

Fondamenti di Elaborazione Numerica dei Segnali  
Note estratte dal Libro del Prof. Rocca  
L'introduzione alla Teoria della Stima è presa dal  
materiale del Prof. Spagnolini per il corso di Elaborazione  
Numerica dei Segnali (Laurea Magistrale TLC)

a cura di S. Tubaro

January 19, 2016

# Processi casuali e stime campionarie

## 1 Processi casuali discreti

Consideriamo una sequenza di variabili aleatorie reali  $\{x(n)\}$ ,  $n = 0 \dots N - 1$  e supponiamo che la densità di probabilità congiunta delle variabili sia gaussiana e che le medie siano tutte nulle. Se le medie non fossero nulle, potrebbero essere sottratte a ciascuna v.a. e gestite separatamente come entità deterministiche. Indichiamo con  $\mathbf{x}$  un vettore colonna le cui  $N$  componenti sono gli elementi della sequenza  $\{x(n)\}$  e con  $\mathbf{C}_x$  la matrice di autocovarianza (ed anche di autocorrelazione), di dimensioni  $[N \times N]$ , del vettore (e quindi della sequenza) e cioè il valore atteso del prodotto  $\mathbf{x}\mathbf{x}^T$ :

$$\mathbf{C}_x = E[\mathbf{x}\mathbf{x}^T] \quad (1)$$

dove  $T$  indica la trasposizione.  $\mathbf{C}_x$ , che verrà spesso chiamata anche matrice di covarianza (o di correlazione), risulta ovviamente simmetrica.

Assumendo il vettore casuale  $\mathbf{x}$  ottenuto campionando a passo regolare un processo casuale il cui l'asse della variabile indipendente sia stato normalizzato in modo da avere il passo di campionamento unitario avremo che le componenti del vettore  $\{x(n)\}$ ,  $n = 0 \dots N$  rappresentano  $N$  "campioni" del processo casuale. Se il processo casuale da cui è estratto il vettore è stazionario possiamo definire per questo processo una funzione di autocorrelazione (coincidente con l'autocovarianza se il valor medio è nullo) funzione di un solo parametro, ed in particolare possiamo considerare la versione campionata (passo 1) di questa autocorrelazione:  $r_x(m) = E[x(n)x(n-m)]$ . In questo caso abbiamo che la matrice  $\mathbf{C}_x$  può essere scritta sulla base della funzione di autocorrelazione come:

$$\mathbf{C}_x = \begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) & \dots & r_x(1-N) \\ r_x(1) & r_x(0) & r_x(-1) & \dots & \dots \\ r_x(2) & r_x(1) & r_x(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_x(N-1) & \dots & \dots & \dots & r_x(0) \end{bmatrix} \quad (2)$$

Si osservi che operando con dati (segnali) "veri", per tutti i valori attesi sopraindicati sarà solo possibile fare delle medie campionarie, e quindi fare solo delle stime, più o meno attendibili e significative.

La matrice  $\mathbf{C}$  è simmetrica e cioè, indicando con  $T$  la trasposizione si ha:

$$\mathbf{C}_x^T = \mathbf{C}_x \quad (3)$$

Inoltre risulta che gli elementi sulla diagonale principale e tutte le altre diagonali sono uguali fra loro. Ciò corrisponde a dire che la matrice è di "Toeplitz".

Si definisce che la densità di probabilità del vettore casuale  $\mathbf{x}$  è gaussiana se può essere scritta nel seguente modo:

$$p(\mathbf{x}) = \frac{1}{(2\pi)^{N/2} |\mathbf{C}_x|^{1/2}} \exp\left(-\frac{\mathbf{x}^T \mathbf{C}_x^{-1} \mathbf{x}}{2}\right) \quad (4)$$

dove  $|\mathbf{C}_x|$  rappresenta il determinante della matrice di covarianza. La precedente relazione è corretta nel caso di variabili a media nulla. Se così non fosse nella espressione ad esponente dove appare  $\mathbf{x}$  andrebbe sostituito  $\mathbf{x}$  meno il vettore delle medie.

Nel seguito sono considerati alcuni casi nei quali il vettore  $\mathbf{x}$  appare intuitivamente gaussiano e verificheremo che effettivamente lo è.

Se gli elementi del vettore  $\mathbf{x}$  sono i.i.d. e cioè indipendenti e identicamente distribuiti (p.e. appartengono ad una sequenza "bianca") si ha:

$$E[x(n)x(n-m)] = \sigma_x^2 \delta(m)$$

la matrice di covarianza  $\mathbf{C}_x$  è diagonale e può essere scritta come:

$$\mathbf{C}_x = \sigma_x^2 \mathbf{I} \quad (5)$$

dove con  $\mathbf{I}$  abbiamo indicato la matrice identica e con  $\sigma_x^2$  il valore q.m. degli elementi della sequenza. In questo caso la 4 risulta immediatamente poichè la densità di probabilità di un vettore le cui componenti sono delle variabili indipendenti è il prodotto delle loro densità.

$$\begin{aligned} p(\mathbf{x}) &= \prod_{i=1}^N p(x_i) = \frac{1}{(2\pi)^{N/2} \sigma_x^N} \exp\left(-\frac{1}{2\sigma_x^2} \sum x_i^2\right) = \\ &= \frac{1}{(2\pi)^{N/2} |\mathbf{C}_x|^{1/2}} \exp\left(-\frac{\mathbf{x}^T \mathbf{C}_x^{-1} \mathbf{x}}{2}\right) \end{aligned} \quad (6)$$

L'estensione al caso in cui gli elementi di  $\mathbf{x}$  siano indipendenti, ma con varianze diverse è immediato.

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

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

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

$$E[\mathbf{y}\mathbf{y}^T] = \mathbf{C}_y = \mathbf{A}E[\mathbf{x}\mathbf{x}^T]\mathbf{A}^T = \sigma_x^2 \mathbf{A}\mathbf{A}^T$$

ricordando la 5 possiamo scrivere:

$$\mathbf{x}^T \mathbf{C}_x^{-1} \mathbf{x} = \mathbf{y}^T (\mathbf{A}^{-1})^T \frac{1}{\sigma_x^2} \mathbf{I} \mathbf{A}^{-1} \mathbf{y} = \frac{1}{\sigma_x^2} \mathbf{y}^T (\mathbf{A}^{-1})^T \mathbf{A}^{-1} \mathbf{y} = \mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{y}$$

e sostituendo nella 6 si ottiene la densità di probabilità 4 a meno di un fattore di scala. Tale fattore può essere risolto normalizzando ad 1 l'integrale della probabilità su tutto il volume infinito  $N$  - dimensionale dopo la trasformazione lineare. Per fare ciò bisogna ricordare che effettuando una trasformazione i volumi vengono mantenuti se si divide per il determinante dello Jacobiano della trasformazione di coordinate da  $\mathbf{x}$  a  $\mathbf{y}$ . Nel caso di trasformazione lineare, lo Jacobiano corrisponde alla matrice di trasformazione stessa ( $\mathbf{A}$ ). Abbiamo quindi:

$$\begin{aligned}
p(\mathbf{y}) &= \frac{1}{(2\pi)^{N/2} |\mathbf{C}_x|^{1/2} |\mathbf{A}|} \exp\left(-\frac{\mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{y}}{2}\right) \\
&= \frac{1}{(2\pi)^{N/2} \sigma_x^N |\mathbf{A}|} \exp\left(-\frac{\mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{y}}{2}\right) \\
&= \frac{1}{(2\pi)^{N/2} |\sigma_x \mathbf{A}|} \exp\left(-\frac{\mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{y}}{2}\right) \\
&= \frac{1}{(2\pi)^{N/2} \sqrt{|\sigma_x^2 \mathbf{A} \mathbf{A}^T|}} \exp\left(-\frac{\mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{y}}{2}\right) \\
&= \frac{1}{(2\pi)^{N/2} |\mathbf{C}_y|^{1/2}} \exp\left(-\frac{\mathbf{y}^T \mathbf{C}_y^{-1} \mathbf{y}}{2}\right) \tag{7}
\end{aligned}$$

e quindi risulta la 4:

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

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

$$\mathbf{A} = \mathbf{U} \mathbf{Q} \mathbf{U}^{-1}$$

dove  $\mathbf{U}$  è la matrice degli autovettori di  $\mathbf{A}$ , orto-normale e quindi con determinante unitario. Quindi la trasformazione indotta da  $\mathbf{A}$  è decomponibile nella cascata di una rotazione  $N$ -dimensionale (la matrice  $\mathbf{U}$ ), uno scalamento di coordinate diverso da coordinata a coordinata (la matrice diagonale  $\mathbf{Q}$  che ha come elementi della diagonale principale gli autovalori di  $\mathbf{A}$ ,  $q_i$ ) ed una rotazione di coordinate opposta alla precedente (la matrice  $\mathbf{U}^{-1} = \mathbf{U}^T$ ). 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:

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

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).

E' opportuno ricordare come un vettore aleatorio  $\mathbf{x}$  gaussiano con media  $\mu_{\mathbf{x}}$  e matrice di covarianza  $\mathbf{C}_{\mathbf{x}}$  è indicato come  $\mathbf{x} \sim \mathcal{N}(\mu_{\mathbf{x}}, \mathbf{C}_{\mathbf{x}})$ . Conseguentemente un vettore  $\mathbf{y}$  a media nulla e matrice di autocorrelazione  $\mathbf{C}_{\mathbf{y}}$  sarà indicato come  $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_{\mathbf{y}})$ .

## 1.1 Estensione a dati complessi

Un vettore aleatorio di dimensione  $N$  scritto come  $\mathbf{x} = \mathbf{x}_C + j\mathbf{x}_I$  si dice gaussiano quando il vettore reale  $\begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix}$ , vettore di dimensione  $2N$  composto dalla parte reale ed immaginaria del vettore  $\mathbf{x}$ , è gaussiano.

Assumendo di avere avere valori medi nulli, la matrice di (auto)covarianza di questo vettore di dimensione  $2N$  è scritta come

$$\begin{aligned} \mathbf{C} &= E \left\{ \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix} \begin{bmatrix} \mathbf{x}_R \\ \mathbf{x}_I \end{bmatrix}^T \right\} = E \begin{bmatrix} E[\mathbf{x}_R \mathbf{x}_R^T] & E[\mathbf{x}_R \mathbf{x}_I^T] \\ E[\mathbf{x}_I \mathbf{x}_R^T] & E[\mathbf{x}_I \mathbf{x}_I^T] \end{bmatrix} = \begin{bmatrix} \mathbf{C}_R & \mathbf{C}_{RI} \\ \mathbf{C}_{IR} & \mathbf{C}_I \end{bmatrix} \\ &= \begin{bmatrix} \mathbf{C}_R & \mathbf{C}_{RI} \\ \mathbf{C}_{RI}^T & \mathbf{C}_I \end{bmatrix} \end{aligned} \quad (8)$$

dove il segno  $T$  indica la semplice trasposizione.

E' da notare che, mentre per caratterizzare un vettore aleatorio gaussiano a valori reali occorre conoscere solo la matrice di correlazione (ed il vettore dei valori medi), per la determinazione della statistica di un vettore gaussiano a valori complessi non basta conoscere, oltre ai valori medi, la sua matrice di covarianza. Infatti quest'ultima è ovviamente definita come:

$$\begin{aligned} \mathbf{C}_x &= E[\mathbf{x}\mathbf{x}^H] = E[(\mathbf{x}_R + j\mathbf{x}_I)(\mathbf{x}_R + j\mathbf{x}_I)^H] = E[(\mathbf{x}_R + j\mathbf{x}_I)(\mathbf{x}_R^T - j\mathbf{x}_I^T)] \\ &= (\mathbf{C}_R + \mathbf{C}_I) + j(\mathbf{C}_{IR} - \mathbf{C}_{RI}) \end{aligned}$$

dove  $\mathbf{x}^H$  è il vettore ottenuto trasponendo e applicando il coniugio al vettore  $\mathbf{x}$  ed il segno  $T$  indica sempre la semplice trasposizione. Normalmente si richiede di conoscere anche la cosiddetta *pseudo-covarianza* definita da:

$$\begin{aligned} \tilde{\mathbf{C}}_x &= E[\mathbf{x}\mathbf{x}^T] = E[(\mathbf{x}_R + j\mathbf{x}_I)(\mathbf{x}_R + j\mathbf{x}_I)^T] = E[(\mathbf{x}_R + j\mathbf{x}_I)(\mathbf{x}_R^T + j\mathbf{x}_I^T)] \\ &= (\mathbf{C}_R - \mathbf{C}_I) + j(\mathbf{C}_{IR} + \mathbf{C}_{RI}) \end{aligned}$$

Infatti si ottiene:

$$\begin{aligned} \mathbf{C}_R &= \frac{1}{2} \text{Re}[\mathbf{C}_x + \tilde{\mathbf{C}}_x], \quad \mathbf{C}_I = \frac{1}{2} \text{Re}[\mathbf{C}_x - \tilde{\mathbf{C}}_x] \\ \mathbf{C}_{RI} &= \frac{1}{2} \text{Im}[-\mathbf{C}_x + \tilde{\mathbf{C}}_x], \quad \mathbf{C}_{IR} = \frac{1}{2} \text{Im}[\mathbf{C}_x + \tilde{\mathbf{C}}_x] \end{aligned}$$

Una particolare classe di vettori aleatori gaussiani a valori complessi è costituita dai cosiddetti *vettori gaussiani propri*. Essi sono definiti dalla condizione

$$\tilde{\mathbf{C}}_x = \mathbf{0} \Rightarrow \mathbf{C}_R = \mathbf{C}_I; \quad \mathbf{C}_{IR} = -\mathbf{C}_{RI} \quad (9)$$

E' opportuno notare che poichè  $\mathbf{C}_{RI} = \mathbf{C}_{IR}^T$  (vedi 8, si ricordi che  $T$  è sempre la semplice trasposizione), risulta, applicando la 9:  $\mathbf{C}_{RI} = -\mathbf{C}_{RI}^T$ . In totale si può dire che le parti reali ed immaginarie di un vettore aleatorio complesso gaussiano proprio  $(\mathbf{x}_R, \mathbf{x}_I)$  hanno identiche

matrici di covarianza e matrice di covarianza incrociata antisimmetrica. In particolare la diagonale principale di quest'ultima matrice sarà costituita tutta da valori nulli. Sulla base di tutto ciò si desume che per un vettore gaussiano complesso proprio (indicato come  $\mathbf{x} \approx \mathcal{CN}(\mathbf{m}_x, \mathbf{C}_x)$ ) basta conoscere la matrice covarianza,  $\mathbf{C}_x$ , e l'eventuale vettore dei valori medi per descriverne la statistica.

Una ulteriore particolare classe di vettori gaussiani complessi è costituita dai vettori che godono della cosiddetta proprietà di simmetria circolare. Un vettore gaussiano complesso è *circularmente simmetrico* se per ogni  $\theta \in [0, 2\pi)$  la statistica di  $\mathbf{x}$  è la stessa di  $e^{j\theta}\mathbf{x}$ . Devono perciò coincidere il vettore dei valori medi e le matrici di covarianza e pseudo-covarianza. Basta, in particolare, che risulti, per ogni  $\theta$ :

$$\begin{aligned} E[e^{j\theta}\mathbf{x}] &= e^{j\theta}E[\mathbf{x}] = E[\mathbf{x}] \\ E[(e^{j\theta}\mathbf{x})(e^{j\theta}\mathbf{x}^T)] &= e^{j2\theta}\tilde{\mathbf{C}}_x = \tilde{\mathbf{C}}_x \end{aligned}$$

$T$  rappresenta sempre la semplice trasposizione.

Queste due relazioni comportano che

$$\begin{aligned} E[\mathbf{x}] &= \mathbf{0} \\ \tilde{\mathbf{C}}_x &= \mathbf{0} \end{aligned}$$

Va ricordato che un vettore gaussiano complesso proprio a valor medio nullo è circularmente simmetrico e viene indicato come  $\mathbf{x} \approx \mathcal{CN}(\mathbf{0}, \mathbf{C}_x)$ .

Per questo tipo di vettori gaussiani complessi la densità di probabilità si potrà quindi esprimere sempre con la relazione:

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

a patto di interpretare  $\mathbf{x}^T$  non come la semplice trasposizione di  $\mathbf{x}$ , ma come la composizione di trasposizione e coniugio (spesso indicato nei testi come  $\mathbf{x}^H$ ). Per evitare confusioni cercheremo di utilizzare l'espressione

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

La correttezza della espressione può essere immediatamente verificata (confrontandola con quella che verrebbe scritta se considerassimo il vettore reale di dimensione  $2N$  ottenuto concatenando parte reale ed immaginaria) nel caso il vettore considerato non sia solo circolare, ma anche con  $\mathbf{C}_{RI} = \mathbf{0}$ . La scelta di utilizzare l'espressione precedente permette di usare la stessa relazione per descrivere in uguale forma la statistica di vettori gaussiani reali e complessi (se hanno simmetria circolare). Nel seguito quando si farà riferimento a vettori gaussiani complessi si avrà sempre a che fare con vettori a simmetria circolare.

Si ricordi comunque che in ogni caso la statistica dei moduli delle componenti di un vettore gaussiano complesso non è gaussiana.

Dato un vettore  $\mathbf{x} \approx \mathcal{CN}(\mathbf{0}, \mathbf{C})$  ed applicandogli una trasformazione lineare reversibile  $A$  si ottiene  $\mathbf{y} = \mathbf{A}\mathbf{x}$  con  $\mathbf{y} \approx \mathcal{CN}(\mathbf{0}, \mathbf{A}\mathbf{C}\mathbf{A}^H)$  dove  $H$  rappresenta sempre la combinazione di trasposizione e coniugio.

Assumendo il vettore casuale  $\mathbf{x}$  ottenuto campionando a passo regolare un processo casuale complesso il cui l'asse della variabile indipendente sia stato normalizzato in modo da avere il passo di campionamento unitario avremo che le componenti del vettore  $\{x_n\}$ ,  $n = 0 \dots N$  rappresentano  $N$  "campioni" complessi del processo casuale. Se il processo casuale da cui è estratto il vettore è stazionario possiamo definire per questo processo una funzione di autocorrelazione (od anche autocovarianza se il valor medio è nullo), ed in particolare possiamo considerare la versione campionata (passo 1) di questa autocorrelazione:  $r_x(m) = E[x(n)x(n-m)^H]$  ( $H$  rappresenta il complesso coniugato). Anche in questo caso abbiamo che la matrice  $\mathbf{C}_x$  può essere scritta sulla base della funzione di autocorrelazione come:

$$\mathbf{C}_x = \begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) & \dots & r_x(1-N) \\ r_x(1) & r_x(0) & r_x(-1) & \dots & \dots \\ r_x(2) & r_x(1) & r_x(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_x(N-1) & \dots & \dots & \dots & r_x(0) \end{bmatrix} \quad (10)$$

La matrice risulta ovviamente a simmetria Hermitiana:  $r_x(-i)$  corrisponde a  $r_x(i)$  complesso coniugato ed inoltre è di "Toeplitz".

## 1.2 Covarianza di sequenze filtrate

Consideriamo una sequenza di campioni  $\{y(n)\}_{n=0}^{N-1}$  ottenuta da un processo casuale ed ordinata in un vettore  $\mathbf{y}$ . Tale sequenza è generata dalla convoluzione tra una sequenza (ordinata in un vettore  $\mathbf{x}$ ) associata ad un processo bianco (e quindi anche a media 0) ed un filtro che ha risposta all'impulso lunga  $M$  campioni,  $\mathbf{h} = \{h(n)\}_{n=0}^{M-1}$ . La convoluzione è valutata per  $N$  campioni e sia anche  $N \gg M$ . Ponendo  $r(\tau) = E[y(n)y(n-\tau)^H]$ , la matrice di covarianza del vettore  $\mathbf{C}_y$  sarà:

$$\mathbf{C}_y = \begin{bmatrix} r_y(0) & r_y(-1) & r_y(-2) & \dots & r_y(1-N) \\ r_y(1) & r_y(0) & r_y(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_y(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_y(N-1) & \dots & \dots & \dots & r_y(0) \end{bmatrix} \quad (11)$$

$$\begin{aligned} y(n) &= \sum_{m=0}^{M-1} x(n-m)h(m); \\ r(\tau) &= E[y(n)y(n-\tau)^H] \\ &= E \left[ \sum_{m=0}^{M-1} x(n-m)h(m) \sum_{l=0}^{M-1} x^H(n-\tau-l)h^H(l) \right] \end{aligned}$$

sulla base del fatto che  $x(n)$  è bianco  $E[x(n-m)x^H(n-\tau-l)]$  è diverso da zero (ed uguale a  $\sigma_x^2$ ) solo per  $n-m = n-\tau-l$  cioè quando  $l = m - \tau$ . Sulla base di ciò

$$\begin{aligned} r(\tau) &= \sigma_x^2 \sum_{m=1}^M h(m)h(m-\tau)^H \\ &= r(0) h(\tau) * h(-\tau)^H = \sigma_x^2 h(\tau)h(-\tau)^H = \sigma_x^2 r_h(\tau) \end{aligned}$$

Dove  $r_h(\tau)$  è l'autocorrelazione deterministica della risposta all'impulso del filtro considerato e  $H$  indica in questo caso il semplice complesso coniugato.

Le righe di  $\mathbf{C}_y$  contengono la replica di  $r_y(\tau)$  funzione di autocorrelazione della sequenza  $y(n)$  sfalsata progressivamente di un passo ogni riga.  $\mathbf{C}_y$  avrà cioè struttura di Toeplitz, e cioè con gli elementi della diagonale principale tutti eguali tra loro ed eguali ad  $r_0$ , gli elementi delle subdiagonali ancora eguali tra loro ma eguali a  $r_1, r_2$  ecc. Si osservi, che nel caso di variabili casuali complesse,  $r(-n) = r(n)^H$  e quindi la matrice di covarianza risulta a simmetria hermitiana. Poichè  $\mathbf{y}$  è ottenuto filtrando un processo bianco la sua autocovarianza si riduce a zero per  $\tau = M$  ben prima dell' $N$ -esimo campione, e cioè :

$$r_y(M) = 0; \quad M \ll N \quad (12)$$

gli elementi della matrice  $\mathbf{C}_y$  saranno nulli al di fuori di una banda di diametro  $2M$  attorno alla diagonale principale (banded matrix, in inglese). Infine va ricordato che la struttura di  $\mathbf{C}_y$  dipende solo dalle caratteristiche del filtro di coloritura ( $h(m)$ ) essendo bianco (cioè con densità spettrale di potenza costante) il processo che alimenta il filtro stesso.

### 1.3 Matrici di covarianza: autovalori ed autovettori

Si consideri una matrice di covarianza  $\mathbf{C}_x$  Si ha:

$$\mathbf{C}_x = E[\mathbf{x} \mathbf{x}^H]$$

la matrice è simmetrica se la sequenza di partenza è reale; altrimenti è a simmetria Hermitiana (la trasposta è eguale alla coniugata). Ricordiamo che una matrice  $\mathbf{C}_x$  ha autovalori  $\lambda_i$  e autovettori colonna  $\mathbf{e}_i$  tali che:

$$\mathbf{C}_x \mathbf{e}_i = \lambda_i \mathbf{e}_i \quad (13)$$

In forma matriciale si ha:

$$\mathbf{E}^H \mathbf{C}_x \mathbf{E} = \mathbf{\Lambda}; \quad \mathbf{E} = [ \mathbf{e}_1 \quad \mathbf{e}_2 \quad \dots \quad \mathbf{e}_{N-1} \quad \mathbf{e}_N ]; \quad \mathbf{E}^{-1} = \mathbf{E}^H \quad (14)$$

dove la matrice **ortonormale**  $\mathbf{E}$  corrisponde agli  $N$  autovettori colonna  $\mathbf{e}_i$ , mentre la matrice  $\mathbf{E}^H$  agli autovettori riga  $\mathbf{e}_i^H$ . Essa è sia l'inversa sia la trasposta (con eventuale coniugio) di  $\mathbf{E}$ . Infine la matrice  $\mathbf{\Lambda}$  è diagonale, e contiene gli autovalori della matrice  $\mathbf{C}_x$ .

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



Mostriamo ora come tutti i  $\lambda_i$  siano positivi o al più nulli e come, conseguentemente, la matrice  $\mathbf{C}$  sia semidefinita positiva. Sia  $\mathbf{a}$  una qualsiasi combinazione lineare degli elementi del vettore  $\mathbf{x}$ . Il valore quadratico medio di questa combinazione lineare è ovviamente maggiore od eguale a zero ( $E[\mathbf{a}^H \mathbf{x} (\mathbf{a}^H \mathbf{x})^H] \geq 0$ ). Si ha quindi:

$$E[\mathbf{a}^H \mathbf{x} (\mathbf{a}^H \mathbf{x})^H] = \mathbf{a}^H E[\mathbf{x} \mathbf{x}^H] \mathbf{a} = \mathbf{a}^H \mathbf{C}_x \mathbf{a} \geq 0 \quad (15)$$

La relazione  $\mathbf{a}^H \mathbf{C}_x \mathbf{a} \geq 0$  è esattamente la definizione di matrice semidefinita positiva.

Se  $\mathbf{a}$  è un autovettore della matrice  $\mathbf{C}_x$  risulta:

$$\mathbf{e}_i^H \mathbf{C}_x \mathbf{e}_i = \lambda_i \geq 0 \quad (16)$$

e quindi tutti gli autovalori di una matrice di covarianza sono maggiori o eguali a 0. Inoltre, poichè per matrici con simmetria Hermitiana è sempre garantito che gli autovettori, associati ad autovalori distinti, siano fra loro ortogonali risulta:

$$\mathbf{e}_i^H \mathbf{e}_j = 0 \rightarrow \mathbf{e}_i^H \mathbf{e}_j = 0 \rightarrow \mathbf{e}_i^H \mathbf{C}_x \mathbf{e}_j = \lambda_j \mathbf{e}_i^H \mathbf{e}_j = \lambda_i \mathbf{e}_i^H \mathbf{e}_j = 0$$

Consideriamo ora una sequenza  $\{x(n)\}$  complessa di cui si conoscono i valori entro una finestra di  $N$  campioni. Essa viene convoluta circolarmente con una risposta all'impulso  $\{h(n)\}$  (in generale complessa) con lunghezza minore od uguale ad  $N$ . E' facile verificare come valgano le seguenti relazioni

$$y(n) = x(n) \otimes h(n) = \sum_k (x(n) * h(n)) * \delta(n - kN) \longleftrightarrow Y(k) = X(k) \cdot H(k)$$

dove

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j2\pi \frac{nk}{N}} = \sum_{n=0}^{N-1} x(n) W_N^{nk} \quad (17)$$

rappresenta la DFT su  $N$  punti della sequenza  $\{x(n)\}_{n=0}^{N-1}$ . In forma matriciale si ha

$$\mathbf{X} = \mathbf{W}_N \mathbf{x}$$

con

$$\mathbf{W}_N = \begin{bmatrix} W_N^{00} & W_N^{10} & W_N^{20} & \dots & W_N^{(N-1)0} \\ W_N^{01} & W_N^{11} & W_N^{21} & \dots & W_N^{(N-1)1} \\ W_N^{02} & W_N^{12} & W_N^{22} & \dots & W_N^{(N-1)2} \\ \dots & \dots & \dots & \dots & \dots \\ W_N^{0(N-2)} & W_N^{1(N-2)} & W_N^{2(N-2)} & \dots & W_N^{(N-1)(N-2)} \\ W_N^{0(N-1)} & W_N^{1(N-1)} & W_N^{2(N-1)} & \dots & W_N^{(N-1)(N-1)} \end{bmatrix}$$

Si noti che  $\mathbf{W}_N$  è una matrice complessa simmetrica (non Hermitiana). La trasformata inversa è definita come

$$\mathbf{x} = \frac{1}{N} \mathbf{W}_N^H \mathbf{X}$$

conseguentemente

$$\mathbf{W}_N^H \mathbf{W}_N = N \mathbf{I}_N = \mathbf{W}_N \mathbf{W}_N^H$$

ovvero  $\mathbf{W}_N$  è una base ortogonale nello spazio complesso.

La convoluzione circolare sopraindicata nel dominio trasformato è data da

$$\mathbf{y} = \frac{1}{N} \mathbf{W}_N^H \mathbf{Y} = \frac{1}{N} \mathbf{W}_N^H (\mathbf{H}\mathbf{X})$$

con

$$\mathbf{H} = \text{diag}\{H(0), H(1), \dots, H(N-1)\}$$

Si assuma il vettore di ingresso  $\{x(n)\}$  come generato da una sequenza di variabili gaussiane complesse iid a valor medio nullo che costituiscano un vettore gaussiano proprio circolarmente simmetrico  $\mathbf{x} \sim \mathcal{CN}(\mathbf{0}, \sigma_x^2 \mathbf{I}_N)$ . Applicando la trasformazione lineare  $\mathbf{W}_N$  e ricordando quanto detto per la trasformazione lineare di vettori di variabili aleatorie complesse gaussiane a simmetria circolare si ha:  $\mathbf{X} \sim \mathcal{CN}(\mathbf{0}, N\sigma_x^2 \mathbf{I}_N)$ . Dove  $\mathbf{X}$  rappresenta la DFT del vettore considerato e può essere vista anch'essa come un vettore gaussiano circolarmente simmetrico in cui le singole componenti sono ancora iid ma con la matrice autocorrelazione/covarianza data da  $\mathbf{C}_X = \mathbf{C}_X = N\sigma_x^2 \mathbf{I}_N$ . Considerando  $\mathbf{X} = \mathbf{X}_R + j\mathbf{X}_I$  si avrà che  $\mathbf{X}_R$  è indipendente da  $\mathbf{X}_I$  inoltre risulta

$$\mathbf{X}_R = \begin{bmatrix} X_R[0] \\ X_R[1] \\ \vdots \\ X_R[N-1] \end{bmatrix} \sim N\left(\mathbf{0}, \frac{N\sigma_x^2}{2} \mathbf{I}_N\right)$$

$$\mathbf{X}_I = \begin{bmatrix} X_I[0] \\ X_I[1] \\ \vdots \\ X_I[N-1] \end{bmatrix} \sim N\left(\mathbf{0}, \frac{N\sigma_x^2}{2} \mathbf{I}_N\right)$$

Passando ora al vettore  $\mathbf{y}$  abbiamo che la sua matrice di autocorrelazione data da:

$$\mathbf{C}_y = \begin{bmatrix} r_y(0) & r_y(-1) & r_y(-2) & \dots & r_y(1-N) \\ r_y(1) & r_y(0) & r_y(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_y(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_y(N-1) & \dots & \dots & \dots & r_y(0) \end{bmatrix}$$

non è solo a simmetria Hermitiana ( $\mathbf{C} = \mathbf{C}^H$ ) e di Toeplitz (gli elementi su ciascuna diagonale sono uguali fra loro), ma anche circolante: ogni riga della matrice è ottenuta da quella

precedente tramite lo spostamento a destra di un elemento, e l'ultimo elemento della riga precedente diventa il primo della successiva. Inoltre per per la prima riga della matrice, come riga precedente va presa l'ultima riga della matrice stessa. Ciò è dovuto al fatto che  $\mathbf{y}$  è stato ottenuta attraverso una convoluzione circolare e quindi è come se fossimo in presenza di una ripetizione periodica del vettore stesso. Ciò porta al fatto che  $r_y(1 - N) = r_y(1)$ ,  $r_y(2 - N) = r_y(2)$  e così via. In totale

$$\mathbf{C}_y = \begin{bmatrix} r_y(0) & r_y(-1) & r_y(-2) & \dots & r_y(1 - N) \\ r_y(1 - N) & r_y(0) & r_y(-1) & \dots & \dots \\ r_y(2 - N) & r_y(1) & r_y(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_y(-1) & \dots & \dots & \dots & r_y(0) \end{bmatrix}$$

La DFT del vettore  $\mathbf{y}$  ( $\mathbf{Y}$ ) può essere vista come ottenuta per trasformazione lineare da  $\mathbf{X}$ :  $\mathbf{Y} = \mathbf{H}\mathbf{X}$ . Ricordando, ancora una volta, quanto detto per la trasformazione lineare di vettori di variabili aleatorie complesse gaussiane a simmetria circolare si ha

$$\mathbf{C}_Y = E[\mathbf{Y}\mathbf{Y}^H] = N\sigma_x^2\mathbf{H}\mathbf{H}^H$$

Poichè  $\mathbf{H}$  è diagonale, lo sarà anche  $\mathbf{H}\mathbf{H}^H$  e quindi anche  $\mathbf{C}_Y$ . Sulla base di ciò è quindi possibile scrivere

$$\mathbf{C}_Y = E[\mathbf{Y}\mathbf{Y}^H] = \mathbf{W}_N \cdot \mathbf{C}_y \cdot \mathbf{W}_N^H = N\sigma_x^2 \text{diag}\{|H(0)|^2, |H(1)|^2, \dots, |H(N-1)|^2\}$$

Quest'ultima relazione mostra come le colonne (e quindi per simmetria anche le righe) della matrice di anti-trasformazione di Fourier siano proporzionali agli autovettori della matrice di autocorrelazione-autocovarianza  $\mathbf{C}_y$ . Per fare in modo che i vettori trovati siano non solo proporzionali agli autovettori di  $\mathbf{C}_y$ , ma coincidano con essi, il fattore  $N$  presente nella IDFT va ripartito fra trasformata ed antitrasformata si ha quindi

$$\mathbf{V}_N^H \cdot \mathbf{C}_{yy} \cdot \mathbf{V}_N = N\sigma_x^2 \text{diag}\{|H(0)|^2, |H(1)|^2, \dots, |H(N-1)|^2\}$$

ove le colonne (e le righe) di  $\mathbf{V}_N$  sono date da

$$\mathbf{v}_k = \frac{1}{\sqrt{N}} \left[ 1, \exp(j\frac{2\pi k}{N}), \dots, \exp(j\frac{2\pi k(N-1)}{N}) \right]^T$$

E' immediato verificare che l'energia (modulo quadro) del vettore sopraindicato è unitaria (vettore normale).

Nel caso il vettore  $\mathbf{x}$  sia costituito da variabili gaussiane iid a valor medio nullo *reali*, la sua DFT presenterà una simmetria di tipo complesso coniugato e conseguentemente solo  $N/2 + 1$  ( $N$  pari) o  $(N + 1)/2$  ( $N$  dispari) coefficienti della DFT devono essere considerati (gli altri sono semplicemente i complessi coniugati degli altri). Sulla base di ciò consideriamo come vettore trasformato  $\mathbf{X}$  quello che contiene gli  $N/2 + 1$  ( $N$  pari) o  $(N + 1)/2$  ( $N$  dispari) elementi significativi della DFT. Di essi quello di indice 0 (frequenza 0) e indice  $N/2$  (frequenza di Nyquist,  $f_c/2$ ), esistente se  $N$  è pari, sono reali e modellizzabili entrambi come

variabili casuali gaussiane reali ( $\mathcal{N}(0, N\sigma_x^2)$ ). Gli altri coefficienti sono complessi e godono delle stesse proprietà indicate precedentemente ed in particolare il vettore che li rappresenta è modellizzabile come  $\mathcal{CN}(\mathbf{0}, N\sigma_x^2\mathbf{I})$ . Se ora si considera  $y(n) = x(n) \circledast h(n)$  con  $h(n)$  reale, il vettore rappresentante la sua DFT,  $\mathbf{Y}$  (con  $N/2+1$  o  $(N+1)/2$  elementi per la sua simmetria complessa coniugata) gode delle stesse proprietà viste nel caso di dati complessi, dal vettore  $\mathbf{Y}$  vanno però esclusi gli elementi di indice 0 e  $N/2$  (se presente) che sono anch'essi reali e vanno "gestiti separatamente". Se ora pensiamo di costruire una matrice reale diagonale che abbia come elementi tutte le parti reali ed immaginarie degli  $N/2 + 1$  ( $(N+1)/2$ ) elementi significativi di  $\mathbf{Y}$ , tale matrice avrà dimensione  $N \times N$  poichè gli elementi corrispondenti alla frequenza 0 e di Nyquist (se presente) sono puramente reali, e costituirà la matrice degli autovalori di  $\mathbf{C}_y$ . I corrispondenti autovettori sono le componenti reali ed immaginarie delle colonne di  $\mathbf{W}_N$ , cioè tratti di seni e coseni, scalati in modo da avere energia (modulo quadro) unitario.

Quando il vettore  $\mathbf{y}$  è ottenuto considerando  $N$  campioni successivi di un segnale qualsiasi (reale o complesso) comunque a media nulla, la sua matrice di autocorrelazione-autocovarianza non sarà circolante, gli autovalori saranno comunque non negativi e gli autovettori saranno comunque ortogonali tra loro: non saranno però proporzionali alle righe di  $\mathbf{W}_N$  (o alle loro componenti reali ed immaginarie, nel caso di  $\mathbf{y}$  reale), ma assomiglieranno abbastanza ad esse. Inoltre la matrice di autocorrelazione-autocovarianza  $\mathbf{Y}$  sarà solo "quasi diagonale". Tutto ciò è dovuto al fatto che, anche facendo una modellizzazione semplice del segnale sotto esame, non si potrà supporre che il vettore  $\mathbf{y}$  sia generato da un vettore gaussiano iid a media nulla ( $\mathbf{x}$ ) attraverso una convoluzione circolare, ma bensì si dovrà considerare una convoluzione lineare. E' però vero che convoluzione circolare e lineare tendono ad assomigliarsi molto se il vettore considerato è di lunghezza molto molto superiore alla durata della risposta all'impulso considerata ( $\mathbf{h}$ ).

Va inoltre notato che se la matrice di autocorrelazione è ottenuta in modo campionario cioè utilizzando  $\tilde{r}_y(k) = \frac{1}{N-k} \sum_{n=k}^{N-1} y(n+k)y(n)^H$ , anche assumendo  $y(n)$  ergodico, la matrice di autocorrelazione così calcolata ( $\tilde{\mathbf{C}}_y$ ) sarà solo una approssimazione della "vera" matrice di autocorrelazione data da  $\mathbf{C}_y = E[\mathbf{C}\mathbf{C}^H]$ .

Infine qualora si calcolino autovettori ed autovalori di  $\mathbf{C}_y$  tramite librerie/tool matematici, l'ordinamento è normalmente per valori decrescenti degli autovettori e quindi, anche nel caso di matrice effettivamente circolante, l'ordinamento non è per "frequenze" crescenti degli autovettori. Ciò a maggior ragione nel caso in cui la matrice di autocorrelazione non sia circolante e quindi non si possa associare una "frequenza" a ciascun autovettore.

## 1.4 Un esempio numerico

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

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}; \quad E[\mathbf{x}\mathbf{x}^H] = \mathbf{C}_x = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} = \mathbf{I}$$

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

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

ed otteniamo il vettore  $\mathbf{y}$  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:

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

dove

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

Calcoliamo ora  $\mathbf{C}_y$ , la matrice di autocovarianza di  $\mathbf{y}$ , circolante per la ciclicità della convoluzione:

$$\mathbf{C}_y = E[\mathbf{y} \mathbf{y}^H] = \mathbf{A} \mathbf{C}_x \mathbf{A}^H = \begin{bmatrix} 2 & 1 & 0 & 1 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 1 & 0 & 1 & 2 \end{bmatrix}$$

Il filtraggio passa - basso fatto con la convoluzione con il filtro  $\mathbf{f}$  comporta uno zero alla frequenza di Nyquist, visto che il numero dei campioni della sequenza è pari. Infatti il vettore  $\mathbf{F}$  trasformata di Fourier di  $\mathbf{f}$ , è:

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

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

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

Si ha ancora:

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

Peraltro:

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

e quindi

$$E [\mathbf{Y} \mathbf{Y}^H] = \mathbf{C}_Y = 4 \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$$

Da un'analisi diretta, è peraltro evidente che la matrice  $\mathbf{C}_Y$  ha un autovalore nullo. Infatti, ad esempio la sua ultima colonna può essere ottenuta come combinazione lineare delle altre 3. L'autovettore corrispondente a questo autovalore nullo è la sinusoide alla frequenza di Nyquist.

$$\mathbf{C}_Y \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix} = 0 \times \begin{bmatrix} 1 \\ -1 \\ 1 \\ -1 \end{bmatrix}$$

Per avere gli altri autovalori ed autovettori di  $\mathbf{C}_Y$ , si può ricorrere alla trasformazione di Fourier discreta che diagonalizza la matrice. In questo caso la matrice trasformazione è

$$\mathbf{W} = \begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -j & -1 & j \\ 1 & -1 & 1 & -1 \\ 1 & j & -1 & -j \end{bmatrix}$$

siamo in presenza di dati puramente reali e quindi gli autovettori della matrice  $\mathbf{C}_Y$ , sono corrispondenti dalle componenti reali ed immaginarie delle prime  $N/2 + 1 = 3$  colonne di  $\mathbf{W}$  normalizzate per avere modulo unitario. In particolare essi saranno proporzionali a: prima colonna (reale), parte reale seconda colonna, parte immaginaria seconda colonna, terza colonna (reale)

$$\tilde{\mathbf{V}} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 1 & 0 & -1 & -1 \\ 1 & -1 & 0 & 1 \\ 1 & 0 & 1 & -1 \end{bmatrix}$$

Effettuando la normalizzazione si ottiene:

$$\mathbf{V} = \begin{bmatrix} \frac{1}{2} & \frac{1}{\sqrt{2}} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & -\frac{1}{\sqrt{2}} & -\frac{1}{2} \\ \frac{1}{2} & -\frac{1}{\sqrt{2}} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{\sqrt{2}} & -\frac{1}{2} \end{bmatrix}$$

Infine la matrice degli autovettori si ottiene come

$$\mathbf{\Lambda} = \mathbf{V}^T \mathbf{C}_y \mathbf{V} = \begin{bmatrix} 4 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

## 2 Introduzione al problema della stima

All'interno della teoria della stima il caso classico che viene affrontato è quello in cui l'osservazione è affetta da rumore additivo. In funzione del contesto l'osservazione  $x[i]$  può essere indicata come *misura*, segnale acquisito, *dato* od *osservazione*. In ogni caso le  $x[i]$  rappresentano le quantità disponibili da cui stimare i parametri (valori) incogniti.

Nel caso di rumore additivo l'osservazione è data da:

$$x[i] = s[i; \theta] + w[i]$$

e risulta funzione di un set di parametri  $\theta$  tale che  $\theta \in \times$  (l'insieme dei parametri ammissibili) e di un termine additivo  $w(i)$  (rumore). Entrambi questi termini (parametri e rumore) sono non osservabili direttamente. Un problema classico della teoria della stima è rappresentato dalla necessità di stimare  $\theta$  a partire da  $N$  osservazioni  $\mathbf{x}_N = \{x[i]\}_{i=1}^N$  (ottenute nei modi più diversi a seconda del problema considerato). Il pedice  $N$  nella rappresentazione del vettore vuole ricordare la dimensionalità del vettore stesso, ma nel seguito verrà frequentemente omesso. Lo stimatore rappresenta la funzione (in generale non-lineare) che lega le osservazioni  $\mathbf{x}_N$  alla stima  $\hat{\theta}$  di  $\theta$ :

$$\hat{\theta} = \hat{\theta}[\mathbf{x}_N],$$

la stima è una vc in quanto funzione di  $N$  vc  $\mathbf{x}_N$ . L'accuratezza di uno stimatore si esprime valutando l'errore di stima rappresentato dalla differenza

$$\delta\hat{\theta}(\mathbf{x}_N) = \hat{\theta}(\mathbf{x}_N) - \theta,$$

si noti che anch'esso è una vc. Tra le proprietà che ci si attende da uno stimatore sono da considerare le seguenti:

- **Polarizzazione:** quando la vc  $\delta\hat{\theta}(\mathbf{x}_N)$  ha valore medio nullo (ovvero in media il valore stimato coincide con il valore  $\theta$ ) si parla di stimatore non-polarizzato. In alternativa si ha uno stimatore polarizzato (proprietà questa non sempre gradita). Anche se uno stimatore è polarizzato ci si attende almeno che la polarizzazione svanisca per  $N \rightarrow \infty$  (stimatore asintoticamente non-polarizzato).
- **Consistenza:** aumentando il numero delle misure lo stimatore diventa sempre più accurato. In sintesi, per uno stimatore consistente, per  $N \rightarrow \infty$ , si deve avere  $\delta\hat{\theta}(\mathbf{x}_N) \rightarrow 0$  secondo qualche norma. Di solito si utilizza l'errore quadratico medio  $MSE = E[|\delta\hat{\theta}(\mathbf{x}_N)|^2]$  come norma in quanto comodo per le derivazioni analitiche e facilmente controllabile per simulazione.

Sono possibili due modelli per la stima del parametro  $\theta$  a seconda che quest'ultimo sia deterministico (o non-casuale) di cui non è noto il valore "vero" del parametro  $\theta_o$  (approccio non-Bayesiano) oppure  $\theta$  sia una vc di cui è nota la sola pdf  $p[\theta]$  detta anche pdf a-priori (approccio Bayesiano):

- **Approccio non-Bayesiano** ( $\theta$  è una variabile deterministica o parametro) - E' possibile ottenere la pdf delle misure condizionata al parametro  $\theta$  *non noto* e verificare



quale valore del parametro la massimizza. La pdf funzione del parametro  $\theta$  è nota come funzione di verosimiglianza (o *likelihood*) del parametro

$$p[\mathbf{x}; \theta]$$

La stima a massima verosimiglianza si basa sulla “conferma” dalle osservazioni del modello parametrico della pdf.

- **Approccio Bayesiano** ( $\theta$  è una vc) - Nota la pdf a-priori del parametro  $p[\theta]$  è possibile ottenere la pdf a-posteriori mediante il prodotto con la pdf condizionale  $p[\mathbf{x}|\theta]$  (ovvero la pdf delle osservazioni condizionata al valore del parametro  $\theta$ ):

$$p[\theta|\mathbf{x}] = \frac{p[\mathbf{x}|\theta]p[\theta]}{p[\mathbf{x}]} = \gamma p[\mathbf{x}|\theta]p[\theta],$$

$\gamma$  è una costante di proporzionalità che segue dalla normalizzazione della pdf  $p[\theta|\mathbf{x}]$ . Dalla  $p[\theta|\mathbf{x}]$  è possibile cercare il valore più probabile (stima che massimizza la probabilità a-posteriori - MAP) o il suo valore medio (stima MMSE) e definire da questi i corrispondenti stimatori.

Va poi ricordato come, nel caso la stima venga espressa come combinazione lineare delle osservazioni, si parli di (**stimatori lineari**)

$$\hat{\theta} = \mathbf{A}\mathbf{x}_N$$

con  $\mathbf{A} \in R^{p \times N}$  o  $R^{p \times N}$ , con  $p$  numero totale di parametri da stimare e  $N$  numero delle osservazioni a disposizione. Gli stimatori lineari sono, in generale, semplici e quindi molto diffusi, hanno inoltre proprietà interessanti e, talvolta, hanno anche il pregio dell’ottimalità.

**Cramér-Rao bound (CRB):** Un problema che si presenta è quello di confrontare tra di loro due (o più) stimatori per decidere quale adottare. L’analisi (e il confronto) dell’errore quadratico medio (MSE) fra valore vero del vettore di parametri e quello stimato rappresenta un buon indicatore nella maggior parte dei casi, scegliere lo stimatore con il MSE più basso a parità di numero di osservazioni  $N$  (o più *efficiente* visto che usa meglio le osservazioni disponibili) è sempre una buona scelta (anche se non è sempre la più economica). Tuttavia sorgono alcune domande: fino a che punto spingere la ricerca di uno stimatore efficiente? esiste forse uno stimatore più efficiente di quello che stò usando? Sono domande delicate perchè pongono il problema di sapere se lo stimatore di cui dispongo è o meno il *migliore* e/o se vale la pena investire per cercarne uno ancor più efficiente. La risposta a questo problema viene dalla teoria in quanto è possibile valutare il limite teorico del MSE (o della covarianza) sotto il quale non è possibile scendere. In altre parole, è possibile valutare il limite di Cramer-Rao (o Cramer-Rao bound - CRB)  $MSE_{CRB}$  tale che

$$E[|\delta\hat{\theta}(\mathbf{x}_N)|^2] \geq MSE_{CRB}.$$

Se lo stimatore è tale per cui  $E[|\delta\hat{\theta}(\mathbf{x}_N)|^2] \simeq MSE_{CRB}$  allora si dice che lo stimatore è **efficiente** (asintoticamente efficiente se  $E[|\delta\hat{\theta}(\mathbf{x}_N)|^2] \simeq MSE_{CRB}$  per  $N \rightarrow \infty$ ) e *non* è

*necessario cercarne di meglio* (a meno di farlo per contenere il costo computazionale). Se  $E[||\delta\hat{\theta}(\mathbf{x}_N)||^2] \gg MSE_{CRB}$  c'è ampio spazio di miglioramento. Il limite di Cramer-Rao è stato definito originariamente per stimatori non-Bayesiani (in cui cioè  $\theta$  è deterministico e non noto), ma sono state sviluppate sue estensioni anche al caso di stimatori Bayesiani. Dettagli ed applicazioni del limite di Cramer-Rao sono presentati nel corso di Elaborazione Numerica dei Segnali della Laurea Magistrale in Ing delle TLC.

### 3 Modelli lineari

Un modello lineare è tale quando le osservazioni hanno una dipendenza lineare dai parametri

$$\begin{bmatrix} x[0] \\ \vdots \\ x[N-1] \end{bmatrix} = \begin{bmatrix} h_{11} & \dots & h_{1p} \\ \vdots & \ddots & \vdots \\ h_{N1} & \dots & h_{Np} \end{bmatrix} \begin{bmatrix} \theta_1 \\ \vdots \\ \theta_p \end{bmatrix} + \begin{bmatrix} w[0] \\ \vdots \\ w[N-1] \end{bmatrix}$$

si noti che la numerazione dei campioni in  $\mathbf{x}$  non è rilevante ai fini del ragionamento ma dipende solo dal problema specifico. In modo compatto la relazione precedente può essere scritta come

$$\mathbf{x} = \mathbf{H}\theta + \mathbf{w}$$

La *matrice delle osservazioni*  $\mathbf{H} \in R^{N \times p}$  (o in generale  $\mathbf{H} \in C^{N \times p}$ ) è nota dal problema, il rumore è caratterizzato dalla pdf anch'essa nota e il vettore  $\theta \in R^p$  (o in generale  $\theta \in C^p$ ) ordina l'insieme delle grandezze da stimare.

Considerando dati e parametri reali, se il rumore è gaussiano allora

$$\mathbf{w} \sim N(\mu_w, \mathbf{C}_w)$$

in cui  $\mu_w$  e  $\mathbf{C}_w$  sono note dal problema. In questo caso

$$\mathbf{x} \sim N(\mu_w + \mathbf{H}\theta, \mathbf{C}_w)$$

#### 3.1 Modello di regressione lineare

In questo caso i dati hanno valori legati, attraverso una costante, al valore della variabile indipendente cui si riferiscono. A tali valori si aggiunge un termine di disturbo. In particolare dato l'insieme dei valori della variabile indipendente cui i dati si riferiscono ( $\{t_1, t_2, \dots, t_N\}$ ), si ha:

$$x(t_n) = A + Bt_n + w(t_n) \quad n = 1, 2, \dots, N.$$

In particolare, ci si riconduce al modello compatto per  $\mathbf{x} = [x(t_1), x(t_2), \dots, x(t_N)]^T$  definendo

$$\theta = [A, B]^T$$

$$\mathbf{H} = \begin{bmatrix} 1 & t_1 \\ 1 & t_2 \\ \vdots & \vdots \\ 1 & t_N \end{bmatrix}$$

la matrice delle osservazioni è ulteriormente semplificata nel caso di campionamento uniforme  $t_k = (k - 1) \cdot \Delta t$

$$\mathbf{H} = \begin{bmatrix} 1 & 0 \\ 1 & \Delta t \\ \vdots & \vdots \\ 1 & (N - 1) \cdot \Delta t \end{bmatrix}.$$

La stima dei parametri della retta  $\theta = [A, B]^T$  note le osservazioni è indicato come problema di *regressione lineare*. Esistono una serie di problemi quali: la stima di frequenza in segnali sinusoidali, dei ritardi (p.e., nei sistemi radar) e delle direzioni di arrivo in schiere lineari uniformi che sono riconducibile al problema della stima dei parametri di regressione lineare.

### 3.2 Filtraggio lineare

Si consideri un filtro con risposta all'impulso  $h[k]$  e si misuri da questo l'uscita rumorosa soggetta a ingresso  $u[n]$

$$x[n] = h[n] * u[n] + w[n]$$

questo è certamente un sistema lineare in cui i parametri sono rappresentati dalla risposta all'impulso del filtro (nei problemi di *identificazione*) o dall'ingresso (nei problemi di *equalizzazione*). Nella definizione dei due modelli si assume il filtro causale e con supporto di  $p$  campioni:

$$x[n] = \sum_{k=0}^{p-1} h[k]u[n - k] + w[n].$$

A seconda di come venga impostato il problema legato al filtraggio lineare si possono scrivere relazioni diverse, ma comunque riconducibili a:

$$\mathbf{x} = \mathbf{h} * \mathbf{u} + \mathbf{w} = \mathbf{H}\mathbf{u} + \mathbf{w} = \mathbf{U}\mathbf{h} + \mathbf{w}$$

a seconda che si voglia stimare la sequenza di ingresso o la risposta all'impulso. Va notato che il prodotto di convoluzione viene scritto come il prodotto fra una matrice ed un vettore colonna. Il vettore colonna contiene una delle due sequenze da convolvere, mentre le righe della matrice contengono versioni traslate della seconda sequenza.

## 4 Stimatore BLUE per modelli lineari

Polarizzazione ed errore quadratico medio (MSE) rappresentano due parametri in grado di caratterizzare l'accuratezza di uno stimatore. E' quindi possibile definire alcuni stimatori semplicemente imponendo che lo stimatore stesso abbia MSE minimo. Data la dipendenza del MSE dalla varianza e polarizzazione un approccio classico assume che lo stimatore sia non-polarizzato e a minima varianza. Si ha quindi il generico stimatore a minima varianza non-polarizzato (*minimum variance unbiased* - MVU) oppure lo stimatore lineare ottimo non-polarizzato (*best linear unbiased estimator* - BLUE) che è un caso particolare di MVU. E' necessario puntualizzare che lo stimatore ad MSE minimo potrebbe avere polarizzazione

non nulla. Ma sicuramente restringere le possibilità di scelta per gli stimatori a solo quelli caratterizzati da polarizzazione nulla è un approccio molto comune.

Si assuma di disporre di un modello lineare in rumore gaussiano (AGN) in cui tutti i termini siano reali

$$\begin{aligned}\mathbf{x} &= \underbrace{[\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_p]}_{\mathbf{H}}\theta + \mathbf{w} \\ \mathbf{w} &\sim N(\mu_w, \mathbf{C}_w)\end{aligned}$$

e si voglia trovare lo stimatore lineare non polarizzato a MSE minimo (stimatore BLUE)

$$\hat{\theta} = \mathbf{A}\mathbf{x} + \mathbf{b}$$

in cui  $\mathbf{A}$  e  $\mathbf{b}$  sono incognite e vanno trovate imponendo proprio l'assenza di polarizzazione ed il vincolo di minimizzazione del MSE.

$$E[\hat{\theta}] = \mathbf{A}E[\mathbf{x}] + \mathbf{b} = \theta$$

ricordando che

$$E[\mathbf{x}] = \mathbf{H}\theta + \mu_w$$

si può ricavare la condizione (o vincolo nel problema di ottimizzazione) di non-polarizzazione

$$\mathbf{A}\mathbf{H}\theta + \mathbf{A}\mu_w + \mathbf{b} = \theta$$

che significa

$$\begin{aligned}\mathbf{A}\mathbf{H} &= \mathbf{I} \leftrightarrow \mathbf{a}_k^T \mathbf{h}_\ell = \delta_{k-\ell} \\ \mathbf{A}\mu_w + \mathbf{b} &= \mathbf{0}\end{aligned}$$

dove  $\mathbf{a}_k^T$  è la k-esima riga della matrice  $\mathbf{A}$  (infatti  $\mathbf{A}^T = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_p] \in R^{N \times p}$ ). In particolare, il secondo vincolo è banale e dipende dal primo (in particolare da  $\mathbf{A}$ ) per cui ci si concentra sul primo vincolo (se si vuole si pensi a  $\mu_w = \mathbf{0}$  e quindi  $\mathbf{b} = \mathbf{0}$ ). La minimizzazione della varianza si ha minimizzando simultaneamente gli MSE ( $MSE_k$ ) relativi ad ogni componente del vettore dei parametri. La soluzione diventa (vedi ad esempio *Fundamentals of Statistical Signal Processing, vol. 1, Estimation Theory, Steven M. Kay, Prentice Hall, 1993*)

$$\hat{\theta} = (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{x}$$

con matrice di covarianza

$$cov[\hat{\theta}] = (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}.$$

Si noti che l'espressione per il calcolo di  $\hat{\theta}$ , nel caso in cui  $\mathbf{C}_w = \sigma_w^2 \mathbf{I}$  si riduce a

$$\hat{\theta} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x}$$

in cui  $(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$  è la cosiddetta pseudoinversa di  $\mathbf{H}$  e corrisponde alla soluzione del problema  $\mathbf{H}\theta = \mathbf{x}$  con il vettore  $\theta$  incognito quando il numero delle equazioni è superiore a quello delle incognite (soluzione ai minimi quadrati). Nel caso la matrice  $\mathbf{H}$  sia quadrata ed invertibile ovviamente  $(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \mathbf{H}^{-1}$ .

#### 4.0.1 Esempio: stimatore BLUE del valore medio

Si assume di disporre di  $N$  misure correlate di un processo con lo stesso valore medio e diversa varianza

$$\mathbf{x} \sim N(a\mathbf{1}, \mathbf{C}_x)$$

si vuole stimare il valore medio mediante uno stimatore lineare:

$$\hat{a} = \alpha^T \mathbf{x} = \sum_{k=1}^N \alpha_k x_k.$$

Per imporre polarizzazione nulla risulta

$$E[\hat{a}] = \sum_{k=1}^N \alpha_k E[x_k] = a \sum_{k=1}^N \alpha_k \rightarrow \sum_{k=1}^N \alpha_k = 1$$

mentre la varianza è

$$\text{var}[\hat{a}] = \alpha^T \text{cov}(\mathbf{x}) \alpha = \alpha^T \mathbf{C}_x \alpha.$$

Minimizzando la varianza con il vincolo  $\alpha^T \mathbf{1} = 1$  (non polarizzazione) si ricava

$$\alpha_{opt} = \frac{\mathbf{C}_x^{-1} \mathbf{1}}{\mathbf{1}^T \mathbf{C}_x^{-1} \mathbf{1}} \rightarrow \hat{a} = \frac{\mathbf{1}^T \mathbf{C}_x^{-1} \mathbf{x}}{\mathbf{1}^T \mathbf{C}_x^{-1} \mathbf{1}}.$$

Il valore minimo della varianza per  $\alpha_{opt}$  diventa

$$\text{var}[\hat{a}]_{\min} = (\mathbf{1}^T \mathbf{C}_x^{-1} \mathbf{1})^{-1}$$

Alcuni casi particolari:

- $\mathbf{C}_x = \sigma_x^2 \mathbf{I}$  (tutti i campioni hanno medesima confidenza e sono tra di loro incorrelati):

$$\alpha_{opt} = \frac{\mathbf{1}}{\mathbf{1}^T \mathbf{1}} = \frac{1}{N} \mathbf{1} \rightarrow \hat{a} = \frac{1}{N} \sum_{k=1}^N x_k$$

si ha la media campionaria in quanto non c'è ragione alcuna di privilegiare alcuni campioni rispetto ad altri.

- $\mathbf{C}_x = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_N^2\}$  (i campioni hanno diverso livello di confidenza ma sono tra di loro incorrelati):

$$\begin{aligned} \mathbf{C}_x^{-1} \mathbf{1} &= \left[ \frac{1}{\sigma_1^2}, \frac{1}{\sigma_2^2}, \dots, \frac{1}{\sigma_N^2} \right]^T \\ \mathbf{1}^T \mathbf{C}_x^{-1} \mathbf{1} &= \sum_{n=1}^N \frac{1}{\sigma_n^2} \\ \alpha_{k,opt} &= \frac{1/\sigma_k^2}{\sum_{n=1}^N 1/\sigma_n^2} \end{aligned}$$

i campioni vengono sommati pesando maggiormente i campioni con varianza più piccola (e quindi più affidabile) e penalizzando i campioni con varianza elevata. *Si pensi a voler mediare i dati da un exit-poll elettorale per minimizzare la varianza della media, nessuno farebbe una media pesando in modo uguale gli exit-poll da un piccolo seggio rurale (con un basso campione statistico) con quelli di una grande città (con grande affluenza).*

- $\mathbf{C}_x = \text{diag}\{2\sigma^2, \sigma^2, \dots, \sigma^2, 2\sigma^2\}$  (due campioni hanno varianza doppia rispetto agli altri e sono tutti incorrelati)

$$\alpha_{k,opt} = \begin{cases} \frac{1}{(N-1)} & \text{per } k \neq 1, N \\ \frac{1/2}{(N-1)} & \text{per } k = 1, N \end{cases}$$

## 5 Stima a massima verosimiglianza (ML)

La stima ML si basa sulla massimizzazione della pdf delle osservazioni  $\mathbf{x}$  parametrizzata in funzione di  $\theta$

$$\hat{\theta} = \theta_{ML} = \arg \max_{\theta} p[\mathbf{x}; \theta].$$

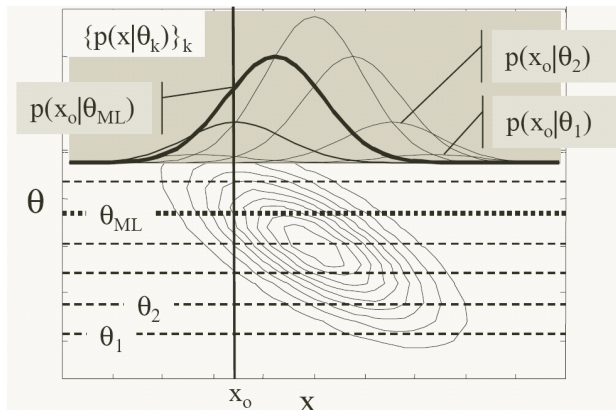


Figure 1: Stimateur ML: Description graphique de l'approche.

La logica è da ricercare nella rappresentazione della pdf  $p[x; \theta]$  in figura per una osservazione  $x = x_o$ . La pdf è funzione sia dell'osservazione che del parametro  $\theta$  (si pensi ad una descrizione parametrica della pdf), per una specifica osservazione  $x = x_o$  si ricerca tra le possibili pdf parametriche rispetto a  $\theta$  quella che ha la maggiore probabilità. Il parametro che corrisponde al massimo della probabilità è il parametro che meglio (o con maggiore

probabilità) descrive la pdf data l'osservazione  $x = x_o$ . La stima  $\theta_{ML}$  è quindi quella che massimizza la probabilità data l'osservazione disponibile (o in altre parole, dall'osservazione  $x$  identifica la pdf visto che questa è parametrica). Il pregio della stima ML è che rappresenta un metodo diretto per la scrittura dello stimatore che, viste le premesse, ha buone caratteristiche in termini di accuratezza. La difficoltà della stima ML risiede nella scrittura della pdf (che per la stima ML è indicata come funzione di verosimiglianza) e nella sua massimizzazione in quanto spesso non-lineare.

**Modello con rumore additivo:** Dato il modello in cui la dipendenza dai parametri non è lineare

$$\mathbf{x} = \mathbf{s}(\theta) + \mathbf{w}$$

(si noti che in questo caso la quantità  $\mathbf{s}(\theta)$  rappresenta una relazione funzionale tra parametri deterministici e quindi non contiene vc, in generale non è sempre così) e la pdf  $p_w[\mathbf{w}]$  è nota, la funzione di verosimiglianza si scrive

$$p[\mathbf{x}; \theta] = p_w[\mathbf{x} - \mathbf{s}(\theta)].$$

Se  $\mathbf{w}$  ha valore medio nullo (se  $\mu_w \neq \mathbf{0}$  questo può essere facilmente accorpato come termine additivo nel modello) la relazione sopra equivale a dire che la pdf di  $\mathbf{x}$  è uguale alla pdf di  $\mathbf{w}$  in cui si è aggiunto un valore medio pari a  $\mathbf{s}(\theta)$ . In alternativa (e in modo più generale) si può scrivere la pdf delle osservazioni in cui l'insieme dei parametri  $\theta$  rientra nella definizione della pdf  $p[\mathbf{x}; \theta]$ . Si noti che di solito si preferisce calcolare il  $\log(\cdot)$  della verosimiglianza in quanto la funzione monotona mantiene inalterata la posizione dell'estremante utile per la stima ML e rimuove la dipendenza esponenziale (questo è particolarmente comodo quando si ha a che fare con una pdf gaussiana) da cui si definisce log-verosimiglianza

$$\mathcal{L}(\mathbf{x}; \theta) = \ln p(\mathbf{x}; \theta).$$

e da questa la stima ML

$$\theta_{ML}(\mathbf{x}) = \arg \max_{\theta} \mathcal{L}(\mathbf{x}; \theta).$$

Alcune proprietà rendono vantaggioso l'uso dello stimatore ML, tra queste ci sono le proprietà asintotiche della stima ML, in particolare la stima ML è non-polarizzata (di sicuro asintoticamente) e l'errore di stima è gaussiano con covarianza pari a quella indicata dal limite di Cramer-Rao  $\mathbf{C}_{CRB}$ , in sintesi

$$\theta_{ML}(\mathbf{x}) \underset{N \rightarrow \infty}{\sim} N(\theta, \mathbf{C}_{CRB}).$$

La definizione dello stimatore ML dipende dalle applicazioni, dai modelli e dal rumore. Di solito il modello gaussiano è quello più ricorrente e permette derivazioni generali.

## 5.1 Modello gaussiano $\mathbf{x} \sim N(\mu(\theta), \mathbf{C}(\theta))$

L'analisi del modello gaussiano generale in cui valore medio  $\mu(\theta)$  e covarianza  $\mathbf{C}(\theta)$  sono funzione di un insieme di parametri è di grande interesse pratico in quanto molti problemi si possono ricondurre a questo modello. Di seguito viene presentato il problema, relativamente semplice, in cui le osservazioni sono reali ed il termine di rumore presente nelle osservazioni stesse è gaussiano e non dipende dei parametri da stimare. Problemi ML più complessi ed il caso di dati/parametri complessi verranno considerati nei corsi di Laurea Magistrale.

### 5.1.1 Modello $\mathbf{x} = \mathbf{s}(\theta) + \mathbf{w}$ con $\mathbf{w} \sim N(\mathbf{0}, \mathbf{C}_w)$

Nel modello gaussiano la pdf è

$$p(\mathbf{x}; \theta) = |2\pi\mathbf{C}_w|^{-1/2} \exp \left\{ -\frac{1}{2} [\mathbf{x} - \mathbf{s}(\theta)]^T \mathbf{C}_w^{-1} [\mathbf{x} - \mathbf{s}(\theta)] \right\}$$

mentre la log-verosimiglianza

$$\mathcal{L}(\mathbf{x}; \theta) = -\frac{N}{2} \ln 2\pi - \frac{1}{2} \ln |\mathbf{C}_w| - \frac{1}{2} [\mathbf{x} - \mathbf{s}(\theta)]^T \mathbf{C}_w^{-1} [\mathbf{x} - \mathbf{s}(\theta)]$$

per cui  $\arg \max_{\theta} p(\mathbf{x}; \theta) = \arg \max_{\theta} \mathcal{L}(\mathbf{x}; \theta)$ . Da questo segue che la stima ML si ottiene minimizzando (si noti il segno meno nell'espressione della log-verosimiglianza) l'errore pesato

$$\theta_{ML} = \arg \min_{\theta} \{ [\mathbf{x} - \mathbf{s}(\theta)]^T \mathbf{C}_w^{-1} [\mathbf{x} - \mathbf{s}(\theta)] \} = \arg \min_{\theta} \|\mathbf{x} - \mathbf{s}(\theta)\|_{\mathbf{C}_w^{-1}}^2.$$

Si noti che la funzione  $\|\mathbf{x} - \mathbf{s}(\theta)\|_{\mathbf{C}_w^{-1}}^2$  non è quadratica (tranne che per modelli lineari) e quindi l'ottimizzazione potrebbe non essere banale.

Per il **modello lineare**  $\mathbf{x} = \mathbf{H}\theta + \mathbf{w}$  con  $\mathbf{w} \sim N(\mathbf{0}, \mathbf{C}_w)$  la stima ML diventa

$$\theta_{ML} = \arg \min_{\theta} \{ [\mathbf{x} - \mathbf{H}\theta]^T \mathbf{C}_w^{-1} [\mathbf{x} - \mathbf{H}\theta] \}$$

si ha minimizzando una forma quadratica da cui (ammettendo che esista l'inversa, tipicamente se il numero di misure è superiore ai parametri)

$$\theta_{ML} = (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{x}$$

ovvero lo stimatore è lineare e coincide con BLUE, la covarianza è

$$\text{cov}[\theta_{ML}] = (\mathbf{H}^T \mathbf{C}_w^{-1} \mathbf{H})^{-1}.$$

## 5.2 Un esempio un po' particolare di stima ML

### 5.2.1 Stima Valore Quadratico Medio Campioni indipendenti

Sia data una sequenza di lunghezza finita di campioni con valori corrispondenti alle uscite di variabili casuali reali gaussiane a media zero e varianza  $\sigma^2$ . Dai valori dei campioni di questa sequenza (in totale  $M$ ) si voglia valutare/stimare  $\sigma^2$  (valore quadratico medio) a priori non noto. Si voglia inoltre valutare la dispersione di tale stima. Si voglia utilizzare la tecnica a Massima Verosimiglianza (ML) a questo scopo, anche se il problema è un po' diverso da quelli presi in considerazione nelle sottosezioni precedenti. Indichiamo con  $\mathbf{x}$  il vettore che raccoglie tutti i campioni ed assumiamo che i campioni siano indipendenti. In questo caso la pdf del vettore è data da:

$$p(\mathbf{x}; \sigma^2) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp \left[ -\frac{1}{2\sigma^2} \sum_{n=0}^{M-1} x_n^2 \right]$$



la log-verosimiglianza risulta

$$\ln p(\mathbf{x}; \sigma^2) = -\frac{M}{2} \ln 2\pi - \frac{M}{2} \ln \sigma^2 - \frac{1}{2\sigma^2} \sum_{n=0}^{M-1} x_n^2$$

il massimo della log-verosimiglianza si ottiene differenziando la stessa e ponendo il risultato uguale a zero:

$$\begin{aligned} \frac{\partial \ln p(\mathbf{x}; \sigma^2)}{\partial \sigma^2} &= -\frac{M}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{M-1} x_n^2 \\ -\frac{M}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{n=0}^{M-1} x_n^2 &= 0 \rightarrow \hat{\sigma}^2 = \frac{1}{M} \sum_{n=0}^{M-1} x_n^2 \end{aligned}$$

Per evitare successivi confusioni metodologiche indichiamo  $\hat{\sigma}^2$  con  $\hat{S}$  e quindi

$$\hat{S} = \frac{1}{M} \sum_{n=0}^{M-1} x_n^2$$

La stima della varianza della sequenza  $\mathbf{x}$  è una variabile casuale, la cui distribuzione è nota come  $\chi^2$  con  $M$  gradi di libertà; infatti  $\hat{S}$  è solo approssimativamente gaussiana, ha valor medio pari a  $\sigma^2$ , e la sua 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 (21),  $\sigma_{\hat{S}}$  decresce con  $\sqrt{M}$ , e quindi lentamente. Calcoleremo ora  $\sigma_{\hat{S}}$  nell'ipotesi che i campioni  $x_m$  siano incorrelati;

$$E[\hat{S}] = \frac{1}{M} E\left[\sum_{m=0}^{M-1} x_m^2\right] = \frac{1}{M} \sum_{m=0}^{M-1} E[x_m^2] = \sigma^2$$

$$\sigma_{\hat{S}}^2 = E[\hat{S}^2] - E[\hat{S}]^2 = \frac{1}{M^2} \sum_{n,m=0}^{M-1} E[x_n^2 x_m^2] - \sigma^4 \quad (18)$$

$$\begin{aligned} &= \frac{ME[x_m^4] + (M^2 - M)E_{n \neq m}[x_n^2 x_m^2]}{M^2} - \sigma^4 = \\ &= \frac{3M\sigma^4 + (M^2 - M)\sigma^4}{M^2} - \sigma^4 = \frac{2\sigma^4}{M} \end{aligned} \quad (19)$$

dove abbiamo utilizzato le relazioni:

$$E[x_m^4] = 3\sigma^4; E[x_m^2 x_n^2] = \sigma^4 \text{ se } (m \neq n); \quad (20)$$

valide per variabili gaussiane reali a v.m. nullo, come sono le  $\{x_m\}$ . Conseguentemente possiamo scrivere

$$\sigma_{\hat{S}} = \sqrt{\frac{2}{M}} \sigma^2 = \sqrt{\frac{2}{M}} E[\hat{S}] \quad (21)$$

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  $\pm 0.05\sigma^2$ . Partiamo dall'ipotesi (ben verificata) che nell'intorno del suo valor medio  $\widehat{S}$  sia ben approssimabile con una gaussiana. 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  $k = 1.016 \times \sqrt{2} = 1.43$ . Di conseguenza

$$1.43\sigma_{\widehat{S}} = 0.05\sigma^2 \rightarrow \sigma_{\widehat{S}} = \frac{0.05}{1.43}\sigma^2$$

e quindi:

$$\sigma_{\widehat{S}} = \sqrt{\frac{2}{M}}\sigma^2 = \frac{0.05}{1.43}\sigma^2 \rightarrow \sqrt{\frac{2}{M}} = \frac{0.05}{1.43} \rightarrow M \simeq 1635$$

### 5.2.2 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:

$$y = \mu x + w; \quad E[x^2] = E[y^2] = \sigma^2; \quad (22)$$

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

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

$$E[x^4] = 3\sigma^4$$

$$E[xy] = \mu\sigma^2; \quad E[x^2y^2] = (1 + 2\mu^2)\sigma^4 \quad (23)$$

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:

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

congruentemente con l'equazione (18) essa sarà caratterizzato dalla varianza

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

e la varianza **NON** decresce a zero anche se  $M$  tende ad  $\infty$ .

Supponiamo ora che i campioni corrispondano ad una sequenza bianca filtrata con un filtro passabasso con un polo in  $z = \rho = 1 - \varepsilon$ . L'autocorrelazione dell'uscita (indicata con  $y$ ) è:

$$E[y_m y_{m+k}] = \rho^{|k|} \sigma^2; \quad \rho \leq 1$$

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

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

Dopo qualche calcolo si trova che:

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

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

$$M_{eq} \simeq M \frac{1-\rho^2}{1+\rho^2}; \text{ se } \rho = 1 - \varepsilon \rightarrow M_{eq} \sim M \cdot \varepsilon$$

## 6 Stima ai minimi quadrati (LS)

Quando si vuole stimare un insieme di parametri *senza* disporre della conoscenza delle pdf interessate ma conoscendo il meccanismo di generazione allora può risultare conveniente minimizzare la somma dei quadrati degli errori di stima. Tale tecnica, indicata come *stima ai minimi quadrati* o *least squares* (LS), è molto diffusa per la semplicità in quanto richiede solo di sapere come è generata la componente deterministica dell'osservazione e per questo rappresenta più una tecnica di ottimizzazione numerica che non una stima nel senso proprio.

Per la stima LS, si assuma noto il modello di generazione

$$x[i] = s[i; \theta] + w[i] \quad \text{con } i = 1, 2, \dots, N$$

in cui non è nota la pdf di  $w[i]$  (contrariamente alla stima ML), la stima LS è basata sulla minimizzazione della somma degli errori tra misura e generazione da modello

$$\theta_{LS} = \arg \min \left\{ \sum_{i=1}^N (x[i] - s[i; \theta])^2 \right\}$$

ovvero sulla minimizzazione del funzionale

$$J(\theta) = \sum_{i=1}^N (x[i] - s[i; \theta])^2 = (\mathbf{x} - \mathbf{s}(\theta))^T (\mathbf{x} - \mathbf{s}(\theta)).$$

Si noti che il valore che ottimizza il funzionale  $J(\theta)$  può non essere noto in forma chiusa o può essere derivabile solo in forma numerica ricercando il minimo con tecniche di ottimizzazione numerica.

Una soluzione più generale è rappresentata dalla stima LS pesata (*weighted LS*) in cui si introduce una matrice definita positiva  $\mathbf{W} \in R^{N \times N}$  con cui vengono pesati in misura maggiore o minore il contributo degli errori sul funzionale complessivo:

$$J(\theta) = (\mathbf{x} - \mathbf{s}(\theta))^T \mathbf{W} (\mathbf{x} - \mathbf{s}(\theta)).$$

Si noti che il funzionale è lo stesso che si avrebbe nella stima ML per rumore gaussiano con covarianza nota  $\mathbf{C}_w$ , per l'uguaglianza basta porre  $\mathbf{W} = \mathbf{C}_w^{-1}$  e quindi per questa scelta la stima LS e la stima ML coincidono. Conseguentemente se si ipotizza un modello lineare la stima LS coincide con quella ricavata nella sottosezione precedente e con lo stimatore BLUE.

## 7 Stima Bayesiana

Talvolta prima di procedere alla stima sono disponibili alcune informazioni sui valori più o meno verosimili per i parametri  $\theta$ . Diventa quindi ragionevole nella stima di  $\theta$  limitare l'intervallo di ricerca ai soli valori ammissibili e/o essere guidati nella stima dalle conoscenze a-priori sui valori maggiormente probabili.

In particolare, dato il modello con rumore additivo

$$\mathbf{x} = \mathbf{s}(\theta) + \mathbf{w},$$

è necessario pensare che ad ogni realizzazione (k-esima) l'insieme di osservazioni  $\mathbf{x}_k \in R^N$  sia il risultato di uno specifico valore dei parametri  $\theta = \theta_k$  e che i parametri possano variare da un insieme di osservazioni all'altro. La variabilità delle osservazioni è quindi duplice: da un lato variano per effetto delle realizzazioni del disturbo  $\mathbf{w}_k$  e dall'altro per le realizzazioni dei parametri  $\theta_k$ . L'insieme dei parametri  $\theta$  è quindi una vc di cui non è noto il valore ma la pdf  $p[\theta]$ . La conoscenza della pdf rappresenta l'informazione a priori sulla probabilità associata ad ogni set di parametri ed è naturale che venga utilizzata nella definizione dello stimatore.

Nello specifico, il MSE può risultare condizionato ad una particolare realizzazione di  $\theta$  e quindi  $\theta$  rappresenta una variabile deterministica non nota ( *$\theta$  ha sempre quel valore e questo non cambia nelle realizzazioni ma, visto che non è scelto a-caso la conoscenza della pdf  $p[\theta]$  non serve a nulla*), in questo caso si ricade negli stimatori ML. Se però si vuole ottimizzare il MSE al variare sia delle realizzazioni di rumore che dei parametri  $\theta$  allora si vuole ricavare uno stimatore che sia ottimale sempre.

Usando il MSE complessivo come criterio di ottimizzazione questo essere calcolato sia rispetto alle osservazioni  $\mathbf{x}$  che rispetto alla vc  $\theta$ , in sintesi il MSE totale (somma del MSE sulle singole variabili, purchè omogenee) risulta

$$MSE = E_{x,\theta}[(\theta - \hat{\theta}(\mathbf{x}))^H (\theta - \hat{\theta}(\mathbf{x}))] = \iint (\theta - \hat{\theta}(\mathbf{x}))^H (\theta - \hat{\theta}(\mathbf{x})) p[\mathbf{x}, \theta] d\mathbf{x} d\theta$$

ed è indipendente dal valore del parametro  $\theta$  (talvolta questo indicato come MSE Bayesiano). Si noti che confrontato con il MSE per la stima di parametri questo è:  $MSE = E_x[(\theta - \hat{\theta}(\mathbf{x}))^H (\theta - \hat{\theta}(\mathbf{x}))] = \int (\theta - \hat{\theta})^H (\theta - \hat{\theta}) p[\mathbf{x}|\theta] d\mathbf{x}$ , ovvero la pdf  $p[\mathbf{x}|\theta]$  è solo parametrizzata rispetto ai parametri  $\theta$  e quindi il MSE può dipendere dal valore  $\theta$ . Dalla regola di Bayes la

pdf congiunta si può calcolare combinando la pdf condizionata al parametro con la pdf del parametro stesso (o pdf a-priori), in sintesi

$$p[\mathbf{x}, \theta] = p[\mathbf{x}|\theta]p[\theta]$$

(non si dimentichi che  $p[\mathbf{x}|\theta]$  è la sezione della pdf congiunta  $p[\mathbf{x}, \theta]$  moltiplicata per un fattore di scala per garantirne la normalizzazione (area unitaria).

Sempre dalla regola di Bayes è conveniente calcolare la pdf del parametro condizionata all'osservazione

$$p[\theta|\mathbf{x}] = \frac{p[\mathbf{x}|\theta]p[\theta]}{p[\mathbf{x}]},$$

questa quantità è spesso indicata come pdf a-posteriori (ovvero la pdf del parametro a valle di una specifica osservazione). La stima del parametro è possibile seguendo due strade:

- **Stima massima probabilità a-posteriori (MAP):** ovvero si stima come parametro il valore che massimizza la probabilità data l'osservazione ovvero

$$\theta_{MAP} = \arg \max_{\theta} p[\theta|\mathbf{x}].$$

- **Stima a minimo errore quadratico medio (MMSE):** dalla definizione e limitandoci al caso di dati e parametri reali

$$MSE = \int \left[ \int (\theta - \hat{\theta}(\mathbf{x}))^T (\theta - \hat{\theta}(\mathbf{x})) p[\theta|\mathbf{x}] d\theta \right] p[\mathbf{x}] d\mathbf{x}$$

per la minimizzazione si deve annullare il gradiente rispetto a  $\hat{\theta}$ , è sufficiente calcolare il gradiente del solo termine in parentesi (che poi rappresenta il MSE condizionato ad una specifica osservazione  $\mathbf{x}$ )

$$\frac{\partial}{\partial \hat{\theta}} \int (\theta - \hat{\theta})^2 p[\theta|\mathbf{x}] d\theta = -2 \int (\theta - \hat{\theta}) p[\theta|\mathbf{x}] d\theta = \mathbf{0}$$

si ha lo stimatore MMSE

$$\hat{\theta}_{MMSE} = \int \theta p[\theta|\mathbf{x}] d\theta = E[\theta|\mathbf{x}]$$

noto anche come stimatore a media condizionata.

Lo stimatore MAP e MMSE operano sempre sulla pdf a-posteriori calcolando rispettivamente il massimo e la media. Talvolta la stima MMSE, pur baricentrale come valore rispetto alla pdf a-posteriori (vedi figura), può corrispondere ad un valore poco probabile (se non impossibile) che comunque ne minimizza l'errore rispetto ad una stima MAP che predilige il valore più probabile e comunque un valore possibile. E' banale verificare (anche graficamente dalla stessa figura) che la stima MAP e MMSE coincidono per pdf a-posteriori  $p[\theta|\mathbf{x}]$  con simmetria pari rispetto ad un valore massimo. E' di cruciale importanza la scelta della pdf a-priori per le prestazioni dello stimatore, questo si ha da una buona conoscenza dei meccanismi di generazione della variabilità dei parametri. Un caso relativamente semplice è quello che vede il problema della stima Bayesiana per una variabile in un rumore additivo, non necessariamente gaussiano.

## 7.1 Stima Bayesiana per $x = \theta + w$ con $w \sim \mathcal{N}(0, \sigma_w^2)$

Sia dato il modello  $x = \theta + w$  in cui si vuole stimare il parametro  $\theta$  da una sola osservazione nell'assunzione che il rumore sia gaussiano:  $w \sim \mathcal{N}(0, \sigma_w^2)$ . Qui verrà discusso solo il caso in cui il parametro ha una pdf gaussiana (il più semplice). Il caso generale (con pdf qualunque) è lasciato ai corsi di Laurea Magistrale.

### 7.1.1 Modello gaussiano: $\theta \sim \mathcal{N}(\bar{\theta}, \sigma_\theta^2)$

La stima ML è rappresentata dal valore della stessa osservazione per ogni realizzazione

$$\begin{aligned}\theta_{ML} &= x \\ \text{var}[\theta_{ML}] &= \text{MSE}_{ML} = \sigma_w^2.\end{aligned}$$

Se si assume che da una realizzazione all'altra (ciascuna da un solo campione) cambia il parametro  $\theta$ , se questo ha una pdf gaussiana  $\theta \sim \mathcal{N}(\bar{\theta}, \sigma_\theta^2)$  in cui  $\bar{\theta}$  e  $\sigma_\theta^2$  sono parametri noti della pdf, la pdf congiunta è  $p[x, \theta] = p[x|\theta]p[\theta]$  mentre la pdf a-posteriori data l'osservazione è

$$p[\theta|x] = \frac{p[x|\theta]p[\theta]}{p[x]} = \underbrace{\Gamma \exp\left(-\frac{(x-\theta)^2}{2\sigma_w^2}\right)}_{p[x|\theta]} \cdot \underbrace{\exp\left(-\frac{(\theta-\bar{\theta})^2}{2\sigma_\theta^2}\right)}_{p[\theta]}$$

dove  $\Gamma$  rappresenta un termine di normalizzazione (irrelevante per la stima in quanto indipendente da  $\theta$ ). Riorganizzando l'argomento del termine esponenziale si ha che la pdf

$$p[\theta|x] = \frac{1}{\sqrt{2\pi}\sigma_o} \exp\left(-\frac{(\theta - g(x))^2}{2\sigma_o^2}\right)$$

è gaussiana (proprietà questa di validità generale quando la pdf dei parametri è gaussiana e il modello è lineare, vedi seguito), i termini sono

$$\begin{aligned}g(x) &= \frac{\sigma_w^2}{\sigma_w^2 + \sigma_\theta^2} \bar{\theta} + \frac{\sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2} x = \bar{\theta} + \frac{\sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2} (x - \bar{\theta}) \\ \sigma_o^2 &= \frac{\sigma_w^2 \sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2}\end{aligned}$$

La massimizzazione della pdf  $p[\theta|x]$  gaussiana si ha per

$$\theta_{MAP} = g(x) = \bar{\theta} + \frac{\sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2} (x - \bar{\theta}) = \theta_{MMSE},$$

dove l'uguaglianza con la stima MMSE segue dalla simmetria della pdf  $p[\theta|x]$ . Si osservi che per  $\sigma_\theta^2 \rightarrow 0$  l'informazione a-priori domina sulla stima e  $\theta_{MAP} \rightarrow \bar{\theta}$  indipendentemente dalla misura mentre per  $\sigma_\theta^2 \rightarrow \infty$  la pdf a-priori diventa irrilevante (in questi casi si parla di *informazione a-priori diffusa*) e  $\theta_{MAP} = x$  come per la stima ML. Infatti in assenza di informazione a-priori (disporre di una pdf a-priori in cui tutti i valori sono equiprobabili equivale ad avere una pdf a-priori senza rilevante contenuto informativo) la stima Bayesiana

converge alla stima non-Bayesiana in quanto non si ha alcun apporto dell'informazione a-priori alla stima del parametro.

Volendo calcolare il MSE in questo secondo caso si deve calcolare l'errore sia rispetto alla variabilità delle osservazioni che del parametro e quindi (mettendo in pedice, per chiarezza, le variabili rispetto a cui è calcolato il valore atteso)

$$\begin{aligned}
MSE_{MAP/MMSE} &= E_{\theta,w}[(\theta_{MAP} - \theta)^2] = E \left[ \left( \frac{\sigma_w^2}{\sigma_w^2 + \sigma_\theta^2} \bar{\theta} + \frac{\sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2} (\theta + w) - \theta \right)^2 \right] \\
&= E \left[ \left( \frac{\sigma_w^2}{\sigma_w^2 + \sigma_\theta^2} (\bar{\theta} - \theta) + \frac{\sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2} w \right)^2 \right] = \left( \frac{\sigma_w^2}{\sigma_w^2 + \sigma_\theta^2} \right)^2 \sigma_\theta^2 + \left( \frac{\sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2} \right)^2 \sigma_w^2 \\
&= \frac{\sigma_w^2 \sigma_\theta^2}{\sigma_w^2 + \sigma_\theta^2} \leq \sigma_w^2 = MSE_{ML}
\end{aligned}$$

In sintesi, il MSE per la stima MAP/MMSE è sempre inferiore alla stima ML in quanto nella stima MMSE si sfrutta in modo costruttivo l'informazione a-priori per la riduzione dell'incertezza media.

Ovviamente vale la relazione  $MSE_{MAP/MMSE} \rightarrow MSE_{ML}$  per  $\sigma_\theta^2 \rightarrow \infty$ .

## 7.2 Stima MMSE per processi gaussiani

Nella pratica è diffusa l'assunzione gaussiana nella modellizzazione dei parametri e delle osservazioni. Il principale vantaggio è legato alla possibilità di calcolare lo stimatore MMSE in forma esplicita. Poichè è possibile mostrare che se la pdf congiunta  $p[\mathbf{x}, \theta]$  è gaussiana

$$\begin{bmatrix} \mathbf{x} \\ \theta \end{bmatrix} \sim N \left( \begin{bmatrix} \mu_x \\ \mu_\theta \end{bmatrix}, \begin{bmatrix} \mathbf{C}_{xx} & \mathbf{C}_{x\theta} \\ \mathbf{C}_{\theta x} & \mathbf{C}_{\theta\theta} \end{bmatrix} \right)$$

allora anche la pdf condizionata  $p[\theta|\mathbf{x}]$  è gaussiana. In particolare la pdf condizionata ha valore medio e covarianza:

$$\theta|\mathbf{x} \sim N \left( \mu_\theta + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - \mu_x), \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \right).$$

Da questa segue immediatamente l'espressione dello stimatore MMSE

$$\theta_{MMSE} = E[\theta|\mathbf{x}] = \mu_\theta + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - \mu_x)$$

che si calcola da caso a caso a seconda del modello. Si noti che lo stimatore è lineare nelle osservazioni  $\mathbf{x}$  a cui si deve sottrarre preventivamente l'informazione a priori sul valore medio  $\mu_x$ .

In particolare, per un modello con rumore additivo (certamente il più diffuso e il più ricorrente)

$$\mathbf{x} = \mathbf{s}(\theta) + \mathbf{w} \quad \text{con } \mathbf{w} \sim N(0, \mathbf{C}_{ww})$$

il calcolo delle covarianze e dei valori medi risulta (assumendo l'indipendenza tra rumore e parametri):

$$\begin{aligned}
\mu_x &= E_{\theta,w}[\mathbf{x}] = E_\theta[\mathbf{s}(\theta)] \\
\mathbf{C}_{xx} &= E_{\theta,w}[(\mathbf{x} - \mu_x)(\mathbf{x} - \mu_x)^T] = E_\theta[\mathbf{s}(\theta)\mathbf{s}(\theta)^T] + \mathbf{C}_{ww} - \mu_x \mu_x^T \\
\mathbf{C}_{\theta x} &= E_{\theta,w}[(\theta - \mu_\theta)(\mathbf{x} - \mu_x)^T] = E_\theta[(\theta - \mu_\theta)(\mathbf{s}(\theta) - \mu_x)^T].
\end{aligned}$$

Si noti che i momenti rispetto ai parametri  $E_{\theta}[\cdot]$  sono le quantità da calcolare per derivare lo stimatore MMSE. Se il modello delle osservazioni è lineare il calcolo dei momenti si semplifica come illustrato di seguito.

### 7.3 Stima MMSE per modelli lineari e processi gaussiani

I modelli lineari rappresentano un caso particolarmente favorevole per la stima MMSE. Si assuma che valga il modello lineare in rumore gaussiano

$$\begin{aligned}\mathbf{x} &= \mathbf{H} \cdot \theta + \mathbf{w} \\ \mathbf{w} &\sim N(0, \mathbf{C}_{ww}),\end{aligned}$$

si calcolano ora i momenti richiesti

$$\begin{aligned}\mu_x &= E_{\theta,w}[\mathbf{H} \cdot \theta + \mathbf{w}] = \mathbf{H}\mu_{\theta} \\ \mathbf{C}_{\theta x} &= E_{\theta,w}[(\theta - \mu_{\theta})(\mathbf{H} \cdot (\theta - \mu_{\theta}) + \mathbf{w})^T] = \mathbf{C}_{\theta\theta}\mathbf{H}^T \\ \mathbf{C}_{xx} &= E_{\theta,w}[(\mathbf{H} \cdot (\theta - \mu_{\theta}) + \mathbf{w})(\mathbf{H} \cdot (\theta - \mu_{\theta}) + \mathbf{w})^T] \\ &= \mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_{ww}\end{aligned}$$

da cui la relazione generale

$$\begin{aligned}\theta_{MMSE} &= \mu_{\theta} + \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_{ww})^{-1}(\mathbf{x} - \mathbf{H}\mu_{\theta}) \\ cov(\theta|\mathbf{x}) &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta\theta}\mathbf{H}^T(\mathbf{H}\mathbf{C}_{\theta\theta}\mathbf{H}^T + \mathbf{C}_{ww})^{-1}\mathbf{H}\mathbf{C}_{\theta\theta}\end{aligned}$$

che evidenzia la struttura lineare (nelle osservazioni) dello stimatore.

**Esempio:** Stima MMSE della componente continua per il modello  $\mathbf{x} = \theta\mathbf{1} + \mathbf{w}$  con  $\mathbf{w} \sim N(0, \sigma_w^2\mathbf{I})$  e  $\theta \sim \mathcal{N}(\bar{\theta}, \sigma_{\theta}^2)$  è

$$\theta_{MMSE} = \bar{\theta} + \sigma_{\theta}^2\mathbf{1}^T(\mathbf{1}\sigma_{\theta}^2\mathbf{1}^T + \sigma_w^2\mathbf{I})^{-1}(\mathbf{x} - \bar{\theta}\mathbf{1})$$

utilizzando un particolare lemma relativo alla inversione di matrici la relazione precedente può essere riscritta come:

$$\theta_{MMSE} = \alpha\bar{\theta} + (1 - \alpha)\underbrace{\mathbf{1}^T\mathbf{x}/N}_{\bar{x}} \quad \text{con } \alpha = \frac{\sigma_w^2/N}{\sigma_w^2/N + \sigma_{\theta}^2}$$

ovvero la combinazione ottimale tra il valore medio del parametro noto a-priori e la media campionaria dalle misure, per  $N \rightarrow \infty$  l'influenza dell'informazione a-priori decresce e si ha  $\theta_{MMSE} \rightarrow \bar{x}$ .

### 7.4 Stima MMSE lineare (LMMSE) e ortogonalità

La stima MMSE  $E[\theta|\mathbf{x}]$  può risultare in alcuni casi troppo complicata in quanto i) per derivare  $E[\theta|\mathbf{x}]$  si richiede la conoscenza delle statistiche dei processi (questo non è sempre possibile), ii) la valutazione della media condizionata  $E[\theta|\mathbf{x}]$  può essere molto complicata.



A questo proposito si cerca uno stimatore da porre in alternativa allo stimatore MMSE in cui la stima i) sia ottenuta come combinazione lineare delle osservazioni e ii) necessiti della conoscenza di pochi momenti (ad esempio solo i momenti di ordine 1 e 2) delle pdf associate alle grandezze in gioco. Tale stimatore, pur subottimo, è indicato come stimatore MMSE lineare (LMMSE).

Lo stimatore LMMSE è quindi dato dalla combinazione lineare

$$\hat{\theta}_k(\mathbf{x}) = a_{k,0} + \sum_{i=1}^N a_{k,i}x[i] = a_{k,0} + \mathbf{a}_k^T \mathbf{x}$$

mentre per l'insieme di parametri diventa

$$\begin{bmatrix} \hat{\theta}_1(\mathbf{x}) \\ \hat{\theta}_2(\mathbf{x}) \\ \vdots \\ \hat{\theta}_p(\mathbf{x}) \end{bmatrix} = \begin{bmatrix} a_{1,0} \\ a_{2,0} \\ \vdots \\ a_{p,0} \end{bmatrix} + \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \vdots \\ \mathbf{a}_p^T \end{bmatrix} \cdot \mathbf{x} \rightarrow \hat{\boldsymbol{\theta}}(\mathbf{x}) = \mathbf{a}_0 + \mathbf{A}\mathbf{x} \quad (24)$$

Per ottenere la matrice dei coefficienti  $\mathbf{A}$  si impongono essenzialmente due vincoli:

- Non-polarizzazione:

$$E[\theta_k - \hat{\theta}_k(\mathbf{x})] = 0 \rightarrow E[\theta] - \mathbf{A}E[\mathbf{x}] - \mathbf{a}_0 = \mathbf{0} \rightarrow \mathbf{a}_0 = E[\theta] - \mathbf{A}E[\mathbf{x}]$$

$$\mathbf{a}_0 = \mu_\theta - \mathbf{A}\mu_x$$

oppure viene riscritto lo stimatore accorpando il valore medio sui parametri e sulle osservazioni, entrambi noti:

$$\underbrace{\hat{\boldsymbol{\theta}}(\mathbf{x}) - \mu_\theta}_{\delta\hat{\boldsymbol{\theta}}(\mathbf{x})} = \mathbf{A} \underbrace{(\mathbf{x} - \mu_x)}_{\delta\mathbf{x}}$$

- Minimizzazione del MSE per ogni parametro:

$$\arg \min_{\mathbf{A}} E[(\theta_k - \hat{\theta}_k(\mathbf{x}))^2], \text{ per } \forall k = 1, \dots, p$$

Per non introdurre ulteriori complicazioni sulla notazione si assume che alle grandezze sia stato preventivamente sottratto il valore medio, ovvero si opera su  $\delta\hat{\boldsymbol{\theta}}(\mathbf{x})$  e  $\delta\mathbf{x}$  mantenendo nella notazione  $\hat{\boldsymbol{\theta}}(\mathbf{x})$  e  $\mathbf{x}$ , in pratica è come assumere che  $\mu_\theta = \mathbf{0}$  e  $\mu_x = \mathbf{0}$ . Il valore di  $\mathbf{a}_k$  che corrisponde al minimo del MSE si trova annullando la derivata

$$\begin{aligned} \frac{\partial}{\partial a_{k,i}} E[(\theta_k - \hat{\theta}_k(\mathbf{x}))^2] &= 0 \quad \text{per } \forall i \\ E[(\theta_k - \hat{\theta}_k(\mathbf{x}))x[i]] &= 0 \quad \text{per } \forall i \end{aligned}$$

che considerando l'intero vettore dei dati può essere scritta come

$$E[(\theta_k - \hat{\theta}_k(\mathbf{x}))\mathbf{x}^T] = E[(\theta_k - \mathbf{a}_k^T \mathbf{x})\mathbf{x}^T] = \mathbf{0}_{1 \times N} \quad \text{per } \forall k$$

questa relazione evidenzia che il minimo del MSE è trovato cercando la **condizione di ortogonalità** tra errore di stima  $(\theta_k - \hat{\theta}_k(\mathbf{x}))$  e osservazioni. Dall'espressione precedente è evidente come la soluzione sia espressa come un sistema lineare di  $N$  equazioni nelle  $N$  incognite (vettore  $\mathbf{a}_k$ ) la cui soluzione è (si ricordi la condizione  $\mu_\theta = \mathbf{0}$  e  $\mu_x = \mathbf{0}$ )

$$\mathbf{a}_k = (E[\mathbf{x}\mathbf{x}^T])^{-1} \cdot E[\theta_k \mathbf{x}]$$

Per l'insieme di parametri la condizione di ortogonalità tra errore di stima e osservazioni diventa

$$E[(\theta - \hat{\theta}(\mathbf{x}))\mathbf{x}^T] = \underset{p \times N}{\mathbf{0}}.$$

Sviluppando la condizione di ortogonalità con  $\hat{\theta}(\mathbf{x}) = \mathbf{A}\mathbf{x}$  (si ricordi che  $\mathbf{a}_0 = 0$  per le assunzioni viste sopra) si ha

$$E[(\theta - \mathbf{A}\mathbf{x})\mathbf{x}^T] = \mathbf{0} \rightarrow \mathbf{A} = \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1}.$$

Aggiungendo ora il termine per la rimozione della polarizzazione si ha la soluzione generale dello stimatore LMMSE

$$\begin{aligned} \theta_{LMMSE} &= \mu_\theta + \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} (\mathbf{x} - \mu_x) \\ cov[\theta_{LMMSE}] &= \mathbf{C}_{\theta\theta} - \mathbf{C}_{\theta x} \mathbf{C}_{xx}^{-1} \mathbf{C}_{x\theta} \end{aligned}$$

Si noti come per lo stimatore LMMSE sia necessario disporre dei valori medi ( $\mu_\theta$  e  $\mu_x$ ), della matrice di covarianza dell'osservazione  $\mathbf{C}_{xx}$  e della cross-covarianza  $\mathbf{C}_{\theta x}$  tra osservazione e parametri. Se il numero di parametri si riducesse a  $p = 1$  allora  $\mathbf{C}_{\theta x}$  rappresenterebbe il vettore di cross-covarianza tra osservazioni e parametro.

Si noti che lo stimatore LMMSE è identico allo stimatore MMSE per processi gaussiani, da cui l'ottimalità solo in questo caso. Lo stimatore lineare LMMSE fornisce una stima sub-ottima nel caso di pdf qualunque pur rappresentando, tra le stime lineari, quella a minimo MSE.

## 8 Applicazioni di Teoria della Stima ai problemi di Elaborazione Numerica dei Segnali

Le metodologie di stima viste nella sezioni precedente hanno notevoli applicazioni nell'ambito dell'elaborazione numerica dei segnali.

### 8.1 Predizione Lineare

In moltissime situazioni vi è la necessità di stimare il valore di un segnale o di una serie temporale ad un certo istante a partire dalla conoscenza dei campioni (valori) precedenti. Si pensi ad esempio al caso in cui si voglia predire il valore che una azione di borsa avrà domani sulla base del suo valore di oggi e dei giorni precedenti.

In particolare si consideri il caso di segnali reali con media nulla ed in cui si voglia stimare  $x(n)$  attraverso una combinazione lineare dei campioni precedenti  $\hat{x}(n) = -\sum_{k=1}^N b_k x(n-k)$  (la ragione del segno - sarà evidente nel seguito). Il problema può essere impostato come un caso di stima LMMSE con un solo parametro da stimare. In questo caso posto  $\theta = x(n)$ ,  $\mathbf{x} = [x(n-N), x(n-N+1), \dots, x(n-1)]^T$  (il vettore include i campioni del segnale per valori crescenti della variabile indipendente), se definiamo  $\mathbf{b}^T = [b_N, b_{N-1}, \dots, b_1]$  si ha  $\hat{x}(n) = -\mathbf{b}^T \mathbf{x}$ . La soluzione al problema è

$$\mathbf{b}^T = -\mathbf{C}_{\theta\mathbf{x}} \mathbf{C}_{\mathbf{xx}}^{-1} = -\mathbf{C}_{x(n)\mathbf{x}} \mathbf{C}_{\mathbf{xx}}^{-1}$$

dove

$$\begin{aligned} \mathbf{C}_{x(n)\mathbf{x}} &= [E[x(n)x(n-N)], E[x(n)x(n-N+1)], \dots, E[x(n)x(n-1)]] \\ &= [r_x(N), r_x(N-1), \dots, r_x(1)] \\ \mathbf{C}_{\mathbf{xx}} &= \begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) & \dots & r_x(1-N) \\ r_x(1) & r_x(0) & r_x(-1) & \dots & \dots \\ r_x(2) & r_x(1) & r_x(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_x(N-1) & \dots & \dots & \dots & r_x(0) \end{bmatrix} \end{aligned}$$

$\mathbf{C}_{x(n)\mathbf{x}}$  e  $\mathbf{C}_{\mathbf{xx}}$  sono normalmente ottenute attraverso stime campionarie. Qualora si sia in presenza di segnali le cui caratteristiche statistiche siano variabili nel tempo, queste stime andranno di pian piano aggiornate per inseguire le variazioni statistiche.

L'errore di predizione è definito come  $\epsilon(n) = x(n) - \hat{x}(n) = x(n) + \mathbf{b}^T \mathbf{x}$ . Si noti come il segno - nella definizione dello stimatore porti ad un segno + per il calcolo dell'errore di predizione. Questo errore ha un significato molto importante. Supponiamo infatti di essere nella condizione in cui  $x(n)$  sia modellizzabile come un rumore bianco che passi attraverso un filtro tutti poli:  $1/A(z)$ .

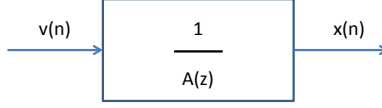


Figure 2:  $x(n)$  generato da una struttura tutti poli.

Si abbia  $A(z) = \sum_{k=0}^N a_k z^{-k}$  con  $a_0 = 1$ . Conseguentemente risulta

$$x(n) = -\sum_{k=1}^N a_k x(n-k) + v(n)$$

e

$$\begin{aligned} r_x(m) &= E[x(n)x(n-m)] \\ &= -E\left[\left(\sum_{k=1}^N a_k x(n-k)\right)x(n-m)\right] + E[v(n)x(n-m)] \\ &= -\sum_{k=1}^N a_k r_x(m-k) + \sigma_v^2 \delta(m) \end{aligned}$$

Se l'ultima relazione è scritta in forma matriciale per  $r_x(m)$  con  $m = N, N-1, \dots, 1$  risulta

$$\begin{aligned} \mathbf{C}_{x(n)\mathbf{x}} &= \begin{bmatrix} E[x(n)x(n-N)] \\ E[x(n)x(n-(N-1))] \\ \vdots \\ E[x(n)x(n-1)] \end{bmatrix} = \begin{bmatrix} r_x(N) \\ r_x(N-1) \\ \vdots \\ r_x(1) \end{bmatrix} \\ &= -[a_N, a_{N-1}, \dots, a_1] \begin{bmatrix} r_x(0) & r_x(-1) & r_x(-2) & \dots & r_x(1-N) \\ r_x(1) & r_x(0) & r_x(-1) & \dots & \dots \\ r_x(2) & r_x(1) & r_x(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_x(N-1) & \dots & \dots & \dots & r_x(0) \end{bmatrix} \\ &= -\mathbf{a}^T \mathbf{C}_{\mathbf{x}\mathbf{x}} \end{aligned}$$

Queste equazioni prendono il nome, normalmente, di relazioni di "Yule-Walker".

Quanto scritto sopra porta a dire che il filtro usato per generare l'errore di predizione corrisponde esattamente all'inverso del filtro di "coloritura" che genera  $x(n)$ . Inoltre l'errore di predizione è bianco come lo era il segnale  $v(n)$  (ed avrà la sua stessa varianza).

Questo tipo di approccio è molto utilizzato per stimare le caratteristiche di un sistema che abbia generato un particolare segnale quando si possa assumere che tale filtro (sistema) sia tutti poli e la sua eccitazione sia bianca. Tipico caso è il segnale vocale in cui il segnale di pressione generato alla glottide può essere assunto, con buona approssimazione, come

bianco ed il passaggio di tale segnale attraverso il tratto vocale è ben rappresentato da una funzione di trasferimento tutti poli.

Se l'ordine del predittore è inferiore a quello del filtro di "coloritura" lo sbiancamento del segnale sarà solo parziale, mentre nel caso di ordine del predittore superiore a quello del filtro tutti i poli lo sbiancamento risulterà sempre completo con gli ultimi coefficienti di predizione a zero. Operando in casi concreti, ove le funzioni di correlazione sono calcolate in modo campionario, eccedere con l'ordine del predittore riduce la qualità della stima.

Talvolta la situazione da affrontare è quella illustrata nella figura seguente:

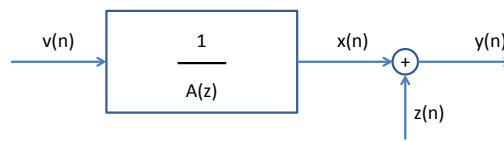


Figure 3:  $y(n)$  segnale osservato è l'uscita di una struttura tutti poli a cui si aggiunge del rumore.

Assumiamo che  $z(n)$  sia un rumore bianco con varianza  $\sigma_z^2$ , inoltre consideriamo tutti i campioni  $v(n)$  incorrelati da quelli di  $z(n)$ . In questo caso il segnale  $x(n)$  non è osservabile, mentre le osservazioni ( $\mathbf{y}$ ) sono costituite da  $N$  campioni di  $y(n)$  ad esempio  $\{y(n-N), \dots, y(n-1)\}$ , ma potremmo considerare anche il campione corrente di  $y(n)$ , cioè l'insieme di  $N$  campioni dato da  $\{y(n-N+1), \dots, y(n)\}$ . Volendo considerare una situazione pratica, possiamo pensare ad un coro che canti in una grande chiesa, il segnale raccolto dal microfono è quello del coro (riverberato dall'ambiente) a cui si somma il brusio dell'uditorio. La riverberazione dell'ambiente può essere vista, in prima approssimazione, come il filtraggio con sistema tutti poli. Si possono cercare di risolvere due problemi:

1. Eliminare il brusio (quindi stimare  $x(n)$ );
2. Cercare di ottenere il segnale del coro depurato dal brusio e dalla riverberazione ( $v(n)$ ).

La stima di  $x(n)$  sarà sempre calcolata come  $\hat{x}(n) = \mathbf{b}^T \mathbf{y}$ , in cui

$$\mathbf{b}^T = -\mathbf{C}_{x(n)\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1}$$

dove, tenuto conto che  $y(n) = x(n) + z(n)$ , si ha:

$$\begin{aligned}
\mathbf{C}_{x(n)\mathbf{y}} &= [E[x(n)y(n-N)], E[x(n)y(n-N+1)], \dots, E[x(n)y(n-1)]] \\
&= [r_x(N), r_x(N-1), \dots, r_x(1)] \\
\mathbf{C}_{\mathbf{y}\mathbf{y}} &= \begin{bmatrix} r_y(0) & r_y(-1) & r_y(-2) & \dots & r_y(1-N) \\ r_y(1) & r_y(0) & r_y(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_y(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_y(N-1) & \dots & \dots & \dots & r_y(0) \end{bmatrix} \\
&= \begin{bmatrix} r_x(0) + \sigma_z^2 & r_x(-1) & r_x(-2) & \dots & r_x(1-N) \\ r_x(1) & r_x(0) + \sigma_z^2 & r_x(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_x(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_x(N-1) & \dots & \dots & \dots & r_x(0) + \sigma_z^2 \end{bmatrix} \\
&= \mathbf{C}_{\mathbf{x}\mathbf{x}} + \sigma_z^2 \mathbf{I}
\end{aligned}$$

$\mathbf{C}_{\mathbf{y}\mathbf{y}}$  può essere stimata in modo campionario dai segnale osservato ( $y(n)$ ). La stima di  $\mathbf{C}_{x(n)\mathbf{y}}$  coinvolge  $x(n)$  non noto, ma osservando le relazioni precedenti (in particolare la prima colonna di  $\mathbf{C}_{\mathbf{y}\mathbf{y}}$ ) si constata che le componenti di  $\mathbf{C}_{x(n)\mathbf{y}}$  coincidono con  $r_y(m)$  per  $m = N, \dots, 1$  nel caso il vettore delle osservazioni sia dato da  $\{y(n-N), \dots, y(n-1)\}$ .

Nel caso il vettore delle osservazioni sia  $\{y(n-N+1), \dots, y(n)\}$  si ha:

$$\begin{aligned}
\hat{x}(n) &= -\mathbf{b}^T \mathbf{y} \\
\mathbf{b}^T &= -\mathbf{C}_{x(n)\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} \\
\mathbf{C}_{x(n)\mathbf{y}} &= [E[x(n)y(n-N+1)], E[x(n)y(n-N+2)], \dots, E[x(n)y(n)]] \\
&= [r_x(N-1), r_x(N-2), \dots, r_x(0)] \\
\mathbf{C}_{\mathbf{y}\mathbf{y}} &= \begin{bmatrix} r_y(0) & r_y(-1) & r_y(-2) & \dots & r_y(1-N) \\ r_y(1) & r_y(0) & r_y(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_y(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_y(N-1) & \dots & \dots & \dots & r_y(0) \end{bmatrix} \\
&= \begin{bmatrix} r_x(0) + \sigma_z^2 & r_x(-1) & r_x(-2) & \dots & r_x(1-N) \\ r_x(1) & r_x(0) + \sigma_z^2 & r_x(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_x(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_x(N-1) & \dots & \dots & \dots & r_x(0) + \sigma_z^2 \end{bmatrix} \\
&= \mathbf{C}_{\mathbf{x}\mathbf{x}} + \sigma_z^2 \mathbf{I}
\end{aligned}$$

In questo caso  $r_x(0)$  non è direttamente disponibile a partire da  $\mathbf{C}_{\mathbf{y}\mathbf{y}}$ , ma si può ottenere da un elemento qualsiasi della diagonale principale di questa matrice se è noto  $\sigma_z^2$ , ad esempio stimando il valore di questa varianza nei momenti di silenzio del coro.

Quando si vuole stimare  $v(n)$  l'approccio è sempre lo stesso, in questo caso si può pensare di usare un vettore delle osservazioni che include il campione corrente, si ha:

$$\begin{aligned}
\hat{v}(n) &= -\mathbf{b}^T \mathbf{y} \\
\mathbf{b}^T &= -\mathbf{C}_{v(n)\mathbf{y}} \mathbf{C}_{\mathbf{y}\mathbf{y}}^{-1} \\
\mathbf{C}_{v(n)\mathbf{y}} &= [E[v(n)y(n-N+1)], E[v(n)y(n-N+2)], \dots, E[v(n)y(n)]] \\
&= [0, 0, \dots, \sigma_v^2] \\
\mathbf{C}_{\mathbf{y}\mathbf{y}} &= \begin{bmatrix} r_y(0) & r_y(-1) & r_y(-2) & \dots & r_y(1-N) \\ r_y(1) & r_y(0) & r_y(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_y(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_y(N-1) & \dots & \dots & \dots & r_y(0) \end{bmatrix} \\
&= \begin{bmatrix} r_x(0) + \sigma_z^2 & r_x(-1) & r_x(-2) & \dots & r_x(1-N) \\ r_x(1) & r_x(0) + \sigma_z^2 & r_x(-1) & \dots & \dots \\ r_y(2) & r_y(1) & r_x(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_x(N-1) & \dots & \dots & \dots & r_x(0) + \sigma_z^2 \end{bmatrix} \\
&= \mathbf{C}_{\mathbf{x}\mathbf{x}} + \sigma_z^2 \mathbf{I}
\end{aligned}$$

Nel caso  $\sigma_z^2 = 0$  siamo esattamente nella prima situazione che abbiamo analizzato (equazioni di Yule-Walker) l'unica differenza è che ora stimiamo direttamente l'errore di predizione. Si noti che in questo caso il valore dato a  $\sigma_v^2$  è irrilevante, semplicemente scala i valori dei coefficienti usati dallo stimatore.

Quando  $\sigma_z^2 \neq 0$  il segnale che si stimerà non potrà essere completamente bianco perchè la matrice di covarianza  $\mathbf{C}_{\mathbf{y}\mathbf{y}}$  è diversa da  $\mathbf{C}_{\mathbf{x}\mathbf{x}}$ .

I problemi che sono stati affrontati finora possono essere risolti anche con approcci basati sull'analisi nel dominio di Fourier (dominio delle frequenze). Il grosso vantaggio è che in molte situazioni possiamo assumere la matrice di covarianza dei dati trasformati come diagonale (vedi quanto detto sui processi casuali gaussiani ed il loro passaggio attraverso sistemi lineari ovvero filtri) e quindi l'espressione  $\mathbf{b} = \mathbf{C}_{\text{oss}}^{-1} \mathbf{C}_{\theta-\text{oss}}$  diventa estremamente semplice da risolvere.

Consideriamo ora lo schema rappresentato in figura 4, in cui il sistema lineare considerato ( $H(z)$ ) può avere sia poli sia zeri.

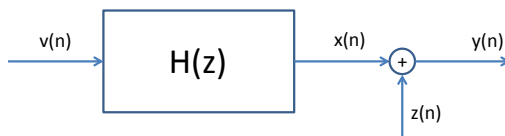


Figure 4:  $y(n)$  segnale osservato è l'uscita di un sistema LTI ( $H(z)$ ) a cui si aggiunge del rumore.

Pensando di lavorare nel dominio delle pulsazioni possiamo scrivere

$$\begin{aligned} Y(\omega) &= V(\omega)H(\omega) + Z(\omega) \\ E[|V(\omega)|^2] &= \sigma_v^2 \\ E[|Z(\omega)|^2] &= \sigma_z^2 \\ \hat{V}(\omega) &= B(\omega)Y(\omega) \end{aligned}$$

in linea di principio la variabile  $\omega$  va considerata come continua nell'intervallo  $-\pi/2, \pi/2$  (intervallo fra campioni consecutivi unitario). Se il numero dei campioni di segnale considerato è molto superiore alla lunghezza della risposta all'impulso di  $H(\omega)$  possiamo anche considerare la variabile  $\omega$  come discreta (usiamo la DFT al posto della DTFT),  $E[|V(\omega)|^2]$  e  $E[|Z(\omega)|^2]$  rappresentano le densità spettrali di potenza di  $v(n)$  e  $z(n)$  (si assume che  $v(n)$  e  $z(n)$  siano stazionari almeno in senso lato). In questo caso si voglia stimare, con la tecnica LMMSE, un filtro  $B(\omega)$  che permetta di ottenere la miglior stima possibile di  $V(\omega)$  (e quindi anche di  $v(n)$ ).

Il principio di ortogonalità che sta alla base dello stimatore LMMSE comporta che vada cercata la condizione per la quale sia nulla la correlazione fra l'errore di stima ed i dati. Nel dominio delle frequenze ciò significa che sia nullo il valore atteso del prodotto fra la trasformata dell'errore di predizione e la coniugata della trasformata del dato:

$$E[(V(\omega) - B(\omega)Y(\omega))Y^*(\omega)] = 0$$

quindi (ricordando che  $v(n)$  e  $z(n)$  sono fra loro incorrelati)

$$B(\omega) = \frac{E[V(\omega)Y^*(\omega)]}{E[Y(\omega)Y^*(\omega)]} = \frac{\sigma_v^2 H^*(\omega)}{\sigma_v^2 |H(\omega)|^2 + \sigma_z^2} = \frac{H^*(\omega)}{H(\omega)H^*(\omega) + \frac{\sigma_z^2}{\sigma_v^2}} = \frac{1}{H(\omega)} \frac{1}{1 + \frac{\sigma_z^2/\sigma_v^2}{|H(\omega)|^2}}$$

quindi, pulsazione per pulsazione, l'approccio in frequenza determina  $B(\omega) = 1/H(\omega)$  se  $\frac{\sigma_z^2}{\sigma_v^2} \ll |H(\omega)|^2$  (il brusio, alla pulsazione considerata, è molto più basso del coro), quando invece succede il contrario il filtro  $B(\omega)$  tende a dare in uscita una versione fortemente attenuata (quasi nulla) di  $Y(\omega)$  (in caso di brusio troppo alto ci si tappa le orecchie).

Nel caso più generale di rumore additivo non bianco, per il quale risulta

$$E[|Z(\omega)|^2] = N(\omega)$$

si ha

$$B(\omega) = \frac{E[W(\omega)Y^*(\omega)]}{E[Y(\omega)Y^*(\omega)]} = \frac{\sigma_v^2 H^*(\omega)}{\sigma_v^2 |H(\omega)|^2 + N(\omega)} = \frac{H^*(\omega)}{H(\omega)H^*(\omega) + \frac{N(\omega)}{\sigma_v^2}} = \frac{1}{H(\omega)} \frac{1}{1 + \frac{N(\omega)}{\sigma_v^2 |H(\omega)|^2}}$$

In questo caso per le pulsazioni in cui  $\sigma_v^2 |H(\omega)|^2 \gg N(\omega)$  (la potenza del segnale utile è preponderante su quella di rumore) si ha  $B(\omega) = 1/H(\omega)$ , nel caso esattamente contrario  $B(\omega) \approx 0$  e poi vi sono tutti i casi intermedi dove il filtro ottimo ha l'andamento dettato dalla formula completa su riportata.

Il filtro  $B(\omega)$  qui considerato prende normalmente il nome di filtro di Wiener. In questa analisi si è assunto di conoscere  $H(\omega)$ , mentre nel caso dell'analisi nel dominio "del tempo" si era assunto di conoscere  $\mathbf{R}_x$  e quindi  $|H(\omega)|^2$ . Va notato che se si fa l'ipotesi che  $H(\omega)$  abbia fase minima è possibile fattorizzare  $|H(\omega)|^2$  in modo da ottenere  $H(\omega)$  e  $H^*(\omega) = H(-\omega)$ .



## 8.2 Riduzione del rumore

La tecnica fin qui presentata può essere utilizzata anche per realizzare sistemi di cancellazione del rumore, quando si ha a disposizione, oltre al segnale da "ripulire" anche un segnale secondario ben correlato con il rumore. Si pensi, ad esempio, al pilota di un elicottero che riceve in cuffia la voce del suo secondo a cui è sovrapposto il rumore delle pale e del motore. E' facile però avere un segnale secondario acquisito nelle vicinanze del rotore che raccoglie solo "rumore". Filtrando questo segnale secondario e sottraendolo al segnale principale è possibile renderlo incorrelato (entro il possibile) con il rumore e quindi migliore sensibilmente ciò che il pilota ascolta (vedi fig. 5). Ovviamente si assume che la voce del copilota ed il rumore dell'elicottero siano fra loro incorrelati.

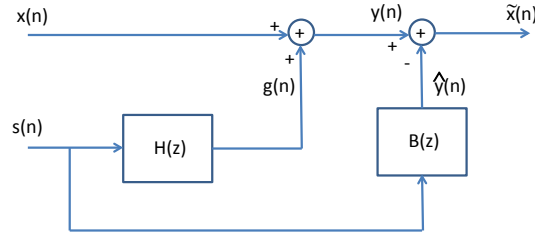


Figure 5: Schema per il sistema di cancellazione del rumore:  $y(n)$  è il segnale rumoroso, mentre  $s(n)$  è il segnale secondario.

In altre parole si vuole costruire un segnale  $\hat{y}(n)$  quanto più simile possibile a  $y(n)$  (segnale primario), ma solo sulla base di  $s(n)$  (segnale secondario) e dei suoi campioni precedenti. In questo modo  $y(n) - \hat{y}(n)$  assomiglierà quanto più possibile a  $x(n)$ .

Ponendo  $\theta = y(n)$ ,  $\mathbf{s} = [s(n - N + 1), s(n - N + 2), \dots, s(n)]^T$  e  $\mathbf{b} = [b_{N-1}, b_{N-2}, \dots, b_0]$  si ha  $\hat{y}(n) = \mathbf{b}^T \mathbf{s}$ , in cui:

$$\begin{aligned} \mathbf{b} &= \mathbf{C}_{y(n)\mathbf{s}} \mathbf{C}_{\mathbf{ss}}^{-1} \\ \mathbf{C}_{y(n)\mathbf{s}} &= [E[y(n)s(n - N + 1)], E[y(n)s(n - N + 2)], \dots, E[y(n)s(n)]] \\ \mathbf{C}_{\mathbf{ss}} &= \begin{bmatrix} r_s(0) & r_s(-1) & r_s(-2) & \dots & r_s(1 - N) \\ r_s(1) & r_s(0) & r_s(-1) & \dots & \dots \\ r_s(2) & r_s(1) & r_s(0) & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots \\ r_s(N - 1) & \dots & \dots & \dots & r_s(0) \end{bmatrix} \end{aligned}$$

## 8.3 Riduzione adattativa del rumore

Facendo ancora riferimento all'esempio precedente (il pilota di elicottero che sente in cuffia la voce del suo co-pilota insieme al rumore del mezzo), ogniqualvolta il copilota sposta la sua testa cambia il legame fra il segnale secondario (acquisito in prossimità del rotore) ed il termine di disturbo acquisito dal microfono del copilota e quindi cambia  $\mathbf{C}_{y(n)\mathbf{s}}$ , mentre

ogni qualvolta cambia il regime di funzionamento del rotore cambia  $\mathbf{C}_{ss}$ . Ciò significa che il vettore  $\mathbf{b}$  deve essere continuamente aggiornato sulla base dei cambiamenti di  $\mathbf{C}_{y(n)\mathbf{s}}$  e di  $\mathbf{C}_{ss}$  a loro volta continuamente stimati in modo campionario sulla base dei campioni precedenti di  $y(n)$  ed  $s(n)$ . Fra l'altro il continuo aggiornamento di  $\mathbf{b}$  richiede una continua inversione di  $\mathbf{C}_{ss}$  ( $\mathbf{b} = \mathbf{C}_{y(n)\mathbf{s}}\mathbf{C}_{ss}^{-1}$ ). Il sistema diventa quindi difficilmente implementabili e sono stati sviluppati metodi alternativi più facilmente realizzabili.

Nella tecnica di "Widrow-Hoff" si ha

$$\mathbf{b}_{n+1} = \mathbf{b}_n + \tilde{\gamma}\epsilon(n)\mathbf{s}_n = \mathbf{b}_n + \tilde{\gamma}(y(n) - \hat{y}(n))\mathbf{s}_n \quad (25)$$

dove  $\mathbf{b}_n$  rappresenta il valore del vettore  $\mathbf{b}$  all'istante  $n$ , mentre  $\mathbf{s}_n$  è invece il vettore che raccoglie i campioni da  $s(n)$  a  $s(n - N + 1)$ . Il fattore  $\gamma$ , sempre molto piccolo, assicura un "lento" aggiornamento di  $\mathbf{b}$ . La lentezza è necessaria a garantire la stabilità del sistema.

Il gradiente dell'errore quadratico medio ( $Q(n) = E[|\epsilon(n)|^2] = E[|y(n) - \hat{y}(n)|^2]$ ) rispetto a  $\mathbf{b}_n$  (cioè alle sue componenti) è  $\nabla_{\mathbf{b}}Q(n) = -2E[\epsilon(n)\mathbf{s}_n]$ . Per cui se nell'equazione 25 ci fosse il termine  $E[\epsilon(n)\mathbf{s}_n]$  l'equazione 25 corrisponderebbe all'aggiornamento del vettore dei coefficienti nella direzione della minimizzazione dell'errore quadratico medio. Qualora il gradiente fosse nullo ci si troverebbe nel punto di predizione ottima e non ci sarebbe nessun aggiornamento dei pesi. Nella realizzazione pratica, non essendo possibile usare l'operatore di media di insieme  $E[\cdot]$  bisogna accettare di usare una stima molto mediocre, e cioè semplicemente  $\epsilon(n)\mathbf{s}_n$  (il valore istantaneo dell'errore per il vettore che rappresenta i campioni del segnale di riferimento). Questo obbliga ad un aggiornamento lento per non creare instabilità.

Se osserviamo l'aggiornamento ogni  $K$  istanti, invece che ad ogni istante, trascurando la variazione del filtro in questo intervallo  $K$  visto che  $\gamma$  è comunque piccolo, abbiamo:

$$\mathbf{b}_{n+K} = \mathbf{b}_n + \gamma \frac{1}{K} \sum_{k=0}^{K-1} \epsilon(n+k)\mathbf{s}_{n+k}; \quad \gamma = K \tilde{\gamma} \quad (26)$$

In questo caso il termine di aggiornamento può essere visto come una buona stima del gradiente, infatti:

$$\frac{1}{K} \sum_{k=0}^{K-1} \epsilon(n+k)\mathbf{s}_{n+k} \simeq E[\epsilon(n)\mathbf{s}_n] \quad (27)$$

Possiamo poi ritenere valida l'equazione (26), che riscriviamo, decidendo anche di porre  $K = 1$ .

$$\mathbf{b}_{n+1} = \mathbf{b}_n + \gamma E[\epsilon_n\mathbf{s}_n] \quad (28)$$

A questo punto è possibile studiare l'evoluzione del filtro, che è divenuta deterministica per l'approssimazione fatta in precedenza (27, 28):

$$\mathbf{b}_{n+1} = \mathbf{b}_n + \gamma E[\mathbf{s}_n(y(n) - \mathbf{s}_n^T \mathbf{b}_n)] \quad (29)$$

Indicando, ancora una volta, con  $\mathbf{C}_{ss}$  la matrice di covarianza dei campioni del segnale di riferimento e con  $\mathbf{C}_{y(n)\mathbf{s}}$  il vettore della correlazione tra segnale principale e campioni del segnale riferimento:

$$\mathbf{C}_{ss} = E[\mathbf{ss}^T]; \quad \mathbf{C}_{y(n)\mathbf{s}} = E[y(n)\mathbf{s}]$$

Sostituendo nell'equazione (29):

$$\mathbf{b}_{n+1} = (\mathbf{I} - \gamma \mathbf{C}_{ss}) \mathbf{b}_n + \gamma \mathbf{C}_{y(n)s}$$

Per  $n \rightarrow \infty$ ,  $\mathbf{b}_{n+1} = \mathbf{b}_n$  e quindi:

$$\mathbf{b}_\infty = \mathbf{C}_{ss}^{-1} \mathbf{C}_{y(n)s}$$

Ricordando che  $\mathbf{C}_{ss}$  e  $\mathbf{C}_{ss}^{-1}$  sono matrici simmetriche (avendo ipotizzato segnali reali), la relazione precedente può essere espressa anche come

$$\mathbf{b}_\infty = \mathbf{C}_{y(n)s} \mathbf{C}_{ss}^{-1}$$

A regime quindi il sistema si comporta come il classico stimatore LMMSE.

La differenza tra il filtro corrente  $\mathbf{b}_n$  ed il filtro a regime  $\mathbf{b}_\infty$  è il filtro errore  $\mathbf{g}_n$  che evolve nel tempo con l'equazione:

$$\mathbf{b}_{n+1} - \mathbf{b}_\infty = \mathbf{g}_{n+1} = (\mathbf{I} - \gamma \mathbf{C}_{ss}) \mathbf{g}_n = (\mathbf{I} - \gamma \mathbf{C}_{ss})^{n+1} \mathbf{g}_0 \quad (30)$$

Condizione necessaria perchè il filtro errore converga a zero è che i moduli di tutti gli autovalori della matrice  $\mathbf{I} - \gamma \mathbf{C}_{ss}$  siano inferiori ad uno, e cioè  $|1 - \gamma \lambda_n| < 1$ . Questo implica:

$$0 < \gamma < \frac{2}{\lambda_{max}}; \rightarrow \frac{2}{\lambda_{max}} > \frac{2}{\sum_{n=1}^N \lambda_n} = \frac{2}{Tr[\mathbf{C}_{ss}]} = \frac{2}{N\sigma_s^2} \quad (31)$$

e quindi:

$$\gamma \leq \frac{2}{N\sigma_s^2} \leq \frac{2}{\lambda_{max}} \quad (32)$$

Il semplice vincolo

$$\gamma \leq \frac{2}{N\sigma_s^2} \quad (33)$$

(dove il termine  $\sigma_s^2$  può essere facilmente stimato, ed aggiornato nel tempo, osservando il segnale di riferimento) è sicuramente una condizione conservativa per il problema che stiamo considerando.

Purtroppo il fatto di aver accettato  $E[\epsilon_n \mathbf{s}_n] = \epsilon_n \mathbf{s}_n$  (cioè aver approssimato il valore atteso di  $\epsilon_n \mathbf{s}_n$  con il suo valore istantaneo) fa sì che quando abbiamo trovato imponga solo che la media del filtro errore ( $\mathbf{g}_n$ ) converga a zero. Se vogliamo che anche la sua dispersione converga a zero il valore di  $\gamma$  deve essere almeno 10 volte più basso.

In caso di stazionarietà dei segnali, quando il sistema va a regime (la predizione è la migliore possibile), l'errore quadratico medio di predizione sarà incorrelato con i dati ( $\mathbf{s}$ ) (e con ogni loro combinazione lineare), quindi :

$$\begin{aligned} E[\epsilon^2] &= E[(y(n) - \mathbf{b}_\infty^T \mathbf{s}(n))(y(n) - \mathbf{b}_\infty^T \mathbf{s}(n))] \\ &= E[(y(n) - \mathbf{b}_\infty^T \mathbf{s}(n))y(n)] \\ &= \sigma_y^2 - \mathbf{b}_\infty^T E[\mathbf{s}(n)y(n)] = \sigma_y^2 - (\mathbf{C}_{ss}^{-1} \mathbf{C}_{y(n)s})^{-1} \mathbf{C}_{y(n)s} \\ &= \sigma_y^2 - \mathbf{C}_{y(n)s}^T \mathbf{C}_{ss}^{-1} \mathbf{C}_{y(n)s} \end{aligned} \quad (34)$$

Si osservi che l'approccio di Widrow e Hoff permette di realizzare la riduzione adattativa del disturbo in modo molto semplice senza necessità di stimare la covarianza del segnale di riferimento ( $\mathbf{C}_{ss}$ ) e  $\mathbf{C}_{y(n)s}$  ovviamente al prezzo di un sistema che richiederà del tempo per convergere.

Per ottimizzare il tempo di convergenza, nel caso di segnali stazionari converrà iniziare con un valore di  $\gamma$  "abbastanza" grande e poi ridurlo tanto più il sistema tenda alla convergenza. Nel caso invece di segnali non stazionari (caso molto frequente) la situazione è più complessa, da un lato si vorrebbe un  $\gamma$  piccolo per inseguire bene il segnale utile, ma al tempo stesso si vorrebbe  $\gamma$  grande per convergere rapidamente nel caso di variazione della statistica. Anche questo tema sarà approfondito nel corso di Elaborazione Numerica dei Segnali (Laurea Magistrale TLC).

## 9 Stima spettrale

La stima spettrale comprende tutte le tecniche per stimare la densità spettrale di potenza (PSD) di un processo casuale (stazionario almeno in senso lato) a partire dall'analisi di un tratto di una sua realizzazione. Parametro di qualità è l'errore tra la densità spettrale di potenza stimata e quella effettiva (o dualmente tra stima dell'autocorrelazione ed autocorrelazione vera del processo). Si distingue tra stima spettrale parametrica e stima spettrale non-parametrica a seconda che si assuma esista o meno un "meccanismo" di generazione del processo. Prima di procedere ad esaminare le diverse tecniche è necessaria l'analisi di alcuni concetti preliminari

### 9.1 PSD per osservazioni limitate (tempo continuo)

L'autocorrelazione e la PSD per processi complessi e tempo-continui sono date da

$$\begin{aligned} r_x(\tau) &= E[x^*(t)x(t+\tau)] = E[x(t)x^*(t-\tau)]; \quad \text{dove } r_{xx}(-\tau) = r_{xx}^*(\tau) \\ S_{xx}(\omega) &= \int_{-\infty}^{\infty} r_x(\tau)e^{-j\omega\tau} d\tau \end{aligned}$$

Si assuma ora che l'osservazione sia limitata alla realizzazione entro l'intervallo di osservazione  $[-T/2, T/2]$ . Valutando la trasformata di Fourier del segnale nell'intervallo suddetto si ottiene

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

da cui la densità spettrale di energia

$$\begin{aligned} |X(\omega)|^2 &= \int_{-T/2}^{T/2} \int_{-T/2}^{T/2} x(t) x^*(\xi) e^{j\omega\xi} e^{-j\omega t} dt d\xi \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x(t) \text{rect}(t/T) x^*(\xi) \text{rect}(\xi/T) e^{j\omega\xi} e^{-j\omega t} dt d\xi \quad (\text{per } t - \xi = \tau): \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x^*(\xi) x(\xi + \tau) \text{rect}((\xi + \tau)/T) \text{rect}(\xi/T) e^{-j\omega\tau} d\xi d\tau \end{aligned}$$

È ragionevole pensare che la densità spettrale di energia normalizzata rispetto all'intervallo di osservazione  $T$  rappresenti una stima della PSD

$$\hat{S}_x(\omega) = \frac{|X(\omega)|^2}{T} \quad [W/Hz]$$

Una prima verifica della ragionevolezza dello stimatore è la valutazione della sua polarizzazione e, nel caso, della congruenza con la grandezza da stimare (la potenza è non negativa e uno stimatore che restituisca una potenza negativa non ha alcun senso).

Per analizzare la polarizzazione dello stimatore calcoliamo

$$\begin{aligned} E[\hat{S}_{xx}(\omega)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \underbrace{\frac{1}{T} \text{rect}((\xi + \tau)/T) \text{rect}(\xi/T) d\xi}_{\frac{1}{T} \text{rect}(\tau/T) * \text{rect}(\tau/T)} \times r_x(\tau) e^{-j\omega\tau} d\tau \\ &= \int_{-\infty}^{\infty} w_T(\tau) \times r_x(\tau) e^{-j\omega\tau} d\tau = S_{xx}(\omega) * W_T(\omega) \end{aligned}$$

dove:

$$w_T(\tau) = 1 - \frac{|\tau|}{T}, |\tau| < T \longleftrightarrow W_T(\omega) = T \left( \frac{\sin \omega T/2}{\omega T/2} \right)^2$$

rappresenta la finestra triangolare per segnale tempo continuo. Si noti che  $S_x(\omega) \geq 0$  per definizione, come pure risulta  $W_T(\omega) \geq 0$ . In conseguenza lo stimatore  $\hat{S}_x(\omega)$  risulta uno stimatore congruente (non è mai negativo) ma polarizzato visto che  $E[\hat{S}_x(\omega)] \neq S_x(\omega)$ . Se poi si assume che si annulli già per valori dell'argomento ben inferiori a  $T$  allora  $E[|X(\omega)|^2] \simeq T \cdot S_{xx}(\omega)$ . A questa stessa conclusione si può arrivare assumendo  $r_x(\tau)$  di estensione finita e considerando  $T \rightarrow \infty$ : si ha  $W_T(\omega) = \delta(\omega)$  visto che  $w_T(\tau) \rightarrow 1$  e quindi:

$$\lim_{T \rightarrow \infty} \hat{S}_{xx}(\omega) = S(\omega)$$

ovvero lo stimatore è asintoticamente non polarizzato.

Dimensionalmente si verifica che la densità spettrale di energia, relativa all'intervallo  $T$ , può essere scritta, utilizzando il teorema di Parseval, come:

$$\int_{-\infty}^{\infty} |X(\omega)|^2 \frac{d\omega}{2\pi} = \int_{-T/2}^{T/2} |x(t)|^2 dt = T \times P_x$$

dove con  $P_x$  abbiamo indicato la potenza istantanea media del segnale  $x(t)$ . Quindi

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

e conseguentemente  $\hat{S}_x(\omega)$  corrisponde effettivamente ad una stima della PSD del segnale e cioè alla ripartizione della potenza del segnale nelle varie bande spettrali. Si osservi che  $\hat{S}_x(\omega)$  non è necessariamente simmetrico rispetto ad  $\omega = 0$  in quanto il segnale  $x(t)$  può essere complesso.

Nel caso di segnali discreti la situazione è leggermente diversa. Considerando, ad esempio, un segnale  $x(t)$  che rappresenti una tensione, nel tempo-discreto (dopo campionamento) viene preservata la dimensionalità

$$\begin{aligned} x(t) &\rightarrow [V]; \\ x[n] &= x(n\Delta t) \rightarrow [V]. \end{aligned}$$

Le trasformate di Fourier tempo-continuo hanno dimensionalità

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

mentre le trasformate tempo-discreto hanno invece la stessa dimensionalità del segnale

$$X[k] = \sum_{n=0}^{N-1} x[n]W_N^{-nk} \rightarrow [V]$$

I campioni della trasformata tempo-continua si ottengono da quelli della trasformata tempo-discreto moltiplicando per il passo di campionamento  $\Delta t$  (sempre che il campionamento sia stato scelto correttamente). Assumendo  $\omega_k = \frac{1}{\Delta t k}$  con  $k = 0 \dots N-1$  si ha:

$$X(\omega_k) = \int_{-\infty}^{\infty} x(t)e^{-j\omega_k t} dt \simeq \Delta t \sum_{n=0}^{N-1} x[n]W_N^{-nk} = \Delta t \cdot X[k]$$

La di stima della PSD abbiamo visto essere data da

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

per avere la stessa dimensionalità lavorando nel tempo-discreto

$$\frac{|X(\omega_k)|^2}{T} \simeq \frac{|X[k]|^2 (\Delta t)^2}{N \Delta t} = \frac{\Delta t}{N} |X[k]|^2 =$$

è quindi necessario scalare prendere i valori del modulo quadro della trasformata tempo-discreto ( $|X[k]|^2$ ) dividerli per  $N$  (numero di campioni considerati) e moltiplicare per l'effettivo passo di campionamento  $\Delta t$ .

## 10 Stima spettrale non-parametrica

Si assuma di disporre di una realizzazione limitata di un processo stazionario a media nulla ( $E[x[n]] = 0$ ) e sequenza di autocorrelazione  $r[m] = E[x[n]^* x[n+m]]$ , si vuole stimare da questa la PSD del processo senza imporre alcun vincolo a-priori sulla forma della PSD. Lo stimatore che ne risulta è così importante che ha un nome specifico: *periodogramma*. Verrà derivato il periodogramma e se ne valuteranno le caratteristiche qualitative (polarizzazione e varianza). Nel seguito le definizioni non conterranno il pedice che contraddistingue il segnale in quanto tale notazione non rilevante ai fini della chiarezza.

## 10.1 Stima campionaria dell'autocorrelazione

Data una realizzazione  $\{x[n]\}_{n=0}^{N-1}$  sono possibili due stime campionarie della sequenza di autocorrelazione (o per brevità autocorrelazione):

$$\tilde{r}[m] = \frac{1}{N - |m|} \sum_{n=0}^{N-1} x[n]^* x[n+m] \quad \text{per } m \in (-N, N) \rightarrow \quad E[\tilde{r}[m]] = \begin{cases} r[m] & \text{per } |m| \leq (N - 1) \\ 0 & \text{altrove} \end{cases}$$

$$\hat{r}[m] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]^* x[n+m] \quad \text{per } m \in (-N, N) \rightarrow \quad E[\hat{r}[m]] = \begin{cases} \frac{N-|m|}{N} r[m] & \text{per } |m| \leq (N - 1) \\ 0 & \text{altrove} \end{cases}$$

la prima rappresenta una stima non-polarizzata (la verifica è semplice in quanto, per ogni valore di  $m$ , nel calcolo del valore atteso si sommano proprio  $N - |m|$  campioni ovvero tanti quanti indicati nella normalizzazione a denominatore) mentre la seconda rappresenta una stima polarizzata (in quanto si divide sempre per  $N$  indipendentemente dal numero di termini non nulli della sommatoria). Si osserva che entrambe le sequenze rappresentano una stima della funzione di autocorrelazione ma hanno supporto limitato in quanto  $\tilde{r}[m] = \hat{r}[m] = 0$  per  $|m| \notin [0, N - 1]$ . Si può pensare ad esse come sequenze di autocorrelazione di lunghezza infinita moltiplicate per una finestra rettangolare di lunghezza complessiva  $2N - 1$  campioni.

Calcolando la trasformata di Fourier delle due stime campionarie dell'autocorrelazione segue

$$\begin{aligned} \tilde{R}(\omega) &= \sum_{m=-(N-1)}^{N-1} \tilde{r}[m] e^{-j\omega m} \\ \hat{R}(\omega) &= \sum_{m=-(N-1)}^{N-1} \hat{r}[m] e^{-j\omega m} \end{aligned} ;$$

usando queste come stima della PSD si vuole verificare dal calcolo del valore atteso (polarizzazione) la corrispondenza con quanto auspicabile come stimatore della PSD:

$$\begin{aligned} E[\tilde{R}(\omega)] &= \sum_{m=-(N-1)}^{N-1} E[\tilde{r}[m]] e^{-j\omega m} = \sum_n w^{(R)}[m] \cdot r[m] e^{-j\omega m} = W^{(R)}(\omega) * S(\omega) \\ E[\hat{R}(\omega)] &= \sum_{m=-(N-1)}^{N-1} E[\hat{r}[m]] e^{-j\omega m} = \sum_n w^{(B)}[m] \cdot r[m] e^{-j\omega m} = W^{(B)}(\omega) * S(\omega) \end{aligned} \quad (35)$$

dove e' indicato con  $w^{(R)}[m]$  e  $w^{(B)}[m]$  la finestra rettangolare e triangolare (o finestra di Bartlett), rispettivamente, di  $2N - 1$  campioni. Le trasformate (DFT) sono:

$$\begin{aligned} w^{(R)}[m] = 1 \quad |m| < N &\longleftrightarrow W^{(R)}(\omega) = \frac{\sin[\omega(2N-1)/2]}{\sin[\omega/2]} \\ w^{(B)}[m] = \frac{N-|m|}{N} \quad |m| < N &\longleftrightarrow W^{(B)}(\omega) = \frac{1}{N} \left( \frac{\sin[\omega N/2]}{\sin[\omega/2]} \right)^2 \end{aligned}$$

Si osserva dall'analisi della (35) che la trasformata di Fourier dello stimatore non-polarizzato della funzione di autocorrelazione ha una stima della PSD che può assumere valori negativi (nota che la trasformata  $W^{(R)}(\omega)$  può assumere valori negativi). Pertanto lo stimatore non-polarizzato dell'autocorrelazione, per quanto la mancanza di polarizzazione nella stima rappresenti un indicatore della mancanza di errori sistematici, non è utilizzabile per la stima della PSD. Lo stimatore polarizzato dell'autocorrelazione ha trasformata che è semidefinita positiva ed è per questo impiegato come stimatore della PSD. Si può notare che, per  $N \rightarrow \infty$ ,

la finestra triangolare assume un valore costante pari a 1 e quindi la sua trasformata diventa impulsiva. Di conseguenza:

$$\lim_{N \rightarrow \infty} E[\widehat{R}(\omega)] = S(\omega)$$

(stimatore asintoticamente non-polarizzato). Sulla base di questa premessa si vuole ora trovare un modo efficiente per stimare la PSD, anche se polarizzata, impiegando la DFT. Si giunge quindi a definire il periodogramma come stimatore della densità spettrale di potenza.

## 10.2 Periodogramma

Sia  $X(\omega) = \sum_{n=0}^{N-1} x[n]e^{-j\omega n}$  trasformata di Fourier della sequenza limitata  $x[0], \dots, x[N-1]$ , si definisce periodogramma la funzione

$$S_N(\omega) = \frac{1}{N} |X(\omega)|^2 = \frac{1}{N} \sum_{n=0}^{N-1} \sum_{k=0}^{N-1} x[n]x[k]^* e^{-j\omega n} e^{j\omega k} \quad (36)$$

L'espressione può essere vista come il prodotto della DFT della sequenza  $x(n)$  (assunta di lunghezza limitata) con la DFT di se stessa ribaltata e coniugata. Ricordando che

$$\widehat{r}[m] = \frac{1}{N} \sum_{n=0}^{N-1} x[n]^* x[n+m] = \frac{1}{N} \sum_{l=0}^{N-1} x[l]x[l-m]^*$$

(l'autocorrelazione è la convoluzione della sequenza di interesse con se stessa risulta e coniugata)  $\widehat{R}(\omega) = \frac{1}{N} |X(\omega)|^2$  e quindi, come già osservato per  $\widehat{R}(\omega)$ , il periodogramma rappresenta uno stimatore polarizzato della PSD:

$$E[S_N(\omega)] = \frac{1}{N} \left( \frac{\sin[\omega N/2]}{\sin[\omega/2]} \right)^2 \circledast S(\omega) \implies \lim_{N \rightarrow \infty} E[S_N(\omega)] = S(\omega) \quad (37)$$

dove la convoluzione circolare è calcolata sul periodo di  $2\pi$ .

Dal punto di vista computazionale è spesso più conveniente stimare la PSD calcolando il periodogramma ( $|X(\omega)|^2$ ) piuttosto che stimare la funzione di autocorrelazione e poi trasformare.

Può sorgere un dubbio sull'utilizzare la (36) per calcolare la stima della PSD. La (36) può essere vista come la trasformata della convoluzione della sequenza con se stessa ribaltata e coniugata. Ma la convoluzione considerata è di tipo circolare. Al contrario il calcolo della stima dell'autocorrelazione è equivalente ad una convoluzione lineare. In realtà è possibile bordare la sequenza  $\{x[n]\}$  con  $N$  zeri prima di trasformare in modo tale che la periodizzazione della convoluzione (corrispondente al prodotto in  $\omega$ ) su  $2N$  campioni restituisca, entro  $2N - 1$  campioni, un risultato identico a quello della convoluzione lineare. I valori della DFT calcolata su  $2N$  campioni e valutati quindi alle frequenze di passo  $2\omega_k = \pi k/2N$  non cambiano rispetto ai valori della DFT calcolata su  $N$  campioni. L'aggiunta degli zeri nei tempi corrisponde ad una interpolazione nelle frequenze in cui si passa da  $N$  campioni a  $2N$  campioni (in cui, a passo alterno,  $N$  valori sono interpolati ed altri  $N$  corrispondenti alla DFT su  $N$  campioni). In questo caso il raddoppio della lunghezza della sequenza corrisponde all'interpolazione del valore intermedio con un interpolatore che rispetta il dato. Da questo segue che la DFT su  $N$  campioni restituisce il periodogramma corretto indipendentemente dalla periodizzazione.



### 10.3 Varianza del periodogramma

Si è visto che il periodogramma è uno stimatore della PSD solo asintoticamente non-polarizzato. Per valutarne l'accuratezza è necessario anche conoscerne la varianza.

Assumendo per semplicità di scrittura la sequenza come reale si ha:

$$\text{var}[S_N(\omega)] = \frac{1}{N^2} \sum_{s=0}^{N-1} \sum_{t=0}^{N-1} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} E[x[s] \cdot x[t] \cdot x[u] \cdot x[v]] e^{-j\omega(s-t+u-v)} - E^2[S_N(\omega)]$$

Si osserva che il calcolo della varianza richiede di valutare momenti di ordine 4 e quindi non è affatto semplice. Il calcolo diventa un po' più semplice se si considera il processo che genera la sequenza a disposizione come bianco e con varianza  $\sigma_x^2$ . Con questa assunzione, utilizzando  $\omega$  continuo (cioè la DTFT), e con passaggi tutt'altro che banali si può arrivare al seguente risultato:

$$\text{var}[S_N(\omega)] = \sigma_x^4 \left[ 1 + \left( \frac{\sin \omega N}{N \sin \omega} \right)^2 \right]$$

Quindi la varianza del periodogramma non tende a zero per  $N \rightarrow \infty$ , da cui segue che il periodogramma non rappresenta uno stimatore consistente della PSD del processo  $\{x_n\}$ . Nel corso di "Digital Signal Processing", Laurea Specialistica in Ing. TLC, si approfondisce come ottenere questo risultato.

Alla luce dell'espressione della varianza del periodogramma su riportata si può dire che questa varianza è dell'ordine di  $\sigma_x^4$ , in modo indipendente da  $N$ . Quindi il semplice periodogramma non è particolarmente adatto ad essere utilizzato come stimatore spettrale, non solo per la polarizzazione, ma perché lo scarto della stima è dello stesso ordine di grandezza della quantità da stimare. All'atto pratico, l'effetto che si osserva è la comparsa di rapide fluttuazioni al posto di un andamento "piatto" (nell'ipotesi di processo bianco).

Passando all'uso della DFT per il periodogramma ( $\omega$  campionato,  $N$  valori sull'intervallo  $2\pi$ ) la varianza diventa (sempre per sequenze reali):

$$\text{var}[S_N(\omega_k)] = \sigma_x^4 \left[ 1 + \left( \frac{\sin 2\pi k}{N \sin 2\pi k/N} \right)^2 \right] = \sigma_x^4 (1 + \delta(k) + \delta(k - N/2)), \quad (38)$$

pertanto questa assume un valore doppio ( $2\sigma_x^4$ ) solo alla frequenza  $\omega_k = 0$  e alla frequenza di Nyquist  $\omega_{N/2} = \pi$  (se esiste).

Ovviamente anche quando viene usata la DFT il periodogramma è uno stimatore *inconsistente* della PSD in quanto la varianza della stima non decresce all'aumentare della lunghezza del dato a disposizione per la stima. Nel caso dell'uso della DFT, la motivazione di questo comportamento è di facile intuizione, all'aumentare della lunghezza ( $N$ ) del dato a disposizione si aggiunge risoluzione (il passo in frequenza decresce con  $N$  in quanto proporzionale a  $1/N$ ) e quindi non aumenta il numero di campioni mediamente a disposizione per ogni elemento del vettore stimato). Nel caso di sequenze reali si hanno a disposizione "in media" due campioni per ogni valore della PSD (non ridondante) in frequenza.

Per ridurre la varianza del periodogramma è necessario fare la media di più realizzazioni indipendenti del periodogramma ottenute su finestre temporali disgiunte sul segnale. Ad esempio, se si dispone di  $M$  sequenze da  $N$  campioni (ottenute segmentando una osservazione

da  $MN$  campioni in  $M$  sottosequenze di durata  $N$ ) la media di  $M$  periodogrammi relativi a sequenze indipendenti consente una riduzione della varianza di un fattore  $1/M$ .

*Osservazione 1:* La relazione (38) ha validità per solo rumore gaussiano bianco reale. Nella pratica se il rumore è reale ma non bianco si approssima la varianza con la relazione indicata di seguito

$$\text{var}[S_N(\omega_k)] \simeq \{S(\omega_k)\}^2 (1 + \delta(k) + \delta(k - N/2)) \quad (39a)$$

in cui si trascura la polarizzazione (ovvero se  $N \gg 1$ ). Si noti che comunque, per sequenze reali, la varianza è assunta doppia alla cella  $k = 0$  e  $N/2$  (se disponibile, cioè  $N$  pari).

*Osservazione 2:* Il periodogramma è una trasformazione non-lineare delle osservazioni e rappresenta una variabile casuale che dipende dalla realizzazione. Se la sequenza  $\{x[n]\}$  è reale valgono le regole di simmetria della trasformata di Fourier da cui  $S_N(\omega_k) = S_N(\omega_{-k})$  e quindi non è possibile ridurre (dimezzare) la varianza calcolando la stima della PSD facendo la media tra  $S_N(\omega_k)$  e  $S_N(\omega_{-k})$ . Tuttavia se la sequenza  $\{x[n]\}$  è complessa allora le relazioni di simmetria si perdono e  $S_N(\omega_k) \neq S_N(\omega_{-k}) = S_N(\omega_{N-k})$ .

## 10.4 La densità di probabilità del periodogramma

Si consideri una sequenza  $\mathbf{x}$  di lunghezza  $N$  vista come una parte di una realizzazione associata ad un processo stazionario gaussiano bianco. Ricordando quanto già visto nelle sezioni precedenti (in particolare 1.3) ed assumendo  $N$  pari, si ha

$$\begin{aligned} \mathbf{x} &\sim N(0, \sigma_x^2 \mathbf{I}_N) \\ \mathbf{X} &= (\mathbf{W}_{N(R)} + j\mathbf{W}_{N(I)})\mathbf{x} = \underbrace{\mathbf{W}_{N(R)}\mathbf{x}}_{\mathbf{x}_R} + j\underbrace{\mathbf{W}_{N(I)}\mathbf{x}}_{\mathbf{x}_I} \\ \begin{bmatrix} X_R[0] \\ X_R[1] \\ \vdots \\ X_R[N/2] \end{bmatrix} &\sim N(0, \frac{N\sigma_x^2}{2} \text{diag}\{2, \underbrace{1, 1, \dots, 1}_{N/2-1}, 2\}) \\ \begin{bmatrix} X_I[0] \\ X_I[1] \\ \vdots \\ X_I[N/2] \end{bmatrix} &\sim N(0, \frac{N\sigma_x^2}{2} \text{diag}\{0, \underbrace{1, 1, \dots, 1}_{N/2-1}, 0\}) \end{aligned}$$

in cui si è limitata l'analisi ai soli campioni indipendenti (per la simmetria della trasformata di Fourier per sequenze reali). Nel caso di  $N$  dispari non vi sarà il termine di  $\mathbf{X}$  corrispondente alla frequenza di Nyquist. Nel seguito si considera solo  $N$  pari, sia perchè questo è il caso più frequente, sia perchè l'estensione di quanto si illustra al caso di  $N$  dispari è immediata.

Pensando al modulo quadro della trasformata (che corrisponde al periodogramma a meno della divisione per  $N$ ) si può scrivere

$$|X[k]|^2 = \begin{cases} X_R[k]^2 \sim \chi_1^2 & \text{per } k = 0, N/2 \\ X_R[k]^2 + X_I[k]^2 \sim \chi_2^2 & \text{per } k \neq 0, N/2 \end{cases}$$

ovvero (per  $k \neq 0, N/2$ )  $|X[k]|^2 \sim \chi_2^2$  (in quanto somma del quadrato di due vc gaussiane indipendenti) e quindi la pdf è esponenziale. Per  $k = 0, N/2$  la pdf è il quadrato di una gaussiana (funzione Chi Quadro con 1 grado di libertà). Ricordando come valutare la pdf di una funzione di variabile casuale, posto  $y = X_R[k]^2$  e  $\sigma_y^2 = N\sigma_x^2$  risulta

$$f_y(y) = \frac{e^{-\frac{y}{2\sigma_y^2}}}{\sigma_y\sqrt{2\pi y}}, \text{ con } y > 0.$$

Ovviamente risulta (per ogni  $k$ )

$$E[|X[k]|^2] = N\sigma_x^2$$

Inoltre (sempre per ogni possibile  $k$ ), ricordando che  $E[X_R^4] = E[X_I^4] = 3\frac{2N^2\sigma_x^4}{4}$  e  $E[X_R^2X_I^2] = E[X_R^2]E[X_I^2] = N^2\sigma_x^4$  si ha

$$var[|X[k]|^2] = E[(X_R^2[k] + X_I^2[k])^2] - N^2\sigma_x^4 = N^2\sigma_x^4[1 + \delta(k) + \delta(k - N/2)]$$

confermando che la varianza del periodogramma è doppia alla continua e al Nyquist ( $N$  pari).

Per sequenze complesse circolari si ha

$$\mathbf{x} \sim CN(0, \sigma_x^2 \mathbf{I}_N) \rightarrow \begin{cases} \mathbf{x}_R \sim CN(0, \frac{\sigma_x^2}{2} \mathbf{I}_N) \\ \mathbf{x}_I \sim CN(0, \frac{\sigma_x^2}{2} \mathbf{I}_N) \end{cases}$$

$$\mathbf{X} = (\mathbf{W}_{N(R)} + j\mathbf{W}_{N(I)})(\mathbf{x}_R + j\mathbf{x}_I) = \underbrace{\mathbf{W}_{N(R)}\mathbf{x}_R - \mathbf{W}_{N(I)}\mathbf{x}_I}_{\mathbf{x}_R} + j \underbrace{\mathbf{W}_{N(I)}\mathbf{x}_R + \mathbf{W}_{N(R)}\mathbf{x}_I}_{\mathbf{x}_I}$$

$$\begin{aligned} \mathbf{X}_R &\sim N(0, \frac{N\sigma_x^2}{2} \mathbf{I}_N) \\ \mathbf{X}_I &\sim N(0, \frac{N\sigma_x^2}{2} \mathbf{I}_N) \end{aligned} \rightarrow \mathbf{X} \sim CN(0, N\sigma_x^2 \mathbf{I}_N)$$

$$|X[k]|^2 = X_R[k]^2 + X_I[k]^2 \sim \chi_2^2 \rightarrow \begin{aligned} E[|X[k]|^2] &= N\sigma_x^2 \\ var[|X[k]|^2] &= (N\sigma_x^2)^2 \end{aligned}$$

Rispetto alla sequenza reale non esiste differenza di varianza tra i campioni del periodogramma, inoltre essi sono tutti tra di loro indipendenti (non esistono le simmetrie che vi erano nel caso di sequenza reale).

Quando la sequenza considerata non è gaussiana bianca, valutare le caratteristiche statistiche del periodogramma diventa molto più complicato e non è qui considerato, rimane comunque vero quanto indicato in 39a.

## 10.5 Polarizzazione, risoluzione e varianza

### 10.5.1 Polarizzazione e risoluzione

Si è visto che il periodogramma è uno stimatore polarizzato della PSD. L'entità della polarizzazione dipende da due termini: lunghezza della finestra di osservazione (in quanto lo stimatore è asintoticamente non-polarizzato) e forma di una eventuale funzione di pesatura  $w[n]$  dei dati (finestratura). La finestratura introduce, in generale, una riduzione della risoluzione spettrale (l'uso di una finestra corrisponde, nel dominio delle frequenze, alla convoluzione con la trasformata della finestra stessa caratterizzata dalla presenza di un ben specifico lobo principale che introduce smussamento) e una polarizzazione locale (*spectral leakage*) dovuta ai lobi laterali della trasformata della finestra. Per valutare entrambi questi effetti si consideri cosa "accade" alla frequenza  $\omega_o$  ai valori assunti dalla PSD e dal valore atteso del periodogramma,  $S(\omega_o)$  e  $E[S_N(\omega_o)]$  rispettivamente. La convoluzione (circolare) da cui si ottiene il valore atteso del periodogramma (37) può essere riscritta dividendo il contributo dovuto al lobo principale dai contributi dovuti ai lobi secondari

$$E[S_N(\omega_o)] = \int_{\omega \in L_p} S(\omega) W_N(\omega_o - \omega) d\omega + \int_{\omega \in L_s} S(\omega) W_N(\omega_o - \omega) d\omega$$

dove  $L_p$  rappresenta l'intervallo di frequenze, intorno a  $\omega_o$ , relativo al lobo principale di  $W_N(\omega)$  (trasformata della finestra  $w_n$ ) mentre  $L_s$  è l'intervallo complementare, cioè quello relativo ai lobi secondari. Si supponga, per il momento, trascurabile il contributo dovuto al secondo integrale. Si osserva che nel calcolo di  $E[S_N(\omega_o)]$  intervengono tutti i valori di  $S(\omega)$  in un intorno di  $\omega_o$  di larghezza pari all'ampiezza del lobo principale di  $W_N(\omega)$ . A causa della finestra il periodogramma non risolve contributi diversi all'interno dell'intervallo di larghezza  $L_p$ . Meta' della larghezza del lobo principale rappresenta il termine di *risoluzione spettrale*:  $\Delta f_s = L_p/2$ . La risoluzione migliora (si riduce  $\Delta f_s$  o la larghezza del lobo principale) se si aumenta la lunghezza della sequenza di osservazione  $N$  in quanto  $\Delta f_s = \alpha/N$ . Il contributo del secondo integrale attribuisce alla frequenza  $\omega_o$  tutta la densità spettrale opportunamente pesata dai lobi laterali. In altre parole i lobi laterali contaminano la stima dello spettro alla frequenza  $\omega_o$  con contributi provenienti da altre zone dello spettro. Questo effetto di polarizzazione a larga banda dei contributi di  $S(\omega)$  è noto come *spectral leakage* e può essere ridotto scegliendo una finestra con lobi secondari di ampiezza contenuta (nel caso si deve finestrare la stima non polarizzata dell'autocorrelazione  $\tilde{r}[m]$  con una finestra la cui trasformata di Fourier sia definita positiva, una finestra triangolare potrebbe essere adeguata così come si ha implicitamente nella stima polarizzata dell'autocorrelazione la cui trasformata di Fourier è il periodogramma).

Nel caso di spettro bianco (o localmente piatto) entrambi i contributi non danno alcun contributo. Se lo spettro non è bianco la polarizzazione è più marcata. In particolare, nelle zone di transizione la polarizzazione è dominata dallo smussamento della transizione a causa della limitata risoluzione spettrale  $\Delta f_s$  (o dalla larghezza del lobo principale). Nel caso di spettro con ampia escursione dinamica, il valore del periodogramma in banda attenuata è prevalentemente dovuto ai contributi dovuti ai lobi laterali (nota che in questo caso la dinamica stimata non può essere maggiore del rapporto tra il lobo principale e i lobi laterali della finestra).

### 10.5.2 Riduzione della varianza nel periodogramma

Per la riduzione della varianza si possono mediare piu' periodogrammi relativi a sottosequenze indipendenti. Il periodogramma complessivo diventa la media di  $M$  periodogrammi semplici su  $N$  campioni:

$$S_{N,M}(\omega_k) = \frac{1}{M} \sum_{m=0}^{M-1} \left\{ \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n + mN] e^{-j\omega_k n} \right|^2 \right\}$$

Evidentemente e' ancora una stima polarizzata (non scompare la convoluzione con la trasformata della finestra triangolare) e la risoluzione spettrale dipende dalla lunghezza del singolo segmento. La varianza complessiva si riduce di un fattore  $1/M$  se i segmenti sono indipendenti. Nota che se la finestra di osservazione di un processo e' data, quindi si dovra' scegliere come soluzione di compromesso tra un'alta risoluzione (che spinge a scegliere  $N$  grande) e la necessita' di avere una bassa varianza (con  $M$  grande). La tecnica WOSA e' simile alla segmentazione appena esaminata ma con finestre a raccordo dolce, ne segue una perdita di risoluzione spettrale ma una ridotta polarizzazione dovuta a contaminazioni a larga banda. Nel caso di finestre "dolci" e rastremate a "0" la sovrapposizione delle finestre e' necessaria per sfruttare tutta l'informazione contenuta nelle osservazioni. Tuttavia la varianza ha ancora una riduzione  $\div 1/M$  in quanto le stime non sono tra di loro indipendenti. Si puo' ridurre la varianza calcolando il periodogramma su  $N$  campioni e poi smussando lo spettro con la convoluzione mediante una opportuna funzione in frequenza  $W(\omega_k)$  (nota che la larghezza della funzione  $W(\omega_k)$  e' necessariamente maggiore della risoluzione introdotta dall'implicita finestatura nei dati):

$$\bar{S}_N(\omega_k) = \frac{1}{N} \sum_{n=0}^{N-1} S_N(\theta_n) W(\omega_k - \theta_n)$$

Il filtro di smussamento puo' essere applicato sia direttamente nelle frequenze (nota che la convoluzione e' circolare) che come finestra della funzione di autocorrelazione nei tempi. Scegliendo di finestrare la funzione di autocorrelazione (questa tecnica, concettualmente simile allo smussamento del periodogramma, venne proposta da Blackmann e Tukey nel 1958) e' necessario notare che non tutte le finestre hanno trasformata positiva. Una semplice procedura per ottenere trasformate sempre positive consiste nell'uso di una finestra ottenuta convolvendo una finestra a spettro non strettamente positivo con se stessa e normalizzando l'ampiezza. La polarizzazione puo' essere approssimata dalla relazione:  $E[\bar{S}_N(\omega)] = 1/N \times S(\omega) \otimes W_N(\omega) \otimes W(\omega) \simeq 1/N \times \sum_{n=0}^{N-1} S(\theta_n) W(\omega_k - \theta_n)$ ; in altre parole il periodogramma smussato e' asintoticamente non-polarizzato purché  $\sum_{k=0}^{N-1} W(\omega_k)/N = w_0 = 1$ . Per la

varianza si ha:

$$\begin{aligned}
 \text{var}[\bar{S}_N(\omega_k)] &= E \left[ \frac{1}{N} \sum_{n=0}^{N-1} (S_N(\theta_n) - E[S_N(\theta_n)]) W(\omega_k - \theta_n) \right]^2 \\
 &= \frac{1}{N^2} \sum_{n=0}^{N-1} \text{var}[S_N(\theta_n)] W^2(\omega_k - \theta_n) = \\
 &= \frac{\sigma_x^4}{N^2} \left[ \sum_{n=0}^{N-1} W^2(\omega_n) + W^2(\omega_k) + W^2(\omega_{k-N/2}) \right]
 \end{aligned}$$

Si nota, tuttavia, che la presenza della componente continua o la componente alla frequenza di Nyquist introducono una degradazione delle prestazioni dello stimatore (in termini di varianza) quando si opera con sequenze reali. A questo proposito può essere conveniente rimuovere dalla media spettrale queste componenti o pesarle opportunamente. Questo tema sarà approfondito nel corso di Elaborazione Numerica dei Segnali (Laurea Magistrale TLC).