecnica WOSA

E' necessario evitare polarizzazioni dovute a eccessive infiltrazioni spettrali (spectral leaks): queste sono dovute, ad esempio, al troncamento nei tempi di una sinusoide effettuato con una finestra rettangolare. Ciò corrisponde nelle frequenze alla convoluzione con un *sinc* e origina componenti spettrali a frequenze anche lontane dalla frequenza della sinusoide di partenza. Quindi, per formare le sottosequenze, è necessario **finestrare i dati** con finestre abbastanza "dolci" come ad esempio coseno rialzato per evitare questo fenomeno, specialmente se il segnale è disturbato da forti segnali monocromatici. Così facendo, tuttavia, la risoluzione spettrale si riduce e soprattutto si pesano poco alcuni dati; quindi, la dispersione della stima spettrale è maggiore del necessario perché non si utilizzano in egual modo tutti i dati. Allora, non volendo o non potendo fare la cosa ottimale e cioè trattare tutti i dati con una finestra unica per poi smussare l'unico periodogramma ottenuto, la tecnica consigliata è quella di sovrapporre le finestre (Window Overlap) con le quali i dati

## 1.7 MODIFICHE DEL PERIODOGRAMMA

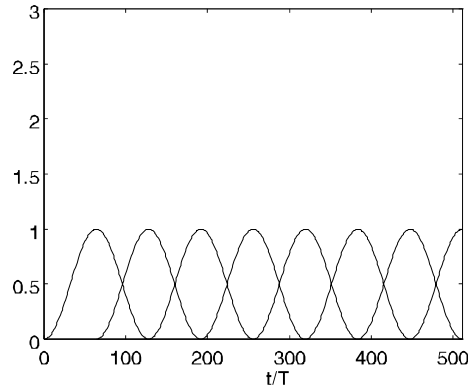


Figura 1.8: Disposizione delle finestre nella tecnica WOSA.

vengono sezionati (vedi figura 10.8) in modo da trascurare il minor numero di dati possibile. Ancora, è possibile tenere conto, **in media**, della finestra con cui sono stati sezionati i dati, al momento della stima dello spettro medio. Infatti, per smussare quest'ultimo spettro si moltiplica la funzione di autocorrelazione per una seconda finestra, in generale più stretta della finestra utilizzata per limitare la lunghezza dei dati. Questa seconda finestra può essere sagomata in modo opportuno da compensare la polarizzazione che in media la prima finestra ha generato sui dati. Questa tecnica di analisi spettrale è nota con il nome di Window Overlap Spectral Analysis (WOSA; Nuttall; Proc. IEEE 1982) e assomiglia molto alla tecnica parametrica MA che vedremo al prossimo capitolo.

## 1.8 Appendice 1

### 1.8.1 Processi casuali discreti

Consideriamo una sequenza di variabili complesse  $\{x_n\}$ ,  $n = 0 \dots N$  e supponiamo che la densità di probabilità congiunta delle variabili (parti reali ed immaginarie) sia gaussiana. Indichiamo con  $\underline{\mathbf{x}}$  un vettore colonna le cui  $N$  componenti sono gli elementi della sequenza  $\{x_n\}$  e con  $\underline{\mathbf{R}}$  la matrice di covarianza, di dimensioni  $[N \times N]$ , del vettore (e della sequenza) e cioè:

$$\underline{\mathbf{R}} = E[\underline{\mathbf{x}}\underline{\mathbf{x}}^*] \quad (1.27)$$

L'asterisco indica la trasposizione e, nel caso di vettori complessi, anche il coniugio. Dimostriamo ora che la densità di probabilità del vettore casuale  $\underline{\mathbf{x}}$  è:

$$p(\underline{\mathbf{x}}) = \frac{1}{(2\pi)^{N/2} |\underline{\mathbf{R}}|^{1/2}} \exp \left( -\frac{\underline{\mathbf{x}}^* \underline{\mathbf{R}}^{-1} \underline{\mathbf{x}}}{2} \right) \quad (1.28)$$

Nel caso semplice che la sequenza  $\{x_n\}$  è bianca, e quindi

$$E[x_m x_{m-i}^*] = \sigma_x^2 \delta_i$$

allora la matrice di covarianza  $\underline{\mathbf{R}}_x$  è diagonale:

$$\underline{\mathbf{R}}_x = \sigma_x^2 \underline{\mathbf{I}} \quad (1.29)$$

dove con  $\underline{\mathbf{I}}$  abbiamo indicato la matrice identica e con  $\sigma_x^2$  il valore q.m. degli elementi della sequenza (dei moduli, se la sequenza è complessa). In questo caso la 1.28 risulta immediatamente poichè la densità di probabilità di un vettore le cui componenti sono delle variabili indipendenti (in questo caso gaussiane a v.m. 0) è il prodotto delle loro densità.

$$\begin{aligned} p(\underline{\mathbf{x}}) &= \prod_{i=1}^N p(x_i) = \frac{1}{(2\pi)^{N/2} \sigma_x^N} \exp \left( -\frac{1}{2\sigma_x^2} \sum x_i^2 \right) = \\ &= \frac{1}{(2\pi)^{N/2} |\underline{\mathbf{R}}_x|^{1/2}} \exp \left( -\frac{\underline{\mathbf{x}}^* \underline{\mathbf{R}}_x^{-1} \underline{\mathbf{x}}}{2} \right) \end{aligned} \quad (1.30)$$

Passiamo ora al caso generale: facciamo una trasformazione di coordinate dal vettore  $\underline{\mathbf{x}}$  ad un vettore  $\underline{\mathbf{y}}$  e quindi moltiplichiamo il vettore  $\underline{\mathbf{x}}$  per una matrice  $\underline{\mathbf{A}}$  quadrata ed invertibile (dimentichiamo le singolarità):

$$\underline{\mathbf{y}} = \underline{\mathbf{A}}\underline{\mathbf{x}} \rightarrow \underline{\mathbf{x}} = \underline{\mathbf{A}}^{-1}\underline{\mathbf{y}}$$

La matrice di covarianza di  $\underline{\mathbf{y}}$  è:

$$E[\underline{\mathbf{y}}\underline{\mathbf{y}}^*] = \underline{\mathbf{R}}_y = \underline{\mathbf{A}}E[\underline{\mathbf{x}}\underline{\mathbf{x}}^*]\underline{\mathbf{A}}^* = \sigma_x^2 \underline{\mathbf{A}}\underline{\mathbf{A}}^*$$

ed è anche vero che:

$$\underline{\mathbf{x}}^* \underline{\mathbf{R}}_x^{-1} \underline{\mathbf{x}} = \frac{1}{\sigma_x^2} \underline{\mathbf{y}}^* (\underline{\mathbf{A}}^{-1})^* \underline{\mathbf{A}}^{-1} \underline{\mathbf{y}} = \underline{\mathbf{y}}^* \underline{\mathbf{R}}_y^{-1} \underline{\mathbf{y}}$$

e sostituendo nella 1.30 si ottiene la densità di probabilità 1.28. Per normalizzare ad 1 l'integrale della probabilità su tutto il volume infinito  $N$ -dimensionale dopo la trasformazione di coordinate, si deve dividere per il determinante della matrice Jacobiana di trasformazione delle coordinate da  $\underline{\mathbf{x}}$  a  $\underline{\mathbf{y}}$ :

$$|\underline{\mathbf{J}}| = \sigma_x^N |\underline{\mathbf{A}}| = \sqrt{\sigma_x^{2N} |\underline{\mathbf{A}}|^2} = |\underline{\mathbf{R}}_y|^{1/2}$$

e quindi risulta la 1.28:

$$p(\underline{\mathbf{y}}) = \frac{1}{(2\pi)^{N/2} |\underline{\mathbf{R}}_y|^{1/2}} \exp\left(-\frac{\underline{\mathbf{y}}^* \underline{\mathbf{R}}_y^{-1} \underline{\mathbf{y}}}{2}\right)$$

La dimostrazione del fatto che dopo una trasformazione di coordinate, gli elementi di volume sono scalati per il determinante della matrice di trasformazione, è semplice in casi semplici. Partiamo dalla constatazione che una matrice  $\underline{\mathbf{A}}$  può venire diagonalizzata con lo sviluppo in autovettori - autovalori. Si ha cioè:

$$\underline{\mathbf{A}} = \underline{\mathbf{U}}\underline{\mathbf{Q}}\underline{\mathbf{U}}^*$$

dove  $\underline{\mathbf{U}}$  è la matrice degli autovettori di  $\underline{\mathbf{A}}$ , ortogonale, e quindi con determinante unitario. Quindi la trasformazione indotta da  $\underline{\mathbf{A}}$  è decomponibile nella cascata di una rotazione  $N$ -dimensionale (la matrice  $\underline{\mathbf{U}}$ ), uno scalamento di coordinate diverso da coordinata a coordinata (la matrice diagonale  $\underline{\mathbf{Q}}$  che ha come elementi della diagonale principale gli autovalori di  $\underline{\mathbf{A}}$ ,  $q_i$ ) ed una rotazione di coordinate opposta alla precedente (la matrice  $\underline{\mathbf{U}}^* = \underline{\mathbf{U}}^{-1}$ ). Lo scalamento dei volumi è ovviamente pari al prodotto dei moduli dei fattori di scala lungo le coordinate, e quindi al prodotto dei moduli degli autovalori, che è appunto il modulo del determinante. Si osservi che nella formula compare appunto:

$$|\underline{\mathbf{R}}_y|^{1/2} = \sigma_x^N \sqrt{|\underline{\mathbf{U}}| |\underline{\mathbf{Q}}|^2 |\underline{\mathbf{U}}^*|} = \sigma_x^N \prod_{n=1}^N |q_n|$$

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

Nel caso di autovalori nulli, la trasformazione è singolare, la nuova densità di probabilità non *occupa* tutto lo spazio, e infatti lungo alcune dimensioni corrisponderà ad un impulso ideale posto sullo zero (uscita sistematicamente nulla).

### 1.8.2 Stima a massima verosimiglianza (da Wikipedia)

Il metodo della massima verosimiglianza in statistica è un procedimento matematico per determinare uno stimatore. Il metodo consiste nel massimizzare la funzione di verosimiglianza, definita in base alla probabilità di osservare una data realizzazione campionaria, condizionatamente ai valori assunti dai parametri oggetto di stima. Il metodo è stato originariamente sviluppato dal genetista e statistico sir Ronald Fisher, tra il 1912 e il 1922.

#### Filosofia del metodo della massima verosimiglianza

Dato un parametro  $\theta$ , caratterizzato da una distribuzione di probabilità  $\mathcal{D}$ , con densità di probabilità  $\mathcal{L}_D$ , e dato un campione di dati osservati  $\{x_i\}_{i=1}^N$  di dimensione  $N$  funzione del parametro  $\theta$ , si può calcolare la probabilità associata ai dati osservati.

$$P\left(\{x_i\}_{i=1}^N \mid \theta\right) = \mathcal{L}_D\left(\theta \mid \{x_i\}_{i=1}^N\right) \quad (1.31)$$

Il parametro  $\theta$  è ignoto, sebbene sia noto che  $\theta$  è estratto dalla distribuzione  $\mathcal{D}$ . Utilizziamo i dati a nostra disposizione,  $\{x_i\}_{i=1}^N$ , per valutare  $\theta$ . Il metodo della massima verosimiglianza ricerca il valore più *verosimile* di  $\theta$ , ossia ricerca, tra tutti i possibili valori di  $\theta$ , il valore del parametro che *massimizza* la probabilità di aver ottenuto il campione dato. Da un punto di vista matematico,  $\mathcal{L}_D\left(\theta \mid \{x_i\}_{i=1}^N\right)$  è detta funzione di verosimiglianza, e lo stimatore di massima verosimiglianza è ottenuto come:

$$\hat{\theta} = \arg \max_{\theta} \mathcal{L}_D(\theta \mid x_1, x_2, \dots, x_N) \quad (1.32)$$

**Esempio**

Si consideri un campione  $\{x_i\}_{i=1}^N$  di variabili casuali identicamente e indipendentemente distribuite, con distribuzione gaussiana a media  $\mu$  e varianza  $\sigma^2$ , ignote. La funzione di verosimiglianza associata è:

$$\mathcal{L}_D(\mu, \sigma^2 | x_1, x_2, \dots, x_N) = \prod_{i=1}^N p(x_i) = \frac{1}{(2\pi)^{N/2} \sigma^N} \exp \left[ -\frac{\sum_{i=1}^N (x_i - \mu)^2}{2\sigma^2} \right] \quad (1.33)$$

La massimizzazione della funzione di verosimiglianza è equivalente a massimizzarne il logaritmo che a meno di costanti risulta essere:

$$\log(\mathcal{L}_D(\theta | x_1, x_2, \dots, x_N)) = -N \log \sigma - \frac{\sum_{i=1}^N (x_i - \mu)^2}{2\sigma^2}$$

I parametri  $\mu, \sigma$  sono determinati risolvendo il problema di massimo:

$$\begin{aligned} \frac{\sum_{i=1}^N (x_i - \hat{\mu})}{\hat{\sigma}^2} &= 0 \rightarrow \hat{\mu} = \sum \frac{x_i}{N} \\ -N\hat{\sigma}^2 - \sum_{i=1}^N (x_i - \hat{\mu})^2 &= 0 \rightarrow \hat{\sigma}^2 = \frac{\sum_{i=1}^N (x_i - \hat{\mu})^2}{N} \end{aligned}$$

L'esempio illustra alcune proprietà degli stimatori di massima verosimiglianza. È immediato verificare che  $\hat{\mu}$  è uno stimatore non polarizzato e cioè il suo valore atteso coincide con il valore vero.

$$E[\hat{\mu}] = \mu$$

Invece,  $\hat{\sigma}^2$  non gode di tale proprietà.

Si verifica che:

$$E[\hat{\sigma}^2] = \frac{N-1}{N} \sigma^2$$

Val la pena d'altra parte di osservare che lo stimatore di massima verosimiglianza è comunque uno stimatore asintoticamente non polarizzato; infatti per  $N \rightarrow \infty$ :

$$\lim_{N \rightarrow \infty} E[\hat{\sigma}^2] = \sigma^2$$

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

In particolare, si dimostra che qualunque stimatore di massima verosimiglianza è asintoticamente ( $N \rightarrow \infty$ ) non polarizzato e asintoticamente normalmente distribuito.

### Casi patologici

Può accadere che il valore dello stimatore di massima verosimiglianza non appartenga allo spazio dei parametri. Lo stimatore di massima verosimiglianza, inoltre, può non essere unico. Ancora, l'operazione di massimo può essere difficile da effettuare: si pensi al caso di variabili a distribuzione uniforme, o a masse impulsive. Per esempio, si voglia stimare una percentuale di popolazione, p.e. i fumatori. Indicando con  $p$  la probabilità di appartenere a tale categoria, considerando una variabile casuale  $x_i$  che valga 1 se la persona  $i$ -esima fuma e zero altrimenti, la densità di probabilità della variabile  $x_i$  è

$$\begin{aligned} p(x) &= (1-p)\delta(0) + p\delta(1); \quad E[x] = p; \quad E[x^2] = p; \\ E[x_i x_j] &= p \text{ se } i = j; \quad E[x_i x_j] = p^2 \text{ se } i \neq j; \end{aligned} \quad (1.34)$$

Una stima non polarizzata di  $p$  è facilmente ottenuta come:

$$\hat{p} = \sum \frac{x_i}{N}; \quad E[\hat{p}] = \sum \frac{E[x_i]}{N} = p$$

La varianza di questo stimatore è:

$$\begin{aligned} \text{var}(\hat{p}) &= E[\hat{p}^2] - E[\hat{p}]^2 = \sum_{i,j} \frac{E[x_i x_j]}{N^2} - p^2 = \\ &= \frac{Np + (N^2 - N)p^2}{N^2} - p^2 = \frac{p(1-p)}{N} < \frac{1}{4N} \end{aligned} \quad (1.35)$$

Per avere con una significativa probabilità che l'errore commesso nella valutazione di  $p$  sia inferiore p.e. a 1%, si deve ragionevolmente dimensionare  $N$  in modo che

$$\frac{1}{\sqrt{4N}} \ll 0.01 \rightarrow N \gg 2500$$

Poi, si può rendere più precisa questa valutazione almeno dal punto di vista statistico, imponendo la probabilità che l'errore sia inferiore ad un certo valore, come fatto nella sezione 10.2.1. Arrivare a questo semplice risultato dalla stima a massima verosimiglianza sarebbe piuttosto complesso.

### 1.8.3 La stima lineare MMSE

Un caso molto frequente cui si applicano bene i metodi probabilistici è quello di un modello lineare in presenza di una distribuzione di probabilità a priori dei parametri da stimare. Si abbia quindi che il dato osservato sia il vettore:

$$\underline{\mathbf{y}} = \underline{\mathbf{A}}\underline{\mathbf{x}} + \underline{\mathbf{n}} \quad (1.36)$$

dove  $\underline{\mathbf{x}}$ ,  $\underline{\mathbf{n}}$  sono vettori incogniti gaussiani indipendenti (a valor medio nullo, per semplicità) le cui densità di probabilità siano:

$$p(\underline{\mathbf{x}}) = \frac{1}{(2\pi)^{N/2} |\underline{\mathbf{R}}_x|^{1/2}} \exp\left(-\frac{\underline{\mathbf{x}}^* \underline{\mathbf{R}}_x^{-1} \underline{\mathbf{x}}}{2}\right)$$

$$p(\underline{\mathbf{n}}) = \frac{1}{(2\pi)^{N/2} |\underline{\mathbf{R}}_n|^{1/2}} \exp\left(-\frac{\underline{\mathbf{n}}^* \underline{\mathbf{R}}_n^{-1} \underline{\mathbf{n}}}{2}\right)$$

La densità di probabilità di  $\underline{\mathbf{x}}$ , dato  $\underline{\mathbf{y}}$ , usando il teorema di Bayes, è:

$$p(\underline{\mathbf{x}}|\underline{\mathbf{y}}) = \frac{p(\underline{\mathbf{y}}|\underline{\mathbf{x}})p(\underline{\mathbf{x}})}{p(\underline{\mathbf{y}})} \quad (1.37)$$

$$p(\underline{\mathbf{x}}|\underline{\mathbf{y}}) = \frac{1}{p(\underline{\mathbf{y}})} \frac{1}{(2\pi)^N (|\underline{\mathbf{R}}_x| |\underline{\mathbf{R}}_n|)^{1/2}} \exp\left[-\frac{(\underline{\mathbf{y}} - \underline{\mathbf{A}}\underline{\mathbf{x}})^* \underline{\mathbf{R}}_n^{-1} (\underline{\mathbf{y}} - \underline{\mathbf{A}}\underline{\mathbf{x}})}{2} - \frac{\underline{\mathbf{x}}^* \underline{\mathbf{R}}_x^{-1} \underline{\mathbf{x}}}{2}\right] \quad (1.38)$$

Cerchiamo la stima che minimizza l'errore di stima quadratico medio (MMSE : Minimum Mean Square Error); si verifica che lo stimatore ottimo da questo punto di vista è la media condizionata,  $E(\underline{\mathbf{x}}|\underline{\mathbf{y}})$ . Nel caso di variabili congiuntamente gaussiane questo valore è quello che massimizza la funzione di probabilità a posteriori  $p(\underline{\mathbf{x}}|\underline{\mathbf{y}})$  e quindi troviamo, osservando che  $p(\underline{\mathbf{y}})$  non dipende da  $\underline{\mathbf{x}}$ :

$$\hat{\underline{\mathbf{x}}} = \arg \min_{\underline{\mathbf{x}}} [(\underline{\mathbf{y}} - \underline{\mathbf{A}}\underline{\mathbf{x}})^* \underline{\mathbf{R}}_n^{-1} (\underline{\mathbf{y}} - \underline{\mathbf{A}}\underline{\mathbf{x}}) + \underline{\mathbf{x}}^* \underline{\mathbf{R}}_x^{-1} \underline{\mathbf{x}}] \quad (1.39)$$

Il minimo di questa forma quadratica nel vettore incognito  $\underline{\mathbf{x}}$  si trova annullando le derivate vettoriali ed è una funzione lineare di  $\underline{\mathbf{y}}$ :

$$\hat{\underline{\mathbf{x}}} = \underline{\mathbf{R}}_x \underline{\mathbf{A}}^* (\underline{\mathbf{A}} \underline{\mathbf{R}}_x \underline{\mathbf{A}}^* + \underline{\mathbf{R}}_n)^{-1} \underline{\mathbf{y}} = (\underline{\mathbf{A}}^* \underline{\mathbf{R}}_n^{-1} \underline{\mathbf{A}} + \underline{\mathbf{R}}_x^{-1})^{-1} \underline{\mathbf{A}}^* \underline{\mathbf{R}}_n^{-1} \underline{\mathbf{y}} \quad (1.40)$$



## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

Nel caso particolare in cui:

$$\underline{\mathbf{R}}_x = \sigma_x^2 \underline{\mathbf{I}}, \quad \underline{\mathbf{R}}_n = \sigma_n^2 \underline{\mathbf{I}}$$

si ha:

$$\hat{\underline{\mathbf{x}}} = \underline{\mathbf{A}}^* \left( \underline{\mathbf{A}} \underline{\mathbf{A}}^* + \frac{\sigma_n^2}{\sigma_x^2} \underline{\mathbf{I}} \right)^{-1} \underline{\mathbf{y}} \quad (1.41)$$

congruentemente con quanto si vedrà al capitolo sulle direzioni di arrivo e con i metodi di minimo errore di predizione o di minimo errore quadratico medio. Si osservi che il numero delle componenti del vettore  $\underline{\mathbf{x}}$  e del vettore osservato  $\underline{\mathbf{y}}$  possono essere diversi e quindi può essere più conveniente l'uso dell'una o dell'altra delle due formulazioni di 1.40. Quindi questa soluzione si presta alla stima di una costante o di un comportamento polinomiale sommati a rumore, o infiniti altri casi.

### 1.8.4 Sequenze correlate e loro spettri

Supponiamo ora di convolvere una sequenza di variabili indipendenti (bianca)  $\{x_n\}$  con un filtro  $\{h_n\}$  in modo da *colorarla*; la matrice di covarianza della sequenza  $\{y_n\} = \{x_n\} * \{h_n\}$  sarà  $\underline{\mathbf{R}}_y$ .

$$\underline{\mathbf{R}}_y = \begin{bmatrix} r_0 & r_{-1} & r_{-2} & \dots & 0 \\ r_1 & r_0 & r_{-1} & \dots & 0 \\ r_2 & r_1 & r_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & r_0 & r_{-1} \\ 0 & 0 & \dots & r_1 & r_0 \end{bmatrix}$$

dove con  $r_0, r_1, r_{-1}$  ecc. abbiamo indicato:

$$r_k = E[y_m y_{m-k}^*] = E \left[ \left( \sum_i x_{m-i} h_i \right) \left( \sum_l x_{m-k-l} h_l \right)^* \right] = \sigma_x^2 \sum_i h_i h_{i-k}^*;$$

$$r_k = r_{-k}^*$$

La matrice  $\underline{\mathbf{R}}_y$  è sia hermitiana che di Toeplitz e cioè la diagonale principale e tutte le sue parallele contengono elementi eguali. Calcoliamo ora la

matrice di covarianza della trasformata di Fourier di una sequenza reale e bianca  $\{x_n\}$  e di lunghezza limitata  $N$ , e cioè la matrice di covarianza della sequenza  $\{X_k\}$  dove:

$$X_k = \sum_{n=0}^{N-1} x_n W^{-nk} \quad (1.42)$$

Indicando ora con  $\underline{\mathbf{X}}$  il vettore le cui componenti sono i campioni della sequenza  $\{X_k\}$  si ha:

$$\underline{\mathbf{X}} = \underline{\mathbf{W}} \underline{\mathbf{x}} \quad (1.43)$$

dove  $\underline{\mathbf{W}}$  è una matrice che corrisponde alla trasformazione di Fourier diretta (attenzione! non è una matrice hermitiana!). La matrice di covarianza relativa al vettore  $\underline{\mathbf{X}}$  è:

$$\underline{\mathbf{R}}_{\underline{\mathbf{X}}} = E[\underline{\mathbf{X}} \underline{\mathbf{X}}^*] = E[\underline{\mathbf{W}} \underline{\mathbf{x}} \underline{\mathbf{x}}^* \underline{\mathbf{W}}^*] = \underline{\mathbf{W}} E[\underline{\mathbf{x}} \underline{\mathbf{x}}^*] \underline{\mathbf{W}}^* = \underline{\mathbf{W}} \underline{\mathbf{R}} \underline{\mathbf{W}}^* \quad (1.44)$$

Indicando con  $A_k$  e  $B_k$  la parte reale ed immaginaria di  $X_k$ , si può verificare che, poichè la sequenza  $\{x_n\}$  è bianca e reale:

$$X_k = A_k + jB_k;$$

$$E[A_k B_h] = 0; \forall k, h; \quad E[A_k A_h] = E[B_k B_h] = 0; \quad |k| \neq |h| \quad (1.45)$$

$$E[A_k^2] = \frac{N}{2} \sigma_x^2 [1 + \delta(k) + \delta(k - N/2)] \quad (1.46)$$

$$E[B_k^2] = E[A_k^2]; \quad k = 1, \dots, N/2 - 1$$

$$E[B_k^2] = 0; \quad k = 0, N/2$$

Infatti, le relazioni precedenti si possono dedurre dal fatto che:

$$\begin{aligned} E[X_k X_h^*] &= \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} E[x_n x_m^*] W^{-nk} W^{mh} = \\ &= \sigma_x^2 \sum_{n=0}^{N-1} \sum_{m=0}^{N-1} \delta(n - m) W^{-nk} W^{mh} = \sigma_x^2 \sum_{m=0}^{N-1} W^{m(h-k)} = \\ &= N \sigma_x^2 \times \delta(h - k); \end{aligned}$$

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

Relazioni simili si possono scrivere per le parti reali ed immaginarie degli  $X_k$ . Se invece la sequenza  $\{x_n\}$  fosse bianca e complessa, si avrebbe:

$$E[A_k B_h] = 0; \quad E[A_k A_h] = E[B_k B_h] = 0; \quad k \neq h \quad (1.47)$$

$$E[A_k^2] = E[B_k^2] = N\sigma_x^2/2$$

Si possono ottenere gli stessi risultati in modo diretto, osservando che:

$$\begin{aligned} \underline{\mathbf{X}} &= \underline{\mathbf{W}} \underline{\mathbf{x}}; \quad \underline{\mathbf{x}} = \frac{1}{N} \underline{\mathbf{W}}^* \underline{\mathbf{X}} = \frac{1}{N} \underline{\mathbf{W}}^* \underline{\mathbf{W}} \underline{\mathbf{x}} \rightarrow \\ \frac{1}{N} \underline{\mathbf{W}}^* \underline{\mathbf{W}} &= \frac{1}{N} \underline{\mathbf{W}} \underline{\mathbf{W}}^* = \underline{\mathbf{I}} \rightarrow (\underline{\mathbf{W}}^*)^{-1} = \frac{1}{N} \underline{\mathbf{W}} \\ \underline{\mathbf{R}}_{\mathbf{X}} &= \underline{\mathbf{W}} \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{W}}^* = N\sigma_x^2 \underline{\mathbf{I}} \end{aligned} \quad (1.48)$$

Quindi la trasformata di Fourier di una sequenza complessa incorrelata (bianca) è bianca (i campioni sono incorrelati); se la sequenza di partenza è reale, vanno sempre tenute in conto le simmetrie.

Ora, convolviamo **periodicamente** la sequenza  $\{x_n\}$  di lunghezza limitata  $N$ , con un qualsiasi filtro  $\{h_n\}$  di eguale lunghezza  $N$  (o diventato tale dopo un'aggiunta di zeri). La matrice di covarianza della sequenza  $\{y_n\} = \{x_n\} * \{h_n\}$  diventa circolante e cioè tutte le sue righe sono eguali a meno di uno slittamento progressivo verso destra.

$$\underline{\mathbf{R}}_{\mathbf{y}} = \begin{bmatrix} r_0 & r_{-1} & 0 & \dots & r_1 \\ r_1 & r_0 & r_{-1} & \dots & 0 \\ 0 & r_1 & r_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & r_{-1} \\ r_{-1} & 0 & \dots & r_1 & r_0 \end{bmatrix}$$

La trasformata di Fourier discreta della sequenza  $\{x_n\}$ ,  $\{X_k\}$  verrà moltiplicata per  $\{H_k\}$ , Trasformata di Fourier discreta della sequenza  $\{h_n\}$ . Si ha cioè, se:

$$\{x_{n+N}\} = \{x_n\}; \quad \{h_{n+N}\} = \{h_n\}; \quad (1.49)$$

$$\{y_n\} = \{h_n\} * \{x_n\} \rightarrow \{y_{n+N}\} = \{y_n\} \quad (1.50)$$

$$X_k = \sum_{n=0}^{N-1} x_n W^{-nk}, \quad H_k = \sum_{n=0}^{N-1} h_n W^{-nk}; \quad (1.51)$$

$$Y_k = X_k H_k; \quad y_n = \frac{1}{N} \sum_{k=0}^{N-1} Y_k W^{nk} \quad (1.52)$$

Indicando con vettori sia la sequenza  $\{X_k\}$  che la  $\{Y_k\}$  e con una moltiplicazione matriciale l'effetto dell'applicazione della funzione di trasferimento  $\{H_k\}$  risulta:

$$\underline{\mathbf{X}} = \begin{bmatrix} X_0 \\ X_1 \\ \dots \\ X_{N-1} \end{bmatrix}; \quad \underline{\mathbf{Y}} = \begin{bmatrix} Y_0 \\ Y_1 \\ \dots \\ Y_{N-1} \end{bmatrix}; \quad \underline{\mathbf{H}} = \begin{bmatrix} H_0 & 0 & 0 & 0 & 0 \\ 0 & H_1 & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & 0 & H_{N-2} & 0 \\ 0 & 0 & 0 & \dots & H_{N-1} \end{bmatrix}$$

$$\underline{\mathbf{Y}} = \underline{\mathbf{H}} \underline{\mathbf{X}}$$

La matrice di covarianza della sequenza  $\underline{\mathbf{Y}}$ ,  $\underline{\mathbf{R}}_{\mathbf{Y}}$  sarà anch'essa diagonale, ma gli elementi della diagonale principale avranno ora valore  $N\sigma_x^2 |H_k|^2$ .

$$\underline{\mathbf{R}}_{\mathbf{Y}} = E [\underline{\mathbf{H}} \underline{\mathbf{X}} \underline{\mathbf{X}}^* \underline{\mathbf{H}}^*] = \underline{\mathbf{H}} \underline{\mathbf{R}}_{\mathbf{X}} \underline{\mathbf{H}}^* =$$

$$= N\sigma_x^2 \begin{bmatrix} |H_0|^2 & 0 & \dots & 0 & 0 \\ 0 & |H_1|^2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & |H_{N-2}|^2 & 0 \\ 0 & 0 & \dots & 0 & |H_{N-1}|^2 \end{bmatrix} \quad (1.53)$$

Questo implica che le componenti della Trasformata di Fourier discreta di una qualsiasi sequenza colorata (filtrata)  $\{y_n\}$  ottenuta convolvendo **periodicamente** una sequenza bianca con una sequenza qualsiasi  $\{h_n\}$  di pari lunghezza, sono incorrelate, anche se di diversa varianza.

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

In altre parole, i campioni di una sequenza stazionaria colorata sono correlati ed i campioni della loro Trasformata di Fourier discreta sono incorrelati, ma la loro varianza cambia con la frequenza.

Vediamo di nuovo ora la struttura della matrice di covarianza della sequenza  $\{y_n\}$  convoluzione periodica delle due sequenze  $\{x_n\}$  e  $\{h_n\}$ , nel caso semplice in cui solo  $r_0, r_1$  siano diversi da 0;

$$\underline{\mathbf{R}}_{\mathbf{y}} = E[\underline{\mathbf{y}} \underline{\mathbf{y}}^*] \quad (1.54)$$

$$\underline{\mathbf{R}}_{\mathbf{y}} = \begin{bmatrix} r_0 & r_{-1} & 0 & \dots & r_1 \\ r_1 & r_0 & r_{-1} & \dots & 0 \\ 0 & r_1 & r_0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & r_{-1} \\ r_{-1} & 0 & 0 & \dots & r_0 \end{bmatrix} \quad (1.55)$$

La matrice di covarianza di una sequenza stazionaria  $\underline{\mathbf{R}}_{\mathbf{y}}$  ha struttura **hermitiana** e cioè è simmetrica se è reale o a simmetria coniugata se complessa. Inoltre, ha struttura di **Toeplitz** e cioè gli elementi lungo la diagonale e le subdiagonali sono costanti. Inoltre, a causa del fatto che la convoluzione è stata fatta in modo periodico, la matrice  $\underline{\mathbf{R}}_{\mathbf{y}}$  è **circolante** e cioè ogni riga è ottenuta dalla precedente con lo slittamento di un posto e mettendo l'ultimo campione al posto del primo. La  $m$ -esima riga della matrice  $\underline{\mathbf{R}}_{\mathbf{y}}$  è quindi l'autocorrelazione circolare del segnale  $\{y_n\}$ ,  $r_{mn} = E[y_n y_{h+m-n}^*]$ ; ricordiamo inoltre la periodicità  $N$  delle sequenze  $\{y_n\}$ ,  $\{h_n\}$ .

Se la sequenza  $\{h_n\}$  ha appunto solo due campioni diversi da zero, la  $\underline{\mathbf{R}}_{\mathbf{y}}$  ha struttura tridiagonale, come mostrato nell'equazione (1.55). Si osserva pertanto che, se la sequenza  $\{h_n\}$  è corta rispetto ad  $N$ , la periodicizzazione cambia solo pochi elementi della matrice di covarianza.

Questo fatto ci porta a concludere che in pratica, convolvendo sequenze lunghe di campioni indipendenti con filtri corti, la matrice di covarianza risultante dalla convoluzione periodica è poco differente da quella risultante dalla convoluzione aperiodica: quindi gli elementi della trasformata di Fourier di una sequenza ottenuta finestrando un segnale

complesso stazionario, bianco o colorato che sia, tendono ad essere indipendenti. Se il segnale è reale, ci si deve ricordare delle relazioni di parità e disparità delle parti reali ed immaginarie.

### 1.8.5 Matrici di Toeplitz: autovalori

Si consideri una matrice di covarianza  $\underline{\mathbf{R}}_{\mathbf{x}}$  di Toeplitz (e cioè con le diagonali parallele alla diagonale principale composte da elementi eguali). La matrice  $\underline{\mathbf{R}}_{\mathbf{x}}$  essendo di covarianza e cioè

$$\underline{\mathbf{R}}_{\mathbf{x}} = E[\underline{\mathbf{x}} \underline{\mathbf{x}}^*]$$

è simmetrica se la sequenza di partenza è reale; altrimenti è a simmetria Hermitiana (la trasposta è eguale alla coniugata). E' intuitivamente attraente collegare gli autovalori e gli autovettori di matrici di questo tipo allo spettro di potenza del processo cui appartiene il vettore casuale  $\underline{\mathbf{x}}$  che ha appunto  $\underline{\mathbf{R}}_{\mathbf{x}}$  come matrice di covarianza. Ricordiamo che una matrice  $\underline{\mathbf{R}}_{\mathbf{x}}$  ha autovalori  $\lambda_i$  (che vedremo corrispondere alle energie spettrali) e autovettori  $\underline{\mathbf{e}}_i$  (corrispondenti alle sinusoidi di varia frequenza) tali che:

$$\underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{e}}_i = \lambda_i \underline{\mathbf{e}}_i \quad (1.56)$$

In forma matriciale si ha:

$$\underline{\mathbf{E}} \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{E}}^* = \underline{\mathbf{\Lambda}}; \quad \underline{\mathbf{E}}^{-1} = \underline{\mathbf{E}}^* \quad (1.57)$$

dove la matrice **ortogonale**  $\underline{\mathbf{E}}$  corrisponde agli  $N$  autovettori colonna, la matrice  $\underline{\mathbf{E}}^*$  agli autovettori riga (è sia l'inversa che la trasposta della precedente, con coniugio se hermitiana) ed infine la matrice  $\underline{\mathbf{\Lambda}}$  è diagonale, e contiene lo "spettro" degli autovalori della matrice  $\underline{\mathbf{R}}_{\mathbf{x}}$ .

$$\underline{\mathbf{\Lambda}} = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 & 0 \\ 0 & \lambda_2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \lambda_{N-1} & 0 \\ 0 & 0 & \dots & 0 & \lambda_N \end{bmatrix}$$

$$\underline{\mathbf{E}} = \begin{bmatrix} \underline{\mathbf{e}}_1 & \underline{\mathbf{e}}_2 & \dots & \underline{\mathbf{e}}_{N-1} & \underline{\mathbf{e}}_N \end{bmatrix}$$

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

Dimostriamo ora che  $\lambda_i \geq 0$  e quindi che la matrice  $\underline{\mathbf{R}}$  è semidefinita positiva. Infatti, sia  $\underline{\mathbf{a}}^* \underline{\mathbf{x}}$  una qualsiasi combinazione lineare degli elementi del vettore  $\underline{\mathbf{x}}$ . Il suo valore quadratico medio è maggiore od eguale a zero:

$$E[\underline{\mathbf{a}}^* \underline{\mathbf{x}} \underline{\mathbf{a}}^* \underline{\mathbf{x}}] = \underline{\mathbf{a}}^* E[\underline{\mathbf{x}} \underline{\mathbf{x}}^*] \underline{\mathbf{a}} = \underline{\mathbf{a}}^* \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{a}} \geq 0 \quad (1.58)$$

Se  $\underline{\mathbf{a}}$  è un autovettore della matrice  $\underline{\mathbf{R}}_{\mathbf{x}}$  risulta:

$$\underline{\mathbf{e}}_i^* \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{e}}_i = \lambda_i \geq 0 \quad (1.59)$$

e quindi tutti gli autovalori di una matrice di covarianza, come le componenti di uno spettro di potenza, sono maggiori o eguali a 0. Inoltre, per la simmetria della matrice, due autovettori  $\underline{\mathbf{e}}_i, \underline{\mathbf{e}}_j$  corrispondenti a due autovalori  $\lambda_i, \lambda_j$  distinti, sono ortogonali, così come due sinusoidi campionate di frequenza diversa, nel senso che la media del loro prodotto è nulla. Infatti:

$$\underline{\mathbf{e}}_i^* \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{e}}_j = \lambda_j \underline{\mathbf{e}}_i^* \underline{\mathbf{e}}_j = \lambda_i \underline{\mathbf{e}}_i^* \underline{\mathbf{e}}_j \rightarrow \underline{\mathbf{e}}_i^* \underline{\mathbf{e}}_j = 0$$

Ancora consideriamo due variabili casuali  $z_1, z_2$  combinazione lineare degli elementi della sequenza  $\{x_n\}$ :

$$z_1 = \underline{\mathbf{a}}_1^* \underline{\mathbf{x}}; \quad z_2 = \underline{\mathbf{a}}_2^* \underline{\mathbf{x}}$$

Risulta ora:

$$E[z_1 z_2] = \underline{\mathbf{a}}_1^* \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{a}}_2$$

e le due variabili casuali  $z_1, z_2$  sono incorrelate se  $\underline{\mathbf{a}}_1^* \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{a}}_2 = 0$  e cioè se i due vettori  $\underline{\mathbf{a}}_1, \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{a}}_2$  sono ortogonali; in questo caso i due vettori  $\underline{\mathbf{a}}_1, \underline{\mathbf{a}}_2$  sono definiti come ortogonali coniugati e se ne discuterà nel capitolo 20.

Supponiamo ora che la funzione di autocorrelazione della sequenza  $\{x_n\}$  tenda a 0 per valori di  $\tau \ll N$ ; la circolarizzazione della matrice  $\underline{\mathbf{R}}_{\mathbf{x}}$  non cambia la matrice in modo significativo (la cambia solo negli angoli).

Consideriamo ora la relazione tra una sequenza complessa filtrata con una convoluzione periodica e la sua trasformata di Fourier discreta:

$$\underline{\mathbf{Y}} = \underline{\mathbf{W}} \underline{\mathbf{y}} \quad (1.60)$$

dove con  $\underline{\mathbf{W}}$  abbiamo indicato la matrice della trasformazione di Fourier diretta. Inoltre:

$$\underline{\mathbf{R}}_{\mathbf{Y}} = E[\underline{\mathbf{Y}} \underline{\mathbf{Y}}^*] = \underline{\mathbf{H}} \underline{\mathbf{R}}_{\mathbf{X}} \underline{\mathbf{H}}^* = \underline{\mathbf{W}} E[\underline{\mathbf{y}} \underline{\mathbf{y}}^*] \underline{\mathbf{W}}^* = \underline{\mathbf{W}} \underline{\mathbf{R}}_{\mathbf{y}} \underline{\mathbf{W}}^* \quad (1.61)$$

dove con  $\underline{\mathbf{R}}_{\mathbf{Y}}$  abbiamo indicato la matrice diagonale che contiene le ampiezze, moltiplicate per  $N$ , del periodogramma della sequenza  $\{y_n\}$ ; infatti abbiamo dimostrato in precedenza che  $E[\underline{\mathbf{Y}} \underline{\mathbf{Y}}^*]$  è diagonale. Possiamo anche scrivere che:

$$\underline{\mathbf{R}}_{\mathbf{y}} = \frac{1}{N} \underline{\mathbf{W}}^* \underline{\mathbf{R}}_{\mathbf{Y}} \underline{\mathbf{W}} \frac{1}{N} = \frac{1}{\sqrt{N}} \underline{\mathbf{W}}^* \frac{\underline{\mathbf{R}}_{\mathbf{Y}}}{N} \underline{\mathbf{W}} \frac{1}{\sqrt{N}} \quad (1.62)$$

e cioè che la sequenza  $\{y_n\}$ , ottenuta con una qualsiasi convoluzione periodica di un filtro e di una sequenza a spettro bianco, ha una matrice di covarianza circolante  $\underline{\mathbf{R}}_{\mathbf{y}}$  che è diagonalizzata dalla trasformazione di Fourier; in altre parole, la sequenza  $\{Y_k\}$  trasformata di Fourier della sequenza  $\{y_n\}$ , ha matrice di covarianza  $\underline{\mathbf{R}}_{\mathbf{Y}}$  diagonale. In particolare, la matrice

$$\frac{1}{\sqrt{N}} \underline{\mathbf{W}}^*$$

è ortogonale, in quanto è inversa della sua trasposta e coniugata. Infatti, le sinusoidi sono qui normalizzate per il fattore  $1/\sqrt{N}$  in modo da avere energia unitaria.

Quindi, gli autovettori della matrice di covarianza  $\underline{\mathbf{R}}_{\mathbf{y}}$  di una sequenza ottenuta da una convoluzione periodica di una sequenza bianca ed un filtro qualsiasi sono le sinusoidi discrete di frequenza  $k/NT$ ;  $0 \leq k \leq N-1$  ed ampiezza  $1/\sqrt{N}$ ; gli autovalori sono le ampiezze del periodogramma.

Nel caso di una matrice di covarianza di un segnale qualsiasi e non di una matrice di covarianza circolante, gli autovalori sono comunque non negativi e gli autovettori sono comunque ortogonali tra loro: non saranno però sinusoidi vere e proprie, ma gli assomiglieranno abbastanza. L'ordinamento dello spettro degli autovalori è diverso però da quello dello spettro di frequenza di una sequenza: lo spettro di frequenza di una sequenza è ordinato sulla base della frequenza degli autovettori (l'indice  $k$ ); lo spettro degli autovalori di una matrice è in genere ordinato per autovalori decrescenti, visto che non è possibile definire una vera e propria frequenza per ogni autovettore.



### 1.8.6 Un esempio numerico

Si consideri una sequenza di  $N = 4$  campioni casuali, a valor medio nullo e varianza unitaria e si indichi con  $\underline{\mathbf{x}}$  il vettore corrispondente:

$$\underline{\mathbf{x}} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}; \quad E[\underline{\mathbf{x}}\underline{\mathbf{x}}^*] = \underline{\mathbf{R}}_{\mathbf{x}} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \underline{\mathbf{I}}$$

Convolviamo periodicamente  $\underline{\mathbf{x}}$  con il filtro:

$$\underline{\mathbf{f}} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix}$$

ed otteniamo il vettore  $\underline{\mathbf{y}}$  le cui componenti sono:

$$y_i = x_i + x_{i+1}; \quad i = 0, \dots, 3; \quad x_4 = x_0 \quad \text{per la ciclicità}$$

In termini matriciali si ha:

$$\underline{\mathbf{y}} = \underline{\mathbf{A}} \underline{\mathbf{x}}$$

dove

$$\underline{\mathbf{A}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

Calcoliamo ora  $\underline{\mathbf{R}}_{\mathbf{y}}$ , la matrice di autocovarianza di  $\underline{\mathbf{y}}$ , circolante per la ciclicità della convoluzione:

$$\underline{\mathbf{R}}_{\mathbf{y}} = E[\underline{\mathbf{y}}\underline{\mathbf{y}}^*] = \underline{\mathbf{A}} \underline{\mathbf{R}}_{\mathbf{x}} \underline{\mathbf{A}}^* = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{bmatrix}$$

come si potrebbe ottenere direttamente dalla funzione di autocorrelazione della sequenza  $\{y_i\}$ ,  $[1 \ 2 \ 1]$  periodicizzata anch'essa. Il filtraggio passa - basso fatto con la convoluzione con il filtro  $\underline{\mathbf{f}}$  comporta uno zero alla frequenza di Nyquist, visto che il numero dei campioni della sequenza è pari. Infatti il vettore  $\underline{\mathbf{F}}$  trasformata di Fourier di  $\underline{\mathbf{f}}$ , è:

$$\underline{\mathbf{F}} = \underline{\mathbf{W}} \underline{\mathbf{f}} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} = \begin{bmatrix} 2 \\ 1-j \\ 0 \\ 1+j \end{bmatrix}$$

Il vettore  $\underline{\mathbf{Y}}$ , trasformata di Fourier di  $\underline{\mathbf{y}}$  si ottiene moltiplicando il vettore  $\underline{\mathbf{X}}$  trasformata di Fourier di  $\underline{\mathbf{x}}$  per una matrice  $\underline{\Phi}$  che ha lungo la diagonale principale gli elementi di  $\underline{\mathbf{F}}$ . Si ha cioè:

$$\underline{\mathbf{Y}} = \underline{\mathbf{W}} \underline{\mathbf{y}} = \begin{bmatrix} 2 & 0 & 0 & 0 \\ 0 & 1-j & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1+j \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_3 \\ X_4 \end{bmatrix} = \underline{\Phi} \underline{\mathbf{W}} \underline{\mathbf{x}}$$

Si ha ancora:

$$E[\underline{\mathbf{Y}} \underline{\mathbf{Y}}^*] = \underline{\Phi} E[\underline{\mathbf{X}} \underline{\mathbf{X}}^*] \underline{\Phi}^*$$

Peraltro:

$$E[\underline{\mathbf{X}} \underline{\mathbf{X}}^*] = 4\underline{\mathbf{I}}$$

e quindi

$$E[\underline{\mathbf{Y}} \underline{\mathbf{Y}}^*] = \underline{\mathbf{R}}_{\mathbf{Y}} = 4 \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

Da un'analisi diretta, è peraltro evidente che la matrice  $\underline{\mathbf{R}}_{\mathbf{Y}}$  ha un autovalore nullo e l'autovettore corrispondente è la sinusoide alla frequenza

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

di Nyquist.

$$\underline{\mathbf{R}}_{\mathbf{y}} \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = 0 \times \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

Per avere gli altri autovalori ed autovettori di  $\underline{\mathbf{R}}_{\mathbf{y}}$ , si può anche ricorrere alla trasformazione di Fourier discreta che diagonalizza la matrice, come già osservato. Se

$$\begin{aligned} \underline{\mathbf{Y}} &= \underline{\mathbf{W}} \underline{\mathbf{y}}; \\ E[\underline{\mathbf{Y}} \underline{\mathbf{Y}}^*] &= \underline{\mathbf{W}} \underline{\mathbf{R}}_{\mathbf{y}} \underline{\mathbf{W}}^* = \underline{\mathbf{R}}_{\mathbf{Y}} \\ \frac{\underline{\mathbf{R}}_{\mathbf{Y}}}{N} &= 1 \times \begin{bmatrix} |F_0|^2 & 0 & 0 & 0 \\ 0 & |F_1|^2 & 0 & 0 \\ 0 & 0 & |F_2|^2 & 0 \\ 0 & 0 & 0 & |F_3|^2 \end{bmatrix} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix} \end{aligned}$$

La matrice  $\underline{\mathbf{R}}_{\mathbf{y}}$  è diagonalizzata dalla trasformazione di Fourier poichè:

$$\begin{aligned} \underline{\mathbf{y}} &= \frac{1}{N} \underline{\mathbf{W}}^* \underline{\mathbf{Y}} \rightarrow \underline{\mathbf{W}}^{-1} = \frac{1}{N} \underline{\mathbf{W}}^* \\ \underline{\mathbf{R}}_{\mathbf{y}} &= \frac{1}{N} \underline{\mathbf{W}}^* \underline{\mathbf{R}}_{\mathbf{Y}} \underline{\mathbf{W}} \frac{1}{N} = \frac{1}{\sqrt{N}} \underline{\mathbf{W}}^* \frac{\underline{\mathbf{R}}_{\mathbf{Y}}}{N} \underline{\mathbf{W}} \frac{1}{\sqrt{N}} \end{aligned}$$

Gli autovettori normalizzati della matrice  $\underline{\mathbf{R}}_{\mathbf{y}}$  sono appunto le sinusoidi (le righe o le colonne della matrice  $\underline{\mathbf{W}}$ ) la cui ampiezza è  $1/\sqrt{N}$  perchè la somma dei quadrati sia unitaria. Pertanto, gli autovalori della matrice  $\underline{\mathbf{R}}_{\mathbf{y}}$  sono proprio gli elementi della diagonale principale della matrice  $\underline{\mathbf{R}}_{\mathbf{Y}}/N$  (il periodogramma) ed in questo caso:

$$[4, 2, 0, 2]$$

Infatti:

$$\begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix} = 4 \times \begin{bmatrix} 1/2 \\ 1/2 \\ 1/2 \\ 1/2 \end{bmatrix}$$

$$\begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ j \\ -1 \\ -j \end{bmatrix} = 2 \times \begin{bmatrix} 1 \\ j \\ -1 \\ -j \end{bmatrix}$$

### 1.8.7 Stime degli autovalori di matrici di covarianza

La matrice di covarianza si ottiene dall'espressione teorica:

$$\underline{\mathbf{R}} = E [\underline{\mathbf{x}} \underline{\mathbf{x}}^*]$$

In realtà poi non è possibile disporre del valore atteso  $E[\ ]$ , ma se ne possono solo fare delle stime; in altre parole ci dobbiamo accontentare di una stima non polarizzata  $\hat{\underline{\mathbf{R}}}$  ottenuta combinando  $M$  realizzazioni del vettore casuale  $\underline{\mathbf{x}}_i$ ;  $i = 1 \dots M$ .

$$\hat{\underline{\mathbf{R}}} = \frac{1}{M} \sum_{i=1}^M \underline{\mathbf{x}}_i \underline{\mathbf{x}}_i^*; \quad \text{Tr}(\hat{\underline{\mathbf{R}}}) = \sum_{n=1}^N \hat{\lambda}_n = \frac{1}{M} \sum_{i=1}^M \underline{\mathbf{x}}_i^* \underline{\mathbf{x}}_i \quad (1.63)$$

$$E[\hat{\underline{\mathbf{R}}}] = \underline{\mathbf{R}}; \quad E[\text{Tr}(\hat{\underline{\mathbf{R}}})] = \frac{M}{M} N \sigma_x^2 = N \sigma_x^2$$

Le dimensioni del vettore  $\underline{\mathbf{x}}_i$  sono  $[1 \times N]$  e la matrice  $\underline{\mathbf{R}}$  ha quindi dimensioni  $[N \times N]$ . Tutto è ragionevole se:

$$M \gg N$$

Allora, le stime non sono molto disperse attorno ai valori attesi e il risultato derivante dall'equazione 1.63 è accettabile. In particolare,  $\hat{\lambda}_n$ ,  $n = 1 \dots N$ , autovalori di  $\hat{\underline{\mathbf{R}}}$ , sono delle ragionevoli stime dei  $\lambda_n$ ,  $n = 1 \dots N$ , autovalori di  $\underline{\mathbf{R}}$ , e similmente per gli autovettori corrispondenti. Tuttavia, non sempre è possibile essere in questa situazione e può accadere che invece addirittura sia:

$$M < N$$

Cosa accade allora? È evidente che  $N - M > 0$  autovalori della matrice sono proprio nulli, perchè la matrice  $\hat{\underline{\mathbf{R}}}$  è somma di  $M$  diadi (prodotti  $\underline{\mathbf{x}}_i \underline{\mathbf{x}}_i^*$ ) e il suo rango è non superiore a  $M$ . Peraltro, la traccia della matrice

## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

$\hat{\mathbf{R}}$  e quindi la somma dei suoi autovalori (contano solo quelli non nulli)  $\hat{\lambda}_n$ , ha sempre valore atteso  $N\sigma_x^2$  e dispersione ridotta trattandosi di un numero solo risultato della media di  $MN$  quadrati di numeri non tanto dipendenti tra loro. Quindi, gli  $M$  autovalori non nulli di  $\hat{\mathbf{R}}$  sono *strizzati* ed hanno cioè un valore atteso medio anche molto più grande in quanto si devono spartire tra  $M$  valori la potenza totale di  $N$  campioni.

$$E[\hat{\lambda}_n] = \frac{N}{M}\sigma_x^2; n = 1..M; \quad \hat{\lambda}_n = 0; n = M + 1, \dots, N;$$

In casi intermedi in cui  $M$  non è grande come si desidererebbe, la matrice  $\hat{\mathbf{R}}$  tende ad essere comunque deficiente in rango; quindi, gli autovalori stimati più piccoli sono più piccoli del valore corrispondente, mentre il valore atteso della stima dell'autovalore più grande diventa più grande del valore corretto.

$$E[\hat{\lambda}_{\max}] > \lambda_{\max}; \quad E[\hat{\lambda}_{\min}] < \lambda_{\min}$$

Per cui, anche se la stima  $\hat{\mathbf{R}}$  della matrice  $\mathbf{R}$  non è polarizzata, le stime degli autovalori lo sono. Sono disponibili dei risultati teorici (Everson, R., Roberts, S.: *Inferring the eigenvalues of covariance matrices from limited noisy data*; IEEE Trans. Sig. Proc. Luglio 2000, pp. 2083 -2091) che forniscono la densità di probabilità di questi autovalori.

### Fluttuazione dei corsi dei titoli di borsa

Questo problema diventa di notevole interesse nel caso dell'analisi delle fluttuazioni del corso di titoli in borsa, p.e. per la costruzione di portafogli di titoli a rischio minimo. In questi casi, si cerca di verificare se le fluttuazioni del corso di due titoli siano anticorrelate o correlate o incorrelate. Il tipico esempio è quello delle azioni di compagnie di aerolinee e di compagnie petrolifere: se sale il prezzo del petrolio, i titoli delle aerolinee scendono e quelli delle petrolifere salgono, e quindi le fluttuazioni dei corsi dei titoli rispettivi saranno anticorrelate; infatti, ci si attende che i profitti delle une scendano e quelli delle altre salgano. Allora, chi volesse rendersi indipendente dalle fluttuazioni del costo del petrolio, comprenderebbe un giardinetto composto da titoli sia delle une che delle altre, in modo da minimizzare quel rischio. Gli autovettori della matrice di

covarianza forniscono giardinetti di titoli le cui fluttuazioni sono incorrelate. Poi ci saranno giardinetti a massima fluttuazione (l'autovettore corrispondente all'autovalore massimo) ecc.

Il problema è però che ci sono migliaia di titoli in borsa ( $N$ ), e pochissimi anni ( $M$ ) su cui fare delle statistiche, peraltro tempo varianti. Quindi, non si avrebbe comunque nessuna utilità a risalire troppo indietro nel tempo per arricchire le statistiche; ancora, non si ha molta utilità a sovracampionare temporalmente (p.e. rilevando i corsi ogni ora od ogni giorno), perchè poi si ha a che fare con campioni correlati tra loro e scarsamente utili per ridurre la varianza della stima. Gli analisti finanziari sono praticamente sempre nella condizione  $M < N$  e quindi devono ricordare bene che le loro stime della matrice di covarianza delle fluttuazioni di migliaia di titoli sono matrici di rango sempre ridottissimo e le stime degli autovalori sono polarizzate. Una soluzione a questo problema è la parametrizzazione e quindi la riduzione della dimensionalità della matrice di covarianza. Si veda p.e. il lavoro: Ledoit, O. and Wolf, M. , *Honey, I Shrunk the Sample Covariance Matrix* (June 2003), UPF Economics and Business Working Paper No. 691). Si possono raggruppare i titoli in comparti (aziende di semiconduttori, supermercati, aerolinee, assicurative, petrolifere, acciai, ecc.) e si studia la matrice di covarianza di dimensioni molto più ridotte che corrisponde alla covarianza delle fluttuazioni medie dei titoli di un comparto rispetto a quelle dei titoli di un altro comparto. Poi, all'interno di ogni comparto, si studiano le caratteristiche rispettive delle varie aziende. Se va bene per una, ci si aspetta che vada bene per tutte, e viceversa ... Quindi accade in borsa il fenomeno apparentemente strano per cui se p.e. un'azienda di semiconduttori o di farmaceutici dichiara profitti (o perdite) eccezionali, tutto il comparto ne risente positivamente (o negativamente). Eppure, sarebbe ragionevole indurre che per i meccanismi di concorrenza in un mercato limitato, il valore di un comparto dovrebbe essere bilanciato, per cui se ad un'azienda va meglio, alle altre dovrebbe andare peggio. Ma non è così: solo in momenti successivi sarà compito degli analisti di borsa ricostruire esattamente la posizione concorrenziale di un'azienda rispetto alle altre e ribilanciare i pesi delle aziende nel comparto, stabilendo se un'azienda sia meglio gestita di un'altra, ecc. Peraltro, possono poi nascere compagnie diversificate che operano contemporaneamente in diversi comparti e possono compensare, all'interno dei loro bilanci, profitti e perdite degli uni e degli altri. Questi conglomerati sono essere meno trasparenti e più

difficili da studiare e da categorizzare.

## 1.9 Appendice 2

### 1.9.1 Autocorrelazione e spettro di potenza

La definizione dell'autocorrelazione nel dominio del tempo continuo per segnali complessi è:

$$r(\tau) = E[x(t)x^*(t - \tau)] = E[x^*(t)x(t + \tau)]; \quad r(-\tau) = r^*(\tau)$$

Per definizione, lo spettro di potenza  $S(\omega)$  è:

$$S(\omega) = \int_{-\infty}^{\infty} r(\tau) e^{-j\omega\tau} d\tau$$

Infatti, considerando un intervallo di tempo di osservazione pari a  $T_{tot}$  :

$$X(\omega) = \int_{-T_{tot}/2}^{T_{tot}/2} x(t) e^{-j\omega t} dt; \quad X^*(\omega) = \int_{-T_{tot}/2}^{T_{tot}/2} x^*(t) e^{j\omega t} dt$$

$$|X(\omega)|^2 = \int_{-T_{tot}/2}^{T_{tot}/2} \int_{-T_{tot}/2}^{T_{tot}/2} x(t) x^*(\xi) e^{j\omega\xi} e^{-j\omega t} dt d\xi$$

Ponendo  $t - \xi = \tau$ ;

$$|X(\omega)|^2 = \int_{-T_{tot}/2}^{T_{tot}/2} d\xi \int_{-T_{tot}/2-\xi}^{T_{tot}/2-\xi} x(\xi + \tau) x^*(\xi) e^{-j\omega\tau} d\tau$$

Passando alla media di insieme:

$$\begin{aligned} E[|X(\omega)|^2] &= \int_{-T_{tot}/2}^{T_{tot}/2} d\xi \int_{-T_{tot}/2-\xi}^{T_{tot}/2-\xi} r(\tau) e^{-j\omega\tau} d\tau \simeq \\ &\simeq S(\omega) \int_{-T_{tot}/2}^{T_{tot}/2} d\xi = S(\omega) \times T_{tot} \end{aligned}$$

perchè si può pensare che  $r(\tau)$  si annulli già per valori dell'argomento ben inferiori a  $T_{tot}$ . Volendo una maggiore precisione, si può tenere conto della finestratura in tempi ( $T_{tot}$ ) con una finestratura triangolare della

funzione di autocorrelazione (vedi eq. 10.23). Passando al limite per  $T_{tot} \rightarrow \infty$ :

$$\lim_{T_{tot} \rightarrow \infty} \frac{E[|X(\omega)|^2]}{T_{tot}} = S(\omega)$$

Lo spettro di energia è per il teorema di Parseval e per un tempo  $T_{tot}$  limitato:

$$\int_{-\infty}^{\infty} |X(\omega)|^2 \frac{d\omega}{2\pi} = \int_{-T_{tot}/2}^{T_{tot}/2} |x(t)|^2 dt = T_{tot} P_m$$

dove con  $P_m$  abbiamo indicato la potenza media (il valore quadratico medio) del segnale  $x(t)$ . Quindi

$$\int_{-\infty}^{\infty} \frac{|X(\omega)|^2}{T_{tot}} \frac{d\omega}{2\pi} = \int_{-\infty}^{\infty} S(\omega) \frac{d\omega}{2\pi} = \frac{1}{T_{tot}} \int_{-T_{tot}/2}^{T_{tot}/2} |x(t)|^2 dt = P_m$$

e quindi  $S(\omega)$  corrisponde allo spettro di potenza del segnale e cioè alla ripartizione della potenza del segnale nelle varie bande spettrali. Si osservi che  $S(\omega)$  non è necessariamente simmetrico rispetto ad  $\omega$  in quanto il segnale  $x(t)$  può essere complesso.

### 1.9.2 Dimensione delle variabili continue e discrete

Se i segnali  $x(t)$  sono delle tensioni, indicheremo tale dimensionalità con:

$$x(t) \rightarrow [V];$$

Nel discreto

$$x_n = x(nT) \rightarrow [V]$$

Le trasformate continue hanno dimensionalità:

$$X(\omega) = \int_{-\infty}^{\infty} x(t) e^{-j\omega t} dt \rightarrow [VT]$$

Le trasformate discrete hanno invece la stessa dimensionalità del segnale

$$X_k = \sum_{n=0}^{N-1} x_n W_N^{-nk} \rightarrow [V]$$



## CAPITOLO 1 ANALISI SPETTRALE NON PARAMETRICA

e pertanto i campioni della trasformata continua si ottengono da quelli della trasformata discreta moltiplicando per  $T$ , sempre se il campionamento è corretto:

$$X(\omega_k) = \int_{-\infty}^{\infty} x(t) e^{-j\omega_k t} dt \simeq T \sum_{n=0}^{N-1} x_n W_N^{-nk} = X_k T$$

La funzione di autocorrelazione ha dimensione  $[V^2]$  e cioè è una potenza (trascuriamo il termine dovuto alla resistenza  $[\Omega^{-1}]$ ). Infine lo spettro di potenza di una variabile continua  $S(\omega)$

$$S(\omega) = \frac{|X(\omega)|^2}{T_{tot}} \rightarrow \left[ \frac{V^2 T^2}{T_{tot}} \right] = [V^2 T]$$

Nel discreto, il periodogramma è collegato allo spettro di potenza dalla relazione:

$$S(\omega_k) = \frac{E[|X(\omega_k)|^2]}{T_{tot}} \simeq \frac{|X_k|^2 T^2}{NT} = \frac{T}{N} |X_k|^2$$

### 1.9.3 Teorema del limite centrale

A titolo di esempio, mostriamo come il teorema del limite centrale sia una conseguenza immediata del meccanismo delle convoluzioni ripetute; si consideri una variabile casuale  $x_i$ , caratterizzata da avere solo due valori possibili ed equiprobabili con probabilità 0.5:

$$x_i = \pm \frac{1}{\sqrt{N}}; \quad p_x(x_i = 1/\sqrt{N}) = p_x(x_i = -1/\sqrt{N}) = .5$$

La trasformata di Fourier della densità di probabilità è:

$$P(u) = \cos\left(\frac{u}{\sqrt{N}}\right)$$

La variabile casuale  $a$ , somma di  $N$  valori della  $x_i$ , ha media nulla e varianza unitaria; infatti:

$$a = \sum_{i=1}^N x_i; \quad E[a^2] = N \sum E[x_i^2] = 1; \quad E[x_i x_j] = 0; i \neq j$$

La densità di probabilità di  $a$  è approssimativamente gaussiana; essa è ottenuta convolvendo  $N$  volte con se stessa la densità di probabilità

$p_x(x)$ . La sua trasformata di Fourier è quindi la potenza  $N$ -esima della  $P(u)$ :

$$P_a(u) = [P(u)]^N = \cos^N \left( \frac{u}{\sqrt{N}} \right)$$

$$\cos^N \left( \frac{u}{\sqrt{N}} \right) \simeq \left( 1 - \frac{u^2}{2N} \right)^N \simeq e^{-u^2/2} ; \quad u \ll \sqrt{N}$$

L'antitrasformata di Fourier, gaussiana a varianza unitaria è quindi:

$$p_a(\xi) = \frac{1}{\sqrt{2\pi}} e^{-\xi^2/2}$$

Si osservi che il fatto che la variabile  $x$  assuma solo due valori discreti, implica che anche la variabile  $a$  sia discreta, a passo  $1/\sqrt{N}$ ; infatti, la sua trasformata  $P_a(u)$ , essendo trasformata di una variabile campionata, è periodica di periodo  $2\pi\sqrt{N}$  e quindi il periodo cresce con  $N$ . Peraltro, l'approssimazione fatta ponendo  $u \ll \sqrt{N}$ , implica un filtraggio passa-basso e quindi lo smussamento della distribuzione di  $a$  e quindi ci impone di trascurare il fatto che essa assuma solo dei valori discreti ad intervalli  $1/\sqrt{N}$ .



## Capitolo 2

# Analisi Spettrale Parametrica

### 2.1 Introduzione

Si consideri una sequenza di campioni indipendenti  $\{x_n\}$  e con essa si alimenti un filtro  $H(\omega)$ ; se la trasformata della sequenza  $\{x_n\}$  è indicata con  $X(\omega)$ , quella dell'uscita sarà:

$$Y(\omega) = X(\omega)H(\omega)$$

$$E[|Y(\omega)|^2] = E[|X(\omega)|^2] \times |H(\omega)|^2 \quad (2.1)$$

$$H(\omega) = \frac{B(\omega)}{A(\omega)} \quad (2.2)$$

La densità spettrale di potenza bilatera del processo bianco  $\{x_n\}$  è costante perché per sequenze di campioni indipendenti di lunghezza  $N$ , le componenti della trasformata di Fourier discreta sono tutte di eguale valore quadratico medio (assumendo  $T = 1$ ):

$$\frac{E[|X_k^2|]}{N} = \sigma_x^2 \quad (2.3)$$

L'analisi spettrale parametrica consiste nel determinare i parametri del filtro  $H(\omega)$  (fig. 11.1) in modo che lo spettro in uscita del filtro assomigli quanto possibile allo spettro del segnale  $\{y_n\}$  da analizzare. Avremo analisi spettrale Moving Average (MA) se il filtro ha una trasformata  $z$  caratterizzata da tutti zeri; avremo il caso Auto Regressivo (AR) quando il filtro è tutti poli e cioè autoregressivo: infine avremo il caso misto

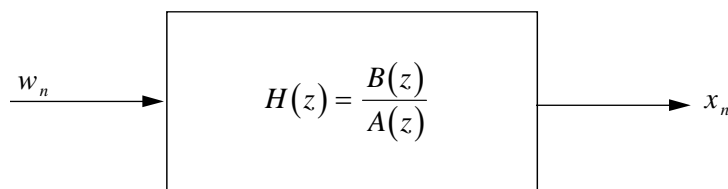


Figura 2.1: Filtro i cui parametri vengono calcolati nell'analisi spettrale AR e MA

ARMA (Auto Regressive Moving average) nel caso piú generale di poli e zeri.

Vedremo che l'analisi spettrale parametrica tutti zeri praticamente coincide con le tecniche non parametriche modificate delle ultime sezioni del capitolo precedente. Le tecniche spettrali AR invece sono di natura ben diversa ed innovativa ed ad esse destineremo grande attenzione per tutta una serie di motivi che chiariremo in seguito.

L'analisi spettrale ARMA poi, é di uso meno frequente anche perché come noto, un qualsiasi filtro può essere rappresentato con soli zeri o soli poli e una rappresentazione mista serve solo per una descrizione della stessa (o di una simile) funzione di trasferimento con un numero inferiore di parametri.

La validità di una descrizione AR o MA o ARMA dipende comunque dalla qualità del modello così ottenuto.

## 2.2 Analisi tutti zeri (MA)

Si consideri un filtro FIR con risposta all'impulso  $\{b_h\}$  la cui trasformata  $z$  é caratterizzata da tutti zeri (fig. 11 .2):

$$B(z) = \sum_{h=0}^H b_h z^{-h} \quad (2.4)$$

Sia  $\{x_n\}$  la sequenza bianca all'ingresso del filtro ed  $\{y_n\}$  la sequenza, spettralmente colorata, in uscita. La funzione di autocorrelazione della

## 2.2 ANALISI TUTTI ZERI (MA)

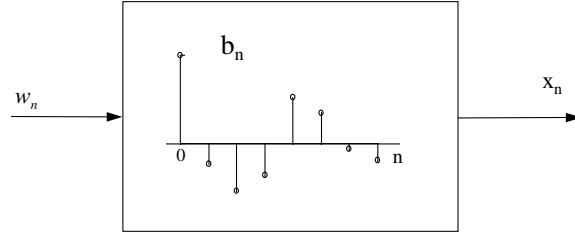


Figura 2.2: Filtro FIR con risposta impulsiva  $\{b_n\}$

sequenza  $\{y_n\}$  é:

$$r_{y,m} = E \left[ \sum_{h,k=0}^H b_h b_k x_{n-h} x_{n+m-k} \right] = \sigma_x^2 \sum_{h=0}^H b_h b_{m+h} \quad (2.5)$$

dove si è utilizzato il fatto che:

$$E[x_{n-h} x_{n+m-k}] = \sigma_x^2 \delta(h + m - k) \quad (2.6)$$

Ricordiamo che il nostro problema è quello di determinare un filtro  $\{b_h\}$  o la sua trasformata  $B(z)$  (la soluzione non è unica) a partire da una stima dell'autocorrelazione dei dati. Per ora supponiamo che la stima disponibile sia molto buona, per cui praticamente possiamo supporre di avere la funzione di autocorrelazione "vera"  $r_m$ .

Passando alla trasformata  $z$  si vede che  $R(z)$ , trasformata  $z$  della  $r_m$  é (fig. 11 .3):

$$R(z) = \sigma_x^2 B(z) B(z^{-1}) \quad (2.7)$$

Si osservi che  $R(z)$  é un polinomio nella variabile  $z^{-1}$  che insieme ad ogni radice, ha anche la radice reciproca. Se le radici sono sul cerchio unitario, saranno a coppie. Quindi, un modo per determinare un filtro  $B(z)$ , assegnata la funzione di autocorrelazione  $R(z)$ , é quello di trovare le radici di  $R(z)$  e per esempio assegnare a  $B(z)$  tutte le  $H$  radici all'esterno del cerchio unitario e a  $B(z^{-1})$  tutte le altre (all'interno). In questo modo,  $B(z^{-1})$  risulta a fase minima; altrimenti possiamo scegliere  $B(z^{-1})$  a fase massima o fase mista, in  $2^H$  modi diversi (fig. 11 .4).

Si osservi che non basta che la  $R(z)$  sia un qualsiasi polinomio simmetrico, perchè si possa individuare un filtro  $B(z)$ . Infatti si avrebbero

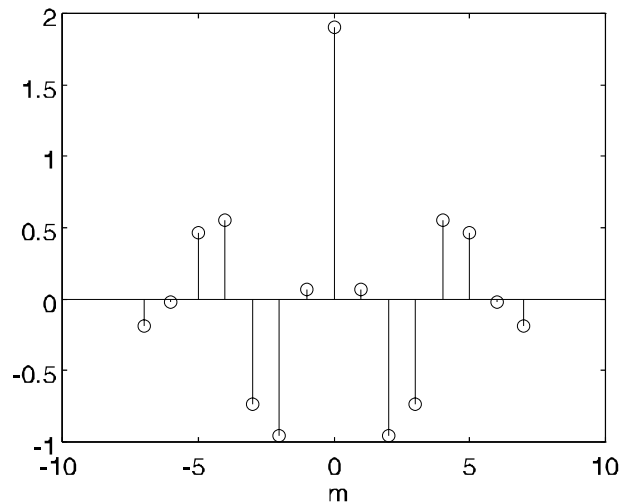


Figura 2.3: Funzione di autocorrelazione del filtro di fig.9.2

coppie di zeri reciproci ma si potrebbero avere anche zeri semplici sul cerchio unitario (fig. 11.5); in questo caso non sarebbe possibile trovare la  $B(z)$  perchè non si potrebbe associare un'altra radice in posizione reciproca. Ma in questo caso, il polinomio simmetrico  $R(z)$  **non rappresenterebbe** una funzione di autocorrelazione in quanto la trasformata di Fourier della  $r_{y,m}$  e cioè i valori assunti dalla  $R(z)$  sul cerchio unitario, conterebbero cambi di segno al passaggio per gli zeri e quindi anche valori negativi, mentre invece uno spettro di potenza, Trasformata di Fourier di una funzione di autocorrelazione vera, non sarà mai negativo.

In conclusione, se lo spettro di un polinomio simmetrico è tutto positivo,

- le radici sono sempre a coppie e non compaiono radici isolate,
- si può trovare un filtro  $B(z)$  in  $2^H$  modi diversi, se  $H$  è l'ordine del filtro.

Una sequenza simmetrica invece, può avere zeri semplici (e non doppi) sul cerchio unitario e quindi non corrispondere ad una funzione di autocorrelazione.

### 2.2.1 Troncamento della funzione di autocorrelazione

Si vede pertanto che se si ha:

## 2.2 ANALISI TUTTI ZERI (MA)

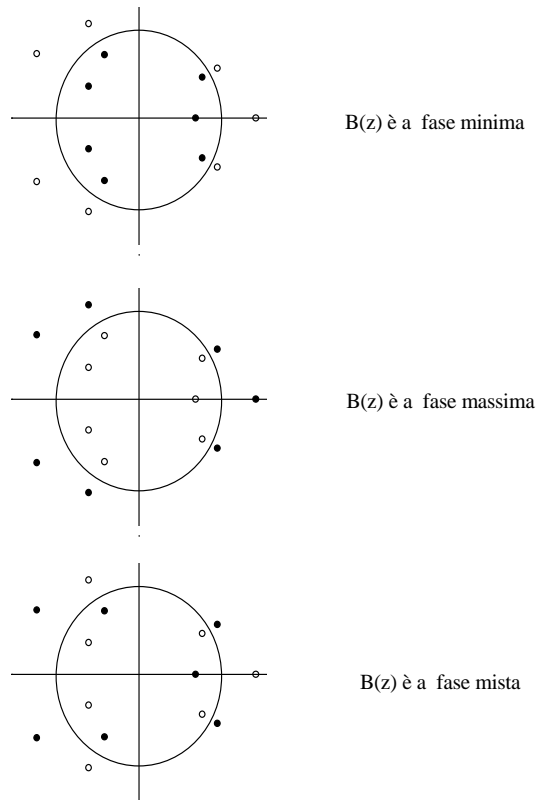


Figura 2.4: Esempi di assegnazione degli zeri della  $R(z)$  alla  $B(z)$  (zeri anneriti)

- una funzione di autocorrelazione (spettro positivo)
- nulla per  $m > H$ .

allora è sempre possibile trovare  $2^H$  filtri di lunghezza  $H + 1$  che alimentati da sequenze bianche hanno in uscita quella autocorrelazione.

Tuttavia, sono disponibili solo stime delle funzioni di autocorrelazione: se la stima è fatta con la correlazione con se stessa della sequenza bordata di zeri, allora la sua trasformata di Fourier, il periodogramma, è sempre positivo. Tuttavia, in questo caso, la lunghezza del filtro risulta eccessiva perchè, a causa della dispersione della stima, i campioni dell'autocorrelazione non saranno mai nulli.

Se si desidera limitare la lunghezza  $H$  del filtro, si deve rastremare a 0 la funzione di autocorrelazione in  $H$  campioni, finstrandola in modo che



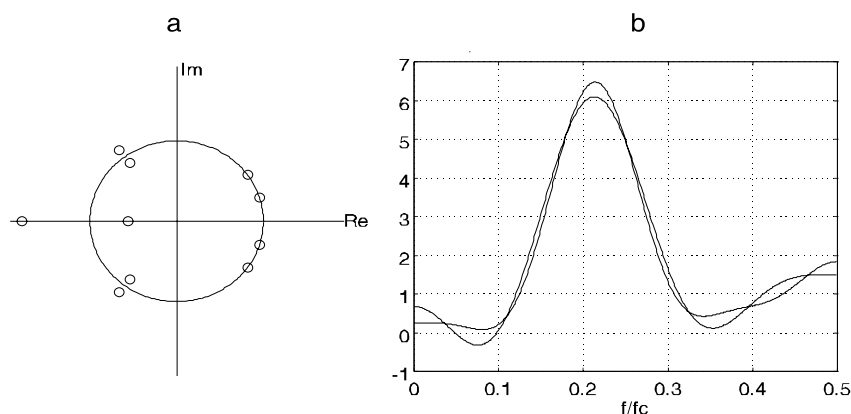


Figura 2.5: a) Zeri dell'autocorrelazione di fig 9.3 troncata con finestra rettangolare b) Spettro di potenza derivante dalla finestratura rettangolare; è sovrapposto lo spettro vero

lo spettro si mantenga positivo, e quindi moltiplicando per una finestra che in pratica è sempre triangolare (fig. 11.6). Usando finestre più efficienti come il coseno rialzato, non si potrebbe garantire che la correlazione finestrata sia a spettro positivo e pertanto che la ricerca del filtro  $B(z)$  possa poi andare a buon fine.

In conclusione, per fare una stima parametrica tutti zeri, è necessario finestrare la funzione di autocorrelazione con una finestra triangolare di lunghezza  $2H$ ; quanto maggiore è  $H$  tanto maggiore sarà la risoluzione spettrale della stima parametrica. In sostanza, si vede che questa tecnica di stima spettrale coincide con quella dello smussamento del periodogramma presa in esame nel precedente capitolo: l'unica differenza sta nel fatto che ora è possibile dare un significato fisico alla semilunghezza della finestra sull'autocorrelazione (e cioè la lunghezza dell'antitrasformata del filtro usato per lo smussamento dello spettro).

## 2.3 Analisi tutti poli (AR)

Questa tecnica di stima spettrale è molto importante per tutta una serie di motivi che emergeranno via via che procederà lo studio: per ora si osservi che avendo i filtri IIR risposta all'impulso illimitata, possono

## 2.3 ANALISI TUTTI POLI (AR)

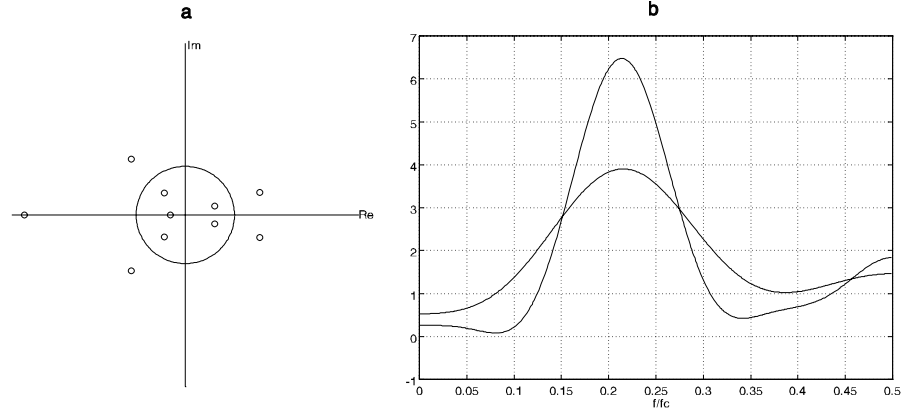


Figura 2.6: a) Zeri della autocorrelazione di fig. 9.3 troncata con finestra triangolare b) Spettro di potenza derivante dalla finestrazione triangolare; è sovrapposto lo spettro vero

produrre spettri di grande risoluzione spettrale con un numero limitato di parametri. Pertanto si prestano alla descrizione di fenomeni che hanno un lungo tempo di coerenza, in cui cioè il processo si decorrela assai lentamente.

Sia  $\{x_n\}$  la sequenza in uscita di un filtro IIR di ordine  $N$ ,  $1/A_N(z)$  alimentato da una sequenza bianca  $\{w_n\}$ .

$$x_n = -\sum_{h=1}^N a_{h,N} x_{n-h} + w_n; \quad X(z) = \frac{W(z)}{A_N(z)}; \quad (2.8)$$

$$A_N(z) = 1 + \sum_{h=1}^N a_{h,N} z^{-h}$$

L'autocorrelazione della sequenza di uscita, che calcoliamo moltiplicando l'equazione (2.8) per il campione  $x_{n-m}^*$  anticipato di  $m$  campioni e coniugato e poi considerando il valore atteso,  $r_m$  è:

$$r_m = E[x_n x_{n-m}^*] = -E\left[\sum_{h=1}^N a_{h,N} x_{n-h} x_{n-m}^*\right] + E[w_n x_{n-m}^*] = \quad (2.9)$$

$$= -\sum_{h=1}^N a_{h,N} r_{m-h} + \sigma_w^2 \delta(m); \quad \forall m \quad (2.10)$$

## CAPITOLO 2 ANALISI SPETTRALE PARAMETRICA

Infatti la sequenza  $\{w_n\}$  è incorrelata con i campioni precedenti  $x_{n-m}$ ;  $m > 0$ . Le equazioni (2.10) sono note come le equazioni di Yule - Walker e forniscono **tutta l'autocorrelazione dell'uscita**, dati i coefficienti del filtro  $A_N(z)$  ed il v.q.m. del segnale d'ingresso  $\sigma_w^2$ . Si osservi che nel caso di segnale e o coefficienti  $a_{h,N}$  complessi, le equazione (2.10) vanno scritte tenendo conto che  $r_{-m} = r_m^*$ . Per motivi mnemonici, può essere utile osservare che la formula (2.10) assomiglia ad una convoluzione tra le sequenze  $r_m$  ed  $a_{h,N}$ ; non è una convoluzione per la limitazione degli indici. L'equazione corrispondente a  $m = 0$  contiene anche la varianza del segnale di ingresso  $\sigma_w^2$ ; peraltro, il filtro non dipende da questo fattore di scala e si può sempre riscalarare la funzione di autocorrelazione ponendo  $r_0 = 1$  risolvendo poi le equazioni (Y-W) per  $m = 1, \dots, N$ . Il fatto che l'equazione (2.10) sia definita per tutti i valori di  $m$ , implica che assegnata qualsiasi  $N - pla$  consecutiva di valori di  $r_m$ , poi possiamo determinare i valori di  $r_m$  per qualsiasi altro valore.

Ancora, le equazioni di Yule Walker si possono ottenere in forma matriciale in modo assai semplice indicando i seguenti vettori con i corrispondenti simboli:

$$\underline{\mathbf{a}}_N = \begin{bmatrix} a_{1,N} \\ a_{2,N} \\ a_{3,N} \\ \dots \\ a_{N,N} \end{bmatrix} ; \underline{\boldsymbol{\xi}}_N = \begin{bmatrix} x_{n-1}^* \\ x_{n-2}^* \\ x_{n-3}^* \\ \dots \\ x_{n-N}^* \end{bmatrix} ; \underline{\mathbf{r}}_N = \begin{bmatrix} r_x(1) \\ r_x(2) \\ r_x(3) \\ \dots \\ r_x(N) \end{bmatrix} = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \dots \\ r_N \end{bmatrix}$$

Allora si ha:

$$x_n = -\underline{\boldsymbol{\xi}}_N^* \underline{\mathbf{a}}_N + w_n \quad (2.11)$$

E moltiplicando a sinistra per  $\underline{\boldsymbol{\xi}}_N$  e considerando il valore atteso, visto che  $E[\underline{\mathbf{x}}_N w_n] = \underline{\mathbf{0}}$ :

$$E[\underline{\boldsymbol{\xi}}_N x_n] = \underline{\mathbf{r}}_N = E[\underline{\boldsymbol{\xi}}_N (-\underline{\boldsymbol{\xi}}_N^* \underline{\mathbf{a}}_N + w_n)] = -E[\underline{\boldsymbol{\xi}}_N \underline{\boldsymbol{\xi}}_N^*] \underline{\mathbf{a}}_N = -\underline{\mathbf{R}} \underline{\mathbf{a}}_N$$

La matrice dei coefficienti delle equazioni (Y-W) è una matrice di Toeplitz; infatti è simmetrica (o Hermitiana, per sequenze complesse) e tutti gli elementi appartenenti alla stessa diagonale o subdiagonale sono

## 2.4 LA STIMA SPETTRALE AR

eguali tra loro. La matrice é pertanto caratterizzata da  $N$  numeri. Il vettore dei termini noti é anche esso formato da campioni della funzione di autocorrelazione,  $r_m$ ;  $m = 1, \dots, N$  col segno cambiato. Riscrivendo in forma matriciale le equazioni di Yule Walker infatti si ottiene:

$$\begin{bmatrix} r_0 & r_{-1} & r_{-2} & \dots & r_{-N+1} \\ r_1 & r_0 & r_{-1} & \dots & r_{-N+2} \\ r_2 & r_1 & r_0 & \dots & r_{-N+3} \\ \dots & \dots & \dots & \dots & \dots \\ r_{N-1} & r_{N-2} & r_{N-3} & \dots & r_0 \end{bmatrix} \begin{bmatrix} a_{1,N} \\ a_{2,N} \\ a_{3,N} \\ \dots \\ a_{N,N} \end{bmatrix} = \underline{\mathbf{R}}\mathbf{a}_N = -\underline{\mathbf{r}}_N = - \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \dots \\ r_N \end{bmatrix} \quad (2.12)$$

ed in termini vettoriali:

$$\underline{\mathbf{R}}\mathbf{a}_N = -\underline{\mathbf{r}}_N \quad (2.13)$$

Per completezza, si aggiunge la formulazione completa, facilmente verificabile, in cui compare anche la prima equazione che contiene la varianza della sequenza bianca di ingresso  $\sigma_w^2$ . Si stia attenti al fatto che in questa equazione non sempre le incognite sono contenute nel vettore  $\mathbf{a}_N$ ; a volte è la funzione di autocorrelazione a non essere nota, oppure il valore di  $\sigma_w^2$ , ecc.

$$\begin{bmatrix} r_0 & r_{-1} & r_{-2} & \dots & r_{-N} \\ r_1 & r_0 & r_{-1} & \dots & r_{-N+1} \\ r_2 & r_1 & r_0 & \dots & r_{-N+2} \\ \dots & \dots & \dots & \dots & \dots \\ r_N & r_{N-1} & r_{N-2} & \dots & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_{1,N} \\ a_{2,N} \\ \dots \\ a_{N,N} \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (2.14)$$

## 2.4 La stima spettrale AR

Una volta trovati gli  $a_{h,N}$  a partire per esempio dai valori assunti dalla funzione di autocorrelazione per  $m = 1, \dots, N$  é immediato determinare

## CAPITOLO 2 ANALISI SPETTRALE PARAMETRICA

le componenti dello spettro continuo del segnale  $\{x_i\}$ ; infatti la funzione di autocorrelazione, che ne è la trasformata, è determinata per tutti i valori dell'indice temporale  $m$ ; pari a quella assegnata per  $m = 1 : N$  e poi desunta dalle equazioni di Yule Walker per tutti gli altri valori.

$$E[|X(\omega)|^2] = \frac{E[|W(\omega)|^2]}{|A(e^{j\omega T})|^2} = \frac{\sigma_w^2 T}{|1 + \sum_{h=1}^N a_{h,N} e^{-j\omega h T}|^2}; \quad (2.15)$$

Per dare un significato fisico a questa stima è necessario garantire che il filtro  $1/A_N(z)$  sia stabile e pertanto che il polinomio  $A_N(z)$  sia a minima fase; si dimostrerà nei prossimi capitoli che ciò accade **se e solo se** i campioni  $r_m$  che si utilizzano nelle equazioni di Yule Walker sono campioni di una funzione di autocorrelazione (a spettro positivo).

Di nuovo, per avere la stima spettrale AR di ordine  $N$ , non serve **tutta** la funzione di autocorrelazione ma ne bastano i primi  $N$  campioni; gli infiniti altri sono calcolabili ancora in base alle equazioni di Y - W a partire dal polinomio  $A_N(z)$ . In altre parole, il troncamento della funzione di autocorrelazione non comporta problemi; semplicemente, i valori dell'autocorrelazione predetti in base alle equazioni Y-W, non coincideranno con quelli effettivi, quando la stima spettrale è fatta con un ordine (il valore di  $N$ ) troppo basso.

## Capitolo 3

### La predizione lineare

Si consideri una sequenza  $\{x_i\}$  a spettro non uniforme; per fissare le idee, si pensi a  $\{x_i\}$  come ad una sequenza uscente da un filtro tutti poli con funzione di trasferimento  $1/A_{N_0}(z)$ , a sua volta alimentato da una sequenza a spettro bianco  $\{w_i\}$  (fig12.1). Si voglia stimare l' $i$ -esimo campione

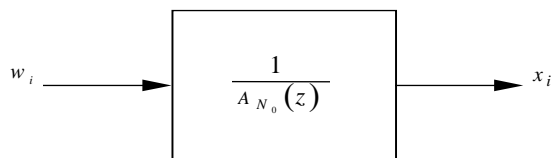


Figura 3.1: Filtro tutti poli

della sequenza, una volta che ne siano noti tutti i campioni precedenti, fino a quello di posto  $i - 1$ . In particolare, si desidera che la predizione sia combinazione lineare degli  $N$  campioni precedenti della sequenza con coefficienti che indicheremo con  $a_{n,N}$ ;  $n = 1, \dots, N$ ; indicando con  $\hat{x}_i$  il valore predetto del campione  $i$ -esimo:

$$\hat{x}_i = - \sum_{n=1}^N a_{n,N} x_{i-n} \quad (3.1)$$

### CAPITOLO 3 LA PREDIZIONE LINEARE

Indicheremo con  $\varepsilon_{i,N}$  l'errore di predizione corrispondente ad un predittore di ordine  $N$  e cioè realizzato con  $N$  coefficienti:

$$\varepsilon_{i,N} = x_i - \hat{x}_i = x_i + \sum_{n=1}^N a_{n,N} x_{i-n} \quad (3.2)$$

La determinazione dei coefficienti  $a_{n,N}$  viene fatta in modo ottimo, minimizzando il valore atteso del quadrato dell'errore:

$$E[|\varepsilon_{i,N}|^2] = E[(x_i + \sum_{n=1}^N a_{n,N} x_{i-n})(x_i + \sum_{n=1}^N a_{n,N} x_{i-n})^*] \quad (3.3)$$

Per trovare gli  $a_{n,N}$  basta minimizzare rispetto ad essi il valore dell'errore q.m. e quindi imporre che:

$$\frac{\partial E[|\varepsilon_{i,N}|^2]}{\partial a_{k,N}} = 2E[(x_i + \sum_{n=1}^N a_{n,N} x_{i-n})x_{i-k}^*] = 2E[\varepsilon_{i,N}x_{i-k}^*] = 0; \quad k = 1, \dots, N. \quad (3.4)$$

Queste equazioni corrispondono alla **condizione di ortogonalità**. Se l'errore di predizione è minimo, deve essere incorrelato con i dati utilizzati per calcolarlo; infatti, se correlazione vi fosse, si potrebbero utilizzare i dati ancora meglio per ridurre l'errore ancora di più, contraddicendo l'ipotesi che l'errore fosse minimo. Inoltre, le equazioni (3.4) non sono altro che le equazioni di Yule Walker.

$$E[\varepsilon_{i,N}x_{i-k}^*] = r_k + \sum_{n=1}^N a_{n,N} r_{k-n} = 0; \quad k = 1, \dots, N. \quad (3.5)$$

Indichiamo ora con  $E_N(z)$  la trasformata  $z$  dell'errore di predizione  $\{\varepsilon_{i,N}\}$ , con  $X(z)$  la trasformata  $z$  della sequenza  $\{x_i\}$ , e con  $A_N(z)$  il filtro causale tutti zeri lungo  $N+1$  campioni (e che ha quindi una trasformata  $z$  costituita da un polinomio di  $N$  grado nella variabile  $z^{-1}$ ) corrispondente alla funzione di trasferimento tra la sequenza  $\{x_i\}$  e l'errore di predizione  $\{\varepsilon_{i,N}\}$  abbiamo:

$$E_N(z) = A_N(z)X(z) \quad (3.6)$$

$$A_N(z) = 1 + \sum_{n=1}^N a_{n,N} z^{-n} \quad (3.7)$$

### 3.1 L'ERRORE DI PREDIZIONE

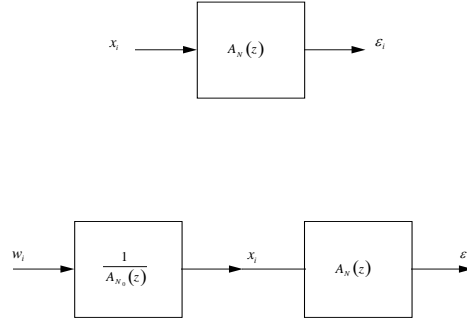


Figura 3.2: Progressivo sbiancamento di uno segnale con spettro tutti poli

Si osservi che  $a_{0,N}$  è unitario;  $A_N(z)$  è quindi un polinomio **monico**. Per la coincidenza delle equazioni di Yule - Walker, il polinomio  $A_N(z)$  è quello stesso che abbiamo trovato a denominatore della funzione di trasferimento tutti poli che sagoma lo spettro bianco di un rumore nella stima spettrale autoregressiva rendendolo al massimo simile allo spettro della sequenza  $\{x_i\}$ ,  $E[X(\omega)X^*(\omega)]$ . In quel caso, si colorava uno spettro bianco; in questo caso invece si sbianca lo spettro della sequenza  $\{x_i\}$ . Quindi, al crescere dell'ordine  $N$  dello stimatore, lo spettro di  $E_N(z)$  tende ad essere tanto più uniforme in frequenza (bianco) perchè la stima spettrale autoregressiva diventa sempre più accurata.

### 3.1 L'errore di predizione

Indichiamo con:

$$\sigma_{\varepsilon,N}^2 = E[(\varepsilon_{i,N})^2] \quad (3.8)$$

Usando ancora le equazioni di Yule Walker e l'incorrelazione tra errore e dati vediamo che

$$E[|\varepsilon_{i,N}|^2] = E[\varepsilon_{i,N}x_i^*] = r_0 + \sum_{n=1}^N a_{n,N}r_n$$



## CAPITOLO 3 LA PREDIZIONE LINEARE

Vista la stazionarietà delle statistiche, si tralascerà l'indice  $i$  nelle medie. Se il predittore è di ordine 0, l'errore di predizione coincide con il segnale da predire, e il suo valore quadratico medio coincide con il valore  $\sigma_x^2 = E[x_i^2]$ . Al crescere dell'ordine del predittore  $N$ , il valore quadratico medio dell'errore di predizione deve decrescere, o quantomeno non crescere. Infatti, si usa sempre più informazione per fare la predizione, che deve quindi migliorare progressivamente.

$$\sigma_x^2 \geq \sigma_{\varepsilon, N}^2 \geq \sigma_{\varepsilon, N+1}^2 \quad (3.9)$$

Questa proprietà sarà poi confermata dall'equazione (3.18). Peraltro, il valore quadratico medio dell'errore di predizione non potrà diventare nullo a meno che non sia possibile, per le caratteristiche statistiche della sequenza  $\{x_i\}$  osservata, prevederne i valori futuri in modo esatto. Comunque, il miglioramento deve progressivamente rallentare e quindi i dati del passato sempre più remoto saranno sempre meno utili per migliorare la predizione. Se sono poco utili, sono incorrelati con il valore attuale dell'errore di predizione. Infatti, si è già osservato che la sequenza degli errori di predizione tende ad essere sempre più bianca, al crescere di  $N$ .

Questo concetto, affermato in modo intuitivo ora, verrà poi dimostrato analiticamente poi.

### 3.2 Predizione lineare e analisi AR

L'analisi spettrale autoregressiva ci ha fornito un metodo per calcolare un filtro tutti poli,  $1/A_N(z)$  che, applicato ad un rumore bianco, ne sagoma lo spettro in modo da riprodurre lo spettro della sequenza osservata. La predizione lineare ci fornisce un filtro tutti zeri  $A_N(z)$  che tende a sbiancare la sequenza  $\{x_n\}$  osservata, sempre meglio al crescere di  $N$ . Questi filtri sono entrambi derivati con le stesse equazioni e dagli stessi parametri e cioè i primi  $N$  valori della funzione di autocorrelazione della sequenza  $\{x_n\}$ ,  $r_0, \dots, r_N$  ( $r_0$ , in sostanza, funziona da fattore di scala); quindi essi tendono ad essere l'uno il reciproco dell'altro, e lo saranno se dello stesso ordine. Supponiamo infatti di avere un processo bianco, reale,  $\{w_i\}$  all'ingresso di un filtro tutti poli  $1/A_{N_0}(z)$  e ne sia  $\{x_i\}$  l'uscita. Se ora applichiamo alla sequenza  $\{x_i\}$  un predittore lineare di ordine  $N$ ,  $N \leq N_0$  l'errore di predizione  $\varepsilon_{i, N}$  sarà reale ma non perfettamente bianco; lo sarà perfettamente se  $N \geq N_0$ . In quelle condizioni,  $\varepsilon_{i, N_0} = w_i$ .

## 3.2 PREDIZIONE LINEARE E ANALISI AR

Una sequenza  $\{x_n\}$  prodotta *colorando* una sequenza bianca con un filtro tutti poli di ordine  $N_0$ , sarà perfettamente sbiancabile con un predittore di ordine  $N \geq N_0$ .

Se l'ordine del predittore cresce ancora, non succede altro perchè gli ulteriori campioni utilizzati per la predizione sono incorrelati con il segnale da predire e quindi il predittore non cambia ulteriormente. Se invece l'ordine  $N$  del predittore è  $N < N_0$  lo sbiancamento non è perfetto e la sequenza  $\varepsilon_{i,N}$  non avrà uno spettro uniforme. Troveremo che il predittore di ordine  $N$  corrisponde alla struttura riverberante costituita dai primi  $N$  tronchi di linea della struttura effettiva (di ordine  $N_0$ ). Quindi, alla crescita di  $N$ , corrisponde la progressiva identificazione dei successivi coefficienti di riflessione della struttura riverberante, da  $c_1$  a  $c_{N_0}$ .

### 3.2.1 Allungamento del predittore

Vogliamo trovare un criterio per determinare i coefficienti del predittore di ordine  $N+1$  a partire da quelli del predittore di ordine  $N$ . In sostanza, non dobbiamo fare altro che risolvere le equazioni di Yule Walker, per l'ordine  $N+1$ .

Accrescere  $N$  di 1 implica utilizzare anche  $x_{i-N-1}$  per stimare meglio  $x_i$ ; tuttavia, parte dell'informazione contenuta in  $x_{i-N-1}$  è già nota attraverso i campioni utilizzati per la predizione. E' ragionevole depurare  $x_{i-N-1}$  da quanto sappiamo già e utilizzare solo quello che di nuovo (rispetto a prima) è contenuto in questo campione. In questo modo, il processo di miglioramento della predizione risulterà più comprensibile.

Stimiamo quindi  $x_{i-N-1}$  utilizzando proprio quei campioni **successivi**  $x_{i-n}; n = 1, N$  che abbiamo usato per la predizione di  $x_i$ . Questa stima lineare ottima, fatta *all'indietro*, per la stazionarietà del processo, utilizza gli stessi coefficienti di predizione, ma applicati alla sequenza ribaltati nel tempo. La stima ottima **all'indietro** di  $x_{i-N-1}$  è  $\tilde{x}_{i-N-1}$ :

$$\tilde{x}_{i-N-1} = - \sum_{n=1}^N a_{n,N} x_{i-N-1+n} \quad (3.10)$$

L'errore di predizione all'indietro  $\beta_{i,N}$  è:

$$\beta_{i,N} = x_{i-N-1} - \tilde{x}_{i-N-1} \quad (3.11)$$

### CAPITOLO 3 LA PREDIZIONE LINEARE

Ora  $\beta_{i,N}$  è incorrelato con  $x_{i-n}; n = 1, N$  per costruzione ed è quindi l'*innovazione* contribuita dal campione  $x_{i-N-1}$ ; il nuovo stimatore di  $\{x_i\}$  che fa uso di questa innovazione è:

$$\hat{x}_{i,N+1} = \hat{x}_{i,N} - c_{N+1}\beta_{i,N} \quad (3.12)$$

$$\varepsilon_{i,N+1} = x_i - \hat{x}_{i,N} + c_{N+1}\beta_{i,N} = \varepsilon_{i,N} + c_{N+1}\beta_{i,N} \quad (3.13)$$

dove  $c_{N+1}$  è un numero da determinare minimizzando l'errore q.m.; questo numero è stato indicato con  $c_{N+1}$  proprio perchè coincide con l'  $N+1$  coefficiente di riflessione della struttura riverberante corrispondente al filtro che ha prodotto la sequenza  $\{x_i\}$ . Calcoliamo  $c_{N+1}$  applicando il principio di ortogonalità; l'errore di predizione deve quindi essere incorrelato con il nuovo dato  $\beta_{i,N}$ :

$$E[\varepsilon_{i,N+1}\beta_{i,N}] = 0 \quad (3.14)$$

$$c_{N+1} = -\frac{E[\beta_{i,N}\varepsilon_{i,N}]}{E[(\beta_{i,N})^2]} \quad (3.15)$$

Infatti

$$E[(\beta_{i,N})^2] = E[(\varepsilon_{i,N})^2] \quad (3.16)$$

e per la diseguaglianza di Schwartz, segue:

$$|c_{N+1}| \leq 1 \quad (3.17)$$

Inoltre:

$$E[(\varepsilon_{i,N+1})^2] = E[(\varepsilon_{i,N})^2] \times (1 - c_{N+1}^2) = \sigma_x^2 \prod_{n=1}^{N+1} (1 - c_n^2) \leq E[(\varepsilon_{i,N})^2] \quad (3.18)$$

Abbiamo ora la possibilità di calcolare direttamente  $c_{N+1}$  dalla funzione di autocorrelazione; infatti, profittando dell'incorrelazione tra errore di predizione e stima:

$$\begin{aligned} E[(\varepsilon_{i,N})^2] &= E[(x_i - \hat{x}_{i,N})(x_i - \hat{x}_{i,N})] = \\ &= E[(x_i - \hat{x}_{i,N})x_i] = r_0 + \sum_{n=1}^N a_{n,N}r_n \end{aligned} \quad (3.19)$$

### 3.3 LA RICORSIONE DI LEVINSON

Similmente:

$$E[\varepsilon_{i,N}\beta_{i,N}] = r_{N+1} + \sum_{n=1}^N a_{n,N}r_{N+1-n} \quad (3.20)$$

$$c_{N+1} = -\frac{r_{N+1} + \sum_{n=1}^N a_{n,N}r_{N+1-n}}{r_0 + \sum_{n=1}^N a_{n,N}^*r_n} = -\frac{r_{N+1} + \widetilde{\mathbf{r}}_N^* \mathbf{a}_N}{r_0 + \widetilde{\mathbf{r}}_N^* \widetilde{\mathbf{a}}_N} \quad (3.21)$$

dove:

$$\widetilde{\mathbf{r}}_N^* = \begin{bmatrix} r_{-N} \\ \cdot \\ \cdot \\ r_{-2} \\ r_{-1} \end{bmatrix} \quad \mathbf{a}_N = \begin{bmatrix} a_{1,N} \\ a_{2,N} \\ \cdot \\ \cdot \\ a_{N,N} \end{bmatrix} ; \quad \widetilde{\mathbf{a}}_N = \begin{bmatrix} a_{N,N}^* \\ \cdot \\ \cdot \\ a_{2,N}^* \\ a_{1,N}^* \end{bmatrix} ;$$

### 3.3 La ricorsione di Levinson

Come abbiamo visto nel paragrafo precedente, il predittore di ordine  $N + 1$  si deriva da quello di ordine  $N$  utilizzando l'errore di predizione all'indietro e cioè l'innovazione apportata dal campione di posto  $i - N - 1$  rispetto agli altri campioni già utilizzati per la predizione. L'errore di predizione è quindi:

$$\varepsilon_{i,N+1} = x_i - \hat{x}_{i,N} + c_{N+1}\beta_{i,N} = \varepsilon_{i,N} + c_{N+1}\beta_{i,N} \quad (3.22)$$

In termini di trasformata  $z$ , indicando con

$$E_N(z) = A_N(z)X(z) \quad (3.23)$$

$$E_{N+1}(z) = E_N(z) + c_{N+1}B_N(z) \quad (3.24)$$

dove abbiamo indicato con  $B_N(z)$  la trasformata  $z$  dell'errore di predizione all'indietro e cioè:

$$B_N(z) = z^{-N-1}A(z^{-1})X(z)$$

si ricava:

$$A_{N+1}(z) = A_N(z) + c_{N+1}z^{-N-1}A_N(z^{-1}) = A_N(z) + c_{N+1}z^{-1} \widetilde{A}_N(z) \quad (3.25)$$

### CAPITOLO 3 LA PREDIZIONE LINEARE

Abbiamo riottenuto la ricorsione di Levinson. **La progressiva identificazione di  $A_{N_0}(z)$  avviene attraverso la progressiva identificazione dei coefficienti di riflessione della struttura riverberante corrispondente.**

Infatti, per  $N > N_0$  le equazioni di Yule Walker ci garantiscono la corretta analisi spettrale ed un errore di predizione perfettamente bianco, se la sequenza di partenza era l'uscita di una struttura riverberante di ordine  $N_0$  alimentata da una sequenza bianca. Accrescendo progressivamente l'ordine del predittore, a partire da  $N = 1$ , calcoliamo in successione i coefficienti di riflessione, che alla fine risultano essere quelli corretti, perchè lo sbiancatore (la funzione di trasferimento tra la sequenza  $\{x_i\}$  e la sequenza errore di predizione  $\{\varepsilon_i\}$ ) diventa perfetto. Quindi, i coefficienti di riflessione erano stati individuati correttamente sin dal momento in cui erano stati calcolati.

Pertanto, se si utilizzano gli stessi campioni della funzione di autocorrelazione, i coefficienti di riflessione che risultano dalla predizione lineare sono gli stessi della struttura riverberante che ha colorato la sequenza bianca  $\{w_i\}$  iniziale. Ovviamente, se invece il processo nasce da meccanismo completamente diverso, è possibile che questi coefficienti non abbiano gran significato; tuttavia, l'esempio delle Direzioni di Arrivo che verrà proposto in un capitolo successivo dimostra che questa tecnica di stima spettrale ha una validità che si estende a numerosissimi campi applicativi.

Per avere una motivazione intuitiva di questo importante risultato, si pensi ad una struttura riverberante disadattata da un lato, eccitata da una sequenza bianca dalla parte disadattata. Si supponga poi che il sistema sia eccitato da una successione quasi continua di impulsi indipendenti, e che un coefficiente di riflessione  $c_1$  corrisponda alla riflessione di una frazione  $c_1$  di questi impulsi. E' comprensibile che il primo campione della funzione di autocorrelazione dell'uscita dipenda solo dal primo coefficiente di riflessione  $c_1$ ; una frazione  $c_1$  dei campioni del segnale in ingresso si riflettono e riemergono all'uscita dopo un ritardo pari a  $T$  sovrapponendosi al campione successivo (o alla sua parte che è stata trasmessa direttamente attraverso la prima interfaccia). Nasce così la correlazione a ritardo  $T$ ,  $r_1$ . Vediamo quale frazione del segnale viene ritardata di  $2T$  e cioè contribuisce a formare  $r_2$ : contribuiscono:

- i campioni dell'ingresso che hanno subito una riflessione alla seconda interfaccia (una frazione pari a  $c_2$ );

### 3.4 IL FILTRO A TRALICCIO NON RICORSIVO

- quella frazione dei campioni che dopo avere fatto una prima volta il percorso ingresso - prima interfaccia, lo ripete. Questa frazione è pari a  $c_1^2$ .

Questa frazione di segnale è però correlata anche con il segnale che abbiamo trovato dopo solo un intervallo  $T$ , e quindi aveva fatto solo una volta il percorso  $c_1, c_0$ ; se pertanto cerchiamo di rimuovere questa componente cercando quanto è decorrelato con il campione precedente, la componente di correlazione residua (il coefficiente di correlazione parziale o PARCOR come viene chiamato in alcune trattazioni) è legato solo al coefficiente di riflessione  $c_2$  e così via.

In altre parole, la correlazione tra gli errori di predizione in avanti ed all'indietro è legata solo alla frazione di segnale che ha percorso il tratto fino alla  $N$ -esima interfaccia senza subire riflessioni intermedie ma venendo solo trasmesso; quindi la sua ampiezza è proporzionale al solo coefficiente di riflessione  $c_N$ . Un altro risultato importante è che, per la ricorsione di Levinson, ripetendo la dimostrazione fatta al capitolo 10, i polinomi  $A_N(z)$  sono a fase minima, se i coefficienti di riflessione sono minori di 1, cosa peraltro dimostrata nella sezione 12.2.

#### 3.3.1 Vantaggi computazionali

La ricorsione di Levinson comporta notevoli vantaggi computazionali nella soluzione delle equazioni di Yule - Walker: infatti la soluzione di un sistema di  $N$  equazioni in  $N$  incognite comporta un numero di moltiplicazioni proporzionale a  $N^3$ ; invece, se si utilizza la ricorsione il costo diventa proporzionale a  $N^2$ . Infatti, allungando il predittore dobbiamo calcolare l' $N + 1$ -esimo coefficiente di riflessione utilizzando la formula (3.15); questo comporta  $2N$  moltiplicazioni ed una divisione; inoltre, dobbiamo calcolare i nuovi  $\{a_{h,N+1}\}$  e questo comporta altre  $N$  moltiplicazioni; quindi, circa  $3N$  moltiplicazioni per ogni passo e pertanto circa un totale di  $1.5 \times N^2$  moltiplicazioni invece di una proporzionalità ad  $N^3$ , con indubbi vantaggi.

### 3.4 Il filtro a traliccio non ricorsivo

Possiamo ora costruire un sistema numerico per il calcolo degli errori di predizione in avanti ed all'indietro, basato appunto sulla ricorsione di

### CAPITOLO 3 LA PREDIZIONE LINEARE

Levinson. Questo sistema sarà poi utilizzabile come uno sbiancatore, a minima fase, di sequenze colorate.

Supponiamo di disporre di  $\varepsilon_{i,N}$  e  $\beta_{i,N}$ ; possiamo ottenere, tenendo conto del ritardo crescente del campione stimato all'indietro, senza contraddizione con la (3.22):

$$\varepsilon_{i,N+1} = \varepsilon_{i,N} + c_{N+1}\beta_{i,N} \quad (3.26)$$

$$\beta_{i,N+1} = \beta_{i-1,N} + c_{N+1}\varepsilon_{i-1,N} \quad (3.27)$$

Ne nasce la struttura numerica di figura 12 .3 nota come **filtro a traliccio non ricorsivo**; è evidente la somiglianza con il filtro a traliccio ricorsivo corrispondente ad una struttura riverberante: le differenze sono nei segni dei collegamenti incrociati e nei sensi della propagazione del segnale nei due rami. Il filtro a traliccio non ricorsivo serve non solo da

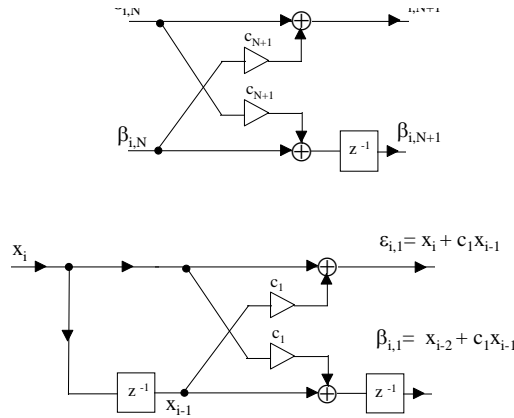


Figura 3.3: Filtro a traliccio non ricorsivo: cella intermedia e cella iniziale

sbiancatore della sequenza, fornendo in uscita  $\varepsilon_{i,N}$ , ma anche da predittore  $\{\hat{x}_{i,N}\}$ ; la predizione è la differenza tra ingresso ed uscita del filtro

### 3.4 IL FILTRO A TRALICCIO NON RICORSIVO

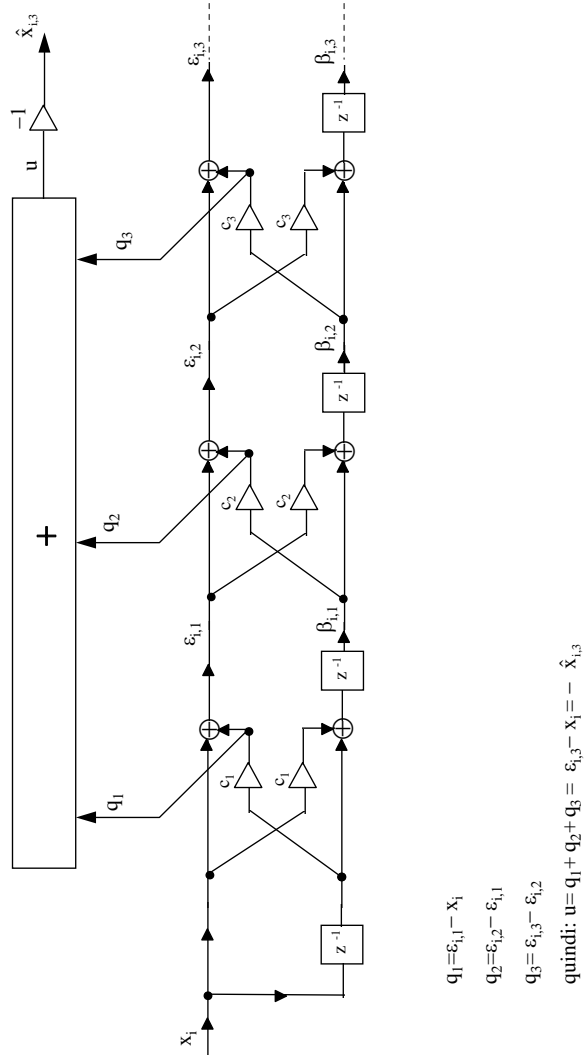


Figura 3.4: Filtro a traliccio non ricorsivo: uso come predittore



## CAPITOLO 3 LA PREDIZIONE LINEARE

ed è funzione solo dei dati passati; essa è disponibile **prima** dell'arrivo del prossimo campione della sequenza e quindi prima che sia disponibile  $\varepsilon_{i,N}$ .

I coefficienti di riflessione possono essere derivati dai valori della funzione di autocorrelazione, come già visto; però possono anche essere ottenuti dalla correlazione mutua tra gli errori di predizione in avanti e all'indietro, normalizzati. In questo modo il filtro a traliccio ricorsivo assume la forma di figura 12.5: l'elemento interessante di questa configurazione è che il sistema può essere reso adattativo, stimando la correlazione tra  $\varepsilon_{i,N}, \beta_{i,N}$  con stimatori che p.e. utilizzano solo gli ultimi  $M$  campioni, **obliando** quelli passati, con dei criteri di progressività che vedremo nel capitolo sui sistemi adattativi.

Il filtro a traliccio nella configurazione di figura 12.5 è perciò un sistema autonomo, che può calcolare, al variare del tempo e delle statistiche dell'ingresso, i coefficienti di riflessione e pertanto lo spettro dell'ingresso, la predizione ottima ed infine l'errore di predizione e cioè la sequenza sbiancata.

### 3.5 Descrizioni alternative dell'autocorrelazione

Nei paragrafi precedenti abbiamo visto come i primi  $N$  valori di una funzione di autocorrelazione  $r_0 \dots r_N$  di una qualsiasi sequenza colorata  $\{x_n\}$  siano esprimibili in funzione dei coefficienti del polinomio  $A_N(z)$  per la predizione  $a_{n,N}$ ,  $n = 1, \dots, N$ :

$$A_N(z) = 1 + \sum_{n=1}^N a_{n,N} z^{-n}$$

nonchè al valore quadratico medio dell'errore di predizione residuo  $E[(\varepsilon_{n,N}^2)]$ ; ancora, sono esprimibili in funzione degli  $N$  coefficienti di riflessione  $c_i$   $i = 1, \dots, N$ , ancora insieme a  $E[(\varepsilon_{n,N}^2)]$ . Un terzo modo di descrivere lo spettro è dato dalle posizioni degli zeri di  $A_N(z)$  ed ancora altri modi saranno visti nel capitolo 15, con le LSP (Line Spectral Pairs). Di tutte queste descrizioni alternative di una funzione di autocorrelazione (o di uno spettro) quella dei coefficienti di riflessione è la più vicina ad una descrizione fisica. Difatti, basta assicurarsi che  $|c_i| < 1$ ,  $i = 1, \dots, N$  perchè

### 3.5 DESCRIZIONI ALTERNATIVE DELL'AUTOCORRELAZIONE

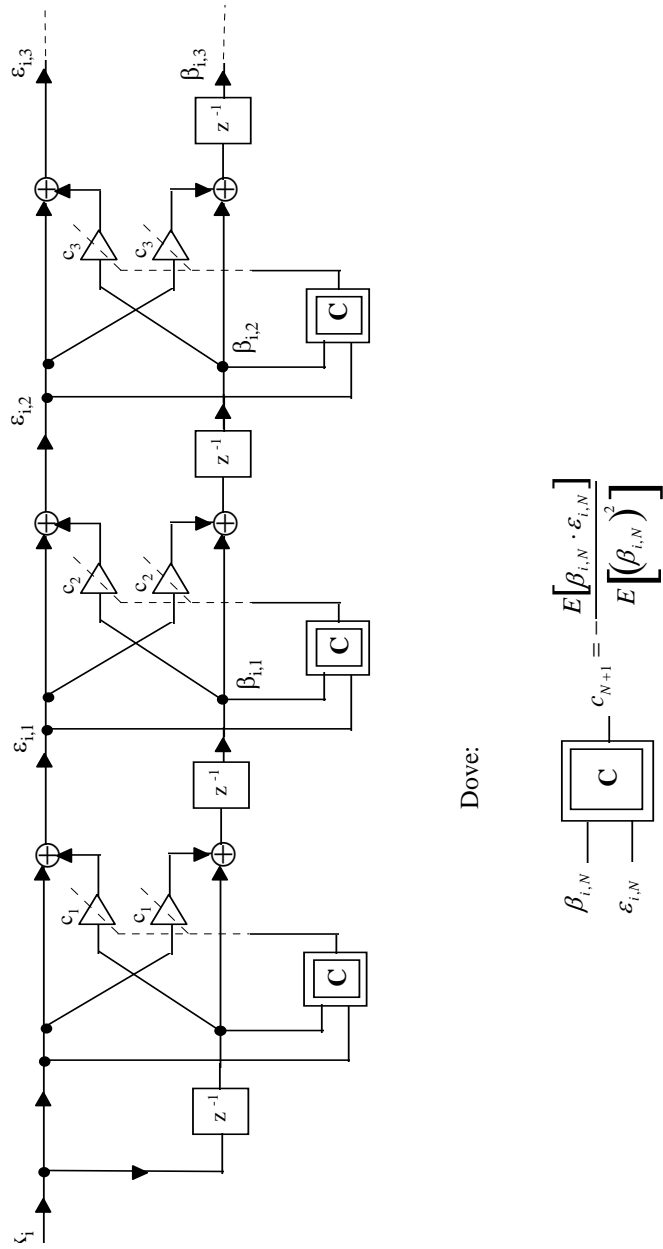


Figura 3.5: Filtro a traliccio adattativo

## CAPITOLO 3 LA PREDIZIONE LINEARE

automaticamente risulti che la matrice di covarianza  $\mathbf{R}$  sia semi definita positiva e cioè che per tutti i suoi autovalori risulti  $\lambda_i \geq 0$ . Se poi è comunque sempre possibile passare da un modo di descrizione ad un altro, è anche vero che i coefficienti di riflessione sono meno sensibili ad effetti quantizzazione rispetto agli  $a_{n,N}$ , come visto nel capitolo 7. Infine, il difetto della descrizione mediante gli  $r_i$  è la difficile verifica della positività della matrice di covarianza  $\mathbf{R}$ , e quindi le derivanti difficoltà nel troncamento di una serie di valori dell'autocorrelazione (la finestatura vista nel capitolo 10) che può portare a valori negativi dello spettro e quindi a qualche  $\lambda_i < 0$ ; abbiamo già osservato che la soluzione a questo problema, nel caso della stima spettrale AR, è l'estrapolazione della sequenza delle autocorrelazioni con le equazioni di Yule Walker, cosa che poi consente di avere uno spettro positivo. Tutte queste difficoltà sono risolte dall'uso dei coefficienti di riflessione; infatti, con il metodo di stima spettrale AR di Burg, molto usato e che sarà visto nel prossimo paragrafo, la stima spettrale avviene con la stima diretta dei  $c_i$ .

## 3.6 Ordine del predittore

### 3.6.1 Risoluzione spettrale

Se i dati a disposizione hanno lunghezza infinita e vi è stazionarietà statistica, l'analisi spettrale non pone problemi: c'è modo di determinare la funzione di autocorrelazione con la qualità desiderata. Il periodogramma potrà dare risultati con la risoluzione spettrale e la dispersione statistica volute, senza dovere fare compromessi tra l'una e l'altra. Il problema dell'analisi spettrale è quindi **sempre e solo** collegato ad una limitata disponibilità di dati. E' interessante valutarne gli effetti nel caso dell'analisi spettrale parametrica. L'affidabilità del modello da estrarre dai dati diminuisce, se ne aumenta la complessità. Al crescere dell'ordine  $N$  del predittore (o dello stimatore) tende a peggiorare il condizionamento della matrice  $\mathbf{R}$ ; inoltre, si dovranno stimare un numero crescente di campioni dell'autocorrelazione dei dati e l'effetto degli errori di questa stima sarà anch'esso crescente. Similmente, nel caso del periodogramma, l'affidabilità dei risultati decresce con il crescere del numero di parametri da stimare; il numero di parametri qui coincide con il numero delle celle

### 3.6 ORDINE DEL PREDITTORE

in cui è suddiviso l'asse delle frequenze e pertanto vi è una relazione diretta tra affidabilità della stima e risoluzione spettrale.

In questo, la stima AR ha una grande differenza con il periodogramma, come si vedrà nel caso delle direzioni di arrivo (DOA): **la risoluzione spettrale AR non dipende sostanzialmente dalla lunghezza del predittore**, ma piuttosto dalla vicinanza al cerchio unitario degli zeri del polinomio  $A_N(z)$ . Questa vicinanza dipende sia dall'entità del rumore che accompagna i dati sia dalla larghezza di banda del segnale stimato; inoltre la dispersione della stima dei campioni dell'autocorrelazione utilizzati per costruire  $\mathbf{R}$  porta alla dispersione delle posizioni degli zeri di  $A_N(z)$  e quindi ad errori di stima delle posizioni dei poli del modello AR ma non necessariamente a picchi spettrali meno pronunciati.

Nel caso di un segnale sinusoidale di frequenza ignota (*il caso dell'orologiaio*), se il rapporto segnale - disturbo è buono, se il segnale corrisponde ad una sola sinusoide, e se inoltre la stima della funzione di correlazione è buona, basterebbero anche 2 soli campioni (ma ben stimati) della funzione di autocorrelazione, per avere una risoluzione spettrale al limite infinita. Qui la differenza con la stima spettrale fatta con il periodogramma è veramente molto grande. Infatti, c'è l'ipotesi essenziale che vi sia una sola sinusoide: la stima spettrale si riduce alla stima della frequenza di quella sinusoide. Se il numero di sinusoidi fosse  $m > 1$ , avremmo invece bisogno almeno di  $m + 1$  campioni della funzione di autocorrelazione per stimare le  $m$  frequenze ed il valore q.m. del segnale. Infine, osserviamo ancora che questo è il tipo di analisi spettrale che potrebbe essere fatto per misurare la frequenza di un singolo oscillatore: l'unica variabile da stimare è la frequenza. Poi, si possono usare le equazioni di Yule Walker per estrapolare la funzione di autocorrelazione su tutto l'asse dei tempi completando l'analisi.

#### 3.6.2 Lunghezza dei dati e criterio di Akaike

Se sono disponibili  $M$  campioni dei dati, l'ordine  $N$  del predittore sarà necessariamente  $N < M$ . Volendo, però,  $N$  può essere fatto crescere fino ad  $M - 1$ , senza rischi di instabilità, utilizzando la via della stima diretta dei  $c_i$  piuttosto che degli  $r_i$ . Un metodo per accrescere senza rischi l'ordine del predittore è il metodo di Burg (o della massima entropia). Invece di stimare i campioni della funzione di autocorrelazione, si stimano appunto direttamente i coefficienti di riflessione e per garantire che siano

### CAPITOLO 3 LA PREDIZIONE LINEARE

sempre minori di 1, si usa la formula:

$$c_{h+1} = -2 \frac{E[\varepsilon_{i,h} \beta_{i,h}]}{E[\varepsilon_{i,h}^2 + \beta_{i,h}^2]} \quad (3.28)$$

dove, non potendo realizzare l'operatore di valore atteso,  $E[.]$ , lo si sostituisce con la media su tutti i dati:

$$c_{h+1} = -\frac{\sum_{m=0}^M \varepsilon_{m,h} \beta_{m,h}}{\sum_{m=0}^M \frac{(\varepsilon_{m,h}^2 + \beta_{m,h}^2)}{2}} \quad (3.29)$$

Poi si calcolano gli errori di predizione di ordine  $h+1$  e così via, eventualmente fino ad avere  $h = M-1$ ; allora, ovviamente l'errore di predizione diventa eguale a zero. Il metodo di Burg (ma sarebbe la stessa cosa per la corrispondente stima spettrale AR) è indicato come metodo a massima entropia perchè si verifica che:

- assegnati i primi  $N$  valori della funzione di autocorrelazione;
- determinando tutti gli altri valori della funzione di autocorrelazione mediante le equazioni di Yule Walker;
- trasformando con Fourier la funzione di autocorrelazione e trovando la densità spettrale;
- calcolando l'entropia di questa densità spettrale (il valore medio del logaritmo);
- **si verifica che tale entropia è massima** tra tutte le densità spettrali (tutte positive) caratterizzate dall'avere i primi  $N$  valori dell'autocorrelazione eguali a quelli assegnati.

**In altre parole, nella stima AR lo spettro è il più frastagliato possibile, compatibilmente con i valori assegnati della funzione di autocorrelazione. Al contrario, nella stima MA, lo spettro è il più liscio possibile (la funzione di autocorrelazione viene troncata appena possibile, compatibilmente con i campioni dell'autocorrelazione misurati e con l'avere un spettro tutto positivo).**

Tuttavia, pur disponendo così di un metodo che garantisce di poter avere predittori di ordine elevatissimo, è sbagliato accrescere troppo  $N$  rispetto ad  $M$ ; i coefficienti di riflessione trovati sono solo una pessima stima di quelli *veri*. Con un'altra realizzazione del processo, sarebbero tutt'altri. Esiste un criterio per limitare l'ordine del predittore, noto come

### 3.7 CONDIZIONAMENTO DI $\underline{\mathbf{R}}$

criterio di Akaike: si studia l'andamento del cosiddetto Final Prediction Error:

$$\overline{\varepsilon_F^2} = \frac{M + N - 1}{M - N - 1} \sum_{m=0}^M \varepsilon_{m,N}^2 \quad (3.30)$$

e si cerca il valore di  $N$  per cui lo FPE è minimo; se  $N \ll M$  lo FPE decrescerà con  $N$ , come consueto; poi, se il numero di parametri del modello diventa esagerato, e  $N$  si avvicina troppo ad  $M$ , lo FPE tende a crescere di nuovo. Spesso è prudente mantenere  $N$  alquanto inferiore a quanto proposto da Akaike.

### 3.7 Condizionamento di $\underline{\mathbf{R}}$

Abbiamo visto come il predittore  $\underline{\mathbf{a}}_N$  venga determinato risolvendo, eventualmente con la ricorsione di Levinson, il sistema:

$$\underline{\mathbf{R}}\underline{\mathbf{a}}_N = -\underline{\mathbf{r}}_N \quad (3.31)$$

Abbiamo anche visto che, se gli elementi della matrice  $\underline{\mathbf{R}}$  derivano dai campioni di una funzione di autocorrelazione, allora il polinomio

$$A_N(z) = 1 + \sum_{n=1}^N a_{n,N} z^{-n} \quad (3.32)$$

è a fase minima; infatti, è denominatore della funzione di trasferimento corrispondente alla struttura riverberante che ha sagomato spettralmente il processo, troncata all'ennesimo coefficiente di riflessione. Tuttavia  $\underline{\mathbf{R}}$  può essere singolare, anzi in generale dovrebbe proprio esserlo. Infatti, se la sequenza  $\{x_n\}$  che osserviamo è proveniente da un campionamento senza equivocazione di un segnale passa basso, il contenuto spettrale alle frequenze vicine alla frequenza di Nyquist dovrà essere praticamente nullo; ora poichè l'errore di predizione

$$\varepsilon_{i,N} = x_i + \sum_{n=1}^N a_{n,N} x_{i-n} \quad (3.33)$$

ha spettro  $E[|E(z)|^2]$  bianco, se  $N$  è abbastanza grande, poichè

$$E[|E(z)|^2] = E[|X(z)|^2] |A_N(z)|^2 \quad (3.34)$$

## CAPITOLO 3 LA PREDIZIONE LINEARE

allora  $|A_N(z)|^2$  è molto grande, almeno vicino alla frequenza di Nyquist, dove  $E[|X(z)|^2]$  è molto piccolo. Ciò implica che la matrice  $\underline{\mathbf{R}}$  è prossima alla singolarità perchè il vettore che è soluzione del sistema ha un modulo grande, malgrado che sia termini noti che coefficienti siano dello stesso ordine di grandezza.

**Pertanto, non possiamo risolvere le equazioni di Y-W senza precauzioni.**

Da un altro punto di vista, gli autovalori  $\lambda_i$  della matrice  $\underline{\mathbf{R}}$ , se circolante, corrispondono alle energie delle componenti della Trasformata Discreta di Fourier della sequenza. Anche se la matrice  $\underline{\mathbf{R}}$  non è circolante, ci attendiamo che gli autovalori siano collegati alle ampiezze dello spettro di potenza della sequenza  $\{x_n\}$ . Se la sequenza è passa basso, alcuni dei valori spettrali saranno pressochè nulli, e così saranno alcuni tra gli autovalori. Ancora, ci attendiamo che la  $\underline{\mathbf{R}}$  sia singolare o almeno mal condizionata.

Per risolvere il problema del cattivo condizionamento, ci comporteremo come se vi fosse un disturbo sovrapposto alla sequenza osservata  $\{x_i\}$ , come vedremo nel prossimo paragrafo.

### 3.8 Stima con rumore

Supponiamo che la sequenza osservata  $\{x_i\}$  sia disturbata da un rumore bianco  $\{z_i\}$ ; supponiamo inoltre che  $\{x_i\}$  sia una sequenza colorata; ad esempio potrebbe essere l'uscita di una struttura riverberante con funzione di trasferimento  $1/A(z)$  alimentata da una sequenza bianca  $\{w_i\}$ ; per un'idea intuitiva, si pensi a  $\{w_i\}$  come al suono generato da un *coro di voci bianche* (tanto per ricordare che si tratta di una sequenza bianca e gaussiana) in una *cattedrale* che causa delle riverberazioni corrispondenti alla funzione di trasferimento  $1/A(z)$ ; (peraltro, una cattedrale vera corrisponde non ad una sola struttura riverberante ma al parallelo di tante di queste strutture e pertanto la funzione di trasferimento che si potrebbe misurare avrebbe non solo poli ma anche zeri; tuttavia, per proseguire l'esempio, supponiamo che la funzione di trasferimento sia costituita solo da poli). *Un ascoltatore rileva  $\{y_i\}$ , e cioè ascolta il coro  $\{w_i\}$  riverberato attraverso la cattedrale  $1/A(z)$  ma è disturbato dal brusio dei vicini  $\{z_i\}$* ; si pongono due problemi nuovi:

- (i) ricostruire  $\{x_i\}$  usando  $N$  campioni del segnale riverberato e dis-

### 3.8 STIMA CON RUMORE

turbato (la sequenza  $\{y_i\}$ ), attenuando l'effetto del disturbo  $\{z_i\}$ ; si vuole cioè fare una stima ottima  $\{\hat{x}_{i,N}\}$  della sequenza  $\{x_i\}$  combinando linearmente  $N$  valori della sequenza osservata  $\{y_i\}$ ; per mantenere l'omogeneità con la trattazione precedente in cui si considera la predizione, supponiamo per ora che i campioni utilizzati per la stima siano tutti precedenti il campione  $y_i$  e siano cioè  $y_{i-1}, \dots, y_{i-N}$ . Si potrebbe tuttavia generalizzare il problema, utilizzando per la stima anche il campione  $y_i$  o addirittura i campioni successivi trasformando il problema da quello di stima predittiva a quello di un'interpolazione.

(ii) stimare direttamente il segnale di ingresso  $\{w_i\}$  (*il coro*) con uno stimatore lungo  $N$  campioni, e cioè determinare  $\{\hat{w}_{i,N}\}$ .

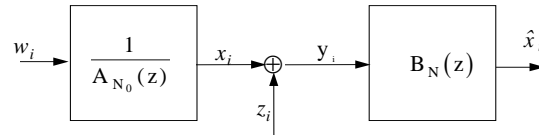


Figura 3.6: Somma di rumore bianco ad un processo colorato

Passando al calcolo, indichiamo con  $\{y_i\}$  i campioni riverberati  $\{x_i\}$  disturbati dal rumore additivo  $\{z_i\}$ :

$$y_i = x_i + z_i \quad (3.35)$$

Vogliamo stimare la sequenza  $\{x_i\}$ , combinando tra loro linearmente  $N$  campioni della sequenza  $\{y_i\}$ : questo problema è immediatamente riconducibile a quanto fatto finora, minimizzando l'errore di stima quadratico medio. Si applica quindi il principio di ortogonalità tra errore di stima (piuttosto che l'errore di predizione, ma in sostanza è la stessa cosa; cambia solo la posizione dei campioni assegnati rispetto a quella del campione da stimare) e dati; sia  $\{\hat{x}_{i,N}\}$  il segnale stimato, funzione lineare dei vari campioni della sequenza  $\{y_i\}$ :

$$\hat{x}_{i,N} = - \sum_{n=1}^N b_{n,N} y_{i-n} \quad (3.36)$$

L'errore di stima è:

$$\gamma_{i,N} = x_i - \hat{x}_{i,N} \quad (3.37)$$



### CAPITOLO 3 LA PREDIZIONE LINEARE

e se ne impone l'incorrelazione con i dati, per garantirne l'ottimalità:

$$E[\gamma_{i,N} y_{i-k}] = 0; \quad k = 1, \dots, N \quad (3.38)$$

È opportuno osservare che l'errore di stima  $\{\gamma_{i,N}\}$  non è disponibile, poichè la sequenza  $\{x_i\}$  non è nota; incidentalmente, si ricordi che l'errore di predizione  $\{\varepsilon_{i,N}\}$  della sequenza  $\{y_i\}$  è invece sempre disponibile, anche se con il ritardo di un campione. Utilizzando l'incorrelazione tra  $\{x_i\}$  e  $\{z_i\}$ , lo stimatore

$$\underline{\mathbf{b}}_N = \{b_{i,N}\}; \quad i = 1, \dots, N \quad (3.39)$$

è dato dal sistema:

$$\begin{bmatrix} r_0 + \sigma_z^2 & r_{-1} & r_{-2} & \dots & r_{-N+1} \\ r_1 & r_0 + \sigma_z^2 & r_{-1} & \dots & r_{-N+2} \\ r_2 & r_1 & r_0 + \sigma_z^2 & \dots & r_{-N+3} \\ \dots & \dots & \dots & \dots & \dots \\ r_{N-1} & r_{N-2} & r_{N-3} & \dots & r_0 + \sigma_z^2 \end{bmatrix} \begin{bmatrix} b_{1,N} \\ b_{2,N} \\ b_{3,N} \\ \dots \\ b_{N,N} \end{bmatrix} = - \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \dots \\ r_N \end{bmatrix} \quad (3.40)$$

e cioè

$$(\underline{\mathbf{R}} + \sigma_z^2 \underline{\mathbf{I}}) \underline{\mathbf{b}}_N = -\underline{\mathbf{r}}_N \quad (3.41)$$

Quindi, se al segnale da stimare è sommato del rumore, alla matrice  $\underline{\mathbf{R}}$  si somma la matrice identica, moltiplicata per la varianza del rumore  $\sigma_z^2$ . Gli autovalori della matrice  $\underline{\mathbf{R}} + \sigma_z^2 \underline{\mathbf{I}}$  sono  $\lambda_i + \sigma_z^2$ ; se la matrice  $\underline{\mathbf{R}}$  è una matrice di covarianza,  $\lambda_i \geq 0$ ; quindi  $\lambda_i + \sigma_z^2 > 0$ .

In conclusione, il problema (i) della stima di una sequenza  $\{x_i\}$ , disturbata da un rumore additivo  $\{z_i\}$ , effettuata con uno stimatore di lunghezza  $N$ , (la ricerca di  $\{\hat{x}_{i,N}\}$ ) è sempre ben condizionato; infatti, non si devono effettuare amplificazioni illimitate perchè porterebbero a segnali troppo influenzati dal disturbo. Si osservi che per risolvere il problema (i) basta conoscere la matrice di covarianza  $\underline{\mathbf{R}}$  del segnale osservato  $\{y_i\}$ , il valore quadratico medio del rumore aggiunto, ed il vettore  $\underline{\mathbf{r}}_N$ ; inoltre, non è necessario che la stima sia riferita al campione attuale; in altre parole, si può decidere di stimare un campione precedente della sequenza  $\{x_i\}$ ; ciò altera il vettore dei termini noti  $\underline{\mathbf{r}}_N$  senza alterare la

### 3.8 STIMA CON RUMORE

matrice  $\underline{\mathbf{R}}$ . In quel caso, la ricorsione di Levinson non sarebbe più applicabile per la determinazione della soluzione, ma l'errore di stima sarebbe più piccolo, perchè utilizzeremmo per fare la stima dati utili anche se successivi al tempo di generazione del segnale.

Possiamo ora risolvere il problema (ii) e da quella via dare un ulteriore significato anche all'errore di predizione; stimiamo con uno stimatore lungo  $N$  campioni, la sequenza  $\{w_i\}$  che alimenta il filtro tutti poli che genera la sequenza  $\{x_i\}$ ; (*ricostruiamo il coro di voci bianche, senza la riverberazione della cattedrale e attenuando quanto possibile il brusio dei vicini*). Supponiamo che la stima questa volta venga fatta utilizzando anche il campione corrente  $y_i$ , e quindi la sommatoria corrispondente allo stimatore inizi da  $n = 0$  e non da  $n = 1$ . I coefficienti dello stimatore sono sempre ottenuti mediante il principio di ortogonalità tra errore di stima e dati, per minimizzare l'errore quadratico medio.

$$y_i = x_i + z_i; \quad x_i = - \sum_{n=1}^{N_0} a_{n,N_0} x_{i-n} + w_i \quad (3.42)$$

$$\hat{w}_{i,N} = - \sum_{n=0}^N c_{n,N} y_{i-n}; \quad (3.43)$$

$$E[(w_i + \sum_{h=0}^N c_{h,N} y_{i-h}) y_{i-k}] = 0; \quad 0 \leq k \leq N \quad (3.44)$$

Mentre la matrice del sistema è sempre la matrice di autocorrelazione della variabile  $\{y_i\}$  e cioè  $\underline{\mathbf{R}} + \sigma_z^2 \underline{\mathbf{I}}$ , il vettore dei termini noti è  $\underline{\mathbf{r}}'$  con componenti  $r'_k = -E[w_i y_{i-k}]$  e pertanto ha tutte le componenti nulle per la bianchezza della sequenza  $\{w_i\}$  tranne la prima eguale a  $\sigma_w^2$ .

$$\begin{bmatrix} r_0 + \sigma_z^2 & r_{-1} & r_{-2} & \dots & r_{-N} \\ r_1 & r_0 + \sigma_z^2 & r_{-1} & \dots & r_{-N+1} \\ r_2 & r_1 & r_0 + \sigma_z^2 & \dots & r_{-N+2} \\ \dots & \dots & \dots & \dots & \dots \\ r_N & r_{N-1} & r_{N-2} & \dots & r_0 + \sigma_z^2 \end{bmatrix} \begin{bmatrix} c_{0,N} \\ c_{1,N} \\ c_{2,N} \\ \dots \\ c_{N,N} \end{bmatrix} = - \begin{bmatrix} \sigma_w^2 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (3.45)$$

Vi sono due casi possibili.

### 3.8.1 Il caso $\sigma_z^2 = 0$

Si ritrovano esattamente le equazioni di Yule - Walker che, nel caso in cui il termine noto sia eguale a  $\sigma_{\varepsilon}^2$ , o a meno di un fattore di scala, danno come soluzione la sequenza  $1, a_{1,N}, \dots, a_{N,N}$  la cui trasformata  $z$  è il solito polinomio  $A_N(z)$ . Pertanto, in questo caso, la sequenza stimata  $\{\hat{w}_{i,N}\}$  coincide con  $\{\varepsilon_{i,N}\}$  errore di predizione della sequenza  $\{y_i\}$ ; ne deduciamo che  $\{\varepsilon_{i,N}\}$  (a meno di un possibile fattore di scala pari a  $\sigma_{\varepsilon}^2/\sigma_w^2$ ) è una stima della sequenza di ingresso della struttura tutti poli che ha generato la sequenza  $\{x_i\}$ ; si comprenderà ancora meglio perchè  $\{\varepsilon_{i,N} = \hat{w}_{i,N}\}$  tenda ad essere una sequenza bianca, non appena lo stimatore abbia una lunghezza sufficiente. È infatti la stima di una sequenza bianca. Inoltre, si può ancora applicare la ricorsione di Levinson.

### 3.8.2 Il caso $\sigma_z^2 \neq 0$

Se  $\sigma_z^2 \neq 0$ , allora, anche per  $N \rightarrow \infty$  la sequenza  $\{\hat{w}_{i,N}\}$  non diventa bianca a differenza della sequenza  $\{\varepsilon_{i,N}\}$ . Quindi, per  $\sigma_z^2 \neq 0$ ,  $\{\varepsilon_{i,N} \neq \hat{w}_{i,N}\}$ ; in particolare la sequenza  $\{\hat{w}_{i,N}\}$  tenderà ad avere valori spettrali bassi alle frequenze per le quali la sequenza  $\{x_i\}$  ha spettro di potenza inferiore al livello del disturbo additivo  $\{z_i\}$ . Per capire ciò, pensiamo al caso in cui *il brusio dei vicini diventi assordante*: la sequenza  $\{y_i\}$  è praticamente coincidente con la sequenza  $\{z_i\}$ ; quindi è imprevedibile e viene a coincidere con la sequenza  $\{\varepsilon_{i,N}\}$ . Invece la sequenza  $\{\hat{w}_{i,N}\}$  tende a zero; infatti, *il coro è impercettibile ed è opportuno tapparsi le orecchie*.

Per comprendere meglio questo effetto, è opportuno ripetere l'intera analisi nel dominio delle frequenze, come faremo nel prossimo paragrafo, o nel dominio degli autovettori, come verrà fatto nel paragrafo ancora successivo.

### 3.8.3 Predizione, stima e sbiancamento: un riassunto

Appare opportuno riassumere i vari casi di predizione o stima di una serie temporale di autocorrelazione assegnata sia essa discendente da un processo AR, che da un processo AR + rumore, che qualsiasi. Le equazioni che seguono corrispondono alla predizione del processo  $\{y_i\}$ , consistente

### 3.8 STIMA CON RUMORE

in un processo AR  $\{x_i\}$ , cui è sovrapposto del rumore  $\{z_i\}$ , a varianza  $\sigma_z^2$ . Il valore q.m. dell'errore di predizione è indicato con  $\sigma_{\varepsilon,N}^2$ ; le incognite del sistema sono appunto  $\sigma_{\varepsilon,N}^2$  e gli  $N$  coefficienti  $a_{n,N}$ . Facendo la predizione del campione attuale ( $i$ -esimo) usando gli  $N$  precedenti si ha un errore  $\varepsilon_{i,N}$ :

$$\begin{aligned} y_i &= x_i + z_i \\ r_n &= E [x_i x_{i-n}^*] \\ \hat{y}_i &= - \sum_{n=1}^N a_{n,N} y_{i-n} \\ \varepsilon_{i,N} &= y_i - \hat{y}_i = y_i + \sum_{n=1}^N a_{n,N} y_{i-n} \end{aligned}$$

Le  $N$  equazioni per il calcolo degli  $a_{n,N}$  sono derivanti dall'ortogonalità:

$$E [(y_i - \hat{y}_{i,N}) y_{i-m}^*] = E \left[ \left( y_i + \sum_{n=1}^N a_{n,N} y_{i-n} \right) y_{i-m}^* \right] = 0 \quad m = 1, \dots, N$$

Un'altra equazione (qui messa come la prima, come viene fatto spesso in letteratura, ma con effetti a volte di confusione) invece dà il valore q.m. dell'errore di stima  $\sigma_{\varepsilon,N}^2$ :

$$\begin{aligned} E [(y_i - \hat{y}_{i,N}) y_i^*] &= E \left[ \left( y_i + \sum_{n=1}^N a_{n,N} y_{i-n} \right) y_i^* \right] = \sigma_{\varepsilon,N}^2 \\ \begin{bmatrix} r_0 + \sigma_z^2 & r_{-1} & r_{-2} & \dots & r_{-N} \\ r_1 & r_0 + \sigma_z^2 & r_{-1} & \dots & r_{-N+1} \\ r_2 & r_1 & r_0 + \sigma_z^2 & \dots & r_{-N+2} \\ \dots & \dots & \dots & \dots & \dots \\ r_N & r_{N-1} & r_{N-2} & \dots & r_0 + \sigma_z^2 \end{bmatrix} \begin{bmatrix} 1 \\ a_{1,N} \\ a_{2,N} \\ \dots \\ a_{N,N} \end{bmatrix} &= \begin{bmatrix} \sigma_{\varepsilon,N}^2 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix} \end{aligned} \quad (3.46)$$

Al crescere di  $\sigma_z^2$ , gli  $a_{n,N}$  tendono a 0 e  $\sigma_{\varepsilon,N}^2$  diventa sempre più prossimo a  $\sigma_z^2$ ; la predizione è inutile, se il dato è coperto da rumore bianco. Facciamo ora la stima della sequenza  $\{w_i\}$  di varianza supposta nota  $\sigma_w^2$ , modello di un possibile ingresso di un sistema autoregressivo di ordine  $N$ ,

### CAPITOLO 3 LA PREDIZIONE LINEARE

alla cui uscita  $\{x_i\}$  sarebbe aggiunto rumore bianco di varianza supposta nota  $\sigma_z^2$ . Per questa stima usiamo anche il valore attuale della sequenza osservata  $y_i$ ; si noti il cambio di segno nell'espressione dello stimatore:

$$\hat{w}_{i,N} = \sum_{n=0}^N c_{n,N} y_{i-n}$$

Questa volta le incognite sono i  $c_{n,N}$ , in numero di  $N + 1$ . L'ortogonalità ci dà:

$$E \left[ (w_i - \hat{w}_{i,N}) y_{i-m}^* \right] = E \left[ \left( w_i - \sum_{n=0}^N c_{n,N} y_{i-n} \right) y_{i-m}^* \right] = 0 \quad m = 0, \dots, N$$

$$\begin{bmatrix} r_0 + \sigma_z^2 & r_{-1} & r_{-2} & \dots & r_{-N+1} \\ r_1 & r_0 + \sigma_z^2 & r_{-1} & \dots & r_{-N+2} \\ r_2 & r_1 & r_0 + \sigma_z^2 & \dots & r_{-N+3} \\ \dots & \dots & \dots & \dots & \dots \\ r_{N-1} & r_{N-2} & r_{N-3} & \dots & r_0 + \sigma_z^2 \end{bmatrix} \begin{bmatrix} c_{0,N} \\ c_{1,N} \\ c_{2,N} \\ \dots \\ c_{N,N} \end{bmatrix} = \begin{bmatrix} \sigma_w^2 \\ 0 \\ 0 \\ \dots \\ 0 \end{bmatrix}$$

Il valore q.m. dell'errore di stima  $\sigma_{\varepsilon,N}^2$  è:

$$E \left[ (w_i - \hat{w}_{i,N}) w_i^* \right] = E \left[ \left( w_i - \sum_{n=0}^N c_{n,N} y_{i-n} \right) w_i^* \right] = \sigma_{\varepsilon,N}^2 = \sigma_w^2 (1 - c_{0,N})$$

Al crescere di  $\sigma_z$ , tutti i  $c_{n,N}$  tendono a 0; la stima diventa impossibile, il processo stimato ha ampiezza tendente a 0 e la varianza dell'errore tende a  $\sigma_w^2$ . Invece se  $\sigma_z = 0$  i due sistemi di equazioni visti sono identici.

Facciamo ora la stima del campione attuale (*i-esimo*) della sequenza  $\{x_i\}$  uscita del sistema autoregressivo usando gli  $N$  campioni precedenti della sequenza  $\{y_i\}$ . Attenzione al segno dello stimatore:

$$\hat{x}_{i,N} = - \sum_{n=1}^N b_{n,N} y_{i-n}$$

### 3.8 STIMA CON RUMORE

Le  $N$  equazioni per il calcolo degli  $b_{n,N}$  sono derivanti dall'ortogonalità:

$$E[(x_i - \hat{x}_{i,N}) y_{i-m}^*] = E\left[\left(x_i + \sum_{n=1}^N b_{n,N} y_{i-n}\right) y_{i-m}^*\right] = 0 \quad m = 1, \dots, N$$

$$\begin{bmatrix} r_0 + \sigma_z^2 & r_{-1} & r_{-2} & \dots & r_{-N} \\ r_1 & r_0 + \sigma_z^2 & r_{-1} & \dots & r_{-N+1} \\ r_2 & r_1 & r_0 + \sigma_z^2 & \dots & r_{-N+2} \\ \dots & \dots & \dots & \dots & \dots \\ r_N & r_{N-1} & r_{N-2} & \dots & r_0 + \sigma_z^2 \end{bmatrix} \begin{bmatrix} b_{1,N} \\ b_{2,N} \\ b_{2,N} \\ \dots \\ b_{N,N} \end{bmatrix} = - \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_N \end{bmatrix}$$

Un'altra equazione ci dà il valore q.m. dell'errore di stima  $\sigma_{\varepsilon,N}^2$ :

$$E[(x_i - \hat{x}_{i,N}) x_i^*] = E\left[\left(x_i + \sum_{n=1}^N b_{n,N} y_{i-n}\right) x_i^*\right] = \sigma_{\varepsilon,N}^2$$

$$\sigma_{\varepsilon,N}^2 = r_0 + \sum_{n=1}^N b_{n,N} r_{-n}$$

Facciamo infine la stima della sequenza  $\{x_i\}$  uscita del sistema autoregressivo usando sia gli  $N$  campioni precedenti che il campione attuale della sequenza  $\{y_i\}$ . Al solito, si noti la convenzione di segno diversa se si passa da predizione a stima,

$$\hat{x}_i = \sum_{n=0}^N d_{n,N} y_{i-n}$$

Le  $N + 1$  equazioni per il calcolo dei  $d_{n,N}$  sono sempre derivanti dall'ortogonalità:

$$E[(x_i - \hat{x}_i) y_{i-m}^*] = E\left[\left(x_i - \sum_{n=0}^N d_{n,N} y_{i-n}\right) y_{i-m}^*\right] = 0 \quad m = 0, \dots, N$$

### CAPITOLO 3 LA PREDIZIONE LINEARE

$$\begin{bmatrix} r_0 + \sigma_z^2 & r_{-1} & r_{-2} & \dots & r_{-N} \\ r_1 & r_0 + \sigma_z^2 & r_{-1} & \dots & r_{-N+1} \\ r_2 & r_1 & r_0 + \sigma_z^2 & \dots & r_{-N+2} \\ \dots & \dots & \dots & \dots & \dots \\ r_N & r_{N-1} & r_{N-2} & \dots & r_0 + \sigma_z^2 \end{bmatrix} \begin{bmatrix} d_{0,N} \\ d_{1,N} \\ d_{2,N} \\ \dots \\ d_{N,N} \end{bmatrix} = \begin{bmatrix} r_0 \\ r_1 \\ \dots \\ r_N \end{bmatrix}$$

e di nuovo, per  $\sigma_z^2$  crescente. i  $d_{n,N}$  tendono a 0.

In generale, è bene sottolineare che i problemi di stima (o predizione) sono sempre caratterizzati dalla determinazione di un vettore di coefficienti di predizione  $\underline{\alpha}_N$  mediante il prodotto dell'inverso (se esiste) della matrice di covarianza dei dati  $\underline{\mathbf{R}}$  (**sempre la stessa!**) per un opportuno vettore di correlazioni  $\underline{\rho}_N$  tra il segnale desiderato e i dati. Si ha quindi sempre una relazione del tipo:

$$\underline{\mathbf{R}} \underline{\alpha}_N = \underline{\rho}_N \rightarrow \underline{\alpha}_N = \underline{\mathbf{R}}^{-1} \underline{\rho}_N$$

ed il valore q.m. dell'errore di predizione o di stima è:

$$E[(x_i - \hat{x}_i) x_i^*] = r_0 - \underline{\rho}_N^* \underline{\mathbf{R}}^{-1} \underline{\rho}_N$$

## 3.9 La deconvoluzione nel dominio delle frequenze

Il problema della stima di un processo può essere risolto sia nel dominio dei tempi che in quello delle frequenze. Il vantaggio di operare nel dominio delle frequenze, veramente grande, sta nel fatto che la **matrice di covarianza dei dati è diagonale** qualora si tratta di dati risultanti dalla trasformazione di Fourier discreta di sequenze stazionarie colorate o bianche. Certamente, come visto in precedenza, il filtraggio deve essere fatto in modo periodicizzato nei tempi. Poi, operando nel dominio delle frequenze, l'inversione della matrice di covarianza dei dati  $\underline{\mathbf{R}}$  di ordine  $[N \times N]$  necessaria per ottenere il vettore dei coefficienti  $\underline{\alpha}_N$ , a partire dal vettore delle correlazioni  $\underline{\rho}_N$

$$\underline{\mathbf{R}} \underline{\alpha}_N = \underline{\rho}_N; \rightarrow \underline{\alpha}_N = \underline{\mathbf{R}}^{-1} \underline{\rho}_N$$

### 3.9 LA DECONVOLUZIONE NEL DOMINIO DELLE FREQUENZE

diventa semplicissima, perchè corrispondente a  $N$  equazioni ad una sola incognita. In questo caso, tuttavia, non sarà in generale possibile limitare il dominio su cui è definito il filtro, come è possibile fare nel dominio dei tempi, in cui si conoscono quali campioni vengono utilizzati per fare la stima (ad esempio, gli  $N$  campioni che precedono il campione da stimare). Ora l'asse dei tempi si estende da  $-\infty$  ad  $\infty$ . Passando all'analisi del problema, sia data la sequenza:

$$y_i = x_i + z_i = \sum_{n=-N/2}^{N/2} w_{i-n} h_n + z_i \quad (3.47)$$

dove  $\{x_i\}$  e  $\{z_i\}$  sono due sequenze, l'una,  $\{x_i\}$  corrispondente al segnale (*le voci bianche*  $w_i$  riverberate da una *cattedrale* con risposta all'impulso  $h_n$  che ora può essere qualsiasi e non necessariamente corrispondente ad un sistema AR) e l'altra  $\{z_i\}$  corrispondente a rumore additivo che per semplicità considereremo a spettro bianco (*il brusio dei vicini*). In quanto segue, la frequenza può essere rappresentata sia da un indice continuo (la solita pulsazione  $\omega$ ) sia, pensando alle trasformazioni di Fourier discrete, da un indice discreto  $k$ . Per semplicità, useremo l'indice continuo. Supponiamo noti sia lo spettro di potenza della sequenza  $\{w_i\}$ , e cioè  $E[|W(\omega)|^2] = \sigma_w^2 T$ , sia **la funzione di trasferimento del filtro**  $H(\omega)$  che ha operato sulla sequenza (*la riverberazione della cattedrale*); nei casi precedentemente discussi, bastava invece conoscere l'autocorrelazione del filtro e quindi  $|H(\omega)|^2$ . La sequenza  $\{z_i\}$  (*il brusio dei vicini*) è anche ignota, e se ne conoscono solo le caratteristiche spettrali. Si vuole stimare

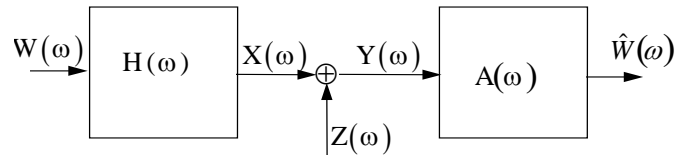


Figura 3.7: Deconvoluzione nel dominio delle frequenze

$\{w_i\}$  da  $\{y_i\}$ , nel dominio della frequenza. Pertanto, indicando con  $A(\omega)$  la funzione di trasferimento del filtro che utilizzeremo per fare la stima



### CAPITOLO 3 LA PREDIZIONE LINEARE

della sequenza  $\{w_i\}$  e indicando con lettere maiuscole le trasformate delle sequenze (fig. 12.7):

$$Y(\omega) = W(\omega)H(\omega) + Z(\omega); \quad E[|W(\omega)|^2] = \sigma_w^2(\omega)T \quad (3.48)$$

$$E[|Z(\omega)|^2] = \sigma_z^2(\omega)T \quad (3.49)$$

$$\widehat{W}(\omega) = A(\omega)Y(\omega) \quad (3.50)$$

Il principio di ortogonalità (derivante dall'ottimalità del risultato desiderato) comporta che sia nulla la correlazione tra l'errore di stima ed il dato (e quindi nelle frequenze sia nullo il valore atteso del prodotto tra la trasformata dell'errore di predizione e la coniugata della trasformata del dato); la pulsazione  $\omega$  è un parametro che non trascriveremo.

$$E[(W - AY)Y^*] = 0; \quad (3.51)$$

$$A = \frac{E[WY^*]}{E[YY^*]} = \frac{\sigma_w^2 H^*}{\sigma_w^2 |H|^2 + \sigma_z^2} = \frac{H^*}{HH^* + Q}; \quad (3.52)$$

$$Q(\omega) = \frac{\sigma_z^2(\omega)}{\sigma_w^2(\omega)}; \quad (3.53)$$

Si vede quindi che il risultato della stima nel dominio delle frequenze è un filtro  $A(\omega) = 1/H(\omega)$ , se  $Q \ll |H(\omega)|^2$  e cioè se il disturbo è trascurabile. Quando invece il disturbo è più ampio del segnale ( $Q \gg |H(\omega)|^2$ ), il filtro tende a dare uscita nulla (*alle frequenze in cui il segnale è coperto dal brusio, non ne tentiamo la ricostruzione*).

Quindi, possiamo ritornare sulla questione discussa nel precedente paragrafo e cioè chiarire bene la differenza tra  $\{\varepsilon_{i,N}\}$  errore di predizione della sequenza  $\{y_i\}$  e  $\{\widehat{w}_{i,N}\}$  stima della sequenza bianca  $\{w_i\}$  che alimenta il sistema riverberante che crea  $\{x_i\}$ . In assenza di rumore additivo  $\{z_i\}$  non vi sono differenze: invece, se  $\sigma_z^2 \neq 0$  la stima  $\{\widehat{w}_{i,N}\}$  del segnale originario  $\{w_i\}$  sarà diversa dall'errore di predizione  $\{\varepsilon_{i,N}\}$  ed in particolare di ampiezza ridotta alle frequenze dove è più sensibile il rumore.

L'operazione così realizzata, e cioè la stima della sequenza  $\{w_i\}$ , si indica con il nome di **deconvoluzione**; infatti la sequenza  $\{w_i\}$  viene osservata dopo la convoluzione con il filtro  $\{h_n\}$  e l'aggiunta di rumore bianco;

### 3.9 LA DECONVOLUZIONE NEL DOMINIO DELLE FREQUENZE

scopo della stima è quello di recuperare la sequenza  $\{w_i\}$  rimuovendo l'effetto del filtro  $H(\omega)$  con l'operazione appunto di deconvoluzione. Come si è visto, la cosa è fattibile applicando un filtro inverso, purchè il livello del segnale filtrato non scenda sotto il livello del rumore: a quelle frequenze, la stima va a zero.

#### 3.9.1 Deconvoluzione predittiva nelle frequenze

Operando nel dominio delle frequenze, non appare immediatamente possibile stabilire il supporto del filtro, cioè definire i campioni per i quali i coefficienti del filtro sono diversi da zero: quindi non sarebbe possibile trovare il filtro ottimo che determina l'errore di predizione, perchè questo implica che il filtro sia causale, cosa che appare difficile imporre. Esiste peraltro la tecnica di fattorizzazione spettrale (Wiener, Kolmogorov) che permette di risolvere il problema dell'errore di predizione anche nel dominio delle frequenze. Il principio è semplice. Con un filtro di lunghezza infinita, la sequenza degli errori di predizione deve:

- (a) essere bianca;
- (b) essere ottenuta dalla sequenza  $\{y_i\}$  attraverso un filtro causale.

Ora, se lo spettro della sequenza  $\{y_i\}$  è  $E[|Y(\omega)|^2]$ , il modulo della funzione di trasferimento del filtro sbiancatore sarà:

$$|A(\omega)|^2 = \frac{1}{E[|Y(\omega)|^2]} \quad (3.54)$$

Ne conosciamo quindi la caratteristica di ampiezza e sappiamo anche, dalle considerazioni fatte nel dominio dei tempi, che la sequenza di campioni che corrisponde al filtro predittore deve essere non solo causale (un predittore utilizza solo i dati "passati") ma anche a minima fase; in realtà avevamo dedotto la proprietà di minima fase solo per  $N$  finito qualsiasi; supponiamo ora che questa proprietà di minima fase si mantenga anche per  $N$  infinito. Pertanto, come visto nel capitolo 7, la caratteristica di fase della funzione di trasferimento sarà trasformata di Hilbert del logaritmo della caratteristica di ampiezza. In conclusione il predittore causale ottenuto dall'analisi nel dominio delle frequenze e di lunghezza infinita (in teoria) ha funzione di trasferimento, in fase ed ampiezza:

$$|A_\infty(\omega)| = \sqrt{\frac{1}{E[|Y(\omega)|^2]}} \quad (3.55)$$

$$\angle A_\infty(\omega) = -\frac{1}{2} \mathcal{H}\{\log E[|Y(\omega)|^2]\} \quad (3.56)$$

Questo risultato, scoperto negli anni della II guerra mondiale ed applicato alla predizione delle tracce radar per il pilotaggio delle batterie contraeree, era stato tenuto gelosamente segreto.

Sono da evitare zeri dello spettro  $E[|Y(\omega)|^2]$ , in particolare alla frequenza di Nyquist, per evitare singolarità del logaritmo e quindi problemi con la trasformazione di Hilbert. Se poi gli zeri spettrali sono ineludibili perchè troppo significativi, non è più possibile fare una stima spettrale AR che diventa troppo lunga ma si deve passare ad una stima spettrale ARMA. In altre parole gli zeri vanno tenuti in conto separatamente, e eventualmente sbiancati in precedenza con i poli opportuni (piuttosto che con infiniti zeri). Per uno studio dell'analisi spettrale ARMA, si rimanda ai testi nella bibliografia proposta nella prefazione.

### 3.10 Le equazioni Y - W nel dominio degli autovettori

Ritornando ora nel dominio dei tempi, trasformeremo l'analisi fatta nel dominio delle frequenze in un'analisi nel dominio degli autovettori della matrice di covarianza  $\underline{\mathbf{R}}$ . Abbiamo visto che, se sulla diagonale principale della matrice  $\underline{\mathbf{R}}$  è addizionata *luce bianca* (*white light, in inglese*)  $\sigma_z^2$ , il sistema diventa ben condizionato, in quanto tutti gli autovalori di  $\underline{\mathbf{R}}$  diventano positivi; si comprende l'importanza di garantire che la matrice  $\underline{\mathbf{R}}$  sia semidefinita positiva; altrimenti, l'aggiunta di *luce bianca*, nel caso di autovalori negativi, poteva portare all'instabilità invece che alla stabilità. Quindi, una tecnica di garantire il condizionamento della matrice quindi la soluzione delle equazioni di Y - W è appunto quella di addizionare *luce bianca* anche se poi non vi è nessun rumore  $\{z_i\}$  sovrapposto al segnale  $\{x_i\}$ . Sviluppando in autovalori ed autovettori di  $\underline{\mathbf{R}}$  sia  $\underline{\mathbf{R}}$  che  $\underline{\alpha}_N$  e  $\underline{\rho}_N$ , si ha, partendo dall'equazione di stima generica:

$$\begin{aligned} \underline{\mathbf{R}} \underline{\alpha}_N &= \underline{\rho}_N \\ \underline{\mathbf{R}} \underline{\alpha}_N &= \sum_{n=1}^N \lambda_n \underline{\mathbf{e}}_n \underline{\mathbf{e}}_n^* \underline{\alpha}_N = \sum_{n=1}^N \underline{\mathbf{e}}_n \underline{\mathbf{e}}_n^* \underline{\rho}_N = \underline{\rho}_N \end{aligned} \quad (3.57)$$

### 3.10 LE EQUAZIONI Y - W NEL DOMINIO DEGLI AUTOVETTORI

Indicando con

$$\underline{\mathbf{e}}_n^* \underline{\boldsymbol{\alpha}}_N = a_{n,N} \quad \underline{\mathbf{e}}_n^* \underline{\boldsymbol{\rho}}_N = \rho_{n,N} \quad (3.58)$$

abbiamo che:

$$\begin{aligned} \lambda_n a_{n,N} &= \rho_{n,N} \\ a_{n,N} &= \frac{\rho_{n,N}}{\lambda_n} \end{aligned} \quad (3.59)$$

Addizionando  $\sigma_z^2 \mathbf{I}$  ad  $\underline{\mathbf{R}}$  abbiamo

$$a_{n,N} = \frac{\rho_{n,N}}{\lambda_n + \sigma_z^2} \quad (3.60)$$

Questa equazione corrisponde all'equazione (3.52) tranne che l'una è stata ottenuta nel dominio delle frequenze e l'altra nel dominio dei tempi (autovettori di  $\underline{\mathbf{R}}$  e correlazione tra segnale desiderato e dati). In altre parole, di nuovo si trova che sono attenuate le componenti della soluzione  $q_{n,N}$  lungo gli autovettori  $\underline{\mathbf{e}}_n$  con autovalori  $\lambda_n$  piccoli rispetto al disturbo  $\sigma_z^2$ . Se ora  $\sigma_z^2$  è abbastanza piccola, ad esempio  $r_0/100$  alteriamo, anche se percentualmente di poco, tutte le componenti  $a_{n,N}$  comprese quelle corrispondenti agli autovalori più grandi di  $\underline{\mathbf{R}}$ , che non avevano motivo di essere alterate. In compenso, l'operazione è computazionalmente facile, perchè basta sommare una matrice identica alla matrice  $\underline{\mathbf{R}}$ . Una soluzione diversa, che prevede però di conoscere tutti gli autovalori di  $\underline{\mathbf{R}}$ , è quella di sostituire a  $\lambda_i \rightarrow \max[\lambda_i, \sigma_z^2]$  (la tecnica è denominata *water filling* dall'effetto di riempire i buchi spettrali sostituendoli con una costante). In questo modo si alterano solo le componenti della soluzione relative agli autovalori minori di  $\sigma_z^2$ , lasciando immutate le altre.

Se ora supponiamo di non conoscere perfettamente la matrice di covarianza e quindi nemmeno i  $\lambda_n$ , abbiamo una soluzione ancora diversa. Infatti, nell'equazione 3.59 deve comparire un termine di rumore  $\nu_n$  per modellare questa incertezza. Omettendo i pedici in  $n$ ,  $N$  per semplicità:

$$\lambda a + \nu = \rho \quad (3.61)$$

e la stima ai minimi quadrati della variabile  $a$  è:

$$\hat{a} = \mu \rho$$

Minimizzando l'errore q.m. di stima si ha:

$$\min_{\mu} E [(\mu \rho - a)^2] \rightarrow E [(\mu \rho - a) \rho] = 0$$

### CAPITOLO 3 LA PREDIZIONE LINEARE

$$\mu = \frac{E[\rho a]}{E[\rho^2]} = \frac{\lambda \sigma_q^2}{\lambda^2 \sigma_q^2 + \sigma_{\nu}^2}$$

$$\hat{a}_{n,N} = \mu \rho_{n,N} = \frac{\lambda_n}{\lambda_n^2 + \frac{\sigma_{\nu}^2}{\sigma_q^2}} \rho_{n,N} \quad (3.62)$$

dove con  $\sigma_{\nu}^2$  abbiamo indicato la varianza del disturbo  $\nu_n$ . In termini matriciali, abbiamo:

$$\underline{\alpha}_N = \{ \underline{\mathbf{R}}^2 + \frac{\sigma_{\nu}^2}{\sigma_q^2} \underline{\mathbf{I}} \}^{-1} \underline{\mathbf{R}} \underline{\rho}_N \quad (3.63)$$

I vantaggi di questa ultima soluzione per le equazioni di Yule Walker sono che:

- il sistema non è mai singolare;
- non abbiamo bisogno di determinare autovalori ed autovettori di  $\underline{\mathbf{R}}$ ;
- le componenti della soluzione peggio condizionate e cioè quelle corrispondenti agli autovalori più piccoli tendono ad essere poco amplificate invece di essere amplificate al massimo come nel caso dell'equazione (3.52).

## Capitolo 4

### Filtraggio adattativo

Uno dei numerosi motivi della diffusione delle tecniche di trattamento numerico dei segnali è la possibilità di modificare i filtri utilizzati in funzione delle caratteristiche spettrali del dato, variabili con il tempo e/o con lo spazio. Un tipico esempio di sistema tempo variante è il tratto vocale, che sagoma lo spettro del segnale generato dalle corde vocali. La stima spettrale di un segnale vocale dovrà quindi essere "adattativa" nel tempo.

Un esempio intuitivo di un sistema adattativo è quello del cancellatore di disturbo: si pensi ad un pilota di elicottero, che ascolta un segnale  $s_i$  ma è disturbato dal rumore del rotore; possiamo pensare di aggiungere negli auricolari del pilota un segnale che compensi ed annulli questo rumore esterno: per fare ciò è necessario ricostruire elettronicamente sia in fase che in ampiezza il rumore disturbante, per poi applicarlo con il segno cambiato. La struttura numerica che si può adottare è presentata in figura 13.1: anzitutto, viene prelevata una sequenza (indicata poi come il disturbo  $\{d_i\}$ ) quanto possibile correlata con il rumore del rotore ascoltato dal pilota  $\{n_i\}$ , ma indipendente dalla sequenza di segnale  $\{s_i\}$  che desideriamo arrivi indisturbata. Ad esempio, la sequenza  $\{d_i\}$  potrebbe essere prelevata direttamente con un sensore posto vicino al rotore; si indichi poi con  $\{y_i\}$  il segnale disturbato che arriva all'orecchio del pilota. Si desidera rimuovere da  $\{y_i\}$  tutta l'energia correlata con il disturbo misurato  $\{d_i\}$ . Così facendo, si **decorrela** il segnale ascoltato dal pilota  $\{y_i\}$  dal disturbo misurato  $\{d_{i-n+1}\}$ ,  $n = 1, \dots, N$ . In altre parole, il residuo  $\varepsilon_{i,N}$  che si ottiene sottraendo da  $\{y_i\}$  il disturbo stimato  $\{\hat{n}_i\}$  deve essere la sequenza  $\{s_i\}$  per quanto possibile **incorrelata**

## CAPITOLO 4 FILTRAGGIO ADATTATIVO

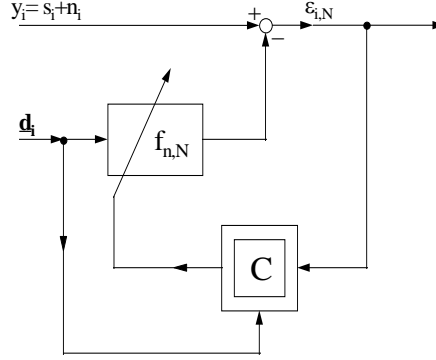


Figura 4.1: Filtro adattativo

con la sequenza di vettori  $\underline{d}_i$  aventi componenti  $\{d_{i-n+1}\}$ ,  $n = 1, \dots, N$ . A sua volta, la sequenza  $\{\hat{n}_i\}$  è ottenuta filtrando il disturbo misurato  $\{d_{i-n+1}, n = 1, \dots, N\}$  con un filtro opportuno  $\{f_{i,n}\}$ ,  $n = 1, \dots, N$ , che sarà tempo variante (ecco il perchè dell'indice  $i$ ), ad esempio per compensare i ritardi tempo - varianti dovuti ai movimenti della testa del pilota.

$$y_i = s_i + n_i; \quad \varepsilon_i = s_i + n_i - \hat{n}_i = y_i - \sum_{n=1}^N f_{i,n} d_{i-n+1} \quad (4.1)$$

Quando si avrà che

$$E[\varepsilon_i d_{i-n+1}] = 0 \quad n = 1, \dots, N \quad (4.2)$$

avremo fatto il possibile per levare il rumore del rotore dall'auricolare. Difatti, nel segnale residuo  $\varepsilon_i = s_i$  non ci dovrebbe più essere correlazione con il disturbo  $\{n_i\}$ , sempre che il filtro sia di durata sufficiente ( $N$  abbastanza grande) e che il disturbo misurato  $\{d_i\}$  sia legato al disturbo da rimuovere  $\{n_i\}$  attraverso una funzione di trasferimento che non ha zeri spettrali e senza che vi sia ulteriore disturbo sovrapposto; allora infatti il disturbo stimato  $\{\hat{n}_i\}$  assomiglia molto al disturbo vero  $\{n_i\}$ .

Questo non è che un possibile esempio: un altro è quello di un elettrocardiogramma fetale; l'elettrocardiogramma di un nascituro è misurabile

#### 4.1 LA TECNICA DI WIDROW E HOFF

solo se combinato con quello della madre; per isolare il primo, è necessario misurare un segnale correlato soprattutto al secondo e poi sottrarlo, sempre tenendo conto della variabilità delle condizioni dell'esperimento. Ancora, si pensi alla cancellazione di un'eco che arriva ad un auricolare telefonico; in questo caso, l'eco nasce soprattutto dall'accoppiamento fonico tra il telefono ed il microfono dell'ascoltatore lontano. Il segnale generato localmente viene trasmesso all'ascoltatore, rientra nel microfono lontano e ritorna come eco al parlatore. Per sopprimere quest'eco, basta decorrelare il segnale in arrivo dal segnale in partenza. La funzione di trasferimento è tempo variante; cambierà, ad esempio, se una barriera (una mano) riflette le radiazioni acustiche dal telefono al microfono dell'apparecchio lontano. Ma ancora, si può pensare che anche un predittore possa essere un sistema adattativo: in questo caso cercheremo di decorrelare il segnale attuale  $s_i$  da una sua versione ritardata di un campione e cioè porremo:

$$d_{i-n+1} = s_{i-n}; \quad n = 1, \dots, N$$

### 4.1 La tecnica di Widrow e Hoff

Per presentare in modo sintetico l'evoluzione del filtro utilizzeremo la notazione vettoriale (supponendo il segnale complesso, indichiamo con \* le operazioni di trasposizione e coniugio, dove necessario); indichiamo poi con i vettori  $\underline{\mathbf{f}}_i$ ,  $\underline{\mathbf{d}}_i$  gli  $N$  coefficienti del filtro e gli ultimi  $N$  campioni della sequenza  $\{d_i\}$  all'istante  $iT$ :

$$\underline{\mathbf{f}}_i = \{f_{i,n}\}; \quad n = 1, \dots, N \quad (4.3)$$

$$\underline{\mathbf{d}}_i = \{d_{i-n+1}\} \quad n = 1, \dots, N \quad (4.4)$$

La sottrazione del disturbo stimato viene quindi presentata così:

$$\varepsilon_i = s_i + n_i - \hat{n}_i = y_i - \hat{n}_i = y_i - \sum_{n=1}^N f_{i,n} d_{i-n+1} = y_i - \underline{\mathbf{f}}_i^* \underline{\mathbf{d}}_i \quad (4.5)$$

Inoltre aggiorniamo il filtro ad ogni istante con l'equazione (Widrow e Hoff):

$$\underline{\mathbf{f}}_{i+1} = \underline{\mathbf{f}}_i + \tilde{\gamma} \varepsilon_i^* \underline{\mathbf{d}}_i \quad (4.6)$$

Osserviamo anzitutto che  $\underline{\mathbf{f}}_i$  è un vettore casuale, di cui cerchiamo di determinare l'evoluzione nel tempo. Dimosteremo ora che, purchè  $\tilde{\gamma}$  sia un



## CAPITOLO 4 FILTRAGGIO ADATTATIVO

fattore abbastanza piccolo,  $\underline{\mathbf{f}}_i$  convergerà al filtro ottimo in qualche modo, ancora da illustrare. Anzitutto, è opportuno chiarire il perchè si utilizza l'equazione di Widrow e Hoff. Indichiamo con  $Q_i$  l'errore quadratico medio (media di insieme), all'istante  $i$ -esimo:

$$Q_i = E[|\varepsilon_i|^2] \quad (4.7)$$

ed il suo gradiente rispetto al filtro  $\underline{\mathbf{f}}_i$ :

$$\nabla_{\underline{\mathbf{f}}} Q_i = -2E[\varepsilon_i^* \underline{\mathbf{d}}_i] \quad (4.8)$$

pertanto l'equazione (4.6) corrisponde all'aggiornamento del vettore filtro  $\underline{\mathbf{f}}_i$  nella direzione della minimizzazione dell'errore quadratico medio. Non essendo possibile ricorrere all'operatore di media d'insieme  $E[.]$ , bisogna accettarne una stima stocastica anche molto mediocre, e cioè  $\varepsilon_i^* \underline{\mathbf{d}}_i$ . Per evitare fenomeni di instabilità, come vedremo meglio al prossimo paragrafo, è necessario che il passo nella direzione del gradiente sia piccolo. Questo implica che l'aggiornamento del filtro  $\underline{\mathbf{f}}_i$  è lento. Se osserviamo l'aggiornamento ogni  $K$  istanti, invece che ad ogni istante, trascurando la variazione del filtro in questo intervallo  $KT$  visto che  $\tilde{\gamma}$  è comunque piccolo, abbiamo:

$$\underline{\mathbf{f}}_{i+K} = \underline{\mathbf{f}}_i + \gamma \frac{1}{K} \sum_{k=0}^{K-1} \varepsilon_{i+k}^* \underline{\mathbf{d}}_{i+k}; \quad \gamma = K \tilde{\gamma} \quad (4.9)$$

Quindi, se il sistema è stazionario, o almeno lo è durante l'aggiornamento del filtro, la stima stocastica del gradiente è ragionevole, purchè l'aggiornamento sia lento. Se quindi approssimiamo con:

$$\frac{1}{K} \sum_{k=0}^{K-1} \varepsilon_{i+k}^* \underline{\mathbf{d}}_{i+k} \simeq E[\varepsilon_i^* \underline{\mathbf{d}}_i] \quad (4.10)$$

possiamo ritenere valida l'equazione (4.9) che riscriviamo, decidendo anche di porre  $K = 1$ .

$$\underline{\mathbf{f}}_{i+1} = \underline{\mathbf{f}}_i + \gamma E[\varepsilon_i^* \underline{\mathbf{d}}_i] \quad (4.11)$$

Possiamo ora studiare l'evoluzione del filtro, che è divenuta deterministica per l'approssimazione fatta in precedenza (4.10):

$$\underline{\mathbf{f}}_{i+1} = \underline{\mathbf{f}}_i + \gamma E[\underline{\mathbf{d}}_i (y_i^* - \underline{\mathbf{d}}_i^* \underline{\mathbf{f}}_i)] \quad (4.12)$$

#### 4.1 LA TECNICA DI WIDROW E HOFF

Indicando con  $\underline{\mathbf{R}}$  la matrice di covarianza dei dati e con  $\underline{\mathbf{r}}$  il vettore della correlazione tra segnale e dati:

$$\underline{\mathbf{R}} = E[\underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*]; \quad \underline{\mathbf{r}} = E[y_i^* \underline{\mathbf{d}}_i] \quad (4.13)$$

Sostituendo nell'equazione (4.12):

$$\underline{\mathbf{f}}_{i+1} = \underline{\mathbf{f}}_i + \gamma E[\varepsilon_i \underline{\mathbf{d}}_i] \quad (4.14)$$

$$\underline{\mathbf{f}}_{i+1} = (\underline{\mathbf{I}} - \gamma \underline{\mathbf{R}}) \underline{\mathbf{f}}_i + \gamma \underline{\mathbf{r}} \quad (4.15)$$

Per  $i \rightarrow \infty$ ,  $\underline{\mathbf{f}}_{i+1} = \underline{\mathbf{f}}_i$  e quindi:

$$\underline{\mathbf{f}}_\infty = \underline{\mathbf{R}}^{-1} \underline{\mathbf{r}} \quad (4.16)$$

La differenza tra la media di insieme del filtro corrente  $\underline{\mathbf{f}}_i$  ed il filtro desiderato  $\underline{\mathbf{f}}_\infty$  è il filtro errore  $\underline{\mathbf{g}}_i$  (media di insieme) che evolve nel tempo con l'equazione:

$$\underline{\mathbf{f}}_{i+1} - \underline{\mathbf{f}}_\infty = \underline{\mathbf{g}}_{i+1} = (\underline{\mathbf{I}} - \gamma \underline{\mathbf{R}}) \underline{\mathbf{g}}_i = (\underline{\mathbf{I}} - \gamma \underline{\mathbf{R}})^{i+1} \underline{\mathbf{g}}_0 \quad (4.17)$$

Condizione necessaria perchè la media di insieme del filtro errore converga a zero è che il modulo degli autovalori della matrice  $\underline{\mathbf{I}} - \gamma \underline{\mathbf{R}}$  sia inferiore ad uno, e cioè  $|1 - \gamma \lambda_n| < 1$ . Questo implica:

$$0 < \gamma < \frac{2}{\lambda_{max}}; \rightarrow \frac{2}{\lambda_{max}} > \frac{2}{\sum_{n=1}^N \lambda_n} = \frac{2}{Tr \underline{\mathbf{R}}} = \frac{2}{N \sigma_d^2} \quad (4.18)$$

pertanto imponendo:

$$\gamma \leq \frac{2}{N \sigma_d^2} \leq \frac{2}{\lambda_{max}} \quad (4.19)$$

siamo certi di ottenere il risultato di stabilità richiesto. Con  $Tr \underline{\mathbf{R}}$  abbiamo indicato la traccia della matrice  $\underline{\mathbf{R}}$  e cioè la somma degli elementi della sua diagonale principale; ciò corrisponde a  $N$  volte la varianza del disturbo stesso. Il risultato ottimale e cioè la convergenza più veloce si otterrà quando tutti gli autovalori sono eguali, come si vedrà nella prossima sezione, e poi ponendo  $\gamma = 1/\lambda_{max}$ .

Si osservi che l'equazione di Widrow e Hoff, nella sua grande semplicità, ha un significato notevole: in sostanza, il sistema fa la stima  $\underline{\mathbf{f}}_i$  della funzione di trasferimento e poi, per migliorarla, utilizza l'esperienza e quindi sia l'errore commesso  $\varepsilon_i$  che l'ingresso  $\underline{\mathbf{d}}_i$  che ha condotto a quell'errore con quella stima. Questo è un meccanismo di autocorrezione valido per chiunque.

## 4.2 Dispersione degli autovalori

La convergenza a zero del filtro errore  $\underline{\mathbf{g}}_i$  dipende dagli autovalori  $\lambda_n$  e quindi le componenti del filtro errore lungo le direzioni degli autovettori di  $\underline{\mathbf{R}}$  con autovalori piccoli si aggiornano lentamente; peraltro, non è possibile accrescere troppo  $\gamma$ , limitato dall'autovalore più grande. Infatti, indicando con  $g_{i,n}$  la componente di  $\underline{\mathbf{g}}_i$  lungo  $\underline{\mathbf{e}}_n$ , autovettore corrispondente all' $n$ -esimo autovalore  $\lambda_n$  della matrice  $\underline{\mathbf{R}}$ :

$$g_{i,n} = \underline{\mathbf{e}}_n^* \underline{\mathbf{g}}_i = (1 - \gamma \lambda_n)^{i+1} g_{0,n} \quad (4.20)$$

Se ad esempio il disturbo è stato filtrato passa basso, come già visto in precedenza, la matrice  $\underline{\mathbf{R}}$  sarà mal condizionata e qualcuno degli autovalori sarà piccolo. La componente del filtro errore verso quell'autovettore e in sostanza alle frequenze corrispondenti al minimo dello spettro, si aggiornerà lentamente, visto che è poco eccitata dal disturbo  $\underline{\mathbf{d}}_i$ , nullo o quasi in quella banda. Peggio ancora, può accadere che a causa di problemi numerici queste componenti del filtro errore possano crescere oltre misura senza controllo. Il problema non apparirebbe grave, se il sistema si mantiene lineare, visto che queste componenti del filtro errore agiscono su componenti di disturbo trascurabili, e quindi il loro effetto è trascurabile. Tuttavia, può accadere che i campioni del filtro (le prese, o *taps* in inglese), crescendo senza controllo nella banda poco controllata, si avvicinino ai valori di trabocco per i coefficienti, rendendo così il problema non lineare e distruggendo la capacità del filtro ad aggiornarsi. Il problema del cosiddetto *tap wandering* ossia "girovagare delle prese" deve essere considerato con cura nel progetto di qualsiasi filtro adattativo.

Per evitarlo, e quindi per fare in modo che tutte le componenti del filtro errore si annullino alla stessa velocità si dovrebbe fare in modo che la matrice  $\underline{\mathbf{R}}$  sia proporzionale alla matrice identica, e cioè che la sequenza  $\{d_i\}$  sia un processo a spettro bianco. Quindi, prima di utilizzare la sequenza  $\{d_i\}$  per fare la cancellazione, è opportuno sbiancarla. Questo implica stimare il suo spettro; alternatively, si può utilizzare un filtro a traliccio adattativo per effettuare lo sbiancamento. Un'altra tecnica efficace è quella di decomporre il segnale in bande con un banco di filtri, ad esempio polifase; per ogni banda, in cui si può pensare che lo spettro del segnale sia uniforme, si effettua un filtraggio adattativo, con un coefficiente di aggiornamento diverso.

### 4.3 RAPIDITÀ DELLA CONVERGENZA

Le operazioni di sbiancamento sono anch'esse tempo varianti come quelle di sottrazione ottima. E' naturale porsi il problema di accelerare al massimo tali procedure di stima per potere garantire in tutte le condizioni un rapido ritorno ad elevate reiezioni del disturbo. Peraltro, come vedremo nel prossimo paragrafo, accelerare la convergenza implica una maggiore sensibilità al rumore; *se si è troppo sensibili alla moda ...* .

### 4.3 Rapidità della convergenza

In questo paragrafo analizzeremo le conseguenze dell'approssimazione fatta con l'equazione (4.10); in particolare, verrà delineata la necessità di un compromesso nella determinazione del passo  $\gamma$  di aggiornamento del filtro da adattare: infatti vi è sempre una sequenza di segnale  $\{s_i\}$ , sovrapposta ma indipendente dal disturbo che si vuole cancellare; questo segnale influenza in modo casuale il filtro  $\underline{f}_i$  (in questo paragrafo indichiamo con  $\underline{f}_i$  il vettore casuale e non la sua media) e impedisce che esso raggiunga stabilmente il valore desiderato  $\underline{f}_\infty$  anche per  $i \rightarrow \infty$ , anche se le statistiche non cambiano (in assenza di fenomeni transitori, cioè). Ancora, si deve tenere conto della casualità del disturbo misurato  $\underline{d}_i$  e quindi dell'aggiornamento del filtro stesso. Tuttavia, se  $\gamma$  fosse molto piccolo, l'aggiornamento sarebbe molto lento e queste due diverse variabili casuali agirebbero solo attraverso la loro media, fatta su un numero grande di campioni; gli effetti di dispersione avranno scarsa influenza, decrescente con  $\gamma$ . A regime quindi, il filtro avrà piccoli scostamenti dal valore asintotico corrispondente alla media di insieme. Invece, se  $\gamma$  non è piccolo, anzi si avvicina al limite teorico, avremo che il filtro  $\underline{f}_i$ , **pur convergendo più velocemente in media di insieme** al valore di regime (4.17), **avrà a regime una dispersione maggiore**, per la grande sensibilità ai valori (casuali) sia della sequenza di segnale  $\{s_i\}$  che del disturbo  $\underline{d}_i$ ; per un'idea intuitiva, si ricordino i detti popolari "ondeggianti come una banderuola" ( $\gamma$  grande) e l'altro "inerte come un sasso" ( $\gamma$  piccolo).

In conclusione, si comprende come all'inizio della convergenza il passo debba essere grande; il passo si deve poi ridurre progressivamente fino ad avere un valore modestissimo in condizioni di regime quasi - statico (lontananza dal transitorio iniziale) e convergenza praticamente raggiunta. Naturalmente, se c'è un nuovo cambiamento e si ripropone la necessità

## CAPITOLO 4 FILTRAGGIO ADATTATIVO

di un aggiornamento, il filtro deve riacquisire la sua capacità di adeguarsi e quindi  $\gamma$  deve crescere di nuovo.

### 4.3.1 Dispersione del filtro errore $\underline{g}_i$

L'analisi della dispersione del filtro errore porterà per il passo di aggiornamento a risultati più restrittivi di quelli ottenuti con l'analisi dell'evoluzione della media di insieme del filtro errore. Un modo semplice per analizzare il problema è il seguente. Si consideri un sistema monodimensionale del tipo (il parametro  $f$  è incognito):

$$y_i = f d_i + s_i ; \quad f_{i+1} = f_i + \gamma \varepsilon_i d_i ; \quad \varepsilon_i = y_i - f_i d_i ;$$

in cui, al solito, si misura il disturbo  $d_i$  per poi valutare il fattore  $f$ . Si suppongano  $s, d$  indipendenti per cui il filtro asintotico è  $f_\infty = 0$ , cui si arriva a partire da una condizione iniziale  $f_0 = g_0 \neq 0$ . La dinamica effettiva del filtro errore è la seguente:

$$g_{i+1} = g_i + \gamma (s_i - g_i d_i) d_i = g_i (1 - \gamma d_i^2) + \gamma d_i s_i ; \quad (4.21)$$

Ora, a regime, quando il filtro errore è ormai ridotto a zero **in media**, levandogli gli indici temporali:

$$\sigma_g^2 = E \left[ (g (1 - \gamma d^2) + \gamma d s)^2 \right] = E \left[ g^2 (1 - \gamma d^2)^2 \right] + \gamma^2 E [d^2 s^2] \quad (4.22)$$

Qualora  $d$  sia gaussiana,  $E[d^4] = 3\sigma_d^4$  e quindi:

$$\begin{aligned} \sigma_g^2 &= \sigma_g^2 (1 + 3\gamma^2 \sigma_d^4 + 2\gamma \sigma_d^2) + \gamma^2 \sigma_d^2 \sigma_s^2 \\ \sigma_g^2 (2 - 3\gamma \sigma_d^2) &= \gamma \sigma_s^2 \\ \sigma_g^2 &= \frac{\gamma \sigma_s^2}{2 - 3\gamma \sigma_d^2} \sim \frac{\gamma \sigma_s^2}{2} ; \text{ se } \gamma \ll \ll \frac{2}{3\sigma_d^2} < \frac{2}{\sigma_d^2} \end{aligned} \quad (4.23)$$

A regime, il valore atteso del filtro errore è nullo, ma non il suo valore quadratico medio, che cresce con  $\gamma$ . E' possibile ora fare un'analisi di un modo sistematico con cui si deve ridurre il passo di aggiornamento, a partire dall'equazione (4.21) in cui però approssimiamo  $d_i^2$  con il suo valore atteso:

$$g_{i+1} = g_i (1 - \gamma \sigma_d^2) + \gamma d_i s_i = \rho g_i + \mu_i$$

### 4.3 RAPIDITÀ DELLA CONVERGENZA

Questa è una semplice equazione alle differenze finite che, sovrapponendo gli effetti dell'ingresso  $\mu_i$  e della condizione iniziale  $g_0$ , variabili casuali indipendenti, ci dà che, se  $\gamma$  è fisso, la dispersione  $\sigma_g^2$  varierà nel tempo come:

$$\begin{aligned}\sigma_{g,i}^2 &\sim \rho^{2i} \sigma_{g,0}^2 + \frac{\sigma_\mu^2}{1-\rho^2} \sim \sigma_{g,0}^2 e^{-2i\gamma\sigma_d^2} + \frac{\gamma\sigma_s^2}{2}; \\ \rho^{2i} &= e^{2i\log(1-\gamma\sigma_d^2)} \sim e^{-2i\gamma\sigma_d^2} \\ \frac{\sigma_\mu^2}{1-\rho^2} &= \frac{\gamma^2\sigma_s^2\sigma_d^2}{(1-\rho)(1+\rho)} \sim \frac{\gamma^2\sigma_s^2\sigma_d^2}{\gamma\sigma_d^2(1+\rho)} \sim \frac{\gamma\sigma_s^2}{2}\end{aligned}\quad (4.24)$$

Quindi, per valori crescenti di  $\gamma$ , cresce sia la rapidità di convergenza che l'asintoto. L'involuppo inferiore di queste curve parametrizzate da  $\gamma$  si trova annullando la derivata delle curve rispetto al parametro (l'intersezione tra due curve caratterizzate da valori molto vicini del parametro) e quindi eliminandolo.

$$\frac{\partial \sigma_{g,i}^2}{\partial \gamma} = -2i\sigma_d^2\sigma_{g,0}^2 e^{-2i\gamma\sigma_d^2} + \frac{\sigma_s^2}{2} = 0$$

Ora  $\gamma$  è funzione dell'indice  $i$  :

$$\gamma(i) = \frac{1}{2\sigma_d^2 i} \ln \left( \frac{\sigma_d^2}{\sigma_s^2} 4\sigma_{g,0}^2 i \right)$$

e il risultato ottimizzato è quindi:

$$\sigma_{g,i}^2 \sim \frac{1}{i} \frac{\sigma_s^2}{4\sigma_d^2} \left[ 1 + \ln \left( 4i\sigma_{g,0}^2 \frac{\sigma_d^2}{\sigma_s^2} \right) \right];$$

Si comprende come, all'inizio dell'adattamento,  $\gamma$  debba essere vicino al limite massimo ma poi debba scendere a valori molto più bassi non appena il sistema si avvicina all'equilibrio, proprio per ridurre  $\sigma_g^2$ . Inoltre, per avere stabilità del filtro errore oltre che in media (vedi 4.17) anche in dispersione, il passo  $\gamma$  deve essere ben più piccolo del valore trovato in base all'equazione di evoluzione delle medie. Infine, se il sistema parte da condizioni iniziali ignote, ma poi è stazionario, il valore quadratico medio dell'errore con cui viene determinato il filtro deve potersi ridurre con il tempo in modo approssimativamente iperbolico.

Aumentando il livello di complessità e passando ad un'analisi multidimensionale ( $N > 1$ ), possiamo quindi analizzare il filtro errore  $\underline{\mathbf{g}}_i$

## CAPITOLO 4 FILTRAGGIO ADATTATIVO

(e cioè la differenza tra il filtro effettivo  $\underline{\mathbf{f}}_i$  (vettore casuale) ed il filtro desiderato a regime,  $\underline{\mathbf{f}}_\infty$ , vettore deterministico) come un vettore casuale di cui si cerca media e varianza, al variare del tempo. La situazione non è stazionaria, in quanto la media (di insieme) del filtro errore convergerà a 0, sempre se  $\gamma < \gamma_{max}$ . Non così la sua dispersione statistica, come vedremo nel seguito. Anzitutto, riscriviamo l'equazione di aggiornamento del filtro  $\underline{\mathbf{f}}_i$ , visto come un vettore casuale:

$$\underline{\mathbf{f}}_{i+1} = \underline{\mathbf{f}}_i + \gamma \varepsilon_i^* \underline{\mathbf{d}}_i = \underline{\mathbf{f}}_i + \gamma (y_i^* - \underline{\mathbf{d}}_i^* \underline{\mathbf{f}}_i) \underline{\mathbf{d}}_i \quad (4.25)$$

L'errore di stima del disturbo  $\varepsilon_i$  è somma di due termini, l'uno dovuto al segnale sovrapposto  $s_i$  e l'altro al disturbo residuo dovuto all'imperfetto filtraggio, perchè il filtro errore  $\underline{\mathbf{g}}_i$ , anche esso casuale, non è nullo:

$$\begin{aligned} \varepsilon_i &= s_i + n_i - \hat{n}_i = y_i - \underline{\mathbf{d}}_i^* \underline{\mathbf{f}}_i = y_i - \underline{\mathbf{d}}_i^* (\underline{\mathbf{g}}_i + \underline{\mathbf{f}}_\infty) = \\ &= y_i - \underline{\mathbf{d}}_i^* \underline{\mathbf{f}}_\infty - \underline{\mathbf{d}}_i^* \underline{\mathbf{g}}_i \simeq s_i - \underline{\mathbf{d}}_i^* \underline{\mathbf{g}}_i \end{aligned} \quad (4.26)$$

sempre che si supponga  $n_i \simeq \underline{\mathbf{d}}_i^* \underline{\mathbf{f}}_\infty$  e cioè che il filtro a regime  $\underline{\mathbf{f}}_\infty$  sia efficace. Per ottenere dei risultati semplici, dobbiamo fare delle ipotesi semplificative, che tuttavia non oscurano la fisica del problema. In particolare supponiamo che, da un istante al successivo, i vettori  $\underline{\mathbf{d}}_i$  siano indipendenti tra loro e dal segnale  $s_i$ .

$$E[\underline{\mathbf{d}}_i \underline{\mathbf{d}}_j^*] = \underline{\mathbf{0}}; \quad i \neq j$$

Inoltre, per semplificare le formule, supporremo che le componenti di  $\underline{\mathbf{d}}_i$  siano indipendenti tra loro e pertanto che la matrice  $\underline{\mathbf{R}}$  sia proporzionale alla matrice identica:

$$\underline{\mathbf{R}} = E[\underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*] = \sigma_d^2 \underline{\mathbf{I}}$$

Calcoleremo la media di insieme del vettore  $\underline{\mathbf{g}}_{i+1}$ ,

$$\underline{\mathbf{g}}_{i+1} = (\underline{\mathbf{I}} - \gamma \underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*) \underline{\mathbf{g}}_i + \gamma \underline{\mathbf{d}}_i s_i \quad (4.27)$$

tenendo conto dell'indipendenza tra il segnale  $s_i$  ed il disturbo  $\underline{\mathbf{d}}_i$ .

Prima però di passare a qualsiasi operazione di media, osserviamo che la matrice casuale  $\underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*$  o una sua media mobile a pesi positivi  $\sum_i p_i \underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*$  è sempre semidefinita positiva in quanto per qualsiasi vettore  $\underline{\mathbf{a}}$ ,  $\underline{\mathbf{d}}_i$ , si ha

$$\sum p_i (\underline{\mathbf{a}}^* \underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^* \underline{\mathbf{a}}) = \sum p_i (\underline{\mathbf{d}}_i^* \underline{\mathbf{a}})^2 \geq 0 \quad \text{se } p_i \geq 0$$

### 4.3 RAPIDITÀ DELLA CONVERGENZA

Questo è un fatto importante perchè non permette il cambio di segno del  $\gamma$  nella proiezione su qualsiasi autovettore della matrice  $\underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*$  e quindi possibili instabilità.

Passando alle medie di insieme:

$$\begin{aligned} E[\underline{\mathbf{g}}_{i+1}] &= E[(\underline{\mathbf{I}} - \gamma \underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*) \underline{\mathbf{g}}_i] + E[\gamma \underline{\mathbf{d}}_i s_i] = \\ &= (\underline{\mathbf{I}} - \gamma \underline{\mathbf{R}}) E[\underline{\mathbf{g}}_i] = (\underline{\mathbf{I}} - \gamma \underline{\mathbf{R}})^{i+1} \underline{\mathbf{g}}_0 \end{aligned} \quad (4.28)$$

In quest'ultima equazione abbiamo utilizzato l'ipotesi che ogni nuovo valore di  $\underline{\mathbf{d}}_i$  è indipendente dai precedenti e quindi anche indipendente da  $\underline{\mathbf{g}}_i$ . Quindi la media del loro prodotto è il prodotto delle medie. Ancora, si è supposto che il valore iniziale del filtro errore  $\underline{\mathbf{g}}_0$  non sia una variabile casuale; esso è peraltro sconosciuto.

La **media di insieme** del filtro errore pertanto si evolve e converge a zero secondo l'equazione determinata precedentemente (4.17) utilizzando l'ipotesi di lenta variazione del filtro ( $\gamma$  piccolo) e stazionarietà delle statistiche del disturbo. Calcoliamo ora il valore quadratico medio del filtro errore; di nuovo, per evitare complicazioni ulteriori, calcoliamo questo v.q.m. a regime ( $i = \infty$ ), e cioè quando la media di insieme del filtro errore ha raggiunto il valore nullo, desiderato. La varianza a regime pertanto corrisponde al valore quadratico medio delle componenti, a regime. Utilizziamo quindi la formula (4.27) di cui calcoliamo il valore quadratico medio di entrambi i membri, utilizzando l'indipendenza tra  $\underline{\mathbf{d}}_i$ ,  $\underline{\mathbf{g}}_i$ ,  $s_i$  e tralasciando l'indice ( $i = \infty$ ), per snellire. Inoltre, riscriviamo l'equazione di aggiornamento **non sostituendo** immediatamente all'espressione  $\underline{\mathbf{d}}_i \underline{\mathbf{d}}_i^*$  la sua media di insieme e cioè la matrice  $\underline{\mathbf{R}}$ , proprio per tenere conto della casualità dell'aggiornamento:

$$\begin{aligned} E[\underline{\mathbf{g}}^* \underline{\mathbf{g}}] &= E[\underline{\mathbf{g}}^* (\underline{\mathbf{I}} - \gamma \underline{\mathbf{d}} \underline{\mathbf{d}}^*) (\underline{\mathbf{I}} - \gamma \underline{\mathbf{d}} \underline{\mathbf{d}}^*) \underline{\mathbf{g}}] + E[\gamma^2 s^2 \underline{\mathbf{d}}^* \underline{\mathbf{d}}] = \\ &= E[\underline{\mathbf{g}}^* (\underline{\mathbf{I}} - 2\gamma \underline{\mathbf{d}} \underline{\mathbf{d}}^* + \gamma^2 \underline{\mathbf{d}} \underline{\mathbf{d}}^* \underline{\mathbf{d}} \underline{\mathbf{d}}^*) \underline{\mathbf{g}}] + E[\gamma^2 s^2 \underline{\mathbf{d}}^* \underline{\mathbf{d}}] \end{aligned} \quad (4.29)$$

e quindi:

$$N\sigma_g^2 = N\sigma_g^2(1 - 2\gamma\sigma_d^2 + \gamma^2(N + K - 1)\sigma_d^4) + \gamma^2 N\sigma_d^2\sigma_s^2$$

Infatti,

$$E[\underline{\mathbf{d}} \underline{\mathbf{d}}^*] = \sigma_d^2 \underline{\mathbf{I}}$$



## CAPITOLO 4 FILTRAGGIO ADATTATIVO

e, per l'indipendenza tra  $\underline{\mathbf{d}}_i, \underline{\mathbf{g}}_i$ :

$$E[\underline{\mathbf{g}}^* \underline{\mathbf{d}} \underline{\mathbf{d}}^* \underline{\mathbf{g}}] = E[\underline{\mathbf{g}}^* E[\underline{\mathbf{d}} \underline{\mathbf{d}}^*] \underline{\mathbf{g}}] = N \sigma_g^2 \sigma_d^2$$

$$E[\underline{\mathbf{d}} \underline{\mathbf{d}}^* \underline{\mathbf{d}} \underline{\mathbf{d}}^*] = E[\underline{\mathbf{d}}^* \underline{\mathbf{d}} \times \underline{\mathbf{d}} \underline{\mathbf{d}}^*] = (N + K - 1) \sigma_d^4 \underline{\mathbf{I}}$$

il  $k$ -esimo elemento della diagonale principale dell'ultima matrice è eguale a:

$$E\left[\sum_{i=1}^N d_i^2 \times d_k^2\right] = (N - 1) \sigma_d^4 + K \sigma_d^4$$

Abbiamo indicato con  $K$  il rapporto tra il valore medio della quarta potenza e il quadrato della varianza:

$$K = \frac{E[d_k^4]}{\sigma_d^4} \quad (4.30)$$

$K$  è la **curtosi** di una variabile casuale a valor medio nullo e per variabili gaussiane  $K = 3$ . Quindi, in condizioni di regime, lontano dai transitori, quando la varianza  $\sigma_g^2$  del filtro errore è dovuta **solo** all'effetto della varianza  $\sigma_s^2$  del segnale  $\{s_i\}$  che provoca aggiornamenti casuali del filtro, si ha:

$$\sigma_g^2 = \frac{\gamma \sigma_s^2}{2 - \gamma(N + K - 1) \sigma_d^2} \quad (4.31)$$

Se  $\gamma$  è piccolo,

$$\sigma_g^2 \simeq \frac{\gamma \sigma_s^2}{2} \quad (4.32)$$

**In sostanza, a regime, la varianza del filtro errore cresce con la varianza del segnale  $\sigma_s$  e con il passo di aggiornamento.** Quindi, una volta raggiunta in pratica la convergenza,  $\gamma$  va ridotto. Inoltre, la dispersione del filtro errore può essere illimitata ( $\sigma_g^2 \rightarrow \infty$ ), se non si verifica che:

$$\gamma \leq \frac{2}{(N + K - 1) \sigma_d^2} < \frac{2}{N \sigma_d^2}$$

In sostanza, ci sono valori del passo di aggiornamento  $\gamma$  per cui il filtro errore convergerebbe (in media) a zero, ma poi ha una dispersione enorme.

## 4.4 Ottimizzazione del passo di aggiornamento

Riassumendo, se il ritmo di aggiornamento del filtro è veloce ( $\gamma \leq \gamma_{\max}$ ), i parametri del filtro si aggiornano rapidamente (*le signore alla moda si vestono in modo ricercato e sempre variabile*), ma poi, a regime, i parametri sono affetti da disturbo (*le persone troppo chic si vestono in modi strani..*). Se invece il filtro si aggiorna *lentamente*, a regime sarà poco disturbato (*le vere signore vestono in modo classico...*). Ne risulta inoltre un altro limite per il passo: infatti, come si è visto:

$$\gamma < \gamma_{\max} = \frac{2}{(N + K - 1)\sigma_d^2} \quad (4.33)$$

Se  $N \gg 1$ , ci si potrebbe avvicinare alla condizione di massima velocità di aggiornamento del filtro media di insieme ( $\gamma = \gamma_{\max}$ ); però, la dispersione a regime del filtro errore tenderebbe ad infinito a causa dell'inevitabile presenza del segnale ( $\sigma_s^2 \neq 0$ ) che impedisce al filtro di inseguire solo il disturbo. Se  $\gamma \ll \gamma_{\max}$ , il filtro si aggiorna molto lentamente ma poi a regime la dispersione è minore perché si è meno sensibili all'influenza del segnale casuale mediata e rimossa a causa del limitatissimo passo di aggiornamento.

Pertanto, all'inizio dell'aggiornamento, quando si è lontani dalla situazione di regime (il filtro errore ha valori grandi), sarà opportuno porre  $\gamma \simeq \gamma_{\max}/2$  per accelerare la convergenza anche a costo di un'elevata dispersione a regime; appena il filtro errore si è ridotto, bisogna ridurre  $\gamma$  per evitare inutili dispersioni dei coefficienti del filtro attorno ad un valore medio vicino a valori già buoni. Gergalmente, in inglese, questo modo di operare è chiamato *gear shifting: scalare le marce*; l'aggiornamento deve essere veloce all'inizio del transitorio e lento alla fine, quando si è prossimi alla situazione asintotica. Si osservi che nei periodi iniziali del transitorio (dopo una variazione delle statistiche), ci si attende un cambiamento celere (*rivoluzione*). La media del filtro errore è grande e c'è grande differenza tra il comportamento atteso e quello effettivo del filtro. Sono possibili variazioni repentine del filtro, apparentemente casuali ( $\gamma \simeq \gamma_{\max} \rightarrow \sigma_g^2 \simeq \infty$ ). Nelle situazioni lontane dai transitori (*conservazione*) il sistema si adegua ordinatamente ma forse troppo lentamente ( $\gamma \ll \gamma_{\max} \rightarrow \sigma_g^2 \simeq 0$ ).

## CAPITOLO 4 FILTRAGGIO ADATTATIVO

Se poi le statistiche del sistema cambiano continuamente, ci sarà un valore ottimale di  $\gamma$  dipendente sia dalla velocità di variazione delle statistiche che dall'ampiezza del rumore. In generale, il valore ottimo per  $\gamma$  sarà quello che minimizza il valore atteso (quadratico medio) del filtro errore: *se il traffico è irregolare, è bene rimanere in seconda ...*

Si osservi che per  $N = 1$  la situazione è peggiore, e si deve mantenere  $\gamma \ll \gamma_{\max}/3 < \gamma_{\max}$ , nel caso di variabili gaussiane. Infatti, sono possibili valori istantanei elevati del disturbo che mandano il sistema in instabilità. Non appena  $N$  cresce e l'aggiornamento è basato su stime fatte su un numero maggiore di variabili casuali indipendenti, c'è minore dispersione della stima, la situazione si stabilizza immediatamente, e  $\gamma$  può essere accresciuto. Questo effetto di instabilità dipende dalla possibilità di avere forti picchi istantanei in qualche componente  $d_{i,n}$  di  $\underline{\mathbf{d}}_i$ ; si ricordi che il fattore  $(N + K - 1)/N$  per le variabili gaussiane vale  $(N + 2)/N$  e vale 3 per  $N = 1$ ; la situazione peggiora ancora con distribuzioni con curtosi più elevata che non la gaussiana (e cioè se vi è probabilità significativa di avere picchi elevati); infatti, sono possibili densità di probabilità con varianza finita e curtosi illimitata. La situazione invece migliora per variabili casuali *più tranquille* e cioè caratterizzate da distribuzioni con code limitate o addirittura assenti come accade per la distribuzione rettangolare per la quale  $K = 9/5 < 3$ ; il valore minimo possibile per la curtosi,  $K = 1$ , è raggiunto solo nel caso di una variabile casuale che assume solo i valori  $\pm q$ . In questo caso e solo in questo caso si può mantenere  $\gamma \leq 1$  anche per  $N = 1$ ; in altre parole, il segnale di pilotaggio è così tranquillo che si può rimanere sempre *con le marce alte*; quest'ultimo non è un caso irrilevante in quanto corrisponde ad un segnale di comunicazioni del tipo BPSK.

## Capitolo 5

### Le direzioni di arrivo

Fino ad ora abbiamo visto l'analisi spettrale di sequenze in cui l'indice rappresenta l'asse temporale; studieremo ora situazioni in cui l'indice si riferisce a posizioni nello spazio. In particolare vogliamo affrontare il problema della misura delle Direzioni di Arrivo su una schiera di sensori (Directions of Arrival: DOA) di segnali temporalmente monofrequenziali (o a spettro molto limitato).

Si disponga di un cavo lungo il quale sono regolarmente disposti dei sensori e siano  $\{x_i\}$ ,  $i = 1, \dots, N$ , i segnali da essi misurati. La distanza tra il sensore  $i$ -esimo e l'inizio del cavo sia  $d_i = (i - 1)\Delta$ ; supponiamo che esistano  $M$  sorgenti di radiazioni, monofrequenziali o a banda stretta, a grande distanza dal cavo; l'angolo tra la direzione della sorgente  $m$  e la normale al cavo (vedi fig. 14.1) è indicato con  $\theta_m$ ; quindi la radiazione arriva al ricevitore  $\{x_i\}$  con il ritardo rispetto al ricevitore  $i = 1$  pari a  $\tau_i = (i - 1) \times \Delta \sin \theta_m / c$ . Inoltre, indichiamo con  $A_m e^{j\omega n T}$  l'ampiezza sinusoidale della radiazione e la sua evoluzione nel tempo, e con  $c$  la velocità di propagazione delle onde. Il segnale misurato dal sensore  $i$  all'istante  $n$  è (per semplicità abbiamo posto  $M = 1$ ):

$$x_{i,n} = A \exp(j\omega n T + j \frac{(i - 1)\omega \Delta \sin \theta_m}{c}) = A e^{j\omega n T} \times e^{jk_m(i-1)\Delta}; \quad (5.1)$$

$$k_m = \frac{\omega \sin \theta_m}{c}$$

Il segnale misurato nella stessa posizione al variare del tempo è una sinusoidale complessa di pulsazione  $\omega$ ; il segnale misurato allo stesso istante

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

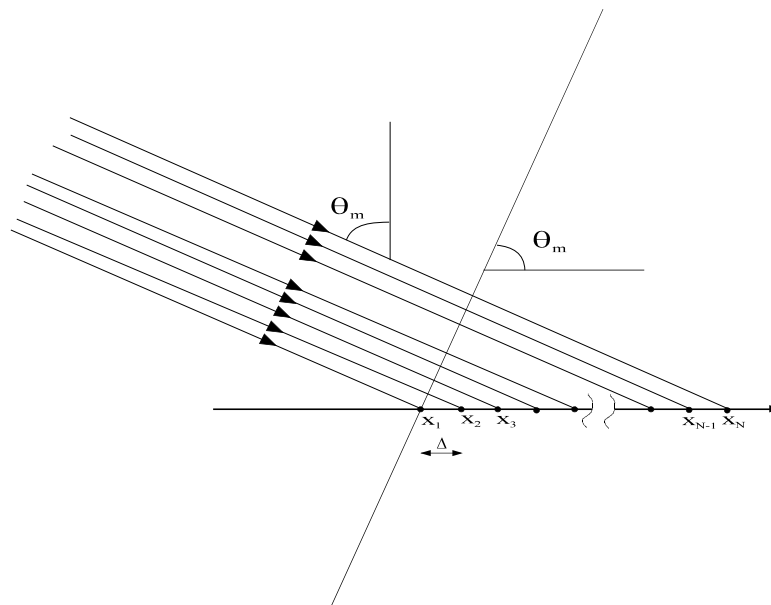


Figura 5.1: Schiera di sensori

$nT$  da tutta la schiera di sensori (al variare della posizione nello spazio) è una **sinusoide complessa**, nel dominio degli spazi, avente numero d'onda  $k_m$  (per le sinusoidi spaziali, è la grandezza corrispondente alla pulsazione delle sinusoidi temporali). Misurare la direzione di arrivo equivale a misurare il numero d'onda della sinusoide complessa registrata.

Fare un'analisi spettrale della sequenza  $\{x_i\}$ , per un dato istante di tempo  $n$ , corrisponde quindi a determinare numero e direzione di arrivo delle radiazioni incidenti sul cavo. Si osservi che se la radiazioni in arrivo non sono monofrequenziali ma la loro banda è sufficientemente ampia, è opportuno farne una decomposizione di Fourier lungo l'asse dei tempi, sensore per sensore, e poi analizzare ogni singola componente isofrequenziale nel dominio degli spazi. In questo modo si avranno più misure (alle varie frequenze temporali) della stessa direzione di arrivo, misure che potranno essere opportunamente combinate.

Si osservi ancora che in questo caso, l'ipotesi di avere segnali sinusoidali (nello spazio) è sostanzialmente corretta. Le tecniche che utilizzeremo per la misura della frequenza ed ampiezza di queste sinusoidi saranno quella del periodogramma e quella AR. Ve ne sono altre (Pisarenko, etc.

## 5.1 DOA CON IL PERIODOGRAMMA

) che non verranno discusse.

La distanza massima tra i sensori per evitare equivocazione spaziale è:

$$|k| = \frac{\omega \sin \theta}{c} \leq k_{Nyquist} = \frac{\pi}{\Delta}; \quad \Delta \leq \frac{c}{2f} = \frac{\lambda}{2} \quad (5.2)$$

Per esempio, con segnali a  $20kHz$ , se  $c = 1500m/sec$ , allora  $\Delta \leq 0.0375m$ . Scegliendo questo valore per  $\Delta$  avremo che alla direzione  $\theta = 0$  corrisponde  $k = 0$ ; alle direzioni  $\theta = \pm\pi/2$  corrisponde  $k = \pm k_{Nyquist}$

### 5.1 DOA con il periodogramma

Determineremo ora l'accuratezza della misura delle direzioni di arrivo, utilizzando per l'analisi spettrale la tecnica del periodogramma. In questo caso, indicando con  $L_{tot} = N\Delta$  la lunghezza del cavo la risoluzione spettrale è

$$\delta k = \frac{2\pi}{L_{tot}} \quad (5.3)$$

Ad un incertezza di numero d'onda  $\Delta k$  corrisponde un'incertezza

$$\delta k = \frac{\omega \cos \theta}{c} \delta \theta; \quad \delta \theta = \frac{\lambda}{L_{tot} \cos \theta} \quad (5.4)$$

Questa è la direttività di un'antenna lunga  $L_{tot} \cos \theta$  pari alla proiezione della lunghezza del cavo nella direzione ortogonale alla radiazione. In sostanza, utilizzare il periodogramma per fare l'analisi spettrale coincide con il combinare assieme i segnali in un'antenna sintetica lunga quanto il cavo. Si osservi che l'uso di questa tecnica corrisponde all'ipotesi che radiazioni provenienti da direzioni diverse (potenze in celle disgiunte del periodogramma) siano tra loro indipendenti. Se l'analisi fosse bidimensionale, questa tecnica produrrebbe quindi un'immagine dell'energia illuminante suddivisa per direzioni angolari di arrivo: come un telescopio, pertanto. Infatti, una lente sottile (o uno specchio ricurvo opportunamente) funziona come un trasformatore di Fourier (a 2 dimensioni): l'immagine nel piano focale è la trasformata di Fourier del segnale in arrivo sulla lente. Quindi, per leggere un titolo di giornale ( $5cm$  di risoluzione) da  $300 km$  di distanza serve un telescopio di diametro almeno eguale a:

$$\frac{0.05}{300000} = \frac{\lambda}{D}; \text{ Per il colore verde: } \lambda = 0.5\mu m \rightarrow D = 3m$$

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

Ma la risoluzione può crescere di molto se si possono imporre ipotesi stringenti sul numero di sorgenti, come vedremo nella prossima sezione.

### 5.2 Stima AR: la risoluzione angolare

L'analisi spettrale che si adatta bene al caso in cui si ha un numero limitato di sorgenti lontane è quella parametrica, tutti poli. Infatti il segnale in arrivo è caratterizzato da un numero discreto  $M$  di fronti d'onda, ognuno con la sua ben definita direzione di arrivo. La sequenza  $\{x_i\}$  sarà somma di  $M$  sinusoidi complesse corrispondenti agli  $M$  fronti. Le loro pulsazioni spaziali (i numeri d'onda) sono:

$$k_m = \frac{\omega \sin \theta_m}{c} ; \quad m = 1, \dots, M \quad (5.5)$$

Supporremo  $M = 1$ , che l'ampiezza della radiazione  $A$  sia una variabile casuale complessa di varianza  $E[AA^*] = \sigma_A^2$  e che sui dati sia sovrapposto del rumore bianco  $w_i$ , complesso, di varianza  $\sigma_w^2$ . Non riportiamo il fattore complesso  $e^{j\omega nT}$  avente modulo unitario, comune a tutti i dati. Ponendo quindi:

$$x_i = Ae^{jk_1(i-1)\Delta} + w_i$$

si vede che l'analisi spettrale AR si presta bene, anche se poi non esiste una vera e propria corrispondenza fisica diretta e cioè non siamo nel caso di un segnale generato da un rumore bianco che pilota un filtro IIR.

Per determinare la direzione della sorgente con l'analisi spettrale AR abbiamo bisogno dei campioni della funzione di autocorrelazione spaziale della sequenza  $\{x_i\}$  che sono:

$$E[x_i x_{i-n}^*] = r_n = r_{-n}^* = \sigma_A^2 e^{jk_1 n \Delta} + \sigma_w^2 \delta(n); 0 \leq n \leq N \quad (5.6)$$

Si osservi che possiamo stimare con precisione i campioni della funzione di autocorrelazione facendo delle operazioni di media **nel tempo**; non abbiamo necessità di fare delle medie nello spazio.

La matrice di covarianza  $\mathbf{R}$  è Hermitiana (la trasposta coincide con la coniugata) ed è in questo caso rappresentabile come combinazione lineare di una matrice identica e di una diade, prodotto esterno di due vettori complessi. Si verifica che la matrice inversa è ancora combinazione lineare della matrice identica e della stessa diade. Nel caso più generale  $M \neq 1$ ,

## 5.2 STIMA AR: LA RISOLUZIONE ANGOLARE

e cioè con  $M$  sorgenti indipendenti, la matrice  $\mathbf{R}$  sarebbe somma di  $M$  diadi e la soluzione analitica del problema richiederebbe la soluzione di una matrice  $M \times M$ . Ovviamente, il problema delle direzioni di arrivo è solubile solo se  $M \leq N$ , ed in generale  $M \ll N$ . Lo spettro AR è determinato dalle relazioni di Yule Walker: Per  $M = 1$  la soluzione analitica è semplice; si osservi che in assenza di rumore, l'ordine  $N$  del sistema necessario per misurarne la frequenza potrebbe essere limitato ad 1, (basterebbero due sensori, quindi) in quanto una sinusoidale complessa corrisponde ad un sistema ad un solo polo. Tuttavia, la presenza del disturbo  $\{w_i\}$  fa sì che l'errore di predizione decresca progressivamente con  $N$ , come vedremo, rendendo utili schiere di più sensori. Sia quindi al solito:

$$\mathbf{R}\mathbf{a}_N = -\mathbf{r}_N$$

$$\mathbf{R} = \begin{bmatrix} r_0 & r_{-1} & r_{-2} & \dots & r_{-N+1} \\ r_1 & r_0 & r_{-1} & \dots & r_{-N+2} \\ r_2 & r_1 & r_0 & \dots & r_{-N+3} \\ \dots & \dots & \dots & \dots & \dots \\ r_{N-1} & r_{N-2} & r_{N-3} & \dots & r_0 \end{bmatrix} = \sigma_A^2 \mathbf{c}\mathbf{c}^* + \sigma_w^2 \mathbf{I}; \quad (5.7)$$

$$\mathbf{c} = \begin{bmatrix} e^{jk_1\Delta} \\ e^{j2k_1\Delta} \\ e^{j3k_1\Delta} \\ \dots \\ e^{jNk_1\Delta} \end{bmatrix}; \quad \mathbf{r}_N = \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \dots \\ r_N \end{bmatrix} = \sigma_A^2 \mathbf{c} \quad (5.8)$$

$$\mathbf{R}^{-1} = \frac{1}{\sigma_w^2} \mathbf{I} - \frac{\sigma_A^2}{\sigma_w^2(\sigma_w^2 + \sigma_A^2 \mathbf{c}^* \mathbf{c})} \mathbf{c}\mathbf{c}^* \quad (5.9)$$

In un caso più generale:

$$(\mathbf{A}^{-1} + \mathbf{s}\mathbf{s}^*)^{-1} = \mathbf{A}^{-1} - \frac{\mathbf{A}^{-1} \mathbf{s}\mathbf{s}^* \mathbf{A}^{-1}}{1 + \mathbf{s} \mathbf{A}^{-1} \mathbf{s}^*}$$

Inoltre, indicando con  $\mathbf{a}_N$  il vettore incognito delle equazioni di Yule Walker (vedi al Capitolo sull'Analisi spettrale parametrica) e cioè lo



## CAPITOLO 5 LE DIREZIONI DI ARRIVO

stimatore ottimo del segnale:

$$\underline{\mathbf{a}}_N = -\underline{\mathbf{R}}^{-1}\underline{\mathbf{r}}_N = -\left(\frac{\sigma_A}{\sigma_w^2 + N\sigma_A^2}\right)\underline{\mathbf{c}} \quad (5.10)$$

poichè

$$\underline{\mathbf{c}}^*\underline{\mathbf{c}} = N \quad (5.11)$$

Risulta infine che la soluzione è:

$$a_{n,N} = -\frac{\sigma_A^2}{\sigma_w^2 + N\sigma_A^2} e^{jk_1 n \Delta} \quad (5.12)$$

Il denominatore della funzione di trasferimento che viene identificata con il metodo autoregressivo è:

$$A(z) = 1 + \sum_{n=1}^N a_{n,N} z^{-n}$$

La stima dello spettro di potenza unilatero del segnale visto dalla schiera dei sensori è ottenuta dallo spettro dell'errore di predizione, di cui è nota la potenza ma non il colore e pertanto è supposto bianco, spettro che poi deve essere diviso per la funzione di trasferimento del predittore. La potenza dell'errore di predizione è quella del segnale di ingresso, ridotta del fattore dipendente dagli  $N$  coefficienti di riflessione  $c_n$  ;  $n = 1 : N$  e, per  $\sigma_A^2 \gg \sigma_w^2$ :

$$c_N = a_{N,N} = -\frac{\sigma_A^2}{\sigma_w^2 + N\sigma_A^2} e^{jk_1 N \Delta} ; \quad N = 1, \dots \quad (5.13)$$

$$1 - |c_1|^2 = 1 - \left(\frac{\sigma_A^2}{\sigma_w^2 + 1 \cdot \sigma_A^2}\right)^2 = \frac{2\sigma_w^2\sigma_A^2 + \sigma_w^4}{(\sigma_w^2 + \sigma_A^2)^2} \sim \frac{2\sigma_w^2}{\sigma_A^2}$$

$$\text{Per } N > 1; \quad 1 - |c_N|^2 = 1 - \left(\frac{\sigma_A^2}{\sigma_w^2 + N\sigma_A^2}\right)^2 \sim 1 - \frac{1}{N^2}$$

Pertanto si ha:

$$(1 - |c_1|^2)(1 - |c_2|^2)(1 - |c_3|^2) \dots = \quad (5.14)$$

$$\sim \frac{2\sigma_w^2}{\sigma_A^2} \left(\frac{3}{4}\right) \left(\frac{8}{9}\right) \dots \quad (5.15)$$

## 5.2 STIMA AR: LA RISOLUZIONE ANGOLARE

Poichè:

$$\prod_{i=2}^{\infty} \left(1 - \frac{1}{i^2}\right) = \frac{1}{2}$$

L'errore di predizione q.m. è:

$$E[|\varepsilon_{i,N}|^2] = (\sigma_A^2 + \sigma_w^2) \prod_{n=1}^N (1 - |c_n|^2) \quad (5.16)$$

è pari a circa  $2\sigma_w^2$  per  $N = 1$  e si riduce a  $\sigma_w^2$  per  $N$  crescenti. Lo spettro stimato è:

$$\hat{S}(k) = \frac{2\Delta(\sigma_A^2 + \sigma_w^2) \prod_{n=1}^N (1 - |c_n|^2)}{\left|1 + \sum_{n=1}^N a_{n,N} e^{-jkn\Delta}\right|^2} \quad (5.17)$$

$\hat{S}(k)$  ha un massimo al numero d'onda  $k_1$  dove vale:

$$\begin{aligned} \hat{S}(k_1) &= \frac{2\Delta(\sigma_A^2 + \sigma_w^2) \prod_{n=1}^N (1 - |c_n|^2)}{\left|1 - \frac{N\sigma_A^2}{\sigma_w^2 + N\sigma_A^2}\right|^2} = \\ &= \frac{2\Delta(\sigma_A^2 + \sigma_w^2) \prod_{n=1}^N (1 - |c_n|^2)}{\left(\frac{\sigma_w^2}{\sigma_w^2 + N\sigma_A^2}\right)^2} \sim \Delta \frac{N^2 \sigma_A^4}{\sigma_w^2} \end{aligned}$$

Il coefficiente passa da 4 per  $N = 1$  a 2 per  $N \rightarrow \infty$ . Si osservi che questo massimo cresce con  $N^2$  e con il rapporto segnale rumore; peraltro, nello spettro vero troveremmo un impulso ideale. La qualità della stima deriva però dall'elevata risoluzione spettrale che ora calcoliamo. Infatti, l'intervallo  $\delta k$  di numeri d'onda intorno a  $k_1$  che corrisponde ad una variazione di 3 dB dell'ampiezza  $\hat{S}(k)$  rispetto al massimo della stima spettrale è, per alto rapporto  $SNR$ :

$$\delta k = \frac{4\sigma_w^2}{N\sigma_A^2} \frac{1}{(N+1)\Delta} \quad (5.18)$$

Ciò si dimostra facilmente sviluppando attorno al suo minimo lo spettro del predittore:

$$1 + \sum_{n=1}^N a_{n,N} e^{-j(k-k_1)n\Delta} = 1 - \sum_{n=1}^N \frac{\sigma_A^2}{\sigma_w^2 + N\sigma_A^2} e^{jn\phi} \sim$$

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

$$= \frac{\sigma_w^2}{\sigma_w^2 + N\sigma_A^2} - j\phi \frac{\sigma_A^2 N(N+1)}{2(\sigma_w^2 + N\sigma_A^2)}; \quad \text{se } |\phi| = |(k - k_1)\Delta| \ll 1$$

$$\phi_{3dB} = \frac{\delta k \Delta}{2} \sim \frac{2\sigma_w^2}{N^2\sigma_A^2}$$

Peraltro la potenza totale nella banda a 3 dB é proprio  $\sigma_A^2$ :

$$P_{est} = \frac{4\sigma_w^2}{N^2\sigma_A^2} \frac{1}{\Delta} \Delta \frac{N^2\sigma_A^4}{\sigma_w^2} \frac{1}{2\pi} = \frac{2}{\pi} \sigma_A^2 \quad (5.19)$$

a meno del fattore  $2/\pi$  già incontrato, dovuto all'approssimazione di spettro uniforme (e non 1 polo) nella banda 3dB. La risoluzione angolare (spettrale) è quindi:

$$\Delta\theta = \delta k \frac{c}{\omega \cos \theta} = \frac{4\sigma_w^2}{N\Delta(N+1)\sigma_A^2} \frac{\lambda}{2\pi \cos \theta} \simeq \quad (5.20)$$

$$= \frac{\lambda}{L_{tot} \cos \theta} \times \frac{2}{\pi} \times \frac{\sigma_w^2}{(N+1)\sigma_A^2} \quad (5.21)$$

tanto migliore rispetto a quanto ottenuto con il periodogramma per quanto più elevato è il rapporto Segnale - Rumore; inoltre cresce con il numero  $N$  di sensori utilizzati. Praticamente, e' inutile fare crescere  $N$  oltre 2 o 3. In questo caso, la sinusoide è ricostruita quasi perfettamente e l'errore residuo è solo il rumore additivo, imprevedibile. Invece, per  $N = 1$ , la sinusoide è ricostruita in modo imperfetto.

Si osservi che abbiamo supposto che l'autocorrelazione sia perfettamente nota. In appendice si studia il caso in cui anche l'autocorrelazione deve essere stimata; allora, si vede che la stima del coefficiente di riflessione dipende anch'essa dalla lunghezza dei dati. Nel caso del periodogramma, la risoluzione angolare è dipendente solo dalla lunghezza della schiera di sensori; tuttavia, potrebbe migliorare cercando di interpolare la posizione del massimo del periodogramma. Allora, il risultato diventerebbe dipendente dal rapporto segnale - rumore e la risoluzione verrebbe a dipendere dal fattore  $N^{-3/2}$ .

### 5.2.1 Stima spettrale di Capon

Un'altra stima spettrale derivata da quella AR è quella di Capon; con essa, si cerca il filtro che pur avendo funzione di trasferimento unitaria

## 5.2 STIMA AR: LA RISOLUZIONE ANGOLARE

alla frequenza  $f$ , abbia in uscita la potenza minima. Quindi, la stima spettrale che ne risulta è sistematicamente non polarizzata, ed ha andamento:

$$\hat{S}_C(k) = \frac{1}{\underline{\mathbf{a}}^* \underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}}$$

$$\underline{\mathbf{a}}^* = [1 \quad e^{jk\Delta} \quad e^{j2k\Delta} \quad \dots \quad e^{j(N-1)k\Delta}]$$

La dimostrazione è semplice: si consideri un filtro

$$\underline{\mathbf{h}}^* = [h_0 \quad h_1 \dots h_N]$$

e il risultato  $\{y_n\}$  della convoluzione della sequenza  $\{x_n\}$  con questo filtro. Indicando con  $\underline{\mathbf{x}}_n$  la sequenza

$$\underline{\mathbf{x}}_n^* = [x_n \quad x_{n-1} \dots x_{n-N+1}]$$

abbiamo

$$y_n = \underline{\mathbf{h}}^* \underline{\mathbf{x}}_n$$

Indicando con  $\underline{\mathbf{R}}$  la matrice di covarianza di  $\underline{\mathbf{x}}_n$  il valore q.m. di  $y_n$  è:

$$E[y_n^2] = E[\underline{\mathbf{h}}^* \underline{\mathbf{x}}_n \underline{\mathbf{x}}_n^* \underline{\mathbf{h}}] = \underline{\mathbf{h}}^* \underline{\mathbf{R}} \underline{\mathbf{h}} \quad (5.22)$$

La risposta in frequenza del filtro  $\underline{\mathbf{h}}$  è:

$$H(k) = \sum_{n=1}^N h_n e^{-jkn\Delta} = \underline{\mathbf{h}}^* \underline{\mathbf{a}}$$

Il metodo di Capon usa un filtro passabanda  $\underline{\mathbf{h}}$  che minimizza la potenza in uscita ( $E[y_n^2]$ ) con il vincolo

$$\underline{\mathbf{h}}^* \underline{\mathbf{a}}(k) = 1$$

Si ha cioè:

$$\underline{\mathbf{h}} = \arg \min_{\underline{\mathbf{h}}} \underline{\mathbf{h}}^* \underline{\mathbf{R}} \underline{\mathbf{h}} \quad \text{con il vincolo} \quad \underline{\mathbf{h}}^* \underline{\mathbf{a}}(k) = 1 \quad (5.23)$$

La funzione di trasferimento del filtro  $\underline{\mathbf{h}}$  è quindi a sua volta funzione del numero d'onda  $k$  alla quale si vuole effettuare la stima della densità spettrale del processo  $\{x_n\}$ . La soluzione della (5.23) è:

$$\underline{\mathbf{h}}(k) = \frac{\underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}}{\underline{\mathbf{a}}^* \underline{\mathbf{R}}^{-1} \underline{\mathbf{a}}}$$

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

La banda di questo filtro si può approssimare con l'inverso della sua lunghezza ( $N\Delta$ ) e quindi sostituendo nella (5.22) si ha:

$$\hat{S}_C(k) = \frac{E[y_n^2]}{1/N\Delta} = N\Delta \left( \frac{\underline{\mathbf{R}}^{-1}\underline{\mathbf{a}}}{\underline{\mathbf{a}}^*\underline{\mathbf{R}}^{-1}\underline{\mathbf{a}}} \right)^* \underline{\mathbf{R}} \left( \frac{\underline{\mathbf{R}}^{-1}\underline{\mathbf{a}}}{\underline{\mathbf{a}}^*\underline{\mathbf{R}}^{-1}\underline{\mathbf{a}}} \right) = \frac{N\Delta}{\underline{\mathbf{a}}^*\underline{\mathbf{R}}^{-1}\underline{\mathbf{a}}}$$

In sostanza, la stima spettrale è fatta utilizzando un banco di filtri diversi, frequenza per frequenza, ottimizzato tenendo conto dello spettro stimato, in modo da minimizzare gli effetti di polarizzazione. La risoluzione spettrale è intermedia tra quella, minima, del periodogramma e quella, massima, della stima spettrale AR.

### 5.3 Reiezione direzionale del disturbo

La tecnica della misura delle DOA può avere varie applicazioni: p.e. nella telefonia cellulare, le stazioni fisse possono localizzare le direzioni dei chiamanti, per concentrare nella loro direzione la direttività delle antenne di ricezione e trasmissione; questo meccanismo potrebbe portare ad un decremento delle interferenze quindi all'accrescimento del numero degli utenti. Ancora, sono facilmente immaginabili applicazioni alla difesa: le sorgenti possono essere sommergibili o aerei di cui si deve determinare l'azimuth.

In ogni caso, vi possono però essere sorgenti vicine che *disturbano* il sensore. La presenza di queste sinusoidi spaziali molto ampie *polarizza* la stima spettrale e la rende imprecisa, come già visto quando abbiamo studiato l'effetto delle finestre su dei segnali sinusoidali. Può essere utile rendere il sistema meno sensibile a queste sorgenti vicine; si adottano quindi tecniche per individuare la direzione da cui proviene il segnale disturbante per cercare poi di annullare in quella direzione la direttività dell'antenna sintetizzata con il sistema di sensori. Nel caso specifico, si può mettere un ulteriore sensore  $x_0$  (vedi fig. 14.1) il cui scopo è quello di registrare la radiazione disturbante. Quindi, si cercherà di decorrelare tutti i sensori rispetto a  $x_0$ ; in altre parole, usando  $x_0$  cercheremo di stimare l'uscita di ogni sensore; la nuova uscita del sensore sarà costituita dall' *errore di predizione*. Così facendo, si eliminerà approssimativamente la radiazione disturbante, raggiungendo lo scopo. Indichiamo il segnale misurato con:

$$x_i = A_1 e^{jk_1 i \Delta} + A_2 e^{jk_2 i \Delta}; \quad i = 0, \dots, N; \quad (5.24)$$

### 5.3 REIEZIONE DIREZIONALE DEL DISTURBO

$$E[|A_1|^2] = \sigma_1^2 \gg E[|A_2|^2] = \sigma_2^2; \quad \mu = \frac{\sigma_2^2}{\sigma_1^2} \ll 1 \quad (5.25)$$

Ampiezza e numero d'onda della radiazione disturbante sono indicati con  $A_1, k_1$ ; quelle della radiazione disturbata con  $A_2, k_2$ . Indichiamo con:

$$y_i = x_i - \alpha_i x_0; \quad i = 1, \dots, N; \quad (5.26)$$

l'errore di predizione dell'uscita del sensore  $i$ -esimo utilizzando come predittore il sensore  $x_0$ ; perciò, imponiamo l'ortogonalità tra  $y_i$  e  $x_0$ :

$$\alpha_i = \frac{E[x_i x_0^*]}{E[|x_0|^2]} \quad (5.27)$$

$$\begin{aligned} \alpha_i &= \frac{\sigma_1^2 e^{jk_1 i \Delta} + \sigma_2^2 e^{jk_2 i \Delta}}{\sigma_1^2 + \sigma_2^2} = \\ &= \frac{e^{jk_1 i \Delta} + \mu e^{jk_2 i \Delta}}{1 + \mu} \quad i = 1, \dots, N; \end{aligned} \quad (5.28)$$

L'errore di predizione è:

$$y_i = B_1 e^{jk_1 i \Delta} + B_2 e^{jk_2 i \Delta} \quad (5.29)$$

$$B_1 = \frac{\mu A_1 - A_2}{1 + \mu}; \quad B_2 = -\frac{\mu A_1 - A_2}{1 + \mu}$$

$$E[B_1 B_1^*] = E[B_2 B_2^*] = -E[B_1 B_2^*]$$

Pertanto, se prima dell'operazione di predizione le ampiezze delle due radiazioni erano fortemente differenti, dopo l'ortogonalizzazione, le ampiezze dei due residui sono simili, e quindi la polarizzazione spettrale sarà trascurabile. Sostanzialmente abbiamo realizzato un sistema che ha un guadagno direzionale con un minimo fortemente pronunciato nella direzione della radiazione disturbante. Se il sistema è adattativo, e quindi le operazioni di media di insieme sono invece realizzate con delle medie temporali con un opportuno fattore di oblio, anche se la direzione dell'onda disturbante varia, la direzione del minimo del guadagno dell'antenna equivalente al sistema si sposterà opportunamente.

## 5.4 Appendice

### 5.4.1 Stima lineare ottima di costanti o pendenze

Consideriamo il seguente processo:

$$\underline{\mathbf{z}} = s\underline{\mathbf{c}} + \underline{\mathbf{m}} \quad (5.30)$$

dove  $\underline{\mathbf{c}}$  è un vettore noto,  $s$ , la grandezza da stimare, uno scalare casuale a valor medio nullo e varianza  $\sigma_s^2$ , ed  $\underline{\mathbf{m}}$  un vettore di rumore casuale a valor medio nullo e matrice di covarianza  $\underline{\mathbf{M}}$ :

$$E[\underline{\mathbf{m}}] = \underline{\mathbf{0}}; \quad E[\underline{\mathbf{m}}\underline{\mathbf{m}}^*] = \underline{\mathbf{M}}$$

Risulta:

$$E[\underline{\mathbf{z}}\underline{\mathbf{z}}^*] = \sigma_s^2 \underline{\mathbf{c}}\underline{\mathbf{c}}^* + \underline{\mathbf{M}}$$

Se si desidera stimare  $s$ , lo stimatore ottimo (ai minimi quadrati) è:

$$\hat{s} = \underline{\mathbf{b}}^* \underline{\mathbf{z}}$$

dove  $\underline{\mathbf{b}}$  è un vettore da calcolare imponendo la minimizzazione dell'errore q. m. di stima e quindi l'ortogonalità tra il vettore dei dati  $\underline{\mathbf{z}}$  e l'errore di stima  $\varepsilon = (s - \hat{s})$ :

$$E[\varepsilon \underline{\mathbf{z}}^*] = E[(s - \hat{s}) \underline{\mathbf{z}}^*] = E[(s - \underline{\mathbf{b}}^* \underline{\mathbf{z}}) \underline{\mathbf{z}}^*] = \underline{\mathbf{0}} \rightarrow$$

$$\underline{\mathbf{b}}^* = E[s \underline{\mathbf{z}}^*] E[\underline{\mathbf{z}}\underline{\mathbf{z}}^*]^{-1}$$

$$\underline{\mathbf{b}} = \sigma_s^2 [\sigma_n^2 \underline{\mathbf{M}} + \sigma_s^2 \underline{\mathbf{c}}\underline{\mathbf{c}}^*]^{-1} \underline{\mathbf{c}}$$

Volendo utilizzare la formula di inversione di una matrice costituita dalla combinazione lineare di una matrice identica ed una diade  $\underline{\mathbf{p}}\underline{\mathbf{p}}^*$ , si veda eq. [?], si deve innanzitutto riportarsi a tale caso premoltiplicando la 5.30 per la matrice

$$\underline{\mathbf{R}} = \underline{\mathbf{M}}^{-1/2}$$

che non è altro che la matrice  $\underline{\mathbf{M}}$ , con gli stessi autovettori, dove però gli autovalori originari  $\lambda_{M,i}$  sono sostituiti con l'inverso della loro radice:

$$\lambda_{R,i} = \frac{1}{\sqrt{\lambda_{M,i}}}$$

## 5.4 APPENDICE

Se la matrice  $\underline{\mathbf{M}}$  é definita positiva, i  $\lambda_{R,i}$  sono reali, positivi e finiti. L'equazione 5.30 diventa:

$$\underline{\mathbf{y}} = \underline{\mathbf{Rz}} = \underline{\mathbf{M}}^{-1/2} (s\underline{\mathbf{c}} + \underline{\mathbf{m}}) = s\underline{\mathbf{p}} + \underline{\mathbf{n}}$$

dove ora però:

$$\underline{\mathbf{p}} = \underline{\mathbf{M}}^{-1/2} \underline{\mathbf{c}}$$

$$E[\underline{\mathbf{nn}}^*] = \underline{\mathbf{M}}^{-1/2} E[\underline{\mathbf{mm}}^*] \underline{\mathbf{M}}^{-1/2} = \underline{\mathbf{M}}^{-1/2} \underline{\mathbf{M}} \underline{\mathbf{M}}^{-1/2} = \underline{\mathbf{I}}$$

Risulta allora:

$$\hat{s} = \underline{\mathbf{a}}^* \underline{\mathbf{y}}$$

$$\underline{\mathbf{a}} = \sigma_s^2 [\underline{\mathbf{I}} + \sigma_s^2 \underline{\mathbf{p}} \underline{\mathbf{p}}^*]^{-1} \underline{\mathbf{p}}$$

Usando la 5.9:

$$\underline{\mathbf{a}} = \left[ \sigma_s^2 - \frac{\sigma_s^4 \underline{\mathbf{p}}^* \underline{\mathbf{p}}}{(1 + \sigma_s^2 \underline{\mathbf{p}}^* \underline{\mathbf{p}})} \right] \underline{\mathbf{p}} = \left[ \frac{\sigma_s^2}{1 + \sigma_s^2 \underline{\mathbf{p}}^* \underline{\mathbf{p}}} \right] \underline{\mathbf{p}}$$

C'è polarizzazione poichè:

$$E[\hat{s}^2] = \left( \frac{\sigma_s^2 \underline{\mathbf{p}}^* \underline{\mathbf{p}}}{1 + \sigma_s^2 \underline{\mathbf{p}}^* \underline{\mathbf{p}}} \right)^2 E[s^2] < E[\hat{s}^2]$$

$$E[(s - \hat{s})^2] = E[\varepsilon^2] = \sigma_s^2 (1 - \underline{\mathbf{a}}^* \underline{\mathbf{p}}) = \frac{\sigma_s^2}{1 + \sigma_s^2 \underline{\mathbf{p}}^* \underline{\mathbf{p}}} = \frac{\sigma_s^2}{1 + \sigma_s^2 \underline{\mathbf{c}}^* \underline{\mathbf{M}}^{-1} \underline{\mathbf{c}}}$$

Si osservi infine che per  $\sigma_s^2 \rightarrow \infty$ , la polarizzazione va a 0 e:

$$\lim_{\sigma_s^2 \rightarrow \infty} E[\varepsilon^2] = \frac{1}{\underline{\mathbf{c}}^* \underline{\mathbf{M}}^{-1} \underline{\mathbf{c}}}$$

Questa è l'errore q. m. minimo della stima non polarizzata in presenza di rumore colorato che abbiamo appunto usato nella stima spettrale di Capon. In quel caso il vettore  $\underline{\mathbf{c}}$  è proprio la sinusoide campionata che ha la frequenza posizionata dove si vuole calcolare lo spettro.



## CAPITOLO 5 LE DIREZIONI DI ARRIVO

### Stima di una costante

Nel caso semplice in cui il rumore é bianco e il vettore  $\underline{\mathbf{p}}$  é una costante:

$$\underline{\mathbf{p}} = A \times \underline{\mathbf{1}} \rightarrow E[\varepsilon^2] = \frac{\sigma_s^2 \sigma_n^2}{\sigma_n^2 + \sigma_s^2 |A|^2 N} \rightarrow \frac{1}{|A|^2 N} \text{ per } \sigma_s^2 \rightarrow \infty$$

Si osservi di nuovo che **lo stimatore è polarizzato** in quanto tende sistematicamente a sottostimare:

$$E[(\widehat{s})^2] < \sigma_s^2$$

Inoltre, per  $\sigma_n^2 \rightarrow \infty$ ,  $\widehat{s} \rightarrow 0$ . A parole, se vi è troppo rumore, la stima tende a zero. Invece se  $\sigma_s^2 \rightarrow \infty$ , la stima diventa non polarizzata.

### Stima di una pendenza

Indicando ora con un vettore  $\underline{\mathbf{q}}$  il caso di una variazione lineare tra  $-1$  e  $1$ , e quindi supponendo che il segnale osservato sia compreso tra  $-M$ ,  $M$ , dove  $M$  è una variabile casuale

$$\underline{\mathbf{y}} = M\underline{\mathbf{q}} + \underline{\mathbf{n}}$$

Le componenti del vettore  $\underline{\mathbf{q}}$ ,  $\{q_i\}$  sono:

$$\{q_i\} = \left( \frac{2i}{N} - 1 \right); \quad i = 0, \dots, N$$

$$\underline{\mathbf{q}}^* \underline{\mathbf{q}} = \sum_{i=0}^N \left( \frac{2i}{N} - 1 \right)^2 \sim \frac{N}{3}$$

La varianza della stima della variabile  $M$  è:

$$\sigma_{\widehat{M}}^2 = E[(M - \widehat{M})M] \simeq \frac{\sigma_M^2 \sigma_n^2}{\sigma_n^2 + \sigma_M^2 N/3} \sim \frac{3\sigma_n^2}{N}$$

e di nuovo si tende alla situazione di non polarizzazione se  $\sigma_n^2 \rightarrow 0$ . Si osservi poi che se si vuole stimare la pendenza locale  $m$  e quindi ponendo:

$$\underline{\mathbf{z}} = m\underline{\mathbf{x}} + \underline{\mathbf{n}}$$

$$\{r_i\} = \left(i - \frac{N}{2}\right); \quad i = 0, \dots, N$$

$$\underline{\mathbf{r}}^* \underline{\mathbf{r}} = \sum_{i=0}^N \left(i - \frac{N}{2}\right)^2 \sim \frac{N^3}{12}$$

$$\sigma_{\hat{m}}^2 \sim \frac{12}{N^3} \sigma_n^2$$

### Stima simultanea di una costante e di una pendenza

Se ora siamo nel caso:

$$\underline{\mathbf{y}} = M \underline{\mathbf{q}} + s \underline{\mathbf{p}} + \underline{\mathbf{n}}$$

profittando dell'ortogonalità tra  $\underline{\mathbf{q}}$  e  $\underline{\mathbf{p}}$  e cioè che:

$$\sum_{i=0}^N p_i q_i^* = 0$$

si vede che le analisi fatte nei precedenti paragrafi sono ancora utilizzabili.

### 5.4.2 Esempio: Stima AR della frequenza di un processo casuale quasi monocromatico

Si farà ora un'analisi spettrale autoregressiva di uno spettro colorato in presenza di rumore colorato. Si consideri un rumore bianco  $w_i$  che pilota un filtro passabanda la cui risposta all'impulso è una sinusoide complessa finestrata da una finestra rettangolare lunga  $N_0$  campioni; a questo spettro fortemente colorato, (quasi monocromatico per  $N_0 \rightarrow \infty$ ) si aggiunga un rumore lievemente colorato, p.e. passa - alto. Si abbia cioè:

$$y_i = A \sum_{h=1}^{N_0} w_{i-h} e^{jh\omega T} + n_i - n_{i-1} \quad (5.31)$$

$$E[|w_i|^2] = \sigma_w^2; \quad E[|n_i|^2] = \sigma_n^2; \quad (5.32)$$

Si voglia ora fare la stima spettrale AR della sequenza  $y_i$ ; supponiamo per ora di avere la funzione di autocorrelazione **esatta**  $r_y(n)$ ; sia l'ordine dello stimatore pari ad uno. Calcoliamo quindi:

$$r_y(0) = N_0 |A|^2 \sigma_w^2 + 2\sigma_n^2 \quad (5.33)$$

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

$$r_y(1) = (N_0 - 1)|A|^2\sigma_w^2 e^{j\varphi} - \sigma_n^2; \quad \varphi = \omega T; \quad (5.34)$$

Si vede anche che:

$$c_1 = -\frac{(N_0 - 1)|A|^2\sigma_w^2 e^{j\varphi} - \sigma_n^2}{N_0|A|^2\sigma_w^2 + 2\sigma_n^2} \quad (5.35)$$

Lo zero di  $A_1(z)$  è in

$$\begin{aligned} z_0 &= \frac{(N_0 - 1)|A|^2\sigma_w^2 e^{j\varphi} - \sigma_n^2}{N_0|A|^2\sigma_w^2 + 2\sigma_n^2} = \\ &= \frac{(N_0 - 1)|A|^2\sigma_w^2}{N_0|A|^2\sigma_w^2 + 2\sigma_n^2} e^{j\varphi} - \frac{\sigma_n^2}{N_0|A|^2\sigma_w^2 + 2\sigma_n^2} \end{aligned} \quad (5.36)$$

La polarizzazione della stima spettrale è data dal vettore differenza  $d$  tra  $e^{j\varphi}$  e  $z_0$ ; in prima approssimazione:

$$\begin{aligned} d = e^{j\varphi} - z_0 &\sim \frac{1}{N_0} \left[ \left( 1 + \frac{2\sigma_n^2}{|A|^2\sigma_w^2} \right) e^{j\varphi} + \frac{\sigma_n^2}{|A|^2\sigma_w^2} \right] \\ d &\sim \frac{e^{j\varphi}}{N_0} + \frac{\sigma_n^2}{N_0|A|^2\sigma_w^2} \end{aligned} \quad (5.37)$$

Quindi, invece di essere in  $e^{j\varphi}$ , lo zero dello stimatore (il polo dello spettro ricostruito) viene richiamato verso il centro del cerchio ed anche spostato. La lunghezza finita della risposta all'impulso sinusoidale porta all'allargamento della riga spettrale ricostruita; l'aggiunta del rumore colorato sposta il picco dello spettro stimato verso le frequenze più elevate.

Nel caso invece in cui  $N_0 \rightarrow \infty$ , e vi sia un rumore bianco additivo:

$$y_i = Ae^{ji\varphi} + n_i; \quad r_y(0) = \sigma_A^2 + \sigma_n^2; \quad r_y(1) = \sigma_A^2 e^{j\varphi}$$

$$z_0 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_n^2} e^{j\varphi}$$

### Lunghezza limitata dei dati

Nel paragrafo precedente abbiamo supposto che fosse disponibile l'autocorrelazione esatta; invece, se sono disponibili solo  $M$  dati, si ha solo una stima dell'autocorrelazione e quindi il valore trovato per  $z_0$  (il valore atteso) sarà affetto da un disturbo casuale che vogliamo ora calcolare in modo

## 5.5 STIMA OTTIMA (U. SPAGNOLINI)

approssimato. La dispersione della frequenza e quindi dell'argomento di  $\hat{z}_0$  può essere stimata come segue:

$$\begin{aligned}\arg(\hat{z}_0) &= \arg(\hat{r}_1) = \arg \frac{\sum |A|^2 e^{j\phi} + A(n_i + n_{i-1}) + n_i n_{i-1}}{N} = \\ &= \arg \left( e^{j\phi} + \frac{|A| (n_i + n_{i-1}) + n_i n_{i-1}}{N |A|^2} \right)\end{aligned}$$

e considerando solo l'effetto dei termini in quadratura con  $e^{j\phi}$ :

$$\hat{\phi} \sim \phi + \frac{|A|p}{\sqrt{2}N|A|^2} + \frac{q}{\sqrt{2}N|A|^2} \sim \phi + \frac{p}{\sqrt{2}N|A|} + \frac{q}{\sqrt{2}N|A|^2}$$

$$\sigma_p^2 = 2N\sigma_n^2; \quad \sigma_q^2 = N\sigma_n^4$$

$$\sigma_{\hat{\phi}}^2 \sim \frac{2N\sigma_n^2}{2N^2|A|^2} + \frac{N\sigma_n^4}{2N^2|A|^2} \sim \frac{\sigma_n^2}{N|A|^2}$$

$$\hat{f} = \frac{\hat{\phi}}{2\pi T} \rightarrow \sigma_{\hat{f}}^2 = \frac{\sigma_n^2}{4\pi^2 T^2 N |A|^2} = f_c^2 \frac{\sigma_n^2}{4\pi^2 N |A|^2}$$

In sostanza, con dei ragionevoli rapporti segnale rumore, la maggiore fonte di dispersione della posizione dello zero è proprio la dispersione della stima dell'autocorrelazione.

## 5.5 Stima ottima (U. Spagnolini)

È opportuno chiedersi quale sia la tecnica che porta alla più precisa (o più *efficiente*) stima della frequenza di una sinusoide e quindi della direzione di arrivo su una schiera di sensori.

### 5.5.1 La regressione lineare

Si consideri un insieme di osservazioni  $\{y_i\}$  con  $i = 1, \dots, N$  generate da un modello

$$y_i = m \cdot i + q + w_i \quad (5.38)$$

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

in cui i parametri  $m$  e  $q$  sono deterministici (ma non noti) e  $w_i$  rappresenta una variabile casuale gaussiana indipendente da campione a campione

$$E[w_i w_j] = \sigma_w^2 \delta(i - j) \quad (5.39)$$

e valore medio zero. Si vogliano stimare i due parametri  $m$  e  $q$  dall'insieme delle misure  $\{y_i\}$ . La stima dei parametri  $(\hat{m}, \hat{q})$  è ottenuta dalla minimizzazione dell'errore quadratico:

$$\varepsilon^2 = \left\{ \sum_{i=1}^N (y_i - \hat{m} \cdot i - \hat{q})^2 \right\} \quad (5.40)$$

la soluzione si ottiene annullando le derivate di  $\varepsilon^2$  rispetto alle due incognite  $(\hat{m}, \hat{q})$  come si fa per la predizione lineare. In realtà è possibile e più agevole riferirsi alle variazioni  $\delta m = m - \hat{m}$  e  $\delta q = q - \hat{q}$  da cui segue il sistema:

$$\begin{bmatrix} \sum_{i=1}^N i^2 & \sum_{i=1}^N i \\ \sum_{i=1}^N i & \sum_{i=1}^N 1 \end{bmatrix} \begin{bmatrix} \delta m \\ \delta q \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^N i w_i \\ \sum_{i=1}^N w_i \end{bmatrix} \rightarrow \mathbf{R} \times \boldsymbol{\delta} = \mathbf{w} \quad (5.41)$$

Si deriva immediatamente che questo stimatore non è polarizzato ( $E[\boldsymbol{\delta}] = 0$ ) in quanto le  $w_i$  sono a valore medio zero. La varianza si valuta come segue:

$$E[\boldsymbol{\delta} \boldsymbol{\delta}^T] = E[\mathbf{R}^{-1} \mathbf{w} \mathbf{w}^T \mathbf{R}^{-1}] = \mathbf{R}^{-1} E[\mathbf{w} \mathbf{w}^T] \mathbf{R}^{-1} = \sigma_w^2 \mathbf{R}^{-2} \quad (5.42)$$

in cui si ha:

$$\sigma_w^2 \mathbf{R}^{-1} = \frac{\sigma_w^2}{\sum_{i=1}^N 1 \sum_{i=1}^N i^2 - \left( \sum_{i=1}^N i \right)^2} \begin{bmatrix} \sum_{i=1}^N 1 & -\sum_{i=1}^N i \\ -\sum_{i=1}^N i & \sum_{i=1}^N i^2 \end{bmatrix} \quad (5.43)$$

La varianza dei singoli parametri risulta quindi:

$$\text{var}(\hat{m}) = \sigma_w^2 \frac{12N}{N^2(N^2 - 1)} = \sigma_w^2 \frac{12}{N(N^2 - 1)} \approx \frac{12}{N^3} \sigma_w^2 \quad (5.44)$$

$$\text{var}(\hat{q}) = \sigma_w^2 \frac{2N(N+1)(2N+1)}{N^2(N^2 - 1)} = \sigma_w^2 \frac{2(2N+1)}{N(N-1)} \approx \frac{4}{N} \sigma_w^2 \quad (5.45)$$

## 5.5 STIMA OTTIMA (U. SPAGNOLINI)

Si osserva la dipendenza della varianza di  $\hat{q}$  con l'inverso del numero dei campioni  $N$  e la dipendenza della varianza della pendenza  $\hat{m}$  con  $1/N^3$  (in altre parole il contributo utile dovuto ai campioni lontani pesa sempre di più in quanto l'angolo - o l'inclinazione  $m$ - dipende inversamente dalla distanza). Si può dimostrare che la varianza dei singoli parametri appena calcolata rappresenta anche il limite ottimo (Cramer - Rao); lo stimatore che minimizza l'errore quadratico medio presentato nell'equazione (5.40) è il "migliore" stimatore non polarizzato dei parametri  $m$  e  $q$ . Se le misure sono simmetriche rispetto al campione  $i = 0$  ovvero la sequenza delle misure è:  $\{y_i\}$  con  $i = -(N-1)/2, \dots, (N-1)/2$  cambiano gli estremi delle sommatorie nel calcolo della varianza:

$$\sigma_w^2 \mathbf{R}^{-1} = \sigma_w^2 \begin{bmatrix} \sum_{i=-(N-1)/2}^{(N-1)/2} i^2 & 0 \\ 0 & \sum_{i=-(N-1)/2}^{(N-1)/2} 1 \end{bmatrix} \quad (5.46)$$

$$var(\hat{m}) = \sigma_w^2 \frac{12}{N(N^2 - 1)} \approx \frac{12}{N^3} \sigma_w^2$$

$$var(\hat{q}) = \sigma_w^2 \frac{1}{N}$$

e la varianza nella stima della pendenza  $m$  rimane inalterata mentre la varianza nella stima dell'intercetta  $q$  si riduce di 6dB.

### 5.5.2 Stima ottima di una frequenza

Si consideri un segnale sinusoidale con rumore additivo:

$$x_i = A \exp[j\theta_i] + w_i = A \exp[j(ik\Delta + \phi)] + w_i; \quad \theta_i = ik\Delta + \phi$$

in cui la fase  $\phi$  e il numero d'onda  $k$  sono parametri deterministici (ma non noti) e il rumore è gaussiano bianco:

$$E[w_i w_l^*] = \sigma_w^2 \delta(i - l)$$

È possibile calcolare sia  $k$  che  $\phi$  applicando la regressione lineare alle osservazioni delle fasi dei campioni della sequenza  $\{x_i\}$ . E' semplice dimostrare che, per  $\sigma \ll A$ , le fasi dei campioni della sequenza  $\{x_i\}$  sono rappresentate dal modello lineare con rumore additivo:

$$\angle x_i \simeq \theta_i + n_i = k\Delta i + \phi + n_i; \quad E[n_i^2] = \frac{\sigma_w^2}{A^2}.$$

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

Diventa quindi possibile stimare la fase e la frequenza (e quindi direzione di arrivo) operando una regressione lineare sulle misure delle fasi della sequenza complessa. Si osservi che, per effettuare la regressione sulla sequenza delle fasi e' necessario "svolgere" la sequenza  $\{\angle x_i\}$  e cioè determinarne le fasi assolute e non a meno di multipli di  $2\pi$ . D'altra parte, se la sequenza è campionata correttamente e non c'è rumore di equivocazione, la rotazione di fase tra un campione ed il successivo si mantiene sempre inferiore a  $\pi$  e quindi lo svolgimento sarebbe non ambiguo se non ci fosse rumore. Si osservi poi che se si converte il segnale a frequenza zero utilizzando la frequenza a cui il periodogramma ha il massimo picco, si ottiene un segnale la cui rotazione di fase in tutto l'intervallo di osservazione non supera  $\pm\pi$ . Quindi, ci si può sempre mettere nelle condizioni in cui la rotazione di fase totale non ha bisogno di essere svolta. Infine, **tutto questo discorso è valido se il rumore è piccolo.**

La dispersione della stima della fase e della frequenza è quindi (assumendo simmetrica la sequenza delle misure):

$$\sigma_{\hat{\phi}} = \frac{\sigma_w}{A} \frac{1}{\sqrt{N}}$$
$$\sigma_{\hat{k\Delta}} = \frac{\sigma_w}{A} \frac{2\sqrt{3}}{\sqrt{N(N^2-1)}} \approx \frac{2\sqrt{3}}{N^{3/2}} \frac{\sigma_w}{A}$$

trovando un fattore  $N^{-3/2}$  invece del fattore  $N^{-1}$  trovato per il periodogramma e  $N^{-1/2}$  per la tecnica AR quando si deve stimare l'auto-correlazione dai dati. Peraltro, se si volesse determinare una frequenza cercando la posizione del massimo locale in un periodogramma, anche cercando la frazione di cella corrispondente usando una tecnica di interpolazione, si troverebbe un risultato simile.

### 5.5.3 Discussione sulla stima di una frequenza

Da quanto detto appare chiaro che per stimare una frequenza, sono necessarie due azioni disgiunte:

I: filtrare passa banda attorno alla frequenza di interesse, per ridurre il rumore; il filtro sarà tanto più stretto quanto più lungo è l'intervallo di osservazione.

II: stimare la pendenza della rampa di fase residua, una volta che siamo già vicini al valore effettivo della frequenza; allora, se gli errori di

## 5.6 STIMA DI PROCESSI MULTIDIMENSIONALI

fase sono al di sotto di  $\pi$  in tutto l'intervallo di stima, si effettua una regressione su tutti questi valori e si stima la pendenza residua, come visto nella sezione precedente.

E' giusto domandarsi come mai quest'ultimo metodo sia tanto migliore della stima AR; con questa, si trova la frequenza come:

$$\hat{f} = \frac{\angle \hat{r}_1}{2\pi T}$$

Per dare una risposta intuitiva, si può dire che in questo caso non si cerca di sfruttare l'effetto di braccio di leva (il fattore  $N$ ) dovuto al progressivo sfasamento, ma si cerca solo di stimare al meglio l'effetto della rotazione di un passo.

## 5.6 Stima di processi multidimensionali

Si supponga di volere stimare un processo multidimensionale e cioè un insieme di  $N$  sequenze temporalmente allineate: in particolare si voglia stimare la sequenza di vettori di  $N$  dimensioni p.e. incorrelati sia tra loro che nel tempo:

$$\underline{\mathbf{s}}_i = \begin{bmatrix} s_{1,i} \\ s_{2,i} \\ s_{j,i} \\ \dots \\ \dots \\ s_{N,i} \end{bmatrix} ; j = 1, \dots, N; \quad i = 1, 2, \dots;$$

$$E [\underline{\mathbf{s}}_i \underline{\mathbf{s}}_{i+k}^*] = \sigma_s^2 \delta(i - k) \underline{\mathbf{I}}$$

Il segnale osservato è la sequenza  $\underline{\mathbf{x}}_i$ , e sussiste la seguente relazione matriciale:

$$\underline{\mathbf{x}}_i = \underline{\mathbf{A}} \underline{\mathbf{s}}_i + \underline{\mathbf{n}}_i ; \quad E [\underline{\mathbf{n}}_i \underline{\mathbf{n}}_{i+k}^*] = \sigma_n^2 \delta(i - k) \underline{\mathbf{I}}$$

Quindi, il rumore additivo è costituito da componenti incorrelate e tra loro e nel tempo. La matrice  $\underline{\mathbf{A}}$ , qualsiasi e magari a coefficienti complessi, ha dimensioni  $[N \times N]$ . Possiamo trovare subito la matrice  $\underline{\mathbf{R}}$  di



## CAPITOLO 5 LE DIREZIONI DI ARRIVO

covarianza del vettore  $\underline{\mathbf{x}}$  :

$$\underline{\mathbf{R}} = E [(\underline{\mathbf{A}}\underline{\mathbf{s}} + \underline{\mathbf{n}}) (\underline{\mathbf{A}}\underline{\mathbf{s}} + \underline{\mathbf{n}})^*] = \sigma_s^2 \underline{\mathbf{A}} \underline{\mathbf{A}}^* + \sigma_n^2 \underline{\mathbf{I}}$$

Usando il principio di ortogonalità e sfruttando il formalismo matriciale troviamo subito la stima ottima del segnale  $\underline{\mathbf{s}}$ :

$$\hat{\underline{\mathbf{s}}} = \underline{\mathbf{B}} \underline{\mathbf{x}}; \quad E [(\underline{\mathbf{B}} \underline{\mathbf{x}} - \underline{\mathbf{s}}) \underline{\mathbf{x}}^*] = \underline{\mathbf{0}} \rightarrow \underline{\mathbf{B}} \underline{\mathbf{R}} = E [\underline{\mathbf{s}} \underline{\mathbf{x}}^*] = \sigma_s^2 \underline{\mathbf{A}}^*$$

e la matrice  $\underline{\mathbf{B}}$  è quindi:

$$\underline{\mathbf{B}} = \sigma_s^2 \underline{\mathbf{A}}^* \underline{\mathbf{R}}^{-1}$$

L'errore di stima è:

$$Q = \sum_{n=1}^N E [(\hat{s}_{i,n} - s_{i,n})^2]$$

$$Q = Tr \{E [(\underline{\mathbf{B}} \underline{\mathbf{x}} - \underline{\mathbf{s}}) (\underline{\mathbf{B}} \underline{\mathbf{x}} - \underline{\mathbf{s}})^*]\} = Tr \{E [(\underline{\mathbf{s}} - \underline{\mathbf{B}} \underline{\mathbf{x}}) \underline{\mathbf{s}}^*]\}$$

profittando dell'incorrelazione tra errore di stima e dati per il principio di ortogonalità. Ricordando poi che:

$$E [\underline{\mathbf{x}} \underline{\mathbf{s}}^*] = \sigma_s^2 \underline{\mathbf{A}}$$

risulta che:

$$Q = \sigma_s^2 Tr \{ \underline{\mathbf{I}} - \underline{\mathbf{A}}^* \underline{\mathbf{R}}^{-1} \underline{\mathbf{A}} \}$$

Si osservi che dalla sequenza di vettori  $\underline{\mathbf{x}}_i$  è possibile ricavare la matrice  $\underline{\mathbf{R}}$  e quindi, se  $\sigma_n^2$  è nota, si può determinare la matrice  $\underline{\mathbf{A}} \underline{\mathbf{A}}^*$  ; non è possibile però ricavare  $\underline{\mathbf{A}}$ . Infatti, si può osservare che

$$\underline{\mathbf{A}} \underline{\mathbf{T}} \underline{\mathbf{T}}^* \underline{\mathbf{A}}^* = \underline{\mathbf{A}} \underline{\mathbf{A}}^*$$

purchè:

$$\underline{\mathbf{T}} \underline{\mathbf{T}}^* = \underline{\mathbf{I}}$$

e quindi se la matrice  $\underline{\mathbf{T}}$  è ortogonale, eguale all'inversa della sua trasposta. Allora, se è nota la matrice  $\underline{\mathbf{R}}$  e quindi  $\underline{\mathbf{A}} \underline{\mathbf{A}}^*$  non potremo ricavarne la matrice  $\underline{\mathbf{A}}$ , se non a meno di una matrice ortogonale  $\underline{\mathbf{T}}$  che rimane incognita. Abbiamo visto un problema equivalente per la caratteristica di fase di una funzione di trasferimento, irrecuperabile dall'autocorrelazione dell'uscita da cui otteniamo lo spettro di potenza e quindi le ampiezze.

## 5.6 STIMA DI PROCESSI MULTIDIMENSIONALI

Il problema é solubile se si possono utilizzare informazioni a priori sulla struttura della matrice  $\underline{\mathbf{A}}$ . Ad esempio, se  $\underline{\mathbf{A}}$  é simmetrica (come dire fase zero), allora la matrice

$$\underline{\mathbf{T}} = \underline{\mathbf{I}}; \rightarrow \underline{\mathbf{A}} = \underline{\mathbf{A}}^*$$

gli autovettori di  $\underline{\mathbf{R}}$  coincidono con quelli di  $\underline{\mathbf{A}}$  (le matrici  $\underline{\mathbf{A}}$ ,  $\underline{\mathbf{A}}^*$  commutano) e gli autovalori di  $\underline{\mathbf{R}}$  sono i quadrati di quelli di  $\underline{\mathbf{A}}$  cui va sommato  $\sigma_n^2$ . La formula che dà l'errore quadratico medio si semplifica:

$$Q = \sigma_s^2 \text{Tr} \{ \underline{\mathbf{I}} - \underline{\mathbf{A}}^* \underline{\mathbf{R}}^{-1} \underline{\mathbf{A}} \} = \sigma_s^2 \sum_{n=1}^N \left( 1 - \frac{\lambda_{A,n}^2 \sigma_s^2}{\lambda_{A,n}^2 \sigma_s^2 + \sigma_n^2} \right)$$

$$Q = \sum_{n=1}^N \left( \frac{\sigma_s^2 \sigma_n^2}{\lambda_{A,n}^2 \sigma_s^2 + \sigma_n^2} \right)$$

profittando del fatto che la traccia di una matrice é la somma degli autovalori.

Allo stesso modo, nell'analisi fatta per la deconvoluzione di segnali nel dominio delle frequenze, si é visto che é impossibile determinare la fase della funzione di trasferimento a partire dall'autocorrelazione dell'uscita.

### 5.6.1 L'effetto cocktail e la separazione delle sorgenti

Le considerazioni fatte nel paragrafo precedente hanno un interessante conseguenza nel caso della ricezione acustica binaurale (con le due orecchie) di un segnale costituito da due sorgenti che emettano contemporaneamente. In altre parole, si pensi ad una persona che ascolti due parlatori che parlino assieme. È ben noto che è possibile concentrare l'attenzione su uno dei due parlatori e praticamente cancellare la conversazione dell'altro (il cosiddetto *effetto cocktail*, che permette di seguire una conversazione in mezzo a tante). Tuttavia, la separazione è quasi impossibile se i due parlatori sono sovrapposti su un singolo canale fonico (p.e. parlano contemporaneamente alla televisione).

Questo problema di stima è analizzabile con le convenzioni stabilite nel paragrafo precedente e risulta:

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

$$\underline{\mathbf{x}}_i = \underline{\mathbf{A}} \underline{\mathbf{s}}_i + \underline{\mathbf{n}}_i ; \quad E [\underline{\mathbf{n}}_i \underline{\mathbf{n}}_{i+k}^*] = \sigma_n^2 \delta(i - k) \underline{\mathbf{I}}$$

dove ora i vettori  $\underline{\mathbf{x}}_i$ ,  $\underline{\mathbf{s}}_i$ ,  $\underline{\mathbf{n}}_i$  sono di rango  $[2 \times 1]$  e la matrice  $\underline{\mathbf{A}}$  di rango  $[2 \times 2]$ . Rappresentiamo le due conversazioni con  $\underline{\mathbf{s}}$ , i due segnali ricevuti (le due orecchie) con  $\underline{\mathbf{x}}$  e la matrice di confusione anch'essa a priori ignota con  $\underline{\mathbf{A}}$ . Se si conoscesse  $\underline{\mathbf{A}}$  sarebbe facile, a parte il disturbo, ritrovare le due sorgenti usando le tecniche viste prima; per trovare  $\underline{\mathbf{A}}$  si potrebbe p.e. correlare il segnale misurato con quello della sorgente (se fosse nota, però), facendo:

$$\underline{\mathbf{A}} = E [\underline{\mathbf{x}} \underline{\mathbf{s}}^*] \quad (5.47)$$

Ma  $\underline{\mathbf{s}}$  corrisponde alla conversazione che deve essere stimata è inaccessibile. Certamente si dispone di  $\underline{\mathbf{x}}$  e quindi si dispone della sua matrice di covarianza  $\underline{\mathbf{R}} = E [\underline{\mathbf{x}} \underline{\mathbf{x}}^*]$ : il problema diventa più complesso. Si è visto infatti che dalla matrice di covarianza  $\underline{\mathbf{R}}$  non è possibile ottenere la matrice  $\underline{\mathbf{A}}$  se non a meno di una rotazione di coordinate  $\underline{\mathbf{T}}$ . In altre parole, anche in assenza di rumore non possiamo ottenere di nuovo  $\underline{\widehat{\mathbf{s}}}$  (la stima delle due sorgenti di segnale) da  $\underline{\mathbf{x}}$  ma solo un vettore  $\underline{\widehat{\mathbf{s}}}'$  in cui le componenti di  $\underline{\mathbf{s}}$ , sono mescolate tra loro:

$$\underline{\widehat{\mathbf{s}}}' = \underline{\mathbf{T}} \underline{\widehat{\mathbf{s}}}; \quad E [\underline{\widehat{\mathbf{s}}} \underline{\widehat{\mathbf{s}}}^*] = E [\underline{\widehat{\mathbf{s}}}' \underline{\widehat{\mathbf{s}}}'^*] = \sigma_s^2 \underline{\mathbf{I}}$$

$$\underline{\mathbf{T}} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}; \quad \underline{\mathbf{T}} \underline{\mathbf{T}}^* = \underline{\mathbf{I}} \quad (5.48)$$

Pertanto, il fatto che si possano separare le due conversazioni dimostra che esistono dei meccanismi più efficienti della predizione lineare per fare l'operazioni di separazione delle sorgenti. Infatti, se invece di imporre l'incorrelazione tra gli elementi di  $\underline{\widehat{\mathbf{s}}}$ , possibile qualsiasi sia la matrice  $\underline{\mathbf{T}}$  e quindi inutile per effettuare una separazione delle sorgenti, si impone l'indipendenza tra quelle componenti, il risultato non potrebbe essere altro che quello corretto. In sostanza, si deve ricorrere a statistiche di ordine superiore al secondo (deconvoluzione cieca).

Per fare un esempio, consideriamo due processi identici  $s_1, s_2$  ma indipendenti, non gaussiani, a curtosi  $k \neq 3$ ; ricordiamo che la curtosi é:

$$k = \frac{E [s_1^4]}{(E [s_1^2])^2} = \frac{E [s_2^4]}{(E [s_2^2])^2} \quad (5.49)$$

## 5.6 STIMA DI PROCESSI MULTIDIMENSIONALI

Dopo la stima lineare ottima, noi abbiamo  $\hat{s}'_1, \hat{s}'_2$

$$\underline{\hat{s}'} = \underline{\mathbf{T}} \underline{\hat{s}} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} \hat{s}_1 \\ \hat{s}_2 \end{bmatrix} \quad (5.50)$$

$$\hat{s}'_1 = \hat{s}_1 \cos \theta + \hat{s}_2 \sin \theta \quad (5.51)$$

$$\hat{s}'_2 = -\hat{s}_1 \sin \theta + \hat{s}_2 \cos \theta \quad (5.52)$$

che sono due combinazioni lineari di  $\hat{s}_1, \hat{s}_2$ , **incorrelate ed a eguale potenza**, qualsiasi sia il valore di  $\theta$ , peraltro ignoto. Quello che varia con  $\theta$ , invece, é la curtosì di  $\hat{s}'_1, \hat{s}'_2$  che é:

$$k_m = \frac{E[s_{a,b}^4]}{(E[s_{a,b}^2])^2} \quad (5.53)$$

Allora:

$$k_m = \frac{E[\hat{s}_1^4] (\cos^4 \theta + \sin^4 \theta) + 6 (E[\hat{s}_1^2])^2 \sin^2 \theta \cos^2 \theta}{(E[\hat{s}_1^2])^2} \quad (5.54)$$

$$k_m = \frac{3(k+1) + (k-3) \cos 4\theta}{4} \quad (5.55)$$

e se

$$\theta = m\pi/2; m = 0..3; \quad (5.56)$$

allora:

$$k_m = k \quad (5.57)$$

Peraltro, per  $m = 1, 3$ , si scambiano le sorgenti tra loro e per  $m = 0, 2$  ne cambiamo i segni. In sostanza, abbiamo separato le conversazioni e risolto il problema. Invece per

$$\theta = \pi/4 + m\pi/2; m = 0..3; \rightarrow k_m = \frac{k+3}{2} \quad (5.58)$$

non c'è separazione. La curtosì misurata  $k_m$  oscilla tra due valori:

$$k < k_m < \frac{k+3}{2} \text{ oppure } k > k_m > \frac{k+3}{2} \quad (5.59)$$

## CAPITOLO 5 LE DIREZIONI DI ARRIVO

a seconda che  $k$  sia maggiore o minore di 3. Si vede che, se  $k \neq 3$ , e quindi le due conversazioni sono rappresentate da segnali non gaussiani, la curtosi di  $\hat{s}'_1, \hat{s}'_2$  ha un'oscillazione con  $\theta$  (tra  $k$  e la media aritmetica tra  $k$  e 3) che permette di individuare il valore di  $\theta$  corretto permettendo perciò la separazione dei due segnali a parte errori di segno e di scambio tra  $\hat{s}_1, \hat{s}_2$ . Invece, se i processi  $s_1, s_2$  sono gaussiani, anche le loro somme lo saranno e la separazione sarà impossibile.

In conclusione, misurando la matrice di covarianza  $\mathbf{R}$  del vettore  $\mathbf{x}$ , dopo avere eventualmente sottratto la parte dovuta al rumore, otteniamo la matrice  $\mathbf{A} \mathbf{T} \mathbf{T}^* \mathbf{A}^*$  e quindi, facendo la radice quadrata della matrice, otteniamo la matrice  $\mathbf{A} \mathbf{T}$  che è una matrice ancora simmetrica, come la  $\mathbf{R}$ . Per trovare finalmente la matrice  $\mathbf{T}$  ed arrivare all'effettiva separazione delle sorgenti, dovremo sfruttare la non gaussianità delle sorgenti. Per fare un esempio numerico sia:

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}; \text{ quindi } \mathbf{A} \mathbf{A}^* = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 2 & 4 \end{bmatrix} = \begin{bmatrix} 5 & 11 \\ 11 & 25 \end{bmatrix} \quad (5.60)$$

$$\mathbf{A} \mathbf{T} = \begin{bmatrix} 5 & 11 \\ 11 & 25 \end{bmatrix}^{1/2} = \begin{bmatrix} 1.2 & 1.89 \\ 1.89 & 4.63 \end{bmatrix}; \quad (5.61)$$

Se conosciamo  $\mathbf{A}$ , troveremmo la rotazione come:

$$\mathbf{T} = \mathbf{A}^{-1} \mathbf{A} \mathbf{T} = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}^{-1} \begin{bmatrix} 1.2 & 1.89 \\ 1.89 & 4.63 \end{bmatrix} = \quad (5.62)$$

$$= \begin{bmatrix} -0.51 & 0.85 \\ 0.855 & 0.51 \end{bmatrix} = \begin{bmatrix} -\sin \theta & \cos \theta \\ \cos \theta & \sin \theta \end{bmatrix} \quad (5.63)$$

$$\theta = \arccos 0.855 = 0.545 \quad (5.64)$$

La rotazione di 0.54 radianti è appunto quella che ci ridarebbe la matrice  $\mathbf{A}$  originale invece di quella simmetrizzata. Questo valore di  $\theta$  sarà appunto quello ottenibile dalla massimizzazione (o minimizzazione) della curtosi.

## Capitolo 6

### Il segnale vocale

Una delle principali applicazioni del trattamento numerico dei segnali è quella al segnale vocale e cioè il segnale generato dalla voce umana. Dopo una brevissima presentazione dei meccanismi di fonazione e dell'udito, tratteremo sia la codifica che la sintesi del segnale vocale; infine si farà anche qualche cenno alla codifica del segnale musicale.

Il segnale vocale viene formato nel tratto vocale che va dalle corde vocali alle labbra. E' questo un sistema riverberante, disadattato all'uscita (le labbra) ed abbastanza ben adattato all'ingresso (le corde vocali). La sagoma del sistema riverberante dipende dalla posizione della bocca e della lingua; il segnale di ingresso è generato dalle corde vocali ed è di due tipi: vocalizzato (quando le corde vocali generano una sequenza di impulsi di energia acustica) e non vocalizzato (quando le corde vocali generano un fruscio praticamente gaussiano a spettro abbastanza uniforme). La sequenza di impulsi generata durante la formazione di segnale vocalizzato ha una frequenza di ripetizione che va da circa  $100Hz$  (voce roca, maschi) fino a  $250 - 300Hz$  (voce melodiosa, bambini). Ogni impulso ha una durata di qualche millisecondo e quindi uno spettro che si estende fino a  $8 - 10kHz$ . Il timbro, e cioè la frequenza di emissione di questi impulsi (in inglese, pitch) può avere valori di significato, ad esempio interrogativo, (se crescente col tempo), o dichiarativo (se decrescente).

Gli impulsi attraversano il tratto vocale e sono quindi sagomati spettralmente dalla funzione di trasferimento tutti poli corrispondente. Ovviamente, trattandosi di un sistema distribuito e non di una cascata di tronchi di linea uniformi, non è possibile determinare il numero dei poli equivalente al sistema senza fare delle approssimazioni. Tuttavia, se si

## CAPITOLO 6 IL SEGNALE VOCALE

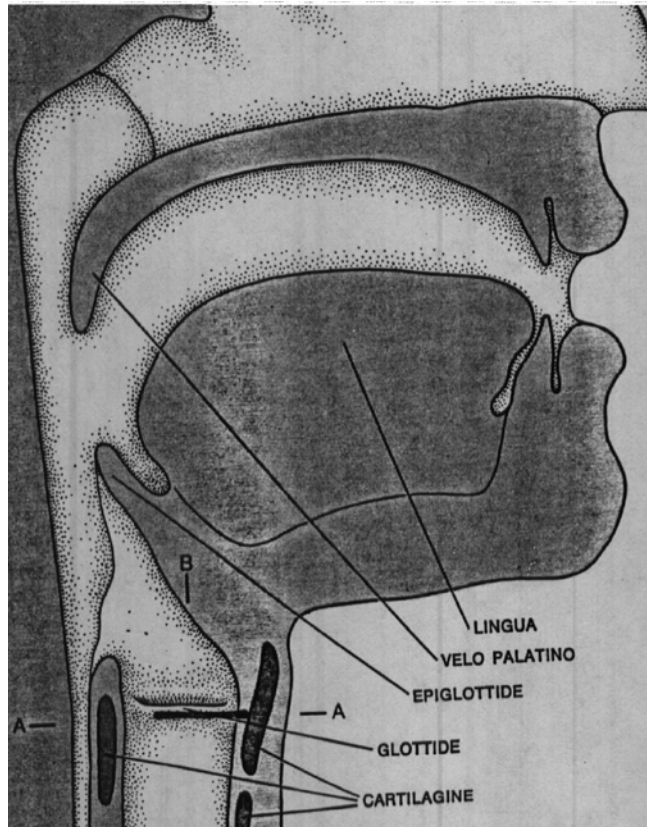


Figura 6.1: Apparato vocale umano

osserva che la lunghezza del tratto vocale è di circa  $20\text{cm}$  e che la velocità del suono è circa  $330\text{m/s}$  si deduce che il ritardo totale di propagazione è di circa  $.6\text{msec}$ . Se ora vogliamo approssimare il fenomeno della fonazione fino a  $4\text{kHz}$  la frequenza di campionamento deve essere almeno  $8\text{kHz}$  e pertanto l'intervallo di campionamento inferiore a  $125\ \mu\text{s}$ ; quindi il ritardo unitario di ogni cella deve essere inferiore a  $62,5\ \mu\text{s}$  ed il numero di celle contenute nel tratto superiore a 10; infatti, in genere si sceglie di rappresentare il tratto vocale con 10 celle e quindi con 10 poli.

Al massimo abbiamo 5 frequenza di risonanza, pertanto; in realtà se ne contano, per ogni vocale, da 3 a 4; queste frequenza di risonanza sono chiamate formanti e la loro posizione caratterizza abbastanza le vocali (vedi Tabella 1); si stia ben attenti a non pensare che distinguere le vocali

porti alla comprensione del parlato; gli inflessi dialettali e numerosi giochi di società insegnano il contrario. Infatti, l'intelligibilità del parlato è soprattutto legata alle consonanti e cioè a transitori tra una vocale (che può essere mantenuta per lungo tempo) ed un'altra.

Tabella I: Frequenze delle formanti dei suoni vocalici inglesi

Vocale	Frequenza	Frequenza	Frequenza
inglese	I formante	II formante	III formante
but	520	1190	2380
bat	660	1720	2410
feet	270	2290	3010
hot	730	1090	2440
food	440	1020	2240
fought	570	840	2410
foot	300	870	2240
bird	490	1350	1690
bit	390	1990	2550
bet	530	1840	2480

I suoni non vocalizzati sono generati dal fruscio emesso dalle corde vocali quando non generano impulsi (si pensi al parlato sottovoce). Infine, i suoni consonantici possono essere di vario tipo: le consonanti sibilanti sono generate dal fruscio emesso dalle corde vocali; ma vi sono consonanti di carattere plosivo (ad esempio b, p, t), corrispondenti a bruschi transitori del tratto vocale, ecc.

I toni nasali sono generati da interferenze tra emissioni acustiche attraverso il tratto vocale ed attraverso il setto nasale: quindi non corrispondono a funzioni di trasferimento di soli poli, ma anche con zeri dovuti al parallelo di due reti tutti poli.



## 6.1 Tecniche di codifica del segnale vocale

Il segnale vocale può essere codificato con tecniche di complessità molto diversa; al limite, si può pensare ad un sistema di riconoscimento del segnale vocale, alla trasmissione alfanumerica del testo riconosciuto, e quindi alla sintesi vocale del testo effettuata in ricezione: questa tecnica, di massima riduzione di ridondanza, porterebbe ad un ritmo di trasmissione di circa  $40 - 50 \text{ bit/sec}$ . Il risparmio è massimo, senza tuttavia perdere la riconoscibilità del testo; si perde invece la possibilità di individuare la personalità e lo stato d'animo del parlatore; inoltre, il ritardo dovuto al tempo necessario al riconoscimento è certamente maggiore dei circa  $10 \text{ msec}$  in genere consentiti. All'estremo opposto, possiamo pensare ad una codifica *trasparente* in cui la forma d'onda del segnale è trasmessa fedelmente al ricevitore: in questo caso, il ritmo di trasmissione può arrivare a  $40 \text{ KHz} \times 16 \text{ bit/campione} = 640 \text{ kbit/sec}$ ; quindi lo stesso segnale può essere trasmesso con 10.000 volte meno bit, ma naturalmente con delle perdite. Come vedremo, utilizzando sia i meccanismi di generazione del segnale, sia i meccanismi di percezione auditiva che rendono impercipienti alcune modifiche al segnale, è possibile avere un'ottima qualità del segnale vocale con  $7 - 8 \text{ kbit/sec}$  ed un'ottima percepibilità con  $3 - 4 \text{ kbit/sec}$ . Un segnale musicale, molto più complesso e per cui è opportuna una maggiore fedeltà, ha bisogno di circa  $150 - 200 \text{ kbit/sec}$ . Naturalmente, queste cifre sono variabili con il tempo e con la tecnologia; tuttavia, il costo di trasmissione del timbro e delle sue variazioni è già di qualche centinaia di  $\text{bit/sec}$ .

## 6.2 Codifica delle forme d'onda

I campioni di un segnale spettralmente colorato quale è il segnale vocale sono correlati tra loro; quindi sono almeno parzialmente predicibili ed è conveniente trasmettere non il segnale ma l'errore di predizione. La convenienza non sta però nello sbiancamento, che non comporta alcuna riduzione della frequenza di campionamento. Essa nasce dal fatto che nel caso più frequente di segnale vocalizzato, l'errore di predizione (approssimativamente il segnale di ingresso generato dalle corde vocali) è una forma d'onda costituita da impulsi ripetuti, con molti campioni pressoché nulli, e che può essere descritta e trasmessa *parsimoniosamente*

## 6.2 CODIFICA DELLE FORME D'ONDA

con un numero ridottissimo di parametri (ad esempio ampiezza, intervallo tra impulsi, e magari anche la forma d'onda dell'impulso, ripetitiva ma lentamente variabile con il tempo).

La possibilità di una codifica efficiente del segnale vocale nasce quindi dalla **non gaussianità** del processo stocastico corrispondente al segnale vocalizzato. Peraltro, i segnali non vocalizzati di durata abbastanza lunga corrispondono sostanzialmente a fruscii: essi sono abbastanza gaussiani, ma in compenso non serve una loro accurata descrizione, in quanto basta trasmetterne lo spettro di potenza piuttosto che la forma d'onda. In questo caso, la forma d'onda del segnale ricostruito non sarà affatto identica o simile a quella del segnale trasmesso, ma sarà incorrelata con esso avendo in comune solo densità di probabilità e spettro di potenza.

La tecnica di codifica del segnale più diffusa (LPC, Linear Predictive Coding) si fonda appunto sulla trasmissione periodica dei parametri del tratto vocale (ad esempio, i 10 coefficienti di riflessione) ed inoltre dei parametri degli impulsi generati dalle corde vocali del parlatore. Tuttavia, a titolo introduttivo, si discuteranno le due tecniche di codifica della forma d'onda più semplici, la Modulazione Delta e l'ADPCM (Adaptive Differential Pulse Code Modulation).

### 6.2.1 La Modulazione Delta

Il modo più elementare per sfruttare la dipendenza statistica dei successivi campioni di un segnale è quello di codificarne le differenze; nella Modulazione Delta questa codifica viene fatta con un solo bit, che quindi comunica se il campione successivo è maggiore o minore di quello precedente. In conseguenza viene aggiornato un contatore per permettere il successivo confronto (vedi figura 15.2). Il segnale di ingresso potrebbe anche essere analogico: in questo caso il sistema funzionerebbe anche da convertitore analogico - numerico. Il disturbo creato dalla Modulazione Delta è ridotto se i campioni successivi sono fortemente correlati e cioè se lo spettro del segnale è passa - basso. Il disturbo è di due tipi:

- rumore di granularità (idle noise o rumore di inattività);
- rumore di inseguimento (slope overload).

Supponiamo che il quanto  $a$  addizionato o sottratto al segnale per l'arrivo di un nuovo bit sia piccolo. Negli intervalli di tempo in cui il segnale varia poco, la trasmissione è fedele; tuttavia se c'è una variazione a gradino di ampiezza  $Ma$ , il sistema avrà bisogno di  $M$  passi per recu-

## CAPITOLO 6 IL SEGNALE VOCALE

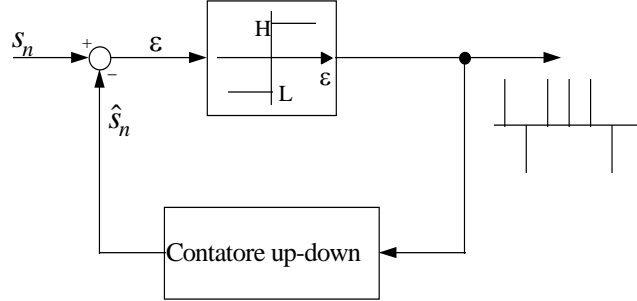


Figura 6.2: Modulatore delta

perare questa differenza (fig. 15.3). La pendenza massima descrivibile dal segnale ricostruito è  $|a|/T$ ; se vi sono pendenze maggiori, siamo nella condizione di inseguimento, il sistema va in "sovraccarico di pendenza" (in inglese *slope overload*) e nascono delle differenze anche grandi. Se invece il quanto  $a$  è abbastanza grande, la pendenza massima sarà sufficiente, ma ne nascerà una sorta di errore di quantizzazione (errore di granularità). In funzione dello spettro del segnale e della frequenza di campionamento vi è un valore ottimale per il quanto; se il segnale è non stazionario, tuttavia, è opportuno rendere la Modulazione Delta adattativa variando il quanto p.e. con la regola di accrescerlo di un fattore  $P$  se due bit successivi hanno lo stesso segno (situazione di inseguimento) e di ridurlo di un fattore  $Q$  se invece hanno segni opposti (situazione di inattività). Si riesce ad ottenere un sistema efficiente se le permanenze di segno sono probabili come le variazioni: dopo  $n$  passi, dovendo il quanto avere circa lo stesso valore, ed essendo la variazione totale di ampiezza:

$$P^{n/2}Q^{n/2} = (PQ)^{n/2} = 1 \quad (6.1)$$

risulta:

$$PQ = 1$$

Agendo in questo modo, si possono avere al limite quanti di ampiezza troppo grande o troppo piccola; per limitare questa escursione si può adottare la tecnica nota come CVSD (Continuously Variable Slope Delta) dove la regola di aggiornamento del quanto è la seguente:

$$a_n = \beta a_{n-1} + D_i ; \quad i = 1, 2 ; \quad D_1 \gg D_2 ; \quad 0 < \beta < 1 \quad (6.2)$$

## 6.2 CODIFICA DELLE FORME D'ONDA

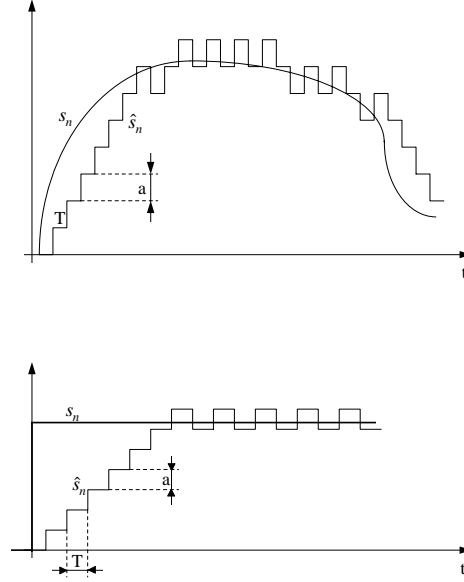


Figura 6.3: Funzionamento della modulazione delta e risposta al gradino

In condizione di inseguimento (bit successivi eguali) l'indice  $i = 1$ , in condizione di inattività (bit successivi diversi)  $i = 2$ . Con un ingresso costante, dopo un numero grande di campioni, si arriva ad avere un quanto di ampiezza minima:

$$a_{min} = \frac{D_2}{1 - \beta}$$

Invece in situazioni di inseguimento prolungate, il quanto cresce fino al valore massimo:

$$a_{max} = \frac{D_1}{1 - \beta}$$

Se  $\beta \simeq 1$  l'aggiornamento del quanto è lento; per progettare il sistema, si sceglie anzitutto il valore per  $\beta$  e quindi  $D_1$ ,  $D_2$  in funzione dei valori desiderati per i quanti minimo e massimo. Si vede ora un problema che pesa sui segnali codificati; essi sono molto vulnerabili agli errori di canale: infatti, se per qualche motivo un bit viene decodificato in modo errato, da quel momento in poi il ricevitore decodificherebbe in modo errato tutti

## CAPITOLO 6 IL SEGNALE VOCALE

i segnali ricevuti; questo mostra che è necessario prevedere dei periodici riassetamenti, ad esempio trasmettendo alcuni campioni non in modo differenziale.

### 6.2.2 L'ADPCM

La modulazione Delta è il caso più semplice di tutta una serie di codificatori differenziali: il sistema può essere migliorato in vari modi:

- predire meglio il prossimo campione, piuttosto che utilizzare un semplice mantentore come nella Modulazione Delta;

- usare un quantizzatore differenziale più duttile, con un numero di livelli di restituzione pari a  $2^m$  con  $m = 3, 4$  ed inoltre con una caratteristica di quantizzazione non uniforme;

- rendere adattativo il sistema, con tecniche di predizione lineare adattativa, quali quelle viste al capitolo 11. Lo schema a blocchi del sistema DPCM (Differential Pulse Code Modulator) è presentato in figura 15.4. La differenza tra la migliore predizione lineare del prossimo campione

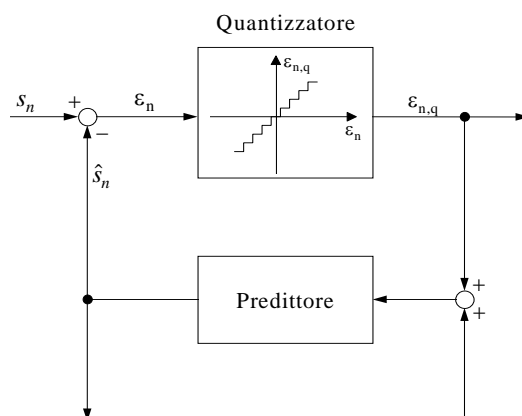


Figura 6.4: Schema a blocchi del DPCM (Differential Pulse Code Modulator)

ed il campione effettivo viene quantizzata con un quantizzatore non uniforme; l'errore di predizione, quantizzato, aggiorna il segnale predetto e aggiorna la base dati per le future predizioni. Un punto importante è

## 6.2 CODIFICA DELLE FORME D'ONDA

che il quantizzatore non è uniforme ma ha una caratteristica di quantizzazione come in figura 15.5. Quando l'ingresso del quantizzatore è grande

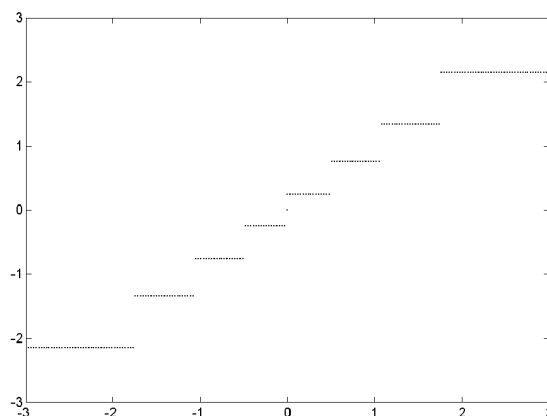


Figura 6.5: Quantizzatore non uniforme a passi crescenti. I livelli di restituzione sono posizionati in modo da minimizzare l'errore massimo

(forte errore di predizione) l'errore di quantizzazione sarà anche grande; questo corrisponde però ad un momento di minore sensibilità percettiva e pertanto l'efficacia è buona. Il vantaggio di un quantizzatore non uniforme è soprattutto evidente per densità di probabilità non uniformi, quale è quella dell'errore di predizione del DPCM, molto vicina ad una Laplaciana, e cioè del tipo:

$$p(x) = \frac{x_0}{2} e^{-|x|/x_0} \quad (6.3)$$

Panter e Dite (1951) hanno dimostrato che l'errore quadratico medio dovuto alla quantizzazione si minimizza se i quanti sono approssimativamente proporzionali all'inverso della radice cubica della densità di probabilità: in questo caso risulta anche che i contributi di tutti i livelli al valore q.m. dell'errore di quantizzazione totale sono circa eguali. In altre parole, i livelli esterni corrispondenti ai valori maggiori della variabile  $x$ , sono sì interessati più di rado dal quantizzatore, ma contribuiscono con errori maggiori all'errore totale, bilanciando la rarità. In conclusione, nel caso del DPCM, gli intervalli di quantizzazione crescono in modo

## CAPITOLO 6 IL SEGNALE VOCALE

approssimativamente esponenziale. Sia il predittore che il quantizzatore possono essere resi adattativi ed in questo caso il sistema viene indicato con il nome di ADPCM (Adaptive DPCM).

Codificatori ADPCM sono normalizzati ed utilizzati ad esempio con ritmo di informazione di  $32 \text{ kbits/sec}$  con un risparmio di un fattore 2 rispetto al consueto  $64 \text{ kbit/sec}$  del PCM (Pulse Code Modulation). In entrambi i casi la frequenza di campionamento è  $8 \text{ KHz}$ ; con il PCM, in questo caso, i campioni vengono trasmessi indipendentemente, usando 8 bit/campione. La caratteristica di quantizzazione usuale del PCM è di tipo logaritmico, ed assegna 5 bit alla mantissa e 3 alla caratteristica.

### 6.2.3 La codifica lineare predittiva

Una tecnica molto usata per la codifica del segnale vocale è quella della codifica lineare predittiva; il segnale campionato a  $8 \text{ KHz}$  viene segmentato in spezzoni della durata di circa 100 campioni ( $12,5 \text{ msec}$ , un intervallo di tempo pari circa a quello del timbro). Per ognuno di questi intervalli viene fatta un'analisi spettrale AR del segnale e vengono determinati 10 coefficienti di riflessione, che vengono trasmessi al ricevitore. L'errore di predizione corrispondente all'uscita del filtro tutti zeri caratterizzato da quei coefficienti di riflessione viene poi anch'esso codificato e trasmesso. Esistono numerose versioni del codificatore e le metodologie di codifica sia dei coefficienti di riflessione che dell'errore di predizione cambiano con il tipo di codificatore. Facciamo una breve valutazione del ritmo di informazione necessario: assegnando 4 bit/coefficiente di riflessione e 10 bit sia per l'ampiezza che per la frequenza del timbro, si arriva ad un totale di  $60 \times 10^3 : 12,5 = 4800 \text{ bits/sec}$ . Inoltre servono segnali per comunicare la situazione vocalizzata/non vocalizzata ecc. Si vede che anche per una versione semplificata del codificatore, sono richiesti comunque circa  $5 \text{ kbit/sec}$ . Esistono numerose possibilità per codificare i coefficienti di riflessione; abbiamo già fatto uso della loro resistenza al rumore di quantizzazione indicando la scarsa sensitività dei filtri a traliccio alla quantizzazione dei coefficienti. Si può fare meglio utilizzando le LSP (Line Spectral Pairs) e cioè le coppie di righe spettrali. Chiudendo il sistema riverberante corrispondente al tratto vocale su un undicesimo coefficiente di riflessione  $c_{11} = \pm 1$  abbiamo diversi effetti:

a) i poli del sistema si portano tutti sul cerchio unitario, visto che il sistema è disadattato dalle due parti, e non c'è perdita di energia;

### 6.3 LA CODIFICA DEL SEGNALE MUSICALE

b) i poli che risultano con il coefficiente  $c_{11} = -1$  si alternano con quelli corrispondenti a  $c_{11} = 1$ .

E' quindi possibile codificare le 10 distanze lungo il cerchio unitario (e quindi differenze di frequenze) che separano i poli dell'uno e dell'altro tipo. In questa maniera si riesce ad ottenere il massimo dell'efficienza nella rappresentazione del tratto vocale. Inoltre, invece di trasmettere ad ogni intervallo di tempo di circa  $12,5 \text{ msec}$  questa informazione, si possono trasmettere solo le variazioni di queste posizioni, quantizzate con tecniche *ADPCM*. L'errore di predizione relativo ai 100 campioni può venire trasmesso scegliendo il più adatto tra numerosi vettori casuali, di 100 campioni, memorizzati sia in trasmissione che in ricezione. Con tecniche di questo tipo (CELP; Codebook Excited Linear Prediction) e con ritmi di trasmissione di circa  $8 \text{ kbit/sec}$  si riesce ad avere una qualità sufficiente.

## 6.3 La codifica del segnale musicale

Il segnale musicale ha banda di circa  $20 \text{ kHz}$  e natura molto più complessa del segnale vocale; se per quest'ultimo sono i meccanismi di generazione a dare lo spunto alle tecniche di codifica, nel caso del segnale musicale ciò non è possibile; si ricorre allora alle caratteristiche del processo uditivo ed in particolare al fenomeno del mascheramento. Se in un segnale è presente un forte tono ad una frequenza  $f_0$ , il rumore che disturba le componenti alle frequenze  $f > f_0$  è molto poco percepibile dall'orecchio umano.

Le tecniche di codifica ne profittano e sono di questo tipo: il segnale viene suddiviso in bande con un banco ad esempio di 32 filtri di larghezza non uniforme (le bande hanno larghezza crescente da qualche centinaio di  $\text{Hz}$  a qualche  $\text{kHz}$ ); i segnali in ogni banda sono quantizzati con tecniche *ADPCM*. I ritmi di trasmissione disponibili per ogni banda sono calcolati, intervallo di tempo per intervallo di tempo (durata di circa  $10 \text{ msec}$ ) sulla base della percettibilità del rumore di quantizzazione; viene cioè fatta un'analisi spettrale (non parametrica) del segnale, trovata la frequenza a cui il segnale ha potenza massima e viene ridotto il numero di bit assegnato a tutte le bande al di sopra di questa.

In questo modo si riesce a rendere molto meno percepibile il rumore, malgrado il notevole risparmio di bit; una buona codifica riesce ad



## CAPITOLO 6 IL SEGNALE VOCALE

avere dei risultati eccellenti con ritmi dell'ordine di 150 - 200  $kbit/sec$ , molto al di sotto del ritmo corrispondente al Compact Disk ( $48KHz \times 16bit \times 2(Stereo) = 1,536Mbit/sec$ ). Si raggiungono pertanto fattori di riduzione dell'ordine di 8, simili a quelli ottenuti per il segnale vocale.

### 6.4 Cenni al riconoscimento del parlato

Le tecniche di riconoscimento del parlato non sono ancora completamente mature, ma è comunque possibile fare un cenno.

Le prime tecniche utilizzate cercavano di analizzare l'evoluzione delle formanti, per riconoscere in qualche modo le vocali pronunziate dal parlatore. Tuttavia, si è visto come queste tecniche fossero abbastanza fragili; si è allora cercato di studiare l'evoluzione temporale dei coefficienti di riflessione ma per fare ciò è necessario:

- individuare inizio e fine di ogni parola pronunciata;

- infatti, il riconoscimento del parlato continuo è molto più difficile del riconoscimento fatto parola per parola.

- fare uno stiramento dinamico dell'asse temporale per compensare effetti di variazione della velocità di fonazione da parlatore a parlatore; ogni parola viene confrontata con ogni altra parola del vocabolario; le due parole a confronto sono *stirate* in modo da avere un'evoluzione temporale quanto possibile parallela. Infine, si cerca la parola più simile alla parola pronunciata.

# Capitolo 7

## Segnali bidimensionali

### 7.1 Le singolarità a 2D

Il passaggio da una a due dimensioni non comporta gravi difficoltà per il trattamento dei segnali, a parte i costi computazionali e la memoria occupata, che cresce notevolmente. Dal punto di vista algoritmico, vi sono alcuni elementi innovativi, che cercheremo di riassumere; tratteremo inizialmente i sistemi continui. La prima differenza significativa tra i sistemi 2D e quelli 1D è nelle singolarità: mentre nei sistemi 1D l'unica singolarità era l'impulso, ora troviamo sia *spilli* e cioè impulsi bidimensionali veri e propri, sia *lama* e cioè segnali impulsivi in una dimensione, ma continui nell'altra. Evidentemente, le singolarità a lama non hanno corrispondenza a 1D. Consideriamo un generico segnale a 2D  $f(x, y)$ : esso può essere **separabile** e cioè

$$f(x, y) = a(x) \times b(y) \quad (7.1)$$

Si vedrà poi che si può sempre rappresentare un segnale qualsiasi come:

$$f(x, y) = \sum_{i=0}^{+\infty} a_i(x) \times b_i(y) \quad (7.2)$$

e quindi la separabilità è un caso particolare od un'approssimazione. Una semplice singolarità a *lama* è un caso particolare di segnale separabile:

$$f_1(x, y) = a_1(x) \times \delta(y) \quad (7.3)$$

## CAPITOLO 7 SEGNALI BIDIMENSIONALI

Questa è una lama diretta lungo l'asse delle  $x$  e di altezza variabile  $a_1(x)$ ; similmente possiamo pensare a singolarità a lama disposte in altro modo nel piano.

La grande differenza tra i sistemi 2D e quelli 1D nasce dal fatto che in 2D la rapidità del transitorio può dipendere dalla direzione del moto sulla superficie; usando un'espressione proverbiale, solo se "*si va sul filo di un rasoio*" è possibile evitare violenti e indesiderati transitori. Una singolarità a *spillo* è il prodotto di due singolarità a lama:

$$\delta(x, y) = \delta(x) \times \delta(y) \quad (7.4)$$

Da queste definizioni si vede subito che moltiplicare per una lama e integrare a 2D equivale a fare un integrale di linea; invece moltiplicando per uno spillo e ancora integrando a 2D, si preleva il valore della funzione, come accade con le singolarità 1D:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) a_1(x) \delta(y) dx dy = \int_{-\infty}^{+\infty} f(x, 0) a_1(x) dx \quad (7.5)$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(x, y) dx dy = f(0, 0) \quad (7.6)$$

Bisogna stare attenti ai fattori di scala delle variabili di cui le  $\delta(\cdot)$  sono funzione. Infatti con semplici sostituzioni di variabili, si vede che:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta(\alpha x, \beta y) dx dy = \frac{f(0, 0)}{\alpha \beta} \quad (7.7)$$

In 2D ancora, risulta che lo *spillo* che nasce dall'incrocio di due singolarità a lama, dipende anche dall'angolo che esse formano tra loro:

$$I = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x, y) \delta[\alpha(x \cos \varphi_0 + y \sin \varphi_0), \beta(x \cos \varphi_1 + y \sin \varphi_1)] dx dy = \quad (7.8)$$

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) \delta(\xi, \eta) d\xi d\eta = \frac{f(0, 0)}{\alpha \beta \sin |\varphi_1 - \varphi_0|}$$

La formula precedente è ottenibile con un cambiamento di coordinate 2D, in cui l'elemento d'area del nuovo sistema è ottenuto da quello vecchio moltiplicando per lo Jacobiano  $J$  (determinante della matrice di trasformazione di coordinate, locale) della trasformazione di coordinate.

## 7.1 LE SINGOLARITÀ A 2D

Ricordiamo qui che per un sistema di due funzioni in due variabili, come nel caso di una trasformazione di coordinate a 2D

$$\begin{aligned}\xi &= \xi(x, y) \\ \eta &= \eta(x, y)\end{aligned}$$

lo Jacobiano della trasformazione è:

$$J = \begin{vmatrix} \frac{\partial \xi}{\partial x} & \frac{\partial \xi}{\partial y} \\ \frac{\partial \eta}{\partial x} & \frac{\partial \eta}{\partial y} \end{vmatrix} = \frac{\partial (\xi, \eta)}{\partial (x, y)}$$

La formula può apparire poco comprensibile, ma basta pensare alla superficie (p.e. in  $m^2$ ) che viene coperta se le due nuove coordinate crescono rispettivamente di  $\delta\xi, \delta\eta$ . Per esempio, nel caso di passaggio da trasformazione di coordinate da cartesiane a polari, la superficie coperta incrementando le nuove variabili è  $\rho\delta\rho\delta\theta$  (*lontano dal centro le aree fabbricabili costano meno*). Infatti, il valore dello Jacobiano è dipendente dalle coordinate:

$$\begin{aligned}x &= \rho \cos \theta; \quad y = \rho \sin \theta \\ dx &= d\rho \cos \theta - \rho \sin \theta d\theta; \quad dy = d\rho \sin \theta + \rho \cos \theta d\theta \\ |J_{cp}| &= \frac{\partial (x, y)}{\partial (\rho, \theta)} = \begin{vmatrix} \frac{\partial x}{\partial \rho} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial \rho} & \frac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos \theta & -\rho \sin \theta \\ \sin \theta & \rho \cos \theta \end{vmatrix} \\ |J_{cp}| &= \rho(\cos^2 \theta + \sin^2 \theta) = \rho = \sqrt{x^2 + y^2}\end{aligned}$$

e pertanto gli elementi di area sono:

$$\begin{aligned}dxdy &\Leftrightarrow \rho d\rho d\theta \\ \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) dxdy &= \int_{-\pi}^{\pi} \int_0^{\infty} f(\rho, \theta) |\rho| d\rho d\theta\end{aligned}$$

indicando in modo poco corretto ma intuibile:

$$f(\rho, \theta) = f(x = \rho \cos \theta, y = \rho \sin \theta)$$

Nel caso della trasformazione inversa da coordinate cartesiane a polari si ha:

$$\begin{aligned}\rho &= \sqrt{x^2 + y^2} \\ \theta &= \arctan \frac{y}{x}\end{aligned}$$

## CAPITOLO 7 SEGNALI BIDIMENSIONALI

$$|J_{pc}| = \begin{vmatrix} \frac{\partial \rho}{\partial x} & \frac{\partial \rho}{\partial y} \\ \frac{\partial \theta}{\partial x} & \frac{\partial \theta}{\partial y} \end{vmatrix}$$

$$|J_{pc}| = \begin{vmatrix} \frac{x}{\sqrt{x^2+y^2}} & \frac{y}{\sqrt{x^2+y^2}} \\ -\frac{y}{x^2+y^2} & \frac{x}{x^2+y^2} \end{vmatrix} = \frac{1}{|\rho|} = \frac{1}{|J_{cp}|}$$

Nel caso dell'incrocio di due lame, si ha invece che lo Jacobiano della trasformazione di coordinate che porta gli assi coordinati sulle due lame è una costante:

$$\xi = \alpha(x \cos \varphi_0 + y \sin \varphi_0); \quad \eta = \beta(x \cos \varphi_1 + y \sin \varphi_1);$$

$$d\xi d\eta = |J| dx dy; \quad |J| = \left| \frac{\partial(\xi, \eta)}{\partial(x, y)} \right| = |\alpha\beta \sin(\varphi_0 - \varphi_1)|$$

La motivazione di questo fattore è resa intuitiva pensando a due lame, di spessore finito  $\Delta$  e di altezza  $h$ , che si intersecano; il volume del prisma di base rombica, loro intersezione e quindi costruito sull'area coperta con l'incremento delle due variabili, è pari a  $\Delta^2 h \sin |\varphi_1 - \varphi_0|$  e cresce se le due lame tendono ad essere parallele (fig. 16.1). Un'importante applicazione

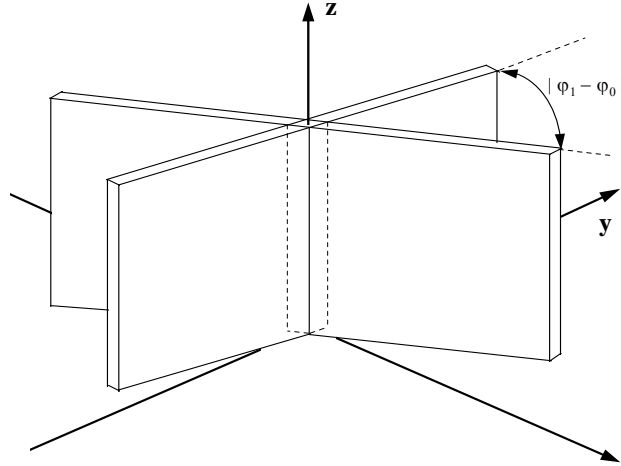


Figura 7.1: Intersezione di due lame di spessore finito

delle singolarità impulsive 2D (a spillo), è la decomposizione di un segnale

2D in somma di impulsi:

$$f(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) \delta(x - \xi, y - \eta) d\xi d\eta$$

## 7.2 Sistemi lineari 2D

Un sistema è caratterizzato da un ingresso ed un'uscita; se il sistema è lineare, vale la sovrapposizione degli effetti e visto che possiamo decomporre l'ingresso in somma di funzioni anche l'uscita sarà scomponibile nella somma delle risposte a quelle funzioni. Consideriamo il caso dell'impulso, posizionato in  $x_0, y_0$ ; in generale, la risposta del sistema è dipendente dalla posizione dell'impulso di ingresso  $h(x, y; x_0, y_0)$ . Risulta allora, che se all'ingresso di questo sistema vi è la funzione  $f(x, y)$  l'uscita è:

$$g(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) h(x, y; \xi, \eta) d\xi d\eta \quad (7.9)$$

In generale i sistemi saranno spazio - varianti e cioè la risposta all'impulso dipenderà dalla posizione dell'impulso stesso; per fare un esempio banale, il suono generato battendo una lastra metallica non uniforme dipende dalla posizione del punto di percussione. Se invece il sistema è **spazio invariante** allora la formula precedente si semplifica perchè:

$$h(x, y; x_0, y_0) = h(x - x_0, y - y_0) \quad (7.10)$$

$$g(x, y) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) h(x - \xi, y - \eta) d\xi d\eta = f(x, y) ** h(x, y) \quad (7.11)$$

Questo è l'**integrale di convoluzione a 2D** che indicheremo con il segno \*\*. La convoluzione di una funzione ed uno spillo centrato nel punto di coordinate  $x_0, y_0$  è la stessa funzione traslata dall'origine nel punto di coordinate  $x_0, y_0$ ; la convoluzione di due lame finite ortogonali è un prisma rettangolare (fig. 16.2); la convoluzione di due *sinc* è un *sinc2D* (fig. 16.3).

### 7.2.1 Le funzioni cilindriche

Una funzione 2D che dipende solo da una variabile è detta una funzione cilindrica: ad esempio:

$$f_1(x, y) = a(x); \quad f_2(x, y) = b(x \cos \theta - y \sin \theta) \quad (7.12)$$

## CAPITOLO 7 SEGNALI BIDIMENSIONALI

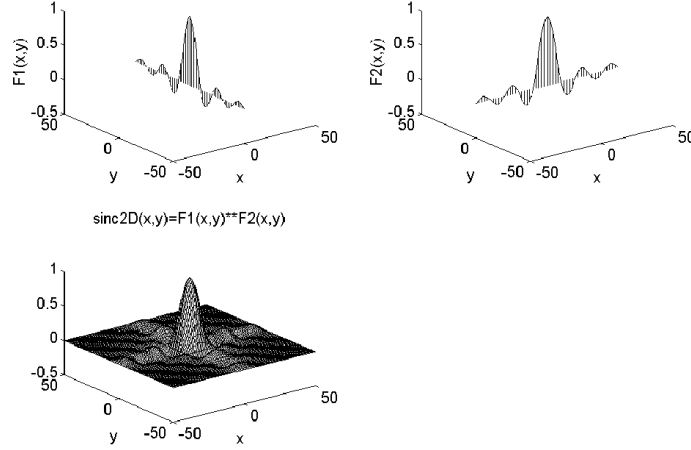


Figura 7.2: Risultato della convoluzione di due lame a forma di *sinc*.

La funzione  $f_1(x, y)$  è costante nella direzione  $y$ , che definiamo come direzione di cilindricità; similmente la funzione  $f_2(x, y)$  ha come direzione di cilindricità la retta che forma l'angolo  $\theta$  con l'asse delle  $y$ . Convolviendo una funzione qualsiasi con una funzione cilindrica si ottiene ancora una funzione cilindrica, con la stessa direzione di cilindricità:

$$\begin{aligned}
 g(x) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) h(x - \xi) d\xi d\eta = \\
 &= \int_{-\infty}^{+\infty} \left( \int_{-\infty}^{+\infty} f(\xi, \eta) d\eta \right) h(x - \xi) d\xi = \int_{-\infty}^{+\infty} p(\xi) h(x - \xi) d\xi
 \end{aligned} \tag{7.13}$$

Indicheremo con:

$$p(\xi) = \int_{-\infty}^{+\infty} f(\xi, \eta) d\eta \tag{7.14}$$

la **proiezione** della  $f(x, y)$  lungo l'asse  $y$ , direzione di cilindricità della  $h(x)$ . Ovviamente, può essere qualsiasi l'azimut di cilindricità (e cioè l'angolo con l'asse  $x$  della direzione lungo cui la funzione  $f(x, y)$  è costante).

### 7.3 Le sinusoidi 2D

Nei sistemi 1D, convolvendo qualsiasi funzione per una sinusoida complessa, otteniamo ancora una sinusoida isofrequenziale, ma di diversa ampiezza complessa. In 2D, avviene la stessa cosa con le funzioni complesse:

$$s(x, y) = e^{j(ux+vy)} \quad (7.15)$$

Infatti:

$$\begin{aligned} g(x, y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) e^{j[u(x-\xi)+v(y-\eta)]} d\xi d\eta = \\ &= e^{j(ux+vy)} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) e^{-j(u\xi+v\eta)} d\xi d\eta = e^{j(ux+vy)} F(u, v) \\ F(u, v) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) e^{-j(u\xi+v\eta)} d\xi d\eta \end{aligned} \quad (7.16)$$

La funzione  $F(u, v)$  è la trasformata di Fourier 2D della funzione  $f(x, y)$  e dà la nuova fase ed ampiezza della sinusoida 2D all'uscita del sistema. E' interessante capire la forma di queste sinusoidi 2D: per fare questo passiamo da coordinate cartesiane a coordinate polari:

$$\begin{aligned} x &= \rho \cos \theta; \quad y = \rho \sin \theta; \quad u = \Omega \cos \varphi; \quad v = \Omega \sin \varphi \\ \rho &= \sqrt{x^2 + y^2}; \quad \theta = \tan^{-1} \frac{y}{x} \quad \Omega = \sqrt{u^2 + v^2}; \quad \varphi = \tan^{-1} \frac{v}{u} \\ s(x, y) &= e^{j(ux+vy)}; \quad s'(\rho, \theta) = e^{j\Omega\rho \cos(\theta-\varphi)} \end{aligned} \quad (7.17)$$

La funzione  $s'(\rho, \theta)$  è un sinusoida di **numero d'onda** (pulsazione spaziale)  $\Omega \cos(\theta - \varphi)$  ed è quindi costante nella direzione  $\theta = \varphi + \pi/2$  (le sinusoidi 2D sono quindi funzioni separabili cilindriche, con direzione di cilindricità  $\varphi + \pi/2$ ) ed ha invece la pulsazione massima  $\Omega$  nella direzione  $\varphi$ ; sostanzialmente è come una *tettoia ondulata*; uniforme nella direzione  $\varphi + \pi/2$ , ondulata nella direzione  $\varphi$  (fig. 16.4).

### 7.4 L'antitrasformazione di Fourier 2D

Come un impulso ideale 1D è somma delle sinusoidi a tutte le frequenze di ampiezza unitaria e fase zero, così un impulso 2D è somma di tutte le



## CAPITOLO 7 SEGNALI BIDIMENSIONALI

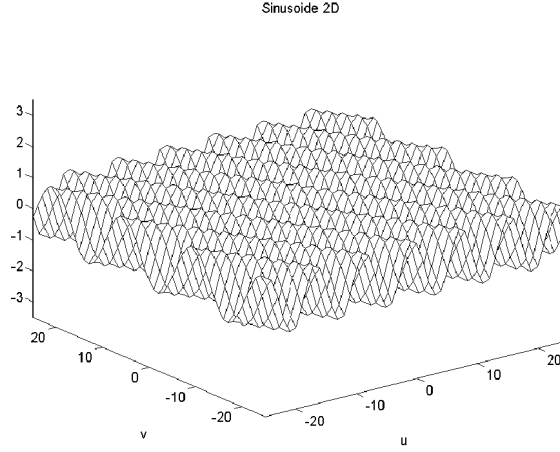


Figura 7.3: Sinusoide 2D (*tettoia ondulata*).

sinusoidi 2D a tutti i numeri d'onda, con fase nulla ed ampiezza costante. Infatti:

$$\begin{aligned}\delta(x, y) &= \delta(x)\delta(y) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{jux} du \times \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{jvy} dv = \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{j(ux+vy)} dudv\end{aligned}\quad (7.18)$$

Combinando la formula di composizione con la formula di trasformazione, se ne deduce la formula di antitrasformazione a 2D:

$$\begin{aligned}f(x, y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) \delta(x - \xi, y - \eta) d\xi d\eta = \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(\xi, \eta) e^{j[u(x-\xi)+v(y-\eta)]} dudvd\xi d\eta = \\ &= \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} F(u, v) e^{j(ux+vy)} dudv\end{aligned}\quad (7.19)$$

Le formule di trasformazione sono utili anche in coordinate polari:

$$f(\rho, \theta) = \frac{1}{4\pi^2} \int_0^{+\infty} \int_{-\pi}^{+\pi} F(\Omega, \varphi) e^{j\Omega\rho \cos(\theta-\varphi)} \Omega d\Omega d\varphi \quad (7.20)$$

#### 7.4 L'ANTITRASFORMAZIONE DI FOURIER 2D

dove è stato inserito lo Jacobiano della trasformazione da coordinate cartesiane a polari,  $\Omega$ . Nel caso dell'impulso ideale, esso è l'antitrasformata di una costante e pertanto si ha:

$$\delta(\rho) = \frac{1}{4\pi^2} \int_0^{+\infty} \int_{-\pi}^{+\pi} e^{j(\Omega\rho \cos(\theta-\varphi))} \Omega d\Omega d\varphi \quad (7.21)$$

Quest'ultima formula può anche essere interpretata osservando che per comporre un impulso ideale, per ogni numero d'onda  $\Omega$  serve un *certo numero* di sinusoidi spaziali, componenti, che siano uniformemente distribuite su tutti gli azimut  $\varphi$ ; questo *numero di sinusoidi* cresce proporzionalmente ad  $\Omega$  così da garantire una densità uniforme delle sinusoidi nel piano.

##### 7.4.1 Funzioni a simmetria circolare

Se la funzione da trasformare è a simmetria circolare, lo sarà anche la trasformata, e viceversa. In questo caso la trasformazione si semplifica e diventa:

$$f(\rho) = \frac{1}{4\pi^2} \int_0^{+\infty} \int_{-\pi}^{+\pi} F(\Omega) e^{j(\Omega\rho \cos(\theta-\varphi))} \Omega d\Omega d\varphi = \quad (7.22)$$

$$\begin{aligned} & \frac{1}{2\pi} \int_0^{+\infty} \Omega F(\Omega) \left( \frac{1}{2\pi} \int_{-\pi}^{+\pi} e^{j(\Omega\rho \cos(\theta-\varphi))} d\varphi \right) d\Omega = \\ & = \frac{1}{2\pi} \int_0^{+\infty} \Omega F(\Omega) J_0(\Omega\rho) d\Omega \end{aligned}$$

dove la parentesi è stata sostituita con il suo valore e cioè la funzione di Bessel di ordine 0,  $J_0(\Omega\rho)$ . La trasformazione 1D corrispondente all'equazione precedente viene indicata come trasformata di **Hankel**. Segue ora un'utile tabella di trasformate a simmetria circolare.

## CAPITOLO 7 SEGNALI BIDIMENSIONALI

<i>Dominio spazi</i>	<i>Frequenze spaziali</i>
$e^{-\pi\rho^2}$	$e^{-\pi\Omega^2}$
$\frac{1}{\sqrt{a^2+\rho^2}}$	$\frac{1}{\Omega}e^{-2\pi a\Omega}$
$\frac{1}{(a^2+\rho^2)^{3/2}}$	$\frac{2\pi}{a}e^{-2\pi a\Omega}$
$\frac{1}{\rho}$	$\frac{1}{\Omega}$
$e^{-a\rho}$	$\frac{2\pi a}{(a^2+4\pi^2\Omega^2)^{3/2}}$
$\frac{e^{-a\rho}}{\rho}$	$\frac{2\pi}{\sqrt{a^2+4\pi^2\Omega^2}}$
$\frac{\delta(\rho)}{2\pi\rho}$	1
$\rho^2 e^{-\pi\rho^2}$	$\left(\frac{1}{\pi} - \Omega^2\right) e^{-\pi\Omega^2}$
$-\rho^2 f(\rho)$	$\frac{d^2 F}{d\Omega^2} + \frac{1}{\Omega} \frac{dF}{d\Omega} = \nabla^2 F$

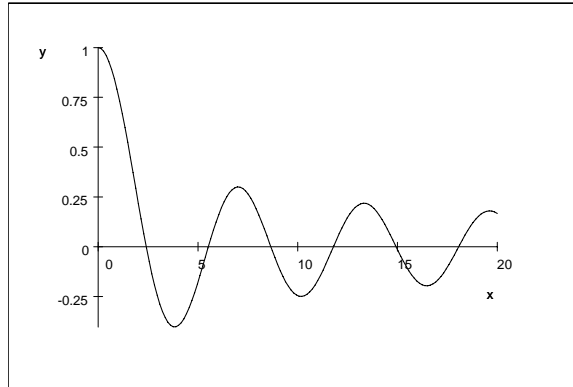
Da: Bracewell (1999) *The Fourier Transform and its applications*,  
McGraw Hill.

La circonferenza  $\delta(\Omega - \Omega_0)$ , ha per antitrasformata di Hankel la funzione  $J_0(\Omega_0\rho)$  che può apparire simile ad un impulso ideale filtrato passa basso con un filtro ideale a simmetria circolare. L'impulso filtrato passa basso ha trasformata:

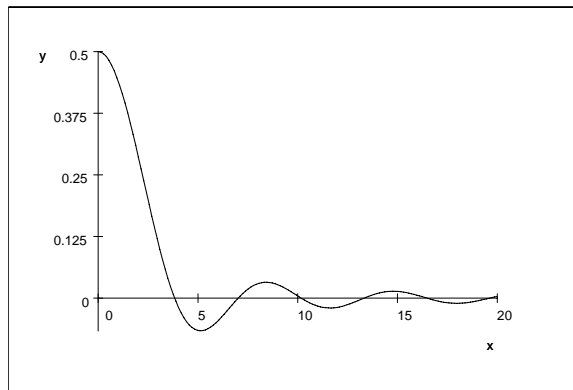
$$\frac{1}{2\pi} \int_0^{\Omega_0} \Omega J_0(\Omega\rho) d\Omega = \frac{\Omega_0 J_1(\Omega_0\rho)}{2\pi\rho} \quad (7.23)$$

Non ci si lasci ingannare: la convergenza a zero delle funzioni  $J_m(x)$  (vedi le figure che seguono)

## 7.4 L'ANTITRASFORMAZIONE DI FOURIER 2D



La funzione di Bessel  $J_0(x)$ .



La funzione di Bessel  $J_1(x)/x$

per  $x \rightarrow \infty$ ,  $m$  intero, è solo come

$$J_m(x) \sim \sqrt{\frac{2}{\pi|x|}} \cos\left(x - \frac{m\pi}{2} - \frac{\pi}{4}\right)$$

e quindi un qualsiasi reticolo regolare di impulsi convoluti con  $J_0(\Omega_0\rho)$  si *impasta* e sparisce o cambia drasticamente forma. Non così nel caso della 7.23. La cosa si può vedere anche nel dominio delle frequenze: convolvere con un reticolo regolare di impulsi equivale (come vedremo nel prossimo capitolo) a moltiplicare per un reticolo regolare di impulsi nelle frequenze spaziali. Questo reticolo, se moltiplicato per la circonferenza  $\delta(\Omega - \Omega_0)$ , trasformata della  $J_0$ , si annulla ovunque, tranne al più in qualche posizione particolare.

## 7.5 Proprietà della trasformazione

Riepiloghiamo alcune delle proprietà della trasformata di Fourier 2D, estensioni di quella 1D; indichiamo rispettivamente con  $f(x, y)$  la funzione nel dominio degli spazi e  $F(u, v)$  la sua trasformata:

$$f(x, y) \leftrightarrow F(u, v) \quad (7.24)$$

$$f(\alpha x, \beta y) \leftrightarrow \frac{1}{\alpha\beta} F\left(\frac{u}{\alpha}, \frac{v}{\beta}\right); \quad (7.25)$$

$$\begin{aligned} \frac{\partial f(x, y)}{\partial x} &\leftrightarrow juF(u, v); \quad \frac{\partial f(x, y)}{\partial y} \leftrightarrow jvF(u, v); \\ \frac{\partial^2 f(x, y)}{\partial x \partial y} &\leftrightarrow -uvF(u, v) \end{aligned} \quad (7.26)$$

Al solito, le relazioni di derivazione nascono da un'integrazione per parti. Se  $f(x, y)$  è reale:

$$F(u, v) = F^*(-u, -v) \quad (7.27)$$

$$f(\rho, \theta + \psi) \leftrightarrow F(\Omega, \varphi + \psi) \quad (7.28)$$

La relazione di Parseval diventa:

$$\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |f(x, y)|^2 dx dy = \frac{1}{4\pi^2} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |F(u, v)|^2 du dv \quad (7.29)$$

Se abbiamo un'equazione differenziale alle derivate parziali, definita su tutto lo spazio (le condizioni al contorno sono all'infinito), essa corrisponde alla seguente equazione algebrica sulla trasformata:

$$\sum_{i,h=0}^{I,H} \frac{\partial^i \partial^h f(x, y)}{\partial x^i \partial y^h} \leftrightarrow \sum_{i,h=0}^{I,H} (ju)^i (jv)^h F(u, v) \quad (7.30)$$

### 7.5.1 Funzioni cilindriche e separabili

La trasformata di Fourier di un impulso 2D posizionato in  $x_0, y_0$  è il piano di fase:

$$\delta(x - x_0, y - y_0) \leftrightarrow e^{-j(ux_0 + vy_0)} \quad (7.31)$$

Una cosinusoide ha invece trasformata:

$$\cos(u_0 x + v_0 y + \varphi) \leftrightarrow \frac{1}{2} [e^{-j\varphi} \delta(u - u_0, v - v_0) + e^{+j\varphi} \delta(u + u_0, v + v_0)] \quad (7.32)$$

## 7.5 PROPRIETÀ DELLA TRASFORMAZIONE

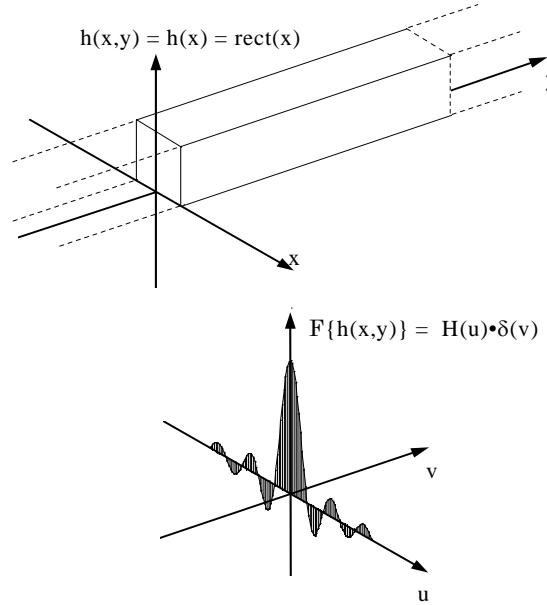


Figura 7.4: Trasformata di Fourier di una funzione cilindrica

La trasformata di Fourier di una funzione cilindrica è una *lama* nella variabile coniugata alla variabile lungo cui la funzione cilindrica è uniforme: ad esempio:

$$h(x, y) = h(x) \leftrightarrow \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(x) e^{-j(ux+vy)} dx dy = H(u) \delta(v) \quad (7.33)$$

La superficie cilindrica  $h(x)$  è costante nella direzione  $y$  e quindi nel dominio coniugato troviamo una funzione impulsiva in quella direzione (lama allungata lungo le  $u$ ). Una lama uniforme, caso particolare di funzione cilindrica, ha per trasformata un'altra lama uniforme disposta lungo la coniugata della variabile ortogonale (una lama lungo le  $x$  ha come trasformata una lama lungo le  $v$ ). La trasformata di una somma di due lame ortogonali è quindi ancora una somma di due lame ortogonali. Ancora, una stella di lame sarà eguale alla sua trasformata (a meno di una rotazione di  $\pi/2$ ). La somma di infinite lame, come si vedrà al capitolo 19, è una funzione del tipo  $1/\rho$  ed ha come trasformata la stessa funzione  $1/\Omega$ .

## CAPITOLO 7 SEGNALI BIDIMENSIONALI

Una funzione separabile, prodotto di due funzioni 1D, ha come trasformata 2D il prodotto delle due trasformate 1D.

### 7.6 Filtri ideali 2D

I filtri ideali a 2D hanno una funzione di trasferimento unitaria in alcune zone del dominio  $u, v$  e nulla altrove. A differenza che ad 1D, la forma del filtro potrà cambiare. Il caso semplice è quello del filtro a base rettangolare e cioè:

$$\begin{aligned} F(u, v) &= 1; |u|, |v| < \frac{\pi}{\Delta} \\ F(u, v) &= 0; |u|, |v| > \frac{\pi}{\Delta} \end{aligned} \quad (7.34)$$

Questo è un filtro separabile: infatti è ottenibile come la cascata di due filtri passa basso ideali cilindrici in  $u, v$  successivamente (fig. 16.6). Un

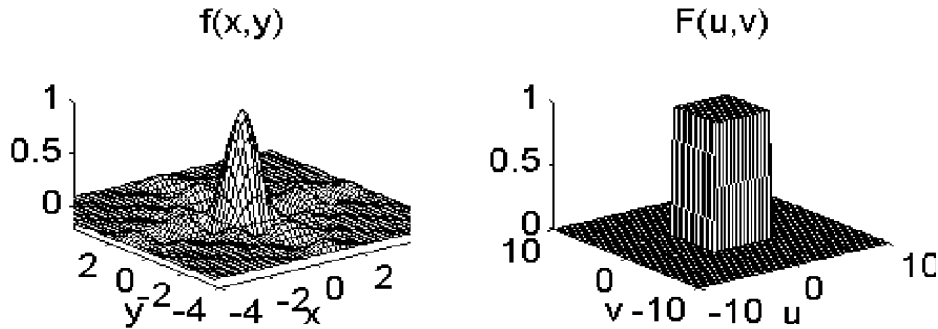


Figura 7.5: Trasformata 2D del *sinc*

filtro solo nelle  $u$  ha funzione di trasferimento a lama lungo le  $x$  e quindi ritroviamo il risultato già visto che:

$$F(u, v) \leftrightarrow f(x, y) = \frac{1}{\Delta^2} \frac{\sin(\pi x/\Delta)}{(\pi x/\Delta)} \frac{\sin(\pi y/\Delta)}{(\pi y/\Delta)} \quad (7.35)$$

Esiste però anche il filtro ideale circolare e cioè:

$$H(\Omega) = 1; \Omega < \Omega_0; \leftrightarrow h(\rho) = \Omega_0 \frac{J_1(\Omega_0 \rho)}{\rho} \quad (7.36)$$

Dove  $J_1(\cdot)$  è la funzione di Bessel di ordine 1 (fig. 16.7).

## 7.7 FILTRI A VENTAGLIO

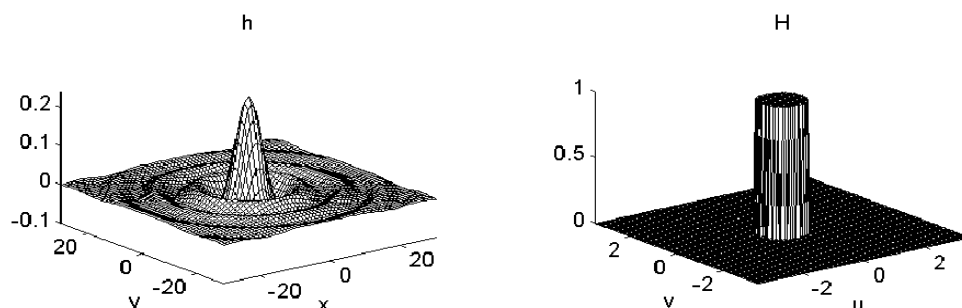


Figura 7.6: Il cilindro, trasformata 2D della funzione di Bessel di ordine 1

## 7.7 Filtri a ventaglio

Un altro tipo di filtri ideali inesistenti a 1D sono i filtri a ventaglio (fig. 16.8) e cioè filtri nulli in un settore angolare ed unitari altrove. Con

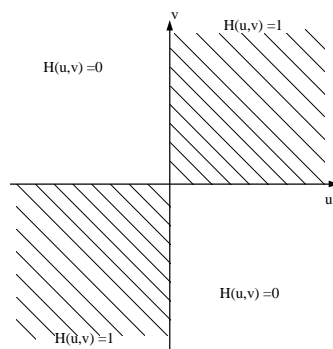


Figura 7.7: Funzione di trasferimento del filtro quadrantale

una cascata di rotazioni e scalamenti un qualsiasi filtro a ventaglio è trasformabile in un filtro quadrantale, nullo nel II e IV quadrante ed unitario altrove. Infatti, basta prima ruotare il filtro a ventaglio in modo che l'asse della zona "passa" sia ad esempio l'asse x. Poi, si scala l'asse y in modo che il settore angolare copra un angolo pari a  $\pi/2$  ed infine lo si ruota di  $\pi/4$  per portare il filtro a coincidere con il I e III quadrante.



## CAPITOLO 7 SEGNALI BIDIMENSIONALI

Questo filtro ha per funzione di trasferimento la somma di due funzioni separabili:

$$H(u, v) = 1; \quad uv > 0; \quad H(u, v) = 0; \quad uv < 0 \quad (7.37)$$

$$H(u, v) = \frac{1}{2}[1 + \text{sign}(u) \times \text{sign}(v)] \leftrightarrow h(x, y) \quad (7.38)$$

$$h(x, y) = \frac{1}{2} \left[ \delta(0, 0) - \frac{1}{\pi x} \times \frac{1}{\pi y} \right] \quad (7.39)$$

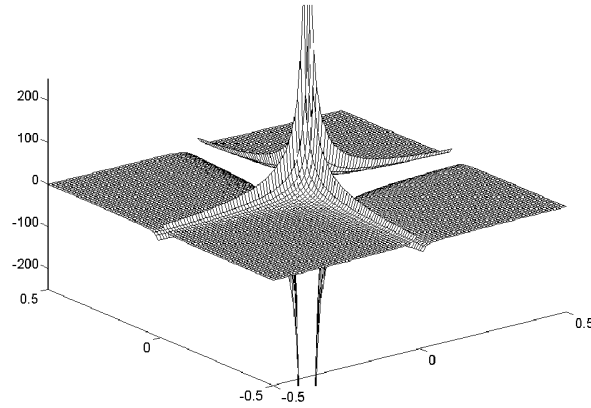


Figura 7.8: Risposta impulsiva del filtro quadrantale

## Capitolo 8

### Filtri 2D di natura fisica

Per formare un'intuizione dei meccanismi del filtraggio 2D, mostreremo ora come diversi fenomeni fisici possano essere interpretati come filtri lineari; in particolare ci riferiremo a delle equazioni differenziali alle derivate parziali ellittiche, iperboliche e paraboliche che origineranno funzioni di trasferimento a fase zero, di pura fase, e miste. Lo studente interessato è invitato a osservare le antitrasformate dei vari filtri qui derivati utilizzando ad esempio *MATLAB*<sup>TM</sup>.

#### 8.1 L'equazione dei potenziali

Si consideri il seguente problema fisico tridimensionale: nel sottosuolo (il semispazio  $z < 0$ ) vi siano dei corpi di densità differente da quella del materiale che li circonda. Inoltre il semispazio  $z > 0$  è vuoto (trascuriamo la densità dell'atmosfera; fig. 17.1). Per determinare la posizione e grandezza dei corpi disomogenei, si vuole determinare l'andamento del potenziale gravitazionale  $f(x, y, z)$  a livello del terreno, e cioè per  $z = 0$ . Se il terreno è poco praticabile, perchè vi sono boschi, paludi ecc. è più agevole misurare il potenziale ad una certa quota  $z_0$ , per esempio con misure eliportate, o aeroportate. Poi, si cercherà un algoritmo (un filtraggio lineare) per calcolare  $f(x, y, z = 0)$  a partire da  $f(x, y, z_0)$  (continuazione in basso di un campo gravitazionale). La funzione potenziale  $f(x, y, z)$  in zone prive di sorgenti soddisfa l'equazione armonica e cioè:

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} + \frac{\partial^2 f}{\partial z^2} = 0 \quad (8.1)$$

## CAPITOLO 8 FILTRI 2D DI NATURA FISICA

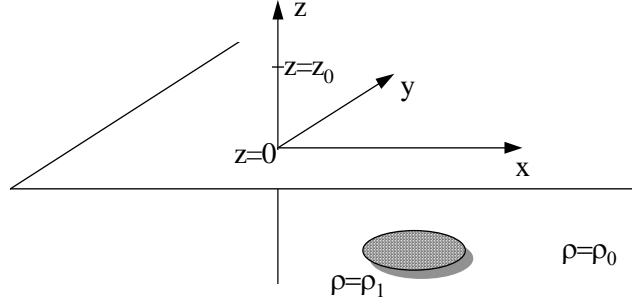


Figura 8.1: Modello di semispazio con materiali a densità diversi

Si desidera appunto, conosciuta  $f(x, y, z_0); \forall x, y$  trovare  $f(x, y, z = 0); \forall x, y$  e viceversa (continuazione in alto). Poichè la funzione  $f(x, y, z_0)$  è conosciuta per tutti i valori della  $x$  e della  $y$ , possiamo trasformare gli elementi dell'equazione (8.1) secondo Fourier e risulta l'equazione differenziale semplice, in cui le variabili coniugate  $u, v$  sono dei parametri:

$$-(u^2 + v^2)F(u, v, z) + \frac{d^2 F(u, v, z)}{dz^2} = 0; \quad F(u, v, z) \leftrightarrow f(x, y, z) \quad (8.2)$$

L'equazione ha la soluzione:

$$F(u, v, z) = A(u, v)e^{-\Delta z \sqrt{u^2 + v^2}} + B(u, v)e^{\Delta z \sqrt{u^2 + v^2}} \quad (8.3)$$

$$\Delta z = z - z_0; \quad \Omega = \sqrt{u^2 + v^2}$$

Imponiamo  $B(u, v) = 0$  per evitare che la soluzione diverga per  $z \rightarrow \infty$ , e poniamo  $A(u, v) = F(u, v, z_0)$ ; troviamo la soluzione:

$$\begin{aligned} F(u, v, z) &= F(u, v, z_0)e^{-(z-z_0)\Omega} = \\ &= F(u, v, z_0)e^{-\Delta z \Omega} = F(u, v, z_0) \times H(u, v, \Delta z); \end{aligned} \quad (8.4)$$

Decomponendo con la trasformazione di Fourier 2D il campo misurato a quota 0 in sinusoidi 2D, possiamo ottenere il campo misurabile a quota  $z_0$  purchè l'ampiezza di ogni componente (appunto  $F(u, v, z_0)$ ) sia scalata del fattore  $e^{-\Delta z \Omega}$ . In sostanza, la continuazione in basso o in alto del campo è ottenuta filtrando linearmente la funzione potenziale con un

## 8.1 L'EQUAZIONE DEI POTENZIALI

filtro 2D  $H(u, v, \Delta z)$  funzione della differenza di quote  $\Delta z$ . Si osservi che la funzione di trasferimento  $H(u, v)$  è a fase zero ed a simmetria circolare, come dettato dalla natura del problema. Il filtro è passa basso se la continuazione è verso l'alto; se ci si allontana dalle sorgenti il potenziale è sempre più smussato fino ad essere uniforme a distanze molto grandi rispetto alle dimensioni dei corpi. Se invece ci si avvicina alle sorgenti, il filtro è passa alto e amplifica esponenzialmente le componenti ai numeri d'onda più elevati. Su questo punto torneremo nel prossimo paragrafo, perchè l'inevitabile presenza di rumore modifica la situazione. Il filtro passa basso ha la sua antitrasformata:

$$H(u, v, \Delta z) = e^{-\Delta z \sqrt{u^2 + v^2}} = e^{-\Delta z \Omega} \quad (8.5)$$

$$e^{-\Delta z \Omega} \leftrightarrow \frac{\Delta z}{(x^2 + y^2 + \Delta z^2)^{3/2}}; \quad \Delta z \geq 0$$

Non così il filtro inverso  $\Delta z < 0$  che è divergente per numeri d'onda  $u, v$  abbastanza grandi. Utilizziamo questo esempio per mostrare come la tecnica di stima lineare viene in aiuto anche in questo caso per condurre a soluzioni significative.

### 8.1.1 Continuazione in basso e predizione lineare

Supponiamo che la misura  $f_m(x, y, z_0)$  del potenziale alla quota  $z_0$  sia contaminata da rumore:

$$f_m(x, y, z_0) = f(x, y, z_0) + n(x, y) \quad (8.6)$$

Vogliamo stimare in modo lineare ottimo  $f(x, y, z = 0)$  e sia  $\hat{f}(x, y, z = 0)$  questa stima. Facciamo la stima nel dominio dei numeri d'onda, per semplicità. Trasformando l'equazione precedente si ha:

$$F_m(u, v, z_0) = F(u, v, 0)e^{-z_0 \Omega} + N(u, v) \quad (8.7)$$

$$\hat{F}(u, v, 0) = H(u, v)F_m(u, v, z_0) \quad (8.8)$$

Vogliamo

$$E[(H(u, v)F_m(u, v, z_0) - F(u, v, 0))F_m^*(u, v, z_0)] = 0 \quad (8.9)$$

## CAPITOLO 8 FILTRI 2D DI NATURA FISICA

e cioè incorrelazione tra errore di stima e dati.

$$H(u, v) = \frac{E[F_0 F_m^*]}{E[|F_m|^2] + \lambda_N^2}; \quad \lambda_N^2 = 4\Delta^2 \sigma_N^2 \quad (8.10)$$

dove  $\Delta$  è l'intervallo di campionamento spaziale, e  $\lambda_N^2$ ,  $\sigma_N^2$  sono rispettivamente la densità spettrale monolatera del rumore ed il suo valore quadratico medio.

$$E[|F_m|^2] = E[|F_0|^2]e^{-2z_0\Omega} + \lambda_N^2; \quad E[F_0 F_m^*] = E[|F_0|^2]e^{-z_0\Omega} \quad (8.11)$$

$$H(u, v) = \frac{e^{-z_0\Omega}}{e^{-2z_0\Omega} + Q}; \quad Q = \frac{\lambda_N^2}{E[|F_0|^2]} \quad (8.12)$$

Ai numeri d'onda per cui  $Qe^{2z_0\Omega} \ll 1$  il filtro opererà come filtro inverso e quindi passa alto; invece se  $Qe^{2z_0\Omega} > 1$  il filtro si comporterà ancora come passa basso (figura 17.2).

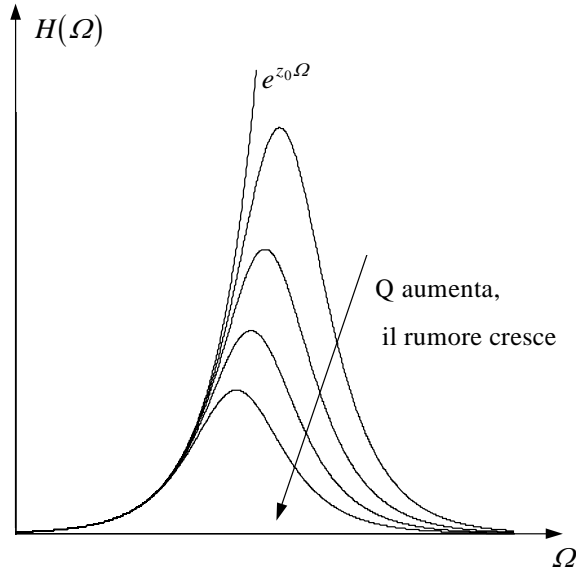


Figura 8.2: Filtri inversi per la continuazione in basso di potenziali

## 8.2 La propagazione di onde

In questo paragrafo discuteremo di un altro problema fisico, che condurrà allo studio di una funzione di trasferimento a fase pura. Il problema è quello della propagazione di onde in spazio libero: ad esempio le onde superficiali su un lago, oppure le onde acustiche nello spazio. L'equazione che governa la propagazione è

$$\frac{\partial^2 f(x, y, z)}{\partial x^2} + \frac{\partial^2 f(x, y, z)}{\partial y^2} + \frac{\partial^2 f(x, y, z)}{\partial z^2} = \frac{1}{c^2} \frac{\partial^2 f(x, y, z)}{\partial t^2} \quad (8.13)$$

dove  $c$  è la velocità di propagazione delle onde. Anzitutto, eliminiamo la variabile  $z$  supponendo che vi sia uniformità in quella direzione e pertanto

$$\frac{\partial^2 f(x, y, z)}{\partial z^2} = 0 \quad (8.14)$$

Si desidera ora, noto ad esempio  $f(x, y, t = 0); \forall x, y$  trovare  $f(x, y, t); \forall x, y$  e quindi l'evoluzione temporale di un campo d'onde, in assenza di sorgenti, a partire da delle condizioni iniziali. Un altro problema che risolveremo (al prossimo paragrafo) è quello del telerilevamento e cioè:

determinare  $f(x, y, t); \forall x, y, t$  dato:  $f(x, y = 0, t); \forall x, t$

Come già fatto nel caso precedente, trasformiamo l'equazione nelle variabili note su tutto il dominio, da  $-\infty \div \infty$ . Poichè, nel primo caso dello studio dell'evoluzione temporale di un campo d'onde, la funzione è nota per tutti i valori della  $x$  e della  $y$ , possiamo trasformare in Fourier per quelle variabili e risulta l'equazione differenziale semplice:

$$-(u^2 + v^2)F(u, v, t) + \frac{1}{c^2} \frac{d^2 F(u, v, t)}{dt^2} = 0; \quad F(u, v, t) \leftrightarrow f(x, y, t) \quad (8.15)$$

$$F(u, v, t) = A(u, v)e^{jct\Omega} + B(u, v)e^{-jct\Omega} \quad (8.16)$$

La soluzione è somma di due termini che corrispondono rispettivamente alle onde che esplodono ed implodono. In particolare, la risposta all'impulso corrispondente alle onde che esplodono e cioè si allontanano dalla sorgente sarà:

$$F_0(u, v, t) = e^{jct\Omega} \quad (8.17)$$

## CAPITOLO 8 FILTRI 2D DI NATURA FISICA

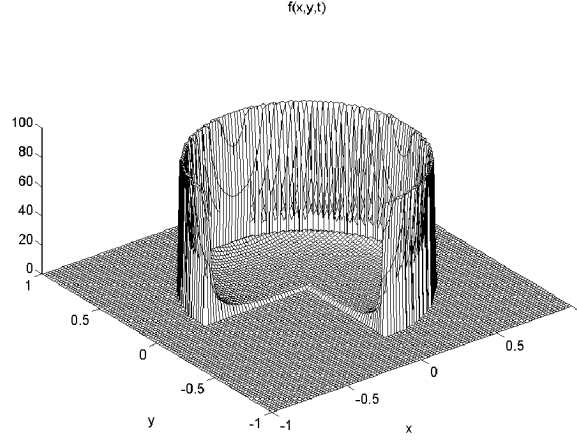


Figura 8.3: Sezione di un'onda circolare bidimensionale (p.e. un *cerchio sull'acqua*)

la cui antitrasformata è (figura 17.3):

$$f(x, y, t) = \frac{ct}{\sqrt{c^2t^2 - y^2 - x^2}}; \quad x^2 + y^2 \leq c^2t^2 \quad (8.18)$$

Questa risposta all'impulso corrisponde p.e. a dei *cerchi nell'acqua*  $t$  secondi dopo avere lanciato un sasso in uno stagno; piú precisamente, si può pensare ad un tuono, dopo un lampo (una sorgente rettilinea di perturbazione, di lunghezza praticamente illimitata in una direzione, che indichiamo con  $z$ ). La perturbazione generata è contenuta in un cerchio di raggio  $ct$  nel caso del sasso nello stagno o in un cilindro di raggio  $ct$  diretto come  $z$  nel caso del lampo. Si osservi poi che vi è differenza nella risposta ad un'eccitazione impulsiva nei casi 2D e 3D. Infatti, nel caso 3D la propagazione dell'onda generata da un impulso puntiforme avviene lungo un fronte sferico, e la perturbazione non permane, dopo che sia passato il primo fronte d'arrivo. Si pensi ora al caso 2D, nel caso della sorgente di perturbazione rettilinea; il piano di osservazione del fenomeno è ottenuto sezionando lo spazio 3D ortogonalmente all'asse  $z$ , direzione della linea delle sorgenti. Disponendo un ricevitore in questo piano, e applicando la sovrapposizione degli effetti, si comprende che questo ricevitore misurerà anzitutto l'arrivo delle perturbazione dovuta

## 8.2 LA PROPAGAZIONE DI ONDE

alla zona di sorgente situata nel piano. Successivamente, il ricevitore misurerà anche gli arrivi generati dalle zone della sorgente lineare situate al di fuori del piano che contiene il ricevitore (si pensi p.e. al *brontolio del tuono*). Quindi la forma del segnale che corrisponde al transitorio del passaggio dell'onda impulsiva in un sistema 2D **non è** una muraglia cilindrica, anche se il segnale si concentra intorno ad essa.

### 8.2.1 Il problema del telerilevamento

Il problema che ora abbiamo è il seguente: supponiamo p.e. di assistere alla rappresentazione di un'opera lirica; possiamo disporre dei microfoni lungo il bordo del palcoscenico (l'asse  $x, y = 0$ ) e vogliamo combinare le loro uscite in modo da potere isolare la voce di qualsiasi cantante (le sorgenti  $s_1, s_2$ ), come se disponessimo di un microfono in un punto qualsiasi della scena ( $x, y \neq 0$ ) (fig. 17.4). Questo problema corrisponde al rilevamento e alla focalizzazione da lontano del segnale generato in un punto qualsiasi dello spazio. Supponendo che il campionamento spaziale (lungo il palcoscenico) e temporale sia fittissimo, misuriamo la funzione  $f(x, y = 0, t); \forall x, t$ ; cerchiamo  $f(x, y, t); \forall x, y, t$ . L'equazione che utilizzeremo è ancora quella della propagazione:

$$-(u^2 - \frac{\omega^2}{c^2})F(u, y, \omega) + \frac{d^2 F(u, y, \omega)}{dy^2} = 0; \quad (8.19)$$

ma ora abbiamo trasformato le variabili  $x, t$ ; le variabili coniugate sono  $u, \omega$ . Risolviamo l'equazione differenziale e troviamo il risultato desiderato e cioè che:

$$F(u, y, \omega) = A(u, \omega)e^{jy\sqrt{\frac{\omega^2}{c^2} - u^2}} + B(u, \omega)e^{-jy\sqrt{\frac{\omega^2}{c^2} - u^2}} \quad (8.20)$$

Di nuovo, le due soluzioni corrispondono ad onde che si avvicinano o si allontanano dall'asse  $y = 0$  (il proscenio): per il problema che vogliamo risolvere, scegliamo le onde che si avvicinano e quindi anticipi per  $y$  crescenti.

$$F(u, y, \omega) = F(u, y = 0, \omega)e^{j\frac{\omega y}{c}\sqrt{1 - \frac{u^2 c^2}{\omega^2}}} \quad (8.21)$$

Trovata  $F(u, y, \omega)$ , la funzione cercata  $f(x, y, t)$  sarà ottenuta con una antitrasformazione di Fourier 2D.

$$f(x, y, t) = f(x, y = 0, t) ** FT^{-1} \left( e^{j\frac{\omega y}{c}\sqrt{1 - \frac{u^2 c^2}{\omega^2}}} \right)$$



## CAPITOLO 8 FILTRI 2D DI NATURA FISICA

Basta cioè convolvere la funzione  $f(x, y, t)$  con un filtro opportuno, dipendente dalla variabile  $y$ .

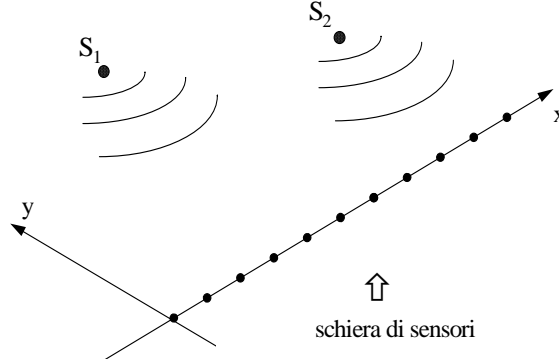


Figura 8.4: Disposizione di microfoni su un palcoscenico per isolare i suoni emessi dai singoli cantanti

### 8.2.2 Le onde piane

Il filtro dell'equazione precedente ha un'espressione complicata, che tuttavia può essere spiegata introducendo il concetto delle onde piane, già implicitamente utilizzato nel capitolo sulle Direzioni di Arrivo (DOA). Nel dominio  $x, t$  le sinusoidi bidimensionali sono segnali del tipo:

$$s(x, t) = Ae^{j(ux+\omega t)} \quad (8.22)$$

Come già visto nel capitolo 12 questi segnali corrispondono ad onde monocromatiche di ampiezza complessa  $A$  e di pulsazione  $\omega$ , i cui fronti d'onda sono delle rette nel piano  $x, y$  che formano un angolo  $\theta$  con l'asse delle ascisse  $x$ . Infatti, due punti a distanza  $2\pi/u$  lungo l'asse delle  $x$ , oscillano in fase e pertanto sono separati da un tempo di propagazione  $\tau = 2\pi/\omega$ ; si ha allora:

$$\begin{aligned} \frac{2\pi}{u} \sin \theta &= c \frac{2\pi}{\omega} = c\tau; \\ \sin \theta &= \frac{uc}{\omega}; \quad \cos \theta = \sqrt{1 - \frac{u^2 c^2}{\omega^2}} \end{aligned} \quad (8.23)$$

## 8.2 LA PROPAGAZIONE DI ONDE

Si osservi che se:

$$|u| > \left| \frac{\omega}{c} \right| \quad (8.24)$$

l'angolo  $\theta$  diventa una variabile immaginaria; le onde non si propagano più, ma diventano evanescenti. Per capire questo fenomeno, basti pensare a più sorgenti puntiformi monocromatiche (a frequenza  $\omega$ ) disposte lungo l'asse  $x$  a distanza  $\delta$  e irradianti indipendentemente; inoltre, la fase dell' $n$ -esima sorgente è ruotata di  $e^{j\omega n\delta}$ . Se la disuguaglianza (8.24) è rispettata, il campo emesso non è rappresentabile come un'onda piana che fa un angolo  $\theta$  con l'asse  $x$ ; le varie sorgenti interferiscono tra loro in modo distruttivo e il campo si riduce progressivamente (in modo esponenziale) via via che ci si allontana dall'asse  $x$ . Nel caso semplice in cui  $\omega = 0$ , siamo di nuovo al caso dell'elettrostatica o della gravità. Infatti nell'equazione 8.13 la derivata rispetto al tempo è nulla, non c'è propagazione, e siamo di nuovo all'equazione armonica.

Ritornando al problema del paragrafo precedente, per ricostruire il campo d'onda  $f(x, y, t)$  in posizioni  $y \neq 0$  a partire da  $f(x, y = 0, t)$  si può farne anzitutto una decomposizione in onde piane. Questa è un'interpretazione della trasformazione di Fourier della  $f(x, y = 0, t)$ ; le ampiezze delle onde piane a pulsazione  $u, \omega$  sono appunto i valori  $F(u, y = 0, \omega)$ . Ogni onda piana si propaga in avanti o all'indietro nello spazio senza variazioni di ampiezza, ma subendo solo uno sfasamento. Essa quindi può venire retropropagata nello spazio (e/o nel tempo) fino alla posizione della sorgente applicando una rotazione di fase che dipende dalla pulsazione  $\omega$  e dalla distanza  $y$  di cui deve essere retropropagata. La rotazione è :

$$\omega\tau = \frac{\omega y \cos \theta}{c} \quad (8.25)$$

Quindi per ricostruire il campo all'ascissa  $y$  basta retropropagare le onde sfasandole del fattore:

$$e^{j\omega\tau} = e^{j\frac{\omega y}{c} \cos \theta} = e^{j\frac{\omega y}{c} \sqrt{1 - \frac{u^2 c^2}{\omega^2}}} \quad (8.26)$$

come risulta anche dalla figura 17.5; la combinazione di tutte le onde piane dopo la loro retropropagazione ci dá la distribuzione delle sorgenti nello spazio e pertanto l'immagine cercata.

L'antitrasformata della funzione

$$e^{-jy\sqrt{\frac{\omega^2}{c^2} - u^2}} \quad (8.27)$$

## CAPITOLO 8 FILTRI 2D DI NATURA FISICA

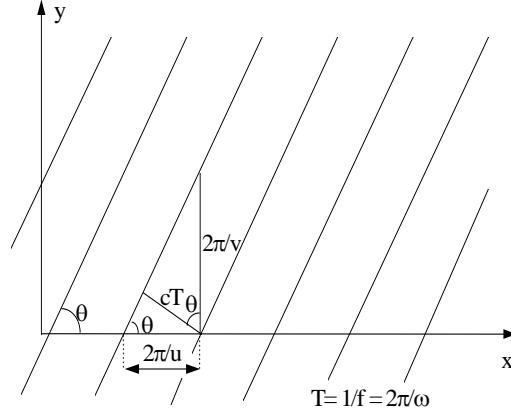


Figura 8.5: Retropropagazione di onde piane

ha la forma presentata in figura 17.6; essa corrisponde a quanto viene osservato lungo la retta  $y = 0$  in corrispondenza di un impulso (una sorgente esplosiva) al tempo ed alla posizione  $x = 0, t = 0, y$ ; il primo arrivo dell'onda all'ascissa  $x$  si ha al tempo:

$$t(x) = \frac{\sqrt{x^2 + y^2}}{c} \quad (8.28)$$

Quindi, nel piano  $x, t$  questi arrivi sono disposti lungo un'iperbole di equazione:

$$x^2 - c^2 t^2 = y^2 = c^2 t_0^2 \quad (8.29)$$

Disponendo ora della funzione  $f(x, y = 0, t)$  possiamo ora determinare la funzione  $f(x, y, t = 0)$  e cioè intuitivamente, *determinare dalle onde che arrivano alla spiaggia la distribuzione delle sorgenti in alto mare* oppure *potere appendere un microfono al collo di Pavarotti, malgrado che tutti i microfoni siano sul proscenio*. Per fare questo partiamo dall'equazione (8.21)

$$F(u, y = 0, \omega) = \int \int f(x, y = 0, t) e^{-j(ux + \omega t)} dt dx$$

$$f(x, y, t) = \int \int F(u, y = 0, \omega) e^{jy\sqrt{\frac{\omega^2}{c^2} - u^2} + jux + j\omega t} \frac{du d\omega}{(2\pi)^2}$$

## 8.2 LA PROPAGAZIONE DI ONDE

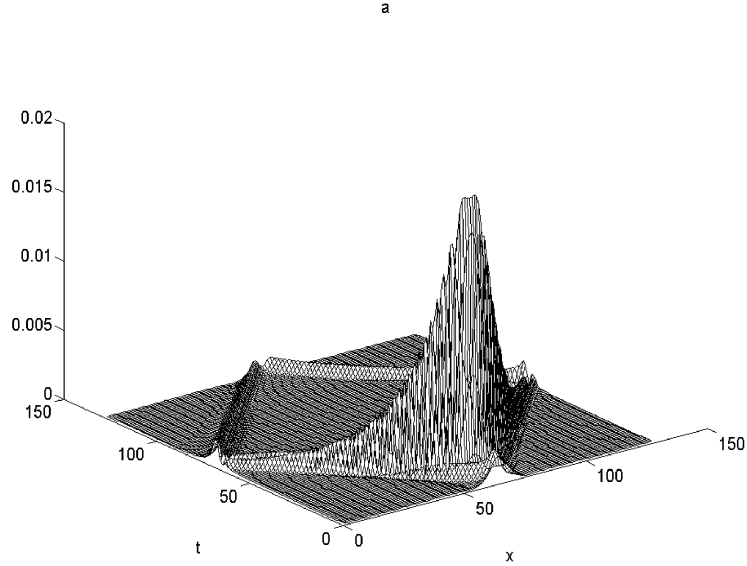


Figura 8.6: L'iperbole corrispondente ai tempi di arrivo di un segnale impulsivo ai vari microfoni sul palcoscenico; sono visibili effetti di periodicizzazione dell'asse delle ascisse dovute all'uso di DFT

(Il microfono al collo di Pavarotti). Ancora si ha:

$$f(x, y, t = 0) = \int \int F(u, y = 0, \omega) e^{jy\sqrt{\frac{\omega^2}{c^2} - u^2} + jux} \frac{dud\omega}{(2\pi)^2} \quad (8.30)$$

(Questo corrisponderebbe alla distribuzione delle posizioni, al largo, dove sono stati lanciati sassi al tempo 0, determinata sulla base delle onde che arrivano sulla spiaggia). Per ottenere, invece dell'espressione complessa (8.30), una semplice antitrasformata di Fourier 2D, basta effettuare la seguente trasformazione di coordinate:

$$\sqrt{\frac{\omega^2}{c^2} - u^2} = v \rightarrow \omega = c\sqrt{u^2 + v^2}; \quad d\omega = \frac{2cv dv}{\sqrt{u^2 + v^2}}$$

$$f(x, y, t = 0) = \int \int F(u, y = 0, \omega = c\sqrt{u^2 + v^2}) e^{j(vy + ux)} \frac{2cv}{\sqrt{u^2 + v^2}} \frac{dudv}{(2\pi)^2}$$

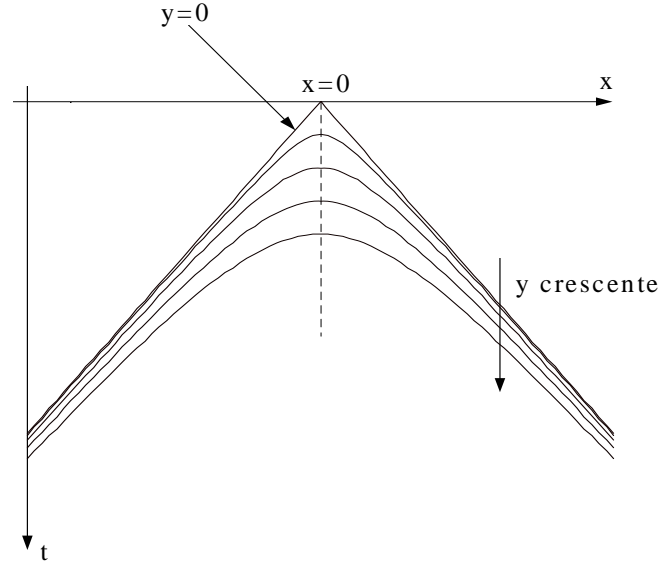


Figura 8.7: Iperboli dei tempi di arrivo ottenute con il principio della fase stazionaria

In altri termini, data la  $F(u, y = 0, \omega)$ , la si interpola nei punti  $u_i, \omega_l$  corrispondenti ai valori di  $u_i, v_l$  richiesti dal campionamento effettuato in queste due coordinate, e la si antitrasforma (Interpolazione di Stolt).

### 8.3 Il principio della fase stazionaria

Se si hanno delle complicate antitrasformazioni da compiere, è utile il principio della fase stazionaria. Consideriamo un integrale di Fourier del tipo:

$$Q(x) = \int_{-\infty}^{\infty} A(x, \xi) e^{jMG(x, \xi)} d\xi \quad (8.31)$$

dove  $M$  è un numero grande; se la fase dell'esponenziale scorre rapidamente l'integrale (una media) avrà valore assai basso; gli unici valori della variabile  $x$  per i quali l'integrale assume valori significativi si hanno quando **la fase è stazionaria** e cioè quando:

$$\frac{\partial G(x, \xi)}{\partial \xi} = 0 \quad (8.32)$$

## 8.4 IL *CHIRP*, LA LENTE SOTTILE E L'OTTICA DI FOURIER

Ovviamente,  $\xi$  può anche essere una variabile multideimensionale. Ad esempio, nel caso dell'antitrasformata di Fourier:

$$\frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-j\frac{\omega y}{c} \sqrt{1 - \frac{u^2 c^2}{\omega^2}} + ux + \omega t} du d\omega \quad (8.33)$$

risulta:

$$G = -t_0 \omega \sqrt{1 - \frac{u^2 c^2}{\omega^2}} + ux + \omega t \quad (8.34)$$

$$\frac{\partial G(x, t; u, \omega)}{\partial \omega} = -\frac{\omega t_0}{\sqrt{\omega^2 - u^2 c^2}} + t = 0 \quad (8.35)$$

$$\frac{\partial G(x, t; u, \omega)}{\partial u} = \frac{t_0}{\sqrt{\omega^2 - u^2 c^2}} uc^2 + x = 0 \quad (8.36)$$

Da cui, eliminando  $\omega, u$ :

$$\left(\frac{x}{c}\right)^2 - t^2 = t_0^2 \quad (8.37)$$

si trova appunto la (8.29). In sostanza, l'antitrasformata è nulla ovunque, a parte dove la 8.37 è verificata, e cioè lungo l'iperbole. Il principio della fase stazionaria si applica se  $\omega t_0 \gg 1$  e quindi se i raggi percorrono un numero elevato di lunghezze d'onda e pertanto non siamo in condizione di "campo vicino".

## 8.4 Il *chirp*, la lente sottile e l'ottica di Fourier

Un altro caso di interesse è quello del *chirp* bidimensionale. Il nome *chirp* viene dal cinguettio di un uccello in cui la frequenza del suono emesso cresce progressivamente. Il chirp nel tempo è quindi corrispondente ad un segnale caratterizzato da una frequenza crescente linearmente con il tempo e quindi da una fase crescente quadraticamente con il tempo. La sua trasformata è ancora un chirp; si ricorda infatti la coppia di trasformate (vedi al capitolo 9):

$$c(t) = \exp\left(-j\frac{\pi t^2}{\tau^2}\right) \leftrightarrow C(f) = \tau \exp\left(-j\pi\tau^2 f^2 - j\frac{\pi}{4}\right)$$

## CAPITOLO 8 FILTRI 2D DI NATURA FISICA

Il *chirp* spaziale bidimensionale corrisponde ad una frequenza spaziale crescente linearmente con lo spazio e quindi ad una fase crescente quadraticamente con la distanza dall'origine.

$$h(\rho) = \frac{\exp(j(x^2 + y^2)/4\sigma^2)}{2\pi\sigma^2} = \frac{\exp(jx^2/4\sigma^2)}{\sqrt{2\pi}\sigma} \cdot \frac{\exp(jy^2/4\sigma^2)}{\sqrt{2\pi}\sigma} \quad (8.38)$$

Si verifica subito che anche la funzione di trasferimento  $H(\Omega)$  è un *chirp* bidimensionale. Infatti, la trasformata è immediatamente trovata, visto che le variabili sono in questo caso separabili, e si ritrova un *chirp* questa volta nello frequenza spaziale. Posto al solito  $\Omega = \sqrt{u^2 + v^2}$ , a parte una costante complessa:

$$H(\Omega) = \exp(-j\sigma^2 u^2) \cdot \exp(-j\sigma^2 v^2) = \exp(-j\sigma^2 \Omega^2) \quad (8.39)$$

L'interesse per il chirp nasce dal fatto che la convoluzione con un *chirp* approssima molto bene una trasformazione di Fourier. Infatti, operando ad 1D per semplicità, si ha che il risultato della convoluzione con il *chirp* di una funzione  $f(t)$  che ha per trasformata  $F(\omega)$  è la funzione  $g(t)$  (che ha trasformata  $G(\omega)$ ):

$$g(t) = \int F(\omega) e^{j(-\omega^2 \sigma^2 + \omega t)} \frac{d\omega}{2\pi}$$

Calcoliamo  $g(t)$  con la tecnica della fase stazionaria. Poichè si ha stazionarietà della fase quando:

$$-2\omega_0 \sigma^2 + t_0 = 0 \rightarrow t_0 = 2\omega_0 \sigma^2$$

si trova, a meno di un fattore di proporzionalità  $\alpha$ ,

$$g(t_0) \simeq \alpha F\left(\omega_0 = \frac{t_0}{2\sigma^2}\right) \quad (8.40)$$

In altre parole, la funzione  $g(t)$  è circa proporzionale alla trasformata di Fourier della  $f(t)$ .

Per una comprensione intuitiva di questo meccanismo, si pensi al chirp come alla successiva trascrizione in un'unica forma d'onda complessiva delle  $N$  sinusoidi finestrate di una trasformazione di Fourier di sequenze lunghe  $N$  campioni, cominciando da quella a frequenza di Nyquist

#### 8.4 IL *CHIRP*, LA LENTE SOTTILE E L'OTTICA DI FOURIER

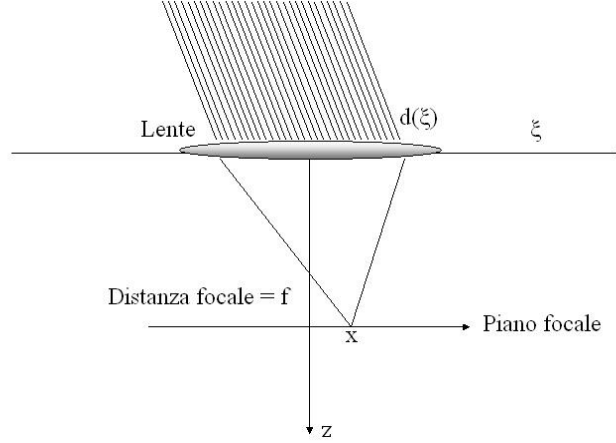


Figura 8.8: Lente sottile: nel piano focale si forma la trasformata di Fourier bidimensionale della funzione di illuminazione  $d(\xi)$ .

negativa, passando per quella a frequenza zero, per finire a quella a frequenza di Nyquist positiva. La lunghezza totale sarebbe di  $N^2$  campioni. Potremo individuare la frequenza di una nuova sinusoide convolvendola con questo chirp lungo e poi trovando il tempo a cui la convoluzione è massima. Decimando il risultato della convoluzione di un fattore  $N$ , e prendendone il modulo, si otterrà circa il periodogramma, ma trascritto in una sequenza di valori. La risoluzione in frequenza crescerà con  $N$ . Ovviamente, nel caso del chirp, la frequenza cambia con continuità invece che a gradini, come nella trascrizione del periodogramma.

Vediamo ora come il caso del chirp si applica alla focalizzazione con delle lenti sottili ed alla formazione delle immagini nel loro piano focale (vedi figura 8.8). Indichiamo con  $d(\xi)$  la funzione di illuminazione e cioè l'ampiezza complessa dell'onda elettromagnetica luminosa sull'apertura della lente. L'effetto della lente è quello di sfasare il segnale in arrivo di una fase pari a:

$$\exp j \frac{\omega}{c} \frac{\xi^2}{2f} ;$$

dove  $\omega$  è al solito la pulsazione dell'onda,  $c$  la velocità della luce,  $f$  la distanza focale della lente. Per chi non ricordasse l'ottica geometrica,



## CAPITOLO 8 FILTRI 2D DI NATURA FISICA

notiamo che:

$$\frac{1}{f} = \left( \frac{c - c_1}{c_1} \right) \left( \frac{1}{R_1} - \frac{1}{R_2} \right)$$

dove  $R_1$ ,  $R_2$ , sono i raggi di curvatura (con i segni relativi) delle due facce della lente e  $c_1$  la velocità della luce nella lente (piu' bassa). Il segnale irraggiato dalla lente, misurato in un piano parallelo alla lente e distante  $z$  da essa è, per sovrapposizione degli effetti:

$$m(x) = \int_{\text{apertura}} d(\xi) \exp \frac{jk\xi^2}{2f} \exp \left[ -jk \left( z + \frac{(x - \xi)^2}{2z} \right) \right] d\xi \quad ; \quad k = \frac{\omega}{c} \quad (8.41)$$

Lo sfasamento additivo:

$$\exp \left[ -jk \left( z + \frac{(x - \xi)^2}{2z} \right) \right]$$

tiene conto (con un'approssimazione del secondo ordine) dello sfasamento dovuto alla propagazione dal punto di ascissa  $\xi$  sulla lente a quello di ascissa  $x$  a distanza  $z$  dalla lente. Quindi:

$$m(x) = e^{-j\frac{kx^2}{z} - jkz} \int_{\text{apertura}} d(\xi) \exp jk \left[ \frac{\xi^2}{2} \left( \frac{1}{f} - \frac{1}{z} \right) + \frac{x\xi}{z} \right] d\xi \quad (8.42)$$

e se  $z = f$  :

$$m(x) = e^{-j\frac{kx^2}{f} - jkf} \int_{\text{apertura}} d(\xi) \exp jk \left( \frac{x\xi}{f} \right) d\xi \quad (8.43)$$

Nel piano focale della lente quindi, troviamo la trasformata di Fourier della funzione di illuminazione a meno del fattore di fase  $\exp \left( -j\frac{kx^2}{f} - jkf \right)$ . Se poi l'oggetto (la funzione di illuminazione) è posizionato nell'altro fuoco della lente, questo fattore di fase si cancella. I due piani focali sono gli unici tra cui la trasformazione di Fourier avviene senza sfocamento. Ancora, ad ogni direzione di arrivo bidimensionale sulla lente (ogni pulsazione spaziale delle onde piane in cui si decompone la funzione di illuminazione sulla lente) corrisponde un preciso punto del piano focale (vedi figura 8.8) dove l'energia proveniente da questa direzione si concentra. L'immagine che si forma nel piano focale corrisponde quindi

#### 8.4 IL *CHIRP*, LA LENTE SOTTILE E L'OTTICA DI FOURIER

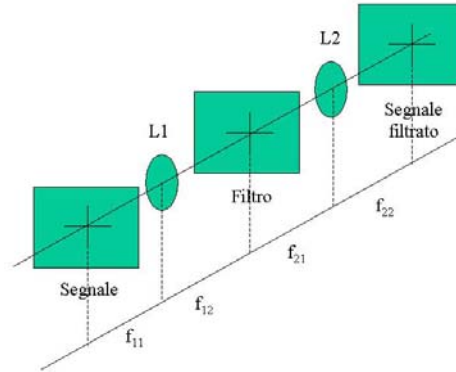


Figura 8.9: Tecniche di filtraggio ottico: il segnale è depositato sulla trasparenza nel primo fuoco della lente L1 ed il filtro su un'altra trasparenza nel secondo fuoco della lente L1 che coincide con il primo fuoco della lente L2. Il segnale filtrato appare nell'immagine nel secondo fuoco della lente L2.

appunto alla distribuzione dell'energia che proviene dalle varie direzioni. Se si osserva solo questa energia, come nel caso dell'occhio o di un telescopio, l'errore di fase di cui nell'equazione 8.43 è irrilevante. In ogni modo, la tecnica di analisi spettrale che viene implicitamente utilizzata nella visione è quindi quella del periodogramma, come peraltro già osservato in occasione della discussione delle DOA.

La disponibilità di una trasformata di Fourier 2D (o anche 1D, basta mettere delle lenti cilindriche invece che sferiche) in tempo reale ha portato a realizzare filtri spaziali con tecniche ottiche; basta mettere l'immagine da trasformare nel primo piano focale di una lente L1, la sagoma del filtro nell'altro piano focale della lente L1 che poi coincide con il primo piano focale di una seconda lente L2 (fig. 8.9). Il segnale filtrato è osservabile nel secondo piano focale della lente L2. I primi sistemi radar ad apertura sintetica utilizzavano proprio questo metodo per effettuare una focalizzazione sia pure approssimativa, ma in tempo reale, del segnale registrato. Questa tecnica appare oggi obsoleta, vista l'espansione delle metodologie numeriche di trattamento, ma potrebbe rinascere in qualsiasi momento, visto il continuo evolversi delle tecnologie e la grande potenza di calcolo che potrebbe garantire.

## 8.5 L'equazione della diffusione

Il quarto caso che prenderemo in considerazione è quello dell'equazione della diffusione, ad esempio termica; essa governa l'evoluzione temporale della temperatura  $T(x, y, t)$  di una piastra indefinita che ha conducibilità termica uniforme  $c$ .

$$\frac{\partial^2 T(x, y, t)}{\partial x^2} + \frac{\partial^2 T(x, y, t)}{\partial y^2} = c \frac{\partial T(x, y, t)}{\partial t} \quad (8.44)$$

Supponiamo anzitutto di conoscere la temperatura della piastra al tempo  $t = 0$ ;  $\forall x, y$  e di volerne determinare l'evoluzione. Trasformiamo nelle variabili  $x, y$  ponendo, al solito  $\Omega = \sqrt{u^2 + v^2}$ :

$$-\Omega^2 T(\Omega, t) = c \frac{dT(\Omega, t)}{dt} \quad (8.45)$$

$$T(u, v, t) = T(u, v, 0) \times e^{-\frac{\Omega^2 t}{c}} \quad (8.46)$$

Si osservi che il valore della funzione di trasferimento  $e^{-\frac{\Omega^2 t}{c}}$  per  $\Omega = 0$ , è unitario ed indipendente dal tempo. Passando alle variabili  $x, y$ :

$$T(x, y, t) = T(x, y, 0) * * \frac{c}{4\pi t} e^{-\frac{\rho^2 c}{4t}} \quad (8.47)$$

La temperatura iniziale è quindi convoluta per un filtro gaussiano di volume costante (la quantità di calore posseduta inizialmente dalla piastra rimane invariata) e invece *diametro* crescente con  $\sqrt{t}$ : è l'effetto della diffusione termica.

Un secondo problema è quello dell'osservazione della temperatura lungo la linea  $y = 0, T(x, y = 0, t), \forall t$  e poi la determinazione della  $T(x, y, t); y > 0$ . Trasformando questa volta in  $x, t$  si ha:

$$(u^2 + j\omega c) \times T(u, y, \omega) = \frac{d^2 T(u, y, \omega)}{dy^2} \quad (8.48)$$

Risolvendo l'equazione differenziale ed eliminando la soluzione divergente per  $y \rightarrow \infty$  abbiamo:

$$T(u, y, \omega) = T(u, y = 0, \omega) \times e^{-y\sqrt{u^2 + j\omega c}} \quad (8.49)$$

## 8.5 L'EQUAZIONE DELLA DIFFUSIONE

Nei tempi - spazi ciò corrisponde a:

$$T(x, y, t) = T(x, y = 0, t) * \frac{y}{2\sqrt{\pi}(t/c)^{3/2}} \times e^{-\frac{y^2}{4t}} \quad (8.50)$$

Si osservi che il flusso di calore dalla superficie verso l'interno del corpo non è nullo. E' interessante anche osservare come l'ultima equazione affermi che la temperatura possa variare per valori anche elevati di  $x, y$  anche in tempi molto piccoli. Ciò è dovuto all'imperfezione della descrizione fisica del processo fatta con l'equazione (8.44) che trascura il tempo finito di propagazione delle variazioni termiche.



## Capitolo 9

### Sequenze 2D

Un segnale bidimensionale può essere campionato con le stesse regole del campionamento a 1D; in particolare, campionare implica replicarne lo spettro. Infatti la trasformata del reticolo bidimensionale di impulsi è un reticolo bidimensionale di impulsi nel dominio delle frequenze spaziali (fig.18.1).

$$\sum_{h,k=-\infty}^{+\infty} \delta(x - hx_0, y - ky_0) = \sum_{h=-\infty}^{+\infty} \delta(x - hx_0) \sum_{k=-\infty}^{+\infty} \delta(y - ky_0) \leftrightarrow \quad (9.1)$$

$$\sum_{l=-\infty}^{+\infty} \delta(u - lu_0) \sum_{m=-\infty}^{+\infty} \delta(v - mv_0) = \sum_{l,m=-\infty}^{+\infty} \delta(u - lu_0, v - mv_0) \quad (9.2)$$

$$u_0 = \frac{2\pi}{x_0}; \quad v_0 = \frac{2\pi}{y_0}; \quad (9.3)$$

Moltiplicare negli spazi per la funzione impulsiva ripetuta implica convolvere nelle frequenze per la sua trasformata e quindi lo spettro di una sequenza a due dimensioni è periodico sia lungo l'asse delle  $u$  che delle  $v$ . Per potere campionare è pertanto sufficiente che il segnale sia contenuto nel rettangolo:

$$|u| < \frac{|u_0|}{2}; \quad |v| < \frac{|v_0|}{2} \quad (9.4)$$

In effetti basta che le repliche non si sovrappongano; pertanto non è necessario che lo spettro stia in un rettangolo; basta che stia in una qualsiasi figura che replicata su un reticolo rettangolare non si sovrapponga. Ad

## CAPITOLO 9 SEQUENZE 2D

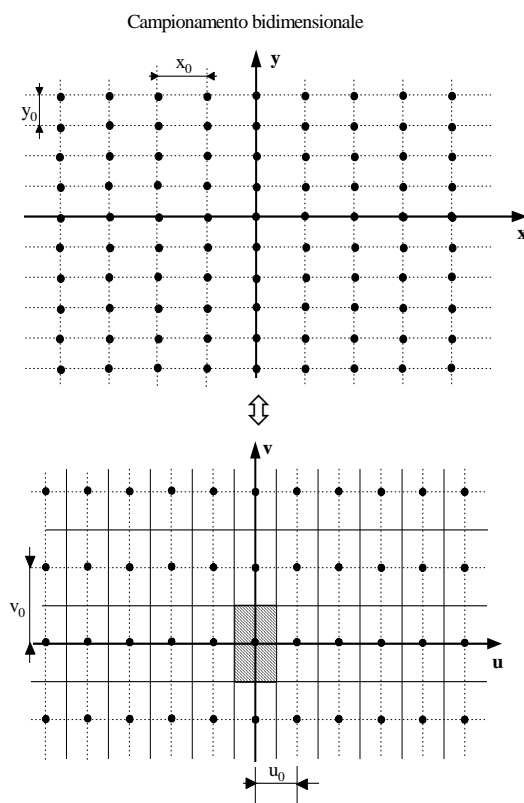


Figura 9.1: Campionamento bidimensionale

esempio possiamo levare da un lato del rettangolo un'area per aggiungerla al lato opposto ecc (vedi figura 18.2.) In sostanza, a 2D si hanno maggiori libertà.

### 9.1 L'equivocazione a 2D

Se lo spettro bidimensionale dell'immagine che viene campionata non è contenuto nel *dominio fondamentale*, le sue repliche rientrano e formano equivocazione. Nella figura 18.3 vediamo lo spettro relativo a una funzione cilindrica campionata (si pensi a dei filari di alberi); le repliche spettrali che escono ad esempio dal bordo destro della mattonella, rientrano dal bordo sinistro, formando una configurazione complessa. In sostan-

## 9.1 L'EQUIVOCAZIONE A 2D

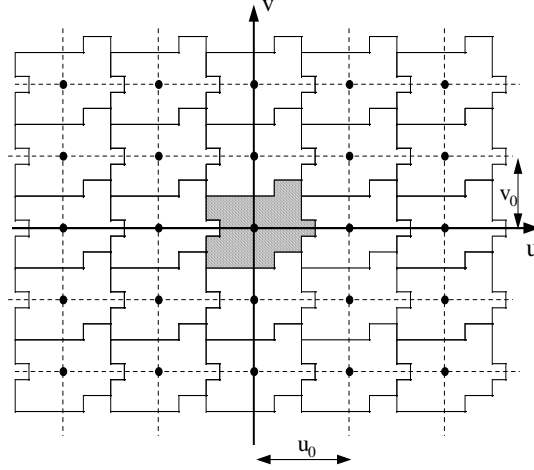


Figura 9.2: Un altro possibile dominio fondamentale per la griglia di campionamento

za, vediamo i filari degli alberi allinearsi sia secondo le direzioni originali che secondo altre direzioni, equivocate, lungo cui appaiono allineamenti fittizi.

### 9.1.1 Effetti di moiré

Il campionamento 2D di segnali di tipo sinusoidale origina effetti di equivocazione spaziale abbastanza complessi, cui viene dato in generale il nome di *moiré*, dal tessuto omonimo che ha un aspetto *marezzato*. Consideriamo una sinusoide 2D, complessa:

$$s(x, y) = A \exp(ju_0x + v_0y)$$

La sinusoide campionata nei punti:

$$x_n = n\Delta; \quad y_m = m\Delta;$$

ha uno spettro replicato

$$S(u, v) = A \sum_{p, q=-\infty}^{\infty} \delta\left(u_0 + p\frac{2\pi}{\Delta}, v_0 + q\frac{2\pi}{\Delta}\right)$$



## CAPITOLO 9 SEQUENZE 2D

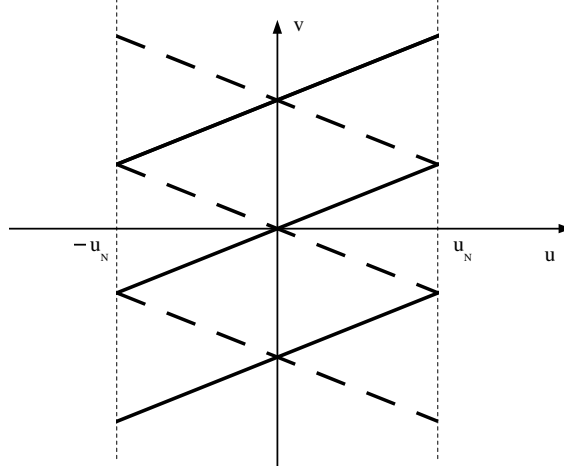


Figura 9.3: Lo spettro di una lama replicato per effetto del campionamento (orizzontale).

Il dominio spettrale fondamentale

$$|u, v| < \frac{\pi}{\Delta} \quad (9.5)$$

limitato dalla pulsazione di Nyquist, contiene solo una sinusoide 2D che coincide con la sinusoide non campionata quando il campionamento è corretto e cioè se:

$$|u_0, v_0| < \frac{\pi}{\Delta} \quad (9.6)$$

Altrimenti, la sinusoide equivocata è

$$u_{eq}, v_{eq} = u_0 + p \frac{2\pi}{\Delta}, v_0 + q \frac{2\pi}{\Delta}$$

dove  $p, q$  sono tali da garantire che la (9.6) sia soddisfatta. Si osservi in particolare che l'azimut della sinusoide 2D equivocata può essere totalmente differente da quello della sinusoide 2D originale.

Un esempio di *moiré* si ottiene dalla sovrapposizione di due reticoli leggermente diversi, che risulta p.e. se si guardano controluce due veli eguali e sovrapposti, poco distanti tra loro. Per effetto della prospettiva, la trama del velo più lontano ci appare più fitta di quella del velo più vicino di un fattore pari a  $D/(D+d)$ , dove  $d$  è la distanza tra i veli e

## 9.1 L'EQUIVOCAZIONE A 2D

$D$  la distanza che ci separa dal velo più vicino. L'immagine che vediamo in trasparenza è circa il prodotto dei due reticoli e pertanto lo spettro del segnale osservato è la convoluzione degli spettri. Quindi, se  $f_0$  è la frequenza corrispondente alla trama di un velo (supponiamo il processo monodimensionale, per semplicità), e  $f_0 D / (D + d)$  l'altra, il battimento (il prodotto) avrà una componente alla frequenza differenza:

$$\Delta f = f_0 - \frac{f_0 D}{D + d} \simeq \frac{d}{D} f_0 \quad \text{se } d \ll D$$

Se il reticolo è 2D, vedremo delle *frange interferenziali* da cui potremmo ricostruire dei dettagli della struttura microscopica dei reticoli di partenza; p.e. conoscendo  $d, D, \Delta f$  potremmo determinare  $f_0$  anche se ciò fosse impossibile ad "occhio nudo". Infine, si osservi che, allontanandosi dai veli, aumenta il periodo della frangia del *moiré*.

### 9.1.2 Il campionamento esagonale

Se il segnale da campionare ha spettro circolare, il reticolo di campionamento quadrato è inefficiente in quanto comporta un dominio fondamentale nelle frequenze spaziali anch'esso quadrato, sostanzialmente. Pertanto, sinusoidi 2D nelle direzioni delle diagonali del quadrato possono avere frequenza  $\sqrt{2}$  volte maggiore che non nelle direzioni dei lati del quadrato stesso. Quindi, a seconda che il quadrato sia iscritto o circoscritto alla circonferenza che delimita lo spettro, o si deve tagliare lo spettro o si deve usare un campionamento troppo fitto. E' utile cercare tecniche di campionamento più isotrope, e cioè aventi un dominio fondamentale più assomigliante ad un cerchio. Ciò si ottiene con il campionamento esagonale che verrà ora presentato. Si consideri un reticolo la cui base sono dei triangoli equilateri (fig. 18.4); questo reticolo ha 3 direzioni di simmetria a  $120^\circ$  e quindi è più simmetrico di quello rettangolare che ne ha due a  $90^\circ$ . Si intuisce come questo reticolo di campionamento possa essere più efficiente di quello quadrato, per funzioni a spettro circolare. Indichiamo con  $a$  il lato del triangolo; il reticolo a base triangolare può essere visto come la combinazione di due reticoli a base rettangolare di lati  $x_0 = a, y_0 = a\sqrt{3}$  rispettivamente; i due reticoli sono sfalsati di mezzo passo lungo le due direzioni. Nel dominio delle frequenze si ha che la

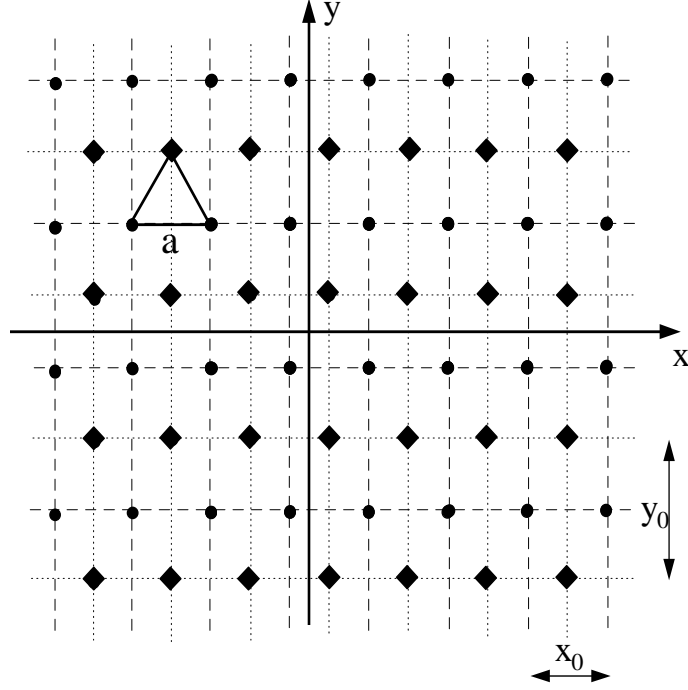


Figura 9.4: Reticolo corrispondente al campionamento esagonale (dominio degli spazi).

trasformata del doppio reticolo di impulsi è:

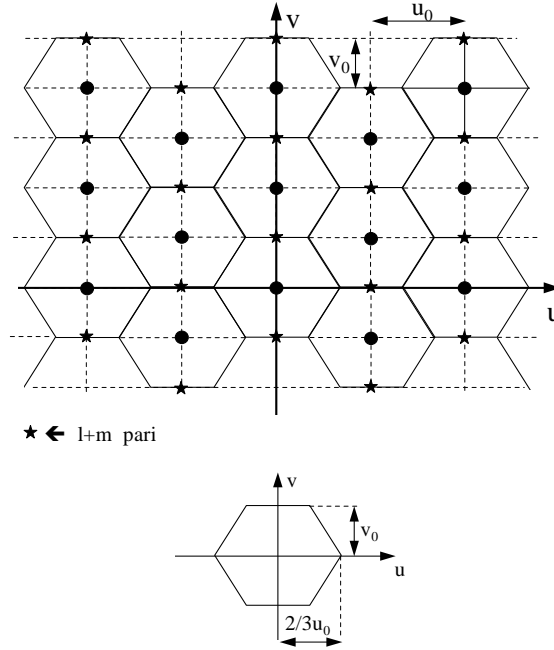
$$H(u, v) = [e^{j(\frac{ux_0}{4} + \frac{vy_0}{4})} + e^{-j(\frac{ux_0}{4} + \frac{vy_0}{4})}] \times \sum_{h,l=-\infty}^{+\infty} \delta(u - hu_0, v - lv_0); \quad (9.7)$$

$$u_0 = \frac{2\pi}{x_0} = \frac{2\pi}{a}; \quad v_0 = \frac{2\pi}{y_0} = \frac{2\pi}{a\sqrt{3}} = \frac{u_0}{\sqrt{3}} \quad (9.8)$$

$$H(u, v) = 2 \cos[(h+l)\frac{\pi}{2}] \times \sum_{h,l=-\infty}^{+\infty} \delta(u - hu_0, v - lv_0); \quad (9.9)$$

La combinazione dei due reticoli elimina le repliche per le quali  $l + h$  è dispari, e raddoppia l'ampiezza di quelle per cui  $l + h$  è pari. La sagoma del dominio base nelle frequenze è quindi un esagono regolare,

## 9.1 L'EQUIVOCAZIONE A 2D



Sagoma del dominio base nelle frequenze

Figura 9.5: Reticolo corrispondente al campionamento esagonale (dominio delle frequenze).

come in figura 18.5. Il rapporto tra la massima frequenza spaziale che non viene equivocata (nelle direzioni  $0^\circ, 60^\circ, 120^\circ$ ) e la minima (nelle direzioni  $30^\circ, 90^\circ, 150^\circ$ ) è

$$\frac{2}{3}u_0 : v_0 = \frac{2}{\sqrt{3}} = 1,154 \quad (9.10)$$

mentre invece per il reticolo quadrato questo rapporto è  $\sqrt{2}$ . L'area per ogni campione è  $l^2$  nel caso del reticolo quadrato di base  $l$ ; invece è di  $a^2\sqrt{3}/2$  nel caso di reticolo triangolare di base  $a$ ; la massima frequenza non equivocabile è  $1/2l$  in un caso e  $1/(a\sqrt{3})$  nell'altro caso. Eguagliando questi due valori, e quindi imponendo di non equivocare questa frequenza qualsiasi sia il suo azimut, e ponendo quindi  $a = 2l/\sqrt{3}$ , si trova che, passando da reticolo quadrato a reticolo esagonale, il numero di campioni necessario decresce del fattore  $2/\sqrt{3} = 1,154$ .

## 9.2 Il campionamento del segnale televisivo

Nella trasmissione analogica di un segnale televisivo, il campionamento del volume  $(x, y, t)$  che corrisponde alla successione di infinite immagini a risoluzione spaziale e temporale illimitata avviene nel modo piuttosto complesso che verrà ora descritto.

Le immagini vengono scandite per righe su campi interallacciati. Il segnale viene cioè campionato verticalmente a distanza  $d$ ; nel sistema analogico PAL, un quadro è formato da 625 linee di cui solo 573 sono visibili. Le righe dispari, distanti perciò tra di loro  $2d$ , sono trasmesse in successione (semiquadro dispari o campo dispari); poi, con un ritardo di 20 msec, viene trasmesso il campo pari e quindi le linee di posizione pari. Dopo ancora 20 msec, si trasmette il campo dispari del quadro 2 e così via al ritmo di 25 quadri al secondo (fig.18.6) Il campionamento del

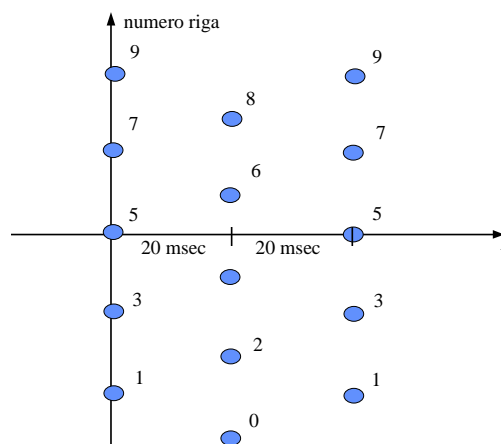


Figura 9.6: Campionamento interlacciato del segnale televisivo: dominio dei tempi

volume è perciò di tipo esagonale sulla coppia di assi  $y, t$  in quanto il reticolo della distribuzione delle righe (verticale) dei semiquadri dispari è sfalsato rispetto a quello dei semiquadri pari. Invece, l'asse  $x$  (le righe del segnale televisivo) non è campionato ma viene trasmesso in continuità nel tempo.

### 9.2.1 Equivocazione spazio - temporale

Vediamo ora cosa accade se il segnale non è opportunamente prefiltrato prima del campionamento spazio - temporale menzionato ed anche cosa accade se vogliamo una ricostruzione del volume  $x, y, t$  sopprimendo le armoniche del pettine di campionamento. L'asse  $x$  non è campionato ed è trasmesso con continuità; il filtraggio lungo questo asse è effettuato automaticamente se non altro dal filtraggio passa basso dovuto alla limitata banda disponibile a radiofrequenza e pertanto dai filtri che precedono la modulazione per la trasmissione.

Il segnale lungo le direzioni sia verticale che temporale è campionato e quindi dovrebbe subire un filtraggio anti equivocazione prima del campionamento. Invece, il prefiltraggio lungo la direzione verticale non avviene affatto se non per quanto è dovuto alla dimensione finita del pennello elettronico che esplora l'immagine e che convolve quindi con un filtro bidimensionale (la distribuzione delle ampiezze della sezione del pennello) l'immagine ripresa. Nel complesso questo filtraggio appare sufficiente, in quanto l'effetto di "moiré" e cioè di conversione in basso di alte frequenze spaziali verticali è abbastanza limitato. Questo effetto è visibile quando un'immagine a righe orizzontali fitte viene campionata dalla scansione televisiva; per l'equivocazione nascono delle inesistenti striscie di bassa frequenza.

Problemi più gravi insorgono invece per il campionamento temporale, decisamente scarso come si vede dalle immagini riprese con la moviola televisiva. Ad esempio, un calciatore ripreso in due quadri televisivi successivi occupa posizioni distanti tra di loro di molto più dell'intervallo tra le righe  $d$ . Infatti, se si muove a 10 m/sec in 40 msec ha percorso 40 cm; se la sua immagine occupa ad esempio una cinquantina di righe lungo la verticale ed uno spazio proporzionale lungo l'orizzontale,  $d$  corrisponde a circa 4 cm di spazio (se è alto 2 m) e perciò si muove di circa 10 "pixels" (picture elements) per quadro. La rotazione di fase per quadro delle componenti alla frequenza corrispondente a quella di Nyquist  $\pi/d$  è circa  $10\pi$  e quindi vi è equivocazione temporale di tutti i segnali che hanno una componente di frequenza spaziale maggiore di  $1/10d$  nella direzione del moto.

Nel fermo immagine non si possono quindi proiettare assieme i due semiquadri perchè corrispondono a due immagini spostate relativamente di un intervallo ben maggiore della risoluzione spaziale. Si proietta un

## CAPITOLO 9 SEQUENZE 2D

semiquadro solo, e da questo si interpola l'altro semiquadro generando un'immagine a risoluzione spaziale ridotta alla metà nella direzione verticale. Inoltre, la telecamera non può integrare la luminanza del segnale televisivo per tutta la durata di un quadro, come si faceva nel passato per migliorare il rapporto segnale - rumore, ma deve riprendere delle "istantanee" per evitare di avere poi delle immagini mosse alla moviola. Ovviamente, se le immagini fossero sempre proiettate in sequenza al loro ritmo, il prefiltraggio temporale effettuato con l'integrazione della telecamera sarebbe utile. La possibilità di dovere usare la moviola costringe ad introdurre del rumore di equivocazione. Peraltro l'osservatore umano, che interpreta le immagini in base a dei modelli a priori e non accetta l'ipotesi che il calciatore od il pallone cambino eccessivamente di forma tra un'immagine e la successiva, trova più accettabile il rumore di equivocazione al prefiltraggio, teoricamente corretto ma praticamente da respingere.

Da quanto osservato, risulta che il segnale televisivo è campionato temporalmente in modo da creare equivocazione e pertanto la ricostruzione dai campioni di un segnale continuo non può essere fatta senza commettere errori. Questo problema si pone quando volessimo interpolare immagini intermedie tra più quadri successivi per creare un segnale rallentato o un segnale a frequenza di quadro superiore (ad esempio 100 Hz) per limitare l'effetto di "sfarfallamento" (l'effetto fastidioso dovuto all'accensione periodica di una luce a frequenza al di sotto della frequenza di taglio del sistema visivo). Ad esempio, tecniche di interpolazione lineare interpolerebbero il moto di un pallone con delle dissolvenze tra un'immagine e la successiva che non potrebbero ricreare l'impressione del moto, che rimarrebbe a scatti; infatti nelle immagini interpolate il pallone scomparirebbe progressivamente da una posizione per riapparire nell'altra e non apparirebbe mai nelle posizioni intermedie. Ciò è possibile solo utilizzando tecniche di interpolazione complesse (a moto compensato) in cui:

a) si individuano i contorni dell'oggetto in moto confrontando le immagini successive;

b) si misura il vettore di moto, si prevede la futura posizione dell'oggetto, e si ricostruisce l'immagine.

Poiché l'interpolazione "a moto compensato" dipende dalle immagini da cui è rilevato il moto, si conclude che è una tecnica di interpolazione non lineare.

## 9.2 IL CAMPIONAMENTO DEL SEGNALE TELEVISIVO

### 9.2.2 Lo spettro del segnale televisivo campionato

Il campionamento spazio temporale sfalsato comporta una particolare struttura dello spettro del segnale televisivo campionato, descritta in figura 18.7. All'origine delle coordinate e' disegnato lo spettro spazio -

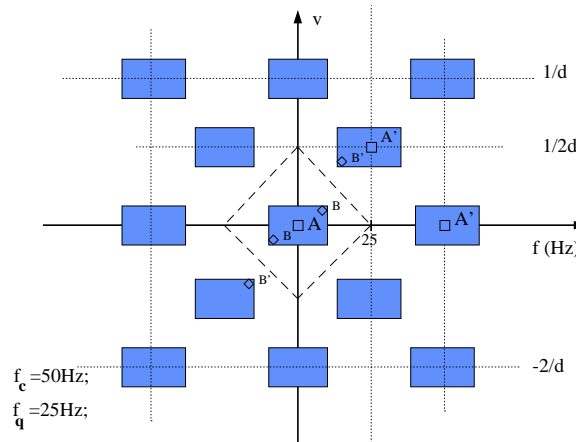


Figura 9.7: Campionamento interlacciato del segnale televisivo: dominio dei numeri d'onda.

temporale di una sequenza; gli assi sono i numeri d'onda verticali (asse verticale) le frequenze temporali (asse orizzontale). I numeri d'onda orizzontali sarebbero riportati ortogonalmente all'asse del foglio, ma non essendo questo asse campionato, non sono riportati affatto, semplificando la figura. Si può pensare pertanto che la figura corrisponda ad un segnale uniforme lungo le righe, variabile da riga a riga e da quadro a quadro.

Anzitutto si osservi che a causa della replica degli spettri, un segnale uniforme verticalmente ( $A$ ) si replica ( $A'$ ) alla frequenza  $f_c$ , frequenza di campo o di semiquadro, pari a  $2f_q$ , cioè il doppio della frequenza di quadro. Questo effetto motiva l'interallacciamento (la proiezione successiva di semiquadri corrispondenti alle righe pari e dispari). Infatti lo spettro corrispondente ad un'immagine p.e. tutta bianca si replica alla frequenza di 50 Hz, appena al di là della frequenza di taglio del sistema visivo umano; l'effetto di "sfarfallamento" che sarebbe stato inaccettabile se le successive immagini fossero state presentate alla frequenza di 25 Hz, è limitato. Similmente, nel cinema, le immagini si susseguono al ritmo di 24 fotogrammi al secondo, ma sono riproiettate due volte sullo schermo.



## CAPITOLO 9 SEQUENZE 2D

Peraltro, si può fare di meglio, e lo dimostra la commercializzazione di sistemi in cui il segnale televisivo è ripresentato sullo schermo al ritmo di 100 Hz, interpolando le immagini intermedie. L'interpolazione viene fatta preferenzialmente con algoritmi "a moto compensato" proprio per evitare gli effetti di equivocazione menzionati in precedenza nel caso della moviola e dei calciatori.

Un dettaglio verticale mobile ( $B$ ) si replica, equivocato, nel punto ( $B'$ ), che è circa a frequenza  $f_c$ . Consideriamo infatti un segnale di alta risoluzione verticale, una riga bianca in campo nero; se la riga è in posizione fissa, non vi sono problemi. Ma se la riga scende di un intervallo di riga per ogni quadro, sono possibili due casi, (a) viene riprodotta sempre (in tutti i quadri, e quindi sarà vista con energia doppia); (b) non sarà mai visibile; una riga suborizzontale in moto discendente (o ascendente) sarà vista apparire e scomparire in modo casuale (equivocazione spazio-temporale). Questo fenomeno si osserva direttamente, p. e. in occasione di una partita di tennis, quando le righe del campo (suborizzontali) sono viste a gradinate ed i gradini tremolano quando la telecamera si muove. E' il solito problema dell'equivocazione, visto all'inizio del testo per i sistemi antincendio.

Dalla figura 18.7 ssi comprende anche come un filtro spazio temporale per la soppressione degli spettri replicati dovuti al pettine di campionamento debba avere il contorno tratteggiato e quindi il filtro non può essere realizzato p.e. con la cascata separabile di filtri spaziali e temporali, che porterebbe ad un filtraggio a base rettangolare.

### 9.3 Ricostruzione 2D

Per ricostruire un segnale continuo bidimensionale si deve convolvere la sequenza dei campioni per la risposta all'impulso di un filtro opportuno; ad esempio un mantentore bidimensionale, la cui risposta all'impulso è il prisma a base quadrata in fig.16.2 o l'interpolatore bilineare; quest'ultimo corrisponde ad interpolare linearmente il segnale prima lungo le righe in corrispondenza dei campioni e poi lungo le colonne per tutti i valori (continui, ora) delle ascisse. Si può anche fare viceversa, il risultato sarà eguale. La risposta all'impulso dell'interpolatore bilineare è quindi costituita dalla convoluzione di due lame triangolari disposte lungo i due assi coordinati e larghe  $2\Delta$ ; si osservi che questa superficie **non** è una

### 9.3 RICOSTRUZIONE 2D

piramide a base quadrata; le superfici dei lati appartengono infatti ad iperboloidi iperbolici (fig.18.8). Le trasformate di Fourier del mantenitore

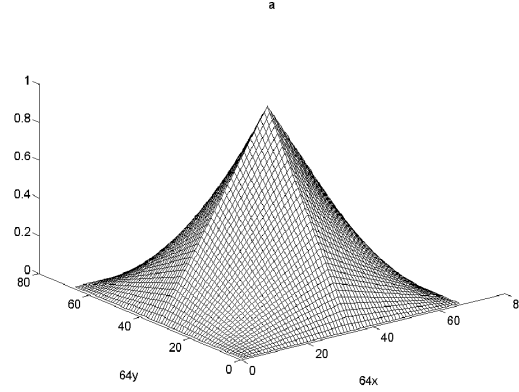


Figura 9.8: L'interpolatore bilineare: i lati della piramide *pigra* sono lembi di iperboloidi iperbolici.

e dell'interpolatore bilineare sono:

$$H_m(u, v) = \text{sinc}(u\Delta)\text{sinc}(v\Delta); H_l(u, v) = [\text{sinc}(u\Delta)\text{sinc}(v\Delta)]^2; \quad (9.11)$$

La piramide *egizia* (fig.18.9) con la stessa base quadrata dell'interpolatore bilineare **non** è un buon interpolatore perchè interpola correttamente lungo le facce ma erratamente lungo gli spigoli; se la base fosse ancora quadrata ma con i vertici a terra sui quattro campioni adiacenti al campione assegnato lungo gli assi coordinati, la superficie interpolerebbe bene lungo gli spigoli e male lungo le facce. Per convincersene, basta pensare che il volume di queste piramidi è  $4\Delta^2 H/3$  o  $2\Delta^2 H/3$  invece di  $\Delta^2 H$ ; nel caso di tutti i campioni eguali il volume totale eccederebbe o sarebbe inferiore a quello corretto. Un interpolatore con facce piane che rispetta i piani (e cioè se i campioni sono disposti su un piano, la superficie interpolante coincide con il piano) è invece la piramide a base esagonale rappresentata in figura 18.10. Per convincersene basta tracciare una diagonale del quadrato base del reticolo per dividerlo in due triangoli. Una faccia della superficie interpolante è il piano che passa per i tre campioni posizionati nei vertici del triangolo cui appartiene il punto da interpolare. Ovviamente, vi sono due possibilità di tracciare la diagonale e quindi due

## CAPITOLO 9 SEQUENZE 2D

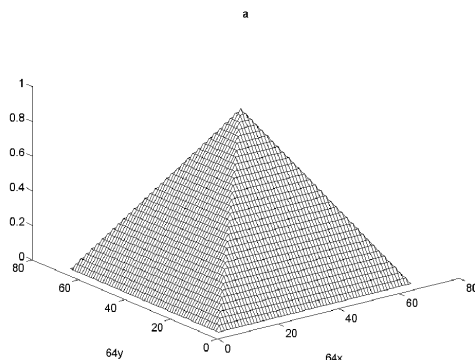


Figura 9.9: La piramide "egizia": un cattivo interpolatore a 2D.

superfici interpolanti che hanno diversa funzione di trasferimento lungo le due diagonali.(fig.18.11). Il loro volume é quello corretto (si veda per ulteriori dettagli il paragrafo sulla triangolazione di Delaunay, alla fine del capitolo).Facendone la media per avere una soluzione piú isotropa, si ottiene per la risposta all'impulso la figura 18.12, dove ancora la superficie interpolante ha facce piane, ma i tre punti che definiscono i vari piani sono i due campioni piú vicini al punto da interpolare ed il campione "supplementare" posto al centro del quadrato base e la cui ampiezza è la media delle ampiezze dei quattro campioni ai vertici del quadrato (0,25). Alla fine, abbiamo approssimato l'interpolatore bilineare con un poliedro a 16 facce piane di forma triangolare (fig. 18.13).Infine, una tecnica molto usata è quella degli splines bicubici: di nuovo si interpola ad esempio prima lungo la direzione orizzontale, questa volta con un interpolatore a bacchetta elastica; quindi, si interpolano lungo la direzione verticale, punto per punto, le funzioni già interpolate, ormai continue lungo la direzione orizzontale.

### 9.4 Trasformata di Fourier discreta a 2D

L'estensione a 2D della trasformazione di Fourier discreta è semplice: la periodicità tuttavia è ora su due assi invece che su uno e pertanto le sequenze antitrasformate sono definite su di una superficie torica piuttosto che su un circolo. Questo è congruente con quanto osservato per le

## 9.4 TRASFORMATATA DI FOURIER DISCRETA A 2D

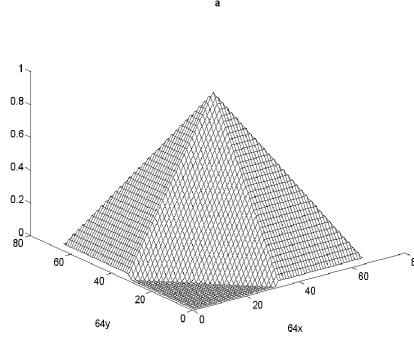


Figura 9.10: Piramide a base esagonale: un possibile interpolatore lineare a 2D anisotropo.

trasformate di segnali campionati negli spazi; il campionamento è ora sia negli spazi che nelle frequenze spaziali (numeri d'onda). L'espressione della trasformazione discreta è la seguente:

$$X(h, k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} x_{n,m} W_N^{-nh} W_M^{-mk}; \quad W_N = e^{j2\pi/N} \quad (9.12)$$

L'antitrasformata è:

$$x(n, m) = \frac{1}{MN} \sum_{h=0}^{M-1} \sum_{k=0}^{N-1} X_{h,k} W_N^{nh} W_M^{mk}; \quad W_M = e^{j2\pi/M} \quad (9.13)$$

La trasformazione si può fare in cascata, prima per righe e poi per colonne:

$$x(n, m) = \frac{1}{MN} \sum_{h=0}^{M-1} W_N^{nh} \left( \sum_{k=0}^{N-1} X_{h,k} W_M^{mk} \right); \quad (9.14)$$

La quantità in parentesi è l'antitrasformata di Fourier effettuata per righe; la trasformata di Fourier 2D avviene quindi prima facendo la trasformazione per righe e poi per colonne (o viceversa). Il costo in termini di moltiplicazioni è:

$$C = \frac{MN}{2} \log_2 N + \frac{MN}{2} \log_2 M = \frac{MN}{2} \log_2 NM \quad (9.15)$$

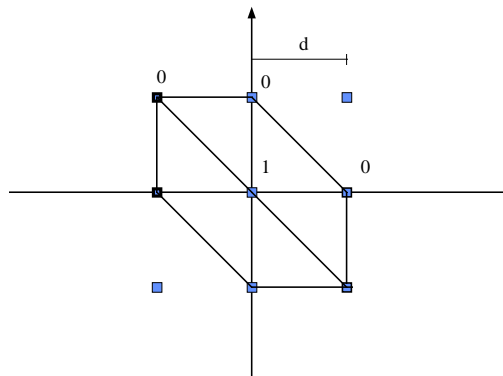


Figura 9.11: L'interpolatore a base esagonale visto in pianta; il comportamento differente nelle due direzioni a  $90^\circ$  è evidente.

## 9.5 Filtraggi con prodotto di trasformate

Con opportune aggiunte di zeri, in modo che le due funzioni abbiano eguale supporto e cioè siano definite sullo stesso dominio, è possibile farne la convoluzione facendo il prodotto delle trasformate. E' opportuno notare tuttavia che l'aggiunta di zeri è più costosa a 2D che a 1D; infatti se un filtro ha un supporto  $M \times M$  ed il segnale  $N \times N$ , per eliminare gli effetti della periodicità, si devono aggiungere circa  $2NM + M^2$  zeri, e cioè in proporzione circa il doppio che a 1D. In altre parole, il costo del *zero padding* è cresciuto con la dimensionalità dei dati. Il costo del filtraggio di una sequenza  $N \times N$  sarà

$$2N^2(2^{\frac{1 + \log_2 N}{2}}) \quad (9.16)$$

tenendo conto delle trasformate per righe e colonne, delle antitrasformate e del prodotto tra le trasformate per il filtraggio.

## 9.6 Filtri 2D

Il teorema fondamentale dell'algebra non si estende a 2 dimensioni; è vero che il prodotto di due polinomi in 2 variabili è la convoluzione bidimensionale dei loro coefficienti, ma non è vero che un polinomio in 2 variabili

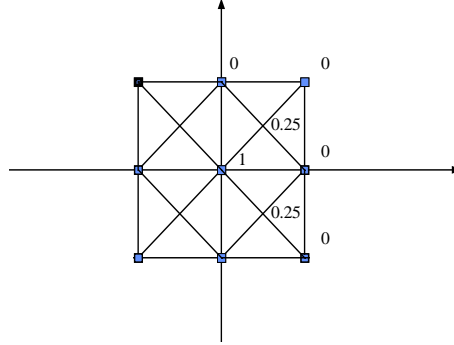


Figura 9.12: Media di due interpolatori a base esagonale orientati reciprocamente a  $90^\circ$ .

è sempre decomponibile nel prodotto di polinomi di grado minore. Quindi l'utilità della trasformata  $z$  a due dimensioni è minore perchè non si estende a 2D il concetto di zeri e quindi di poli e cioè sequenze semplici in cui possono venire decomposte le sequenze più complesse. Pertanto il filtraggio 2D è prevalentemente realizzato con tecniche di prodotto delle trasformate di Fourier discrete oppure come convoluzione 2D con opportune sequenze, p.e. trovate finestrando le antitrasformate delle funzioni di trasferimento desiderate. E' invece utile ricorrere al concetto di separabilità per trasformare un problema 2D nella cascata di  $2k$  problemi 1D, approssimando la sequenza 2D come:

$$h_{m,n} = \sum_{k=1}^K u_{k,m} v_{k,n} \quad (9.17)$$

Aggiungendo zeri, si può sempre far corrispondere la sequenza  $h_{m,n}$  ad una matrice quadrata; questa matrice a sua volta può essere rappresentata in funzione dei suoi autovettori ed autovalori, come somma di diadi;

$$\mathbf{R} = \sum_{n=1}^N \lambda_n \mathbf{e}_n \mathbf{e}_n^* \quad (9.18)$$

Pertanto  $K \leq N$  se  $|m, n| \leq N$ . Una convoluzione bidimensionale per una funzione separabile può essere realizzata come cascata di due convoluzioni monodimensionali per le due funzioni che compaiono nel

## CAPITOLO 9 SEQUENZE 2D

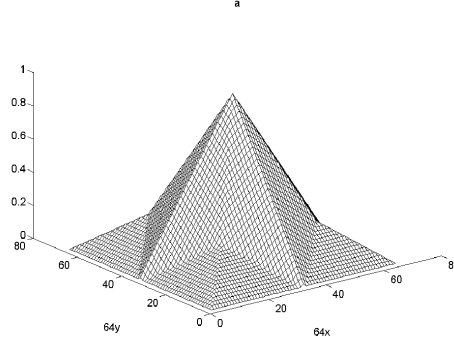


Figura 9.13: Approssimazione dell'interpolatore bilineare con un interpolatore a facce piane.

prodotto con un costo computazionale totale pari a  $2NK$ ; questo modo di operare spesso è assai conveniente perchè si ottiene una buona approssimazione della  $h_{m,n}$  con  $K \ll N$ . Un esempio molto utile è quello della trasformazione di Mc Clellan, che serve a generare filtri a simmetria circolare da filtri 1D.

### 9.6.1 La trasformazione di Mc Clellan

Si voglia realizzare un filtro con funzione di trasferimento a simmetria circolare ed a fase nulla.

$$H(u, v) = H(\Omega); \quad \Omega = \sqrt{u^2 + v^2} \quad (9.19)$$

Visto che la funzione di trasferimento 1D  $H(\Omega)$  è anch'essa a fase 0, abbiamo:

$$H(\Omega) = a_0 + \sum_{n=1}^N a_n \cos(n\Omega\Delta) \quad (9.20)$$

La trasformazione è la seguente:

$$\cos \Omega\Delta \rightarrow -1 + \frac{1}{2}(1 + \cos u\Delta)(1 + \cos v\Delta) \quad (9.21)$$

somma di un impulso e di una funzione separabile. L'operazione di moltiplicazione (nelle frequenze) per  $\cos u\Delta$  corrisponde alla convoluzione

(negli spazi) per la sequenza di due campioni di ampiezza  $1/2$  collocati nei campioni di coordinate  $(-1, 0)$ ;  $(1, 0)$ ; similmente per  $\cos v\Delta$ . Quindi il costo per campione di una convoluzione (la moltiplicazione della trasformata per  $\cos \Omega\Delta$ ) è (a parte la moltiplicazione per il coefficiente) pari a 7 somme. Se ora si vuole realizzare l'intero filtro 2D e quindi convolvere per le antitrasformate di funzioni come  $\cos n\Omega\Delta$  si può operare utilizzando la formula di prostaferesi:

$$\cos n\Omega\Delta = 2 \cos[(n-1)\Omega\Delta] \times \cos \Omega\Delta - \cos[(n-2)\Omega\Delta] \quad (9.22)$$

In questo modo, le convoluzioni corrispondenti al termine  $n$ -esimo dei polinomi trigonometrici non devono essere calcolate direttamente ma sono ottenute come combinazione dei risultati ottenuti per valori minori di  $n$  ripetendo poi la stessa operazione di convoluzione elementare (la trasformazione lineare indicata come  $G(u, v)$  nella figura 18.15). Le isolinee della trasformazione di Mc Clellan per  $n = 1$  sono rappresentate in fig.18.14; per  $n$  maggiori, la zona in cui l'approssimazione è accettabile si riduce in proporzione. e lo schema a blocchi è presentato in figura

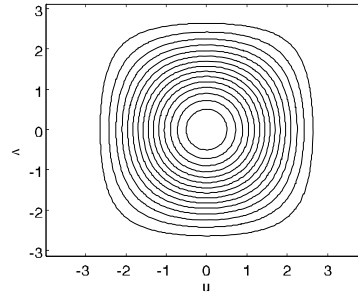


Figura 9.14: Isolinee della trasformazione di Mc Clellan

18.15.

### 9.6.2 Filtri per la continuazione di campo (wavefield continuation)

Come si è visto nel paragrafo sui filtri 2D derivanti dall'equazione del campo gravitazionale, il campo a quota  $z + \Delta z$  è ottenuto da quello a



## CAPITOLO 9 SEQUENZE 2D

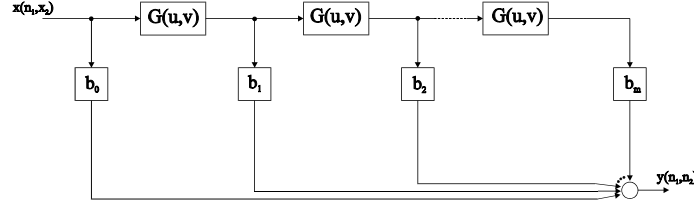


Figura 9.15: Schema a blocchi per il calcolo della trasformazione di McClellan

quota  $z$ , applicando la funzione di trasferimento:

$$\exp(-|\Omega| \Delta z) \quad (9.23)$$

Nel caso del telerilevamento, si è visto che si può rilevare la morfologia del sottosuolo se p.e. si dispongono in superficie ( $z = 0$ ) sorgenti impulsive in tutte le posizioni  $x, y$  e si registra ancora in tutte le posizioni il campo riflesso  $f(x, y, z = 0, t)$ . In questo tipo di esperimento radar, la posizione della sorgente coincide con la posizione del ricevitore. Come avevamo osservato, il campo riflesso misurato coincide con quello che sarebbe misurato in un mezzo a velocità di propagazione dimezzata se i riflettori emettessero un impulso di energia acustica al tempo 0, avente ampiezza proporzionale alla loro riflettività. L'ampiezza delle sorgenti fittizie (e quindi dei riflettori) a profondità  $z$  viene determinata ricostruendo il campo a profondità  $z$  e trovandone il valore al tempo 0. Il campo a profondità  $z$  viene ottenuto da quello in superficie, applicando la funzione di trasferimento (questa volta a 3D, piuttosto che a 2D)

$$\exp\left(jz\sqrt{\frac{\omega^2}{c^2} - \Omega^2}\right); \Omega = u^2 + v^2 \quad (9.24)$$

oppure convolvendo con l'antitrasformata di Fourier a 3D, indicata con  $F_{3D}^{-1}$ .

$$f(x, y, z, t) = f(x, y, z = 0, t) *** F_{3D}^{-1} \left[ \exp\left(jz\sqrt{\frac{\omega^2}{c^2} - \Omega^2}\right) \right] \quad (9.25)$$

## 9.7 INTERPOLAZIONE MEDIANTE STIMA LINEARE

e trovandone il valore al tempo 0. Nel dominio trasformato,

$$F(u, v, z = 0, \omega) = \int \int \int f(x, y, z = 0, t) \exp [-j (ux + vy + \omega t)] dx dy dt \quad (9.26)$$

$$F(u, v, z, \omega) = F(u, v, z = 0, \omega) \exp \left( jz \sqrt{\frac{\omega^2}{c^2} - \Omega^2} \right) \quad (9.27)$$

$$f(x, y, z, t = 0) = \int F(u, v, z, \omega) \frac{d\omega}{2\pi} \quad (9.28)$$

Invece di fare trasformazioni di Fourier a 3D, che implicherebbe la stazionarietà spaziale e quindi la costanza ovunque della velocità di propagazione, si può semplicemente convolvere a 2D (e cioè in  $x, y$ ) la trasformata 1D (nel tempo  $t$ ) del campo riflesso, con il vantaggio non trascurabile di avere uno strumento utilizzabile anche quando la velocità di propagazione è funzione di  $x, y$ , ed anche di  $z$ .

$$f(x, y, z, \omega) = f(x, y, z = 0, \omega) * * F_{2D}^{-1} \left[ \exp \left( jz \sqrt{\frac{\omega^2}{c^2} - \Omega^2} \right) \right] \quad (9.29)$$

Filtri del tipo:

$$F_{2D}^{-1} [\exp (-|\Omega| \Delta z)], \quad F_{2D}^{-1} \left[ \exp \left( jz \sqrt{\frac{\omega^2}{c^2} - \Omega^2} \right) \right] \quad (9.30)$$

sono appunto realizzabili con la trasformazione di McClellan, che può essere facilmente resa spazio variante.

## 9.7 Interpolazione mediante stima lineare

Una tecnica di interpolazione utile è quella che fa uso della stima lineare; i valori della funzione al di fuori dei punti del reticolo vengono stimati con uno stimatore lineare. Per discutere di questa tecnica è opportuno fare seguire un breve riassunto delle proprietà dei processi casuali bidimensionali e della loro stima.

### 9.7.1 Processi casuali a 2D

Supponiamo di avere un segnale 2D,  $s(x, y)$  realizzazione di un processo stocastico, a valor medio nullo, campionato su un reticolo di lato  $\Delta$ ; si può supporre inoltre di non avere a disposizione  $s_{m,n}$  ma una versione disturbata da rumore bianco

$$s'_{m,n} = s(m\Delta, n\Delta) + z_{m,n}; \quad \forall m, n; \quad (9.31)$$

La sequenza 2D, disturbante, è costituita da campioni indipendenti  $z_{m,n}$  e risulta:

$$\begin{aligned} E[z_{m,n}z_{h,k}] &= \sigma_z^2; \quad m = h; \quad n = k; \\ E[z_{m,n}z_{h,k}] &= 0; \quad m \neq h; \quad n \neq k \end{aligned} \quad (9.32)$$

La sequenza di segnale è ottenuta p.e. convolvendo un'altra sequenza "bianca"  $w_{m,n}$  con un filtro  $h_{m,n}$  e si ha:

$$s_{m,n} = w_{m,n} * h_{m,n} = \sum_{k,l=-N/2}^{N/2} w_{m-k,n-l} h_{k,l} \quad (9.33)$$

La funzione di autocorrelazione della sequenza 2D  $s'_{m,n}$  è:

$$E[s'_{m,n}s'_{m+k,n+l}] = r'_{k,l} = \sigma_w^2 \sum_{i,j=-N/2}^{N/2} h_{k,l} h_{i+k,j+l} + \sigma_z^2 \delta(k, l) \quad (9.34)$$

### 9.7.2 Spettri di potenza a 2D

Similmente a quanto accade a 1D, lo spettro 2D di una sequenza si ottiene per trasformazione di Fourier discreta della funzione di autocorrelazione. Nel caso di rumore bianco a 2D si ha:

$$S_x(h, k) = F\{r_x(n, m)\} = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} r_{n,m} W_N^{-nh} W_M^{-mk} = \sigma_x^2; \quad (9.35)$$

Similmente ancora a quanto accade a 1D, se la sequenza  $x(i, j)$  con trasformata di Fourier discreta  $X(h, k)$  è convoluta con un filtro a risposta all'impulso  $h(i, j)$  e avente trasformata di Fourier discreta  $H(h, k)$ , lo spettro cambia come segue:

$$y(i, j) = x(i, j) * h(i, j) \quad (9.36)$$

## 9.7 INTERPOLAZIONE MEDIANTE STIMA LINEARE

$$Y(h, k) = X(h, k) \times H(h, k)$$

$$E[|Y(h, k)|^2] = E[|X(h, k)|^2] \times |H(h, k)|^2 \quad (9.37)$$

In particolare, si possono considerare sequenze incorrelate in una direzione e correlate lungo la direzione ortogonale. In questo caso, la funzione di correlazione è a lama, e lo spettro è cilindrico.

### 9.7.3 Stima lineare a 2D

Noti i valori di  $s'_{n,m}$  sul reticolo di campionamento si vuole ora stimare  $s(x, y)$  in tutti i punti del piano. Costruendo lo stimatore come combinazione lineare degli  $(H+1)^2$  valori assunti dalla sequenza  $s'_{n,m}$  in un intorno  $H \times H$  del punto da stimare, risulta:

$$\widehat{s}(x, y) = \sum_{n,m=-H/2}^{+H/2} a_{n,m}(x, y) s'_{n,m} \quad (9.38)$$

I coefficienti  $a_{n,m}(x, y)$  sono funzione continua di  $x, y$  e sono determinabili in base al principio di ortogonalità.

$$E[\{s(x, y) - \widehat{s}(x, y)\} s'_{n,m}] = 0; \quad |n, m| \leq H/2 \quad (9.39)$$

Questo sistema è detto il sistema di equazioni **normali** o di Wiener - Hopf ed è:

$$\sum_{p,q=-H/2}^{+H/2} a_{p,q}(x, y) r'[(p-n)\Delta, (q-m)\Delta] = \quad (9.40)$$

$$r(x-n\Delta, y-m\Delta); \quad |n, m| \leq H/2$$

$$r'(p, q) = r(x=p\Delta, y=q\Delta) + \sigma_w^2 \delta(p, q)$$

Generalizzando la stima al caso di reticolo irregolare, si possono indicare con un solo indice  $i = 1, \dots, N$  gli  $N$  punti in cui la funzione  $s(x, y)$  è assegnata. Definiamo quindi il vettore dei dati come:

$$\underline{\mathbf{s}} = \begin{bmatrix} s(x_1, y_1) \\ s(x_2, y_2) \\ \dots \\ s(x_N, y_N) \end{bmatrix}$$

## CAPITOLO 9 SEQUENZE 2D

La stima della superficie in un punto qualsiasi  $(x, y)$  è combinazione lineare dei valori della funzione in tutti gli  $N$  punti dati con un vettore di coefficienti  $\underline{\mathbf{a}}(x, y)$ , dipendente dalla posizione del punto da stimare:

$$\widehat{s}(x, y) = \underline{\mathbf{s}}^* \underline{\mathbf{a}}(x, y)$$

Il vettore dei coefficienti  $\underline{\mathbf{a}}(x, y)$  viene dal sistema di equazioni normali:

$$\underline{\mathbf{R}} \underline{\mathbf{a}}(x, y) = \underline{\mathbf{r}}(x, y)$$

Gli elementi della matrice di covarianza  $\underline{\mathbf{R}}$  sono:

$$r_{m,n} = r(x_m - x_n, y_m - y_n) + \sigma_w^2 \delta(m, n); \quad m, n = 1, \dots, N$$

e il vettore  $\underline{\mathbf{r}}(x, y)$  è la correlazione tra il valore della funzione nel punto da stimare e quello nei punti dove è assegnata:

$$\underline{\mathbf{r}} = \begin{bmatrix} r(x - x_1, y - y_1) \\ r(x - x_2, y - y_2) \\ \dots \\ r(x - x_N, y - y_N) \end{bmatrix}$$

Si verifica che l'interpolatore **non** onora il dato, se non nel caso particolare in cui  $\sigma_w^2 = 0$ . Quindi, l'interpolatore può funzionare anche come filtro passa basso, per smussare il segnale misurato qualora sia affetto da rumore a spettro bianco. Il vantaggio dell'interpolazione realizzata con la stima lineare è appunto nel fatto che le formule possono essere generalizzate a situazioni di campionamento irregolare ed a rumore con spettro qualsiasi, p.e. più uniforme in una direzione che in un'altra. Non tutte le funzioni  $r(x, y)$  sono possibili funzioni di autocorrelazione; è necessario che la loro trasformata sia tutta positiva. Solo in questo caso è possibile garantire che, aggiungendo luce bianca  $\sigma_w^2$  non compaiano zeri nello spettro degli autovalori di  $\underline{\mathbf{R}}$  e quindi il sistema delle equazioni normali sia ben condizionato. Ad esempio sono utilizzabili come funzioni di autocorrelazione funzioni del tipo:

$$r_1(x, y) = \frac{1}{d_0^2} e^{-\frac{|x|}{|d_0|} - \frac{|y|}{|d_1|}}; \quad r_2(x, y) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2}\right) \quad (9.41)$$

## 9.7 INTERPOLAZIONE MEDIANTE STIMA LINEARE

rispettivamente separabile (la prima) e separabile ed a simmetria circolare (la seconda, se  $\sigma_x^2 = \sigma_y^2$ ). I rispettivi spettri di potenza saranno:

$$P_1(u, v) = \frac{1}{1 + u^2 d_0^2} \times \frac{1}{1 + v^2 d_1^2} \quad (9.42)$$

$$P_2(u, v) = \exp(-u^2 \sigma_x^2 - v^2 \sigma_y^2) \quad (9.43)$$

Non andrebbe bene un prisma; andrebbe bene la forma dell'interpolatore bilineare.

### 9.7.4 Stima lineare ed interpolazione

Si osservi che la tecnica della stima lineare, nel caso di  $\sigma_w^2 = 0$ , viene a coincidere con la tecnica di interpolazione discussa nel capitolo 2 in cui la funzione interpolata è combinazione lineare di tante funzioni lisce  $g(t - t_j)$  centrate agli istanti  $t_j$  dove sono posizionati i campioni. I coefficienti della combinazione lineare sono tali da fare sì che la funzione interpolata rispetti i dati. La stessa cosa accade in questo caso; invece di un asse temporale, abbiamo due assi spaziali ed inoltre il ruolo della  $g(t)$  è ora sostenuto dalla funzione di autocorrelazione  $r(x, y)$ . Infatti, si può verificare che la funzione interpolata è somma di tante funzioni di autocorrelazione centrate sui campioni e con coefficienti tali che la loro somma rispetti il dato; con scrittura vettoriale, abbiamo trovato:

$$\underline{\mathbf{R}} \underline{\mathbf{a}}(x, y) = \underline{\mathbf{r}}(x, y) \quad (9.44)$$

La matrice  $\underline{\mathbf{R}}$  non dipende dalle variabili continue  $x, y$  ma solo dalla disposizione del reticolo di campionamento e ovviamente dalla forma della funzione di autocorrelazione. Inoltre, avendo indicato con  $\underline{\mathbf{s}}$  il vettore dei campioni da interpolare:

$$\widehat{s}(x, y) = \underline{\mathbf{s}}^* \underline{\mathbf{a}}(x, y) = \underline{\mathbf{s}}^* \underline{\mathbf{R}}^{-1} \underline{\mathbf{r}}(x, y) = \underline{\mathbf{q}}^* \underline{\mathbf{r}}(x, y) = \quad (9.45)$$

$$= \begin{bmatrix} q_1 & q_2 & \dots & q_N \end{bmatrix} \begin{bmatrix} r(x - x_1, y - y_1) \\ r(x - x_2, y - y_2) \\ \dots \\ r(x - x_N, y - y_N) \end{bmatrix}$$

## CAPITOLO 9 SEQUENZE 2D

Il vettore

$$\underline{\mathbf{q}} = \underline{\mathbf{R}}^{-1} \underline{\mathbf{s}}$$

non dipende dalle variabili continue  $x, y$  ma solo dai dati, dalla  $\underline{\mathbf{R}}$ , e dalla posizione reciproca dei punti di campionamento. Pertanto, come già detto, la stima  $\hat{s}(x, y)$  è una combinazione lineare di funzioni di autocorrelazione centrate sugli  $N$  campioni, e moltiplicate per coefficienti  $q_1, \dots, q_N$  tali da garantire il rispetto del dato, sempre che  $\sigma_w^2 = 0$ . In questo caso, infatti, per  $x = x_i, y = y_i$  il vettore dei termini noti  $\underline{\mathbf{r}}(x, y)$  è eguale ad una riga della matrice  $\underline{\mathbf{R}}$  e il vettore  $\underline{\mathbf{a}}(x = x_i, y = y_i)$  si riduce ad un vettore tutto di zeri ed un solo 1 nella posizione  $i$ -esima. La funzione stimata è quindi eguale al campione situato in quella posizione. In conclusione, abbiamo ora uno strumento per dare un'interpretazione fisica alla funzione  $g(t)$  incontrata nel capitolo 2.

### 9.8 Poligoni di Voronoi e triangoli di Delaunay

Una distribuzione di campioni irregolare su un piano può essere interpolata con una tecnica molto semplice, generalizzazione immediata del mantentore nel caso di un reticolo regolare. Basta infatti associare ogni punto del piano al campione più vicino: la zona di influenza di ogni campione, poligonale, è chiamata poligono di Voronoi ed è delimitata dagli assi dei segmenti che congiungono ogni campione a quelli più vicini, con un lato di un poligono di Voronoi in comune. Se poi si connette ogni campione ai campioni vicini (e cioè quei campioni che condividono un lato di un poligono di Voronoi) si ottiene una suddivisione del piano in triangoli. Questa suddivisione è chiamata mosaico di Delaunay (Delaunay tessellation). La figura in basso illustra il procedimento. La procedura di triangolazione di Delaunay identifica una tecnica di interpolazione semplicissima: basta imporre che ad ogni triangolo corrisponda un piano che passa per i tre punti: ne risulta una superficie a faccette piane che onora il dato. Nel caso di un reticolo regolare ed una triangolazione di Delaunay uniforme ovunque (ne sono ammesse due differenti tra loro), la risposta all'impulso è una delle due piramidi a base esagonale viste in precedenza.

Nel caso di disposizioni casuali nel piano di  $N$  campioni, si può osser-

## 9.8 POLIGONI DI VORONOI E TRIANGOLI DI DELAUNAY

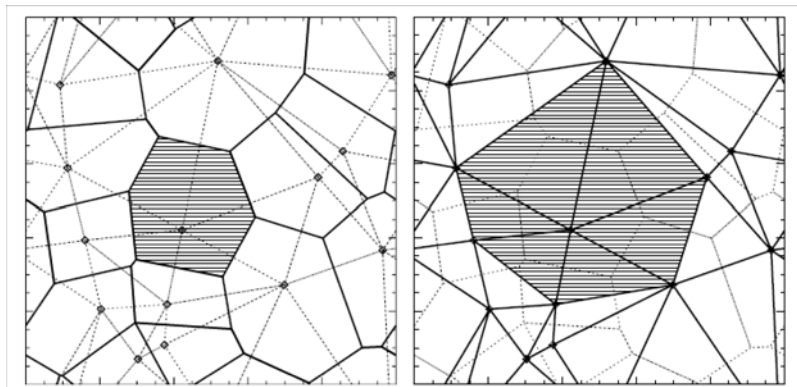


Figura 9.16: A sinistra i poligoni di Voronoi e a destra la triangolazione di Delaunay.

vare che il numero di poligoni di Voronoi é ovviamente eguale al numero  $N$  dei campioni. Quindi, la formula di Eulero

$$Facce + Vertici = Spigoli + 2 \quad (9.46)$$

applicata al poliedro costituito dai poligoni di Voronoi, ci dice che il numero medio di lati dei poligoni di Voronoi é 6. In altre parole, disponendo a caso dei campioni sul piano, ogni campione *confina* in media con altri 6. Infatti, i vertici dei polgoni di Voronoi sono sempre formati dall'incontro di 3 spigoli (a parte casi degeneri, come il reticolo regolare a base rettangolare, la probabilità che esistano punti equidistanti da 4 campioni o più é trascurabile) e quindi:

$$N + V \sim \frac{3V}{2} \rightarrow V \sim 2N \rightarrow S \sim \frac{6N}{2} \quad (9.47)$$

ricordando di non contare 2 volte gli spigoli. I triangoli di Delaunay sono tanti quanti i vertici e quindi il doppio dei campioni. In una simulazione su  $2 \cdot 10^8$  campioni distribuiti nel piano alla Poisson (Brakke), le frequenze relative del numero dei lati dei poligoni di Voronoi (da 3 a 11) sono state misurate così: 0.01, 0.1, 0.26, 0.29, 0.19, 0.09, 0.03, 0.007, 0.0015. La probabilità di avere poligoni con un numero ancora maggiore di lati é trascurabile anche se mai nulla.



### 9.8.1 Il campionamento a 3D

L'interpolazione lineare si generalizza semplicemente anche nel caso di funzioni campionate in modo casuale in spazi a 3 dimensioni. Vediamo anzitutto la generalizzazione del reticolo di campionamento ottimale. A 2D abbiamo visto che il reticolo a base esagonale è più funzionale di quello a base quadrata. Un modo per spiegare questo fatto sta nell'osservazione che, se dobbiamo impaccare sfere 2D (e quindi cerchi) di eguale diametro, la densità massima si ottiene quando i cerchi sono disposti su un reticolo a base esagonale. Se vediamo questo risultato nel dominio delle frequenze 2D, ne risulta che, a pari diametro del cerchio base e quindi a pari frequenza spaziale massima che si può riprodurre senza equivocazione, le repliche sono più vicine tra loro (più dense) se si adotta un reticolo esagonale. Ogni cerchio ne tocca altri sei, e peraltro anche con una distribuzione casuale dei campioni sul piano, avevamo verificato che il numero medio dei campioni vicini era sei. Reciprocamente, il reticolo dei campioni nello spazio sarà il più rado possibile, e quindi l'impaccamento fitto esagonale nelle frequenze porta al reticolo di campionamento ottimo, e cioè il più rado possibile per una data frequenza di taglio.

Similmente a 3D, si è di recente dimostrato quanto è intuitivo e cioè che il massimo impaccamento di sfere si ottiene con una disposizione quale quella di una piramide di palle di cannone; in sostanza, questa disposizione corrisponde alla sovrapposizione di strati di sfere con i centri su reticoli piani esagonali. Gli strati successivi hanno i centri sfalsati in maniera da fare penetrare le sfere dello strato superiore nelle cavità lasciate libere nello strato inferiore. Ogni sfera ne tocca altre 12.

Ancora, si possono estendere a 3D sia la triangolazione di Delaunay che i poligoni (poliedri) di Voronoi. Questi ultimi corrispondono a tutti i punti che sono vicini ad un campione più che a tutti gli altri. I poliedri sono separati da tanti piani in asse tra un campione ed uno vicino. Ogni vertice di un poliedro di Voronoi ha 4 spigoli. Questo vertice è equidistante da quattro campioni ed è il centro della sfera che li tocca. Quindi i piani che si incontrano in un vertice di un poliedro di Voronoi sono i quattro piani in asse tra i campioni equidistanti da quel vertice. Queste quaterne di campioni vicini, e cioè i vertici del tetraedro iscritto nella sfera identificata dai quattro campioni equidistanti dal vertice del poliedro in questione, formano un tetraedro di Delaunay, generalizzazione di un triangolo di Delaunay nella triangolazione piana. Così come si può

## 9.8 POLIGONI DI VORONOI E TRIANGOLI DI DELAUNAY

disporre un piano per 3 punti, si può disporre su quattro punti nello spazio una funzione che varii linearmente in funzione delle tre coordinate. In questo modo, è possibile definire una funzione continua in uno spazio 3D che assume i valori desiderati in corrispondenza dei campioni noti ed è localmente sempre variabile linearmente con le tre coordinate. In corrispondenza delle facce dei tetraedri di Delaunay, questa funzione ha delle discontinuità delle derivate prime.

Per curiosità, il numero medio di campioni adiacenti in uno spazio 3D, e con una distribuzione casuale alla Poisson dei campioni, è stato calcolato da Meijering (1953) in:

$$\frac{48\pi^2}{35} + 2 = 15,53.. \quad (9.48)$$

La relazione di Eulero, anche se generalizzata a 4D, non sarebbe sufficiente per trovare questo valore. È invece necessario fare un complesso studio di geometria aleatoria.



# Capitolo 10

## La Tomografia

### 10.1 Introduzione

Si pensi ad un bambino che osserva una torta con delle candeline; la torta è poggiata su di un tavolo ed il bimbo, che ha gli occhi all'altezza delle candeline, non può distinguere nè il loro numero nè la luminosità. Se però il bambino fa un giro attorno al tavolo su cui è poggiata la torta, vedrà le luci delle candeline allinearsi a volte lungo certe direzioni, a volte lungo altre. A volte ancora vedrà tutte le candeline distribuite lungo il diametro della torta. Ora, la tomografia è la tecnica con cui si riesce a trovare posizione e luminosità delle candeline pur senza mai vedere la torta dall'alto.

Passando dal caso particolare della torta con le candeline al caso generale della tomografia, abbiamo la seguente situazione. Si desidera rilevare una grandezza (ad esempio l'energia luminosa emessa dalle candeline) il cui andamento nel piano è descritto dalla funzione  $f(x, y)$ ; questa grandezza tuttavia non può essere rilevata punto per punto; la cosa sarebbe possibile se disponessimo di un misuratore nel punto di coordinate  $x, y$  ma questo è invece inaccessibile. Stando all'esterno del corpo, invece, è possibile misurare solo **proiezioni** e cioè integrali della funzione  $f(x, y)$  lungo le rette con direzione  $\varphi + \pi/2$  (fig. 10.1); nel caso della torta possiamo misurare la somma di tutta l'energia luminosa generata dalle candeline lungo delle linee che indicheremo con  $\ell$ .

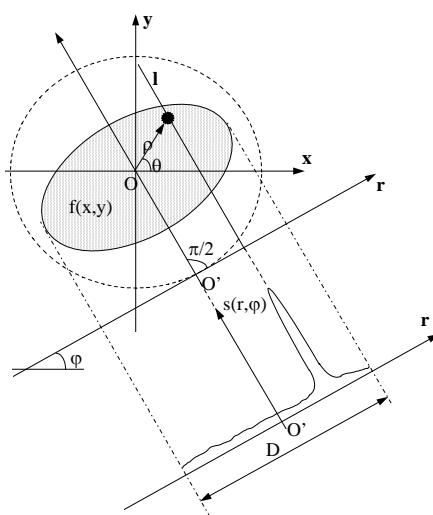
Indicheremo con  $\mathbf{r}$  l'asse lungo il quale sono situati i punti di rilevazione e con  $r$  le sue ascisse; l'asse è ortogonale alle linee  $\ell$  ed è quindi

## CAPITOLO 10 LA TOMOGRAFIA

inclinato di  $\varphi$  rispetto all'asse  $x$ . L'origine delle coordinate dell'asse  $\mathbf{r}$  è posta nel punto  $O'$  proiezione sulla retta  $\mathbf{r}$  dell'origine delle coordinate del piano  $x, y$ . Una proiezione è quindi data dalla funzione

$$s(r, \varphi) = \int_{-\infty}^{\infty} f(\xi, \eta) d\ell$$

Vedremo ora come fare il processo inverso e cioè passare dalle proiezioni  $s(r, \varphi)$ ;  $0 < \varphi < \pi$  all'immagine originale. Questa operazione di **inversione** è indicata con il nome di **tomografia** da  $\tau\acute{\epsilon}\mu\nu\omega = \textit{taglio}$  in greco. Per comprendere il meccanismo dell'inversione tomografica, si rammenti che il problema è lineare e si può procedere per sovrapposizione degli effetti o utilizzare il dominio dei numeri d'onda.



$D$  = dimensione massima dell'oggetto

Figura 10.1: Ripresa di proiezioni dall'esterno di un corpo

## 10.2 Trattamento nei numeri d'onda

Utilizzeremo il teorema detto delle sezioni di proiezione (*projection slices*) della trasformata di Fourier 2D e cioè utilizzeremo il fatto che la trasfor-

## 10.2 TRATTAMENTO NEI NUMERI D'ONDA

mata di Fourier di una proiezione (ad assegnato angolo  $\varphi$  con l'asse  $x$ ), è la *sezione* della trasformata di Fourier bidimensionale  $F(u, v)$  della  $f(x, y)$  che ha eguale angolo  $\varphi$  con l'asse  $u$ . Determiniamo analiticamente l'espressione di una proiezione  $s(r, \varphi)$  e della sua trasformata di Fourier:

$$s(r, \varphi) = \int_{-\infty}^{\infty} f(\xi, \eta) d\ell \quad (10.1)$$

dove  $\ell$  è la retta di coordinate:

$$\ell \rightarrow \rho \cos(\theta - \varphi) = r \quad (10.2)$$

le variabili  $\rho, \theta$  sono le coordinate polari nel piano  $x, y$ :

$$\rho = \sqrt{x^2 + y^2} \quad \theta = \tan^{-1} \frac{y}{x} \quad (10.3)$$

In coordinate cartesiane si avrebbe:

$$\ell \rightarrow x \cos \varphi + y \sin \varphi = r$$

Nel caso di  $\varphi = 0$  l'espressione di una proiezione si semplifica in:

$$s(r, 0) = \int_{-\infty}^{\infty} f(r, y) dy \quad -\infty \leq r \leq \infty \quad (10.4)$$

e la variabile  $r$  coincide con la variabile  $x$ . Se la trasformata di Fourier dell'immagine è:

$$F(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) e^{-j(ux+vy)} dx dy$$

la trasformata di Fourier della proiezione è, indicando con  $\Omega$  la variabile coniugata alla variabile  $r$ :

$$\int_{-\infty}^{\infty} s(r, 0) e^{-j\Omega r} dr = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(r, y) e^{-j\Omega r} dr dy = F(\Omega, v = 0) \quad (10.5)$$

Ruotando gli assi del piano  $x, y$  di un angolo  $\varphi$  ruotano di altrettanto gli assi coordinati della trasformata  $F(u, v)$ . Quindi la trasformata di Fourier di una proiezione  $s(r, \varphi)$  è  $F(\Omega, \varphi)$  se introduciamo le coordinate polari nel dominio dei numeri d'onda  $u, v$ :

$$\Omega = \sqrt{u^2 + v^2}; \quad \varphi = \tan^{-1} \frac{v}{u} \quad (10.6)$$

## CAPITOLO 10 LA TOMOGRAFIA

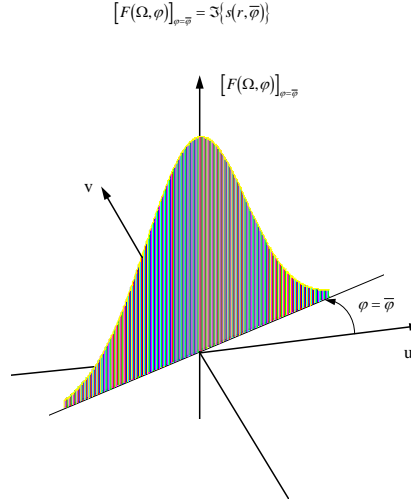


Figura 10.2: Una sezione di una trasformata di Fourier 2D, trasformata di una proiezione

Si osservi che ora  $\Omega \geq 0$ , e una sezione completa della trasformata di Fourier viene ottenuta considerando i valori della  $F(\Omega, \varphi)$  sia per l'angolo  $0 \leq \varphi \leq \pi$  che per l'angolo  $\varphi + \pi$ . In conclusione abbiamo mostrato come la trasformata di Fourier 1D della proiezione  $s(r, \varphi)$  fatta lungo le rette  $\ell$  che fanno un angolo  $\varphi + \pi/2$  con l'asse  $x$  è eguale alla sezione della trasformata di Fourier 2D  $F(u, v)$  dell'immagine  $f(x, y)$  lungo la direzione che fa angolo  $\varphi$  con l'asse  $u$  e cioè  $F(\Omega, \varphi)$ ,  $F(\Omega, \varphi + \pi)$  (fig. 10.2).

$$s(r, \varphi) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\Omega, \varphi) e^{j\Omega r} d\Omega \quad (10.7)$$

I valori di  $\Omega$  negativi sono riferiti all'angolo  $\varphi + \pi$ .

La cosa diventa ovvia se si pensa che una proiezione è ottenuta convolvendo l'immagine originale per una lama di altezza costante, parallela all'asse della proiezione stessa; convolvere con una lama corrisponde a moltiplicare la trasformata di Fourier 2D per la lama ortogonale, trasformata della lama negli spazi. Pertanto, se sono note  $M$  proiezioni disposte regolarmente lungo un arco di  $180^\circ$  sono note  $M$  sezioni della trasformata

## 10.2 TRATTAMENTO NEI NUMERI D'ONDA

di Fourier 2D,

$$F(\Omega, \varphi_m); m = 1, \dots, M; \quad F(\Omega, \varphi_m + \pi); m = 1, \dots, M. \quad |\Omega| \geq 0;$$

Da queste sezioni che suddividono uniformemente l'angolo giro, con una interpolazione fatta nel dominio dei numeri d'onda, è possibile trovare la  $F(u, v)$  su di un reticolo cartesiano regolare; con una antitrasformazione discreta si riottiene la  $f(x, y)$  (fig. 10.3). Si può ora calcolare il numero di proiezioni necessario per fare una corretta interpolazione nel dominio dei numeri d'onda. I campioni della  $F(\Omega, \varphi_m)$  sono massimamente distanti

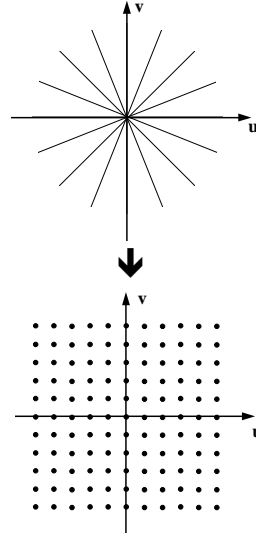


Figura 10.3: Passaggio da coordinate polari a cartesiane

tra loro per i valori massimi di  $|\Omega|$ ; ora se indichiamo con  $L$  la larghezza del lobo principale di un filtro che delimita la risoluzione spaziale dell'immagine, si ha

$$|\Omega| \leq \frac{\pi}{L} \quad (10.8)$$

Se l'oggetto ha dimensione massima  $D$ , i campioni nel dominio spettrale devono essere distanti meno di  $2\pi/D$  perchè le repliche dovute al campionamento nei numeri d'onda non si sovrappongano negli spazi; quindi



## CAPITOLO 10 LA TOMOGRAFIA

(fig. 10.4):

$$\frac{\pi}{L} \times \frac{\pi}{M} \leq \frac{2\pi}{D} \rightarrow M \geq \frac{\pi D}{2L} \quad (10.9)$$

Dall'analisi fatta nel dominio dei numeri d'onda si conclude che è possibile

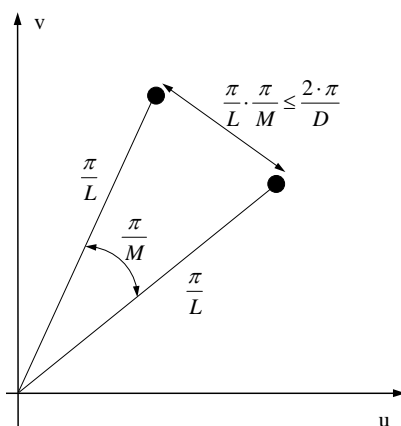


Figura 10.4: Condizione di corretto campionamento nel dominio dei numeri d'onda

effettuare una inversione tomografica:

- se la risoluzione spaziale dell'immagine è stata limitata e lo spettro dell'immagine è sostanzialmente nullo oltre il numero d'onda  $|\Omega| = \pi/L$ ;
- • se il numero delle proiezioni  $M$  è sufficiente e cioè  $M \geq 1,57D/L$  dove  $D$  è il diametro dell'oggetto da tomografare.

Ripeteremo ora l'analisi rimanendo nel dominio degli spazi, per comprendere i meccanismi che dominano la tomografia e così avere gli strumenti concettuali per estendere la trattazione a situazioni di geometria irregolare. Per ora limitiamoci ad osservare che servono apparentemente più equazioni di quante siano le incognite in quanto con una risoluzione  $L$  ed un diametro  $D$  l'oggetto è descritto da  $\simeq \pi(D/2L)^2$  valori diversi di ampiezza (la risoluzione spaziale corrispondente alle differenti luminosità di altrettante *candeline*). Le equazioni disponibili sono invece almeno  $M \times D/L$  e cioè il doppio. Ciò implica che queste equazioni sono linearmente dipendenti tra loro;  $M$  relazioni sono trovate subito osservando che l'"area" di tutte le proiezioni (l'energia luminosa emessa da tutte le candeline) è costante e pari al "volume" racchiuso dalla funzione  $f(x, y)$ .

Questo equivale a osservare che "la componente a frequenza zero" di tutte le proiezioni è costante e pari a  $F(0,0)$ . Similmente, si intuisce che i campioni a bassa frequenza spaziale di tutte le proiezioni sono sì funzione di  $\varphi$ , ma dipendono dai pochi valori indipendenti assunti dalla  $F(u,v)$  attorno all'origine delle coordinate, ecc.

## 10.3 Il sinogramma

Per comprendere il meccanismo della tomografia nel dominio degli spazi, faremo uso del principio di sovrapposizione degli effetti e quindi supporremo che la funzione  $f(x,y)$  corrisponda ad un singolo impulso unitario (*la torta abbia una sola candelina*). Ad esempio, se l'impulso (la luce emessa dalla candelina) occupa l'origine degli assi  $O$ , le proiezioni misurate dall'osservatore avranno un picco all'ascissa  $O'$ , proiezione sull'asse  $\mathbf{r}$  del punto  $O$ , qualsiasi sia il valore di  $\varphi$ . Si riportino tali proiezioni in un'immagine e cioè un grafico tridimensionale (detto **sinogramma** per i motivi che vedremo tra poco) che ha per ascissa  $r$  e per ordinata  $\varphi$ ; lungo la terza dimensione (il livello di grigio dell'immagine), non riportata nella figura 10.5, dovrebbe comparire il valore dell'integrale di linea della funzione  $f(x,y)$  fatto integrando lungo la linea  $\ell$  come si vede nella figura 19.1. Il risultato che si ottiene nel caso di una *candelina* nell'origine è un'immagine con una linea verticale in corrispondenza dell'ascissa 0 (la retta  $r = 0$  nella fig. 10.5) il cui livello di grigio è proporzionale alla luminosità della *candelina*. Se invece sono presenti tre *candeline*, nel sinogramma compaiono, oltre la traccia verticale vista prima, altre due tracce di forma sinusoidale e di livello di grigio proporzionale alla luminosità delle altre due *candeline* (nel caso di fig. 10.6, una delle tre sorgenti (quella in B) ha segno negativo).

Supponiamo ora che le coordinate polari dell'unico impulso che forma la  $f(x,y)$  siano  $\rho_0$  e  $\theta_0$ . In ogni riga del sinogramma e cioè per ogni valore di  $\varphi$  l'unico valore di  $r$  per cui il sinogramma ha valore diverso da 0 oscilla tra  $+\rho_0, -\rho_0$  con una traiettoria sinusoidale (ecco il perchè del nome sinogramma):

$$r = \rho_0 \cos(\theta_0 - \varphi) \quad (10.10)$$

Naturalmente, se invece di un impulso (*una candelina sulla torta*) ve ne sono molti, ad esempio in A, B, C, nella riga del sinogramma che corrisponde alla proiezione di angolo  $\varphi$  sarà registrato l'andamento relativo

## CAPITOLO 10 LA TOMOGRAFIA

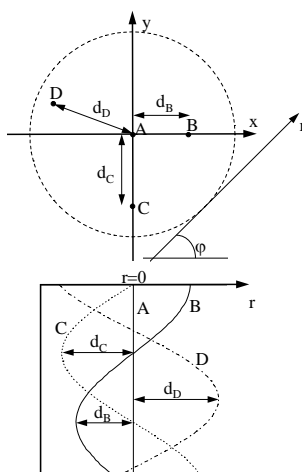


Figura 10.5: Sinogramma

sinogramma non filtrato

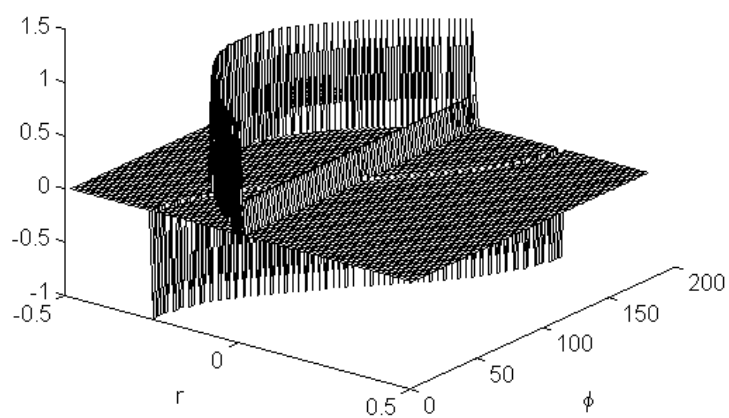


Figura 10.6: Sinogramma di 3 impulsi.

alla somma delle energie luminose di tutte le candeline posizionate ai punti di coordinate polari  $\rho, \theta$  legate dalla relazione:

$$\rho \cos(\theta - \varphi) = r \quad (10.11)$$

Risulta quindi che l'ampiezza del sinogramma è:

$$s(r, \varphi) = \int_{-\infty}^{\infty} f(x, y) d\ell \quad (10.12)$$

e cioè l'integrale della  $f(x, y)$  (rappresentata in coordinate polari come  $f(\rho, \theta)$ ) lungo le linee  $\ell$ .

## 10.4 La retroproiezione

Se si desidera ora passare dal sinogramma  $s(r, \varphi)$  di nuovo all'immagine in coordinate polari  $f(\rho, \theta)$ , si deve effettuare l'operazione di **retroproiezione**. Un modo elementare di vedere il meccanismo di retroproiezione, direttamente corrispondente al nome, è il seguente. Si visualizzi l'immagine  $f(x, y)$  come una distribuzione disuniforme di burro su di una fetta di pane; la proiezione lungo una direzione  $\varphi + \pi/2$  dà l'informazione, ascissa per ascissa, di *quanto burro* sia disposto lungo quella direzione. La ricostruzione più ragionevole dell'immagine che si può fare, non disponendo di ulteriore informazione ma garantendo la congruenza con la proiezione, corrisponde all'operazione di *spalmare il burro sulla fetta di pane*. In altre parole, le ampiezze del segnale vengono prima sommate tra loro lungo la direzione  $\ell$  (formazione della proiezione) e poi ridistribuite sull'immagine (*spalmate*) ancora lungo la stessa direzione (la retroproiezione); si passa così dall'immagine originale  $f(x, y)$  ad una funzione cilindrica uniforme nella direzione  $\varphi + \pi/2$ , pur mantenendo eguale proiezione  $s(r, \varphi)$ . Un impulso, riproiettato, diventa una lama, passante per l'impulso originario, ortogonale alla direzione di proiezione. Completata l'operazione di retroproiezione relativa ad un direzione di proiezione (un azimut), si devono adesso utilizzare tutti gli altri azimut.

L'operazione di retroproiezione viene quindi completata mediando tra loro tutte le retroproiezioni effettuate lungo i vari azimut; ciò equivale a stimare la distribuzione originale del burro mediando tra loro *le quantità di burro rispalmate dalle varie direzioni*. Le stesse operazioni di

## CAPITOLO 10 LA TOMOGRAFIA

retroproiezione dai vari azimut possono essere ottenute in modo equivalente operando come segue. Tutte le informazioni relative allo stesso punto dell'immagine posizionato nel punto  $P$  di coordinate polari  $(\rho_0, \theta_0)$ , provenienti dalle proiezioni della  $f(x, y)$  lungo i vari azimut, sono disposte nel sinogramma lungo la traiettoria sinusoidale descritta dall'equazione (10.10). Mediando i valori del sinogramma lungo questa traiettoria, si trova una stima del valore della  $f(x, y)$ , che indichiamo con  $\hat{f}(x, y)$ . Questa operazione di media coincide con la retroproiezione perchè si attribuisce alla  $f(\rho, \theta)$  il valore medio delle proiezioni fatte sui punti di ascissa  $r = \rho \cos(\theta - \varphi)$  al variare dell'angolo  $\varphi$ . Mediando su tutti gli azimut, la retroproiezione è quindi definibile come:

$$\hat{f}(\rho, \theta) = \frac{1}{2} \times \frac{1}{\pi} \int_0^\pi s[r = \rho \cos(\theta - \varphi), \varphi] d\varphi \quad (10.13)$$

formula che vedremo sarà utile nella derivazione formale della retroproiezione. La divisione per 2 è una conseguenza del sistema di riferimento scelto (ovvero  $0 \leq \varphi < \pi$  mentre  $\rho$  può essere sia positivo che negativo e quindi la retroproiezione è equivalente, in pratica, ad effettuare l'operazione di media su  $2\pi$ ). Vediamo cosa accade con  $M$  proiezioni se la  $f(x, y)$  è costituita da un singolo impulso ad esempio nell'origine  $O$ . Ogni retroproiezione è una lama di ampiezza costante che passa per  $O$ ; la  $\hat{f}(x, y)$  è ottenuta dalla somma di  $M$  lame che passano tutte per  $O$ . Anche per  $M \rightarrow \infty$  la  $\hat{f}(x, y)$  non è l'impulso di partenza  $\delta(x, y)$  ma la funzione:

$$q(x, y) = \frac{1}{\sqrt{x^2 + y^2}} = \frac{1}{\rho} \quad (10.14)$$

la cui trasformata di Fourier 2D è:

$$Q(\Omega) = \frac{1}{\sqrt{u^2 + v^2}} = \frac{1}{|\Omega|} \quad (10.15)$$

Infatti, basta pensare che le lame siano  $M$  e che ogni lama abbia uno spessore  $\delta$ ; il volume totale della media delle lame in una corona circolare di raggio  $\rho$  e larghezza  $\Delta$  è costante e pari a  $M\delta\Delta/M$  (tale volume deve essere normalizzato rispetto al numero delle proiezioni da cui segue la divisione per  $M$ ), quindi l'altezza equivalente in grado di fornire il medesimo volume delle  $M$  retroproiezioni è  $M\delta/2\pi M\rho$  e decresce con  $\rho$ . Ne consegue che la retroproiezione ottenuta ridistribuendo il valore della

## 10.5 LA RETROPROIEZIONE FILTRATA

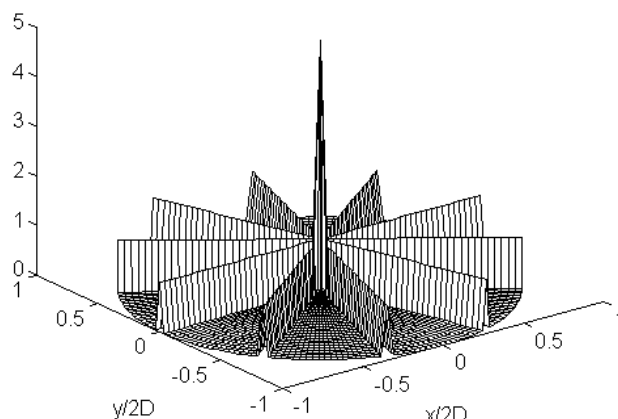


Figura 10.7: Retroproiezione di 5 proiezioni di un impulso nell'origine.

proiezione su tutti i campioni dell'immagine NON restituisce l'immagine di partenza bensì l'immagine convoluta con un filtro la cui risposta all'impulso è  $q(x, y) = 1/\rho$ . Per ottenere di nuovo un impulso "ideale" fino alla frequenza di Nyquist, si deve compensare (deconvolvere) la risposta all'impulso  $q(x, y)$  e cioè moltiplicare (nelle frequenze) per  $|\Omega|$ , ovviamente ancora fino alla frequenza di Nyquist. Il motivo è che ad ogni lama nel dominio degli spazi (ad esempio  $\delta(x)$ ) corrisponde una lama nel dominio dei numeri d'onda (in questo caso  $\delta(v)$ ), e le due funzioni dell'equazione precedente sono entrambe somma di infinite lame che si incrociano nell'origine. Se il numero di proiezioni è limitato, si comprende come la risposta all'impulso del sistema di retroproiezione non sia un impulso ma piuttosto una stella di  $M$  raggi centrata là dove era situato l'impulso originario (fig.10.7, 10.8).

## 10.5 La retroproiezione filtrata

Abbiamo visto come in ogni punto del sinogramma si sommano i contributi relativi ai molti impulsi che compongono l'immagine; facendo la media del sinogramma lungo tutta la traiettoria relativa al punto di interesse  $P_0$  di coordinate polari  $(\rho_0, \theta_0)$ , i contributi degli altri impulsi **non**

## CAPITOLO 10 LA TOMOGRAFIA

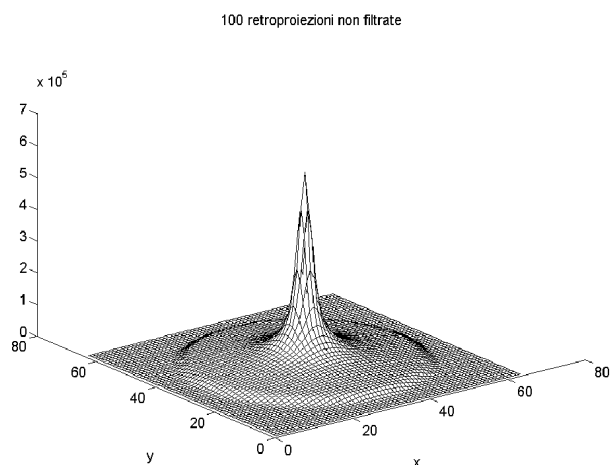


Figura 10.8: Retroproiezione di 100 proiezioni di un impulso.

si elidono. Se però deconvolviamo per la  $q(x, y)$  (prima di sommare lungo le traiettorie sinusoidali nel sinogramma) e quindi convolviamo ogni proiezione con una opportuna funzione  $\tilde{\rho}(r)$  la cui trasformata 1D vedremo essere  $|\Omega|$ , è possibile eliminare l'effetto dell'interferenza dei vari impulsi tra di loro (fig. 10.9). Supponiamo ad esempio che  $\tilde{\rho}(0)$  abbia valore unitario; supponiamo inoltre che  $\tilde{\rho}(r)$  abbia un lobo principale di larghezza circa  $L$  e lobi laterali negativi e positivi di larghezza  $L$  e di ampiezza decrescente in modo che il valor medio della  $\tilde{\rho}(r)$  sia nullo (una specie di sinc). Se la funzione  $f(x, y)$  è costituita da un solo impulso (ad

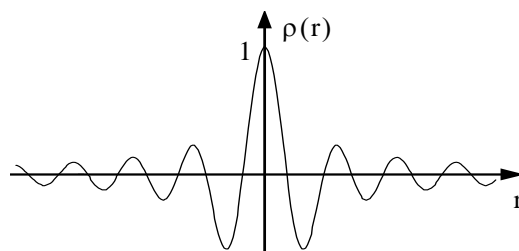


Figura 10.9: Il filtro per le retroproiezioni  $\tilde{\rho}(r)$

## 10.5 LA RETROPROIEZIONE FILTRATA

esempio nell'origine), i valori del sinogramma lungo la traiettoria relativa alla posizione dell'impulso (la retta  $r = 0$ ) non sono modificati dalla convoluzione con  $\tilde{\rho}(r)$ . Vediamo ora che accade se valutiamo l'ampiezza dell'immagine in un altro punto, distante  $\rho_1$  dall'origine ( $\rho_1 \gg L$ ). Per fare ciò dobbiamo sommare i valori assunti dal sinogramma lungo la traiettoria sinusoidale relativa a questo secondo impulso che traversa la retta  $r = 0$  secondo una linea obliqua. Nella somma, si incontrano e si mediano successivamente tutti i lobi positivi e negativi della  $\tilde{\rho}(r)$ , che hanno per ipotesi somma nulla. Se ne conclude che, filtrando le proiezioni con la  $\tilde{\rho}(r)$ , l'interferenza tra impulsi abbastanza lontani viene soppressa (fig. 10.10, 10.11, 10.12, 10.13); ciò conferma che la moltiplicazione per  $|\Omega|$  effettua la deconvoluzione. Non è così se i due impulsi sono più vicini di  $L$ . Allora, sommando lungo la seconda traiettoria, non si incontrano tutti i lobi sia positivi che negativi della funzione  $\tilde{\rho}(r)$ , ma ci si trattiene sempre nel lobo principale e perciò vi sarà interferenza. Da quanto visto fino ad ora, la risposta all'impulso del sistema tomografico con retroproiezione filtrata è prossima a quella ideale, a parte la limitazione a  $L$  della risoluzione spaziale, corrispondente alla limitazione della banda alla frequenza di Nyquist. Le lame che formano la stella nel caso della retroproiezione non filtrata hanno ora lobi positivi e negativi che tendono ad elidersi tra loro; se vi sono molte proiezioni, l'immagine ricostruita ha pertanto un lobo centrale isolato largo circa  $L$ . Vedremo ora l'effetto della limitazione del numero delle proiezioni e cioè l'equivocazione spaziale. Da quanto osservato finora si dedurrebbe che conviene fare  $L$  il più piccolo possibile; vedremo nella prossima sezione perchè questo non è vero.

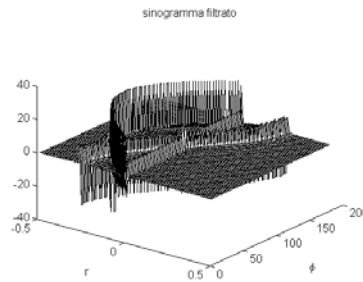


Figura 10.10: Sinogramma filtrato di 3 impulsi (Fig. 10.5).



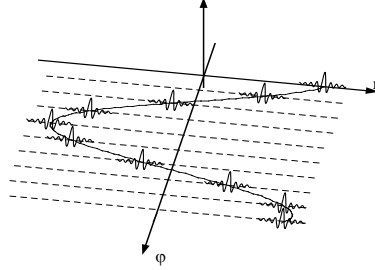


Figura 10.11: Il sinogramma filtrato

### 10.5.1 Equivocazione angolare

Se il campionamento angolare è insufficiente, vedremo ora che vi è interferenza tra gli impulsi se questi sono lontani tra loro: avremo cioè un effetto di equivocazione angolare se la differenza angolare tra due proiezioni successive  $\Delta\varphi$  è

$$\Delta\varphi > 2L/D \quad (10.16)$$

dove  $D$  è la massima dimensione dell'oggetto osservato. Supponiamo infatti di avere due impulsi distanti tra loro  $D$ ; a partire dall'azimut  $\varphi_0$  per cui i due impulsi sono sovrapposti nella proiezione, le loro due proiezioni si distanziano al crescere di  $\varphi$  di

$$D \sin(\varphi - \varphi_0) \simeq D(\varphi - \varphi_0)$$

Se ora  $\varphi_n = n\Delta\varphi$ , passando da una proiezione alla successiva, la  $\tilde{\rho}(r)$  relativa al primo impulso viene campionata dalla traiettoria relativa al secondo impulso per valori dell'ascissa pari a

$$n\Delta x = nD\Delta\varphi$$

Ma se la (10.16) è vera,  $\Delta x > 2L$ , si saltano addirittura lobi della  $\tilde{\rho}(r)$ , la media non sarà più a valore nullo e quindi appare l'interferenza. In conclusione, si verifica che la risposta all'impulso dell'operazione di retro-proiezione con campionamento angolare di intervallo  $\Delta\varphi$  sarà caratterizzata da un lobo centrale di diametro  $L$  intorno a  $P_0$ , da una zona oscura di diametro  $2L/\Delta\varphi$  dove i lobi positivi e negativi delle varie lame si elidono, e infine a distanza ancora maggiore da  $P_0$ , da una stella di tanti raggi quante sono le proiezioni (fig.10.12).

## 10.5 LA RETROPROIEZIONE FILTRATA

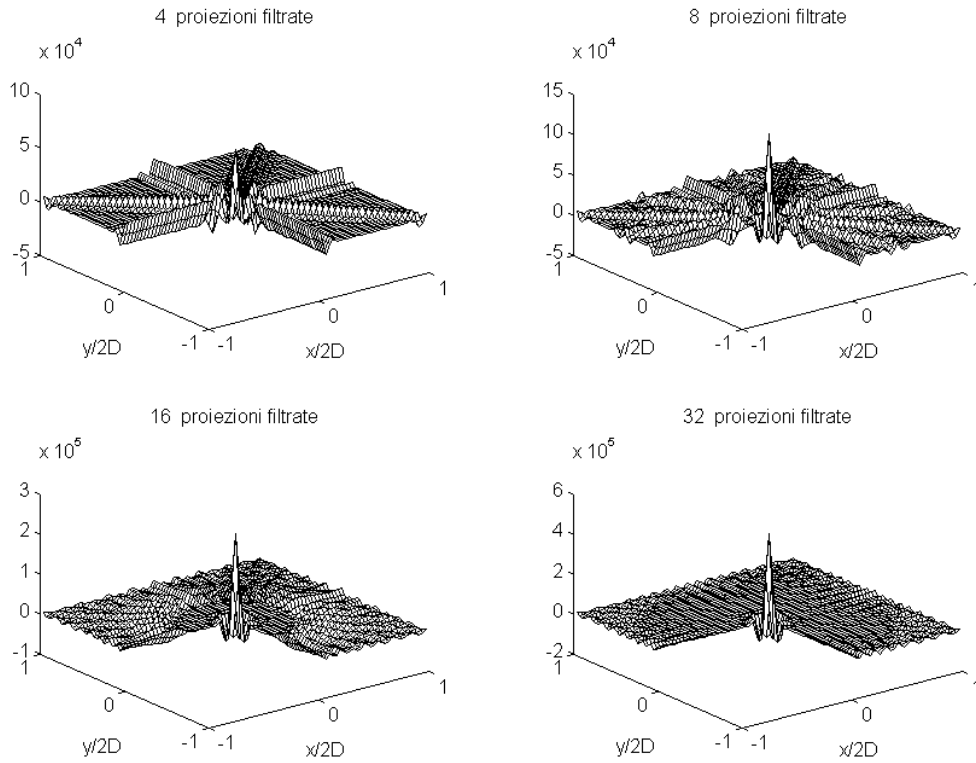


Figura 10.12: Ricostruzione dell'impulso da 4,8,16,32 proiezioni filtrate (particolare della ricostruzione).

### 10.5.2 Retroproiezioni e numeri d'onda

Possiamo ora interpretare l'operazione di retroproiezione avvalendoci del teorema delle *projection slices* e cioè delle sezioni della trasformata di Fourier 2D. Fare la retroproiezione, equivale a comporre l'immagine con tante funzioni cilindriche (le singole proiezioni retroproiettate) disposte lungo i vari azimut di proiezione. Lo spettro 2D dell'immagine così ricostruita è la media di tante lame, di ampiezza proporzionale alla corrispondente sezione della trasformata di Fourier 2D. Finestrando negli spazi, si smussa nel dominio dei numeri d'onda e quindi si ridistribuisce lo spettro da una distribuzione a lame ad un'altra più smussata che cor-

## CAPITOLO 10 LA TOMOGRAFIA

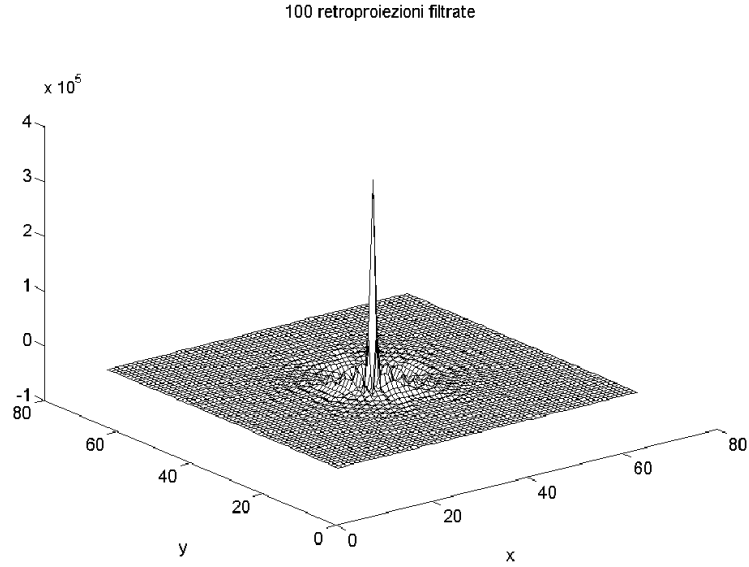


Figura 10.13: Ricostruzione da 100 retroproiezioni filtrate (confrontare con Fig.10.8).

risponde maggiormente al vero spettro 2D. Tuttavia, si deve evitare l'equivocazione e da ciò derivano le stesse condizioni già viste in precedenza sul numero delle proiezioni. Inoltre, le sezioni si affollano al centro delle coordinate del piano  $u, v$  e sono rade alla periferia; è necessario amplificare il risultato in proporzione alla distanza dal centro delle coordinate e quindi alla pulsazione spaziale  $|\Omega|$ . Una osservazione equivalente era già stata fatta prima, quando avevamo visto che la stella di lame corrispondenti alla retroproiezione non filtrata ha spettro  $1/\Omega$ .

Arriveremo ora a queste stesse conclusioni in modo più formale. Definiamo la retroproiezione come:

$$\hat{f}(\rho, \theta) = \frac{1}{2\pi} \int_0^\pi s[r = \rho \cos(\theta - \varphi), \varphi] d\varphi \quad (10.17)$$

Sostituendo l'espressione della proiezione in funzione della antitrasformata di Fourier dell'immagine si ha:

$$\hat{f}(\rho, \theta) = \frac{1}{2\pi} \int_0^\pi \left( \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\Omega, \varphi) e^{j\Omega \rho \cos(\theta - \varphi)} d\Omega \right) d\varphi = \quad (10.18)$$

## 10.5 LA RETROPROIEZIONE FILTRATA

$$= \frac{1}{2\pi} \int_{-\pi}^{\pi} \left( \frac{1}{2\pi} \int_0^{\infty} F(\Omega, \varphi) e^{j\Omega\rho \cos(\theta-\varphi)} d\Omega \right) d\varphi$$

Ma questa è esattamente l'espressione dell'antitrasformata di Fourier 2D **tranne il fattore**  $|\Omega|$ , corrispondente allo Jacobiano della trasformazione delle coordinate da cartesiane a polari:

$$\hat{f}(\rho, \theta) = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_0^{\infty} \frac{F(\Omega, \varphi)}{|\Omega|} e^{j\Omega[\rho \cos(\theta-\varphi)]} \Omega d\Omega d\varphi \quad (10.19)$$

La retroproiezione non filtrata è l'antitrasformata della  $F(\Omega, \varphi)/|\Omega|$ ; la moltiplicazione per la funzione  $1/|\Omega|$  corrisponde alla convoluzione bidimensionale per la sua antitrasformata  $1/\rho$  che ha **volume infinito**; infatti  $1/|\Omega| = \infty$ , nell'origine. In altre parole, ricostruendo l'immagine con le retroproiezioni senza filtraggio con i cosiddetti  $\rho$ -filtri, e cioè senza la convoluzione con la  $\tilde{\rho}(r)$ , otteniamo un'immagine con un'eccessiva riduzione di ampiezza alle frequenze spaziali più elevate. Per ottenere un'immagine corretta si deve pertanto moltiplicare la trasformata della proiezione per la funzione  $|\Omega|$ . Inoltre, il numero d'onda massimo contenuto nelle proiezioni deve essere inferiore a  $\pi/L$ ; si osservi infine che invece di filtrare 2D l'immagine con il filtro  $1/|\Omega|$ , basta filtrare le singole proiezioni con un filtro passa alto la cui funzione di trasferimento è (asse delle pulsazioni bilatero):

$$\tilde{P}(\Omega) = |\Omega| = -j\Omega \times j\text{sign}(\Omega); \quad 0 \leq |\Omega| \leq \pi/L; \quad (10.20)$$

per ottenere il risultato corretto (fig. 10.14). Computazionalmente ciò è molto meno oneroso. Quindi, prima di fare la retroproiezione, nel caso continuo ed a risoluzione spaziale limitata, si dovrebbe fare la derivata delle proiezioni e poi fare seguire alla derivazione una trasformazione di Hilbert; il risultato della retroproiezione sarebbe quello desiderato. Nel dominio degli spazi, il  $\rho$ -filtro è descritto dalla seguente espressione:

$$\tilde{\rho}(r) = \frac{\pi}{L^2} \left[ \text{sinc}\left(\frac{r}{L}\right) - \frac{1}{2} \text{sinc}^2\left(\frac{r}{2L}\right) \right]; \quad (10.21)$$

Nel caso discreto, si osservi che se il numero delle proiezioni è pari a  $M$ , il  $\rho$ -filtro  $\tilde{P}(\Omega)$  deve avere un guadagno non nullo a frequenza zero e pari a  $1/M$ .

## CAPITOLO 10 LA TOMOGRAFIA

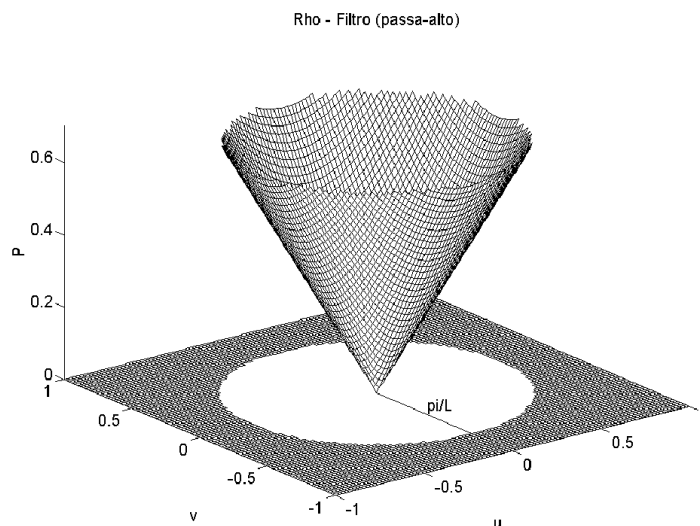


Figura 10.14: Filtro 2D passa alto equivalente al filtraggio delle retroproiezioni.

La tecnica delle retroproiezioni filtrate è computazionalmente più conveniente di quella dell'interpolazione su di un reticolo cartesiano della trasformata di Fourier bidimensionale nota su un reticolo polare, in quanto il mezzo di calcolo può essere utilizzato con maggiore uniformità temporale durante le acquisizioni. Nel caso delle retroproiezioni filtrate, è possibile fare la convoluzione con il filtro  $\tilde{P}(\Omega)$  man mano che le proiezioni sono acquisite; si può anche fare immediatamente la retroproiezione della proiezione filtrata. Non appena viene acquisita e trattata l'ultima proiezione, il risultato è disponibile, senza che sia necessario un picco di utilizzo del mezzo di calcolo. Nel caso dell'interpolazione su reticolo cartesiano della trasformata di Fourier bidimensionale, è invece necessario ritardare l'antitrasformazione bidimensionale, e pertanto un forte uso di potenza di calcolo, ma solo al momento in cui sono disponibili tutte le proiezioni.

## 10.5 LA RETROPROIEZIONE FILTRATA

5 retroproiezioni filtrate nel dominio  $u,v$

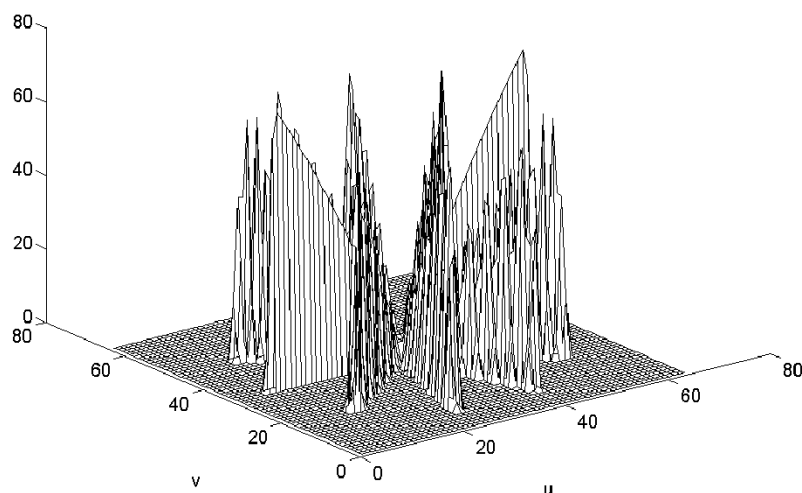


Figura 10.15: 5 retroproiezioni filtrate viste nel dominio dei numeri d'onda

### 10.5.3 Proiezioni incomplete o mosse

E' facile determinare cosa accade se non tutte le proiezioni vengono misurate, se cioè per qualche motivo mancano le *sezioni* dello spettro 2D del dato appartenenti ad un settore angolare  $\Delta\varphi$ . Il risultato è che il segnale viene filtrato da un filtro a ventaglio di larghezza  $\Delta\varphi$  che sopprime le proiezioni appartenenti all'angolo non misurato. Ancora, può essere utile calcolare l'effetto di un movimento del soggetto ripreso durante la tomografia; supponiamo che durante la prima metà delle proiezioni il soggetto si trovi in  $O$  e durante la seconda metà delle proiezioni si trovi in  $O'$ . La risposta è di nuovo semplice; basta pensare di filtrare con il filtro a ventaglio corrispondente alle prime proiezioni l'impulso posizionato in  $O$  e poi con il filtro a ventaglio complementare l'impulso spostato nella seconda posizione  $O'$ . Il risultato sarà la combinazione delle due risposte all'impulso. Se p.e. le proiezioni relative ad una posizione sono corrispondenti ad un'apertura angolare di  $\pi/2$  usando un sistema di coordinate

## CAPITOLO 10 LA TOMOGRAFIA

conveniente, i due filtri a ventaglio sono rappresentati dall'espressione:

$$2 \times H_{1,2}(u, v) = [1 \pm \text{sign}(u)\text{sign}(v)] \leftrightarrow \delta(0, 0) \pm \frac{1}{\pi^2 xy} \quad (10.22)$$

Se le posizioni dei due impulsi sono corrispondenti a:

$$P_{1,2}(u, v) = e^{\pm j(ux_0 + vy_0)} \quad (10.23)$$

il risultato è:

$$H_1(u, v)P_1(u, v) + H_2(u, v)P_2(u, v) \leftrightarrow \quad (10.24)$$

$$\begin{aligned} & \delta(x - x_0, y - y_0) + \frac{1}{\pi^2(x - x_0)(y - y_0)} \\ & + \delta(x + x_0, y + y_0) - \frac{1}{\pi^2(x + x_0)(y + y_0)} \end{aligned}$$

Nel caso di segnale campionato e periodicizzato, le singolarità sono evitate; si utilizzi l'espressione della trasformazione di Hilbert riportata nel Capitolo 7.

### 10.6 Tomografia 3D

È interessante analizzare cosa accade estendendo a tre dimensioni il concetto di tomografia, discussa in 2 dimensioni; infatti, si è supposto che la propagazione avvenga nel piano (*della torta*)  $(x, y)$  detto assiale (infatti TAC: tomografia assiale computerizzata) e cioè nel piano ortogonale alla direzione  $z$ .

Anzitutto vediamo come nella trasformazione di Fourier a 3D, una retta è coniugata al piano che ha la giacitura ortogonale. In altre parole, una singolarità *a bacchetta* si trasforma in una singolarità *a lamina*; infatti il comportamento è uniforme nella direzione della bacchetta ed impulsivo nelle due direzioni ad essa ortogonali; correlativamente, nel dominio coniugato, abbiamo un comportamento impulsivo andando nella direzione della bacchetta, e troviamo una costante andando nelle due direzioni ortogonali. Indicando con  $u, v, w$  le variabili coniugate alle variabili  $x, y, z$ , la trasformazione di Fourier in 3D è:

$$F(u, v, w) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y, z) e^{-j(ux+vy+wz)} dx dy dz$$

$$f(x, y, z) = \frac{1}{(2\pi)^3} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} F(u, v, w) e^{j(ux+vy+wz)} du dv dw$$

Se:

$$f(x, y, z) = \delta(x)\delta(y) \rightarrow F(u, v, w) = \frac{1}{(2\pi)^3} \delta(w)$$

e se

$$f(x, y, z) = \delta(z) \rightarrow F(u, v, w) = \frac{1}{(2\pi)^3} \delta(u)\delta(v)$$

Quindi, convolvendo con *una bacchetta* si moltiplica lo spettro per una lamina; ancora, convolvendo con una lamina, si moltiplica lo spettro per una bacchetta. Tutte le osservazioni fatte per la tomografia 2D si estendono facilmente a 3D e si possono derivare delle relazioni che fissano il numero totale di proiezioni in funzione della risoluzione spaziale desiderata. Per esempio, nel caso di proiezioni lungo delle lamine (convolvendo la funzione  $f(x, y, z)$  con delle lamine e cioè mediando i valori della funzione su interi piani) si ha che un volume di dati si proietta in una funzione 1D; inoltre, se  $D$  è il diametro dell'oggetto da tomografare (supposto sferico) e  $\Delta\varphi$  l'angolo tra le normali a due lamine adiacenti,

$$|\Omega|_{\max} \Delta\varphi \leq \frac{2\pi}{D}; \quad |\Omega|_{\max} = \frac{\pi}{L} \rightarrow \Delta\varphi \leq \frac{2L}{D}$$

Infatti, per non avere equivocazione, il campionamento dello spettro, che avviene lungo una stella di rette, deve effettuarsi a intervalli di numeri d'onda inferiori a  $2\pi/D$ ; il numero di proiezioni in una semisfera è pertanto almeno:

$$M_3 = \frac{2\pi |\Omega|_{\max}^2}{(\Delta\varphi)^2 |\Omega|_{\max}^2} \geq \frac{\pi}{2} \frac{D^2}{L^2} \quad (10.25)$$

indicando con  $L$  la risoluzione spaziale (la larghezza del lobo). Poichè ad ogni proiezione corrispondono  $D/L$  equazioni, si hanno in totale  $\pi D^3/2L^3$  equazioni contro  $\pi D^3/6L^3$  incognite con una ridondanza maggiore che non in 2D.

Un altro modo per intuire che la tomografia 3D può essere fatta proiettando il dato di volume lungo piani (convolvendo il volume con delle



## CAPITOLO 10 LA TOMOGRAFIA

lamine) nasce dall'osservazione che nello spazio ci sono  $\infty^3$  piani ed altrettanti punti. Quindi, il numero di proiezioni sui piani ha lo stesso ordine di infinità del numero dei punti dello spazio e conoscere le proiezioni ci porta a conoscere i valori puntuali, a parte gli effetti di indeterminazione considerati prima.

La situazione cambia facendo proiezioni su rette. Di queste, ce ne è  $\infty^4$  e quindi è inutile misurare le proiezioni su "tutte" le rette: avremmo misure ridondanti rispetto ai dati. Facendo la convoluzione con tutte le rette ortogonali ad un dato piano otteniamo una proiezione bidimensionale (l'oggetto "*schacciato*" lungo la direzione della normale al piano). La trasformata di Fourier 2D di questi dati "*schacciati*" ci dà una sezione 2D dello spettro 3D del dato (un cerchio massimo della sfera). Di queste sezioni 2D ce ne sono  $\infty^2$ , ma ce ne bastano  $\infty$ . Per esempio, basta considerare solo le rette ortogonali ad un asse verticale  $z$  (che sono appunto  $\infty^3$ ). Facendo proiezioni secondo queste linee, si può procedere a 3D operando in coordinate cilindriche e quindi ricadere nelle tecniche tomografiche 2D.

Peraltro, la proiezione lungo delle rette è più ricca di informazione e quindi ci permette di recuperare degli ulteriori parametri locali, come per esempio la variabilità angolare della riflettività locale alle onde o in generale parametri di variabilità angolare (azimut o elevazione) delle proprietà locali dell'oggetto sotto misura. Si pensi al caso di specchietti piani sospesi in un mezzo trasparente (come in un acquario). Illuminando tale volume con dei raggi luminosi provenienti da tutte le direzioni, riusciamo a trovare la distribuzione spaziale sia degli specchietti che dei loro puntamenti.

## 10.7 Tecniche iterative

### 10.7.1 Retroproiezione e gradiente

Si consideri un problema di tipo tomografico: sono assegnate delle misure, raccolte in un vettore  $\underline{\mathbf{p}}$ ; si cerca un vettore di parametri  $\underline{\mathbf{m}}$  tale che le misure fatte sul modello definito da questi parametri corrispondano quanto più possibile ai dati. In formula si ha:

$$\underline{\mathbf{A}}\underline{\mathbf{m}} = \underline{\mathbf{d}} \quad (10.26)$$

La matrice  $\underline{\mathbf{A}}$  è in generale di dimensioni  $[M \times N]$ ; i vettori  $\underline{\mathbf{m}}$  e  $\underline{\mathbf{d}}$  sono di dimensione  $[N]$  ed  $[M]$  rispettivamente. Ad esempio, il modello  $\underline{\mathbf{m}}$  potrebbe corrispondere alla distribuzione delle lentezze (inverse delle velocità) di propagazione di un mezzo qualsiasi, e il vettore dei dati  $\underline{\mathbf{d}}$  corrispondere ai tempi di propagazione di raggi che hanno traversato questo mezzo, nelle varie direzioni. Il tempo di propagazione è proporzionale all'integrale di linea delle lentezze lungo la direzione del raggio e quindi si ritrova il classico problema di tomografia. La limitata accessibilità del mezzo fa sì poi che non sono disponibili raggi in tutte le direzioni e pertanto la matrice  $\underline{\mathbf{A}}$  può essere praticamente qualsiasi. La situazione è molto più semplice quando sono disponibili raggi in tutte le direzioni ed allora la struttura della matrice assume forme particolari. La tecnica consueta di risoluzione di questo sistema è quella della pseudoinversione; si cerca una soluzione ai minimi quadrati e cioè tale che l'energia del residuo  $\varepsilon^2$  sia minima:

$$(\underline{\mathbf{A}}\underline{\mathbf{m}} - \underline{\mathbf{d}})^*(\underline{\mathbf{A}}\underline{\mathbf{m}} - \underline{\mathbf{d}}) = \varepsilon^2 \quad (10.27)$$

Come si vedrà più diffusamente nel prossimo capitolo, il minimo di questa forma quadratica si raggiunge per

$$\underline{\mathbf{m}} = (\underline{\mathbf{A}}^*\underline{\mathbf{A}})^{-1}\underline{\mathbf{A}}^*\underline{\mathbf{d}} \quad (10.28)$$

e cioè applicando al vettore  $\underline{\mathbf{d}}$  la matrice  $(\underline{\mathbf{A}}^*\underline{\mathbf{A}})^{-1}\underline{\mathbf{A}}^*$  detta appunto pseudo inversa della matrice  $\underline{\mathbf{A}}$ . Infatti, se la matrice  $\underline{\mathbf{A}}$  è quadrata e non singolare si ritorna al caso classico. Se poi la matrice  $\underline{\mathbf{A}}^*\underline{\mathbf{A}}$  è troppo vicina ad essere singolare, si può sempre aggiungere della luce bianca e cioè accrescere la diagonale principale aggiungendo  $\lambda\mathbf{I}$  dove  $\mathbf{I}$  è la matrice identità e  $\lambda$  un parametro "piccolo". Si verificherà che questo coincide con il minimizzare  $\varepsilon^2 + \lambda(\underline{\mathbf{m}}^*\underline{\mathbf{m}})^2$ . Cerchiamo di capire tutto ciò in termini di tomografia.

### 10.7.2 Retroproiezione e trasposizione

Anzitutto, determiniamo il significato fisico dell'operazione

$$\underline{\mathbf{m}}_0 = \underline{\mathbf{A}}^*\underline{\mathbf{d}} \quad (10.29)$$

e cioè di applicare al vettore dei dati la trasposta della matrice del modello. Questa operazione è la retroproiezione. Per vederlo, si consideri uno

## CAPITOLO 10 LA TOMOGRAFIA

spazio suddiviso in cellette rettangolari. Ogni componente del vettore dei dati (le proiezioni  $\underline{\mathbf{d}}$ ),  $\{d_i\}$ , è ottenuta sommando i contenuti  $\{m_{j_i}\}$  delle cellette corrispondenti ad un'operazione di proiezione. La matrice  $\underline{\mathbf{A}}$  che corrisponde all'operazione di proiezione, ha degli 1 negli elementi  $a_{i,j_i}$  che corrispondono alle celle  $j_i$  che si proiettano nell'elemento  $i$ -esimo del vettore delle proiezioni. La matrice  $\underline{\mathbf{A}}^*$ , che corrisponde all'operazione di retroproiezione, ha gli stessi 1 nelle celle  $a_{j,i}$  relative agli elementi  $i_j$  del vettore delle proiezioni che sono alterati quando viene alterato il contenuto della cella  $m_j$ . Gli elementi della matrice corrispondenti a celle che non si proiettano in quell'elemento della proiezione conterranno degli zeri; se la proiezione è solo parziale si avranno numeri compresi tra 1 e 0. Il valore corretto del modello (il vettore  $\underline{\mathbf{m}}$ ) si ottiene dall'equazione:

$$(\underline{\mathbf{A}}^* \underline{\mathbf{A}}) \underline{\mathbf{m}} = \underline{\mathbf{m}}_0 \quad (10.30)$$

Leggendo questa equazione si vede che, retroproiettando le proiezioni ottenute con il modello  $\underline{\mathbf{m}}$ , si deve ottenere il dato retroproiettato e cioè  $\underline{\mathbf{m}}_0$ . Si abbia un modello  $\underline{\mathbf{m}}$  con una sola cella contenente un valore non nullo e cioè si voglia determinare la risposta all'impulso del sistema; questa risposta sarà spazio - variante tuttavia, e i risultati pertanto potranno cambiare al variare della posizione dell'impulso. Il risultato dell'applicazione dell'equazione (10.28) corrisponderà ad una approssimazione del modello  $\underline{\mathbf{m}}_0$  in cui avremo sí un picco in corrispondenza dell'impulso di ingresso (la cella a contenuto non nullo) ma non solo in quel punto; molti altri elementi del modello saranno non nulli. Nel caso della tomografia queste due operazioni in cascata corrispondono alla proiezione seguita dalla retroproiezione non filtrata. La risposta all'impulso di questo sistema è una struttura a stella, centrata sull'impulso di ingresso e con tanti raggi quante sono le proiezioni.

Quindi, nel caso della tomografia, l'applicazione della matrice  $(\underline{\mathbf{A}}^* \underline{\mathbf{A}})^{-1}$  corrisponde alla focalizzazione di questa stella di raggi. Questa operazione si realizza filtrando le proiezioni con il filtro  $\tilde{P}(\Omega)$  avente una caratteristica di ampiezza proporzionale alla frequenza radiale  $\Omega$ . Si osservi che la semplicità di quest'ultimo risultato è dovuta alla regolarità del rilevamento.

Il caso generale, quando cioè le geometrie non sono controllate e quindi la matrice  $\underline{\mathbf{A}}$  è completamente libera, è ben più complesso. Si osservi poi che nel caso più generale, la matrice  $\underline{\mathbf{A}}$  non è fatta solo di uni e

zeri, ma la retroproiezione su una cella dovrà tener conto della lunghezza (frazionaria) dell'intersezione tra la cella e il raggio di retroproiezione pesando la retroproiezione in funzione di questo dato. Inoltre, potrebbe darsi che la tomografia che stiamo analizzando è una tomografia di tempi di percorso, in cui cioè il modello ricercato è la distribuzione della lentezza di propagazione (inversa della velocità) dei raggi (acustici od elettromagnetici) all'interno del corpo e le misure sono i tempi di percorso, per l'appunto. In questo caso, i percorsi dei raggi dipendono, per la legge di Snell, dalla distribuzione delle velocità, e pertanto il sistema diventa non lineare, con complicazioni ancora maggiori.

### 10.7.3 Interpretazione stocastica della pseudo-inversione

Il problema della determinazione di  $\underline{\mathbf{m}}$  a partire da  $\underline{\mathbf{d}}$  è anche risolvibile con le ben note tecniche di stima. Infatti, si supponga che siano note le  $M$  misure degli  $N$  parametri, questa volta sommate a rumore:

$$\underline{\mathbf{d}} = \underline{\mathbf{A}}\underline{\mathbf{m}} + \underline{\mathbf{n}}$$

La stima di  $\underline{\mathbf{m}}$  a partire da  $\underline{\mathbf{d}}$  sarà ottenuta in base al principio di ortogonalità:

$$\begin{aligned}\hat{\underline{\mathbf{m}}} &= \underline{\mathbf{B}}\underline{\mathbf{d}} \\ E[(\underline{\mathbf{B}}\underline{\mathbf{d}} - \underline{\mathbf{m}})\underline{\mathbf{d}}^*] &= \underline{\mathbf{0}}\end{aligned}$$

Ma:

$$E[\underline{\mathbf{d}}\underline{\mathbf{d}}^*] = \underline{\mathbf{A}}E[\underline{\mathbf{m}}\underline{\mathbf{m}}^*]\underline{\mathbf{A}}^* + E[\underline{\mathbf{n}}\underline{\mathbf{n}}^*]$$

e supponendo per semplicità

$$E[\underline{\mathbf{m}}\underline{\mathbf{m}}^*] = \sigma_m^2 \underline{\mathbf{I}}_N; \quad E[\underline{\mathbf{n}}\underline{\mathbf{n}}^*] = \sigma_n^2 \underline{\mathbf{I}}_M$$

dove con  $\underline{\mathbf{I}}_N$ ,  $\underline{\mathbf{I}}_M$  è stata indicata la matrice identica negli spazi a  $N$  ed  $M$  dimensioni rispettivamente,

$$E[\underline{\mathbf{d}}\underline{\mathbf{d}}^*] = \sigma_m^2 \underline{\mathbf{A}}\underline{\mathbf{A}}^* + \sigma_n^2 \underline{\mathbf{I}}_M$$

$$E[\underline{\mathbf{m}}\underline{\mathbf{d}}^*] = \sigma_m^2 \underline{\mathbf{A}}^*$$

## CAPITOLO 10 LA TOMOGRAFIA

da cui

$$\begin{aligned}\underline{\mathbf{B}} (\sigma_m^2 \underline{\mathbf{A}} \underline{\mathbf{A}}^* + \sigma_n^2 \underline{\mathbf{I}}_M) &= \sigma_m^2 \underline{\mathbf{A}}^* \\ \underline{\mathbf{B}} &= \underline{\mathbf{A}}^* \left( \underline{\mathbf{A}} \underline{\mathbf{A}}^* + \frac{\sigma_n^2}{\sigma_m^2} \underline{\mathbf{I}}_M \right)^{-1}\end{aligned}\quad (10.31)$$

Ora, usando il cosiddetto lemma di inversione delle matrici, verificabile per sostituzione, per cui:

$$(\underline{\mathbf{B}} \underline{\mathbf{C}} \underline{\mathbf{D}} + \underline{\mathbf{A}})^{-1} = \underline{\mathbf{A}}^{-1} - \underline{\mathbf{A}}^{-1} \underline{\mathbf{B}} (\underline{\mathbf{C}}^{-1} + \underline{\mathbf{D}} \underline{\mathbf{A}}^{-1} \underline{\mathbf{B}})^{-1} \underline{\mathbf{D}} \underline{\mathbf{A}}^{-1}$$

si ha:

$$\underline{\mathbf{B}} = \left( \underline{\mathbf{A}}^* \underline{\mathbf{A}} + \frac{\sigma_n^2}{\sigma_m^2} \underline{\mathbf{I}}_N \right)^{-1} \underline{\mathbf{A}}^* \quad (10.32)$$

e, per un buon rapporto segnale rumore, si ottiene la (10.28). Si osservi che la (10.32) prevede l'inversione di una matrice  $[N \times N]$  mentre invece la (10.31) richiede l'inversione di una matrice  $[M \times M]$ , di dimensioni possibilmente ben maggiori.

### 10.7.4 Interpretazione tomografica del gradiente

Possiamo ora tentare di dare un'interpretazione tomografica delle tecniche di ottimizzazione con il metodo del gradiente e del gradiente coniugato.

Cerchiamo di risolvere l'equazione (10.30) con il metodo del gradiente; a partire da un modello iniziale, ad esempio  $\underline{\mathbf{m}}_0$ , calcoliamo il gradiente della forma quadratica in (10.27) e procediamo lungo questa direzione fino ad arrivare ad un minimo.

$$\underline{\mathbf{m}}_1 = \underline{\mathbf{m}}_0 + \alpha_0 \underline{\mathbf{g}}_0; \quad \underline{\mathbf{g}}_0 = \underline{\mathbf{A}}^* (\underline{\mathbf{A}} \underline{\mathbf{m}}_0 - \underline{\mathbf{d}}) \quad (10.33)$$

In sostanza il nuovo modello  $\underline{\mathbf{m}}_1$  è ottenuto dal precedente aggiungendo un vettore proporzionale al gradiente; questo non è altro che la **retro-proiezione del residuo** e cioè la retroproiezione della differenza tra la proiezione del modello iniziale ed i dati; il valore di  $\alpha_0$  è determinato imponendo che il nuovo gradiente (quindi il nuovo residuo)  $\underline{\mathbf{A}}^* (\underline{\mathbf{A}} \underline{\mathbf{m}}_1 - \underline{\mathbf{d}})$  sia ortogonale al gradiente precedente (e cioè il residuo precedente). Poi si retroproietta il nuovo residuo e si calcola il nuovo gradiente, si determina il passo con la condizione di ortogonalità e così via. In altre

## 10.7 TECNICHE ITERATIVE

parole abbiamo mostrato l'equivalenza tra la retroproiezione del residuo e il metodo del gradiente.

Il metodo del gradiente coniugato che si studierà nel prossimo capitolo invece prevederà la retroproiezione di qualcosa di diverso dal residuo, e cioè la parte del residuo corrispondente ad un modello ortogonale coniugato con il modello utilizzato sino a questo punto. Così si decompone lo spazio dei modelli secondo vettori ortogonali coniugati sotto la norma indotta dalla matrice  $\underline{\mathbf{A}}^* \underline{\mathbf{A}}$  e cioè tali che i modelli corrispondenti siano ortogonali **dopo il filtraggio con il  $\rho$ -filtro**. In questo modo, l'effetto di questo filtraggio può venire tenuto in conto e la soluzione viene determinata progressivamente.



# Capitolo 11

## Tecniche di ottimizzazione

### 11.1 Il gradiente

Supponiamo di volere risolvere il sistema di equazioni lineari:

$$\underline{\mathbf{R}}\underline{\mathbf{a}} = \underline{\mathbf{b}} \quad (11.1)$$

dove  $\underline{\mathbf{R}}$  e  $\underline{\mathbf{a}}$ ,  $\underline{\mathbf{b}}$  sono rispettivamente una matrice  $[N \times N]$  definita positiva e due vettori  $[N \times M]$  corrispondenti alla matrice del sistema ed ai termini incogniti e noti rispettivamente. Nel seguito, come di consueto, supporremo  $M = 1$ . Invece di risolvere il sistema, possiamo cercare il minimo della forma quadratica:

$$\frac{1}{2} \underline{\mathbf{a}}^* \underline{\mathbf{R}}\underline{\mathbf{a}} - \underline{\mathbf{b}}^* \underline{\mathbf{a}} = Q(\underline{\mathbf{a}}) \quad (11.2)$$

Si può minimizzare  $Q(\underline{\mathbf{a}})$  con la tecnica del gradiente. Indicando con  $\underline{\mathbf{a}}_0$  il valore iniziale assegnato al vettore  $\underline{\mathbf{a}}$  e  $\underline{\mathbf{f}}_0$  l'errore iniziale si ha:

$$\underline{\mathbf{a}}_0 = \underline{\mathbf{a}} + \underline{\mathbf{R}}\underline{\mathbf{f}}_0 \quad (11.3)$$

$$\underline{\mathbf{R}}\underline{\mathbf{a}}_0 = \underline{\mathbf{R}}(\underline{\mathbf{a}} + \underline{\mathbf{f}}_0) = \underline{\mathbf{b}} + \underline{\mathbf{f}}_0 \quad (11.4)$$

L'evoluzione con le successive iterazioni della soluzione approssimata può essere rappresentata con la convergenza a 0 della sequenza dei vettori (delle matrici) degli errori  $\underline{\mathbf{f}}_0, \underline{\mathbf{f}}_1, \underline{\mathbf{f}}_N$ . Il gradiente della forma quadratica nel punto di coordinate  $\underline{\mathbf{a}}_0$  è (supponiamo  $M = 1$ ):

$$\underline{\boldsymbol{\varepsilon}}_0 = \nabla_{\underline{\mathbf{a}}} Q(\underline{\mathbf{a}}_0) = \underline{\mathbf{R}}\underline{\mathbf{a}}_0 - \underline{\mathbf{b}} = \underline{\mathbf{R}}\underline{\mathbf{f}}_0 \quad (11.5)$$



## CAPITOLO 11 TECNICHE DI OTTIMIZZAZIONE

Con la tecnica della massima pendenza (steepest descent) aggiorniamo la soluzione all'iterazione  $k$ -esima ( $\underline{\mathbf{a}}_k$ ) aggiungendo un vettore la cui direzione è quella del gradiente  $\underline{\boldsymbol{\varepsilon}}_k$  e la cui ampiezza è determinata con un'ottimizzazione della variabile  $\alpha_k$ :

$$\underline{\mathbf{a}}_{k+1} = \underline{\mathbf{a}}_k + \alpha_k \underline{\boldsymbol{\varepsilon}}_k; \quad \underline{\boldsymbol{\varepsilon}}_k = \underline{\mathbf{R}} \underline{\mathbf{f}}_k \quad (11.6)$$

Ricordando che

$$\underline{\mathbf{a}}_k = \underline{\mathbf{a}} + \underline{\mathbf{f}}_k \quad (11.7)$$

risulta:

$$\underline{\mathbf{f}}_{k+1} = \underline{\mathbf{f}}_k + \alpha_k \underline{\mathbf{R}} \underline{\mathbf{f}}_k = (\underline{\mathbf{I}} + \alpha_k \underline{\mathbf{R}}) \underline{\mathbf{f}}_k \quad (11.8)$$

Volendo mantenere  $\alpha$  costante, risulta:

$$\underline{\mathbf{f}}_{k+1} = (\underline{\mathbf{I}} + \alpha \underline{\mathbf{R}})^{k+1} \underline{\mathbf{f}}_0 \quad (11.9)$$

Per avere stabilità nelle iterazioni, come già visto per il filtraggio adattativo, si ha:

$$\alpha < \frac{2}{Tr \underline{\mathbf{R}}} \quad (11.10)$$

In questo modo, gli autovalori della matrice in parentesi nell'equazione (11.9) sono minori di 1 e quindi l'errore converge a 0, seppure con infiniti passi. Di nuovo, si incontra il problema di una lenta convergenza nelle componenti di  $\underline{\mathbf{f}}_0$  dirette verso gli autovettori di  $\underline{\mathbf{R}}$  con autovalori piccoli; il problema è grave se gli autovalori di  $\underline{\mathbf{R}}$  sono di ampiezze assai differenti tra loro, anche se sono tutti positivi visto che  $\underline{\mathbf{R}}$  è positiva definita per ipotesi. Il valore ottimo di  $\alpha_k$  è determinato minimizzando ad ogni passo la forma quadratica  $Q(\underline{\mathbf{a}})$ :

$$\frac{\partial}{\partial \alpha_k} (\underline{\mathbf{f}}_k + \alpha_k \underline{\mathbf{R}} \underline{\mathbf{f}}_k)^* \underline{\mathbf{R}} (\underline{\mathbf{f}}_k + \alpha_k \underline{\mathbf{R}} \underline{\mathbf{f}}_k) = 0 \quad (11.11)$$

e pertanto:

$$\alpha_k = -\frac{\underline{\mathbf{f}}_k^* \underline{\mathbf{R}}^2 \underline{\mathbf{f}}_k}{\underline{\mathbf{f}}_k^* \underline{\mathbf{R}}^3 \underline{\mathbf{f}}_k} = -\frac{\underline{\boldsymbol{\varepsilon}}_k^* \underline{\boldsymbol{\varepsilon}}_k}{\underline{\boldsymbol{\varepsilon}}_k^* \underline{\mathbf{R}} \underline{\boldsymbol{\varepsilon}}_k} \quad (11.12)$$

Si osservi che i vettori  $\underline{\mathbf{R}} \underline{\mathbf{f}}_k = \underline{\boldsymbol{\varepsilon}}_k$  e  $\underline{\mathbf{R}} \underline{\mathbf{f}}_{k+1} = \underline{\boldsymbol{\varepsilon}}_{k+1}$  sono ortogonali.

$$\underline{\boldsymbol{\varepsilon}}_k = \underline{\mathbf{R}} \underline{\mathbf{f}}_k \perp \underline{\mathbf{R}} \underline{\mathbf{f}}_{k+1} = \underline{\boldsymbol{\varepsilon}}_{k+1} \quad (11.13)$$

## 11.2 IL GRADIENTE CONIUGATO

In altre parole, con la tecnica della discesa lungo la linea di massima pendenza e con passi di lunghezza ottimizzata, si prosegue lungo la direzione di partenza fino a che non si sia su un minimo e perciò tangenti ad una curva di livello. Pertanto, le direzioni successive sono ortogonali tra loro. Questo limita di molto l'efficienza nel caso di dispersione degli autovalori, come verrà meglio chiarito nel paragrafo successivo sul gradiente coniugato.

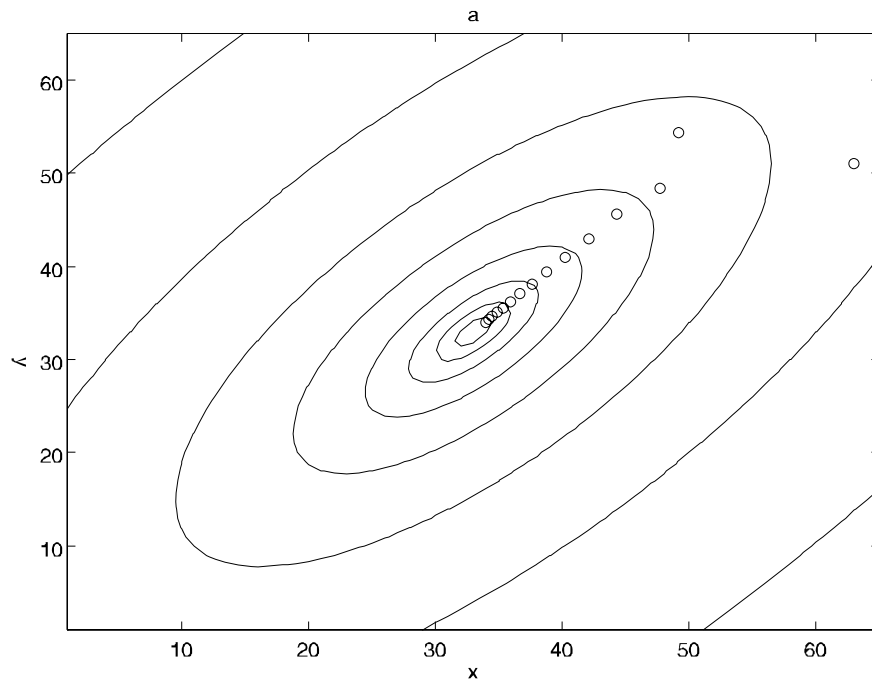


Figura 11.1: Ottimizzazione con la tecnica del gradiente

## 11.2 Il gradiente coniugato

### 11.2.1 Ortogonalità coniugata, incorrelazione e stima lineare

Prima di discutere la tecnica del gradiente coniugato, è opportuno introdurre una interpretazione di tipo stocastico per una maggiore compren-

## CAPITOLO 11 TECNICHE DI OTTIMIZZAZIONE

sione. Si consideri  $\underline{\mathbf{x}}$ , un vettore casuale di  $N$  elementi, caratterizzato da una matrice di covarianza  $\underline{\mathbf{R}}$ . Si abbia cioè

$$E[\underline{\mathbf{x}}\underline{\mathbf{x}}^*] = \underline{\mathbf{R}}$$

La matrice  $\underline{\mathbf{R}}$  è di dimensioni  $[N \times N]$ . Si considerino ora delle combinazioni lineari delle  $N$  componenti  $\{x_n\}$  del vettore  $\underline{\mathbf{x}}$  e cioè

$$y_i = \underline{\mathbf{a}}_i^* \underline{\mathbf{x}} \quad (11.14)$$

ottenute facendo un prodotto scalare tra  $\underline{\mathbf{x}}$  ed un qualsiasi vettore  $\underline{\mathbf{a}}_i$  di  $N$  componenti. Risulta che:

$$E[y_i y_j] = E[\underline{\mathbf{a}}_i \underline{\mathbf{x}} \underline{\mathbf{x}}^* \underline{\mathbf{a}}_j] = \underline{\mathbf{a}}_i^* \underline{\mathbf{R}} \underline{\mathbf{a}}_j \quad (11.15)$$

Le variabili casuali  $y_i, y_j$  sono incorrelate se  $E[y_i y_j] = 0$  e cioè se

$$\underline{\mathbf{a}}_i^* \underline{\mathbf{R}} \underline{\mathbf{a}}_j = 0 \quad (11.16)$$

I due vettori  $\underline{\mathbf{a}}_i, \underline{\mathbf{a}}_j$  sono allora detti essere **ortogonali coniugati** tra di loro sotto la norma della matrice  $\underline{\mathbf{R}}$ . Indicando con  $\{a_{i,n}\}$  le componenti del vettore  $\underline{\mathbf{a}}_i$ , lungo gli autovettori  $\underline{\mathbf{e}}_n$  della matrice  $\underline{\mathbf{R}}$ , risulta:

$$\sum_{n=1}^N \lambda_n a_{i,n} a_{j,n} = 0 \quad (11.17)$$

I due vettori sarebbero anche ortogonali tra loro se

$$\sum_{n=1}^N a_{i,n} a_{j,n} = 0; \quad \lambda_n = \text{cost} \quad (11.18)$$

e quindi se la matrice  $\underline{\mathbf{R}}$  fosse una matrice identica e il processo fosse bianco. Si ricordi che  $\underline{\mathbf{R}}$  è una matrice semi definita positiva visto che  $E[y_i^2] \geq 0$  per qualsiasi vettore  $\underline{\mathbf{a}}_i$  non nullo. L'onnipresenza del rumore ci consente poi di trascurare il caso in cui la matrice  $\underline{\mathbf{R}}$  abbia autovalori nulli. Abbiamo perciò visto come all'incorrelazione delle due variabili casuali  $y_i, y_j$  corrisponda l'ortogonalità coniugata dei vettori deterministici  $\underline{\mathbf{a}}_i, \underline{\mathbf{a}}_j$  sotto la norma della matrice  $\underline{\mathbf{R}}$ . Vediamo ora come anche le tecniche di stima lineare hanno una corrispondenza nel settore della minimizzazione di forme quadratiche. Utilizzando il vettore casuale  $\underline{\mathbf{x}}$ , si

## 11.2 IL GRADIENTE CONIUGATO

cerchi di stimare in modo lineare ottimo una variabile casuale  $\zeta$ ; ciò si ottiene mediante:

$$\hat{\zeta} = \underline{\mathbf{a}}_{\zeta}^* \underline{\mathbf{x}} \quad (11.19)$$

una combinazione lineare delle componenti di  $\underline{\mathbf{x}}$  ottenuta facendone il prodotto scalare con un vettore  $\underline{\mathbf{a}}_z$  da determinare in modo ottimale. Senza ledere la generalità si può stimare anziché  $\zeta$  la variabile casuale *residuale*:

$$z = \zeta - \underline{\mathbf{a}}_0^* \underline{\mathbf{x}} \quad (11.20)$$

dove  $\underline{\mathbf{a}}_0$  è un vettore qualsiasi, tipicamente un'ipotesi iniziale di stimatore, da raffinare poi con il calcolo successivo. Dobbiamo pertanto determinare in modo ottimale  $\underline{\mathbf{a}}_z$  per calcolare  $\underline{\mathbf{a}}_z^* \underline{\mathbf{x}}$  stimatore ottimo della variabile casuale  $z$ .

$$\hat{z} = \underline{\mathbf{a}}_z^* \underline{\mathbf{x}} \quad (11.21)$$

Il vettore  $\underline{\mathbf{a}}_z$  si trova con la minimizzazione dell'errore quadratico medio, oppure con il principio di ortogonalità:

$$E[\varepsilon^2] = E[z - \underline{\mathbf{a}}_z^* \underline{\mathbf{x}}]^2 = E[z^2] + \underline{\mathbf{a}}_z^* \underline{\mathbf{R}} \underline{\mathbf{a}}_z - 2 \underline{\mathbf{a}}_z^* E[z \underline{\mathbf{x}}] \quad (11.22)$$

Il minimo si raggiunge per

$$\underline{\mathbf{R}} \underline{\mathbf{a}}_z = \underline{\mathbf{b}} = E[z \underline{\mathbf{x}}] \quad (11.23)$$

Le componenti di  $\underline{\mathbf{b}}$  e cioè  $\{b_i\}$  sono date da

$$b_i = E[z x_i] \quad (11.24)$$

Lo scalare  $E[\varepsilon^2]$ , i vettori  $\underline{\mathbf{a}}_z$ ,  $\underline{\mathbf{b}}$  e la matrice  $\underline{\mathbf{R}}$  hanno il ruolo che avevano rispettivamente nel paragrafo precedente  $Q(\underline{\mathbf{a}})$ ,  $\underline{\mathbf{a}}$ ,  $\underline{\mathbf{b}}$  e la matrice  $\underline{\mathbf{R}}$ . Inoltre il vettore gradiente del paragrafo precedente corrisponde alla correlazione tra errore di stima e vettore dati:

$$\underline{\varepsilon} = \underline{\mathbf{R}} \underline{\mathbf{a}} - \underline{\mathbf{b}} = E[\varepsilon \underline{\mathbf{x}}] \quad (11.25)$$

Quando il gradiente è nullo, i dati sono incorrelati con l'errore di stima e il problema è risolto.

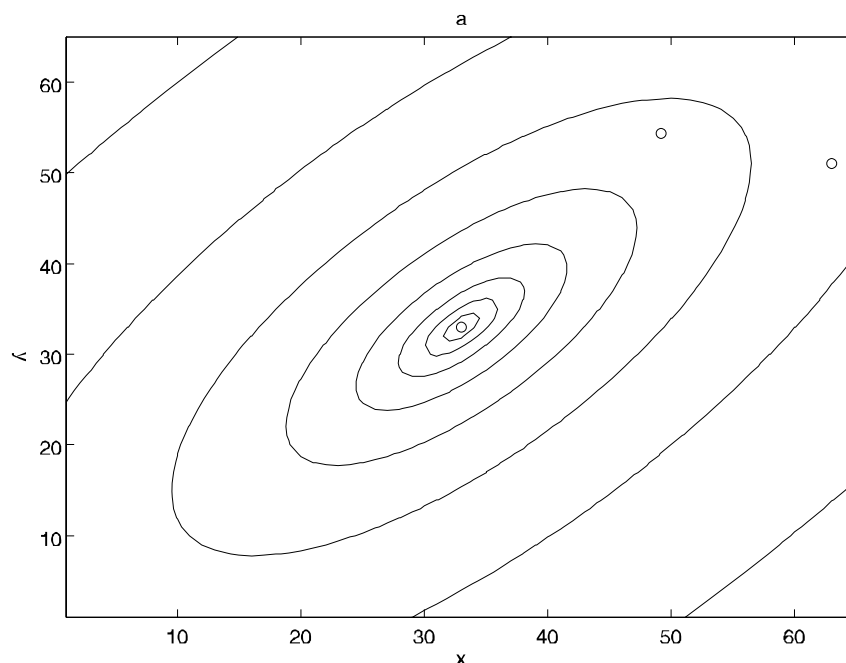


Figura 11.2: Ottimizzazione con la tecnica del gradiente coniugato

### 11.2.2 Connessione con la ricorsione di Levinson

Si voglia ora stimare la variabile casuale  $z$  con una tecnica corrispondente a quella ricorsiva di Levinson. In altre parole, si inizi a stimare  $z$  usando solo la prima componente di  $\underline{x}$  poi le prime due e così via, fino ad utilizzare tutte le  $N$  componenti. Il punto chiave è che, passando dallo stimatore di lunghezza  $m$  a quello di lunghezza  $m + 1$ , utilizziamo, per migliorare la stima, l'errore di predizione "all'indietro" e cioè la differenza tra il valore effettivo della  $m + 1$ -esima componente di  $\underline{x}$  e la sua stima in base alle prime  $m$  componenti di  $\underline{x}$  che già sono state utilizzate per stimare  $z$ . In questo modo arricchiamo la stima fatta in precedenza con l'uso di una nuova variabile casuale che da un lato è **incorrelata con tutti** i dati utilizzati finora (e quindi porta nuova informazione) ma d'altra parte può essere correlata con la variabile casuale da stimare. Quindi, l'errore di stima non può che decrescere e la minimizzazione dell'errore quadratico medio può essere fatta progressivamente su di una variabile

## 11.2 IL GRADIENTE CONIUGATO

alla volta senza distruggere l'ottimizzazione fatta con le variabili precedenti. Inoltre, la determinazione dello stimatore ottimo termina in  $N$  passi, perchè terminerà quando abbiamo utilizzato tutte le  $N$  variabili casuali, tra loro incorrelate, con cui possiamo rappresentare il vettore casuale  $\underline{\mathbf{x}}$ .

Nel caso generale del gradiente coniugato il ragionamento va come segue: la ricerca dello stimatore ottimo  $\underline{\mathbf{a}}_z$  parte dalla determinazione di un primo stimatore di prova  $\underline{\mathbf{a}}_0$ . La prima "direzione"  $\underline{\mathbf{D}}_0$  può essere trovata con il criterio della massima pendenza e quindi calcolando il gradiente dell'errore quadratico medio:

$$\underline{\mathbf{D}}_0 = \nabla \underline{\mathbf{a}} \left( \frac{\underline{\mathbf{a}}^* \underline{\mathbf{R}} \underline{\mathbf{a}}}{2} - \underline{\mathbf{b}}^* \underline{\mathbf{a}} \right) \big|_{\underline{\mathbf{a}}=\underline{\mathbf{a}}_0} = \underline{\mathbf{R}} \underline{\mathbf{a}}_0 - \underline{\mathbf{b}} \quad (11.26)$$

Poi, si cerca la lunghezza del passo minimizzando l'errore quadratico medio e quindi imponendo che la variabile casuale che corrisponde all'errore residuo sia incorrelata rispetto al dato e cioè la variabile casuale  $\underline{\mathbf{D}}_0^* \underline{\mathbf{x}}$  pertanto si calcola il numero  $\alpha_0$  tale che:

$$\alpha_0 = - \frac{(\underline{\mathbf{R}} \underline{\mathbf{a}}_0 - \underline{\mathbf{b}})^* \underline{\mathbf{D}}_0}{\underline{\mathbf{D}}_0^* \underline{\mathbf{R}} \underline{\mathbf{D}}_0} \quad (11.27)$$

Il nuovo stimatore  $\underline{\mathbf{a}}_1$  è:

$$\underline{\mathbf{a}}_1 = \underline{\mathbf{a}}_0 + \alpha_0 \underline{\mathbf{D}}_0 \quad (11.28)$$

Ora un altro passo; vogliamo minimizzare ancora l'errore quadratico medio, muovendoci nella direzione  $\underline{\mathbf{D}}_1$ ; come nella ricorsione di Levinson, vogliamo che la variabile casuale  $\underline{\mathbf{D}}_1^* \underline{\mathbf{x}}$  sia incorrelata con la variabile casuale che abbiamo usato prima, per utilizzare **solamente** dell' "informazione nuova". Per accelerare la minimizzazione dell'errore, la nuova direzione  $\underline{\mathbf{D}}_1$  sarà vicina al gradiente in  $\underline{\mathbf{a}}_1$ , e cioè  $\underline{\mathbf{g}}_1 = \underline{\mathbf{R}} \underline{\mathbf{a}}_1 - \underline{\mathbf{b}}$ , per quanto possibile; deve anche però essere incorrelata (ortogonale coniugata) rispetto alla prima direzione presa  $\underline{\mathbf{D}}_0$ . Quindi:

$$\underline{\mathbf{D}}_1 = \underline{\mathbf{R}} \underline{\mathbf{a}}_1 - \underline{\mathbf{b}} + \beta_1 \underline{\mathbf{D}}_0 = \underline{\mathbf{g}}_1 + \beta_1 \underline{\mathbf{D}}_0 \quad (11.29)$$

dove  $\beta_1$  si trova imponendo l'ortogonalità coniugata tra  $\underline{\mathbf{D}}_1$  e  $\underline{\mathbf{D}}_0$ .

$$\beta_1 = - \frac{\underline{\mathbf{g}}_1^* \underline{\mathbf{R}} \underline{\mathbf{D}}_0}{\underline{\mathbf{D}}_0^* \underline{\mathbf{R}} \underline{\mathbf{D}}_0} \quad (11.30)$$

## CAPITOLO 11 TECNICHE DI OTTIMIZZAZIONE

L'algoritmo procede aggiornando lo stimatore  $\underline{\mathbf{a}}_1$  spingendosi nella nuova direzione  $\underline{\mathbf{D}}_1$  con l'opportuna lunghezza del passo; ciò avviene ortogonalizzando l'errore di stima residuo rispetto al nuovo dato e cioè la variabile casuale  $\underline{\mathbf{D}}_1^* \underline{\mathbf{x}}$  e quindi imponendo l'ortogonalità tra il gradiente  $\underline{\mathbf{g}}_2$  nella nuova posizione  $\underline{\mathbf{a}}_2$  e  $\underline{\mathbf{D}}_1$ :

$$\underline{\mathbf{a}}_2 = \underline{\mathbf{a}}_1 + \alpha_1 \underline{\mathbf{D}}_1 \quad (11.31)$$

$$\alpha_1 = -\frac{\underline{\mathbf{g}}_1^* \underline{\mathbf{D}}_1}{\underline{\mathbf{D}}_1^* \underline{\mathbf{R}} \underline{\mathbf{D}}_1} \quad (11.32)$$

A questo punto dobbiamo individuare un'altra direzione  $\underline{\mathbf{D}}_2$  e cioè un'altra variabile casuale  $\underline{\mathbf{D}}_2 \underline{\mathbf{x}}$  con cui fare un'ulteriore stima del residuo. Per quanto detto scegliamo questa direzione con il solito criterio di stare vicini al gradiente in  $\underline{\mathbf{a}}_2$  pur mantenendo l'ortogonalità coniugata con la direzione  $\underline{\mathbf{D}}_1$  risulta allora:

$$\underline{\mathbf{D}}_2 = \underline{\mathbf{R}} \underline{\mathbf{a}}_2 - \underline{\mathbf{b}} + \beta_2 \underline{\mathbf{D}}_1 \quad (11.33)$$

La variabile  $\beta_2$  viene scelta con il criterio di imporre che la nuova variabile casuale  $\underline{\mathbf{D}}_2^* \underline{\mathbf{x}}$  sia incorrelata con quella usata precedentemente e cioè  $\underline{\mathbf{D}}_1^* \underline{\mathbf{x}}$ . Quindi:

$$\beta_2 = -\frac{\underline{\mathbf{g}}_2^* \underline{\mathbf{R}} \underline{\mathbf{D}}_1}{\underline{\mathbf{D}}_1^* \underline{\mathbf{R}} \underline{\mathbf{D}}_1} \quad (11.34)$$

Se si desidera ancora decrescere il valore quadratico medio dell'errore di stima, ci dobbiamo comunque muovere nella direzione del gradiente dell'errore rispetto al vettore  $\underline{\mathbf{a}}$ . Pur così facendo, ci si muove verso una direzione ortogonale coniugata rispetto a tutte le direzioni precedenti (aggiungere una variabile casuale incorrelata con le variabili casuali finora utilizzate) altrimenti le minimizzazioni precedenti avrebbero potuto essere migliorate portando ad errori inferiori. Il nuovo vettore perciò non può puntare verso una delle direzioni già utilizzate.

Il residuo è **incorrelato** con le componenti per cui è stata fatta la minimizzazione e pertanto punta in direzione ortogonale coniugata ad esse; lo spazio dei parametri si riduce di una dimensione ad ogni passo fino a vuotarsi in  $N$  passi. L'algoritmo quindi prosegue e termina in  $N$  passi, con la determinazione dello stimatore ottimo  $\underline{\mathbf{a}}_z$ . Si osservi che già nei primi passi si era trovata una buona approssimazione dello stimatore; ciò è tanto più vero quanto maggiore è la dispersione degli autovalori della matrice  $\underline{\mathbf{R}}$ .

## 11.3 Gradiente e gradiente coniugato

Sulla base delle osservazioni statistiche fatte nella sezione precedente si comprende come mai il metodo del gradiente coniugato sia assai piú efficiente di quello della massima pendenza. Anzitutto, si ha convergenza all'ottimo in  $N$  passi e non in infiniti come nella massima pendenza. Infatti, l'ottimizzazione è progressivamente fatta rispetto a variabili tra loro incorrelate, fino a che tutte le componenti del vettore  $\underline{x}$  siano perfettamente descritte. Ancora, la convergenza è la piú veloce possibile, e cioè in pratica dopo pochi passi le componenti di  $\underline{a}_z$  relative agli autovettori della matrice  $\underline{R}$  con autovalori piú elevati sono determinate con buona cura. Successivamente sono determinate tutte le altre componenti, fino alla soluzione completa cui si giunge in  $N$  passi, come detto.

Il paragone classico che viene fatto per spiegare la tecnica del gradiente coniugato è quello relativo ad uno sciatore in un canalone. Con la tecnica del gradiente, lo sciatore parte lontano dal fondo del canalone e si avvia lungo la direzione della linea di massimo pendio; poi, prosegue sempre dritto (traversando il canalone) fino a quasi risalire il pendio; si ferma cioè quando gli sci sono tangenti ad una curva di livello (Il punto di partenza è in alto a destra nella fig 20.1; attenzione, le curve di livello **non** sono equispaziate). Poi riparte per il tratto successivo, riorientando gli sci nella direzione di massimo pendio, e quindi in una direzione ortogonale a quella precedente. Ciò lo porta a ritraversare il canalone fino a portarsi un poco piú in basso sul versante opposto; in altre parole lo sciatore fa numerosissimi passi brevi. Questa situazione corrisponde ad una forma quadratica caratterizzata da un forte dispersione degli autovalori (piccolo nella direzione del canalone, grandi nelle direzioni normali ad esso).

Con la tecnica del gradiente coniugato, lo sciatore è forzato ad allontanarsi dalla direzione di massimo pendio procedendo verso una direzione "nuova" rispetto all'ultima direzione presa ed anche a tutte le precedenti. Pertanto, invece di disporre gli sci lungo la direzione di massimo pendio, vi rinuncia e imbocca il canalone facendo passi piú lunghi, anche se in una direzione in cui il pendio è meno ripido. Nel caso della fig. 20.2, si arriva al minimo di quota in due passi, coerentemente con il fatto che se la minimizzazione è fatta con il gradiente coniugato, si arriva al minimo in  $N$  passi nel caso di una forma quadratica  $N$  - dimensionale). Il primo passo è lo stesso che per il metodo del gradiente. Con il gradiente co-



## CAPITOLO 11 TECNICHE DI OTTIMIZZAZIONE

niugato, quindi, la direzione di discesa si allontana progressivamente da quelle di massima concavità per esplorare le direzioni dove le concavità sono minori. Peraltro, le successive ottimizzazioni **non sono in contrasto tra loro** e progressivamente portano alla soluzione. Negli ultimi passi, lo sciatore è nel "campetto" vicino al traguardo e anche se si ferma, non accade nulla di grave. In altre parole, la presenza di errori di stima della matrice di covarianza portano comunque a errori nel posizionamento del minimo globale (*pozzanghera vicino al traguardo*). Peraltro, anche se lo sciatore non arriva esattamente al traguardo (il processo termina prima di  $N$  passi), la perdita di livello è minima.