

Identificazione dei Modelli e Analisi dei Dati

Appunti

Scandella Matteo & Pezzoli Matteo

19 febbraio 2015

Indice

Introduzione	1
I. Processi stocastici	3
Introduzione	4
1. Richiami di statistica e modelli stocastici	5
2. Processi stocastici	9
3. Processi stocastici stazionari	10
3.1. Rappresentazione spettrale	12
3.2. Rappresentazione dinamica	14
3.3. Depolarizzazione	15
4. Tipi di processi	17
4.1. Processi MA (Moving Average)	17
4.2. Processi AR (AutoRegressive)	19
4.2.1. Processi $AR(1)$	20
4.3. Processi ARMA (AutoRegressive Moving Average)	23
4.4. Processi ARMAX (AutoRegressive Moving Average with eXogenous inputs)	24
4.5. Osservazioni teoriche	27
4.6. Curiosità	29
5. Teorema della fattorizzazione spettrale	30
5.1. Enunciato	30
5.2. Esempio	30
6. Filtro passa-tutto	32
7. Forma canonica	34
7.1. Procedimento di canonizzazione	37
II. Stimatori campionari	39
8. Richiami sulla teoria della stima	40
9. Media campionaria	42
10. Covarianza campionaria	43
10.1. Variante non corretta	43

11. Densità spettrale campionaria	45
11.1. Utilizzo del secondo stimatore della funzione di covarianza	46
11.2. Regolarizzazione	46
 III. Predizione	 48
Introduzione	49
12. Predittore ottimo di un processo $MA(n)$	51
13. Predittore ottimo di un processo $ARMA(m, n)$	53
13.1. Predittore a un passo ottimo di un processo $ARMA(m, n)$	56
13.2. Valutazione del predittore	56
14. Predittore ottimo di un processo $ARMAX(m, n, p)$	62
14.1. Predittore a un passo ottimo di un processo $ARMAX(m, n, 1)$	64
15. Predittore $ARMAX(n, m, p)$ a media non nulla	67
15.1. Depolarizzazione	67
15.2. Variabile esogena	70
16. Teoria di Kolmogorov-Wiener	73
 IV. Identificazione modelli	 74
Introduzione	75
17. Identificazione PEM di un modello ARX	77
17.1. Metodo alternativo per impostare il problema	80
18. Identificazione PEM di un modello $ARMAX$	84
18.1. Introduzione ai metodi iterativi per l'identificazione di un processo $ARMAX$	85
18.2. Introduzione al metodo di newton	85
18.3. Metodo del gradiente	87
18.4. Metodo di Newton	87
18.5. Metodi di quasi-Newton	87
18.6. Calcolo del gradiente dell'errore di predizione a un passo	88
18.7. Conclusione	90
19. Analisi asintotica dei metodi PEM	92
20. Problema di identificabilità	96
21. Valutazione dell'incertezza	100
22. Scelta della complessità del modello	103
Introduzione	103
22.1. Cross-validazione	104
22.2. Formule di stima della complessità ottima	105

V. Identificazione segnali non misurabili	107
Introduzione	108
23. Rappresentazione tramite variabili di stato	109
23.1. Caso di rumore non bianco	113
23.2. Osservabilità di sistema	113
23.3. Raggiungibilità di un sistema	114
23.3.1. Raggiungibilità dal rumore	114
24. Filtro di Kalman	116
24.1. Introduzione	116
24.2. Predittore a un passo di Kalman	116
24.3. Predittore a k passi di Kalman	117
24.4. Filtro di Kalman	118
24.5. Predittore e filtro di Kalman per sistemi tempo-varianti	118
24.6. Predittore e filtro di Kalman per sistemi con ingresso esogeno	118
24.7. Predittore e filtro di Kalman asintotico	119
24.8. Filtro di Kalman esteso(EKF)	121
24.9. Identificazione a scatola grigia	123
VI. Esercitazioni	126
25. Esercitazione del 15/10	127
26. Esercitazione del 26/11	146

Introduzione

Le attività ingegneristiche si scontrano con tre grandi problemi:

- Progettazione
Si risolve un problema progettando dal nulla un qualcosa.
- Modellistica
Si descrive con un linguaggio rigoroso, e quindi matematico, il comportamento di un sistema preesistente che va studiato
- Controllo
Si vuole far variare il comportamento a un sistema preesistente

Le tre attività sono correlate fra di loro, infatti non è possibile svolgerle in completa autonomia. In particolare per controllare un sistema occorre prima modellizzarlo.

Esistono due diversi approcci alla modellizzazione: white box e black box.

White box (scatola bianca) Questo tipo di modellazione utilizza principi fisici per cercare di modellizzare al meglio il problema, e viene affrontato in molti corsi in base al tipo di sistema (fisica tecnica, elettrotecnica, elettronica ...).

- Vantaggi:
 - I modelli white box sono facilmente generalizzabili per sistemi simili, in quanto si conosce il significato fisico di tutte le variabili usate.
Ad esempio se si modella un circuito RC, anche cambiando il valore della resistenza o del condensatore il modello rimane valido.
 - I modelli white box sono facilmente interpretabili fisicamente, ogni variabile ha un certo significato noto.
- Svantaggi:
 - La generazione di un modello white box è molto lenta e richiede conoscenze avanzate specifiche del problema. Quindi i costi sono molto elevati.
 - Se un sistema è molto complesso scrivere un modello white box è praticamente impossibile.

Black box (scatola chiusa) Questo tipo di modellazione sfrutta degli appositi esperimenti sul sistema per approssimare un modello. Quindi si forzano degli input e si guarda l'andamento delle uscite ignorando completamente come il sistema sia fatto fisicamente.

- Vantaggi:
 - Questi modelli sono di natura molto accurati dato che sono scritti in modo che approssimino il sistema il meglio possibile.
 - Si basano su poche assunzioni iniziali.
 - Si possono ricavare molto velocemente con varie tecniche.
 - Non è richiesta una conoscenza fisica accurata del problema, anche se un minimo è sempre gradita.

0. Introduzione

- La difficoltà nell'identificazione non aumenta con l'aumentare della complessità del sistema. Quindi si può applicare per ogni sistema nello stesso modo.
- Svantaggi:
 - I modelli non sono generalizzati, per ogni modifica bisogna ripetere l'esperimento dato che non si conosce il significato fisico delle variabili
 - I modelli non sono interpretabili fisicamente.

Confronto Fino a 40 anni fa si usavano solo i modelli white box perchè i sensori costavano molto e alcuni neppure esistevano rendendo la modellistica black box impossibile.

Questo richiedeva elevati costi per un gruppo di fisici che dovevano analizzare il sistema mettendoci anche anni, rendendo il tutto molto costoso, per questo motivo con l'abbassamento dei costi della sensoristica e dell'elettronica l'approccio black box ha preso sempre più piede abbassando drasticamente i costi sia per il tempo che per il personale.

Oggi la modellistica black box è largamente utilizzata ovunque ed è quindi un argomento di studio molto importante che è ancora in fase di ricerca e sviluppo.

Questa prima parte del corso si occupa proprio della modellazione black box, in particolare si studierà solo la parte che è ormai consolidata nella letteratura dato che per molti tipi di sistemi la ricerca è ancora in corso. Quindi si tratteranno solo sistemi dinamici LTI (lineari tempo invarianti) e SISO (single input single output).

Questa disciplina ha una principale differenza rispetto alla statistica che è la presenza di sistemi dinamici e non statici rendendo la teoria molto più complicata e ancora in fase di evoluzione.

Parte I.

Processi stocastici

Introduzione

Lo studio dei processi stocastici serve per poter rappresentare l'incertezza della misura delle uscite del sistema. Infatti quello che solitamente si fa è eseguire un esperimento che permette di ottenere un numero finito di misure di ingresso e uscita :

$$D_S = \{u(1), u(2), \dots, u(n), y(1), y(2), \dots, y(n)\}$$

Le quali sono soggette a vari errori, come quello di misura del sensore, come mostrato nella figura 0.1.

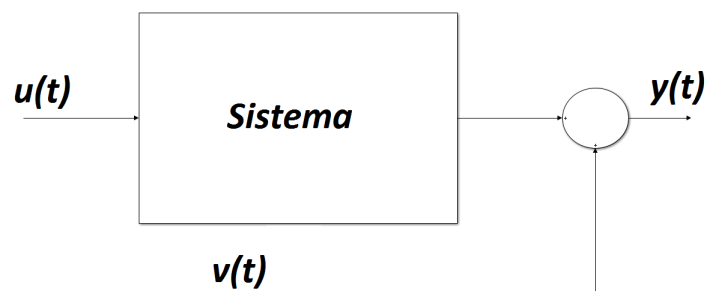


Figura 0.1.: Modello che mostra che la variabile di uscita $y(t)$ è composta anche da una componente $v(t)$ che ne rappresenta l'incertezza

1. Richiami di statistica e modelli stocastici

Definizione Una variabile casuale v è una variabile definita a partire dall'esito di un esperimento casuale.

Definizione La distribuzione di probabilità di una variabile casuale v è la funzione:

$$\begin{aligned} F &\rightarrow [0, 1] \\ F(q) &= P(v \leq q) \end{aligned}$$

dove con $P(v \leq q)$ si intende la probabilità che la variabile v assuma un valore \leq di q .

Definizione La densità di probabilità di una variabile casuale v , con distribuzione F , è la funzione:

$$f(q) = \frac{\partial F(q)}{\partial q}$$

Definizione Il valore atteso di una variabile casuale v , con densità di probabilità f , si definisce come:

$$\mathbb{E}[v] = \int_{-\infty}^{\infty} q \cdot f(q) dq$$

Proprietà L'operatore di valore atteso è lineare e quindi date due variabili casuali v_1 e v_2 si ha che:

$$\mathbb{E}[\alpha \cdot v_1 + \beta \cdot v_2 + \gamma] = \alpha \cdot \mathbb{E}[v_1] + \beta \cdot \mathbb{E}[v_2] + \gamma, \forall \alpha, \beta, \gamma \in \mathbb{R}$$

Definizione La varianza di una variabile casuale v , con densità di probabilità f , si definisce come:

$$\begin{aligned} Var[v] &= \int_{-\infty}^{\infty} (q - \mathbb{E}[v])^2 \cdot f(q) dq \\ &= \mathbb{E}[(v - \mathbb{E}[v])^2] \end{aligned}$$

A volte per comodità di uso anche:

$$\sigma[v] = \sqrt{Var[v]}$$

Osservazioni

- Dato che la varianza è definita come una somma di quadrati è un valore sempre positivo

1. Richiami di statistica e modelli stocastici

- Si può notare che:

$$\begin{aligned}
 Var[v] &= \mathbb{E}[(v - \mathbb{E}[v])^2] = \\
 &= \mathbb{E}[v^2 - 2 \cdot v \cdot \mathbb{E}[v] + \mathbb{E}[v]^2] = \\
 &= \mathbb{E}[v^2] - 2 \cdot \mathbb{E}[v \cdot \mathbb{E}[v]] + \mathbb{E}[\mathbb{E}[v]^2] = \\
 &= \mathbb{E}[v^2] - 2 \cdot \mathbb{E}[v] \cdot \mathbb{E}[v] + \mathbb{E}[v]^2 = \\
 &= \mathbb{E}[v^2] - 2 \cdot \mathbb{E}[v]^2 + \mathbb{E}[v]^2 = \\
 &= \mathbb{E}[v^2] - \mathbb{E}[v]^2
 \end{aligned}$$

- Data una variabile casuale v si ha che:

$$Var[\alpha \cdot v_1 + \beta] = \alpha^2 \cdot Var[v_1], \forall \alpha, \beta \in \mathbb{R}$$

Definizione La variabile casuale v si definisce normale o gaussiana se si distribuisce con la seguente funzione di densità:

$$f(q) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{(q-\mu)^2}{2\sigma^2}}$$

con:

$$\begin{aligned}
 \mathbb{E}[v] &= \mu \\
 Var[v] &= \sigma^2
 \end{aligned}$$

Proprietà della gaussiana Data la variabile casuale $w = \alpha + \beta v$ con $\alpha, \beta \in \mathbb{R}$.

- Linearità:

$$\begin{aligned}
 \mathbb{E}[w] &= \alpha + \beta \cdot \mathbb{E}[v] \\
 Var[w] &= \beta^2 \cdot Var[v]
 \end{aligned}$$

- Linearità in distribuzione:

$$w \sim N(\alpha + \beta \cdot \mathbb{E}[v], \beta^2 \cdot Var[v])$$

Definizione Queste definizioni si possono generalizzare per un vettore di n variabili casuali $\underline{v} = [v_1 \ v_2 \ \dots \ v_n]^T$:

- Distribuzione di probabilità:

$$F(q_1, q_2, \dots, q_n) = P(v_1 \leq q_1, v_2 \leq q_2, \dots, v_n \leq q_n)$$

- Densità di probabilità:

$$f(q_1, q_2, \dots, q_n) = \frac{\partial^n F(q_1, q_2, \dots, q_n)}{\partial q_1 \partial q_2 \dots \partial q_n}$$

- Il valore atteso è un vettore colonna con n elementi definiti come:

$$\mathbb{E}[\underline{v}] = [\mathbb{E}[v_1], \mathbb{E}[v_2], \dots, \mathbb{E}[v_n]]^T$$

- La varianza è una matrice $n \times n$, semidefinita positiva, definita come:

$$Var[\underline{v}] = \int_{\mathbb{R}^n} (q - \mathbb{E}[\underline{v}]) (q - \mathbb{E}[\underline{v}])^T f(q) dq$$

1. Richiami di statistica e modelli stocastici

Definizione Date v_1 e v_2 si definisce il coefficiente di correlazione ρ :

$$\rho = \frac{\mathbb{E}[(v_1 - \mathbb{E}[v_1]) \cdot (v_2 - \mathbb{E}[v_2])]}{\sigma[v_1] \cdot \sigma[v_2]}$$

Questo coefficiente indica quanto le due variabili dipendono l'una dall'altra, infatti:

- Se $v_2 = \alpha \cdot v_1 + \beta$ con $\alpha, \beta \in \mathbb{R}$. allora $\rho = 1$, perchè tra le due variabili c'è una dipendenza lineare perfetta.
- Se $\rho = 0$ allora le variabili si dicono scorrelate e si ha che:

$$\mathbb{E}[(v_1 - \mathbb{E}[v_1]) \cdot (v_2 - \mathbb{E}[v_2])] = 0$$

Definizione Date v_1 e v_2 si definisce la covarianza come:

$$Cov(v_1, v_2) = \mathbb{E}[(v_1 - \mathbb{E}[v_1]) \cdot (v_2 - \mathbb{E}[v_2])]$$

e quindi:

$$\rho = \frac{Cov(v_1, v_2)}{\sigma[v_1] \cdot \sigma[v_2]}$$

se le variabili sono scorrelate allora:

$$Cov(v_1, v_2) = 0$$

Definizione Due variabili casuali v_1 e v_2 , con funzione di densità di probabilità congiunta f , si dicono indipendenti se e solo se:

$$f(q_1, q_2) = f_1(q_1) \cdot f_2(q_2)$$

Teorema Siano v_1 e v_2 due variabili indipendenti, allora v_1 e v_2 sono scorrelate.

Dimostrazione Per dimostrare che le due variabili sono scorrelate bisogna dimostrare che la covarianza sia nulla e quindi:

$$\begin{aligned} Cov(v_1, v_2) &= \mathbb{E}[(v_1 - \mathbb{E}[v_1]) \cdot (v_2 - \mathbb{E}[v_2])] = \\ &= \mathbb{E}[v_1 \cdot v_2 - v_1 \cdot \mathbb{E}[v_2] - v_2 \cdot \mathbb{E}[v_1] + \mathbb{E}[v_1] \cdot \mathbb{E}[v_2]] = \\ &= \mathbb{E}[v_1 \cdot v_2] - \mathbb{E}[v_1] \cdot \mathbb{E}[\mathbb{E}[v_2]] - \mathbb{E}[v_2] \cdot \mathbb{E}[\mathbb{E}[v_1]] + \mathbb{E}[\mathbb{E}[v_1]] \cdot \mathbb{E}[\mathbb{E}[v_2]] = \\ &= \mathbb{E}[v_1 \cdot v_2] - \mathbb{E}[v_1] \cdot \mathbb{E}[\mathbb{E}[v_2]] - \mathbb{E}[v_2] \cdot \mathbb{E}[\mathbb{E}[v_1]] + \mathbb{E}[\mathbb{E}[v_1]] \cdot \mathbb{E}[\mathbb{E}[v_2]] = \\ &= \mathbb{E}[v_1 \cdot v_2] - \mathbb{E}[v_1] \cdot \mathbb{E}[v_2] - \cancel{\mathbb{E}[v_1] \cdot \mathbb{E}[v_2]} + \cancel{\mathbb{E}[v_1] \cdot \mathbb{E}[v_2]} = \\ &= \mathbb{E}[v_1 \cdot v_2] - \mathbb{E}[v_1] \cdot \mathbb{E}[v_2] \end{aligned}$$

Ricaviamo il valore atteso congiunto:

$$\begin{aligned} \mathbb{E}[v_1 \cdot v_2] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q_1 \cdot q_2 \cdot f(q_1, q_2) dq_1 dq_2 = \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} q_1 \cdot q_2 \cdot f_1(q_1) \cdot f_2(q_2) dq_1 dq_2 = \\ &= \int_{-\infty}^{\infty} q_1 \cdot f_1(q_1) dq_1 \cdot \int_{-\infty}^{\infty} q_2 \cdot f_2(q_2) dq_2 = \\ &= \mathbb{E}[v_1] \cdot \mathbb{E}[v_2] \end{aligned}$$

1. Richiami di statistica e modelli stocastici

e quindi:

$$\begin{aligned} \text{Cov}(v_1, v_2) &= \mathbb{E}[v_1 \cdot v_2] - \mathbb{E}[v_1] \cdot \mathbb{E}[v_2] = \\ &= \mathbb{E}[v_1] \cdot \mathbb{E}[v_2] - \mathbb{E}[v_1] \cdot \mathbb{E}[v_2] = \\ &= 0 \end{aligned}$$

Nota Se $v_1 = v_2$ e $v_1 \sim N(0, 1)$ allora $\mathbb{E}[v_1^2] = 1$, ma $\mathbb{E}[v_1] \cdot \mathbb{E}[v_2] = 0$.

2. Processi stocastici

Definizione Un processo stocastico a tempo discreto è una successione di variabili casuali definite tutte a partire dallo stesso esperimento casuale S e ordinate secondo un indice t . Viene indicato con $v(S, t)$.

Osservazioni

- Un processo stocastico non è deterministico, ma bensì casuale, e quindi si possono avere infinite realizzazioni dello stesso processo. Ad esempio la sequenza di misure $y(0), y(1), \dots, y(n)$ è una possibile realizzazione finita (dato che potrebbe andare avanti) del processo presa da un determinato esperimento s_1 ; se si eseguisse un secondo esperimento s_2 si otterrebbe una sequenza di misure diversa dalle precedenti.
- Facendo un'analogia con i termini deterministici si può dire che il processo stocastico corrisponde a un segnale, anche se bisogna tener presente che un processo stocastico può generare infinite realizzazioni (come spiegato nel punto precedente).
- Spesso si può omettere l'esperimento da cui proviene il processo stocastico indicando $v(0), v(1), \dots, v(n)$ al posto di $v(S, 0), v(S, 1), \dots, v(S, n)$

Definizione Dato un processo stocastico $v(S, t)$ si definiscono:

- Valor medio:

$$m(t) = \mathbb{E}[v(S, t)]$$

- Covarianza:

$$\gamma(t_1, t_2) = \mathbb{E}[(v(t_1) - m(t_1)) \cdot (v(t_2) - m(t_2))]$$

La varianza al tempo t si ottiene da:

$$\begin{aligned}\gamma(t, t) &= \mathbb{E}[(v(t) - m(t)) \cdot (v(t) - m(t))] = \\ &= \mathbb{E}[(v(t) - m(t))^2]\end{aligned}$$

3. Processi stocastici stazionari

Definizione Un processo stocastico $v(S, t)$ si dice stazionario in senso forte se e solo se $\forall n \in \mathbb{N}$, scelti t_1, t_2, \dots, t_n , il comportamento della n-upla $v(t_1 + \tau), v(t_2 + \tau), \dots, v(t_n + \tau)$ è lo stesso di quello della n-upla $v(t_1), v(t_2), \dots, v(t_n)$ $\forall \tau \in \mathbb{Z}$. Questo significa che prendendo un istante iniziale t_1 il processo rimane inalterato

Definizione Un processo stocastico $v(S, t)$ si dice stazionario in senso debole se e solo se:

- $m(t) = m = \text{costante}, \forall t$
- $\gamma(t_1, t_2) = \gamma(t_3, t_4)$ se $|t_2 - t_1| = |t_4 - t_3|, \forall t_1, t_2, t_3, t_4$
Quindi la covarianza tra due variabili a tempi diversi dipende solamente dalla distanza tra i due tempi $\tau = |t_2 - t_1| = |t_4 - t_3|$.

Dato che la covarianza dipende solo dalla differenza si può usare la funzione:

$$\gamma(\tau) = \mathbb{E}[(v(t) - m) \cdot (v(t + \tau) - m)]$$

che è costante $\forall t$.

Proprietà di $\gamma(\tau)$

1. $\gamma(0) = \mathbb{E}[(v(t) - m)^2] \geq 0$
2. $|\gamma(\tau)| \leq \gamma(0), \forall \tau$ quindi la funzione è limitata
3. $\gamma(\tau) = \gamma(-\tau)$ quindi la funzione γ è pari

Queste proprietà sono rappresentate nella figura 3.1.

3. Processi stocastici stazionari

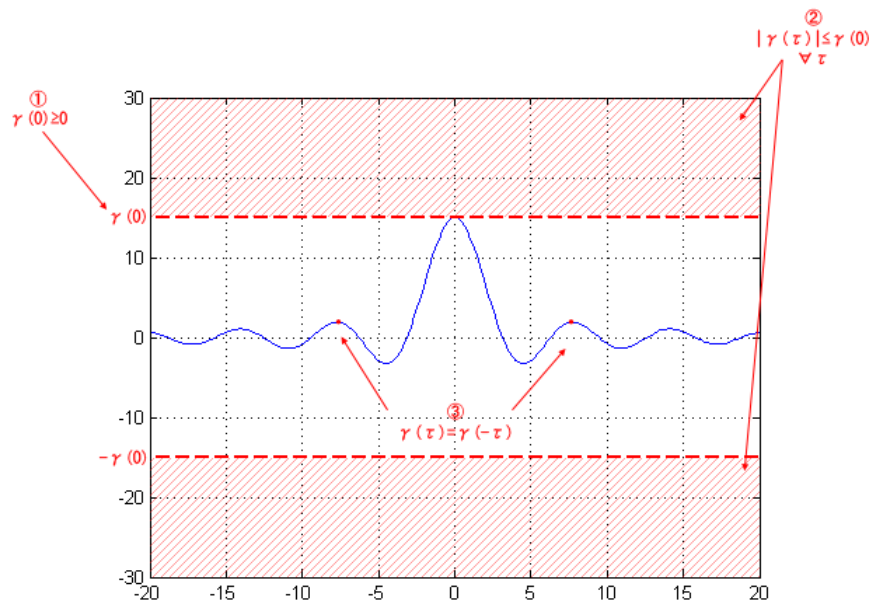


Figura 3.1.: Rappresentazione di una funzione $\gamma(\tau)$ d'esempio

Definizione Dati due processi stocastici $v_1(t)$ e $v_2(t)$ si dicono equivalenti se e solo se hanno lo stesso valore atteso m e la stessa covarianza γ .

Osservazioni

- L'equivalenza tra due processi stocastici non corrisponde all'uguaglianza stretta, ma si può comunque usare la notazione $=$ anche per esprimere solo l'equivalenza.

Definizione Un processo stocastico stazionario $e(t)$ è detto rumore bianco (WN , White Noise) e lo si indica con la notazione $WN(\mu, \lambda^2)$ se:

- $\mathbb{E}[e(t)] = \mu$
- $\gamma(0) = \mathbb{E}[(v(t) - \mu)^2] = \lambda^2, \forall t$
- $\gamma(\tau) = \mathbb{E}[(e(t) - \mu) \cdot (e(t + \tau) - \mu)] = 0 \forall t, \forall \tau \neq 0$

e si indica con $e(t) \sim WN(\mu, \lambda^2)$

Osservazioni

- Il rumore bianco è incorrelato in diversi istanti, e quindi è imprevedibile
- La distribuzione delle variabili casuali non è determinante e possono essere gaussiane, uniformi... in particolare si indica con WNG un rumore bianco gaussiano.

3.1. Rappresentazione spettrale

Definizione Si definisce densità spettrale di potenza di un processo stocastico stazionario $v(t)$ la trasformata discreta di Fourier della funzione $\gamma(\tau)$:

$$\begin{aligned}\Gamma(\omega) &= \mathcal{F}[\gamma(\tau)] = \\ &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) \cdot e^{-j \cdot \omega \cdot \tau}\end{aligned}$$

Nota La funzione $\Gamma(\omega)$ esiste solo per i processi stocastici stazionari in cui $\gamma(\tau) \rightarrow 0$ per $\tau \rightarrow \infty$, ma nei casi che nei casi che si studierà questa ipotesi sarà sempre verificata.

Proprietà di $\Gamma(\omega)$

- $\Gamma(\omega)$ è una funzione reale della variabile reale ω

$$\text{Im}[\Gamma(\omega)] = 0, \forall \omega$$

- $\Gamma(\omega)$ è una funzione positiva

$$\Gamma(\omega) \geq 0, \forall \omega$$

- $\Gamma(\omega)$ è una funzione pari

$$\Gamma(\omega) = \Gamma(-\omega), \forall \omega$$

- $\Gamma(\omega)$ è una funzione periodica di periodo 2π

$$\Gamma(\omega) = \Gamma(\omega + 2 \cdot k \cdot \pi), \forall k \in \mathbb{Z}, \forall \omega$$

Queste proprietà sono rappresentate nell'immagine 3.2.

Osservazione Dato che la funzione è periodica solitamente la si rappresenta solamente nell'intervallo $[-\pi, +\pi]$, perchè rappresentare altre parti non aggiunge informazioni

3. Processi stocastici stazionari

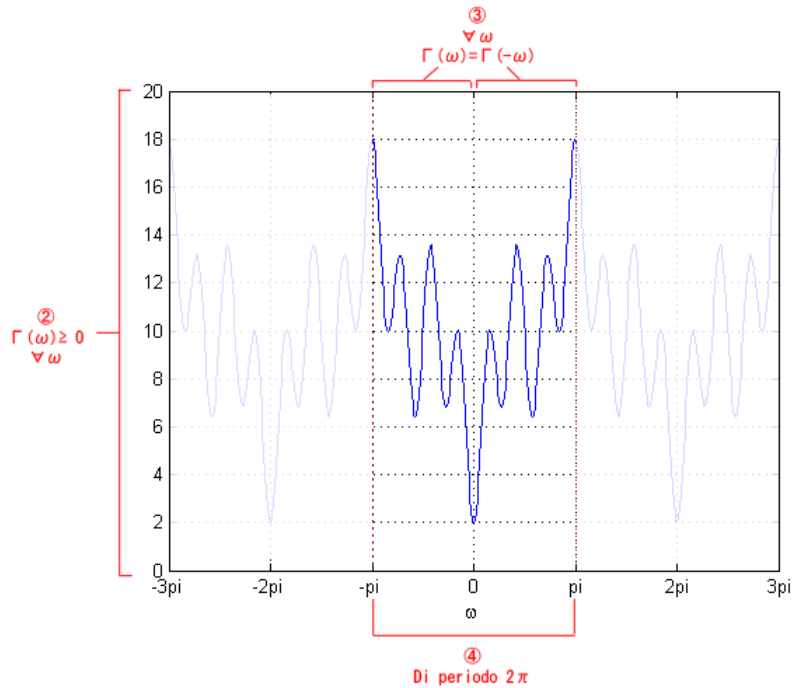


Figura 3.2.: Rappresentazione di una funzione $\Gamma(\omega)$ d'esempio

Nota Si può ricavare $\gamma(\tau)$ da $\Gamma(\omega)$ dalla antitrasformata di Fourier:

$$\begin{aligned}\gamma(\tau) &= \mathcal{F}^{-1}[\Gamma(\omega)] = \\ &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \Gamma(\omega) \cdot e^{j \cdot \omega \cdot \tau} d\omega\end{aligned}$$

da cui si nota che:

$$\begin{aligned}\gamma(0) &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \Gamma(\omega) \cdot e^{j \cdot \omega \cdot 0} d\omega = \\ &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \Gamma(\omega) \cdot d\omega\end{aligned}$$

ossia la varianza del processo stocastico stazionario corrisponde all'area sottesa dal grafico della densità spettrale di potenza, a meno di un prodotto.

Densità spettrale di potenza del rumore bianco

Considerando $e(t) \sim WN(\mu, \lambda^2)$ si ha che:

$$\begin{aligned}\Gamma(\omega) &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) \cdot e^{-j \cdot \omega \cdot \tau} \\ &= \underbrace{\dots}_0 + \underbrace{\gamma(-1) e^{j \cdot \omega}}_0 + \underbrace{\gamma(0)}_{\lambda^2} + \underbrace{\gamma(1) e^{-j \cdot \omega}}_0 + \underbrace{\dots}_0 = \\ &= \lambda^2\end{aligned}$$

Quindi lo spettro, $\Gamma(\omega)$, è una funzione costante, questo comportamento, simile a quello della luce bianca, spiega il motivo del nome del rumore bianco.

3.2. Rappresentazione dinamica

Il rumore bianco $e(t)$ ha uno spettro costante (si può dire che contiene tutte le componenti di frequenza) e quindi tutti i processi stocastici stazionari si possono rappresentare come:

$$\begin{aligned} v(t) &= w_0 \cdot e(t) + w_1 \cdot e(t-1) + \dots = \\ &= \sum_{i=-\infty}^{+\infty} w_i \cdot e(t-i) \end{aligned}$$

Questa espressione è simile alla risposta all'impulso di un sistema a tempo discreto con una funzione di trasferimento $H(z)$ con i coefficienti w_i , e quindi si ha che:

$$v(t) = H(z) \cdot e(t)$$

Questa rappresentazione di $v(t)$ è definita rappresentazione dinamica in quanto sfrutta un sistema dinamico $H(z)$ e permette di modellare l'incertezza nel modo illustrato in figura 3.3.

Questo permette di studiare l'errore tramite un sistema dinamico, modellizzando la parte stocastica con il rumore bianco in entrata.

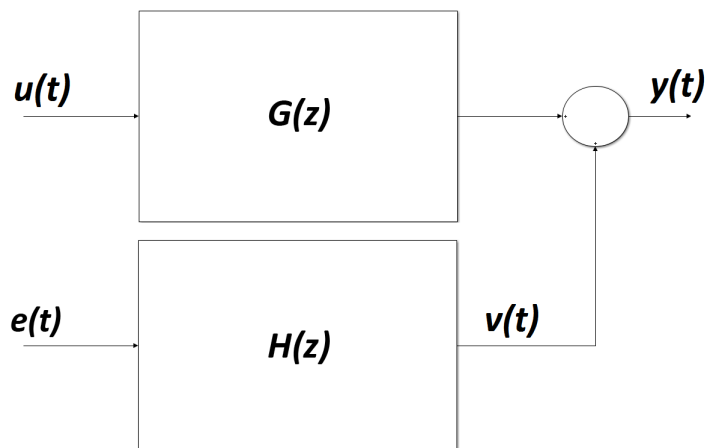


Figura 3.3.: Rappresentazione del sistema con l'incertezza mostrata tramite la rappresentazione dinamica

Bisogna però tenere in considerazione che questo modo di rappresentare è solo un modello e che in realtà non c'è nessun rumore bianco $e(t)$ e nessun sistema $H(z)$. Inoltre, nelle applicazioni trattate, l'incertezza $v(t)$ è sempre stazionaria e quindi il sistema $H(z)$ deve essere stabile in modo che la media del segnale in uscita si sposti di un valore finito rispetto a quella del rumore bianco facendo rimanere $v(t)$ stazionario.

In altre parole, se $H(z)$ fosse un sistema instabile l'uscita $v(t)$ tenderebbe a divergere per ogni variazione dell'ingresso $e(t)$ e quindi non avrebbe media costante, perdendo quindi la proprietà di stazionarietà. Questo significa che se dopo aver stimato i parametri di $H(z)$ si nota che il sistema è instabile, è stato commesso un errore durante la modellizzazione del sistema.

3.3. Depolarizzazione

La depolarizzazione è una tecnica che permette di semplificare il calcolo del parametro $\gamma(\tau)$ di un processo stocastico stazionario $v(t)$ a media non nulla m_v .

Ricordando che

$$\gamma(\tau) = \mathbb{E}[(v(t) - m_v) \cdot (v(t + \tau) - m_v)]$$

si nota che se la media fosse nulla la formula si ridurrebbe a:

$$\gamma(\tau) = \mathbb{E}[v(t) \cdot v(t + \tau)]$$

che è molto più gestibile.

Quindi l'idea è quella di utilizzare un processo ausiliario chiamato $\tilde{v}(t)$ definito come:

$$\tilde{v}(t) = v(t) - m_v$$

di cui si nota che:

$$\begin{aligned} \mathbb{E}[\tilde{v}(t)] &= \mathbb{E}[v(t) - m_v] = \\ &= \mathbb{E}[v(t)] - m_v = \\ &= 0 \end{aligned}$$

e quindi:

$$\begin{aligned} \gamma_{\tilde{v}}(\tau) &= \mathbb{E}[(\tilde{v}(t) - m_{\tilde{v}}) \cdot (\tilde{v}(t + \tau) - m_{\tilde{v}})] = \\ &= \mathbb{E}[\tilde{v}(t) \cdot \tilde{v}(t + \tau)] = \\ &= \mathbb{E}[(v(t) - m_v) \cdot (v(t + \tau) - m_v)] = \\ &= \gamma_v(\tau) \end{aligned}$$

e quindi i processi v e \tilde{v} hanno la stessa funzione di covarianza.

Esempio Capire se il seguente processo è stazionario e calcolarne media m_v e covarianza $\gamma_v(\tau)$.

$$v(t) = 5 \cdot e(t - 1) + e(t), \quad e(t) \sim WN(1, 1)$$

per verificare che sia stazionario bisogna riscriverlo in forma dinamica introducendo l'operatore di ritardo, e quindi:

$$\begin{aligned} v(t) &= e(t) + 5 \cdot e(t - 1) = \\ &= e(t) + 5 \cdot z^{-1} \cdot e(t) = \\ &= e(t) \cdot (1 + 5 \cdot z^{-1}) = \\ &= e(t) \cdot \left(\frac{z + 5}{z} \right) \end{aligned}$$

quindi il sistema ha un solo polo nell'origine, il che lo rende stabile, e di conseguenza il processo risulta stazionario.

Per calcolare la media si usa la definizione:

$$\begin{aligned} m_v &= \mathbb{E}[v(t)] = \\ &= \mathbb{E}[e(t) + 5 \cdot e(t - 1)] = \\ &= \mathbb{E}[e(t)] + 5 \cdot \mathbb{E}[e(t - 1)] = \\ &= 1 + 5 \cdot 1 = \\ &= 6 \end{aligned}$$

3. Processi stocastici stazionari

Dato che la media non è nulla conviene depolarizzare per semplificare i conti della covarianza e quindi si introduce:

$$\begin{cases} \tilde{v}(t) = v(t) - 6 \\ \tilde{e}(t) = e(t) - 1 \end{cases}$$

Riscrivendo il processo si ottiene:

$$\begin{aligned} \tilde{v}(t) + 6 &= 5 \cdot (\tilde{e}(t-1) + 1) + \tilde{e}(t) + 1 \\ \tilde{v}(t) &= 5 \cdot \tilde{e}(t-1) + \tilde{e}(t) + 5 + 1 - 6 \\ \tilde{v}(t) &= 5 \cdot \tilde{e}(t-1) + \tilde{e}(t) \end{aligned}$$

e ricordando che:

$$\gamma(\tau) = \gamma_{\tilde{v}}(\tau)$$

si può applicare la definizione:

$$\begin{aligned} \gamma(0) &= \gamma_{\tilde{v}}(0) = \\ &= \mathbb{E} [\tilde{v}(t)^2] = \\ &= \mathbb{E} [(5 \cdot \tilde{e}(t-1) + \tilde{e}(t))^2] = \\ &= \mathbb{E} [25 \cdot \tilde{e}(t-1)^2 + \tilde{e}(t)^2 + 10 \cdot \tilde{e}(t-1) \cdot \tilde{e}(t)] = \\ &= \underbrace{\mathbb{E} [\tilde{e}(t)^2]}_1 + 25 \cdot \underbrace{\mathbb{E} [\tilde{e}(t-1)^2]}_1 + 10 \cdot \underbrace{\mathbb{E} [\tilde{e}(t) \cdot \tilde{e}(t-1)]}_0 \\ &= 26 \end{aligned}$$

$$\begin{aligned} \gamma(1) = \gamma(-1) &= \gamma_{\tilde{v}}(1) = \\ &= \mathbb{E} [\tilde{v}(t) \cdot \tilde{v}(t-1)] = \\ &= \mathbb{E} [(\tilde{e}(t) + 5 \cdot \tilde{e}(t-1)) \cdot (\tilde{e}(t-1) + 5 \cdot \tilde{e}(t-2))] = \\ &= \mathbb{E} [\tilde{e}(t) \cdot \tilde{e}(t-1) + 5 \cdot \tilde{e}(t-1) \cdot \tilde{e}(t-1) + 5 \cdot \tilde{e}(t) \cdot \tilde{e}(t-2) + 25 \cdot \tilde{e}(t-1) \cdot \tilde{e}(t-2)] = \\ &= \underbrace{\mathbb{E} [\tilde{e}(t) \cdot \tilde{e}(t-1)]}_0 + 5 \cdot \underbrace{\mathbb{E} [\tilde{e}(t-1)^2]}_1 + 5 \cdot \underbrace{\mathbb{E} [\tilde{e}(t) \cdot \tilde{e}(t-2)]}_0 + 25 \cdot \underbrace{\mathbb{E} [\tilde{e}(t-1) \cdot \tilde{e}(t-2)]}_0 \\ &= 5 \end{aligned}$$

$$\begin{aligned} \gamma(2) = \gamma(-2) &= \gamma_{\tilde{v}}(2) = \\ &= \mathbb{E} [\tilde{v}(t) \cdot \tilde{v}(t-2)] = \\ &= \mathbb{E} [(\tilde{e}(t) + 5 \cdot \tilde{e}(t-1)) \cdot (\tilde{e}(t-2) + 5 \cdot \tilde{e}(t-3))] = \\ &= \mathbb{E} [\tilde{e}(t) \cdot \tilde{e}(t-2) + 5 \cdot \tilde{e}(t-1) \cdot \tilde{e}(t-2) + 5 \cdot \tilde{e}(t) \cdot \tilde{e}(t-3) + 25 \cdot \tilde{e}(t-1) \cdot \tilde{e}(t-3)] = \\ &= \underbrace{\mathbb{E} [\tilde{e}(t) \cdot \tilde{e}(t-2)]}_0 + 5 \cdot \underbrace{\mathbb{E} [\tilde{e}(t-1) \cdot \tilde{e}(t-2)]}_0 + 5 \cdot \underbrace{\mathbb{E} [\tilde{e}(t) \cdot \tilde{e}(t-3)]}_0 + 25 \cdot \underbrace{\mathbb{E} [\tilde{e}(t-1) \cdot \tilde{e}(t-3)]}_0 \\ &= 0 \end{aligned}$$

andando avanti saranno tutti nulli, e quindi si ha:

$$\gamma(\tau) = \begin{cases} 26 & \text{per } \tau = 0 \\ 5 & \text{per } \tau = \pm 1 \\ 0 & \text{per } |\tau| > 1 \end{cases}$$

4. Tipi di processi

4.1. Processi MA (Moving Average)

I processi MA sono caratterizzati da:

- $n \in \mathbb{N}$ che è l'ordine del processo
- $n + 1$ costanti c_0, c_1, \dots, c_n che sono i coefficienti

si indicano con $MA(n)$ e si definiscono come:

$$\begin{aligned}v(t) &= c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_n \cdot e(t-n) = \\&= \sum_{i=0}^n c_i \cdot e(t-i) \\ \text{con } e(t) &\sim WN(\mu, \lambda^2)\end{aligned}$$

In questo tipo di processi si ha che il valore al tempo t dipende da tutti i valori assunti dal rumore bianco $e(t)$ negli n istanti precedenti.

Ricordando l'operatore di ritardo z^{-1} :

$$z^{-1}x(t) = x(t-1)$$

si può riscrivere il processo come:

$$\begin{aligned}v(t) &= c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_n \cdot e(t-n) = \\&= c_0 \cdot e(t) + c_1 \cdot e(t) \cdot z^{-1} + \dots + c_n \cdot e(t) z^{-n} = \\&= (c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}) \cdot e(t) = \\&= \frac{c_0 \cdot z^n + c_1 \cdot z^{n-1} + \dots + c_n}{z^n} \cdot e(t)\end{aligned}$$

da cui si ricava che:

$$\begin{aligned}H(z) &= \frac{c_0 \cdot z^n + c_1 \cdot z^{n-1} + \dots + c_n}{z^n} = \\&= c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n} = \\&= C(z)\end{aligned}$$

Da cui si nota che tutti i processi di tipo $MA(n)$ hanno sempre n poli nell'origine e n zeri il cui valore dipende dai coefficienti c_0, c_1, \dots, c_n , per questo motivo il sistema $H(z)$ è sempre stabile e quindi i processi $MA(n)$ sono sempre stazionari.

Calcolo dei parametri caratteristici

Si ha che:

- Valore atteso

$$\begin{aligned}
 m(t) &= \mathbb{E}[v(t)] = \\
 &= \mathbb{E}[c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_n \cdot e(t-n)] = \\
 &= c_0 \cdot \mathbb{E}[e(t)] + c_1 \cdot \mathbb{E}[e(t-1)] + \dots + c_n \cdot \mathbb{E}[e(t-n)] = \\
 &= c_0 \cdot \mu + c_1 \cdot \mu + \dots + c_n \cdot \mu = \\
 &= (c_0 + c_1 + \dots + c_n) \cdot \mu = \\
 &= \mu \cdot \sum_{i=0}^n c_i
 \end{aligned}$$

da cui si nota che se il rumore bianco ha valore atteso nullo allora anche $v(t)$ ha valore atteso nullo.

- Funzione γ , per la depolarizzazione si può supporre che $\mu = 0$.

$$\begin{aligned}
 \gamma(0) &= \mathbb{E}[(v(t) - m(t))^2] = \\
 &= \mathbb{E}[v(t)^2] = \\
 &= \mathbb{E}[(c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_n \cdot e(t-n))^2] = \\
 &= \mathbb{E} \left[\underbrace{c_0^2 \cdot e(t)^2 + c_1^2 \cdot e(t-1)^2 + \dots + c_n^2 \cdot e(t-n)^2}_{\text{quadrati}} + \right. \\
 &\quad \left. + \underbrace{2 \cdot c_0 \cdot c_1 \cdot e(t) \cdot e(t-1) + \dots + 2 \cdot c_{n-1} \cdot c_n \cdot e(t-n+1) \cdot e(t-n)}_{\text{doppi prodotti}} \right] = \\
 &= c_0^2 \cdot \underbrace{\mathbb{E}[e(t)^2]}_{\lambda^2} + c_1^2 \cdot \underbrace{\mathbb{E}[e(t-1)^2]}_{\lambda^2} + \dots + c_n^2 \cdot \underbrace{\mathbb{E}[e(t-n)^2]}_{\lambda^2} + 2 \cdot c_0 \cdot c_1 \cdot \underbrace{\mathbb{E}[e(t) \cdot e(t-1)]}_0 + \dots + \\
 &+ 2 \cdot c_{n-1} \cdot c_n \cdot \underbrace{\mathbb{E}[e(t-n+1) \cdot e(t-n)]}_0 = \\
 &= (c_0^2 + c_1^2 + \dots + c_n^2) \cdot \lambda^2 = \\
 &= \lambda^2 \cdot \sum_{i=0}^n c_i^2
 \end{aligned}$$

$$\begin{aligned}
 \gamma(1) &= \mathbb{E}[(v(t) - m(t)) \cdot (v(t-1) - m(t-1))] = \\
 &= \mathbb{E}[v(t) \cdot v(t-1)] = \\
 &= \mathbb{E}[(c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_n \cdot e(t-n)) \cdot (c_0 \cdot e(t-1) + c_1 \cdot e(t-2) + \dots + c_n \cdot e(t-n-1))]
 \end{aligned}$$

Svolgendo il prodotto ci si accorge che i termini non allo stesso istante valgono 0 mentre quelli allo stesso istante valgono λ^2 per questo motivo:

$$\gamma(1) = \lambda^2 \cdot (c_0 \cdot c_1 + c_1 \cdot c_2 + \dots + c_{n-1} \cdot c_n)$$

4. Tipi di processi

per lo stesso motivo ci si accorge che:

$$\begin{aligned}\gamma(2) &= \lambda^2 \cdot (c_0 \cdot c_2 + c_1 \cdot c_3 + \dots + c_{n-2} \cdot c_n) \\ &\vdots \\ \gamma(n) &= \lambda^2 \cdot c_0 \cdot c_n\end{aligned}$$

e per tutti i termini successivi nessun termine del prodotto è negli stessi istanti e quindi:

$$\gamma(k) = 0, \forall k > n$$

Per questo motivo un processo $MA(n)$ dipende solo dai valori che distano n istanti di tempo e si può dire che dopo un po' diventa imprevedibile come un rumore bianco si dice quindi che il processo $MA(n)$ è parzialmente colorato

4.2. Processi AR (AutoRegressive)

I processi AR sono caratterizzati da:

- $m \in \mathbb{N}$ che è l'ordine del processo
- m costanti a_1, a_2, \dots, a_m che sono i coefficienti del processo AR
- una sola costante c_0 che sarà il coefficiente di $e(t)$

si indicano con $AR(m)$ e si definiscono come:

$$\begin{aligned}v(t) &= a_1 \cdot v(t-1) + a_2 \cdot v(t-2) + \dots + a_m \cdot v(t-m) + c_0 \cdot e(t) = \\ &= c_0 \cdot e(t) + \sum_{i=1}^m a_i \cdot v(t-i) \\ \text{con} \quad e(t) &\sim WN(0, \lambda^2)\end{aligned}$$

In questo tipo di processi il valore al tempo t è dato da una regressione di m valori passati del processo stesso (definita quindi una autoregressione). Come nella sezione 4.1, possiamo esprimerlo anche usando l'operatore ritardo z^{-1}

$$\begin{aligned}v(t) &= c_0 \cdot e(t) + a_1 \cdot v(t-1) + \dots + a_m \cdot v(t-m) \\ v(t) &= c_0 \cdot e(t) + a_1 \cdot v(t) \cdot z^{-1} + \dots + a_m \cdot v(t) \cdot z^{-m} \\ v(t) \cdot (1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}) &= c_0 \cdot e(t) \\ v(t) &= \frac{c_0}{1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}} \cdot e(t) \\ v(t) &= \frac{c_0 \cdot z^m}{(1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}) \cdot z^m} \cdot e(t) \\ v(t) &= \frac{c_0 \cdot z^m}{z^m - a_1 \cdot z^{m-1} - a_2 \cdot z^{m-2} - \dots - a_m} \cdot e(t)\end{aligned}$$

da cui si ricava che:

$$\begin{aligned}H(z) &= \frac{c_0 \cdot z^m}{z^m - a_1 \cdot z^{m-1} - a_2 \cdot z^{m-2} - \dots - a_m} = \\ &= \frac{c_0}{1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}} \\ &= \frac{c_0}{A(z)}\end{aligned}$$

4. Tipi di processi

dove $A(z) = 1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}$.

Da cui si nota che a differenza dei processi $MA(n)$, i processi $AR(m)$ hanno m zeri nell'origine e m poli. Questo comporta che non è sempre stabile: lo è solo quando tutti i poli sono all'interno del cerchio di raggio unitario. Per questo motivo bisogna prestare molta attenzione nello stimare i parametri del processo (ovvero i vari coefficienti a_1, a_2, \dots, a_m) in quanto una scelta sbagliata potrebbe rendere instabile il sistema.

Calcolo dei parametri caratteristici

Ha senso ricavarli solo nel caso in cui il processo $AR(m)$ è stazionario, quindi, si ha che:

- Valore atteso¹

$$\begin{aligned}
 m = \mathbb{E}[v(t)] &= \mathbb{E}[a_1 \cdot v(t-1) + a_2 \cdot v(t-2) + \dots + a_m \cdot v(t-m) + c_0 \cdot e(t)] \\
 \mathbb{E}[v(t)] &= a_1 \cdot \underbrace{\mathbb{E}[v(t-1)]}_{\mathbb{E}[v(t)]} + a_2 \cdot \underbrace{\mathbb{E}[v(t-2)]}_{\mathbb{E}[v(t)]} + \dots + a_m \cdot \underbrace{\mathbb{E}[v(t-m)]}_{\mathbb{E}[v(t)]} + c_0 \cdot \mathbb{E}[e(t)] \\
 \mathbb{E}[v(t)] &= a_1 \cdot \mathbb{E}[v(t)] + a_2 \cdot \mathbb{E}[v(t)] + \dots + a_m \cdot \mathbb{E}[v(t)] + c_0 \cdot \mathbb{E}[e(t)] \\
 \mathbb{E}[v(t)] \cdot (1 - a_1 - a_2 - \dots - a_m) &= c_0 \cdot \mathbb{E}[e(t)] \\
 \mathbb{E}[v(t)] \cdot (1 - a_1 - a_2 - \dots - a_m) &= c_0 \cdot \mu \\
 \mathbb{E}[v(t)] &= \frac{c_0 \cdot \mu}{1 - a_1 - a_2 - \dots - a_m}
 \end{aligned}$$

come nei processi $MA(n)$ se il rumore bianco ha valore atteso nullo $\mu = 0$ il valore atteso di $v(t)$ sarà nullo anch'esso.

- Funzione γ , dato che i passaggi non sono banali ci riduciamo al caso $AR(1)$.

4.2.1. Processi $AR(1)$

Dato il processo $AR(1)$

$$\begin{aligned}
 v(t) &= a \cdot v(t-1) + c_0 \cdot e(t) \\
 \text{con} \quad e(t) &\sim WN(\mu, \lambda^2)
 \end{aligned}$$

Si può ricavare la $H(z)$ introducendo l'operatore di ritardo z^{-1} o usando la formula generale ricavata prima ottenendo:

$$\begin{aligned}
 H(z) &= \frac{c_0 \cdot z^m}{z^m - a_1 \cdot z^{m-1} - a_2 \cdot z^{m-2} - \dots - a_m} = \\
 &= \frac{c_0 \cdot z}{z - a}
 \end{aligned}$$

da cui si nota che la $H(z)$ ha un polo in a . Per garantire la stabilità, e quindi la stazionarietà del processo, è necessario che $|a| < 1$. Se il processo è stazionario si può ricavare la funzione $\gamma(\tau)$. Per farlo si suppone che $\mu = 0$ per la depolarizzazione (ovvero se μ non fosse 0, si applica la depolarizzazione per renderla tale e si procede come segue).

¹ $\mathbb{E}[v(t-1)] = \mathbb{E}[v(t)]$ perchè $v(t)$ è stazionario, quindi ha media costante

4. Tipi di processi

- $\gamma(0)$

$$\begin{aligned}
 \gamma(0) &= \mathbb{E}[(v(t) - m_v) \cdot (v(t-0) - m_v)] \\
 \gamma(0) &= \mathbb{E}[v(t)^2] \\
 \gamma(0) &= \mathbb{E}[(a \cdot v(t-1) + e(t))^2] \\
 \gamma(0) &= \mathbb{E}[a^2 \cdot v(t-1)^2 + e(t)^2 + 2 \cdot a \cdot v(t-1) \cdot e(t)] \\
 \gamma(0) &= \underbrace{a^2 \cdot \mathbb{E}[v(t-1)^2]}_{\gamma(0)} + \underbrace{\mathbb{E}[e(t)^2]}_{\lambda^2} + 2 \cdot a \cdot \underbrace{\mathbb{E}[v(t-1) \cdot e(t)]}_0 \\
 \gamma(0) &= a^2 \cdot \gamma(0) + \lambda^2 \\
 \gamma(0) \cdot (1 - a^2) &= \lambda^2 \\
 \gamma(0) &= \frac{\lambda^2}{1 - a^2}
 \end{aligned}$$

Il termine $\mathbb{E}[v(t-1) \cdot e(t)]$ si può dire che uguale 0 per una dimostrazione formale mostrata nel capitolo 4.5, ma più semplicemente si può dire che il processo stocastico v al tempo $t-1$ non può aver delle relazioni con l'errore in un istante futuro visto che il rumore bianco e è incorrelato con il passato.

- $\gamma(1)$

$$\begin{aligned}
 \gamma(1) &= \mathbb{E}[(v(t) - m(t)) \cdot (v(t-1) - m(t))] = \\
 &= \mathbb{E}[v(t) \cdot v(t-1)] = \\
 &= \mathbb{E}[(a \cdot v(t-1) + e(t)) \cdot v(t-1)] \\
 &= \mathbb{E}[a \cdot v(t-1)^2 + v(t-1) \cdot e(t)] \\
 &= a \cdot \underbrace{\mathbb{E}[v(t-1)^2]}_{\gamma(0)} + \underbrace{\mathbb{E}[v(t-1) \cdot e(t)]}_0 \\
 &= a \cdot \gamma(0)
 \end{aligned}$$

- $\gamma(2)$

$$\begin{aligned}
 \gamma(2) &= \mathbb{E}[(v(t) - m(t)) \cdot (v(t-2) - m(t))] = \\
 &= \mathbb{E}[v(t) \cdot v(t-2)] = \\
 &= \mathbb{E}[(a \cdot v(t-1) + e(t)) \cdot v(t-2)] \\
 &= \mathbb{E}[a \cdot v(t-1) \cdot v(t-2) + v(t-2) \cdot e(t)] \\
 &= a \cdot \underbrace{\mathbb{E}[v(t-1) \cdot v(t-2)]}_{\gamma(1)} + \underbrace{\mathbb{E}[v(t-2) \cdot e(t)]}_0 \\
 &= a \cdot \gamma(1)
 \end{aligned}$$

Infine, generalizzando :

$$\begin{aligned}
 \gamma(0) &= \frac{\lambda^2}{1 - a^2} \\
 \gamma(\tau) &= a \cdot \gamma(\tau - 1) \text{ per } \tau > 0
 \end{aligned}$$

oppure:

$$\gamma(\tau) = a^{|\tau|} \cdot \frac{\lambda^2}{1 - a^2}$$

Queste formule sono dette di Yule-Walker, ed esistono anche per un generico $AR(n)$.

4. Tipi di processi

Osservazioni Dato che $|a| < 1$ (altrimenti non sarebbe stazionario), si ha che $|\gamma(\tau + 1)| < |\gamma(\tau)|$ e dato che $a \neq \pm 1$ $\gamma(0)$ sarà sempre un valore finito. Inoltre si può notare che:

- il processo $\bar{v} = a \cdot v(t - 1) + c_0 \cdot e(t)$ con $0 < a < 1$ ha $\gamma(\tau) > 0, \forall \tau$ e sarà una funzione decrescente che non raggiungerà mai lo 0.
Questo significa che le varie realizzazioni variano lentamente in quanto sono molto correlate fra di loro per intervalli di tempo piccoli, quindi ci si aspetta una realizzazione con componenti a bassa frequenza.
- il processo $\underline{v} = a \cdot v(t - 1) + c_0 \cdot e(t)$ con $-1 < a < 0$ ha una funzione $\gamma(\tau)$ oscillante con segno che varia a ogni τ diminuendo il suo valore assoluto a tutte le iterazioni.
Questo significa che le varie realizzazioni avranno cambi di segno repentini tra un valore e l'altro creando dei segnali con componenti ad alta frequenza.

Questi comportamenti sono rappresentati nella figure 4.1 e 4.2.

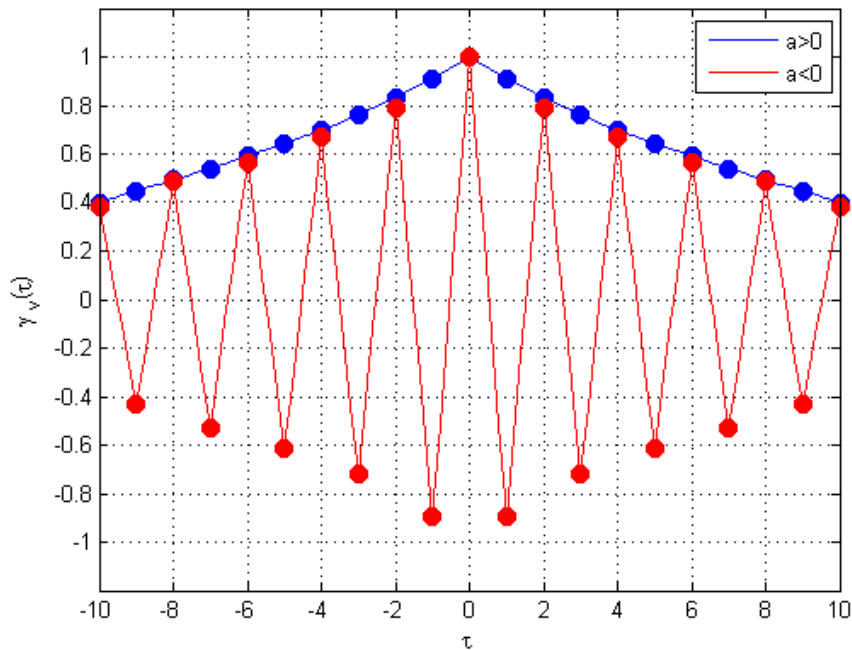


Figura 4.1.: Rappresentazione di una possibile funzione $\gamma_v(\tau)$ per valori positivi e negativi di a

4. Tipi di processi

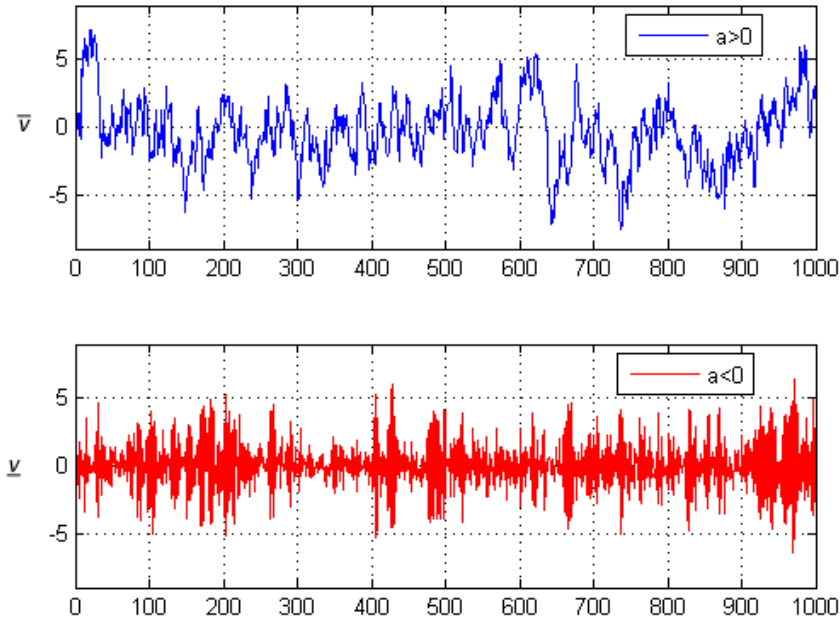


Figura 4.2.: Rappresentazione delle realizzazioni di due processi AR con $a > 0$ (\bar{v}) e $a < 0$ (\underline{v})

4.3. Processi ARMA (AutoRegressive Moving Average)

I processi $ARMA$ sono banalmente l'unione di un processo AR e di un processo MA ; essi sono caratterizzati da:

- $m \in \mathbb{N}$ che è l'ordine del processo AR
- m costanti a_1, a_2, \dots, a_m che sono i coefficienti del processo AR
- $n \in \mathbb{N}$ che è l'ordine del processo MA
- $n + 1$ costanti c_0, c_1, \dots, c_n che sono i coefficienti del processo MA

si indicano con $ARMA(m, n)$ e si definiscono come:

$$\begin{aligned} v(t) = & a_1 \cdot v(t-1) + a_2 \cdot v(t-2) + \dots + a_m \cdot v(t-m) + \\ & + c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_n \cdot e(t-n) + \end{aligned}$$

Si può notare che $ARMA(0, n) = MA(n)$ e che $ARMA(m, 0) = AR(m)$, e quindi i processi visti prima sono dei sottocasi.

Introducendo l'operatore ritardo si può ricavare la $H(z)$:

4. Tipi di processi

$$\begin{aligned}
 v(t) &= a_1 \cdot z^{-1} \cdot v(t) + a_2 \cdot z^{-2} \cdot v(t) + \dots + a_m \cdot z^{-m} \cdot v(t) + \\
 &\quad + c_0 \cdot e(t) + c_1 \cdot z^{-1} \cdot e(t) + \dots + c_n \cdot z^{-n} \cdot e(t) \\
 v(t) \cdot (1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}) &= e(t) \cdot (c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}) \\
 v(t) &= \frac{(c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n})}{(1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m})} \cdot e(t) \\
 v(t) &= \frac{z^m \cdot z^n \cdot (c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n})}{z^m \cdot z^n \cdot (1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m})} \cdot e(t) \\
 v(t) &= z^{m-n} \cdot \frac{z^n \cdot c_0 + c_1 \cdot z^{n-1} + \dots + c_n}{z^m - a_1 \cdot z^{m-1} - a_2 \cdot z^{m-2} - \dots - a_m} \cdot e(t)
 \end{aligned}$$

da cui abbiamo che:

$$\begin{aligned}
 H(z) &= z^{m-n} \cdot \frac{z^n \cdot c_0 + c_1 \cdot z^{n-1} + \dots + c_n}{z^m - a_1 \cdot z^{m-1} - a_2 \cdot z^{m-2} - \dots - a_m} = \\
 &= \frac{c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}}{1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}} = \\
 &= \frac{C(z)}{A(z)}
 \end{aligned}$$

dove $A(z) = 1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}$ è un polinomio di grado m e $C(z) = c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}$ è un polinomio di grado n .

Per fare in modo che il sistema $H(z)$ sia stabile, e quindi che il processo sia stazionario, bisogna che le radici del denominatore siano all'interno del cerchio di raggio unitario; si può notare che questo dipende solo dai coefficienti a_i ossia solo dalla parte autoregressiva.

4.4. Processi ARMAX (AutoRegressive Moving Average with eXogenous inputs)

I processi *ARMAX* sono modelli *ARMA* a cui si aggiunge la presenza di un input esogeno^{II}; sono caratterizzati da:

- $m \in \mathbb{N}$ che è l'ordine del processo *AR*
- m costanti a_1, a_2, \dots, a_m che sono i coefficienti del processo *AR*
- $n \in \mathbb{N}$ che è l'ordine del processo *MA*
- $n + 1$ costanti c_0, c_1, \dots, c_n che sono i coefficienti del processo *MA*
- $p \in \mathbb{N}$ che è l'ordine della variabile esogena
- $p + 1$ costanti b_0, b_1, \dots, b_p che sono i coefficienti della variabile esogena
- una costante $k \in \mathbb{N}$ che indica il ritardo tra ingresso e uscita

si indicano con *ARMAX* ($m, n, k + p$) e si definiscono come:

$$\begin{aligned}
 v(t) &= a_1 \cdot v(t-1) + a_2 \cdot v(t-2) + \dots + a_m \cdot v(t-m) + \\
 &\quad + c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_n \cdot e(t-n) + \\
 &\quad + b_0 \cdot u(t-k) + b_1 \cdot u(t-k-1) + \dots + b_p \cdot u(t-k-p)
 \end{aligned}$$

^{II} una variabile deterministica indipendente dal processo

4. Tipi di processi

Introducendo l'operatore di ritardo si ottiene:

$$\begin{aligned}
 y(t) &= a_1 \cdot z^{-1} \cdot y(t) + a_2 \cdot z^{-2} \cdot y(t) + \dots + a_m \cdot z^{-m} \cdot y(t) + \\
 &\quad + c_0 \cdot e(t) + c_1 \cdot z^{-1} \cdot e(t) + \dots + c_n \cdot z^{-n} \cdot e(t) + \\
 &\quad + b_0 \cdot u(t-k) + b_1 \cdot z^{-1} \cdot u(t-k) + \dots + b_p \cdot z^{-p} \cdot u(t-k) \\
 y(t) \cdot (1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}) &= e(t) \cdot (c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}) + \\
 &\quad + u(t-k) \cdot (b_0 + b_1 \cdot z^{-1} + \dots + b_p \cdot z^{-p}) \\
 y(t) \cdot A(z) &= e(t) \cdot C(z) + u(t-k) \cdot B(z) \\
 y(t) &= \frac{e(t) \cdot C(z) + u(t-k) \cdot B(z)}{A(z)} \\
 y(t) &= \underbrace{\frac{C(z)}{A(z)}}_{H(z)} \cdot e(t) + \underbrace{\frac{B(z)}{A(z)}}_{G(z)} \cdot u(t-k)
 \end{aligned}$$

La rappresentazione dinamica di questo modello è mostrata in figura 4.3

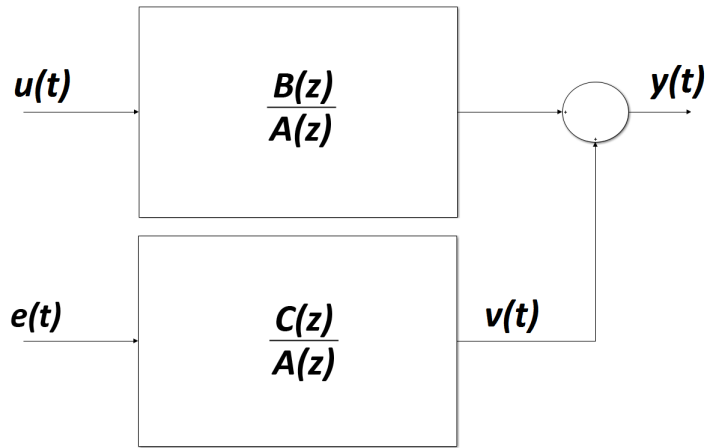


Figura 4.3.: Rappresentazione dinamica di un modello ARMAX

Si può notare che in questo modello viene inglobato anche il sistema reale e l'input $u(t)$. Per questo motivo la variabile stocastica non rappresenta più solo l'incertezza, ma bensì l'uscita del sistema reale e viene quindi definita $y(t)$ e non più $v(t)$.

Dato che l'ingresso $u(t)$ ora influenza il processo, non è possibile dire se $y(t)$ sia stazionario o meno; $y(t)$ sarà stazionario se si verificano le seguenti ipotesi:

- Tutte le radici di $A(z)$ sono all'interno del cerchio di raggio unitario, in modo che $H(z)$ sia stabile
- $u(t)$ è stazionario e quindi costante, dato che non è una variabile casuale

Supporre che l'ingresso deterministico sia costante è un'ipotesi forte che non è quasi mai verificata in quanto questo può assumere qualsiasi forma (rampa, senoide,...). Per questo motivo solitamente il sistema *ARMAX* non è stazionario, ma la sua componente incerta, ossia la parte *ARMA*, deve essere stazionaria.

4. Tipi di processi

Esempio Capire se il seguente processo v è stazionario e calcolarne media m_v e covarianza $\gamma_v(\tau)$.

$$v(t) = \frac{1}{3} \cdot v(t-1) + e(t) + 2 \text{ con } e(t) \sim WN(1, 1)$$

Per capire se è stazionario bisogna scriverlo tramite la rappresentazione dinamica e quindi:

$$\begin{aligned} v(t) &= \frac{1}{3} \cdot v(t-1) + e(t) + 2 \\ v(t) &= \frac{1}{3} \cdot v(t) \cdot z^{-1} + e(t) + 2 \\ v(t) - \frac{1}{3} \cdot v(t) \cdot z^{-1} &= e(t) + 2 \\ v(t) \cdot \left(1 - \frac{1}{3} \cdot z^{-1}\right) &= e(t) + 2 \\ v(t) \cdot \left(\frac{3 \cdot z - 1}{3 \cdot z}\right) &= e(t) + 2 \\ v(t) &= \frac{3 \cdot z}{3 \cdot z - 1} \cdot e(t) + \frac{3 \cdot z}{3 \cdot z - 1} \cdot 2 \end{aligned}$$

Quindi si ha una variabile esogena $u(t) = 2$ e

$$H(z) = \frac{3 \cdot z}{3 \cdot z - 1}$$

$$G(z) = \frac{3 \cdot z}{3 \cdot z - 1}$$

Per poter essere stazionario:

- $u(t)$ deve essere stazionaria, ma essendo costante ha una media costante e quindi è stazionaria.
- $H(z)$ e $G(z)$ sono stabili, ma hanno solo un polo in $p = \frac{1}{3} < 1$ e quindi è stabile.

Se si volesse calcolare il contributo di questa variabile esogena bisognerebbe utilizzare il teorema della risposta in frequenza che afferma: data una sinusoide $u(t) = a \cdot \cos(\omega \cdot t + \varphi)$ si ha un'uscita:

$$w(t) = |G(e^{j\omega})| \cdot a \cdot \cos(\omega \cdot t + \varphi + \angle G(e^{j\omega}))$$

Si usa $G(e^{j\omega})$ perchè si sta lavorando in discreto.

Quindi avendo un input con frequenza nulla, $\omega = 0$, si ha:

$$\begin{aligned} w(t) &= \left| \frac{3 \cdot e^{j \cdot 0}}{3 \cdot e^{j \cdot 0} - 1} \right| \cdot 2 = \\ &= \left| \frac{3 \cdot 1}{3 \cdot 1 - 1} \right| \cdot 2 = \\ &= \frac{3}{2} \cdot 2 = \\ &= 3 \end{aligned}$$

4. Tipi di processi

che è un contributo costante e quindi l'unico effetto è spostare la media del processo stocastico.

Per ricavare la media si può applicare la definizione:

$$\begin{aligned}
 m_v = \mathbb{E}[v(t)] &= \mathbb{E}\left[\frac{1}{3} \cdot v(t-1) + e(t) + 2\right] \\
 m_v &= \frac{1}{3} \cdot \mathbb{E}[v(t-1)] + \mathbb{E}[e(t)] + 2 \\
 m_v - \frac{1}{3} \cdot m_v &= 1 + 2 \\
 m_v \cdot \left(1 - \frac{1}{3}\right) &= 3 \\
 m_v &= 3 \cdot \frac{3}{2} = \frac{9}{2}
 \end{aligned}$$

Per ricavare la funzione di covarianza di depolarizza sia la $v(t)$ che la $e(t)$:

$$\begin{cases} \tilde{v}(t) = v(t) - \frac{9}{2} \\ \tilde{e}(t) = e(t) - 1 \end{cases}$$

quindi il processo può essere scritto come:

$$\begin{aligned}
 v(t) &= \frac{1}{3} \cdot v(t-1) + e(t) + 2 \\
 \tilde{v}(t) + \frac{9}{2} &= \frac{1}{3} \cdot \left(\tilde{v}(t-1) + \frac{9}{2}\right) + \tilde{e}(t) + 1 + 2 \\
 \tilde{v}(t) &= \frac{1}{3} \cdot \tilde{v}(t-1) + \frac{3}{2} + \tilde{e}(t) + 3 - \frac{9}{2} \\
 \tilde{v}(t) &= \frac{1}{3} \cdot \tilde{v}(t-1) + \tilde{e}(t) + \frac{3+6-9}{2} \\
 \tilde{v}(t) &= \frac{1}{3} \cdot \tilde{v}(t-1) + \tilde{e}(t)
 \end{aligned}$$

che è un processo a media nulla, e quindi si può ricavare la funzione di covarianza usando le formule di Yule-Walker:

$$\gamma_v(\tau) = \gamma_{\tilde{v}}(\tau) = \left(\frac{1}{3}\right)^\tau \cdot \frac{1}{1 - \frac{1}{9}} = \frac{9}{8} \cdot \frac{1}{3^\tau} = \frac{3^{2-\tau}}{8}$$

4.5. Osservazioni teoriche

Quando si è introdotta la rappresentazione dinamica di un processo stocastico stazionario v , si è paragonato il rumore bianco a un impulso e si è scritto il processo v come la risposta all'impulso ossia come una combinazione lineare di coefficienti con il rumore bianco dal tempo t fino a $-\infty$, che equivale a un processo $MA(\infty)$. Questo concetto viene dimostrato dal seguente teorema.

Teorema Dato un processo stocastico stazionario $ARMA(n, m)$ può essere scritto come un $MA(\infty)$.

4. Tipi di processi

Esempio Il procedimento per scrivere un processo come un $MA(\infty)$ non è sempre banale, per questo come esempio prendiamo il processo $AR(1)$:

$$\begin{aligned} v(t) &= a \cdot v(t-1) + e(t) \\ \text{con} \quad e(t) &\sim WN(\mu, \lambda^2) \end{aligned}$$

che introducendo l'operatore di ritardo diventa:

$$\begin{aligned} v(t) &= \underbrace{\frac{1}{1 - a \cdot z^{-1}}}_{\text{serie geometrica}} \cdot e(t) = \\ &= \sum_{k=0}^{\infty} (a \cdot z^{-1})^k \cdot e(t) = \\ &= \sum_{k=0}^{\infty} a^k \cdot e(t-k) \end{aligned}$$

che è un $MA(\infty)$.

Questo modo di rappresentare un $AR(1)$ permette di dimostrare formalmente che:

$$\mathbb{E}[v(t-\tau) \cdot e(t)] = 0, \forall \tau > 0$$

ossia che il processo $AR(1)$ non è correlato con il rumore bianco preso in un istante futuro, in quanto:

$$\begin{aligned} \mathbb{E}[v(t-\tau) \cdot e(t)] &= \mathbb{E}\left[\sum_{k=0}^{\infty} a^k \cdot e(t-\tau-k) \cdot e(t)\right] = \\ &= \sum_{k=0}^{\infty} a^k \cdot \mathbb{E}[e(t-\tau-k) \cdot e(t)] = \\ &= \sum_{k=0}^{\infty} a^k \cdot \underbrace{\gamma_e(\tau+k)}_{0 \forall k > 0} = \\ &= \sum_{k=0}^{\infty} a^k \cdot 0 = \\ &= 0 \end{aligned}$$

Metodo alternativo Esiste un metodo che permette di rappresentare un processo $ARMA(m, n)$ come un $MA(\infty)$, che però permette di ricavare solo un numero finito di termini. Prendendo il processo:

$$y(t) = \frac{C(z)}{A(z)} \cdot e(t) \quad \text{con } e(t) \sim WN(\mu, \lambda^2)$$

con:

$$\begin{aligned} C(z) &= c_0 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n} \\ A(z) &= 1 - a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m} \end{aligned}$$

Si può calcolare la divisione tra i due polinomi, che però potrebbe andare avanti all'infinito e quindi ci si ferma a k passi ottenendo:

$$\frac{C(z)}{A(z)} = Q(z) + \frac{R(z)}{A(z)}$$

4. Tipi di processi

dove $Q(z)$ è il risultato della divisione e quindi è un polinomio con potenze negative fino a k , e $R(z)$ è il resto della divisione con potenze negative che partiranno da $k - 1$.

Quindi il processo è uguale a:

$$\begin{aligned} y(t) &= \frac{C(z)}{A(z)} \cdot e(t) = \\ &= \left(Q(z) + \frac{R(z)}{A(z)} \right) \cdot e(t) = \\ &= \underbrace{Q(z) \cdot e(t)}_{MA(k)} + \frac{R(z)}{A(z)} \cdot e(t) \end{aligned}$$

dato che $A(z)$ è un polinomio di grado 0 mentre $R(z)$ è un polinomio di grado $-k$ si può dire che quel rapporto è molto piccolo e quindi che il processo è approssimato da un $MA(k)$ dato da $Q(z) \cdot e(t)$. Se si potesse eseguire la divisione all'infinito si otterrebbe un $MA(\infty)$, ma per ovvie ragioni questo non è possibile. Questa divisione prende il nome di lunga divisione perchè è potenzialmente infinita.

4.6. Curiosità

I modelli *ARMAX* sono i modelli più generali che verranno trattati in questa trattazione e come si è visto permettono di modellizzare l'intero sistema scrivendo le due funzioni di trasferimento

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

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

e si può notare che entrambe le funzioni hanno lo stesso denominatore che dipende dalla parte autoregressiva del modello. Esistono dei modelli ancora più generali detti di Box-Jenkins che permettono di modellare l'intero sistema usando denominatori diversi

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

$$G(z) = \frac{B(z)}{D(z)}$$

questi modelli tuttavia non saranno trattati.

5. Teorema della fattorizzazione spettrale

Si è visto come ricavare lo spettro $\Gamma(\omega)$ partendo dalla funzione di covarianza $\gamma(\tau)$, ma generalmente si tende a rappresentare l'incertezza tramite la rappresentazione dinamica e quindi usando il modello $H(z)$. Il teorema della fattorizzazione spettrale permette di ricavare lo spettro partendo dalla funzione $H(z)$.

5.1. Enunciato

Dato un processo stocastico $v(t) = H(z) \cdot e(t)$, si ha che:

$$\Gamma_v(\omega) = |H(e^{j\omega})|^2 \cdot \Gamma_e(\omega)$$

Osservazioni

- Nella rappresentazione dinamica si ha che $e(t) \sim WN(\mu, \lambda^2)$ di cui è noto lo spettro, $\Gamma_e(\omega) = \lambda^2$, e quindi

$$\Gamma_v(\omega) = |H(e^{j\omega})|^2 \cdot \lambda^2$$

questo rende il calcolo dello spettro molto più semplice rispetto alla serie della definizione.

- Questo teorema ha una formulazione simile a quella del teorema della risposta in frequenza.

5.2. Esempio

Prendiamo in considerazione un processo $MA(1)$, ossia:

$$\begin{aligned} v(t) &= e(t) + c \cdot e(t-1) \\ \text{con} \quad e(t) &\sim WN(0, 1) \end{aligned}$$

dato che per questo tipo di processi è nota la funzione $\gamma(\tau)$ si può applicare sia la definizione che il teorema appena enunciato, quindi per comparativa verranno entrambe esplicitate.

Usando la definizione Ricordando la definizione:

$$\Gamma(\tau) = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) \cdot e^{-j\omega \cdot \tau}$$

si nota che bisogna calcolare una serie infinita, ma i processi $MA(n)$ sono parzialmente colorati e quindi la funzione $\gamma(\tau)$ ha valori non nulli solo per $-n \leq \tau \leq n$. Nel nostro caso si ha che:

$$\begin{aligned} \Gamma(\tau) &= \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) \cdot e^{-j\omega \cdot \tau} = \\ &= \gamma(-1) \cdot e^{-j\omega \cdot (-1)} + \gamma(0) \cdot e^{-j\omega \cdot 0} + \gamma(1) \cdot e^{-j\omega \cdot 1} = \\ &= c \cdot e^{j\omega} + 1 + c^2 + c \cdot e^{-j\omega} = \\ &= 1 + c^2 + c \cdot (e^{j\omega} + e^{-j\omega}) \end{aligned}$$

5. Teorema della fattorizzazione spettrale

ricordando le formule di Eulero:

$$\begin{aligned}e^{j\omega} &= \cos(\omega) + j \cdot \sin(\omega) \\ e^{-j\omega} &= \cos(\omega) - j \cdot \sin(\omega)\end{aligned}$$

si ottiene:

$$\begin{aligned}\Gamma(\tau) &= 1 + c^2 + c \cdot (e^{j\omega} + e^{-j\omega}) = \\ &= 1 + c^2 + c \cdot (\cos(\omega) + \cancel{j \cdot \sin(\omega)} + \cos(\omega) - \cancel{j \cdot \sin(\omega)}) = \\ &= 1 + c^2 + 2 \cdot c \cdot \cos(\omega)\end{aligned}$$

che è una funzione che rispecchia le quattro proprietà espresse nel capitolo 3.1.

Usando il teorema Per prima cosa bisogna ricavare $H(z)$, per farlo si introduce l'operatore di ritardo:

$$\begin{aligned}v(t) &= e(t) + c \cdot e(t-1) = \\ &= e(t) + c \cdot z^{-1} \cdot e(t) = \\ &= e(t) \cdot (1 + c \cdot z^{-1}) =\end{aligned}$$

da cui:

$$H(z) = 1 + c \cdot z^{-1}$$

applicando il teorema, e ricordando che lo spettro di un white noise è costante e uguale alla sua varianza, si ha che:

$$\begin{aligned}\Gamma_v(\omega) &= |H(e^{j\omega})|^2 \cdot \Gamma_e(\omega) = \\ &= |H(e^{j\omega})|^2 \cdot 1 = \\ &= (1 + c \cdot e^{-j\omega}) \cdot (1 + c \cdot e^{j\omega}) = \\ &= 1 + c \cdot e^{-j\omega} + c \cdot e^{j\omega} + c^2 \cdot e^{-j\omega} \cdot e^{j\omega} = \\ &= 1 + c \cdot (e^{-j\omega} + e^{j\omega}) + c^2 \cdot e^0 = \\ &= 1 + c^2 + c \cdot (e^{-j\omega} + e^{j\omega})\end{aligned}$$

che è la stessa espressione di prima, e quindi con gli stessi passaggi si arriva allo stesso risultato di prima:

$$\Gamma_v(\omega) = 1 + c^2 + 2 \cdot c \cdot \cos(\omega)$$

6. Filtro passa-tutto

Il filtro passa-tutto è un filtro di ordine 1 con la seguente forma:

$$T(z) = \frac{1}{a} \cdot \frac{z + a}{z + \frac{1}{a}} \text{ con } a \neq 0 \text{ e } a \in \mathbb{R}$$

Rappresentando modulo e fase di un filtro passa-tutto si ottiene il grafico in figura 6.1.

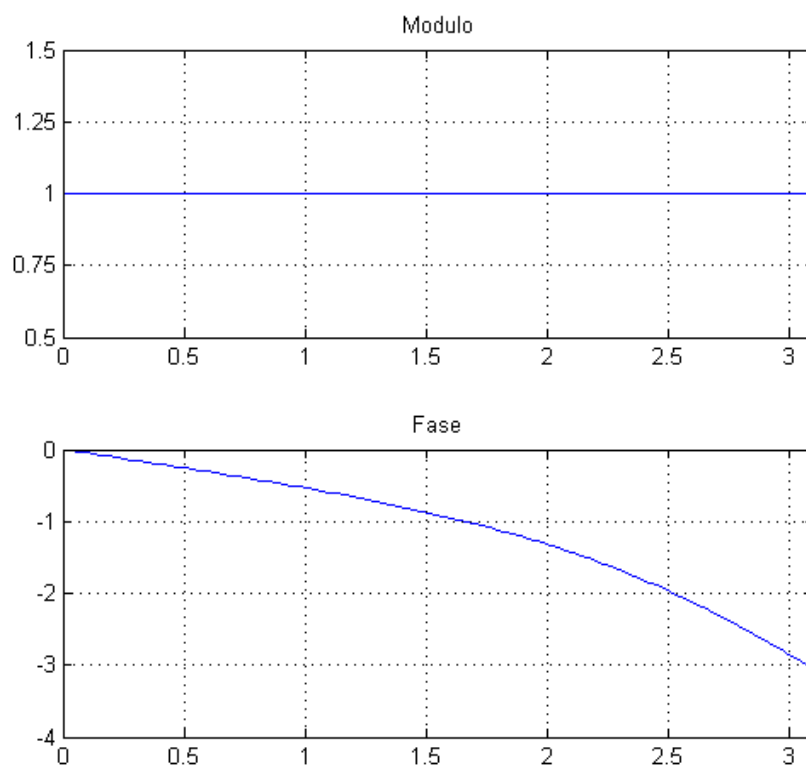


Figura 6.1.: Rappresentazione del diagramma di modulo e fase di un filtro $T(z) = \frac{1}{3} \cdot \frac{z + 3}{z + \frac{1}{3}}$

Ricordando il teorema della fattorizzazione spettrale si ha che:

$$\Gamma_y(\omega) = |T(e^{j\omega})|^2 \cdot \Gamma_e(\omega)$$

6. Filtro passa-tutto

Calcolando il quadrato del modulo si ottiene:

$$\begin{aligned}
 |T(e^{j\omega})|^2 &= \left| \frac{1}{a} \cdot \frac{e^{j\omega} + a}{e^{j\omega} + \frac{1}{a}} \right|^2 = \\
 &= \frac{1}{a} \cdot \frac{e^{j\omega} + a}{e^{j\omega} + \frac{1}{a}} \cdot \frac{1}{a} \cdot \frac{e^{-j\omega} + a}{e^{-j\omega} + \frac{1}{a}} = \\
 &= \frac{1}{a^2} \cdot \frac{1 + a \cdot (e^{j\omega} + e^{-j\omega}) + a^2}{1 + \frac{1}{a} \cdot (e^{j\omega} + e^{-j\omega}) + \frac{1}{a^2}} = \\
 &= \frac{1}{a^2} \cdot \frac{1 + 2 \cdot a \cdot \cos \omega + a^2}{\frac{1}{a^2} \cdot (a^2 + 2 \cdot a \cdot \cos \omega + 1)} = \\
 &= 1
 \end{aligned}$$

e quindi:

$$\begin{aligned}
 \Gamma_y(\omega) &= |T(e^{j\omega})|^2 \cdot \Gamma_e(\omega) = \\
 &= 1 \cdot \Gamma_e(\omega) = \\
 &= \Gamma_e(\omega)
 \end{aligned}$$

Questo vuol dire che il filtro passa-tutto non modifica lo spettro del segnale e quindi il segnale di ingresso e quello di uscita da questo filtro sono equivalenti, non uguali.

In realtà questo filtro ha uno zero e un polo non coincidenti e quindi il segnale d'ingresso viene sfasato, ma questo non compromette l'equivalenza.

È quindi possibile rappresentare i sistemi come in figura 6.2.

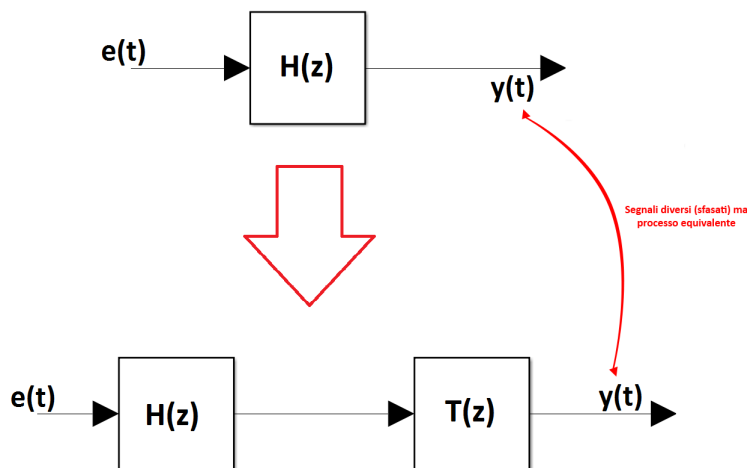


Figura 6.2.: Rappresentazione dinamica mediante l'utilizzo di un filtro passa-tutto

7. Forma canonica

Lo stesso processo si può scrivere in infiniti modi equivalenti, ad esempio prendendo il seguente processo:

$$y_1(t) = \frac{z + \frac{1}{2}}{z - \frac{1}{3}} \cdot e(t) \text{ con } e(t) \sim WN(0, 1)$$

con spettro:

$$\begin{aligned} \Gamma_{y_1}(\omega) &= \left| \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}} \right|^2 \cdot 1 = \\ &= \left| \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}} \right|^2 \end{aligned}$$

- Lo spettro, e quindi il processo, non cambia aggiungendo un ritardo, infatti prendendo il processo:

$$y_2(t) = \frac{z + \frac{1}{2}}{z - \frac{1}{3}} \cdot e(t - 2) \text{ con } e(t) \sim WN(0, 1)$$

$$\begin{aligned} y_2(t) &= \frac{z + \frac{1}{2}}{z - \frac{1}{3}} \cdot e(t - 2) = \\ &= \frac{z + \frac{1}{2}}{z - \frac{1}{3}} \cdot z^{-2} \cdot e(t) \end{aligned}$$

si ottiene uno spettro:

$$\begin{aligned} \Gamma_{y_2}(\omega) &= \left| \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}} \cdot e^{-2j\omega} \right|^2 \cdot 1 = \\ &= \left| \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}} \right|^2 \cdot |e^{-2j\omega}|^2 = \\ &= \Gamma_{y_1}(\omega) \cdot (e^{-2j\omega} \cdot e^{2j\omega}) = \\ &= \Gamma_{y_1}(\omega) \end{aligned}$$

7. Forma canonica

- Lo spettro, e quindi il processo, non cambia aggiungendo aggiungendo poli e zeri cancellabili, infatti prendendo il processo:

$$y_3(t) = \frac{z^2 - \frac{1}{2}}{z^2 - \frac{5}{6} \cdot z + \frac{1}{6}} \cdot e(t) \text{ con } e(t) \sim WN(0, 1)$$

$$\begin{aligned} y_3(t) &= \frac{z^2 - \frac{1}{2}}{z^2 - \frac{5}{6} \cdot z + \frac{1}{6}} \cdot e(t) = \\ &= \frac{\left(z - \frac{1}{2}\right) \cdot \left(z + \frac{1}{2}\right)}{\left(z + \frac{1}{3}\right) \cdot \left(z - \frac{1}{2}\right)} \cdot e(t) = \\ &= y_1(t) \end{aligned}$$

- Lo spettro, e quindi il processo, non cambia aggiungendo aggiungendo fattori moltiplicativi al processo e al rumore in modo che si annullino, infatti prendendo il processo:

$$y_4(t) = \frac{2 \cdot z + 1}{z - \frac{1}{3}} \cdot e(t) \text{ con } e(t) \sim WN\left(0, \frac{1}{4}\right)$$

si ottiene lo spettro:

$$\begin{aligned} \Gamma_{y_4}(\omega) &= \left| 2 \cdot \frac{e^{j \cdot \omega} + \frac{1}{2}}{e^{j \cdot \omega} - \frac{1}{3}} \right|^2 \cdot \frac{1}{4} = \\ &= 4 \cdot \left| \frac{e^{j \cdot \omega} + \frac{1}{2}}{e^{j \cdot \omega} - \frac{1}{3}} \right|^2 \cdot \frac{1}{4} = \\ &= 4 \cdot \Gamma_{y_1}(\omega) \cdot \frac{1}{4} = \\ &= \Gamma_{y_1}(\omega) \end{aligned}$$

- Lo spettro, e quindi il processo, non cambia se si cambiano poli e zeri con i loro reciproci, infatti prendendo il processo:

$$y_5(t) = \frac{z + 2}{z - \frac{1}{3}} \cdot e(t) \text{ con } e(t) \sim WN\left(0, \frac{1}{4}\right)$$

7. Forma canonica

si può applicare un filtro passa-tutto ottenendo:

$$\begin{aligned}
 y_5(t) &= \frac{z+2}{z-\frac{1}{3}} \cdot e(t) = \\
 &= \frac{\cancel{z+2}}{z-\frac{1}{3}} \cdot \frac{z+\frac{1}{2}}{\cancel{z+2}} \cdot 2 \cdot e(t) = \\
 &= 2 \cdot \frac{z+\frac{1}{2}}{z-\frac{1}{3}} \cdot e(t) = \\
 &= \frac{z+1}{z-\frac{1}{3}} \cdot e(t) = \\
 &= y_4 = \\
 &= y_1
 \end{aligned}$$

Dato che esistono infinite rappresentazioni equivalenti dello stesso processo c'è bisogno di definire una forma univoca, detta canonica.

Dato un processo stazionario nella forma:

$$y(t) = \frac{C(z)}{A(z)} \cdot e(t) \text{ con } e(t) \sim WN(\mu, \lambda^2)$$

si dice che è in forma canonica se e solo se:

- $C(z)$ e $A(z)$ hanno lo stesso grado (ovvero si aggiungono ritardi finché non hanno gli stessi gradi)
- $C(z)$ e $A(z)$ sono coprimi (ovvero non ci sono fattori comuni tra i due, quindi non si possono fare altre semplificazioni)
- $C(z)$ e $A(z)$ sono monici (ovvero il coefficiente del termine di grado massimo è uguale a 1)
- $C(z)$ e $A(z)$ hanno radici strettamente interne al cerchio di raggio unitario¹ (ovvero si mettono dei filtri passa-tutto per invertire gli zeri, i poli lo sono già dato che è stabile)

Esempio Prendendo il processo:

$$y(t) = \frac{z+2}{z-\frac{1}{3}} \cdot e(t-2) \text{ con } e(t) \sim WN(0, 1)$$

¹ossia il sistema è a fase minima

7. Forma canonica

Applicando un filtro passa-tutto è possibile rendere il sistema a fase minima:

$$\begin{aligned}y(t) &= \frac{\cancel{z+2}}{z-\frac{1}{3}} \cdot 2 \cdot \frac{z+\frac{1}{2}}{\cancel{z+2}} \cdot e(t-2) = \\&= 2 \cdot \frac{z+\frac{1}{2}}{z-\frac{1}{3}} \cdot e(t-2) = \\&= 2 \cdot \frac{z}{z} \cdot \frac{1+\frac{1}{2} \cdot z^{-1}}{1-\frac{1}{3} \cdot z^{-1}} \cdot e(t-2) = \\&= 2 \cdot \frac{1+\frac{1}{2} \cdot z^{-1}}{1-\frac{1}{3} \cdot z^{-1}} \cdot e(t-2) =\end{aligned}$$

Introducendo il rumore:

$$\eta(t) = 2 \cdot e(t-2)$$

dove:

$$\begin{aligned}\mathbb{E}[\eta(t)] &= \mathbb{E}[2 \cdot e(t-2)] = 2 \cdot \mathbb{E}[e(t-2)] = 2 \cdot 0 = 0 \\Var[\eta(t)] &= Var[2 \cdot e(t-2)] = 4 \cdot Var[e(t-2)] = 4 \cdot 1 = 4\end{aligned}$$

si ottiene:

$$y(t) = \frac{1+\frac{1}{2} \cdot z^{-1}}{1-\frac{1}{3} \cdot z^{-1}} \cdot \eta(t) \text{ con } \eta(t) \sim WN(0, 4)$$

7.1. Procedimento di canonizzazione

Per portare in forma canonica un processo si può seguire l'ordine dei quattro punti che si preferisce, ma in genere si procede così:

1. Si scrive il processo tramite la sua forma dinamica, ottenendo un numeratore $N(z)$ e un denominatore $D(z)$
2. Si scompongono i due polinomi per mettere in evidenza poli e zeri.
3. Si riduce ai minimi termini, ossia si semplificano poli e zeri coincidenti.
4. Si applicano i filtri passa-tutto per portare nel cerchio di raggio unitario gli zeri (i poli devono esserlo di già).
5. Si rimoltiplicano i termini dei due polinomi.
6. Si raccoglie sia al numeratore che al denominatore il termine di grado massimo in modo di scriverli nella forma con potenze negative, e quindi portarli portarli allo stesso grado e anche in forma monica.

7. Forma canonica

7. A questo punto si ridefinisce l'errore in modo che inglobi i termini raccolti nel passo precedente e eventuali ritardi presenti nel rumore.
8. Il processo è ora in forma canonica.

Parte II.

Stimatori campionari

8. Richiami sulla teoria della stima

Per stima si intende ricavare un valore $\hat{\theta}$, detto stimatore, che approssima un parametro nascosto θ tramite una serie di N di misure $\{y_1, y_2, \dots, y_N\}$.

Le N misure sono prese da una variabile casuale e quindi ci possono essere infinite serie di N misure; la stima $\hat{\theta}$ si calcolerà a partire dai dati, ovvero $\hat{\theta} = f(y_1, y_2, \dots, y_N)$, quindi lo stimatore dipende dalla realizzazione usata e questo significa che è una variabile casuale. Se ne può quindi calcolare il valore atteso e la varianza.

Definizione Dato uno stimatore $\hat{\theta}$ di θ si definisce errore di stima:

$$\varepsilon = \hat{\theta} - \theta$$

Definizione Uno stimatore $\hat{\theta}$ di θ si definisce corretto se e solo se:

$$\mathbb{E}[\hat{\theta}] = \theta$$

e quindi:

$$\mathbb{E}[\varepsilon] = \mathbb{E}[\hat{\theta} - \theta] = 0$$

Definizione Uno stimatore $\hat{\theta}$ di θ si definisce asintoticamente corretto se e solo se:

$$\lim_{N \rightarrow +\infty} \mathbb{E}[\hat{\theta}] = \theta$$

e quindi:

$$\lim_{N \rightarrow +\infty} \mathbb{E}[\varepsilon] = \mathbb{E}[\hat{\theta} - \theta] = 0$$

Osservazione

- La proprietà di correttezza garantisce che lo stimatore avrà un valore medio pari al parametro da stimare, mentre l'asintotica correttezza garantisce solo che all'aumentare del numero di campioni ci si avvicina alla correttezza ed è quindi una proprietà più "debole" della semplice correttezza.
- La proprietà di correttezza non basta per capire se uno stimare è buono perchè non spiega se si può migliorare in qualche modo la stima.

Definizione Uno stimatore $\hat{\theta}$ di θ si definisce consistente se e solo se:

$$\begin{aligned} \lim_{N \rightarrow +\infty} \text{Var}(\hat{\theta}) &= 0 \\ \lim_{N \rightarrow +\infty} \mathbb{E}[(\hat{\theta} - \mathbb{E}[\hat{\theta}])^2] &= 0 \end{aligned}$$

Osservazione

- La proprietà di consistenza garantisce che all'aumentare del numero dei campioni la qualità della stima aumenta, e quindi fornisce un modo di migliorare la stima aggiungendo ulteriore informazione.
- Se lo stimatore è corretto si nota che: $Var(\hat{\theta}) = \mathbb{E} \left[\left(\hat{\theta} - \mathbb{E}[\hat{\theta}] \right)^2 \right] = \mathbb{E} \left[\left(\hat{\theta} - \theta \right)^2 \right] = Var(\varepsilon)$
- Uno stimatore si dice buono se è sia corretto (o almeno asintoticamente corretto) che consistente.

Definizione Uno stimatore $\hat{\theta}$ di θ si definisce ottimo se la sua varianza è la minima possibile per una data serie di N di misure $\{y_1, y_2, \dots, y_N\}$.

Osservazione

- Una condizione necessaria, ma non sufficiente, per l'ottimalità è che la stima e l'errore di stima siano scorrelati:

$$\mathbb{E} \left[\hat{\theta} \cdot \varepsilon \right] = 0$$

perchè se così non fosse vorrebbe dire che sono presenti delle informazioni che, nonostante potrebbero essere usate nella stima, non vengono sfruttate, rimanendo così nell'errore. In questi casi bisogna trovare un modo per spostare questa informazione dall'errore alla stima.

9. Media campionaria

La media campionaria è uno stimatore del valore atteso m_v di un processo stazionario v dati N campioni $y(t)$ ottenuti dalla realizzazione $v(t, S)$. Lo stimatore si ricava tramite:

$$\hat{m}_v = \frac{1}{N} \cdot \sum_{t=1}^N y(t)$$

Correttezza

$$\begin{aligned} \mathbb{E}[\hat{m}_v] &= \mathbb{E}\left[\frac{1}{N} \cdot \sum_{t=1}^N y(t)\right] = \\ &= \frac{1}{N} \cdot \sum_{t=1}^N \mathbb{E}[y(t)] = \\ &= \frac{1}{N} \cdot N \cdot \mathbb{E}[y(t)] = \\ &= m_v \end{aligned}$$

Quindi lo stimatore è corretto.

Consistenza

La consistenza è garantita dai seguenti teoremi.

Teorema \hat{m}_v è consistente se e solo se $\gamma(\tau) \rightarrow 0$ per $\tau \rightarrow +\infty$.

Teorema Dato un processo *ARMA* si ha che $\gamma(\tau) \rightarrow 0$ per $\tau \rightarrow +\infty$.

10. Covarianza campionaria

La covarianza campionaria è uno stimatore della funzione $\gamma(\tau)$ di un processo stazionario v a media nulla¹ dati N campioni $y(t)$ ottenuti dalla realizzazione $v(t, S)$. Lo stimatore si ricava tramite:

$$\hat{\gamma}_v(\tau) = \frac{1}{N - |\tau|} \cdot \sum_{t=1}^{N-|\tau|} y(t) \cdot y(t + |\tau|)$$

Osservazioni

- Si può notare che per τ alti diminuisce il numero di termini che si vanno a sommare diminuendo quindi l'accuratezza dello stimatore. Per questo motivo l'accuratezza dello stimatore è alta per valori $|\tau| \ll N$.
- Il valore di τ compare sempre tra valori assoluti perchè la funzione $\gamma(\tau)$ è pari e quindi $\gamma(\tau) = \gamma(-\tau)$

Correttezza

$$\begin{aligned} \mathbb{E}[\hat{\gamma}_v(\tau)] &= \mathbb{E}\left[\frac{1}{N - |\tau|} \cdot \sum_{t=1}^{N-|\tau|} y(t) \cdot y(t + |\tau|)\right] = \\ &= \frac{1}{N - |\tau|} \cdot \sum_{t=1}^{N-|\tau|} \mathbb{E}[y(t) \cdot y(t + |\tau|)] = \\ &= \frac{1}{N - |\tau|} \cdot (N - |\tau|) \cdot \gamma(\tau) = \\ &= \gamma(\tau) \end{aligned}$$

Quindi lo stimatore è corretto.

Consistenza

La consistenza è garantita dal seguente teorema

Teorema $\hat{\gamma}_v(\tau)$ è consistente se e solo se $\gamma(\tau) \rightarrow 0$ per $\tau \rightarrow +\infty$.

10.1. Variante non corretta

Esiste una variante di questo stimatore che si ricava da:

$$\hat{\gamma}'_v(\tau) = \frac{1}{N} \cdot \sum_{t=1}^{N-|\tau|} y(t) \cdot y(t + |\tau|)$$

¹Non è un'ipotesi forte perchè si può sempre depolarizzare stimando la media

questa variante non è corretta infatti:

$$\begin{aligned}
 \mathbb{E} [\hat{\gamma}'_v(\tau)] &= \mathbb{E} \left[\frac{1}{N} \cdot \sum_{t=1}^{N-|\tau|} y(t) \cdot y(t+|\tau|) \right] = \\
 &= \frac{1}{N} \cdot \sum_{t=1}^{N-|\tau|} \mathbb{E} [y(t) \cdot y(t+|\tau|)] = \\
 &= \frac{1}{N} \cdot (N - |\tau|) \cdot \gamma(\tau) = \\
 &= \frac{N - |\tau|}{N} \cdot \gamma(\tau)
 \end{aligned}$$

ma è asintoticamente corretto, infatti:

$$\begin{aligned}
 \lim_{N \rightarrow +\infty} \mathbb{E} [\hat{\gamma}'_v(\tau)] &= \lim_{N \rightarrow +\infty} \frac{N - |\tau|}{N} \cdot \gamma(\tau) = \\
 &= \lim_{N \rightarrow +\infty} \frac{N}{N} \cdot \gamma(\tau) = \\
 &= \gamma(\tau)
 \end{aligned}$$

Questo stimatore è utilizzabile perché l'ipotesi che $N \gg |\tau|$ deve essere verificata anche per l'altro stimatore e quindi non è una assunzione molto restrittiva.

11. Densità spettrale campionaria

Sappiamo dalla definizione che è possibile calcolare lo spettro attraverso la definizione dello stesso:

$$\Gamma(\omega) = \sum_{\tau=-\infty}^{+\infty} \gamma(\tau) \cdot e^{-j\omega\tau}$$

Partendo da questa formula, viene naturale pensare che lo stimatore dello spettro abbia una forma come quella che segue:

$$\hat{\Gamma}(\omega) = \sum_{\tau=-\infty}^{+\infty} \hat{\gamma}_N(\tau) \cdot e^{-j\omega\tau}$$

Ricordando però la formula dello stimatore della funzione di covarianza $\hat{\gamma}_N$:

$$\hat{\gamma}_N(\tau) = \frac{1}{N-|\tau|} \cdot \sum_{t=1}^{N-|\tau|} y(t) \cdot y(t+|\tau|)$$

notiamo che non è possibile fare una sommatoria tra $-\infty$ e $+\infty$, in quanto, dati N campioni, lo stimatore ci può dare solo $\hat{\gamma}_N(\tau)$ per $-(N-1) \leq \tau \leq (N-1)$. Si modifica quindi lo stimatore dello spettro per permetterci di usare lo stimatore della funzione di covarianza al suo interno:

$$\hat{\Gamma}(\omega) = \sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}_N(\tau) \cdot e^{-j\omega\tau}$$

Osservazioni

- Abbiamo approssimato lo stimatore usando uno stimatore al suo interno: $\hat{\gamma}_N(\tau)$
- Abbiamo approssimato ulteriormente troncando la sommatoria

Correttezza

Verifichiamo che sia corretto:

$$\begin{aligned} \mathbb{E}[\hat{\Gamma}(\omega)] &= \mathbb{E}\left[\sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}_N(\tau) \cdot e^{-j\omega\tau}\right] = \\ &= \sum_{\tau=-(N-1)}^{N-1} \underbrace{\mathbb{E}[\hat{\gamma}_N(\tau)]}_{\gamma(\tau)} \cdot e^{-j\omega\tau} \end{aligned}$$

Effettivamente il valore atteso dello stimatore somiglia a $\Gamma(\omega)$, a meno degli indici della sommatoria. Possiamo però notare che per $N \rightarrow +\infty$ il troncamento della sommatoria diventa sempre meno rilevante. Lo stimatore è quindi

11. Densità spettrale campionaria

asintoticamente corretto, in quanto

$$\begin{aligned}
 \lim_{N \rightarrow +\infty} \mathbb{E} [\hat{\Gamma}(\omega)] &= \lim_{N \rightarrow +\infty} \mathbb{E} \left[\sum_{\tau=-(N-1)}^{N-1} \hat{\gamma}_N(\tau) \cdot e^{-j \cdot \omega \cdot \tau} \right] = \\
 &= \lim_{N \rightarrow +\infty} \sum_{\tau=-(N-1)}^{N-1} \underbrace{\mathbb{E} [\hat{\gamma}_N(\tau)]}_{\gamma(\tau)} \cdot e^{-j \cdot \omega \cdot \tau} = \\
 &= \sum_{\tau=-\infty}^{\infty} \gamma(\tau) \cdot e^{-j \cdot \omega \cdot \tau} = \\
 &= \Gamma(\omega)
 \end{aligned}$$

Consistenza

Il calcolo della proprietà di consistenza per questo stimatore è molto complicato, perciò si riportano solo i risultati:

$$\lim_{N \rightarrow +\infty} \mathbb{E} \left[\left(\hat{\Gamma}(\omega) - \Gamma(\omega) \right)^2 \right] = \Gamma(\omega)^2 \neq 0$$

Quindi lo stimatore non risulta consistente. Inoltre vale che:

$$\lim_{N \rightarrow +\infty} \mathbb{E} \left[\left(\hat{\Gamma}(\omega_1) - \Gamma(\omega_1) \right) \cdot \left(\hat{\Gamma}(\omega_2) - \Gamma(\omega_2) \right) \right] = 0, \forall \omega_1, \omega_2 \text{ con } \omega_1 \neq \omega_2$$

Questo vuol dire che anche conoscendo qual'è l'errore dello stimatore per una data ω_1 , non è possibile dire nulla sull'errore ad un'altra ω_2 . Per la prima proprietà però, si sa che a grandi valori di $\Gamma(\omega)$ corrispondono grandi valori dell'errore dello stimatore. Per questi motivi lo stimatore dello spettro ha una varianza che dipende dal parametro nascosto stesso e quindi è difficile valutare la varianza della stima calcolata.

11.1. Utilizzo del secondo stimatore della funzione di covarianza

Come si è visto nel capitolo 10.1, esiste una variante $\gamma'(\tau)$ non corretta dello stimatore della funzione di covarianza. Il vantaggio nell'usarlo nel calcolo dello stimatore dello spettro sta nel fatto che

$$\begin{aligned}
 \hat{\Gamma}'_N(\omega) &= \sum_{\tau=-(N-1)}^{N-1} \gamma'(\tau) \cdot e^{-j \cdot \omega \cdot \tau} = \\
 &= \frac{1}{N} \cdot \left(\sum_{t=1}^N y(t) \cdot e^{-j \cdot \omega \cdot t} \right)^2
 \end{aligned}$$

che è la trasformata di Fourier discreta. Questo significa che è possibile calcolare lo spettro senza passare dalla funzione di covarianza e utilizzando invece un'algoritmo per calcolare la DFT (Discrete Fourier Transform).

Questo stimatore, dato che utilizza uno stimatore della funzione di covarianza non corretto, è meno corretto del precedente, ma è molto più facile da calcolare usando algoritmi noti e veloci.

11.2. Regolarizzazione

Avendo un data set di N campioni, calcolando lo spettro usando lo stimatore appena visto, si ottiene $\Gamma_N(\omega)$. Dividendo però in 4 parti il data set posso calcolare $1\hat{\Gamma}_{\frac{N}{4}}(\omega)$, $2\hat{\Gamma}_{\frac{N}{4}}(\omega)$, $3\hat{\Gamma}_{\frac{N}{4}}(\omega)$ e $4\hat{\Gamma}_{\frac{N}{4}}(\omega)$ e poi la loro media $\hat{\Gamma}_N(\omega) = \frac{1}{4} \cdot \sum_{i=1}^4 i\hat{\Gamma}_{\frac{N}{4}}$,

11. Densità spettrale campionaria

vale che:

$$\frac{1}{4} \cdot \mathbb{E} \left[\left(\hat{\Gamma}(\omega) - \Gamma(\omega) \right)^2 \right] = \mathbb{E} \left[\left(\hat{\Gamma}_N(\omega) - \Gamma(\omega) \right)^2 \right]$$

Generalizzando:

$$\frac{1}{M} \cdot \mathbb{E} \left[\left(\hat{\Gamma}(\omega) - \Gamma(\omega) \right)^2 \right] = \mathbb{E} \left[\left(\hat{\Gamma}_N(\omega) - \Gamma(\omega) \right)^2 \right]$$

con M il numero di parti in cui si è deciso di dividere il data set. Si presenta quindi un trade off: si deve scegliere M abbastanza grande da ridurre l'errore dello stimatore, ma non talmente grande da intaccare la correttezza (che ricordiamo essere migliore per numeri di campioni N più grandi).

Un'esempio di come lo stimatore possa trarre beneficio dalla regolarizzazione è rappresentato in figura 11.1.

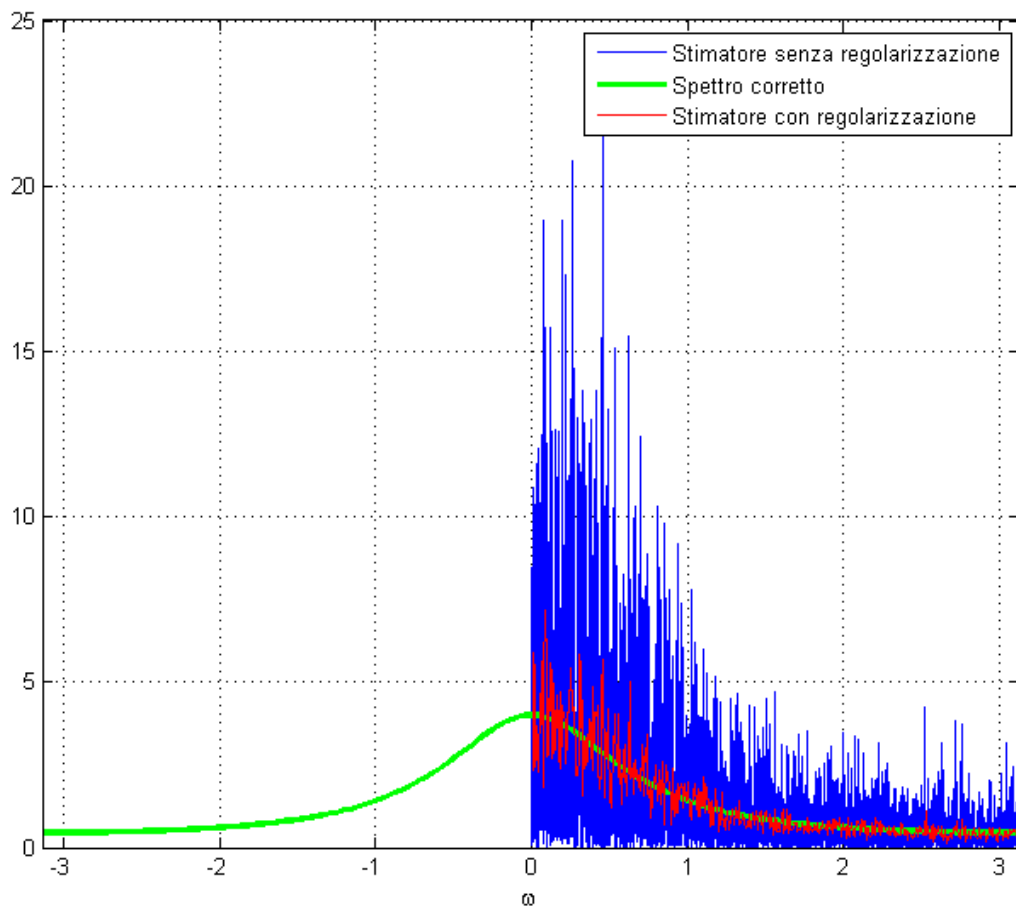


Figura 11.1.: Rappresentazione dello stimatore dello spettro

Parte III.

Predizione

Introduzione

Con predizione o forecasting si intende stimare un dato al tempo $t + k$ conoscendo le misure fino al tempo t , e si indica con $\hat{y}(t + k|t)$.

Questa stima viene definita anche Forecasting e non va confusa con:

- Controllo
Ossia fare in modo che al tempo $t + k$, $y(t + k)$ assuma un valore voluto.
- Identificazione
Ossia stimare i parametri del modello del sistema date le misure fino al tempo t

Il problema di predizione viene illustrato in figura 11.2.

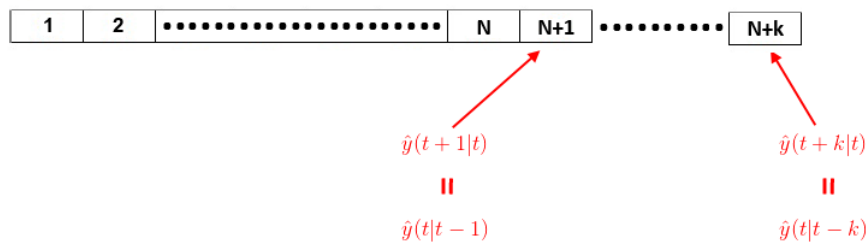


Figura 11.2.: Schema di un problema di predizione

Osservazioni

- Concettualmente parlare di $\hat{y}(t + k|t)$, ossia stimare il valore al tempo $t + k$ conoscendo i dati fino al tempo t , e parlare di $\hat{y}(t|t - k)$, ossia stimare il valore al tempo t conoscendo i dati fino al tempo $t - k$, è la stessa cosa infatti sono entrambe stime a k passi.
- Dato che il predittore è uno stimatore si possono definire:
 - L'errore di predizione: $\varepsilon(t) = y(t) - \hat{y}(t|t - k)$
 - La varianza del predittore $Var[\hat{y}(t|t - k)] = \mathbb{E}[(\hat{y}(t|t - k) - \mathbb{E}[\hat{y}(t|t - k)])^2] = \mathbb{E}[(\hat{y}(t|t - k) - y(t))^2] = Var[\varepsilon(t)]$, se è uno stimatore corretto.

Si consideri il problema di stima $\hat{y}(4|3)$, ossia bisogna stimare $y(4)$ conoscendo i valori: $y(0), y(1), y(2), y(3)$. Lo stimatore è una funzione dei dati conosciuti, e ci possono essere vari modi. Il metodo più banale è fare la media dei valori, ossia:

$$\hat{y}(4|3) = \frac{1}{4} \cdot (y(0) + y(1) + y(2) + y(3))$$

ma potrebbe anche essere ragionevole assegnare dei pesi alle varie misure, contando sul fatto che quelle più vicine temporalmente probabilmente sono più significative di quelle lontane; esistono quindi innumerevoli tipo di predittori. Bisogna, però, tener presente che la proprietà di correttezza è fondamentale e quindi:

$$\mathbb{E}[\varepsilon(t)] = \mathbb{E}[y(t) - \hat{y}(t|t - k)] = 0$$

nel caso della media si ha:

$$\begin{aligned}\mathbb{E}[\varepsilon(t)] &= \mathbb{E}[y(t) - \hat{y}(t|t-k)] \\ &= \mathbb{E}[y(4)] - \frac{1}{4} \cdot \mathbb{E}[y(0) + y(1) + y(2) + y(3)] = \\ &= m - \frac{1}{4} \cdot 4 \cdot m = 0\end{aligned}$$

Nella teoria dei predittori, però, non ci si accontenta di un predittore corretto, ma si vuole trovare quello ottimo, quindi:

- Bisogna rispettare la condizione necessaria: $\mathbb{E}[\hat{y}(t+k|t) \cdot \varepsilon(t)] = 0$
- Bisogna rispettare la definizione di ottimalità: $Var[\hat{y}(t|t-k)] = Var[\varepsilon(t)]$ deve essere la minima possibile.

Esempio Come già detto $\mathbb{E}[\hat{y}(t+k|t) \cdot \varepsilon(t)] = 0$ è solo una condizione necessaria, e questo esempio lo evidenzia. Supponendo di avere un processo $MA(3)$ in forma canonica:

$$y(t) = e(t) + c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + c_3 \cdot e(t-3) \text{ con } e(t) \sim WN(0, \lambda^2)$$

e di utilizzare come predittore:

$$\hat{y}_1(t|t-1) = c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + c_3 \cdot e(t-3)$$

si nota che:

$$\varepsilon_1(t) = e(t)$$

$$\begin{aligned}\mathbb{E}[\hat{y}_1(t+k|t) \cdot \varepsilon_1(t)] &= \mathbb{E}[(c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + c_3 \cdot e(t-3)) \cdot e(t)] = 0 \\ Var(\varepsilon_1(t)) &= Var(e(t)) = \lambda^2\end{aligned}$$

Supponendo ora di usare il predittore:

$$\hat{y}_2(t|t-1) = c_1 \cdot e(t-1) + c_3 \cdot e(t-3)$$

si nota che:

$$\begin{aligned}\varepsilon_2(t) &= e(t) + c_2 \cdot e(t-2) \\ \mathbb{E}[\hat{y}_2(t+k|t) \cdot \varepsilon_2(t)] &= \mathbb{E}[(c_1 \cdot e(t-1) + c_3 \cdot e(t-3)) \cdot (e(t) + e(t-2))] = 0 \\ Var(\varepsilon_2(t)) &= Var(e(t) + c_2 \cdot e(t-2)) = \lambda^2(1 + c_2^2) > \lambda^2 = \varepsilon_1(t)\end{aligned}$$

La varianza del secondo stimatore è maggiore rispetto a quella del primo e quindi il secondo stimatore non è ottimo nonostante la verifica della condizione necessaria. Si può anche notare da questo esempio che non si può dire nulla sull'ottimalità del primo stimatore.

Osservazione Dato $\varepsilon(t) = y(t) - \hat{y}(t|t-k)$, si ricava $y(t) = \varepsilon(t) + \hat{y}(t|t-k)$, dove:

- $\varepsilon(t)$ è la parte imprevedibile al tempo $t-k$, ovvero l'errore
- $\hat{y}(t|t-k)$ è la parte predicibile del processo al tempo $t-k$

12. Predittore ottimo di un processo $MA(n)$

Ricordando la forma di un processo $MA(n)$:

$$\begin{aligned} y(t) &= e(t) + c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + \dots + c_n \cdot e(t-n) \\ \text{con} \quad e(t) &\sim WN(0, \lambda^2) \end{aligned}$$

Per evitare ambiguità si suppone che il processo sia in forma canonica (deve essere monico e quindi $c_0 = 1$, e deve valere che $|c_i| < 1$, $\forall i \leq n$ e avere tutte le radici all'interno del cerchio di raggio unitario)

Supponendo di conoscere i valori assunti dal rumore bianco fino all'istante $t-1$ si può notare che del processo $MA(n)$ è tutto noto escluso l'errore al passo t e quindi potrebbe essere un predittore a un passo:

$$\hat{y}(t|t-1) = c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + \dots + c_n \cdot e(t-n)$$

da cui:

$$\begin{aligned} \varepsilon(t) &= y(t) - \hat{y}(t|t-1) = \\ &= e(t) + c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + \dots + c_n \cdot e(t-n) - (c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + \dots + c_n \cdot e(t-n)) = \\ &= e(t) \end{aligned}$$

Questo stimatore è corretto infatti:

$$\mathbb{E}[\varepsilon(t)] = \mathbb{E}[e(t)] = 0$$

Inoltre la condizione necessaria di ottimalità è verificata:

$$\mathbb{E}[\hat{y}(t|t-1) \cdot \varepsilon(t)] = \sum_{i=1}^n c_i \cdot \underbrace{\mathbb{E}[e(t) \cdot e(t-i)]}_{\text{sempre } 0} = 0$$

e si può dimostrare che la $Var[\varepsilon(t)]$ è minima e quindi lo stimatore è ottimo.

Questo stimatore, però, si basa sull'assunzione di conoscere i valori assunti dal rumore bianco e quindi non è realizzabile in quanto si conoscono solo i valori di y . Si può però notare che:

$$\begin{aligned} y(t) &= (1 + c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n}) \cdot e(t) \\ e(t) &= \frac{1}{1 + c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n}} \cdot y(t) \end{aligned}$$

12. Predittore ottimo di un processo $MA(n)$

Si è quindi trovata un'espressione che trasforma i valori del processo (ovvero $y(t)$) in un rumore bianco (ovvero $e(t)$) che viene chiamata filtro sbiancante, in quanto sbianca il processo. Sostituendo nell'espressione del predittore:

$$\begin{aligned}
 \hat{y}(t|t-1) &= c_1 \cdot e(t-1) + c_2 \cdot e(t-2) + \dots + c_n \cdot e(t-n) = \\
 \hat{y}(t|t-1) &= (c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n}) \cdot e(t) = \\
 \hat{y}(t|t-1) &= \frac{c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n}}{1 + c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n}} \cdot y(t) \\
 \hat{y}(t|t-1) \cdot (1 + c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n}) &= (c_1 + c_2 \cdot z^{-1} + \dots + c_n \cdot z^{-n-1}) \cdot y(t-1) \\
 \hat{y}(t|t-1) &= \underbrace{-c_1 \cdot \hat{y}(t-1|t-2) - \dots - c_n \cdot \hat{y}(t-n|t-n-1)}_{\text{predizioni ai passi precedenti}} + \\
 &\quad + \underbrace{c_1 \cdot y(t-1) + \dots + c_n \cdot y(t-n)}_{\text{dati noti fino a } t-1}
 \end{aligned}$$

In questo modo la predizione dipende da:

- i valori di y fino a $t-1$ che sono noti.
- i valori assunti dal predittore negli istanti precedenti; questo comporta che il calcolo di $\hat{y}(t|t-1)$ diventa ricorsivo e richiede quindi un'inizializzazione, ossia bisogna stabilire quanto vale $y(1|0)$.
Ci sono vari modi per inizializzare, ma un buon modo è usare la media dei campioni, ossia uno stimatore corretto, anche se non ottimo.
- i parametri c_i del modello del processo; questa predizione è ottima se il modello usato rappresenta bene il sistema, ma questo è un problema di identificazione e verrà trattato in seguito.

In modo analogo si può definire il predittore al passo $k \leq n$, ossia $\hat{y}(t|t-k)$, la differenza è che in questo caso si conoscono i valori solo fino al tempo $t-k$ quindi il predittore è:

$$\hat{y}(t|t-k) = c_k \cdot e(t-k) + c_{k+1} \cdot e(t-k-1) + \dots + c_n \cdot e(t-n)$$

Si può notare che se $k > n$ i dati che si possiedono non sono in nessun modo correlati al valore che si vuole predire e quindi è come predire un rumore bianco, ossia la media è la miglior stima.

Questo predittore ha lo stesso problema di quello a un passo e quindi si applica la stessa sostituzione fatta prima, ottenendo:

$$\begin{aligned}
 \hat{y}(t|t-k) &= c_k \cdot e(t-k) + c_{k+1} \cdot e(t-k-1) + \dots + c_n \cdot e(t-n) = \\
 \hat{y}(t|t-k) &= (c_k \cdot z^{-k} + c_{k+1} \cdot z^{-k-1} + \dots + c_n \cdot z^{-n}) \cdot e(t) = \\
 \hat{y}(t|t-k) &= (c_k \cdot z^{-k} + c_{k+1} \cdot z^{-k-1} + \dots + c_n \cdot z^{-n}) \cdot \frac{1}{(1 + c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n})} \cdot y(t) \\
 \hat{y}(t|t-k) &= \frac{c_k \cdot z^{-k} + c_{k+1} \cdot z^{-k-1} + \dots + c_n \cdot z^{-n}}{1 + c_1 \cdot z^{-1} + c_2 \cdot z^{-2} + \dots + c_n \cdot z^{-n}} \cdot y(t)
 \end{aligned}$$

da questo predittore a k passi si può notare che le radici di $C(z)$ devono essere all'interno del cerchio di raggio unitario altrimenti l'instabilità porterebbe la stima a divergere, per questo motivo il processo deve essere scritto in forma canonica.

13. Predittore ottimo di un processo $ARMA(m, n)$

Sia un processo $ARMA(m, n)$ in forma canonica:

$$y(t) = \frac{C(z)}{A(z)} \cdot e(t) \text{ con } e(t) \sim WN(0, \lambda^2)$$

con:

$$C(z) = 1 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}$$

$$A(z) = 1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2} + \dots + a_m \cdot z^{-m}$$

A differenza dei processi MA non è chiaro quale sia la parte nota e quella non nota dato che i termini autoregressivi dipendono dal rumore dei passi precedenti, quindi non si può ricavare un predittore in modo semplice come nel caso MA . Dalla teoria vista precedentemente, però, si sa che un processo $ARMA(m, n)$ può sempre essere visto come un $MA(\infty)$ e questo permetterebbe di utilizzare un predittore simile¹ a quello usato per gli $MA(n)$.

Purtroppo non è facile scriverlo come un $MA(\infty)$, ma si è visto che si può approssimare tramite la lunga divisione a un $MA(k)$ dove k è il numero di passi della divisione, ottenendo:

$$\frac{C(z)}{A(z)} = E(z) + \frac{R(z)}{A(z)}$$

$$C(z) = E(z) \cdot A(z) + R(z)$$

dove $E(z)$ è il quoziente e quindi è un polinomio con potenze negative fino $-(k-1)$ e $R(z)$ è il resto e quindi un polinomio con potenze negative $< -k$.²

Quindi il processo si può scrivere come:

$$y(t) = \frac{C(z)}{A(z)} \cdot e(t) =$$

$$= \left(E(z) + \frac{R(z)}{A(z)} \right) \cdot e(t) =$$

$$= \underbrace{E(z) \cdot e(t)}_{\text{dipendenza dal rumore per } k \text{ passi}} + \underbrace{\frac{R(z)}{A(z)} \cdot e(t)}_{\text{valori noti}}$$

Da quest'ultima espressione si nota che si è messo in evidenza la dipendenza dal rumore bianco fino al passo k , e quindi, analogamente al caso MA si ottiene:

$$\hat{y}(t|t-k) = \frac{R(z)}{A(z)} \cdot e(t)$$

dato che $R(z)$ può contenere solo potenze $\leq -k$ si può dire che:

$$R(z) = \tilde{R}(z) \cdot z^{-k}$$

¹non si può usare esattamente lo stesso perchè questo comporterebbe una somma infinita di termini

²Questa espressione è una equazione diofantea (o diofantina) nell'anello dei polinomi di \mathbb{C} , dove $E(z)$ e $R(z)$ sono le incognite mentre $C(z)$ è il termine noto e $A(z)$ e 1, polinomio neutro nell'anello dei polinomi in \mathbb{C} , sono i coefficienti. Questo dettaglio non è molto importante, ma questa equazione viene comunemente chiamata equazione diofantina.

13. Predittore ottimo di un processo ARMA(m, n)

si può riscrivere il predittore come:

$$\hat{y}(t|t-k) = \frac{\tilde{R}(z)}{A(z)} \cdot e(t-k)$$

In cui si nota che il predittore dipende dai valori dell'errore assunti fino a $t-k$ come ci si aspettava.

Da questo predittore si ottiene:

$$\varepsilon(t) = E(z) \cdot e(t)$$

che è un processo MA(k-1), e da cui si ottiene:

$$\begin{aligned} \mathbb{E}[\hat{y}(t|t-k) \cdot \varepsilon(t)] &= \mathbb{E}\left[\frac{\tilde{R}(z)}{A(z)} \cdot e(t-k) \cdot E(z) \cdot e(t)\right] = \\ &= \frac{\tilde{R}(z)}{A(z)} \cdot E(z) \cdot \mathbb{E}[e(t-k) \cdot e(t)] = \\ &= 0 \end{aligned}$$

e quindi è uno stimatore corretto, inoltre si può dimostrare che questo predittore è ottimo.

Si può notare che questo predittore richiede la conoscenza dei valori assunti dall'errore, ma si può risolvere con la stessa sostituzione fatta nel caso di processo MA:

$$\begin{aligned} y(t) &= \frac{C(z)}{A(z)} \cdot e(t) \\ e(t) &= \frac{A(z)}{C(z)} \cdot y(t) \end{aligned}$$

e quindi il predittore diventa:

$$\begin{aligned} \hat{y}(t|t-k) &= \frac{\tilde{R}(z)}{A(z)} \cdot e(t-k) = \\ &= \frac{\tilde{R}(z)}{\cancel{A(z)}} \cdot \frac{\cancel{A(z)}}{C(z)} \cdot y(t-k) = \\ &= \frac{\tilde{R}(z)}{C(z)} \cdot y(t-k) \end{aligned}$$

Da cui si capisce che il processo deve essere scritto in forma canonica in modo che le radici del polinomio $C(z)$ siano interne al cerchio di raggio unitario, in modo che la stima non diverga.

Esempio Ricavare il predittore ottimo a un passo e a 2 passi del processo:

$$y(t) = \frac{1 + \frac{1}{2} \cdot z^{-1}}{1 + \frac{1}{3} \cdot z^{-1}} \cdot e(t)$$

come prima cosa bisogna verificare che il processo sia rappresentato in forma canonica, ma questo è già in forma canonica. Successivamente si procede con la lunga divisione.

13. Predittore ottimo di un processo ARMA (m, n)

$$\begin{aligned}
 R_1(z) &= \begin{array}{c|c} \begin{array}{cc} 1 & +\frac{1}{2} \cdot z^{-1} \\ -1 & -\frac{1}{3} \cdot z^{-1} \end{array} & \begin{array}{c} 1 + \frac{1}{3} \cdot z^{-1} \\ 1 \end{array} \\ \hline \begin{array}{cc} 0 & +\frac{1}{6} \cdot z^{-1} \end{array} & 1 + \frac{1}{6} \cdot z^{-1} \\ \hline \begin{array}{cc} -\frac{1}{6} \cdot z^{-1} & -\frac{1}{18} \cdot z^{-2} \end{array} & \\ \hline \begin{array}{cc} 0 & -\frac{1}{18} \cdot z^{-2} \end{array} & \end{array} = E_1(z) \\
 \\
 R_2(z) &= \begin{array}{c|c} \begin{array}{cc} 1 & +\frac{1}{2} \cdot z^{-1} \\ -1 & -\frac{1}{3} \cdot z^{-1} \end{array} & \begin{array}{c} 1 + \frac{1}{3} \cdot z^{-1} \\ 1 \end{array} \\ \hline \begin{array}{cc} 0 & +\frac{1}{6} \cdot z^{-1} \end{array} & 1 + \frac{1}{6} \cdot z^{-1} \\ \hline \begin{array}{cc} -\frac{1}{6} \cdot z^{-1} & -\frac{1}{18} \cdot z^{-2} \end{array} & \\ \hline \begin{array}{cc} 0 & -\frac{1}{18} \cdot z^{-2} \end{array} & \end{array} = E_2(z)
 \end{aligned}$$

Quindi se prendiamo i dati dopo un passo, si ottiene:

$$E_1(z) = 1$$

$$R_1(z) = \frac{1}{6} \cdot z^{-1}$$

e quindi:

$$y(t) = \frac{1 + \frac{1}{2} \cdot z^{-1}}{1 + \frac{1}{3} \cdot z^{-1}} \cdot e(t) = \left(1 + \frac{\frac{1}{6} \cdot z^{-1}}{1 + \frac{1}{3} \cdot z^{-1}} \right) \cdot e(t)$$

e quindi si ottiene il predittore:

$$\begin{aligned}
 \hat{y}(t|t-1) &= \frac{R(z)}{C(z)} \cdot y(t) = \\
 &= \frac{\frac{1}{6} \cdot z^{-1}}{1 + \frac{1}{2} \cdot z^{-1}} \cdot y(t)
 \end{aligned}$$

da cui si ottiene:

$$\hat{y}(t|t-1) = -\frac{1}{2} \cdot \hat{y}(t-1|t-2) + \frac{1}{6} \cdot e(t-1)$$

Se invece si vuole il predittore a 2 passi si prende:

$$E_2(z) = 1 + \frac{1}{6} \cdot z^{-1}$$

$$R_2(z) = -\frac{1}{18} \cdot z^{-2}$$

e quindi:

$$y(t) = \frac{1 + \frac{1}{2} \cdot z^{-1}}{1 + \frac{1}{3} \cdot z^{-1}} \cdot e(t) = \left(1 + \frac{1}{6} \cdot z^{-1} + \frac{-\frac{1}{18} \cdot z^{-2}}{1 + \frac{1}{3} \cdot z^{-1}} \right) \cdot e(t)$$

13. Predittore ottimo di un processo ARMA (m, n)

e quindi si ottiene il predittore:

$$\begin{aligned}\hat{y}(t|t-1) &= \frac{R(z)}{C(z)} \cdot e(t) = \\ &= \frac{-\frac{1}{18} \cdot z^{-2}}{1 + \frac{1}{2} \cdot z^{-1}} \cdot e(t)\end{aligned}$$

da cui si ottiene:

$$\hat{y}(t|t-2) = -\frac{1}{2} \cdot \hat{y}(t-1|t-3) - \frac{1}{18} \cdot e(t-2)$$

13.1. Predittore a un passo ottimo di un processo ARMA (m, n)

Per ricavare il predittore a un passo si utilizza il procedimento spiegato per k passi, ipotizzando che il processo sia in forma canonica, si esegue con la lunga divisione:

$$\begin{array}{r|l} C(z) & A(z) \\ -A(z) & 1 \end{array} = E(z)$$

$$R(z) = C(z) - A(z)$$

Quindi il predittore è:

$$\begin{aligned}\hat{y}(t|t-1) &= \frac{R(z)}{A(z)} \cdot e(t) \\ &= \frac{C(z) - A(z)}{A(z)} \cdot e(t)\end{aligned}$$

$$\begin{aligned}\hat{y}(t|t-1) &= \frac{R(z)}{C(z)} \cdot y(t) = \\ &= \frac{C(z) - A(z)}{C(z)} \cdot y(t)\end{aligned}$$

e l'errore vale:

$$\varepsilon(t) = E(z) \cdot e(t) = 1 \cdot e(t) = e(t)$$

13.2. Valutazione del predittore

La qualità di una predizione è definita tramite la sua varianza, ma dato il valore della varianza è difficile dare una valutazione dello stimatore, per questo motivo si definisce l'error to signal ratio o ESR.

Questo valore mette a confronto la varianza del predittore con quella del predittore banale, ossia la media, in modo da dare un'idea immediata del predittore.

$$ESR(k) = \frac{Var[\varepsilon_k(t)]}{Var[y(t)]}$$

dove k è il numero di passi del predittore.

13. Predittore ottimo di un processo ARMA(m,n)

Esempio Si consideri il seguente processo:

$$y(t) = \frac{z+3}{2 \cdot z+1} \cdot e(t-1)$$

con $e(t) \sim WN(0,1)$

Si calcoli:

- $\gamma_y(0)$
- $\hat{y}(t|t-k)$ con $k=1$ e $k=2$
- $Var[\varepsilon(t)^2]$ di entrambi i predittori

Si potrebbe ricavare subito $\gamma_y(0)$, ma dato che per calcolare i predittori avremo bisogno dell'espressione del processo in forma canonica, è meglio ricavarla direttamente (non cambia nulla nel calcolo di $\gamma_y(0)$), quindi:

$$y(t) = \frac{z+3}{2 \cdot \left(z + \frac{1}{2}\right)} \cdot e(t-1)$$

usando un filtro passa-tutto:

$$\begin{aligned} y(t) &= \frac{\cancel{z+3}}{2 \cdot \left(z + \frac{1}{2}\right)} \cdot 3 \cdot \frac{z + \frac{1}{3}}{\cancel{z+3}} \cdot e(t-1) = \\ &= \frac{3}{2} \cdot \frac{z + \frac{1}{3}}{z + \frac{1}{2}} \cdot e(t-1) = \\ &= \frac{3}{2} \cdot \frac{1 + \frac{1}{3} \cdot z^{-1}}{1 + \frac{1}{2} \cdot z^{-1}} \cdot e(t-1) \end{aligned}$$

Introducendo $\eta(t) = \frac{3}{2} \cdot e(t-1)$, si ottiene:

$$y(t) = \frac{1 + \frac{1}{3} \cdot z^{-1}}{1 + \frac{1}{2} \cdot z^{-1}} \cdot \eta(t)$$

con:

$$\eta(t) \sim WN\left(0, \frac{9}{4}\right)$$

13. Predittore ottimo di un processo $ARMA(m, n)$

L'ultima espressione del processo è in forma canonica, quindi si può passare a calcolare $\gamma_y(0)$:

$$\begin{aligned}
 \gamma_y(0) &= \mathbb{E} \left[\left(-\frac{1}{2} \cdot y(t-1) + \eta(t) + \frac{1}{3} \cdot \eta(t-1) \right)^2 \right] = \\
 &= \frac{1}{4} \cdot \underbrace{\mathbb{E} [y(t-1)^2]}_{\gamma_y(0)} + \mathbb{E} [\eta(t)^2] + \frac{1}{9} \cdot \mathbb{E} [\eta(t-1)^2] + \\
 &\quad + \underbrace{\mathbb{E} [y(t-1) \cdot \eta(t)]}_0 + \frac{1}{3} \cdot \underbrace{\mathbb{E} [\eta(t) \cdot \eta(t-1)]}_0 - \frac{1}{3} \cdot \mathbb{E} [y(t-1) \cdot \eta(t-1)] \\
 \gamma_y(0) - \frac{1}{4} \cdot \gamma_y(0) &= \frac{9}{4} + \frac{1}{9} \cdot \frac{9}{4} - \frac{1}{3} \cdot \mathbb{E} \left[\left(-\frac{1}{2} \cdot y(t-2) + \eta(t-1) + \frac{1}{3} \cdot \eta(t-2) \right) \cdot \eta(t-1) \right] \\
 \frac{3}{4} \cdot \gamma_y(0) &= \frac{9}{4} + \frac{1}{4} - \frac{1}{3} \cdot \left(-\frac{1}{2} \cdot \underbrace{\mathbb{E} [y(t-2) \cdot \eta(t-1)]}_0 + \mathbb{E} [\eta(t-1)^2] + \frac{1}{3} \cdot \underbrace{\mathbb{E} [\eta(t-2) \cdot \eta(t-1)]}_0 \right) \\
 \gamma_y(0) &= \frac{4}{3} \cdot \left(\frac{10}{4} - \frac{1}{3} \cdot \frac{9}{4} \right) = \\
 &= \frac{7}{3}
 \end{aligned}$$

Ora si può calcolare il predittore ad un passo $\hat{y}(t|t-1)$, ricordando che il processo è un $ARMA(1, 1)$:

$$\begin{aligned}
 \hat{y}(t|t-1) &= \frac{C(z) - A(z)}{C(z)} \cdot y(t) = \\
 &= \frac{1 + \frac{1}{3} \cdot z^{-1} - 1 - \frac{1}{2} \cdot z^{-1}}{1 + \frac{1}{3} \cdot z^{-1}} \cdot y(t) = \\
 &= \frac{-\frac{1}{6} \cdot z^{-1}}{1 + \frac{1}{3} \cdot z^{-1}} \cdot y(t) = \\
 &= -\frac{1}{3} \cdot \hat{y}(t-1|t-2) - \frac{1}{6} \cdot y(t-1)
 \end{aligned}$$

E anche il predittore a due passi $\hat{y}(t|t-2)$:

13. Predittore ottimo di un processo $ARMA(m, n)$

$$R(z) = \frac{\begin{array}{c|c} 1 & \frac{1}{3} \cdot z^{-1} \\ \hline -1 & -\frac{1}{2} \cdot z^{-1} \end{array}}{\begin{array}{c|c} 0 & -\frac{1}{6} \cdot z^{-1} \\ \hline \frac{1}{6} \cdot z^{-1} & +\frac{1}{12} \cdot z^{-2} \end{array}} \cdot \frac{1 + \frac{1}{2} \cdot z^{-1}}{1 - \frac{1}{6} \cdot z^{-1}} = E(z)$$

$$\begin{aligned} \hat{y}(t|t-2) &= \frac{R(z)}{C(z)} \cdot y(t) = \\ &= \frac{\frac{1}{12} \cdot z^{-2}}{1 + \frac{1}{3} \cdot z^{-1}} \cdot y(t) = \\ &= -\frac{1}{3} \cdot \hat{y}(t-1|t-3) + \frac{1}{12} \cdot y(t-2) \end{aligned}$$

Ora è possibile calcolare le varianze degli errori di predizione:

$$\begin{aligned} Var[\varepsilon_1(t)] &= Var\left[\underbrace{E(z)}_1 \cdot \eta(t)\right] = \\ &= \frac{9}{4} \end{aligned}$$

$$\begin{aligned} Var[\varepsilon_2(t)] &= Var[E(z) \cdot \eta(t)] = \\ &= \mathbb{E}\left[\left(\eta(t) - \frac{1}{6} \cdot \eta(t-1)\right)^2\right] = \\ &= \mathbb{E}[\eta(t)^2] + \frac{1}{36} \cdot \mathbb{E}[\eta(t-1)^2] - \frac{1}{3} \cdot \mathbb{E}[\eta(t) \cdot \eta(t-1)] = \\ &= \frac{9}{4} + \frac{1}{16} = \\ &= \frac{37}{16} \end{aligned}$$

Da cui si vede che la varianza dell'errore di predizione è aumentata. Considerando il caso peggiore possibile, ovvero quando $\varepsilon(t) = y(t)$, si ha che

$$Var[\varepsilon(t)] = Var[y(t)] = \frac{7}{3}$$

13. Predittore ottimo di un processo ARMA (m, n)

In conclusione, un predittore non potrà mai avere varianza dell'errore di predizione maggiore della varianza del processo, e più aumentano i passi del predittore, più la varianza dell'errore si avvicinerà a quella del processo, come si vede in figura 13.1.

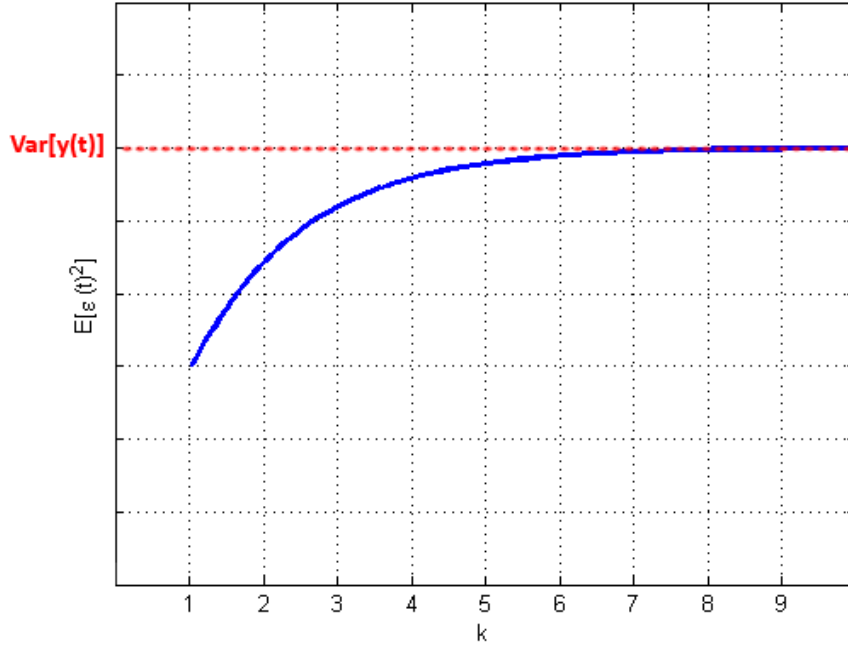


Figura 13.1.: Rappresentazione della varianza dell'errore di predizione in funzione del numero di passi del predittore

Esempio Si supponga di avere le seguenti misure $y(1) = 1, y(2) = \frac{1}{2}, y(3) = -\frac{1}{2}, y(4) = 0, y(5) = -\frac{1}{2}$ e si considerino i modelli:

$$y(t) = \frac{1}{2} \cdot y(t-1) + \frac{1}{4} \cdot y(t-2) + e(t) \quad \text{Modello 1}$$

$$y(t) = e(t) + \frac{1}{2} \cdot e(t-1) \quad \text{Modello 2}$$

Si può notare che il primo modello rappresenta un $AR(2)$, mentre il secondo un $MA(1)$. Ricaviamo ora i predittori ad un passo per entrambi i modelli:

- Modello 1:

$$y(t) = \frac{1}{1 - \frac{1}{2} \cdot z^{-1} - \frac{1}{4} \cdot z^{-2}} \cdot e_1(t)$$

considerandolo come un $ARMA(2, 0)$ si ha che il predittore ad un passo sarà:

$$\begin{aligned} \hat{y}_1(t|t-1) &= \frac{C(z) - A(z)}{C(z)} \cdot y(t) = \\ &= \frac{1 - \left(1 - \frac{1}{2} \cdot z^{-1} - \frac{1}{4} \cdot z^{-2}\right)}{1} \cdot y(t) = \\ &= \frac{1}{2} \cdot y(t-1) + \frac{1}{4} \cdot y(t-2) \end{aligned}$$

13. Predittore ottimo di un processo $ARMA(m, n)$

- Modello 2:

$$y(t) = \frac{1 + \frac{1}{2} \cdot z^{-1}}{1} \cdot e(t)$$

considerandolo come un $ARMA(0, 1)$ si ha che il predittore ad un passo sarà:

$$\begin{aligned}\hat{y}_2(t|t-1) &= \frac{C(z) - A(z)}{C(z)} \cdot y(t) = \\ &= \frac{1 + \frac{1}{2} \cdot z^{-1} - 1}{1 + \frac{1}{2} \cdot z^{-1}} \cdot y(t) = \\ &= -\frac{1}{2} \cdot \hat{y}_2(t-1|t-2) + \frac{1}{2} \cdot y(t-1)\end{aligned}$$

Si possono calcolare ora i $\hat{y}(6|5)$ di entrambi i modelli:

- Modello 1:

$$\begin{aligned}\hat{y}_1(6|5) &= \frac{1}{2} \cdot y(5) + \frac{1}{4} \cdot y(4) = \\ &= \frac{1}{2} \cdot \left(-\frac{1}{2}\right) + \frac{1}{4} \cdot 0 = \\ &= -\frac{1}{4}\end{aligned}$$

- Modello 2:

$$\hat{y}_2(6|5) = -\frac{1}{2} \cdot \hat{y}_2(5|4) + \frac{1}{2} \cdot y(5)$$

Si può capire che la definizione di questo predittore è ricorsiva; quindi, come discusso nel capitolo 12, si inizializza $\hat{y}(1|0)$ al valore della media campionaria, che in questo particolare caso vale $\frac{1}{10}$

$$\begin{aligned}\hat{y}_2(1|0) &= \frac{1}{10} \\ \hat{y}_2(2|1) &= -\frac{1}{2} \cdot \hat{y}_2(1|0) + \frac{1}{2} \cdot y(1) = -\frac{1}{2} \cdot \frac{1}{10} + \frac{1}{2} \cdot 1 = \frac{9}{20} \\ \hat{y}_2(3|2) &= -\frac{1}{2} \cdot \hat{y}_2(2|1) + \frac{1}{2} \cdot y(2) = -\frac{1}{2} \cdot \frac{9}{20} + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{40} \\ \hat{y}_2(4|3) &= -\frac{1}{2} \cdot \hat{y}_2(3|2) + \frac{1}{2} \cdot y(3) = -\frac{1}{2} \cdot \frac{1}{40} + \frac{1}{2} \cdot \left(-\frac{1}{2}\right) = -\frac{21}{80} \\ \hat{y}_2(5|4) &= -\frac{1}{2} \cdot \hat{y}_2(4|3) + \frac{1}{2} \cdot y(4) = -\frac{1}{2} \cdot \left(-\frac{21}{80}\right) + \frac{1}{2} \cdot 0 = \frac{21}{160}\end{aligned}$$

e infine:

$$\hat{y}_2(6|5) = -\frac{1}{2} \cdot \hat{y}_2(5|4) + \frac{1}{2} \cdot y(5) = -\frac{1}{2} \cdot \frac{21}{160} + \frac{1}{2} \cdot \left(-\frac{1}{2}\right) = -\frac{101}{320}$$

14. Predittore ottimo di un processo $ARMAX(m, n, p)$

Sia $y(t)$ un processo $ARMAX(m, n, p)$:

$$y(t) = \frac{B(z)}{A(z)} \cdot u(t-k) + \frac{C(z)}{A(z)} \cdot e(t)$$

con $e(t) \sim WN(0, \lambda^2)$

Si consideri come ipotesi che $\frac{C(z)}{A(z)}$ sia in forma canonica; non si possono invece fare ipotesi su $\frac{B(z)}{A(z)}$, in quanto è il modello del sistema sotto analisi, e non si può canonizzare.

Osservazione In un processo $ARMAX(m, n, p)$ l'ingresso esogeno influenza il valore dell'uscita solo dopo k passi, quindi se si calcola un predittore a h passi con $h < k$ la variabile esogena non influisce sul valore di y e quindi neppure su quello predittore. In questo caso particolare la variabile esogena è come se venisse ignorata e quindi il predittore corrisponde a quello della parte stocastica. Per questo motivo è solito calcolare predittori solo con $h \geq k$ passi e in particolare il predittore a k passi.

Predittore a k passi Si vuole ricavare il predittore a k passi, per ottenerlo si scompone il processo in una parte predicibile e in una non predicibile, come fatto per i precedenti predittori. Quindi si eseguono k passi di lunga divisione ottenendo:

$$\frac{C(z)}{A(z)} = E(z) + \frac{z^{-k} \cdot \tilde{R}(z)}{A(z)}$$

Dove $E(z)$ è il quoziente della lunga divisione e $z^{-k} \cdot \tilde{R}(z)$ è il resto. Sostituendo nell'equazione del processo:

$$y(t) = \underbrace{\frac{B(z)}{A(z)} \cdot u(t-k) + \frac{\tilde{R}(z)}{A(z)} \cdot e(t-k)}_{\text{Componente predicibile}} + \underbrace{E(z) \cdot e(t)}_{\text{Componente non predicibile}}$$

Si può notare infatti che la componente predicibile è data dalla combinazione lineare dell'ingresso esogeno e del rumore fino al tempo $t-k$, mentre la parte non predicibile è una combinazione del rumore dal tempo $t-k+1$ fino al tempo t (basti pensare che $E(z)$ è un $MA(k-1)$, quindi $E(z) \cdot e(t) = c_0 \cdot e(t) + c_1 \cdot e(t-1) + \dots + c_{k-1} \cdot e(t-k+1)$)

Riconosciute le componenti predicibili e non, si ha quindi:

$$\hat{y}(t|t-k) = \frac{B(z)}{A(z)} \cdot u(t-k) + \frac{\tilde{R}(z)}{A(z)} \cdot e(t-k)$$

Analisi predittore Ricavando l'errore si ottiene:

$$\begin{aligned} \varepsilon(t) &= y(t) - \hat{y}(t|t-k) = \\ &= \frac{B(z)}{A(z)} \cdot u(t-k) + \frac{\tilde{R}(z)}{A(z)} \cdot e(t-k) + E(z) \cdot e(t) - \left(\frac{B(z)}{A(z)} \cdot u(t-k) + \frac{\tilde{R}(z)}{A(z)} \cdot e(t-k) \right) = \\ &= \frac{B(z)}{A(z)} \cdot \cancel{u(t-k)} + \frac{\tilde{R}(z)}{A(z)} \cdot \cancel{e(t-k)} + E(z) \cdot e(t) - \frac{B(z)}{A(z)} \cdot \cancel{u(t-k)} - \frac{\tilde{R}(z)}{A(z)} \cdot \cancel{e(t-k)} = \\ &= E(z) \cdot e(t) \end{aligned}$$

14. Predittore ottimo di un processo $ARMAX(m, n, p)$

Quindi, analogamente al caso $ARMA$, è un processo $MA(k-1)$ e quindi, dato che il valore atteso dell'errore è nullo, ha valore atteso nullo e quindi lo stimatore è corretto.

Come al solito si vuole trovare il predittore ottimo:

$$\mathbb{E}[\hat{y}(t|t-k) \cdot \varepsilon(t)] = 0$$

In quanto si avranno solo moltiplicazioni tra l'errore al tempo $\tau \geq t-k+1$ e l'ingresso esogeno ai tempi $\tau \leq t-k$ (scorrelati tra di loro) e moltiplicazioni tra l'errore al tempo $\tau \geq t-k+1$ e l'errore al tempo $\tau \leq t-k$ (scorrelato in quanto rumore bianco). Si ha quindi che il predittore soddisfa la condizione necessaria per l'ottimalità.

Si ha inoltre che $Var[\varepsilon(t)]$ è minima, quindi il predittore è ottimo.

Al solito, non è possibile utilizzare questo stimatore in quanto dipende dal rumore; è possibile però ricavare l'espressione del rumore in funzione dei dati del processo e dell'ingresso esogeno:

$$\begin{aligned} e(t) &= \frac{A(z)}{C(z)} \cdot \left(y(t) - \frac{B(z)}{A(z)} \cdot u(t-1) \right) = \\ &= \frac{A(z)}{C(z)} \cdot y(t) - \frac{\cancel{A(z)}}{C(z)} \cdot \frac{B(z)}{\cancel{A(z)}} \cdot u(t-1) = \\ &= \frac{A(z)}{C(z)} \cdot y(t) - \frac{B(z)}{C(z)} \cdot u(t-1) \end{aligned}$$

Da questo punto si può sostituire questa espressione all'interno dell'equazione del predittore (si noti che il ritardo z^{-k} , che prima era all'interno del rumore, è stato riassorbito nel polinomio $\tilde{R}(z)$):

$$\begin{aligned} \hat{y}(t|t-k) &= \frac{B(z)}{A(z)} \cdot u(t-k) + \frac{R(z)}{A(z)} \cdot \left(\frac{A(z)}{C(z)} \cdot y(t) - \frac{B(z)}{C(z)} \cdot u(t-k) \right) = \\ &= \frac{R(z)}{C(z)} \cdot y(t) + \left(\frac{B(z) \cdot (C(z) - R(z))}{A(z) \cdot C(z)} \right) \cdot u(t-k) \end{aligned}$$

Dall'equazione diofantea è possibile notare che:

$$\begin{aligned} C(z) &= A(z) \cdot E(z) + R(z) \\ C(z) - R(z) &= A(z) \cdot E(z) \end{aligned}$$

E sostituendo nell'equazione del predittore:

$$\begin{aligned} \hat{y}(t|t-k) &= \frac{R(z)}{C(z)} \cdot y(t) + \left(\frac{B(z) \cdot \cancel{A(z)} \cdot E(z)}{\cancel{A(z)} \cdot C(z)} \right) \cdot u(t-k) = \\ &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) \end{aligned}$$

Osservazioni

- Il predittore di un processo $ARMAX(m, n, p)$ ha l'errore uguale a quello del predittore della parte stocastica, questo perché la parte deterministica è completamente predicibile e si conosce precisamente come l'ingresso viene riportato sull'uscita e non crea ulteriore errore.
- Data l'osservazione precedente, l'indice ESR si può calcolare basandosi solo sulla parte $ARMA$ e quindi:

$$ESR(k) = \frac{Var[\varepsilon]}{Var\left[\frac{C(z)}{A(z)} \cdot e(t)\right]}$$

14.1. Predittore a un passo ottimo di un processo $ARMAX(m, n, 1)$

Per ottenere il predittore a un passo bisogna eseguire un passo di lunga divisione ottenendo:

$$E(z) = 1$$

$$R(z) = C(z) - A(z)$$

L'espressione del predittore ad un passo sarà quindi:

$$\hat{y}(t|t-1) = \frac{C(z) - A(z)}{C(z)} \cdot y(t) + \frac{B(z)}{C(z)} \cdot u(t-k)$$

$$\varepsilon(t) = E(z) \cdot e(t) = e(t)$$

Esempio Dato:

$$y(t) = (2 + 6 \cdot z^{-1}) \cdot u(t-2) + \frac{2}{3 + \frac{1}{2} \cdot z^{-1}} \cdot \eta(t-1)$$

con $\eta(t) \sim WN(0, 1)$

Trovare il predittore e $Var[\varepsilon(t)]$

Si procede controllando che la parte $ARMA$ sia stazionaria; dato che il sistema ha un polo in $-\frac{1}{2}$, è stazionario. Ora si può procedere a trasformare il processo in forma canonica (solo per la parte $ARMA$, in quanto la restante parte è una rappresentazione del sistema e non può essere modificata).

Rinominando la parte $ARMA$ in $v(t)$:

$$v(t) = \frac{2}{3} \cdot \frac{1}{1 + \frac{1}{2} \cdot z^{-1}} \cdot \eta(t-1)$$

$$\frac{2}{3} \cdot \eta(t-1) = e(t)$$

con $e(t) \sim WN\left(0, \frac{4}{9}\right)$

Quindi:

$$y(t) = (2 + 6 \cdot z^{-1}) \cdot u(t-2) + \frac{1}{1 + \frac{1}{2} \cdot z^{-1}} \cdot e(t)$$

Dato che il ritardo puro tra ingresso e uscita è di $k = 2$, si calcola il predittore a due passi, ricordando che:

$$C(z) = 1$$

$$A(z) = 1 + \frac{1}{2} \cdot z^{-1}$$

$$\frac{B(z)}{A(z)} = 2 + 6 \cdot z^{-1}$$

$$B(z) = 2 \cdot (1 + 3z^{-1}) \cdot \left(1 + \frac{1}{2} \cdot z^{-1}\right)$$

14. Predittore ottimo di un processo ARMAX (m, n, p)

$$R(z) = \begin{array}{c|c} \begin{array}{cc} 1 & \\ -1 & -\frac{1}{2} \cdot z^{-1} \end{array} & \begin{array}{c} 1 + \frac{1}{2} \cdot z^{-1} \\ 1 - \frac{1}{2} \cdot z^{-1} \end{array} \\ \hline \begin{array}{cc} 0 & -\frac{1}{2} \cdot z^{-1} \\ \frac{1}{2} \cdot z^{-1} & +\frac{1}{4} \cdot z^{-2} \end{array} & \\ \hline \begin{array}{cc} 0 & +\frac{1}{4} \cdot z^{-2} \end{array} & \end{array} = E(z)$$

Il predittore sarà quindi:

$$\begin{aligned} \hat{y}(t|t-2) &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(t)}{C(z)} \cdot u(t-2) = \\ &= \frac{1}{4} \cdot y(t-2) + \frac{2 \cdot (1 + 3z^{-1}) \cdot \left(1 + \frac{1}{2} \cdot z^{-1}\right) \cdot \left(1 - \frac{1}{2} \cdot z^{-1}\right)}{1} \cdot u(t-2) = \\ &= \frac{1}{4} \cdot y(t-2) + 2 \cdot (1 + 3z^{-1}) \cdot \left(1 - \frac{1}{4} \cdot z^{-2}\right) \cdot u(t-2) = \\ &= \frac{1}{4} \cdot y(t-2) + 2 + 6 \cdot z^{-1} - \frac{1}{2} \cdot z^{-2} - \frac{3}{2} \cdot z^{-3} \\ &= \frac{1}{4} \cdot y(t-2) + 2 \cdot u(t-2) + 6 \cdot u(t-3) - \frac{1}{2} \cdot u(t-4) - \frac{3}{2} \cdot u(t-5) \end{aligned}$$

e la varianza dell'errore sarà:

$$\begin{aligned} Var[\varepsilon(t)] &= Var[E(z) \cdot e(t)] = \\ &= \mathbb{E} \left[\left(e(t) - \frac{1}{2} \cdot e(t-1) \right)^2 \right] = \\ &= \mathbb{E} [e(t)^2] + \frac{1}{4} \cdot \mathbb{E} [e(t-1)^2] + \mathbb{E} [e(t) \cdot e(t-1)] = \\ &= \frac{4}{9} + \frac{1}{4} \cdot \frac{4}{9} + 0 = \\ &= \frac{5}{9} \end{aligned}$$

A questo punto si potrebbe calcolare l'ESR di questo predittore, per valutarne la bontà:

$$ESR(2) = \frac{Var[\varepsilon(t)]}{Var \left[\frac{C(z)}{A(z)} \cdot e(t) \right]} = \frac{Var[\varepsilon(t)]}{Var[v(t)]}$$

dove la varianza della parte stocastica è:

14. Predittore ottimo di un processo $ARMAX(m, n, p)$

$$\begin{aligned}
 Var[v(t)] &= \mathbb{E} \left[\left(-\frac{1}{2} \cdot y(t-1) + e(t) \right)^2 \right] \\
 Var[v(t)] &= \frac{1}{4} \cdot Var[v(t)] + \mathbb{E} [e(t)^2] - \mathbb{E} [y(t-1) \cdot e(t)] \\
 \frac{3}{4} \cdot Var[v(t)] &= \frac{9}{4} \\
 Var[v(t)] &= 3
 \end{aligned}$$

e quindi:

$$ESR(2) = \frac{Var[\varepsilon(t)]}{Var[v(t)]} = \frac{\frac{5}{9}}{3} = \frac{5}{18}$$

15. Predittore $ARMAX(n, m, p)$ a media non nulla

Tutti i predittori visti fino a ora avevano come ipotesi che il rumore bianco avesse media nulla, questo garantiva che il predittore fosse corretto, infatti:

$$\varepsilon(t) = E(z) \cdot e(t)$$

che è un preprocesso $MA(k-1)$ dove k è il numero di passi del predittore quindi:

$$\mathbb{E}[\varepsilon(t)] = 0 \leftrightarrow \mathbb{E}[e(t)] = 0$$

questo vuol dire che se il rumore bianco è a media non nulla il predittore usato non ha errore con media nulla e quindi non è corretto. Per risolvere il problema esistono due metodi.

15.1. Depolarizzazione

Questo metodo si basa sulla depolarizzazione, spiegata nella sezione 3.3. Quindi si definisce un processo ausiliario \tilde{y} con media nulla:

$$\tilde{y}(t) = y(t) - m_y$$

di cui si può calcolare lo stimatore corretto $\hat{\tilde{y}}(t|t-k)$ coi metodi visti in precedenza e quindi:

$$\begin{aligned}\hat{\tilde{y}}(t|t-k) &= \frac{R(z)}{C(z)} \cdot \tilde{y}(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) \\ \hat{\tilde{y}}(t|t-k) &= \frac{R(z)}{C(z)} \cdot (y(t) - m_y) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) \\ \hat{\tilde{y}}(t|t-k) &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) - \frac{R(z)}{C(z)} \cdot m_y\end{aligned}$$

Dato che il processo $\tilde{y}(t)$ è traslato di un valore costante rispetto al processo $y(t)$ anche i rispettivi predittori corretti saranno traslati dello stesso valore, e quindi:

$$\hat{\tilde{y}}(t|t-k) = \hat{y}(t|t-k) - m_y$$

sostituendo si ottiene:

$$\begin{aligned}\hat{\tilde{y}}(t|t-k) &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) - \frac{R(z)}{C(z)} \cdot m_y \\ \hat{y}(t|t-k) - m_y &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) - \frac{R(z)}{C(z)} \cdot m_y \\ \hat{y}(t|t-k) &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) - \frac{R(z)}{C(z)} \cdot m_y + m_y \\ \hat{y}(t|t-k) &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) + \left(1 - \frac{R(z)}{C(z)}\right) \cdot m_y\end{aligned}$$

15. Predittore ARMAX (n, m, p) a media non nulla

Quindi il predittore è uguale a quello che si troverebbe se il rumore avesse media nulla con l'aggiunta di un addendo che dipende dalla media del processo. Per il teorema della risposta in frequenza il termine che si aggiunge al predittore è una costante; infatti m_y , essendo costante, può essere vista come una sinusoidale di ampiezza m_y e di frequenza nulla e quindi:

$$\begin{aligned}\hat{y}(t|t-k) &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) + \left| 1 - \frac{R(e^{j0})}{C(e^{j0})} \right| \cdot m_y = \\ &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) + \left| 1 - \frac{R(1)}{C(1)} \right| \cdot m_y\end{aligned}$$

Lo stimatore così ricavato è corretto infatti:

$$\begin{aligned}\mathbb{E}[\varepsilon(t)] &= \mathbb{E}[\hat{y}(t|t-k) - y(t)] = \\ &= \mathbb{E}[\hat{\tilde{y}}(t|t-k) + m_y - y(t)] = \\ &= \mathbb{E}[\hat{\tilde{y}}(t|t-k)] + \mathbb{E}[m_y] - \mathbb{E}[y(t)] = \\ &= 0 + m_y - m_y = \\ &= 0\end{aligned}$$

e la varianza dell'errore è:

$$\begin{aligned}\text{Var}[\varepsilon(t)] &= \text{Var}[\hat{y}(t|t-k) - y(t)] = \\ &= \text{Var}[\hat{\tilde{y}}(t|t-k) + m_y - y(t)] = \\ &= \text{Var}[\hat{\tilde{y}}(t|t-k) - y(t)] = \\ &= \text{Var}[\tilde{\varepsilon}(t)]\end{aligned}$$

e dato che il predittore $\hat{\tilde{y}}(t|t-k)$ è ottimo allora lo è anche $\hat{y}(t|t-k)$.

Esempio Dato:

$$\begin{aligned}y(t) &= e(t) + 4 \cdot e(t-1) = (1 + 4 \cdot z^{-1}) \cdot e(t) \\ \text{con} \quad e(t) &\sim WN(1, 1)\end{aligned}$$

Calcolare il predittore a uno e due passi e le rispettive varianze dell'errore di predizione.

Canonizzazione Per prima cosa si può portare il processo in forma canonica attraverso un filtro passa tutto:

$$\begin{aligned}y(t) &= \cancel{(1 + 4 \cdot z^{-1})} \cdot 4 \cdot \frac{1 + \frac{1}{4} \cdot z^{-1}}{\cancel{1 + 4 \cdot z^{-1}}} \cdot e(t) = \\ &= \left(1 + \frac{1}{4} \cdot z^{-1}\right) \cdot \eta(t) \\ \text{con} \quad \eta(t) &= 4 \cdot e(t)\end{aligned}$$

dove:

$$\begin{aligned}\mathbb{E}[\eta(t)] &= \mathbb{E}[4 \cdot e(t)] = 4 \cdot \mathbb{E}[e(t)] = 4 \cdot 1 = 4 \\ \text{Var}[\eta(t)] &= \text{Var}[4 \cdot e(t)] = 16 \cdot \text{Var}[e(t)] = 16 \cdot 1 = 16\end{aligned}$$

quindi $\eta(t) \sim WN(4, 16)$.

15. Predittore ARMAX (n, m, p) a media non nulla

Depolarizzazione Quindi si procede con la depolarizzazione:

$$\begin{cases} \tilde{y}(t) = y(t) - 5 \\ \tilde{\eta}(t) = \eta(t) - 4 \end{cases}$$

Sostituendo nell'espressione del processo

$$\begin{aligned} y(t) &= \left(1 + \frac{1}{4} \cdot z^{-1}\right) \cdot \eta(t) \\ \tilde{y}(t) - 5 &= \tilde{\eta}(t) - 4 + \frac{1}{4} \cdot \tilde{\eta}(t-1) - 1 \\ \tilde{y}(t) &= \tilde{\eta}(t) + \frac{1}{4} \cdot \tilde{\eta}(t-1) \end{aligned}$$

Predittore a un passo Applicando la formula per il predittore ad un passo di un processo ARMA:

$$\begin{aligned} \hat{\tilde{y}}(t|t-1) &= \frac{C(z) - A(z)}{C(z)} \cdot \tilde{y}(t) = \\ &= \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot \tilde{y}(t-1) \end{aligned}$$

$$\tilde{\varepsilon}(t) = 1 \cdot \tilde{\eta}(t) = \tilde{\eta}(t)$$

Calcolando il predittore $\hat{y}(t|t-1)$, si ottiene:

$$\begin{aligned} \hat{\tilde{y}}(t|t-1) - 5 &= \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot (y(t-1) - 5) \\ \hat{y}(t|t-1) &= \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t-1) + \left(1 - \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}}\right) \cdot 5 = \\ \hat{y}(t|t-1) &= \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t-1) + \left(1 - \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot e^{-j \cdot 0}}\right) \cdot 5 = \\ \hat{y}(t|t-1) &= \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t-1) + \frac{1 + \cancel{\frac{1}{4}} - \cancel{\frac{1}{4}}}{1 + \frac{1}{4}} \cdot 5 = \\ \hat{y}(t|t-1) &= \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t-1) + \frac{4}{5} \cdot 5 = \\ \hat{y}(t|t-1) &= \frac{\frac{1}{4}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t-1) + 4 \end{aligned}$$

15. Predittore ARMAX (n, m, p) a media non nulla

Come visto prima la varianza dell'errore del predittore $\hat{\tilde{y}}(t|t-1)$ è la stessa dell'errore del predittore $\hat{y}(t|t-1)$, quindi:

$$Var[\varepsilon(t)] = Var[\tilde{\varepsilon}(t)] = Var[\tilde{\eta}(t)] = 16$$

Predittore a due passi Il processo è un $MA(1)$, quindi il processo è scorrellato a due passi e quindi è come predire un rumore bianco e quindi il miglior stimatore è il valore atteso del processo e quindi il predittore banale, ma si può raggiungere la stessa conclusione eseguendo il normale procedimento. Quindi si esegue la lunga divisione a due passi sul processo polarizzato.

$$R(z) = \begin{array}{r|l} 1 & +\frac{1}{4} \cdot z^{-1} \\ \hline -1 & \\ \hline 0 & +\frac{1}{4} \cdot z^{-1} \\ & \hline & -\frac{1}{4} \cdot z^{-1} \\ & \hline & 0 \end{array} \quad \begin{array}{l} 1 \\ \\ 1 + \frac{1}{4} \cdot z^{-1} \\ = E(z) \end{array}$$

Quindi il predittore è:

$$\hat{\tilde{y}}(t|t-2) = \frac{R(z)}{C(z)} \cdot \tilde{y}(t) = \frac{0}{1 + \frac{1}{4} \cdot z^{-1}} \cdot \tilde{y}(t) = 0$$

$$\tilde{\varepsilon}(t) = \left(1 + \frac{1}{4} \cdot z^{-1}\right) \cdot \tilde{\eta}(t) = \tilde{y}(t)$$

Calcolando il predittore $\hat{y}(t|t-2)$, si ottiene:

$$\begin{aligned} \hat{y}(t|t-2) - 5 &= 0 \\ \hat{y}(t|t-2) &= 5 \end{aligned}$$

Come visto prima la varianza dell'errore del predittore $\hat{\tilde{y}}(t|t-2)$ è la stessa dell'errore del predittore $\hat{y}(t|t-2)$, quindi:

$$Var[\varepsilon(t)] = Var[\tilde{\varepsilon}(t)] = Var[\tilde{y}(t)]$$

15.2. Variabile esogena

Dato un processo $ARMA$, con rumore bianco a media μ si può definire un rumore bianco a media nulla $\tilde{e}(t)$:

$$\tilde{e}(t) = e(t) - \mu$$

15. Predittore $ARMAX(n, m, p)$ a media non nulla

con:

$$\begin{aligned}\mathbb{E}[\tilde{e}(t)] &= \mathbb{E}[e(t) - \mu] = \mu - \mu = 0 \\ \text{Var}[\tilde{e}(t)] &= \text{Var}[e(t) - \mu] = \text{Var}[e(t)]\end{aligned}$$

Sostituendo nel processo si ottiene:

$$\begin{aligned}y(t) &= \frac{C(z)}{A(z)} \cdot e(t) = \\ &= \frac{C(z)}{A(z)} \cdot (\tilde{e}(t) - \mu) = \\ &= \frac{C(z)}{A(z)} \cdot \tilde{e}(t) - \frac{C(z)}{A(z)} \cdot \mu\end{aligned}$$

Quindi si può intendere il processo $y(t)$ come un $ARMAX$ in cui la variabile esogena è $u(t) = \mu, \forall t$. ottenendo un predittore:

$$\begin{aligned}y(t|t-k) &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{B(z) \cdot E(z)}{C(z)} \cdot u(t-k) = \\ &= \frac{R(z)}{C(z)} \cdot y(t) + \frac{\cancel{C(z)} \cdot E(z)}{\cancel{C(z)}} \cdot \mu = \\ &= \frac{R(z)}{C(z)} \cdot y(t) + E(z) \cdot \mu\end{aligned}$$

per il teorema della risposta in frequenza, considerando μ come una sinusoide con ampiezza μ e frequenza nulla, si ottiene:

$$y(t|t-k) = \frac{R(z)}{C(z)} \cdot y(t) + |E(1)| \cdot \mu$$

Questo predittore è usato usando il predittore $ARMAX$ e quindi ne eredita la correttezza e ottimalità.

Esempio Ricordando il processo usato nell'esempio precedente in forma canonica:

$$\begin{aligned}y(t) &= \left(1 + \frac{1}{4} \cdot z^{-1}\right) \cdot \eta(t) = \eta(t) + \frac{1}{4} \cdot \eta(t-1) \\ \text{con} \quad e(t) &\sim WN(4, 16)\end{aligned}$$

sostituendo il rumore $\tilde{\eta}(t) = \eta(t) - 4$ si ottiene:

$$\begin{aligned}y(t) &= \eta(t) + \frac{1}{4} \cdot \eta(t-1) = \\ &= (\tilde{\eta}(t) + 4) + \frac{1}{4} \cdot (\tilde{\eta}(t-1) + 4) = \\ &= \tilde{\eta}(t) + \frac{1}{4} \cdot \tilde{\eta}(t-1) + 4 + 1 = \\ &= \tilde{\eta}(t) + \frac{1}{4} \cdot \tilde{\eta}(t-1) + 5 = \\ &= \frac{1}{1} \cdot 5 + \frac{1 + \frac{1}{4} \cdot z^{-1}}{1} \cdot \tilde{\eta}(t)\end{aligned}$$

15. Predittore $ARMAX(n, m, p)$ a media non nulla

Considerando la variabile esogena $u(t) = 5$, si ottiene:

$$A(z) = 1$$

$$B(z) = 1$$

$$C(z) = 1 + \frac{1}{4} \cdot z^{-1}$$

Quindi il predittore $ARMAX$, considerando $u(t-1) = 5$, è:

$$\begin{aligned} y(t|t-1) &= \frac{C(z) - A(z)}{C(z)} \cdot y(t) + \frac{B(z)}{C(z)} \cdot u(t-1) = \\ &= \frac{1 + \frac{1}{4} \cdot z^{-1} - 1}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t) + \frac{1}{1 + \frac{1}{4} \cdot z^{-1}} \cdot 5 = \\ &= \frac{\frac{1}{4} \cdot z^{-1}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t) + \frac{1}{1 + \frac{1}{4}} \cdot 5 = \\ &= \frac{\frac{1}{4} \cdot z^{-1}}{1 + \frac{1}{4} \cdot z^{-1}} \cdot y(t) + 4 \end{aligned}$$

Osservazione

- Il predittore ricavato con questo metodo è lo stesso di quello trovato tramite la depolarizzazione, come ci si aspettava dato che si è dimostrato che sono entrambi ottimi.
- Il procedimento usato per risolvere questo esempio è leggermete diverso da quello usato nella parte teorica perchè in quest'ultima si è usata la notazione con l'operatore ritardo mentre nell'esempio si è rimasti nella notazione "normale". I due procedimenti sono equivalenti e portano alla stessa soluzione.

16. Teoria di Kolmogorov-Wiener

I predittori visti in questa sezione sono definiti predittori di Kolmogorov-Wiener e funzionano solo in determinate situazioni:

- Il sistema deve essere a regime, infatti questi predittori non sono ottimi durante il transitorio del sistema;
- Il sistema deve essere lineare a tempo invariante;
- Il sistema deve essere stabile;
- È in grado di prevedere solo il valore dell'uscita del sistema e non delle sue variabili di stato;

Queste limitazioni non sono importanti nella identificazione tramite modelli *ARMAX*, ma in molte applicazioni sono dei grossi limiti, per questo motivo esistono teorie sulla predizioni più approfondite che superano questi limiti come il predittore di Kalman, che verrà illustrato più avanti.

Parte IV.

Identificazione modelli

Introduzione

Tutte le analisi viste fino a questo punto si basano tutte su un punto cardine: la conoscenza del modello del sistema. La disciplina dell'identificazione volge lo sguardo ai problemi in cui il modello del sistema risulta sconosciuto, e va quindi stimato in qualche modo.

Ad esempio, dato il sistema raffigurato in figura 16.1 risolvere un problema di identificazione significa, partendo da dati sperimentali di ingresso u_1, u_2, \dots, u_n e uscita y_1, y_2, \dots, y_n dello stesso ricavare il modello G del sistema. Quindi si dispone di un set di dati $D = \{u_1, u_2, \dots, u_n, y_1, y_2, \dots, y_n\}$ e si vuole ricavare il sistema G che li ha generati.



Figura 16.1.: Rappresentazione del problema di identificazione

I principali passi per affrontare un problema di identificazione sono:

1. Raccolta dati sperimentali

Questa parte è fondamentale: bisogna raccogliere dei dati che contengano più informazione possibile, in modo che la stima del modello sia la migliore. In particolare bisogna scegliere:

- Il tipo di ingresso $u(t)$, bisogna trovare l'ingresso che massimizza il più possibile l'informazione contenuta nei dati.
- Quanti dati prendere, dato che non se ne possono prendere infiniti bisogna sapere quanti dati bastano per poter avere una buona stima.

Questi punti non sono banali: esistono addirittura intere discipline (DOE : Design Of Experiment/Input design) che studiano questi problemi.

2. Scelta della famiglia di modelli

Bisogna decidere che tipo di modello $M(\vartheta)$ usare, ogni modello dipende da un set ϑ di parametri che saranno da stimare. Esistono diversi tipi di modelli che differiscono per vari fattori:

- Tempo discreto o tempo continuo.
- Sistema lineare o non lineare.
- Tempo invariante o tempo variante.
- Dinamico o statico.

In questa trattazione si parlerà solo di identificazione di modelli a tempo discreto, lineari, a tempo invarianti e dinamici ossia la famiglia dei processi *ARMAX*. Una volta scelto di voler usare un modello ARMAX bisogna comunque decidere gli ordini.

3. Scelta della cifra di merito

la cifra di merito è una funzione¹ $J_N(\vartheta) : \mathbb{R}^{\vartheta_n} \rightarrow \mathbb{R}^+$ che definisce la qualità della stima ϑ . Solitamente questa

¹Dove il pedice di J indica il numero di misure fatte e ϑ_n è il numero di parametri da stimare

funzione è definita in modo che la stima migliore sia quella che minimizza tale funzione.

L'approccio che sarà usato in questa trattazione si chiama PEM (Prediction Error Minimization) e consiste nel minimizzare la varianza dell'errore di predizione a un passo:

$$J(\vartheta) = \mathbb{E} \left[(y(t) - \hat{y}(t|t-1, \vartheta))^2 \right]$$

ma dato che si possiedono solo N dati si usa una sua approssimazione, ossia la varianza campionaria:

$$J_N(\vartheta) = \frac{1}{N} \cdot \sum_{t=1}^N (y(t) - \hat{y}(t|t-1, \vartheta))^2$$

ovvero un modello verrà considerato buono in base a quanto è affidabile la sua previsione del futuro.

4. Minimizzazione della cifra di merito:

Una volta stabilita la cifra di merito bisogna trovare un modo per minimizzarla e quindi trovare la stima $\hat{\vartheta}_N$. In base a come la cifra di merito si presenta, gli approcci sono differenti (si faccia riferimento alla figura 16.2 per degli esempi):

- La funzione è quadratica:
in questo caso la funzione è un paraboloide e quindi ha un solo minimo locale, che quindi è anche assoluto, e si può trovare in modo analitico.
- La funzione non è quadratica, ma ha solo un minimo locale:
in questo caso la funzione rimane convessa e quindi è possibile usare metodi numerici (metodo del gradiente, newton o quasi newton) per trovare il minimo, con garanzia di convergenza verso il minimo globale.
- La funzione non è quadratica e ha più minimi locali:
purtroppo in questo caso la funzione diventa concava e quindi non ci sono garanzie di convergenza verso il minimo globale, in quanto un metodo numerico può benissimo convergere verso un minimo locale. Sarà necessaria una buona inizializzazione del metodo.

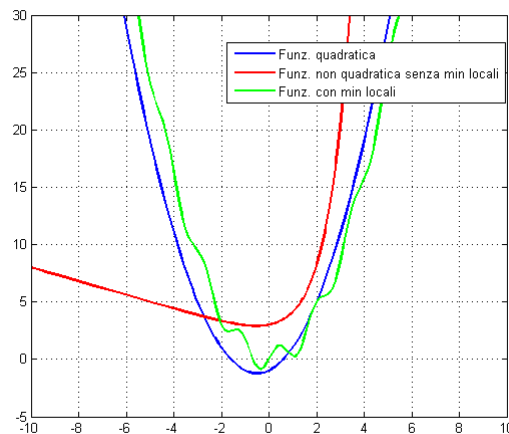


Figura 16.2.: Possibili funzioni della cifra di merito

5. Validazione del modello $M(\hat{\vartheta}_N)$:

Questa fase serve per controllare che il modello trovato rappresenti bene il sistema fisico. Questa fase è importante perchè nelle fasi svolte precedentemente si è stimato $\hat{\vartheta}$ in base ai valori misurati, ma il modello non deve rappresentare bene solo quei dati, ma bensì tutto il modello per qualsiasi tempo, qualsiasi ingresso e qualsiasi realizzazione del rumore.

Questa fase è fondamentale perchè permette di garantire che il modello stimato rappresenti bene il sistema sotto esame e non si limiti a interpolare i dati misurati.

17. Identificazione PEM di un modello ARX

Questo metodo permette di stimare i parametri di un modello $ARX(m, p+1)$ dato il set di N dati $D = \{y_1, \dots, y_N, u_1, \dots, u_N\}$ ossia trovare i parametri di un processo $ARX(m, p+1)$ che approssimi il meglio possibile il modello conoscendo solamente il set di dati D . Si consideri quindi un generico $ARX(m, p+1)$:

$$y(t) = \frac{B(z)}{A(z)} \cdot u(t-1) + \frac{1}{A(z)} \cdot e(t)$$

Dove:

$$\begin{aligned} B(z) &= b_0 + b_1 \cdot z^{-1} + \dots + b_p \cdot z^{-p} \\ A(z) &= 1 + a_1 \cdot z^{-1} + \dots + a_m \cdot z^{-m} \\ C(z) &= 1 \end{aligned}$$

Si noti che il ritardo puro tra ingresso e uscita $k = 1$ non intacca la generalizzazione del modello, in quanto (supponendo di eseguire correttamente l'identificazione) se il sistema presentasse un ritardo puro tra ingresso e uscita $k > 1$, l'identificazione restituirà $b_0 = \dots = b_{k-2} = 0$

Scelta della cifra di merito Dato che si vuole eseguire una identificazione PEM la cifra di merito è la varianza campionaria dell'errore del predittore:

$$J_N(\vartheta) = \frac{1}{N} \cdot \sum_{t=1}^N (y(t) - \hat{y}(t|t-1, \vartheta))^2$$

Ricavando il predittore:

$$\begin{aligned} \hat{y}(t|t-1) &= \frac{B(z)}{C(z)} \cdot u(t-1) + \frac{C(z) - A(z)}{C(z)} \cdot y(t) = \\ &= B(z) \cdot u(t-1) + (1 - A(z)) \cdot y(t) = \\ &= (b_0 + b_1 \cdot z^{-1} + \dots + b_p \cdot z^{-p}) \cdot u(t-1) + (-a_1 \cdot z^{-1} - a_2 \cdot z^{-2} - \dots - a_m \cdot z^{-m}) \cdot y(t) = \\ &= -(a_1 + a_2 \cdot z^{-1} + \dots + a_m \cdot z^{-m+1}) \cdot y(t-1) + (b_0 + b_1 \cdot z^{-1} + \dots + b_p \cdot z^{-p}) \cdot u(t-1) \end{aligned}$$

si nota però che:

- in un processo $AR(m)$ (e quindi anche in un $ARX(m, p+1)$) il predittore $\hat{y}(t|t-1, \vartheta)$ dipende da m misure precedenti $y(t-1), y(t-2), \dots, y(t-m)$
- per poter trasferire la variabile esogena sull'uscita correttamente bisogna conoscere $p+1$ valori precedenti dell'ingresso $u(t-1), u(t-2), \dots, u(t-p-1)$

per questo motivo il predittore non può essere calcolato per tutti i N valori perchè i primi $h = \max(m, p+1)$ valori sono incalcolabili per le osservazioni appena fatte. Per questo motivo la cifra di merito corretta è:

$$J_N(\vartheta) = \frac{1}{N-h} \cdot \sum_{t=h+1}^N (y(t) - \hat{y}(t|t-1, \vartheta))^2$$

Introduzione del vettore delle osservazioni Per risolvere un problema di identificazione bisogna stimare $n_\vartheta = m + p + 1$ parametri e quindi il vettore $\vartheta = [a_1, a_2, \dots, a_m, b_0, b_1, \dots, b_p]^T$ di dimensioni $n_\vartheta \times 1$ ¹. Si definisce un vettore delle osservazioni di dimensioni $n_\vartheta \times 1$ per rappresentare il predittore in una forma più compatta:

$$\varphi(t) = [-y(t-1), -y(t-2), \dots, -y(t-m), u(t-1), u(t-2), \dots, u(t-p-1)]^T$$

Il predittore¹¹ sarà quindi:

$$\hat{y}(t|t-1) = \varphi(t)^T \cdot \vartheta$$

E la cifra di merito sarà:

$$J_N(\vartheta) = \frac{1}{N-h} \cdot \sum_{t=h+1}^N \left(y(t) - \varphi(t)^T \cdot \vartheta \right)^2$$

Siccome la cifra di merito è una funzione quadratica in ϑ , è possibile calcolare il minimo analiticamente:

$$\begin{aligned} \left. \frac{\partial J_N(\vartheta)}{\partial \vartheta} \right|_{\vartheta=\hat{\vartheta}} &= 0 \\ \frac{1}{N-h} \cdot \sum_{t=h+1}^N \left[\left(y(t) - \varphi(t)^T \cdot \hat{\vartheta} \right) \cdot (-\varphi(t)) \right] &= 0 \\ \sum_{t=h+1}^N \left[\left(y(t) - \varphi(t)^T \cdot \hat{\vartheta} \right) \cdot \varphi(t) \right] &= 0 \end{aligned}$$

Il prodotto nella sommatoria è tra uno scalare e un vettore e quindi gode della proprietà commutativa:

$$\begin{aligned} \sum_{t=h+1}^N \left[\varphi(t) \cdot \left(y(t) - \varphi(t)^T \cdot \hat{\vartheta} \right) \right] &= 0 \\ \sum_{t=h+1}^N [\varphi(t) \cdot y(t)] - \sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T \cdot \hat{\vartheta}] &= 0 \\ \sum_{t=h+1}^N [\varphi(t) \cdot y(t)] &= \sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T \cdot \hat{\vartheta}] \\ \sum_{t=h+1}^N [\varphi(t) \cdot y(t)] &= \sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T] \cdot \hat{\vartheta} \end{aligned}$$

Essendo $\sum_{t=h+1}^N \varphi(t)^T \cdot \varphi(t)$ una matrice quadrata (che ipotizziamo invertibile), si ha che:

$$\begin{aligned} \left(\sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T] \right)^{-1} \cdot \sum_{t=h+1}^N [\varphi(t) \cdot y(t)] &= \left(\sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T] \right)^{-1} \cdot \sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T] \cdot \hat{\vartheta} \\ \left(\sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T] \right)^{-1} \cdot \left(\sum_{t=h+1}^N [\varphi(t) \cdot y(t)] \right) &= \hat{\vartheta} \\ \hat{\vartheta} &= \left(\sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T] \right)^{-1} \cdot \left(\sum_{t=h+1}^N [\varphi(t) \cdot y(t)] \right) \end{aligned}$$

¹Tutti i vettori in questo corso sono vettori colonna.

¹¹Questa forma compatta è un prodotto scalare e quindi può anche essere scritto come $\hat{y}(t|t-1) = \vartheta^T \cdot \varphi(t)$, ma scrivere $\hat{y}(t|t-1) = \vartheta \cdot \varphi(t)^T$ o $\hat{y}(t|t-1) = \varphi(t) \cdot \vartheta^T$ è sbagliato perchè queste sono matrici $n_\vartheta \times n_\vartheta$.

17. Identificazione PEM di un modello ARX

Questo metodo viene anche chiamato least square (LS) dato che consiste nel minimizzare la media del quadrato dell'errore.

Esempio Si supponga di avere 10 dati ($N = 10$) e si stimi un modello del tipo:

$$y(t) = \frac{b}{1 + a \cdot z^{-1}} \cdot u(t-1) + \frac{1}{1 + a \cdot z^{-1}} \cdot e(t)$$

$$\begin{aligned} B(z) &= b \\ A(z) &= 1 + a \cdot z^{-1} \\ C(z) &= 1 \end{aligned}$$

ovvero un $ARX(1,1)$

Come visto precedentemente, è necessario avere l'espressione del predittore ad un passo per poter calcolare la cifra di merito:

$$\begin{aligned} \hat{y}(t|t-1) &= \frac{B(z)}{C(z)} \cdot u(t-1) + \frac{C(z) - A(z)}{C(z)} \cdot y(t) = \\ &= b \cdot u(t-1) - a \cdot y(t-1) \end{aligned}$$

Si possono ora definire ϑ e $\varphi(t)$:

$$\begin{aligned} \vartheta &= \begin{bmatrix} a \\ b \end{bmatrix} \\ \varphi(t) &= \begin{bmatrix} -y(t-1) \\ u(t-1) \end{bmatrix} \end{aligned}$$

Dato che $h = \max(1, 1) = 1$, la cifra di merito sarà quindi:

$$\begin{aligned} J_{10}(\vartheta) &= \frac{1}{10-1} \cdot \sum_{t=2}^{10} \left(y(t) - \varphi(t)^T \cdot \vartheta \right)^2 = \\ &= \frac{1}{9} \cdot \sum_{t=2}^{10} (y(t) + a \cdot y(t-1) - b \cdot u(t-1))^2 \end{aligned}$$

Derivando:

$$\begin{cases} \frac{\partial J_{10}(\vartheta)}{\partial a} = \frac{2}{9} \cdot \left(\sum_{t=2}^{10} y(t) + a \cdot y(t-1) - b \cdot u(t-1) \right) \cdot y(t-1) = 0 \\ \frac{\partial J_{10}(\vartheta)}{\partial b} = \frac{2}{9} \cdot \left(\sum_{t=2}^{10} y(t) + a \cdot y(t-1) - b \cdot u(t-1) \right) \cdot (-u(t-1)) = 0 \end{cases}$$

$$\begin{cases} \sum_{t=2}^{10} (y(t) \cdot y(t-1)) + \sum_{t=2}^{10} (a \cdot y(t-1) \cdot y(t-1)) - \sum_{t=2}^{10} (b \cdot u(t-1) \cdot y(t-1)) = 0 \\ - \sum_{t=2}^{10} (y(t) \cdot u(t-1)) - \sum_{t=2}^{10} (a \cdot y(t-1) \cdot u(t-1)) + \sum_{t=2}^{10} (b \cdot u(t-1) \cdot u(t-1)) = 0 \end{cases}$$

$$\begin{cases} a \cdot \sum_{t=2}^{10} (y(t-1)^2) - b \cdot \sum_{t=2}^{10} (u(t-1) \cdot y(t-1)) = - \sum_{t=2}^{10} y(t) \cdot y(t-1) \\ -a \cdot \sum_{t=2}^{10} (y(t-1) \cdot u(t-1)) + b \cdot \sum_{t=2}^{10} (u(t-1)^2) = \sum_{t=2}^{10} y(t) \cdot u(t-1) \end{cases}$$

Scrivendo il sistema in forma matriciale si avrà:

$$\begin{bmatrix} \sum_{t=2}^{10} y(t-1)^2 & -\sum_{t=2}^{10} y(t-1) \cdot u(t-1) \\ -\sum_{t=2}^{10} y(t-1) \cdot u(t-1) & \sum_{t=2}^{10} u(t-1)^2 \end{bmatrix} \cdot \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} -\sum_{t=2}^{10} y(t) \cdot y(t-1) \\ \sum_{t=2}^{10} y(t) \cdot u(t-1) \end{bmatrix}$$

Esiste però anche un metodo alternativo per risolvere questo tipo di problemi, che si basa sull'imporre che il predittore restituisca previsioni esatte per ogni dato.

17.1. Metodo alternativo per impostare il problema

Questo metodo consiste nell'imporre che il predittore sia uguale alla misura fatta:

$$y(t) = \hat{y}(t|t-1), \forall t > h$$

dove $h = \max(m, p+1)$ perchè i predittori per $t \leq h$ non sono calcolabili. Questa espressione si traduce in un sistema:

$$\begin{cases} \hat{y}(h+1|h) & = y(h+1) \\ \hat{y}(h+2|h+1) & = y(h+2) \\ \vdots & \\ \hat{y}(N|N-1) & = y(N) \end{cases}$$

Ricordando che:

$$\begin{aligned} \hat{y}(t|t-1) &= -(a_1 + a_2 \cdot z^{-1} + \dots + a_m \cdot z^{-m+1}) \cdot y(t-1) + (b_0 + b_1 \cdot z^{-1} + \dots + b_p \cdot z^{-p}) \cdot u(t-1) = \\ &= -a_1 \cdot y(t-1) - a_2 \cdot y(t-2) - \dots - a_m \cdot y(t-m) + b_0 \cdot u(t-1) + \dots + b_p \cdot u(t-1-p) \end{aligned}$$

il sistema si può scrivere come:

$$\begin{cases} -a_1 \cdot y(h) - \dots - a_m \cdot y(h+1-m) + b_0 \cdot u(h) + \dots + b_p \cdot u(h-p) & = y(h+1) \\ -a_1 \cdot y(h+1) - \dots - a_m \cdot y(h+2-m) + b_0 \cdot u(h+1) + \dots + b_p \cdot u(h+1-p) & = y(h+2) \\ \vdots & \\ -a_1 \cdot y(N-1) - \dots - a_m \cdot y(N-1-m) + b_0 \cdot u(N-1) + \dots + b_p \cdot u(h+1-p) & = y(N) \end{cases}$$

Da cui si nota che è un sistema a n_θ variabili e $N-h$ equazioni, e quindi generalmente è un sistema con più equazioni che variabili dato che solitamente $N \gg n_\theta$, ossia si prendono più misure del numero di parametri da stimare. Per poterlo scrivere in forma compatta si introduce la matrice:

$$\Phi_{(N-h) \times n_\theta} = \begin{bmatrix} -y(h) & -y(h-1) & \dots & y(h+1-m) & u(h) & u(h-1) & \dots & u(h-p) \\ -y(h+1) & -y(h) & \dots & y(h+2-m) & u(h+1) & u(h) & \dots & u(h+1-p) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -y(N-1) & -y(N-2) & \dots & y(N-1-m) & u(N-1) & u(N-2) & \dots & u(h+1-p) \end{bmatrix}$$

e il vettore:

$$\Upsilon_{(N-h) \times 1} = \begin{bmatrix} y(h+1) \\ y(h+2) \\ \vdots \\ y(N) \end{bmatrix}$$

17. Identificazione PEM di un modello ARX

Il sistema scritto precedentemente diventa:

$$\begin{matrix} \Phi & \cdot & \vartheta & = & \Upsilon \\ (N-h) \times n_{\vartheta} & & n_{\vartheta} \times 1 & & (N-h) \times 1 \end{matrix}$$

Da cui si può notare che Φ è una matrice non quadrata, quindi non invertibile. Questo problema può essere risolto attraverso il calcolo della matrice Φ^\dagger pseudo-inversa di Φ che si ottiene moltiplicando per la trasposta entrambi i termini del sistema:

$$\Phi^T \cdot \Phi \cdot \vartheta = \Phi^T \cdot \Upsilon$$

Facendo l'analisi dimensionale si nota che:

$$\begin{matrix} \Phi^T & \cdot & \Phi & \cdot & \vartheta & = & \Phi^T & \cdot & \Upsilon \\ n_{\vartheta} \times (N-h) & & (N-h) \times n_{\vartheta} & & n_{\vartheta} \times 1 & & n_{\vartheta} \times (N-h) & & (N-h) \times 1 \end{matrix}$$

$$\begin{matrix} \Phi^T \cdot \Phi & \cdot & \vartheta & = & \Phi^T \cdot \Upsilon \\ n_{\vartheta} \times n_{\vartheta} & & n_{\vartheta} \times 1 & & n_{\vartheta} \times 1 \end{matrix}$$

e quindi è un sistema risolvibile dato che $\Phi^T \cdot \Phi$ è una matrice quadrata (che si suppone invertibile), ottenendo:

$$\begin{aligned} (\Phi^T \cdot \Phi)^{-1} \cdot (\Phi^T \cdot \Phi) \cdot \vartheta &= (\Phi^T \cdot \Phi)^{-1} \cdot \Phi^T \cdot \Upsilon \\ \hat{\vartheta} &= \underbrace{(\Phi^T \cdot \Phi)^{-1} \cdot \Phi^T}_{\Phi^\dagger} \cdot \Upsilon \end{aligned}$$

Si può dimostrare che i due metodi spiegati sono equivalenti e, inoltre, che:

$$\Phi^T \cdot \Phi = \sum_{t=h+1}^N [\varphi(t) \cdot \varphi(t)^T]$$

$$\Phi^T \cdot \Upsilon = \sum_{t=h+1}^N [\varphi(t) \cdot y(t)]$$

Esempio Riprendendo l'esempio precedente si ha come predittore:

$$\hat{y}(t|t-1) = b \cdot u(t-1) - a \cdot y(t-1)$$

Quindi imponendo che il predittore sia uguale al dato si ottiene il sistema:

$$\begin{cases} b \cdot u(1) - a \cdot y(1) = y(2) \\ b \cdot u(2) - a \cdot y(2) = y(3) \\ \vdots \\ b \cdot u(9) - a \cdot y(9) = y(10) \end{cases}$$

Ottenendo un sistema sovradimensionato, come previsto nella teoria. Si possono definire le matrici Φ e Υ come segue:

$$\Phi = \begin{bmatrix} -y(1) & u(1) \\ -y(2) & u(2) \\ \vdots & \vdots \\ -y(9) & u(9) \end{bmatrix}_{9 \times 2}$$

$$\Upsilon_{9 \times 1} = \begin{bmatrix} y(2) \\ y(3) \\ \vdots \\ y(10) \end{bmatrix}$$

Il sistema scritto precedentemente diventa:

$$\Phi \cdot \vartheta = \Upsilon$$

Ricavando la soluzione tramite pseudo-inversa si ottiene:

$$\hat{\vartheta} = (\Phi^T \cdot \Phi)^{-1} \cdot \Phi^T \cdot \Upsilon$$

È possibile anche controllare che questo secondo metodo è equivalente al primo proposto:

$$\begin{aligned} \Phi^T \cdot \Phi &= \begin{bmatrix} -y(1) & -y(2) & \cdots & -y(9) \\ u(1) & y(2) & \cdots & u(9) \end{bmatrix} \cdot \begin{bmatrix} -y(1) & u(1) \\ -y(2) & u(2) \\ \vdots & \vdots \\ -y(9) & u(9) \end{bmatrix} = \\ &= \begin{bmatrix} \sum_{t=2}^{10} y(t-1)^2 & -\sum_{t=2}^{10} y(t-1) \cdot u(t-1) \\ -\sum_{t=2}^{10} y(t-1) \cdot u(t-1) & \sum_{t=2}^{10} u(t-1)^2 \end{bmatrix} \\ \Phi^T \cdot \Upsilon &= \begin{bmatrix} -y(1) & -y(2) & \cdots & -y(9) \\ u(1) & y(2) & \cdots & u(9) \end{bmatrix} \cdot \begin{bmatrix} y(2) \\ y(3) \\ \vdots \\ y(10) \end{bmatrix} = \\ &= \begin{bmatrix} -\sum_{t=2}^{10} y(t) \cdot y(t-1) \\ \sum_{t=2}^{10} y(t) \cdot u(t-1) \end{bmatrix} \end{aligned}$$

Esempio Si hanno $N = 5$ dati derivati da $y(t)$ (processo a media nulla):

$$\begin{aligned} y(1) &= \frac{1}{2} \\ y(2) &= 0 \\ y(3) &= -1 \\ y(4) &= -\frac{1}{2} \\ y(5) &= \frac{1}{4} \end{aligned}$$

Si identifichi un modello $AR(1)$:

$$\begin{aligned} y(t) &= a \cdot y(t-1) + e(t) \\ \text{con} \quad e(t) &\sim WN(0, \lambda^2) \end{aligned}$$

Successivamente si usi il modello identificato per calcolare il valore della predizione $\hat{y}(6|5)$.

17. Identificazione PEM di un modello ARX

Per prima cosa è necessario trovare il predittore ad un passo: è possibile notare che la parte predicibile è solamente $a \cdot y(t-1)$, oppure si può applicare la formula per il calcolo di un predittore ad un passo di un modello *ARMA*, con $C(z) = 1$ e $A(z) = 1 - a \cdot z^{-1}$.

$$\hat{y}(t|t-1) = \frac{C(z) - A(z)}{C(z)} \cdot y(t) = \frac{1 - 1 + a \cdot z^{-1}}{1} \cdot y(t) = a \cdot y(t-1)$$

Conoscendo il predittore ad un passo, è possibile calcolare la cifra di merito:

$$\begin{aligned} J_5(\vartheta) &= \frac{1}{5-1} \cdot \sum_{t=2}^5 (y(t) - a \cdot y(t-1))^2 = \\ &= \frac{1}{4} \cdot \sum_{t=2}^5 y(t)^2 + a^2 \cdot y(t-1)^2 - 2 \cdot a \cdot y(t) \cdot y(t-1) = \\ &= \frac{1}{4} \cdot \left[\underbrace{\left(0 + \frac{1}{4} \cdot a^2 + 0\right)}_{t=2} + \underbrace{(1 + 0 + 0)}_{t=3} + \underbrace{\left(\frac{1}{4} + a^2 - a\right)}_{t=4} + \underbrace{\left(\frac{1}{16} + \frac{1}{4} \cdot a^2 + \frac{1}{4} \cdot a\right)}_{t=5} \right] = \\ &= \frac{1}{4} \cdot \left(\frac{6}{4} \cdot a^2 - \frac{3}{4} \cdot a + \frac{21}{16} \right) = \\ &= \frac{3}{8} \cdot a^2 - \frac{3}{16} \cdot a + \frac{21}{64} \end{aligned}$$

Essendo la cifra di merito una funzione quadratica, è possibile calcolare il minimo analiticamente:

$$\begin{aligned} \frac{\partial J_5(\vartheta)}{\partial a} &= 0 \\ \frac{3}{4} \cdot a - \frac{3}{16} &= 0 \\ a &= \frac{1}{4} \end{aligned}$$

Quindi il sistema diventerà:

$$\begin{aligned} y(t) &= \frac{1}{1 - a \cdot z^{-1}} \cdot e(t) = \frac{1}{1 - \frac{1}{4} \cdot z^{-1}} \cdot e(t) \\ \text{con } e(t) &\sim WN(0, \lambda^2) \end{aligned}$$

Per trovare il valore di λ^2 basta ricordare che la varianza del predittore a un passo corrisponde a quella del rumore del processo:

$$\lambda^2 = Var[e(t)] = Var[\varepsilon(t)] = \frac{1}{4} \cdot \sum_{t=2}^{10} (y(t) - \hat{y}(t|t-1))^2 = \frac{3}{8} \cdot \left(\frac{1}{4}\right)^2 - \frac{3}{16} \cdot \frac{1}{4} + \frac{21}{64} = \frac{39}{128}$$

Infine calcolare $\hat{y}(6|5)$ è immediato:

$$\hat{y}(6|5) = \frac{1}{4} \cdot y(5) = \frac{1}{4} \cdot \frac{1}{4} = \frac{1}{16}$$

Osservazione Non è facile capire se le stime di a e λ^2 appena fatte siano delle buone stime, ma dati i valori delle misure queste sono le migliori stime che si possano fare. Bisogna comunque capire se avere solo 5 misure basta e se approssimare con un *AR*(1) vada bene per questo sistema fisico.

18. Identificazione PEM di un modello $ARMAX$

Il problema è molto simile al precedente: dato un set di N dati $D = \{u_1, u_2, \dots, u_n, y_1, y_2, \dots, y_n\}$ bisogna stimare i parametri di un processo $ARMAX(m, n, p + 1)$. Si consideri il generico processo $ARMAX$:

$$y(t) = \frac{B(z)}{A(z)} \cdot u(t-1) + \frac{C(z)}{A(z)} \cdot e(t)$$

Come fatto precedentemente, si consideri il vettore colonna contenente i parametri da stimare (di dimensione $n_\vartheta = m + n + p + 1$):

$$\vartheta = \begin{bmatrix} a_1 \\ \vdots \\ a_m \\ b_0 \\ \vdots \\ b_p \\ c_1 \\ \vdots \\ c_n \end{bmatrix}$$

Usando lo stesso approccio dell'analisi per modelli ARX , si deve trovare il minimo della varianza campionaria dell'errore di predizione per il predittore ad un passo. Dato che:

$$\varepsilon(t) = E(z) \cdot e(t)$$

e siccome $E(z) = 1$ dato che si sta considerando il predittore ad un passo (e quindi è stato effettuato un solo colpo di lunga divisione)

$$\begin{aligned} \varepsilon(t) &= e(t) = \\ &= \frac{A(z)}{C(z)} \cdot \left(y(t) - \frac{B(z)}{A(z)} \cdot u(t-1) \right) = \\ &= \frac{A(z)}{C(z)} \cdot y(t) - \frac{\cancel{A(z)}}{C(z)} \cdot \frac{B(z)}{\cancel{A(z)}} \cdot u(t-1) = \\ &= \frac{A(z)}{C(z)} \cdot y(t) - \frac{B(z)}{C(z)} \cdot u(t-1) \end{aligned}$$

la cifra di merito sarà quindi:

$$\begin{aligned} J_N &= \frac{1}{N} \cdot \sum_{t=1}^N \left(\frac{A(z)}{C(z)} \cdot y(t) - \frac{B(z)}{C(z)} \cdot u(t-1) \right)^2 = \\ &= \frac{1}{N} \cdot \sum_{t=1}^N \varepsilon(t, \vartheta)^2 = \end{aligned}$$

Si può subito notare che in questo caso si hanno parametri da stimare al denominatore (ovvero c_1, \dots, c_n), quindi la cifra di merito non sarà un paraboloide (e probabilmente non sarà nemmeno convessa), quindi è necessario usare metodi iterativi per stimare il minimo assoluto della funzione.

18.1. Introduzione ai metodi iterativi per l'identificazione di un processo ARMAX

Un metodo iterativo si compone in due fasi:

- Scelta di un valore di inizializzazione ϑ^{11}
seguendo l'approccio puramente black box la scelta del ϑ^1 sarebbe casuale, ma in pratica si presenteranno sempre informazioni utili a fare una stima migliore da usare come base per l'algoritmo iterativo, diminuendo drasticamente il numero di passi necessari per trovare la soluzione.
- Una procedura per ricavare il ϑ^i successivo
Nella maggior parte dei casi consiste in un criterio che ricava ϑ^{i+1} conoscendo ϑ^i , che permette di creare un successione che converge a $\hat{\vartheta}_N$

Dopo un certo numero N^{11} di iterazioni si raggiunge un valore ϑ^N che è una buona approssimazione di $\hat{\vartheta}_N$.

Questa procedura però non garantisce di convergere al minimo globale, infatti con una sbagliata inizializzazione il metodo potrebbe convergere a un minimo locale della funzione. Per risolvere questo problema si esegue il metodo M volte usando dei diversi ϑ^1 in modo da ottenere M stime di $\hat{\vartheta}_N$. Se le M stime sono tutte uguali allora si ha ragionevole certezza (ma non si ha la garanzia) di aver trovato il minimo globale, se invece sono diverse allora la miglior stima che si ha è quella con $J(\vartheta^N)$ minore.

Con questo metodo non si può garantire di trovare il minimo globale, ma più si cerca più si aumentano le probabilità di trovarlo creando un trade-off tra tempo e miglior stima.

18.2. Introduzione al metodo di newton

Si può partire dalla seguente idea di base: approssimare $J_N(\vartheta)$ con il suo sviluppo in serie di Taylor al secondo ordine intorno a ϑ_i , trovare il minimo di quest'ultimo e usarlo come ϑ_{i+1} . Una rappresentazione di questo metodo è riportata in figura 18.1.

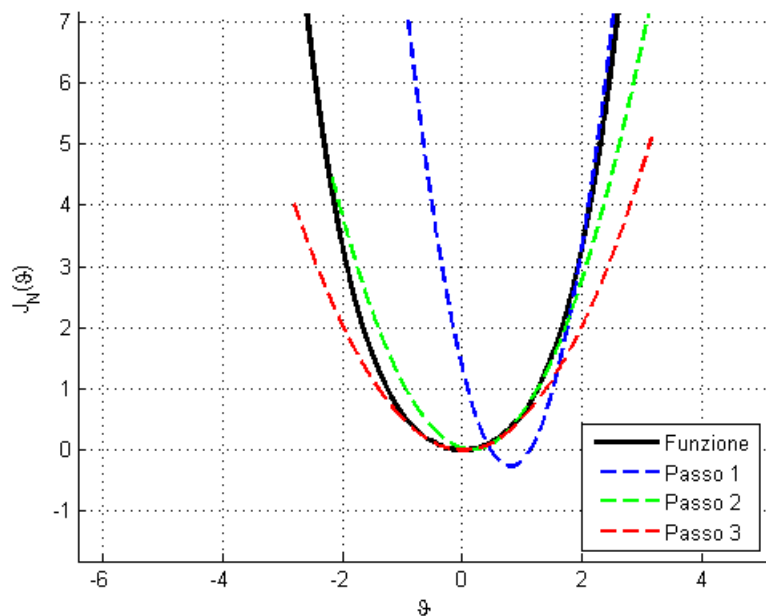


Figura 18.1.: Rappresentazione grafica del metodo di Newton

¹Per la successione del metodo iterativo si mette l'indice come esponente e bisogna stare attenti a non scambiare per una potenza

¹¹Questo N non è lo stesso che indica il numero di misure

18. Identificazione PEM di un modello ARMAX

Tramite l'approssimazione al secondo ordine della serie di Taylor intorno a ϑ^i si ottiene:

$$J_N(\vartheta) \approx V(\vartheta) = J_N(\vartheta^i) + \underbrace{\frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i}}_{\text{Gradiente}} \cdot (\vartheta - \vartheta^i) + \frac{1}{2} \cdot (\vartheta - \vartheta^i)^T \underbrace{\frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i}}_{\text{Hessiana}} \cdot (\vartheta - \vartheta^i)$$

Per trovare il minimo bisogna trovare lo zero della derivata e quindi:

$$\begin{aligned} \frac{\partial V(\vartheta)}{\partial \vartheta} &= \frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i} + \frac{1}{2} \cdot 2 \cdot \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i} \cdot (\vartheta - \vartheta^i) = 0 \\ \frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i} + \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i} \cdot \vartheta - \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i} \cdot \vartheta^i &= 0 \\ \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i} \cdot \vartheta &= \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i} \cdot \vartheta^i - \frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i} \\ \vartheta &= \cancel{\frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i}^{-1}} \cdot \cancel{\frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i}} \cdot \vartheta^i - \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i}^{-1} \cdot \frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i} \\ \vartheta &= \vartheta^i - \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i}^{-1} \cdot \frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i} \end{aligned}$$

Dato che ϑ è lo zero della derivata di $V(\vartheta)$ allora è un punto stazionario di $V(\vartheta)$ che è un'ottima approssimazione di $J_N(\vartheta)$ intorno a ϑ^i e quindi:

- Se la matrice hessiana è definita positiva allora il punto è un minimo e quindi è una migliore approssimazione del minimo di $J_N(\vartheta)$ e quindi si può usare come ϑ^{i+1}
- Se la matrice hessiana ha determinante nullo allora non si può invertire e allora è una situazione da evitare
- Se la matrice hessiana è definita negativa allora il punto è un massimo e quindi se lo si usasse come ϑ^{i+1} la stima peggiorerebbe e quindi è una situazione da evitare

quindi si usa come passo successivo:

$$\vartheta_{i+1} = \vartheta_i - \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \Big|_{\vartheta^i}^{-1} \cdot \frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i}$$

ma il metodo funziona solo quando la matrice hessiana è definita positiva.

Calcolando il gradiente si ottiene:

$$\begin{aligned} \frac{\partial J_N(\vartheta)}{\partial \vartheta} \Big|_{\vartheta^i} &= \frac{\partial}{\partial \vartheta} \left(\frac{1}{N} \cdot \sum_{t=1}^N \varepsilon(t, \vartheta)^2 \right) \Big|_{\vartheta^i} = \\ &= \frac{2}{N} \cdot \sum_{t=1}^N \varepsilon(t) \cdot \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \Big|_{\vartheta^i} \end{aligned}$$

Il calcolo e l'inversione della matrice hessiana è complicato e quindi si approssima e in base al tipo di approssimazione il metodo cambia nome.

18.3. Metodo del gradiente

Nel metodo del gradiente si approssima la matrice hessiana con una costante μ scalare e positiva.

$$\vartheta^{i+1} = \vartheta^i - \mu \cdot \left. \frac{\partial J_N(\vartheta)}{\partial \vartheta} \right|_{\vartheta^i}$$

Vantaggi:

- la matrice hessiana è sempre definita positiva (dato che è costante) e quindi il metodo converge sempre su un minimo e quindi è un metodo robusto
- è molto semplice computazionalmente

Svantaggi:

- è molto lento in quanto la costante corregge la stima sempre nello stesso modo quindi quando la funzione ha un gradiente piccolo il metodo rallenta anche se magari ci si trova lontano dal minimo, per questo motivo a volte si usa una μ variabile

18.4. Metodo di Newton

Nel metodo di Newton l'hessiana non viene approssimata e quindi:

$$\vartheta_{i+1} = \vartheta_i - \left. \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \right|_{\vartheta^i}^{-1} \cdot \left. \frac{\partial J_N(\vartheta)}{\partial \vartheta} \right|_{\vartheta^i}$$

Vantaggi:

- L'hessiana non viene approssimata e quindi questo metodo richiede poche iterazioni e quindi è veloce.

Svantaggi:

- Calcolare l'inversa della matrice hessiana è complicato ed è computazionalmente complesso
- la matrice hessiana potrebbe perdere la proprietà di definizione positiva e quindi il metodo non è robusto

18.5. Metodi di quasi-Newton

Calcolando l'hessiana si ottiene:

$$\begin{aligned} \left. \frac{\partial^2 J_N(\vartheta)}{\partial \vartheta^2} \right|_{\vartheta^i} &= \left. \frac{\partial}{\partial \vartheta} \left(\frac{2}{N} \cdot \sum_{t=1}^N \varepsilon(t, \vartheta) \cdot \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right) \right|_{\vartheta^i} = \\ &= \frac{2}{N} \cdot \left(\sum_{t=1}^N \left. \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right|_{\vartheta^i} \cdot \left(\left. \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right|_{\vartheta^i} \right)^T + \sum_{t=1}^N \varepsilon(t, \vartheta) \cdot \left. \frac{\partial^2 \varepsilon(t, \vartheta)}{\partial \vartheta^2} \right|_{\vartheta^i} \right) \end{aligned}$$

A questo punto si può prendere una versione approssimata della matrice Hessiana scartando il termine in cui compare la derivata seconda dell'errore di predizione; si fa questo per due motivi:

18. Identificazione PEM di un modello ARMAX

- Più l'algoritmo si avvicina al minimo, più piccolo sarà $\varepsilon(t)$, rendendo quindi il termine piccolo anch'esso assottigliando quindi l'errore dovuto all'approssimazione
- Il termine restante è una matrice definita positiva, dato che è un vettore per se stesso. Rendendo il metodo robusto.

Ottenendo:

$$\vartheta_{i+1} = \vartheta_i - M^{-1} \cdot \left. \frac{\partial J_N(\vartheta)}{\partial \vartheta} \right|_{\vartheta^i}$$

dove:

$$M = \frac{2}{N} \cdot \sum_{t=1}^N \left. \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right|_{\vartheta^i} \cdot \left(\left. \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right|_{\vartheta^i} \right)^T$$

La matrice M potrebbe ancora avere determinante nullo e quindi non essere invertibile, per risolverlo si somma un valore piccolo alla matrice, ottenendo:

$$M = \frac{2}{N} \cdot \left(\sum_{t=1}^N \left. \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right|_{\vartheta^i} \cdot \left(\left. \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right|_{\vartheta^i} \right)^T \right) + \delta I$$

con $\delta \in \mathbb{R}^+$ e δ piccolo

Come vantaggi e svantaggi questo metodo garantisce una via di mezzo tra quello di Newton e quello del gradiente.

Osservazione Il metodo esposto non è l'unico metodo della famiglia dei quasi-Newton, infatti rientrano in questa categoria tutti i metodi che approssimano l'hessiana con un matrice definita positiva.

18.6. Calcolo del gradiente dell'errore di predizione a un passo

Sia per il calcolo del gradiente di $J_N(\vartheta)$ che per il calcolo della sua hessiana (o delle sue approssimazioni) è richiesto il calcolo di:

$$\left. \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \right|_{\vartheta^i}$$

Ricordando che:

$$\begin{aligned} \varepsilon(t, \vartheta) &= e(t) = \frac{A(z)}{C(z)} \cdot y(t) - \frac{B(z)}{C(z)} \cdot u(t-1) = \\ &= \frac{1 + a_1 \cdot z^{-1} + \dots + a_m \cdot z^{-m}}{1 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}} \cdot y(t) - \frac{b_0 + b_1 \cdot z^{-1} + \dots + b_p \cdot z^{-p}}{1 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}} \cdot u(t-1) \end{aligned}$$

Si può calcolare le derivate rispetto a $a_1 \dots a_m$:

$$\begin{aligned} \frac{\partial \varepsilon(t)}{\partial a_1} &= \frac{z^{-1}}{C(z)} \cdot y(t) = \frac{1}{C(z)} \cdot y(t-1) = \alpha(t-1) \\ \frac{\partial \varepsilon(t)}{\partial a_2} &= \frac{z^{-2}}{C(z)} \cdot y(t) = \frac{1}{C(z)} \cdot y(t-2) = \alpha(t-2) \\ &\vdots \\ \frac{\partial \varepsilon(t)}{\partial a_m} &= \frac{z^{-m}}{C(z)} \cdot y(t) = \frac{1}{C(z)} \cdot y(t-m) = \alpha(t-m) \end{aligned}$$

Dove:

$$\alpha(t) = \frac{1}{C(z)} \cdot y(t)$$

Rispetto a $b_0 \dots b_p$:

$$\begin{aligned}\frac{\partial \varepsilon(t)}{\partial b_0} &= -\frac{1}{C(z)} \cdot u(t-1) = -\frac{1}{C(z)} \cdot u(t-1) = \beta(t-1) \\ \frac{\partial \varepsilon(t)}{\partial b_1} &= -\frac{z^{-1}}{C(z)} \cdot u(t-1) = -\frac{1}{C(z)} \cdot u(t-2) = \beta(t-2) \\ &\vdots \\ \frac{\partial \varepsilon(t)}{\partial b_p} &= -\frac{z^{-p}}{C(z)} \cdot u(t-1) = -\frac{1}{C(z)} \cdot u(t-p-1) = \beta(t-p-1)\end{aligned}$$

Dove:

$$\beta(t) = \frac{1}{C(z)} \cdot u(t)$$

Il calcolo delle derivate rispetto ai coefficienti $c_1 \dots c_n$ è più complesso, ma può essere semplificato riscrivendo l'equazione in questo modo:

$$\begin{aligned}\varepsilon(t, \vartheta) &= \frac{A(z)}{C(z)} \cdot y(t) - \frac{B(z)}{C(z)} \cdot u(t-1) \\ C(z) \cdot \varepsilon(t, \vartheta) &= A(z) \cdot y(t) - B(z) \cdot u(t-1)\end{aligned}$$

Calcolando la derivata parziale rispetto a c_1 :

$$\frac{\partial [(1 + c_1 \cdot z^{-1} + \dots + c_n \cdot z^{-n}) \cdot \varepsilon(t, \vartheta)]}{\partial c_1} = \frac{\partial [A(z) \cdot y(t) - B(z) \cdot u(t-1)]}{\partial c_1}$$

Siccome al primo membro compare una derivata di un prodotto (anche $\varepsilon(t, \vartheta)$ è funzione di c_1) e al secondo non compare nessun termine di $C(z)$ si ha che:

$$\begin{aligned}z^{-1} \cdot \varepsilon(t) + C(z) \cdot \frac{\partial \varepsilon(t)}{\partial c_1} &= 0 \\ \frac{\partial \varepsilon(t)}{\partial c_1} &= -\frac{1}{C(z)} \cdot z^{-1} \cdot \varepsilon(t)\end{aligned}$$

analogo per le altre c_i :

$$\begin{aligned}\frac{\partial \varepsilon(t)}{\partial c_1} &= -\frac{1}{C(z)} \cdot z^{-1} \cdot \varepsilon(t) = -\frac{1}{C(z)} \cdot \varepsilon(t-1) = \gamma(t-1) \\ \frac{\partial \varepsilon(t)}{\partial c_2} &= -\frac{1}{C(z)} \cdot z^{-2} \cdot \varepsilon(t) = -\frac{1}{C(z)} \cdot \varepsilon(t-2) = \gamma(t-2) \\ &\vdots \\ \frac{\partial \varepsilon(t)}{\partial c_n} &= -\frac{1}{C(z)} \cdot z^{-n} \cdot \varepsilon(t) = -\frac{1}{C(z)} \cdot \varepsilon(t-n) = \gamma(t-n)\end{aligned}$$

Dove:

$$\gamma(t) = -\frac{1}{C(z)} \cdot \varepsilon(t)$$

Infine si possono raggruppare tutte le derivate in un'unico vettore:

$$\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} = \begin{bmatrix} \alpha(t-1) \\ \vdots \\ \alpha(t-m) \\ \beta(t-1) \\ \vdots \\ \beta(t-p-1) \\ \gamma(t-1) \\ \vdots \\ \gamma(t-n) \end{bmatrix}$$

Tutto questo può essere riassunto in un'unico schema rappresentato in figura 18.2. Come si può notare si è creato un filtro, che dipende dal valore di ϑ^i , che dato in ingresso $u(t)$ e $y(t)$ si ottengono tutti i dati necessari per calcolare ϑ^{i+1} .

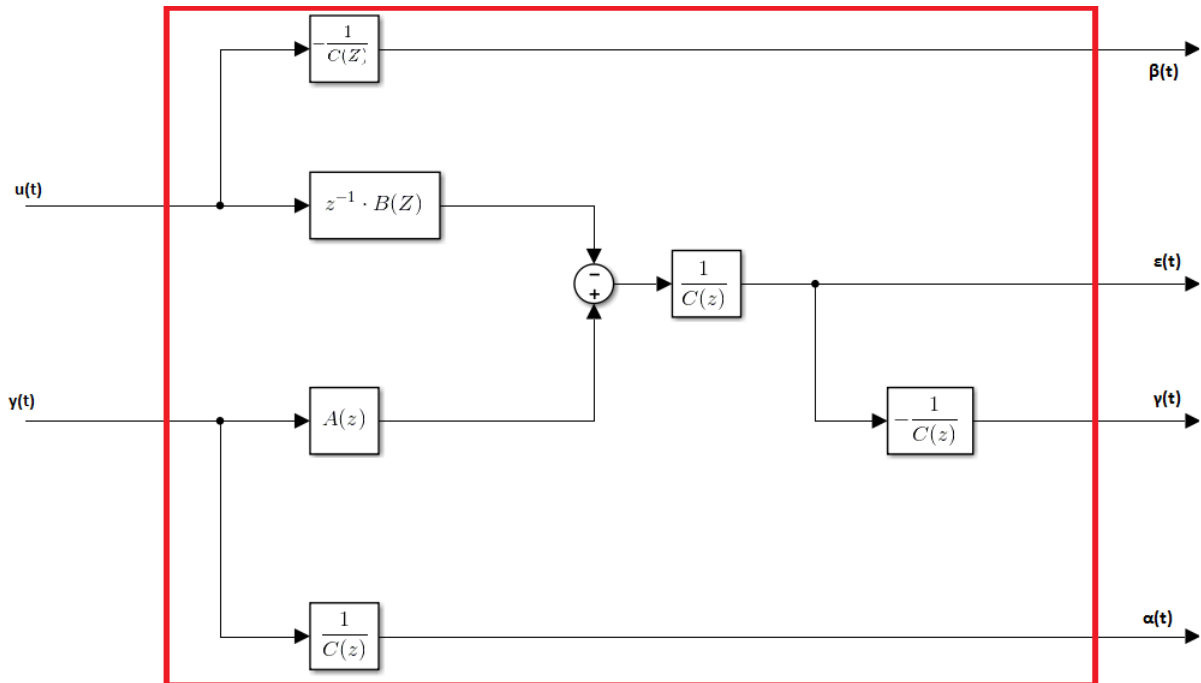


Figura 18.2.: Schema di filtraggio per il calcolo di $\frac{\partial \varepsilon(t)}{\partial \vartheta}$

Nel filtro si può notare che se in una iterazione il polinomio $C(z)$ è instabile porterebbe le uscite α , β e γ a divergere facendo sbagliare il metodo, esiste il metodo di Bauer che risolve questo problema canonizzando in modo efficiente il processo stocastico $\frac{C(z)}{A(z)} \cdot e(t)$ in modo che questa situazione non possa avvenire.

18.7. Conclusione

Il calcolo della stima dei parametri di modello ARMAX non è semplice e richiede l'utilizzo di uno dei tre metodi numerici (Gradiente, Newton o quasi-Newton) visti in precedenza, quindi il calcolo si esegue in modo iterativo, in cui a ogni iterazione si eseguono tre passi:

- Si usa ϑ^i per costruire $A^i(z)$, $B^i(z)$ e $C^i(z)$

18. Identificazione PEM di un modello ARMAX

- Si procede a calcolare $\varepsilon^i(t, \vartheta^i)$ e $\frac{\partial \varepsilon^i(t, \vartheta^i)}{\partial \vartheta}$ tramite il filtro illustrato prima
- Si calcola il successivo valore di ϑ attraverso la formula di uno dei tre metodi.

19. Analisi asintotica dei metodi PEM

Precedentemente si sono visti i metodi di identificazione PEM (a minimizzazione dell'errore di predizione) per modelli ARX e $ARMAX$, ma non si è controllato che i metodi fossero asintoticamente corretti e asintoticamente consistenti, le quali sono proprietà fondamentali per una stima.

Per studiare le proprietà asintotiche si considerino i metodi PEM con un numero di campioni $N \rightarrow \infty$. Dato che questi metodi corrispondono alla minimizzazione della cifra di merito è necessario calcolare quanto essa vale per $N \rightarrow \infty$, ricordando che la cifra di merito è lo stimatore della varianza dell'errore di predizione:

$$J_N(\vartheta) = \frac{1}{N} \cdot \sum_{t=1}^N \varepsilon(t, \vartheta)^2$$

e ricordando che questo stimatore è asintoticamente corretto si ottiene che:

$$J_N(\vartheta) \rightarrow \mathbb{E} [\varepsilon(t, \vartheta)^2] = J(\vartheta)$$

Definendo Δ come l'insieme dei minimi globali di $J(\vartheta)$

$$\Delta = \{\bar{\vartheta} | J(\vartheta) \geq J(\bar{\vartheta}), \forall \vartheta\}$$

si ha che i metodi PEM convergono a un minimo globale di $J(\vartheta)$ e quindi:

$$\hat{\vartheta}_N \rightarrow \bar{\vartheta}^* \in \Delta \text{ per } N \rightarrow \infty$$

Per poter capire se il metodo è corretto bisogna definire quale sia il parametro da stimare; Considerando quindi un sistema $S \in M(\vartheta)$, si può dire che esiste un vettore di parametri ϑ^0 tale che $M(\vartheta^0) = S$, ovvero che rappresenta perfettamente il sistema. Se il metodo di identificazione PEM è asintoticamente corretto, si avrà che il vettore $\vartheta^0 \in \Delta$.

È possibile dimostrare che $\vartheta^0 \in \Delta$. Si consideri l'errore di predizione per un certo ϑ :

$$\varepsilon(t, \vartheta) = y(t) - \hat{y}(t|t-1, \vartheta)$$

aggiungendo e sottraendo il predittore ad un passo del modello vero (vale a dire $\hat{y}(t|t-1, \vartheta^0)$) non cambia niente:

$$\varepsilon(t, \vartheta) = y(t) - \hat{y}(t|t-1, \vartheta) + \hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta^0)$$

Si può però notare che $y(t) - \hat{y}(t|t-1, \vartheta^0)$ corrisponde all'errore di predizione a un passo del modello vero e quindi vale $e(t)$.

$$\varepsilon(t, \vartheta) = e(t) + \hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta)$$

Calcolando la varianza dell'errore, ossia $J(\vartheta)$, si ottiene:

$$\begin{aligned} J(\vartheta) &= \mathbb{E} [\varepsilon(t, \vartheta)^2] = \\ &= \mathbb{E} [(e(t) + \hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta))^2] = \\ &= \mathbb{E} [e(t)^2] + \mathbb{E} [(\hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta))^2] + 2 \cdot \mathbb{E} [e(t) \cdot (\hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta))] = \\ &= \mathbb{E} [e(t)^2] + \mathbb{E} [(\hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta))^2] + 2 \cdot \mathbb{E} [e(t) \cdot \hat{y}(t|t-1, \vartheta^0) - e(t) \cdot \hat{y}(t|t-1, \vartheta)] \end{aligned}$$

Dato che i predittori $\hat{y}(t|t-1, \vartheta^0)$ e $\hat{y}(t|t-1, \vartheta)$ sono entrambi ottimi vale la condizione necessaria di ottimalità per entrambi:

$$\begin{aligned}\mathbb{E}[e(t) \cdot \hat{y}(t|t-1, \vartheta^0)] &= 0 \\ \mathbb{E}[e(t) \cdot \hat{y}(t|t-1, \vartheta)] &= 0\end{aligned}$$

e quindi:

$$2 \cdot \mathbb{E}[e(t) \cdot \hat{y}(t|t-1, \vartheta^0) - e(t) \cdot \hat{y}(t|t-1, \vartheta)] = 0$$

e quindi:

$$\begin{aligned}J(\vartheta) &= \mathbb{E}[e(t)^2] + \mathbb{E}[(\hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta))^2] \\ &= \lambda^2 + \mathbb{E}[(\hat{y}(t|t-1, \vartheta^0) - \hat{y}(t|t-1, \vartheta))^2]\end{aligned}$$

Dato che un quadrato è sicuramente positivo si ha che:

$$J(\vartheta) \geq \lambda^2$$

Dato che il metodo PEM minimizza $J(\vartheta)$ bisogna trovare il valore di ϑ tale per cui $J(\vartheta) = \lambda^2$ e questo avviene nel caso in cui

$$\hat{y}(t|t-1, \vartheta^0) = \hat{y}(t|t-1, \vartheta)$$

e quindi quando:

$$\vartheta = \vartheta^0$$

in conclusione si ha:

$$\forall \vartheta \neq \vartheta^0, J(\vartheta) \geq J(\vartheta^0)$$

Che è esattamente la proprietà che ϑ^0 doveva rispettare per appartenere a Δ . Si ha quindi che $\vartheta^0 \in \Delta$, ovvero i metodi PEM sono in grado di stimare i parametri veri del modello per $N \rightarrow \infty$ (sono quindi asintoticamente corretti). Inoltre, dato che si minimizza la varianza dell'errore, i metodi PEM sono anche asintoticamente consistenti.

Se si è scelto la classe di modelli sbagliata cade l'ipotesi iniziale e quindi i metodi PEM non convergono sui parametri veri.

Conclusione Secondo l'analisi asintotica appena fatta ci si può trovare in situazioni diverse, elencate di seguito:

- $S \in M(\vartheta)$ e $\Delta \equiv \bar{\vartheta} \equiv \vartheta^0$, ovvero la scelta della famiglia di modelli è stata quella giusta e la cifra di merito ha un solo minimo assoluto, ϑ^0 . In questo caso man mano che il numero di dati aumenta, $\hat{\vartheta}_N \rightarrow \vartheta^0$, con garanzia di convergenza. Una rappresentazione grafica di questa situazione è presentata in figura 19.1.

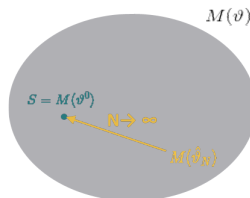


Figura 19.1.: Situazione 1: $S \in M(\vartheta)$ e un solo minimo assoluto della cifra di merito

- $S \in M(\vartheta)$ ma Δ è costituito da più valori, ovvero la scelta della famiglia di modelli è ancora giusta, ma la cifra di merito ha più di un minimo assoluto. In questo caso il metodo converge a un qualsiasi minimo, ma questo non presenta un problema in quanto trovare un'altro minimo assoluto significa trovare una serie di parametri che minimizza la cifra di merito tanto quanto la minimizza il set di parametri ϑ^0 (in parole povere il modello è equivalente dal punto di vista dell'identificazione PEM). Una rappresentazione grafica di questa situazione è presentata in figura 19.2.

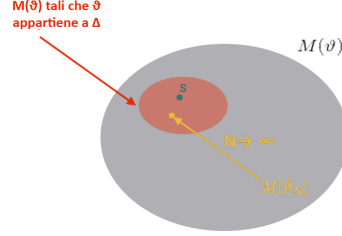


Figura 19.2.: Situazione 2: $S \in M(\vartheta)$ ma più minimi assoluti della cifra di merito

- $S \notin M(\vartheta)$ e $\Delta \equiv \bar{\vartheta}$, ovvero la scelta della famiglia di modelli è sbagliata; Questo significa che la migliore stima che si potrà raggiungere sarà $\bar{\vartheta}$, e ϑ^0 non si potrà mai ottenere, ma questo non è un problema grave perchè molte volte si vuole approssimare un problema con un determinato modello e anche se il sistema in realtà è più complesso si riesce a trovare la miglior approssimazione possibile usando il modello voluto. Una rappresentazione grafica di questa situazione è presentata in figura 19.3.

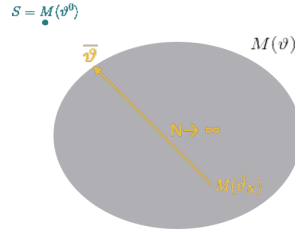


Figura 19.3.: Situazione 3: $S \notin M(\vartheta)$ ma è presente un solo minimo assoluto della cifra di merito

- $S \notin M(\vartheta)$ e Δ è costituito da più valori, in cui la scelta della famiglia di modelli è ancora sbagliata e inoltre sono presenti più minimi assoluti della cifra di merito. La situazione non cambia molto rispetto al punto precedente, in quanto ϑ^0 non potrà comunque essere raggiunto e una qualsiasi $\vartheta^* \in \Delta$ fornirà la stima dei parametri migliore che si può raggiungere (senza cambiare famiglia di modelli). Una rappresentazione grafica di questa situazione è presentata in figura 19.4.

19. Analisi asintotica dei metodi PEM

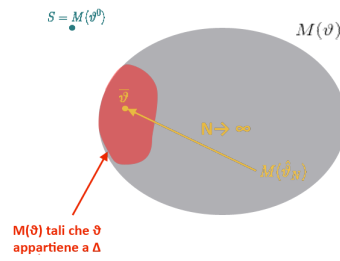


Figura 19.4.: Situazione 4: $S \notin M(\vartheta)$ e ci sono più minimi assoluti della cifra di merito

20. Problema di identificabilità

Nella sezione precedente si è visto che i metodo PEM sono asintoticamente corretti, ma non si è considerato il metodo attraverso il quale si calcola questa stima, il quale spesso comporta problemi che non permettono di arrivare ai valori voluti. Per capire meglio il problema si consideri un generico modello ARX :

$$y(t) = \frac{B(z)}{A(z)}u(t-1) + \frac{1}{A(z)} \cdot e(t)$$

con $e(t) \sim WN(0, \lambda^2)$

Si è visto precedentemente che la stima dei parametri per un generico modello ARX è:

$$\hat{\vartheta}_N = \left(\sum_{t=1}^N \varphi(t) \cdot \varphi(t)^T \right)^{-1} \cdot \sum_{t=1}^N \varphi(t) \cdot y(t)$$

Si definiscono le matrici $S(N)$ e $R(N)$ come:

$$S(N) = \sum_{t=1}^N \varphi(t) \cdot \varphi(t)^T$$

$$R(N) = \frac{1}{N} \cdot S(N) = \frac{1}{N} \cdot \sum_{t=1}^N \varphi(t) \cdot \varphi(t)^T$$

e quindi

$$\begin{aligned} \hat{\vartheta}_N &= S(N)^{-1} \cdot \sum_{t=1}^N \varphi(t) \cdot y(t) \\ \hat{\vartheta}_N &= \left(\sum_{t=1}^N \varphi(t) \cdot \varphi(t)^T \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{t=1}^N \varphi(t) \cdot y(t) = \\ &= \left(\frac{1}{N} \cdot \sum_{t=1}^N \varphi(t) \cdot \varphi(t)^T \right)^{-1} \cdot \frac{1}{N} \cdot \sum_{t=1}^N \varphi(t) \cdot y(t) = \\ &= R(N)^{-1} \cdot \frac{1}{N} \cdot \left(\sum_{t=1}^N \varphi(t) \cdot y(t) \right) \end{aligned}$$

Perchè $\hat{\vartheta}_N$ sia calcolabile in questo modo la matrice $R(N)$ deve essere invertibile e quindi non avere autovalori nulli, ma questo è un problema del modo attraverso il quale si calcola la stima e non del metodo PEM che rimane comunque asintoticamente corretto.

Analizzando la matrice $R(N)$, si può vedere che è il prodotto di un vettore per se stesso, il che la rende una matrice semidefinita positiva e quindi può avere solo autovalori nulli e positivi. Per garantire che la matrice sia effettivamente definita positiva è necessario dimostrare che abbia solo autovalori positivi.

Per semplicità si considera il problema in regime asintotico e quindi per $N \rightarrow \infty$, quindi si ha che:

$$R(N) \rightarrow \bar{R} = \left[\begin{array}{c|c} \bar{R}_y & -\bar{R}_{yu} \\ \hline -\bar{R}_{uy} & \bar{R}_u \end{array} \right]$$

Per capire meglio a cosa corrispondono i quattro blocchi si consideri i seguenti esempi:

- $AR(1)$: ossia con il vettore delle osservazioni è pari a $\varphi(t) = y(t-1)$ ottenendo:

$$\begin{aligned} R(N) &= \frac{1}{N} \cdot \sum_{t=1}^N \varphi(t) \cdot \varphi(t)^T = \\ &= \frac{1}{N} \cdot \sum_{t=1}^N y(t-1)^2 \end{aligned}$$

dato che $N \rightarrow \infty$ e lo stimatore della covarianza è asintoticamente corretto si ha che:

$$\bar{R} = \mathbb{E} [y(t-1)^2] = \gamma_y(0)$$

- $ARX(1,1)$: in questo caso il vettore delle osservazioni è pari a $\varphi(t) = \begin{bmatrix} -y(t-1) & u(t-1) \end{bmatrix}^T$ ottenendo:

$$\begin{aligned} R(N) &= \frac{1}{N} \cdot \sum_{t=1}^N \varphi(t) \cdot \varphi(t)^T = \\ &= \frac{1}{N} \cdot \sum_{t=1}^N \cdot \begin{bmatrix} -y(t-1) \\ u(t-1) \end{bmatrix} \begin{bmatrix} -y(t-1) & u(t-1) \end{bmatrix} = \\ &= \frac{1}{N} \cdot \sum_{t=1}^N \cdot \begin{bmatrix} y(t-1)^2 & -y(t-1) \cdot u(t-1) \\ -y(t-1) \cdot u(t-1) & u(t-1)^2 \end{bmatrix} \end{aligned}$$

dato che $N \rightarrow \infty$ e lo stimatore della covarianza è asintoticamente corretto si ha che:

$$\bar{R} = \begin{bmatrix} \mathbb{E} [y(t-1)^2] & -\mathbb{E} [u(t-1) \cdot y(t-1)] \\ -\mathbb{E} [u(t-1) \cdot y(t-1)] & \mathbb{E} [u(t-1)^2] \end{bmatrix}$$

definendo:

$$\gamma_{uy}(\tau) = \mathbb{E} [u(t) \cdot y(t+\tau)]$$

si ha che:

$$\bar{R} = \begin{bmatrix} \gamma_y(0) & -\gamma_{yu}(0) \\ -\gamma_{yu}(0) & -\gamma_u(0) \end{bmatrix}$$

In generale si ha che:

$$\bar{R}_y = \begin{bmatrix} \gamma_y(0) & \gamma_y(1) & \gamma_y(2) & \cdots & \cdots & \gamma_y(m-1) \\ \gamma_y(1) & \gamma_y(0) & \gamma_y(1) & \gamma_y(2) & \cdots & \gamma_y(m-2) \\ \gamma_y(2) & \gamma_y(1) & \gamma_y(0) & \gamma_y(1) & \cdots & \gamma_y(m-3) \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \gamma_y(m-1) & \gamma_y(m-2) & \cdots & \cdots & \cdots & \gamma_y(0) \end{bmatrix}$$

che è la matrice di covarianza di $y(t)$ di ordine m , ossia una matrice $m \times m$ di Toeplitz¹ con $\gamma_y(0)$ sulla diagonale principale, e man mano che ci si sposta verso l'esterno aumenta la τ .

$$\bar{R}_u = \begin{bmatrix} \gamma_u(0) & \gamma_u(1) & \gamma_u(2) & \cdots & \cdots & \gamma_u(p) \\ \gamma_u(1) & \gamma_u(0) & \gamma_u(1) & \gamma_u(2) & \cdots & \gamma_u(p-1) \\ \gamma_u(2) & \gamma_u(1) & \gamma_u(0) & \gamma_u(1) & \cdots & \gamma_u(p-2) \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \gamma_u(p) & \gamma_u(p-1) & \cdots & \cdots & \cdots & \gamma_u(0) \end{bmatrix}$$

¹Le matrici di Toeplitz sono matrici in cui i valori sulle diagonali parallele alla principale sono costanti

20. Problema di identificabilità

che è la matrice di covarianza di $u(t)$ di ordine $p+1$, ossia una matrice $(p+1) \times (p+1)$ di Toeplitz con $\gamma_u(0)$ sulla diagonale principale, e man mano che ci si sposta verso l'esterno aumenta la τ .

$$\bar{R}_{yu} = \begin{bmatrix} \gamma_{yu}(0) & \gamma_{yu}(1) & \gamma_{yu}(2) & \cdots & \cdots & \gamma_{yu}(p) \\ \gamma_{yu}(1) & \gamma_{yu}(0) & \gamma_{yu}(1) & \gamma_{yu}(2) & \cdots & \gamma_{yu}(p-1) \\ \gamma_{yu}(2) & \gamma_{yu}(1) & \gamma_{yu}(0) & \gamma_{yu}(1) & \cdots & \gamma_{yu}(p-2) \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \gamma_{yu}(m-1) & \gamma_{yu}(m-2) & \cdots & \cdots & \cdots & \gamma_{yu}(0) \end{bmatrix}$$

che è la matrice di covarianza tra $y(t)$ e $u(t)$ di ordine m e $p+1$, ossia una matrice $m \times (p+1)$ e infine $\bar{R}_{uy} = (\bar{R}_{yu})^T$.
Quindi la matrice $R(N)$ è una matrice quadrata simmetrica $(m+p+1) \times (m+p+1)$.

Lemma di Schur Data una matrice M nella forma:

$$M = \left[\begin{array}{c|c} F & K \\ \hline K^T & H \end{array} \right]$$

con F e H simmetriche, M è definita positiva se e solo se:

- H è definita positiva
- $F - K \cdot H \cdot K^T$ è definita positiva

Osservazione

- La seconda condizione del lemma coinvolge tutte le matrici e quindi è una condizione molto complicata da valutare
- La prima condizione è una proprietà della matrice H e quindi della $R_u(N)$, che dipende solo dal valore dell'ingresso, che è una grandezza che viene decisa in fase di progettazione. Per questo motivo è importante scegliere un giusto valore dell'ingresso.

Definizione Si definisce la matrice di autocorrelazione $R_u^{(i)}$ di ordine i :

$$R_u^{(i)} = \begin{bmatrix} \gamma_u(0) & \gamma_u(1) & \gamma_u(2) & \cdots & \cdots & \gamma_u(i-1) \\ \gamma_u(1) & \gamma_u(0) & \gamma_u(1) & \gamma_u(2) & \cdots & \gamma_u(i-2) \\ \gamma_u(2) & \gamma_u(1) & \gamma_u(0) & \gamma_u(1) & \cdots & \gamma_u(i-3) \\ \vdots & & & \ddots & & \vdots \\ \vdots & & & & \ddots & \vdots \\ \gamma_u(i-1) & \gamma_u(i-2) & \cdots & \cdots & \cdots & \gamma_u(0) \end{bmatrix}$$

Da cui si nota che:

$$R(N) = R_u^{(p+1)}$$

Definizione Un segnale $u(t)$ si definisce segnale persistentemente eccitante di ordine N se:

- $R_u^{(1)}, R_u^{(2)}, \dots, R_u^{(N)}$ sono definite positive
- $R_u^{(N+1)}, R_u^{(N+2)}, \dots, R_u^{(N+k)}$ sono semidefinite positive

Overo N risulta essere l'ordine massimo per cui $R_u^{(i)}$ è invertibile.

Conclusione Per permettere l'identificazione tramite metodo PEM la matrice $R(N)$ deve essere invertibile e quindi definita positiva. Perchè questo avvenga bisogna garantire:

- Identificabilità strutturale
Il modello non deve essere sovradimensionato rispetto al sistema, questo problema verrà approfondito nella parte sulla validazione dove si analizzerà il modo per stimare l'ordine del modello corretto.
- Identificabilità sperimentale
Il segnale d'ingresso $u(t)$ deve essere scelto in modo che i dati contengano abbastanza informazione, questo vincolo viene formalizzato dal lemma di Schur in cui si esplicita che una condizione necessaria è che il segnale d'ingresso sia persistentemente eccitante almeno di ordine $p + 1$, questa è la base della disciplina dell'input design.

Osservazione Dato $u(t) \sim WN(0, \lambda^2)$:

$$\overline{R}_u^{(i)} = \begin{bmatrix} \lambda^2 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \lambda^2 & 0 & & \cdots & 0 \\ 0 & 0 & \lambda^2 & & \cdots & 0 \\ \vdots & & & \ddots & & \vdots \\ \vdots & \vdots & \vdots & & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \cdots & \lambda^2 \end{bmatrix} = \lambda^2 \cdot I_i$$

Ovvero sarà una matrice diagonale con valori sulla diagonale positivi $\forall i$. Questo significa che un WN è un segnale persistentemente eccitante di ordine ∞ . Nonostante un rumore bianco sia un segnale che rispetta sempre la condizione necessaria del lemma di Schur, la disciplina dell'input design studia per cercare segnali che, oltre a rispettare la condizione necessaria, minimizzano la varianza di predizione a pari dati. Questo perchè il rumore bianco ideale non è un segnale riproducibile; si usano invece altri tipi di segnali, come la PseudoRandom Binary Sequence (PRBS), il cui spettro è abbastanza "piatto", similmente a quello del rumore bianco ideale.

21. Valutazione dell'incertezza

L'analisi fatta nel capitolo precedente si basava sull'ipotesi che $N \rightarrow \infty$. Purtroppo non è possibile applicare questo approccio a problemi reali, in quanto si dovrà lavorare con un set finito di dati.

Partendo dalle ipotesi:

- Se $S \in M(\vartheta)$ allora $\vartheta^0 \in \Delta$
- Se $\Delta \equiv \bar{\vartheta}$ allora $\bar{\vartheta} = \vartheta^0$

Si può ricavare che la varianza dello stimatore $\hat{\vartheta}$ vale:

$$Var[\hat{\vartheta}] = \frac{1}{N} \cdot \lambda^2 \cdot \bar{C}^{-1}$$

dove:

$$\lambda^2 = \mathbb{E}[e(t)^2] = \mathbb{E}[\varepsilon(t, \vartheta^0)^2] = \mathbb{E}[(y(t) - \hat{y}(t|t-1))^2]$$

$$\bar{C} = \mathbb{E} \left[\left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\vartheta^0} \right) \cdot \left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\vartheta^0} \right)^T \right]$$

Per poter calcolare queste misure servirebbe di conoscere ϑ^0 , ma si può calcolare una stima utilizzando le stime campionarie, ossia:

$$\begin{aligned} \lambda^2 &= \mathbb{E}[\varepsilon(t, \vartheta^0)] \cong \\ &\cong \mathbb{E}[\varepsilon(t, \hat{\vartheta})] \cong \\ &\cong \frac{1}{N} \cdot \sum_{t=1}^{t=N} \varepsilon(t, \hat{\vartheta})^2 = \\ &= J_N(\hat{\vartheta}) \end{aligned}$$

$$\begin{aligned} \bar{C} &= \mathbb{E} \left[\left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\vartheta^0} \right) \cdot \left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\vartheta^0} \right)^T \right] \cong \\ &\cong \mathbb{E} \left[\left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\hat{\vartheta}} \right) \cdot \left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\hat{\vartheta}} \right)^T \right] \cong \\ &\cong \frac{1}{N} \cdot \sum_{t=1}^{t=N} \left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\hat{\vartheta}} \right) \cdot \left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta} \bigg|_{\hat{\vartheta}} \right)^T \end{aligned}$$

Osservazione Per comprendere meglio il significato di \bar{C} si ricava l'hessiana della varianza dell'errore di predizione, ossia di $J(\vartheta)$ ottenendo:

$$\begin{aligned}\frac{\partial J(\vartheta)}{\partial \vartheta} &= \frac{\partial \mathbb{E}[\varepsilon(t, \vartheta)^2]}{\partial \vartheta} = \\ &= \mathbb{E}\left[\frac{\partial \varepsilon(t, \vartheta)^2}{\partial \vartheta}\right] = \\ &= \mathbb{E}\left[2 \cdot \varepsilon(t, \vartheta) \cdot \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right]\end{aligned}$$

$$\begin{aligned}\frac{\partial^2 J(\vartheta)}{\partial \vartheta^2} &= \frac{\partial}{\partial \vartheta} \mathbb{E}\left[2 \cdot \varepsilon(t, \vartheta) \cdot \frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right] = \\ &= \mathbb{E}\left[2 \cdot \left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right) \cdot \left(\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right)^T + \varepsilon(t, \vartheta) \cdot \frac{\partial^2 \varepsilon(t, \vartheta)}{\partial \vartheta^2}\right]\end{aligned}$$

Valutando in $\vartheta = \vartheta^0$ si ottiene:

$$\begin{aligned}\left.\frac{\partial J(\vartheta)}{\partial \vartheta}\right|_{\vartheta^0} &= \mathbb{E}\left[2 \cdot \left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right) \cdot \left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right)^T + \varepsilon(t, \vartheta^0) \cdot \left(\left.\frac{\partial^2 \varepsilon(t, \vartheta)}{\partial \vartheta^2}\right|_{\vartheta^0}\right)\right] = \\ &= \mathbb{E}\left[2 \cdot \left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right) \cdot \left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right)^T + e(t) \cdot \left(\left.\frac{\partial^2 \varepsilon(t, \vartheta)}{\partial \vartheta^2}\right|_{\vartheta^0}\right)\right]\end{aligned}$$

dato che $\left.\frac{\partial^2 \varepsilon(t, \vartheta)}{\partial \vartheta^2}\right|_{\vartheta^0}$ dipende dall'errore di predizione sarà costituito dai termini passati del rumore bianco e quindi il secondo termine è nullo e quindi:

$$\begin{aligned}\left.\frac{\partial J(\vartheta)}{\partial \vartheta}\right|_{\vartheta^0} &= \mathbb{E}\left[2 \cdot \left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right) \cdot \left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right)^T + e(t) \cdot \left(\left.\frac{\partial^2 \varepsilon(t, \vartheta)}{\partial \vartheta^2}\right|_{\vartheta^0}\right)\right] = \\ &= 2 \cdot \mathbb{E}\left[\left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right) \cdot \left(\left.\frac{\partial \varepsilon(t, \vartheta)}{\partial \vartheta}\right|_{\vartheta^0}\right)^T\right] = \\ &= 2 \cdot \bar{C}\end{aligned}$$

quindi:

$$\bar{C} = \frac{1}{2} \cdot \left.\frac{\partial J(\vartheta)}{\partial \vartheta}\right|_{\vartheta^0}$$

ossia \bar{C} è la metà dell'hessiano calcolato nel punto di ottimo.

Conclusione sapendo che:

$$Var[\hat{\vartheta}] = \frac{1}{N} \cdot \lambda^2 \cdot \bar{C}^{-1} = \frac{1}{N} \cdot \lambda^2 \cdot \frac{1}{2} \cdot \left.\frac{\partial J(\vartheta)}{\partial \vartheta}\right|_{\vartheta^0}$$

si nota che:

21. Valutazione dell'incertezza

- $N \uparrow \Rightarrow Var [\hat{\vartheta}] \downarrow$, quindi aumentando i campioni si diminuisce la varianza della stima.
- $\lambda^2 \uparrow \Rightarrow Var [\hat{\vartheta}] \uparrow$, quindi più i dati sono rumorosi meno la stima è affidabile.
- $\overline{C} \uparrow \Rightarrow Var [\hat{\vartheta}] \downarrow$, quindi si vorrebbe una hessiana nel punto di minimo più grossa possibile, questo significa che si vuole avere una grossa variazione nell'intorno del minimo. come mostrato in figura 21.1.

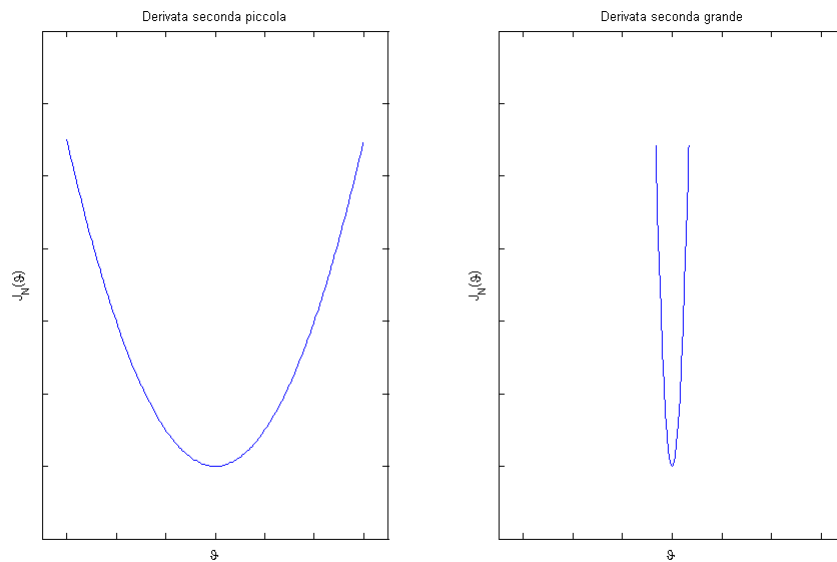


Figura 21.1.: Rappresentazione Hessiane piccole e grandi

22. Scelta della complessità del modello

Introduzione

Tutte le analisi viste finora hanno dato per scontata la scelta dei parametri strutturali del modello: m, n, p e k di un modello $ARMAX(m, n, p + k)$. Analizzandoli in dettaglio:

- k è il ritardo puro tra ingresso e uscita, come già visto prima questo parametro viene stimato automaticamente nella stima PEM se si ipotizza $k = 1$, infatti se k fosse maggiore di 1 si stimerebbero i primi coefficiente di $B(z)$ nulli.
- m, n e p sono i parametri del modello $ARMAX$, si può considerare $m \approx n \approx p$, infatti se fossero diversi si potrebbe porli pari a $\max(m, n, p)$ in questo modo i parametri di grado maggiore uscirebbero nulli ottenendo lo stesso sistema che si otterrebbe ponendo i 3 parametri al giusto valore. Per illustrare meglio il concetto se si vuole identificare un sistema $S = ARMAX(1, 1, 2)$ rappresentato dalla seguente equazione:

$$y(t) = \frac{1 - \frac{1}{2} \cdot z^{-1}}{1 + \frac{1}{3} z^{-1}} \cdot e(t) + \frac{1 + \frac{1}{2} \cdot z^{-1} - \frac{1}{4} \cdot z^{-2}}{1 + \frac{1}{3} z^{-1}} \cdot u(t-1)$$

Seguendo l'ipotesi fatta precedentemente, si cercherebbe di identificare il sistema S con un sistema $Q = ARMAX(2, 2, 2)$. Dopo aver eseguito la procedura di identificazione, si potrà trovare che il sistema verrà identificato come

$$y_Q(t) = \frac{1 - \frac{1}{2} \cdot z^{-1} + 0 \cdot z^{-2}}{1 + \frac{1}{3} z^{-1} + 0 \cdot z^{-2}} \cdot e(t) + \frac{1 + \frac{1}{2} \cdot z^{-1} - \frac{1}{4} \cdot z^{-2}}{1 + \frac{1}{3} z^{-1} + 0 \cdot z^{-2}} \cdot u(t-1)$$

che è un sistema equivalente.

Quindi i 4 parametri si riducono a un solo parametro: $n_\vartheta = n + m + p + 1$.

Si può notare che prendendo $n_\vartheta = N$ l'identificazione PEM si limita a interpolare tutti i dati e quindi si ottiene $J_N(\hat{\vartheta}_N, n_\vartheta) = 0$, in quanto il modello spiegherà alla perfezione tutti i dati, ma in questo modo si insegue una specifica realizzazione del rumore che è completamente scorrelata da tutte le altre realizzazioni. Questo fenomeno prende il nome di overfitting e bisogna evitarlo.

Dall'osservazione di prima si nota che la funzione è decrescente e arriva a 0 quando $n_\vartheta = N$:

$$J_N(\hat{\vartheta}_N, n_\vartheta) \xrightarrow{n_\vartheta \rightarrow \infty} 0$$

quindi la cifra di merito non è un buon indice per decidere il valore di n_ϑ , dato che minimizzando si ottiene overfitting.

Spiegando in altre parole, incrementando il numero di parametri:

- Inizialmente si migliora la spiegazione della fisica del sistema, migliorando la stima del modello e quindi abbassando il valore della cifra di merito;
- In una seconda parte il modello è completamente spiegato e quindi si inizia a cercare di spiegare il rumore inseguendo abbassando il valore della cifra di merito.

Generalmente si vuole usare il valore al confine tra le due zone, ma dato che in entrambe la cifra di merito diminuisce, è impossibile trovarlo usando solamente la cifra di merito. Nella prima zona si dice che si fa underfitting mentre nella seconda overfitting.

22.1. Cross-validazione

La cross-validazione è un metodo per stimare il numero di parametri di un modello. Questo metodo chiede di avere due set di dati $D_N^{(i)}$ (dati di identificazione) e $D_N^{(v)}$ (dati di validazione), quindi solitamente si divide il set di dati in due parti di stessa grandezza. Dopo di che si procede ripetendo la seguente procedura per ogni valore di $n_\vartheta = 1, 2, 3, 4, \dots$:

1. Si ricava la stima $\hat{\vartheta}$ minimizzando la cifra di merito $J_N(\vartheta, n_\vartheta)$ usando i dati del set $D_N^{(i)}$
2. Si ricavano i valori:
 - $J_N(\hat{\vartheta}, n_\vartheta)$ usando i dati del set $D_N^{(i)}$,
 - $J_N(\hat{\vartheta}, n_\vartheta)$ usando i dati del set $D_N^{(v)}$
3. Si analizzano i dati ottenuti.

Un possibile grafico delle due successioni di valori ottenute nel passo 2 è rappresentato in figura 22.1.

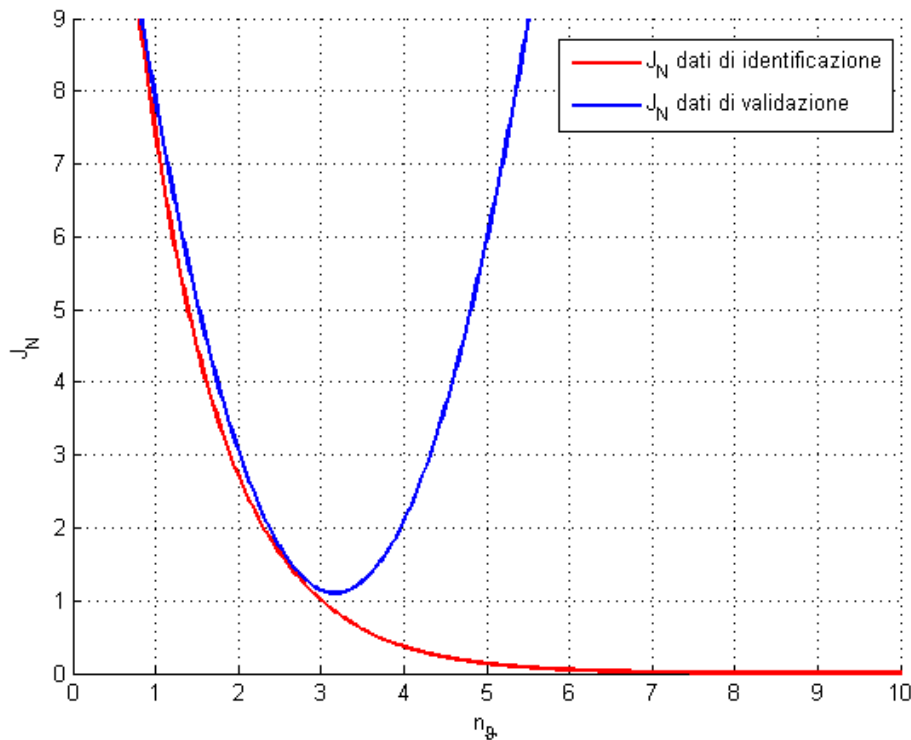


Figura 22.1.: Rappresentazione dell'andamento delle cifra di merito calcolate sui dati di identificazione e di validazione

Dal grafico si può notare che nella prima parte, dove si fa underfitting e si cerca di spiegare la fisica, le due cifre di merito hanno un valore simile, mentre nella seconda parte, dove si fa overfitting e si insegue il rumore, la cifra di merito ottenuta dai dati della validazione inizia ad aumentare. Questo è dovuto al fatto che quando si fa overfitting si insegue la realizzazione del rumore che ha generato i primo set di dati che è scorrelato dalla realizzazione del rumore che ha generato il secondo.

Minimizzando la seconda funzione si può ricavare il punto in cui si finisce di fare underfitting e si inizia a fare overfitting risolvendo il problema di stima di n_ϑ .

La cross-validazione è un metodo molto affidabile che funziona nella quasi totalità dei casi, compresi quelli non lineari, ma ha il problema che richiede molti dati.

22.2. Formule di stima della complessità ottima

Questo metodo consiste nell'utilizzare delle cifre di merito diverse che penalizzano i modelli con n_ϑ in modo che la cifra di merito non tenda a 0, ma dopo un certo punto inizi a crescere e quindi divergere all'infinito. Con queste cifre di merito si può minimizzare direttamente la cifra di merito calcolata usando il set di dati usati nell'identificazione e quindi richiede meno dati rispetto alla cross-validazione, ma è meno affidabile.

Esistono diverse tecniche che usano questo concetto, ad esempio:

- FPE (Final Prediction Error)

$$FPE(n_\vartheta) = \frac{N + n_\vartheta}{N - n_\vartheta} \cdot J_N(\hat{\vartheta}_N, n_\vartheta)$$

Si può notare che:

$$\begin{aligned} n_\vartheta \uparrow &\Rightarrow \frac{N + n_\vartheta}{N - n_\vartheta} \uparrow \\ n_\vartheta \uparrow &\Rightarrow J_N(\hat{\vartheta}_N, n_\vartheta) \downarrow \end{aligned}$$

Quindi per n_ϑ grandi, la cifra di merito è "alzata" dal fattore $\frac{N + n_\vartheta}{N - n_\vartheta}$, penalizzando appunto gli ordini grandi del sistema (e di conseguenza, la grande probabilità di overfitting).

- AIC (Akaike Information Criteria)

$$AIC(n_\vartheta) = 2 \cdot \frac{n_\vartheta}{N} + \ln \left(J_N(\hat{\vartheta}_N, n_\vartheta) \right)$$

In cui si ha che:

$$\begin{aligned} n_\vartheta \uparrow &\Rightarrow 2 \cdot \frac{n_\vartheta}{N} \uparrow \\ n_\vartheta \uparrow &\Rightarrow \ln \left(J_N(\hat{\vartheta}_N, n_\vartheta) \right) \downarrow \end{aligned}$$

Il discorso è quindi identico a quello fatto per il criterio precedente.

- MDL (Minimum Description Length)

$$MDL(n_\vartheta) = \ln(N) \cdot \frac{n_\vartheta}{N} + \ln \left(J_N(\hat{\vartheta}_N, n_\vartheta) \right)$$

Come prima:

$$\begin{aligned} n_\vartheta \uparrow &\Rightarrow \ln(N) \cdot \frac{n_\vartheta}{N} \uparrow \\ n_\vartheta \uparrow &\Rightarrow \ln \left(J_N(\hat{\vartheta}_N, n_\vartheta) \right) \downarrow \end{aligned}$$

Questi tre criteri non sono molto diversi infatti se si calcola il logaritmo di FPE si ottiene:

$$\begin{aligned} \ln(FPE(n_\vartheta)) &= \ln \left(\frac{N + n_\vartheta}{N - n_\vartheta} \cdot J_N(\hat{\vartheta}_N, n_\vartheta) \right) = \\ &= \ln \left(\frac{1 + \frac{n_\vartheta}{N}}{1 - \frac{n_\vartheta}{N}} \cdot J_N(\hat{\vartheta}_N, n_\vartheta) \right) = \\ &= \ln \left(1 + \frac{n_\vartheta}{N} \right) - \ln \left(1 - \frac{n_\vartheta}{N} \right) + \ln \left(J_N(\hat{\vartheta}_N, n_\vartheta) \right) \end{aligned}$$

22. Scelta della complessità del modello

Ricordando che $\ln(x+1) \approx x$ quando $x \approx 0$ e che $\frac{n_\vartheta}{N} \approx 0$, in quanto si cerca di evitare l'overfitting, si ha che:

$$\begin{aligned}
 \ln(FPE(n_\vartheta)) &= \ln\left(1 + \frac{n_\vartheta}{N}\right) - \ln\left(1 - \frac{n_\vartheta}{N}\right) + \ln\left(J_N(\hat{\vartheta}_N, n_\vartheta)\right) = \\
 &= \frac{n_\vartheta}{N} - \left(-\frac{n_\vartheta}{N}\right) + \ln\left(J_N(\hat{\vartheta}_N, n_\vartheta)\right) = \\
 &= \frac{n_\vartheta}{N} + \frac{n_\vartheta}{N} + \ln\left(J_N(\hat{\vartheta}_N, n_\vartheta)\right) = \\
 &= 2 \cdot \frac{n_\vartheta}{N} + \ln\left(J_N(\hat{\vartheta}_N, n_\vartheta)\right) = \\
 &= AIC(n_\vartheta)
 \end{aligned}$$

Quindi dato che se x_0 è un minimo di $f(x)$ allora sarà minimo anche di $\ln(f(x_0))$ i due metodi sono diversi dal punto di vista computazionale, ma il risultato che si ottiene è lo stesso. Inoltre si può anche notare che MDL è uguale a AIC a meno di un coefficiente:

$$\begin{aligned}
 AIC(n_\vartheta) &= 2 \cdot \frac{n_\vartheta}{N} + \ln\left(J_N(\hat{\vartheta}_N, n_\vartheta)\right) \\
 MDL(n_\vartheta) &= \ln(N) \cdot \frac{n_\vartheta}{N} + \ln\left(J_N(\hat{\vartheta}_N, n_\vartheta)\right)
 \end{aligned}$$

Se $\ln(N) > 2$ allora MDL penalizza di più all'aumentare del modello rispetto ad AIC e quindi diminuisce il rischio di fare overfitting aumentando quello di fare underfitting, che è comunque molto meglio. Si può notare che questa condizione è verificata per $N > e^2$ e quindi, dato che N è intero, $N \geq 8$ e quindi è praticamente sempre verificata.

In conclusione è possibile utilizzare i criteri AIC e FPE nel caso di sicurezza nella scelta della famiglia di modelli, dove il rischio di fare overfitting è minore rispetto al caso in cui la scelta della famiglia di modelli è sbagliata. Nel caso non si conosca nulla sul sistema, è preferibile affidarsi alla maggiore robustezza del criterio MDL che penalizzando maggiormente diminuisce la probabilità di fare overfitting, aumentando quella di fare underfitting.

Parte V.

Identificazione segnali non misurabili

Introduzione

Si ha un sistema S di cui si dispongono m misure y_i provenienti da alcuni sensori, ma si sa che il sistema dipende anche da una variabile x di cui non si conosce la misura, come mostrato in figura .

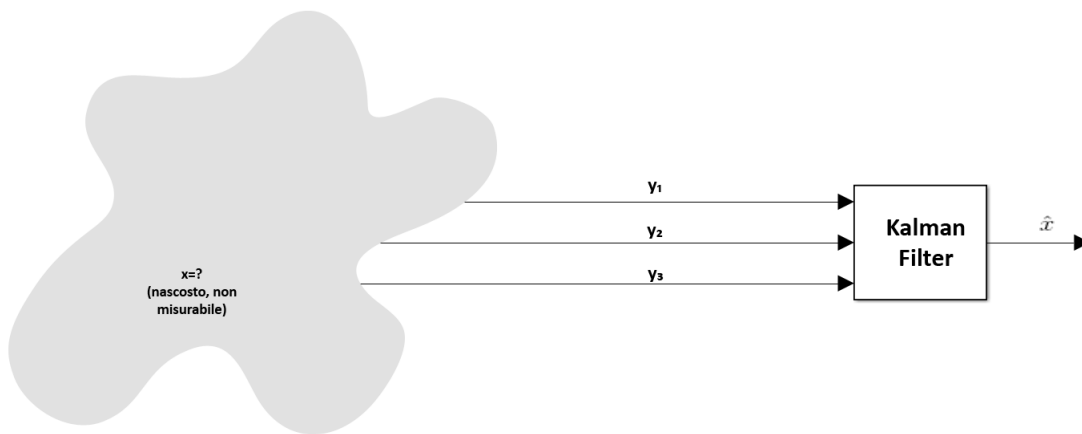


Figura 22.2.: Problema di identificazione di segnali non misurabili risolto con un filtro di Kalman

Si vuole stimare nel modo migliore possibile la variabile x . Questo problema è viene definito problema di virtual sensing perchè consiste nel sostituire un sensore hardware con un algoritmo software.

Esistono varie ragioni per cui il virtual sensing è molto usato:

- Permette di risparmiare soldi nei dispositivi prodotti in massa dato che può evitare l'inserimento di un sensore hardware, che può essere anche costoso.
- In situazioni critiche un sensore virtuale può sostituire un vero sensore che si è rotto, aumentando l'affidabilità del prodotto. Inoltre paragonando il valore del sensore virtuale con quello del vero sensore si può capire se c'è stato un guasto o comunque se ci si trova in una situazione anomala.
- Esistono applicazioni in cui non esiste ancora la tecnologia per produrre quel determinato sensore, capita per lo più per misure in condizioni ambientali critiche (alta/bassa temperatura o pressione, presenza di radiazione, presenza di disturbi elettromagnetici...)

L'algoritmo più usato è il filtro di Kalman e verrà spiegato in seguito, ma non è l'unico presente in letteratura.

23. Rappresentazione tramite variabili di stato

Fino a questo momento si è sempre rappresentato un sistema tramite un modello *ARMAX*:

$$y(t) = \frac{C(z)}{A(z)} \cdot e(t) + \frac{B(z)}{A(z)} \cdot u(t-k) \text{ con } e(t) \sim WN(\mu, \lambda^2)$$

Esiste una rappresentazione leggermente più ricca che è la rappresentazione tramite variabili di stato. Ipotizzando un sistema MIMO con m uscite y_i , n variabili di stato x_i e p variabili di ingresso u_i si può scrivere:

$$\begin{cases} x(t+1) = F \cdot x(t) + \Gamma \cdot u(t) + v_1(t) & \text{con } v_1(t) \sim WN(0, V_1) \\ y(t) = H \cdot x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

dove:

- $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$ è il vettore colonna delle variabili di stato e quindi è una matrice $n \times 1$;

- $y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$ è il vettore colonna delle variabili di uscita e quindi è una matrice $m \times 1$;

- $u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_p \end{bmatrix}$ è il vettore colonna delle variabili di uscita e quindi è una matrice $p \times 1$;

- F è una matrice $n \times n$, ed è la matrice di stato;

- H è una matrice $m \times n$, ed è la matrice di uscita;

- Γ è una matrice $n \times p$, ed è la matrice che porta l'ingresso sullo stato;

- $v_1 = \begin{bmatrix} v_{11} \\ v_{12} \\ \vdots \\ v_{1n} \end{bmatrix}$ è un vettore colonna di rumori bianchi, quindi variabili stocastiche, di dimensione $n \times 1$ (un rumore per

ogni variabile di stato), che sintetizza l'incertezza sul modello. Le caratteristiche fondamentali di questi errori sono:

- $\mathbb{E}[v_1(t)] = 0$

- $\mathbb{E}[v_1(t) \cdot v_1^T(t)] = V_1$, che è la matrice delle varianze-covarianze e quindi è simmetrica

- Se tutti gli errori sono scorellati tra di loro questa matrice è simmetrica

- Può capitare che non esista incertezza sul modello e quindi questa matrice risulterà solo semidefinita positiva, in quanto alcuni rumori avranno varianza nulla

23. Rappresentazione tramite variabili di stato

◦ $\mathbb{E} [v_1(t) \cdot v_1^T(t - \tau)] = 0, \forall t, \tau$, come tutti i rumori bianchi non sono correlati nel tempo

• $v_2 = \begin{bmatrix} v_{21} \\ v_{22} \\ \vdots \\ v_{2m} \end{bmatrix}$ è un vettore colonna di rumori bianchi, quindi variabili stocastiche, di dimensione $m \times 1$ (un rumore per ogni variabile di uscita), che sintetizza l'incertezza della misura. Le caratteristiche fondamentali di questi errori sono:

- $\mathbb{E} [v_2(t)] = 0$
- $\mathbb{E} [v_2(t) \cdot v_2^T(t)] = V_2$, che è la matrice delle varianze-covarianze e quindi è simmetrica
 - Se tutti gli errori sono scorellati tra di loro questa matrice è simmetrica
 - L'incertezza di misura è sempre presente e quindi V_2 è sempre definita positiva
- $\mathbb{E} [v_2(t) \cdot v_2^T(t - \tau)] = 0, \forall t, \tau$, come tutti i rumori bianchi non sono correlati nel tempo

Osservazioni:

- I rumori di modello v_1 e di misura v_2 non sono rappresentati direttamente in un modello *ARMAX* infatti entrambi vengono inglobati nello stesso rumore, per questo motivo questa rappresentazione è più ricca della precedente.
- Si può dimostrare che è sempre possibile passare tra le due rappresentazioni, ma è difficile da dimostrare. Inoltre questa trasformazione non è unica, ossia esistono più rappresentazioni tramite variabili di stato che corrispondono allo stesso *ARMAX* e viceversa.
Dato che la dimostrazione è complicata si esegue questa conversione in un esempio con una variabile di stato.

Esempio

$$\begin{cases} x(t+1) = \frac{1}{2} \cdot x(t) + \frac{1}{4} \cdot u(t) + v_1(t) & \text{con } v_1(t) \sim WN(0, 1) \\ y(t) = x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0, 1) \end{cases}$$

Passando al dominio operativo si ottiene:

$$\begin{cases} z \cdot x(t) = \frac{1}{2} \cdot x(t) + \frac{1}{4} \cdot u(t) + v_1(t) \\ y(t) = x(t) + v_2(t) \end{cases}$$

$$\begin{cases} \left(z - \frac{1}{2}\right) \cdot x(t) = \frac{1}{4} \cdot u(t) + v_1(t) \\ y(t) = x(t) + v_2(t) \end{cases}$$

$$\begin{cases} x(t) = \frac{\frac{1}{4}}{z - \frac{1}{2}} \cdot u(t) + \frac{1}{z - \frac{1}{2}} \cdot v_1(t) \\ y(t) = x(t) + v_2(t) \end{cases}$$

Quindi:

$$\begin{aligned}
 y(t) &= \frac{\frac{1}{4}}{z - \frac{1}{2}} \cdot u(t) + \frac{1}{z - \frac{1}{2}} \cdot v_1(t) + v_2(t) = \\
 &= \frac{\frac{1}{4}}{1 - \frac{1}{2} \cdot z^{-1}} \cdot z^{-1} \cdot u(t) + \frac{z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}} v_1(t) + v_2(t) = \\
 &= \underbrace{\frac{\frac{1}{4}}{1 - \frac{1}{2} \cdot z^{-1}} \cdot u(t-1)}_{\text{parte deterministica}} + \underbrace{\frac{z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}} v_1(t) + v_2(t)}_{\text{parte stocastica}}
 \end{aligned}$$

quindi facile notare che:

$$\frac{B(z)}{A(z)} = \frac{\frac{1}{4}}{1 - \frac{1}{2} \cdot z^{-1}}$$

e quindi:

$$\begin{aligned}
 B(z) &= \frac{1}{4} \\
 A(z) &= 1 - \frac{1}{2} \cdot z^{-1}
 \end{aligned}$$

la parte stocastica invece è più complessa perchè si deve passare da una somma di due rumori, uno bianco e uno colorato, a un unico rumore equivalente, per farlo bisogna calcolare lo spettro del rumore e trovare un rumore con lo stesso spettro, quindi chiamando:

$$w(t) = \frac{z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}} v_1(t) + v_2(t)$$

si ottiene:

$$\begin{aligned}
 \Gamma_w(\omega) &= \left| \frac{z^{-1}}{1 - 0.5 \cdot z^{-1}} \right|_{z=e^{j \cdot \omega}}^2 \cdot \Gamma_{v_1}(\omega) + \Gamma_{v_2}(\omega) = \\
 &= \left(\frac{z^{-1}}{1 - 0.5 \cdot z^{-1}} \cdot \frac{z^1}{1 - 0.5 \cdot z^1} \right) \Big|_{z=e^{j \cdot \omega}} \cdot 1 + 1 = \\
 &= \left(\frac{1}{1 - 0.5 \cdot z^{-1} - 0.5 \cdot z + 0.25} + 1 \right) \Big|_{z=e^{j \cdot \omega}} = \\
 &= \left(\frac{1 + 1.25 - 0.5 \cdot (z^{-1} + z)}{1.25 - 0.5 \cdot (z^{-1} + z)} \right) \Big|_{z=e^{j \cdot \omega}} = \\
 &= \left(\frac{2.25 - 0.5 \cdot (z^{-1} + z)}{1.25 - 0.5 \cdot (z^{-1} + z)} \right) \Big|_{z=e^{j \cdot \omega}}
 \end{aligned}$$

23. Rappresentazione tramite variabili di stato

bisogna trovare un rumore equivalente della forma:

$$w(t) = \frac{C(z)}{A(z)} \cdot e(t) \text{ con } e(t) \sim WN(0,1)$$

dato che $A(z) = 1 - \frac{1}{2} \cdot z^{-1}$ si ha che $C(z)$ deve avere solo una radice e quindi $C(z) = \alpha + \beta \cdot z^{-1}$, e quindi:

$$w(t) = \frac{\alpha + \beta \cdot z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}} \cdot e(t)$$

calcolando lo spettro si ottiene:

$$\begin{aligned} \Gamma_w(\omega) &= \left| \frac{\alpha + \beta \cdot z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}} \right|_{z=e^{j\omega}}^2 \cdot \Gamma_e(\omega) = \\ &= \left(\frac{\alpha + \beta \cdot z^{-1}}{1 - 0.5 \cdot z^{-1}} \cdot \frac{\alpha + \beta \cdot z^1}{1 - 0.5 \cdot z^1} \right) \Big|_{z=e^{j\omega}} \cdot 1 = \\ &= \left(\frac{\alpha^2 + \alpha \cdot \beta \cdot z^{-1} + \alpha \cdot \beta \cdot z + \beta^2}{1 - 0.5 \cdot z^{-1} - 0.5 \cdot z + 0.25} \right) \Big|_{z=e^{j\omega}} = \\ &= \left(\frac{\alpha^2 + \beta^2 + \alpha \cdot \beta \cdot (z^{-1} + z)}{1.25 - 0.5 \cdot (z^{-1} + z)} \right) \Big|_{z=e^{j\omega}} \end{aligned}$$

uguagliando le due espressioni ottenute:

$$\left(\frac{\alpha^2 + \beta^2 + \alpha \cdot \beta \cdot (z^{-1} + z)}{1.25 - 0.5 \cdot (z^{-1} + z)} \right) \Big|_{z=e^{j\omega}} = \left(\frac{2.25 - 0.5 \cdot (z^{-1} + z)}{1.25 - 0.5 \cdot (z^{-1} + z)} \right) \Big|_{z=e^{j\omega}}$$

dato che i denominatori sono uguali basta uguagliare il numeratore che equivale a risolvere il sistema che egualia i termini con lo stesso grado:

$$\begin{cases} \alpha^2 + \beta^2 = 2.25 \\ \alpha \cdot \beta = -0.5 \end{cases}$$

risolvendo il sistema si ottiene che:

$$\begin{aligned} \alpha &= 1.46 \\ \beta &= -0.34 \end{aligned}$$

e quindi si ha che:

$$w(t) = \frac{1.46 - 0.34 \cdot z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}} \cdot e(t)$$

e quindi:

$$\begin{aligned} \begin{cases} x(t+1) = \frac{1}{2} \cdot x(t) + \frac{1}{4} \cdot u(t) + v_1(t) & \text{con } v_1(t) \sim WN(0,1) \\ y(t) = x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0,1) \end{cases} \\ \Updownarrow \\ y(t) = \frac{\frac{1}{4}}{z - \frac{1}{2}} \cdot u(t) + \frac{1.46 - 0.34 \cdot z^{-1}}{z - \frac{1}{2}} \cdot e(t) \text{ con } e(t) \sim WN(0,1) \end{aligned}$$

23.1. Caso di rumore non bianco

Nella rappresentazione tramite variabile di stato i rumori sono tutti bianchi, ma nella realtà può capitare che il rumore che descrive l'incertezza del modello non sia bianco. Si può dimostrare che questi modelli sono abbastanza generali per inglobare anche i casi in cui il rumore non è bianco.

Infatti si può sostituire il rumore colorato con un rumore bianco e una variabile di stato in più, per capire meglio si può considerare il seguente esempio:

$$\begin{cases} x_1(t+1) = a \cdot x_1(t) + \eta(t) \\ y(t) = b \cdot x_1(t) + v_2(t) \end{cases} \quad \text{con } v_2(t) \sim WN(0,1)$$

dove:

$$\eta(t) = \frac{1}{1 - c \cdot z^{-1}} \cdot e(t) \quad \text{con } e(t) \sim WN(0,1)$$

e

$$e(t) \perp v_2(t)$$

Si può scrivere l'equazione alle differenze:

$$\begin{aligned} \eta(t) &= c \cdot z^{-1} \cdot \eta(t) + e(t) \\ \eta(t) &= c \cdot \eta(t-1) + e(t) \\ \eta(t+1) &= c \cdot \eta(t) + e(t+1) \\ \eta(t+1) &= c \cdot \eta(t) + v_1(t) \end{aligned}$$

con $v_1(t) = e(t+1)$ e quindi $v_1(t) \sim WN(0,1)$. Inserendo la variabile di stato:

$$x_2 = \eta(t)$$

si ottiene:

$$\begin{cases} x_1(t+1) = a \cdot x_1(t) + x_2(t) \\ x_2(t+1) = c \cdot x_2(t) + v_1(t) \\ y(t) = b \cdot x_1(t) + v_2(t) \end{cases} \quad \begin{aligned} &\text{con } v_1(t) \sim WN(0,1) \\ &\text{con } v_2(t) \sim WN(0,1) \end{aligned}$$

$$\begin{cases} \overbrace{\begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix}}^{x(t+1)} = \overbrace{\begin{bmatrix} a & 1 \\ c & 0 \end{bmatrix}}^F \cdot \overbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}}^{x(t)} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} \cdot v(t) \\ y(t) = b \cdot x_1(t) + v_2(t) \end{cases} \quad \begin{aligned} &\text{con } v(t) \sim WN\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) \\ &\text{con } v_2(t) \sim WN(0,1) \end{aligned}$$

che rappresenta lo stesso sistema dinamico di prima, ma con una variabile di stato in più e usando solo rumori bianchi.

23.2. Osservabilità di sistema

Definizione Sia il sistema S :

$$\begin{cases} x(t+1) = F \cdot x(t) \\ y(t) = H \cdot x(t) \end{cases}$$

e due valori iniziali x_{01} e x_{02} dello stato e le rispettive uscite $y_1(t)$ e $y_2(t)$. Il sistema S è osservabile (anche detto come "la coppia $F - H$ è osservabile"), se e solo se:

$$y_1(t) = y_2(t) \quad \forall t > 0 \Leftrightarrow x_{01} = x_{02}$$

Osservazioni

- Questa proprietà del sistema dice che osservando i valori assunti dall'uscita si può dedurre che valore iniziale aveva lo stato, e quindi non esistono stati iniziali "nascosti" e sono tutti osservabili dall'uscita.
- Questa proprietà riguarda solo la coppia di matrici F e H per questo motivo nel sistema usato nella definizione non viene rappresentata la variabile esogena e l'errore.
- Questa proprietà vale anche in ambito deterministico.
- Per capire se un sistema è osservabile si può guardare il rango della matrice di osservabilità O :

$$O = \begin{bmatrix} H \\ H \cdot F \\ H \cdot F^2 \\ \vdots \\ H \cdot F^{n-1} \end{bmatrix}$$

$Rango(O) = n \Leftrightarrow$ la coppia $F - H$ è osservabile

23.3. Raggiungibilità di un sistema

Definizione Sia il sistema S :

$$\begin{cases} x(t+1) = F \cdot x(t) + \Gamma \cdot u(t) \end{cases}$$

il sistema S è raggiungibile dall'ingresso $u(t)$ (anche detto come "la coppia $F - \Gamma$ è raggiungibile dall'ingresso $u(t)$) se e solo se partendo da uno stato iniziale nullo esiste sempre un segnale di ingresso $u_1(t)$ che porta lo stato in uno stato finale prescelto in tempo finito.

Osservazioni

- Questa proprietà del sistema dice che l'ingresso $u(t)$ ha in qualche modo effetto su tutti i possibili valori dello stato x e quindi non esistono valori dello stato irraggiungibili facendo variare solamente l'ingresso;
- Questa proprietà riguarda solo la coppia di matrici F e Γ per questo motivo nel sistema usato nella definizione non viene rappresentata l'equazione di uscita e l'errore;
- Questa proprietà vale anche in ambito deterministico;
- Per capire se un sistema è raggiungibile dallo stato $u(t)$ si può guardare il rango della matrice di raggiungibilità R :

$$R = [\Gamma \quad F \cdot \Gamma \quad F^2 \cdot \Gamma \quad \dots \quad F^{n-1} \cdot \Gamma]$$

$Rango(R) = n \Leftrightarrow$ la coppia $F - \Gamma$ è raggiungibile

23.3.1. Raggiungibilità dal rumore

Sia il sistema S :

$$\begin{cases} x(t+1) = F \cdot x(t) + v_1(t) \quad \text{con } v_1(t) \sim WN(0, V_1) \end{cases}$$

fattorizzando la varianza del rumore v_1 si ottiene:

$$V_1 = G \cdot G^T$$

si può riscrivere il sistema:

$$\begin{cases} x(t+1) = F \cdot x(t) + G \cdot w(t) & \text{con } w(t) \sim WN(0, I) \end{cases}$$

infatti:

$$\begin{aligned} Var[G \cdot w(t)] &= \mathbb{E}[(G \cdot w(t)) \cdot (G \cdot w(t))^T] = \\ &= \mathbb{E}[G \cdot w(t) \cdot w(t)^T \cdot G^T] = \\ &= G \cdot \mathbb{E}[w(t) \cdot w(t)^T] \cdot G^T = \\ &= G \cdot Var[w(t)] \cdot G^T = \\ &= G \cdot I \cdot G^T = \\ &= G \cdot G^T = \\ &= V_1 \end{aligned}$$

Scrivendo il rumore in questo modo si può definire la raggiungibilità dal rumore.

Osservazione Esistono varie matrici G che permettono di ottenere quella relazione, ma se il sistema iniziale è raggiungibile dal rumore v_1 qualsiasi matrice si usi non si perde tale proprietà.

Definizione Sia il sistema S :

$$\begin{cases} x(t+1) = F \cdot x(t) + G \cdot w(t) & \text{con } w(t) \sim WN(0, I) \end{cases}$$

il sistema S è raggiungibile dal rumore v_1 (anche detto "la coppia $F-G$ è raggiungibile dal rumore v_1) se e solo se partendo da uno stato iniziale nullo esiste sempre una realizzazione del rumore $w(t)$ che porta lo stato in uno stato finale prescelto in tempo finito.

Osservazione Valgono tutte le considerazioni fatte sulla raggiungibilità dell'ingresso.

24. Filtro di Kalman

24.1. Introduzione

Sia il sistema S MIMO con n variabili di stato, m uscite e nessun ingresso esogeno:

$$\begin{cases} x(t+1) = F \cdot x(t) + v_1(t) & \text{con } v_1(t) \sim WN(0, V_1) \\ y(t) = H \cdot x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

in cui:

- $\mathbb{E}[v_1(t) \cdot v_2(t - \tau)] = \begin{cases} V_{12} & \text{se } \tau = 0 \\ 0 & \text{se } \tau \neq 0 \end{cases}$, quindi i rumori possono essere correlati solo nello stesso istante;
- $\mathbb{E}[x(1)] = x_0 = \begin{bmatrix} x_{01} \\ x_{02} \\ \vdots \\ x_{0n} \end{bmatrix}$ ossia lo stato iniziale ha valore atteso x_0 , ed è un vettore colonna di n elementi;
- $Var[x(1)] = P_0$ ossia lo stato iniziale ha una varianza di P_0 , ed è una matrice $n \times n$ semidefinita positiva;
- $v_1(1) \perp x(1)$ e $v_2(1) \perp x(1)$ ossia lo stato iniziale non è correlato con i rumori iniziali.

Può capitare che lo stato iniziale non sia incerto, in questi casi si può porre $P_0 = 0$.

La teoria di Kalman può essere usata anche come predittore di stato e uscita e, come si vedrà, con varie estensioni può superare i limiti dei predittori di Kolmogorov-Wiener.

24.2. Predittore a un passo di Kalman

Il predittore di Kalman si prefigge di trovare i predittori ottimi $\hat{x}(t+1|t)$ e $\hat{y}(t+1|t)$ del sistema sopradescritto e li si può ottenere dalle seguenti equazioni:

$$\begin{cases} \hat{x}(t+1|t) = F \cdot \hat{x}(t|t-1) + K(t) \cdot e(t) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \\ e(t) = y(t) - \hat{y}(t|t-1) \end{cases}$$

dove $K(t)$ si definisce guadagno del filtro di Kalman, ed è rappresentato dall'espressione:

$$K(t) = (F \cdot P(t) \cdot H^T + V_{12}) \cdot (H \cdot P(t) \cdot H^T + V_2)^{-1}$$

dove $P(t)$ è la varianza dell'errore di predizione dello stato a un passo, ossia:

$$\begin{aligned} P(t) &= Var[x(t) - \hat{x}(t|t-1)] = \\ &= \mathbb{E}[(x(t) - \hat{x}(t|t-1)) \cdot (x(t) - \hat{x}(t|t-1))^T] \end{aligned}$$

e si può ricavare tramite le equazioni alle differenze di Riccati (DRE):

$$P(t+1) = F \cdot P(t) \cdot F^T + V_1 - (F \cdot P(t) \cdot H^T + V_{12}) \cdot (H \cdot P(t) \cdot H^T + V_2)^{-1} \cdot (F \cdot P(t) \cdot H^T + V_{12})^T$$

Osservazioni

- Le dimostrazioni di queste formule non verranno trattate in quanto molto complicate;
- $P(t)$ è una matrice di varianze-covarianze e quindi è semidefinita positiva $\forall t$;
- La DRE è una equazione alle differenze e quindi necessita di inizializzare sia il predittore a un passo dello stato sia la varianza dell'errore di predizione:
 - $\hat{x}(1|0) = E[x(1)] = x_0$
 - $P(1) = Var[x(1)] = P_0$
- Il predittore di Kalman può essere rappresentato tramite lo schema a blocchi in figura 24.1.
- Concettualmente il predittore di Kalman migliora la predizione al passo successivo sommandogli l'errore di predizione pesato dal valore $K(t)$. Questo schema non è nuovo e veniva usato anche dall'osservatore di Luemberg, la grande novità del predittore di Kalman è che il guadagno è calcolato in modo che la predizione sia sempre ottima.
- Per poter funzionare bisogna controllare che il predittore sia stabile e quindi che non diverga.

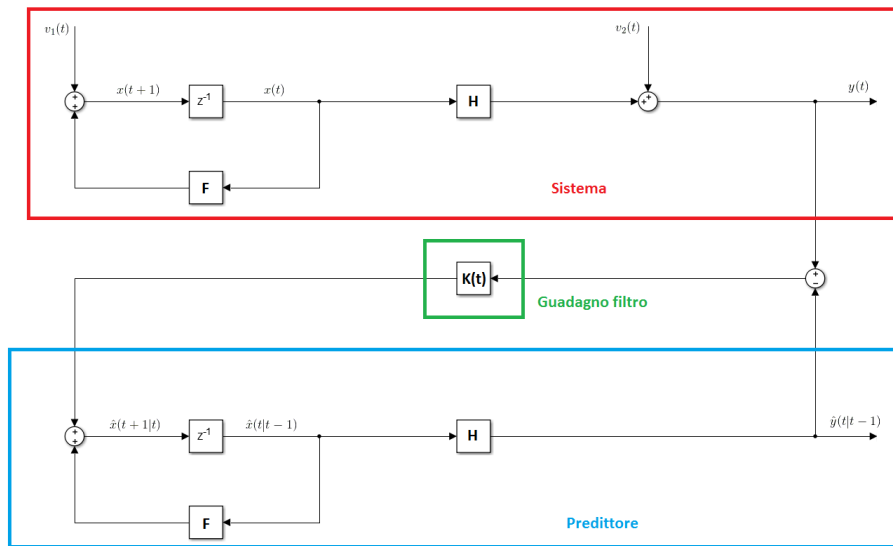


Figura 24.1.: Schema a blocchi del predittore di Kalman

24.3. Predittore a k passi di Kalman

Si possono ricavare i predittori ottimi a k passi di Kalman partendo da quello a un passo usando le seguenti formule:

$$\begin{cases} \hat{x}(t+k|t) = F^{k-1} \cdot \hat{x}(t+1|t) \\ \hat{y}(t+k|t) = H \cdot \hat{x}(t+k|t) \end{cases}$$

Quindi una volta ricavato il predittore a un passo di Kalman è possibile ricavare anche quello a k passi tramite una semplice operazione algebrica.

24.4. Filtro di Kalman

Con filtro di Kalman si intende la stima dello stato nascosto x al tempo t possedendo i valori dell'uscita fino al tempo t , ossia $\hat{x}(t|t)$, e si può ottenere in modo analogo al predittore a più passi, ossia:

$$\left\{ \hat{x}(t|t) = F^{-1} \cdot \hat{x}(t+1|t) \right.$$

Quindi, come per il predittore a più passi, una volta calcolato il predittore a un passo è possibile calcolare il filtro tramite un'espressione algebrica; Si può notare che in questo caso bisogna invertire una matrice, operazione non sempre possibile. Per questi situazioni è possibile usare il seguente sistema:

$$\begin{cases} \hat{x}(t|t) = F \cdot \hat{x}(t-1|t-1) + K_0(t) \cdot e(t) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \\ e(t) = y(t) - \hat{y}(t|t-1) \end{cases}$$

dove il guadagno del filtro $K_0(t)$ vale:

$$K_0(t) = (P(t) \cdot H^T) \cdot (H \cdot P(t) \cdot H^T + V_2)^{-1}$$

dove $P(t)$ ha lo stesso significato di quella del predittore, e quindi si calcola risolvendo la DRE.

24.5. Predittore e filtro di Kalman per sistemi tempo-varianti

In un sistema tempo-variante tutte le matrici cambiano nel tempo e quindi:

$$\begin{cases} x(t+1) = F(t) \cdot x(t) + v_1(t) & \text{con } v_1(t) \sim WN(0, V_1) \\ y(t) = H(t) \cdot x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

Fortunatamente le equazioni del filtro di Kalman:

- Equazione di stato
- Equazione di uscita
- Equazione dell'errore
- Equazione del guadagno
- Equazione DRE

rimangono inalterate, a patto di sostituire F e H con $F(t)$ e $H(t)$.

24.6. Predittore e filtro di Kalman per sistemi con ingresso esogeno

Il sistema trattato nelle sezioni precedenti era privo di ingressi esogeni, quindi ora si valuta cosa cambia nelle equazione introducendo ingressi esogeni. In questo caso si devono possedere anche le misure degli ingressi esogen oltre a quelle delle uscite.

Sia un sistema S MIMO con n variabili di stato, m uscite e p entrate:

$$\begin{cases} x(t+1) = F \cdot x(t) + \Gamma \cdot u(t) + v_1(t) & \text{con } v_1(t) \sim WN(0, V_1) \\ y(t) = H \cdot x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

dove valgono le stesse ipotesi fatte in precedenza.

Il predittore di Kalman diventa:

$$\begin{cases} \hat{x}(t+1|t) = F \cdot \hat{x}(t|t-1) + \Gamma \cdot u(t) + K(t) \cdot e(t) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \\ e(t) = y(t) - \hat{y}(t|t-1) \end{cases}$$

Dove il guadagno $K(t)$ è lo stesso usato nel predittore senza variabili esogene.

Questo sistema può essere rappresentato dallo schema a blocchi in figura 24.2.

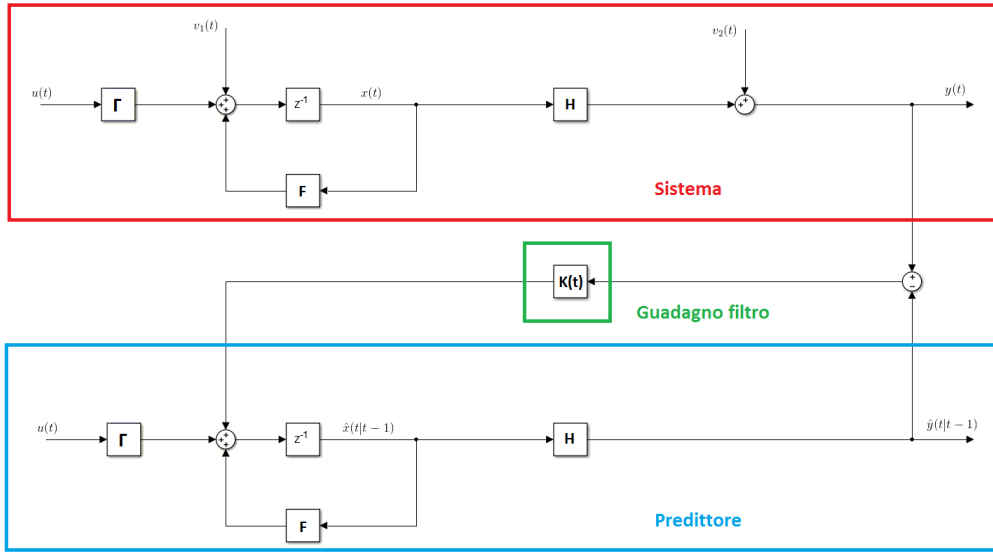


Figura 24.2.: Rappresentazione tramite schema a blocchi del predittore a a un passo di Kalman con ingresso esogeno

La stessa cosa si può applicare alle equazioni del filtro di Kalman:

$$\begin{cases} \hat{x}(t|t) = F \cdot \hat{x}(t-1|t-1) + \Gamma \cdot u(t) + K_0(t) \cdot e(t) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \\ e(t) = y(t) - \hat{y}(t|t-1) \end{cases}$$

24.7. Predittore e filtro di Kalman asintotico

In questa sezione si parlerà sempre del predittore ma analoghi ragionamenti possono essere fatti anche per il filtro. Ricordando le equazioni del predittore:

$$\begin{cases} \hat{x}(t+1|t) = F \cdot \hat{x}(t|t-1) + \Gamma \cdot u(t) + K(t) \cdot (y(t) - \hat{y}(t|t-1)) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \end{cases}$$

da cui si nota che il guadagno $K(t)$ è una quantità variabile, la teoria del predittore asintotico si basa sulla sostituzione di questa quantità variabile con una costante \bar{K} creando un filtro leggermente sub-ottimo, ma che porta dei grossi vantaggi.

Ipotizzando $K(t) = \bar{K}$ il sistema diventa:

$$\begin{cases} \hat{x}(t+1|t) = F \cdot \hat{x}(t|t-1) + \Gamma \cdot u(t) + \bar{K} \cdot (y(t) - \hat{y}(t|t-1)) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \end{cases}$$

$$\begin{cases} \hat{x}(t+1|t) = F \cdot \hat{x}(t|t-1) + \Gamma \cdot u(t) + \bar{K} \cdot y(t) - \bar{K} \cdot \hat{y}(t|t-1) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \end{cases}$$

$$\begin{cases} \hat{x}(t+1|t) = F \cdot \hat{x}(t|t-1) + \Gamma \cdot u(t) + \bar{K} \cdot y(t) - \bar{K} \cdot H \cdot \hat{x}(t|t-1) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \end{cases}$$

$$\begin{cases} \hat{x}(t+1|t) = (F - \bar{K} \cdot H) \cdot \hat{x}(t|t-1) + \Gamma \cdot u(t) + \bar{K} \cdot y(t) \\ \hat{y}(t|t-1) = H \cdot \hat{x}(t|t-1) \end{cases}$$

I vantaggi di questa modifica sono principalmente due:

- Il calcolo del guadagno richiede la soluzione della DRE e di molte inversioni di matrici, e quindi ricalcolarlo in tutte le iterazioni è dispendioso in termini computazionali.
- Nel nuovo sistema si nota che la stabilità dipende solo dagli autovalori della matrice $F - \bar{K} \cdot H$, quindi anche se il sistema fosse instabile, ossia la matrice F ha autovalori fuori dalla cerchio di raggio unitario, il predittore potrebbe risultare stabile dato che il guadagno del filtro modifica il valore degli autovalori.

Per rendere il filtro di Kalman asintotico utilizzabile bisogna però trovare un valore \bar{K} che non peggiori troppo la stima di x . In alcuni sistemi la varianza dell'errore di predizione $P(t)$ varia solo a causa di una sbagliata assunzione nel caso iniziale, ma che il filtro di Kalman eventuale risolve portando la varianza $P(t)$ a diventare costante, ossia:

$$P(1) \rightarrow P(2) \rightarrow \dots \rightarrow \bar{P}$$

in questi casi, dato che il guadagno dipende solamente dalle matrici del sistema (costanti) e da $P(t)$ anch'esso avrà un andamento simile:

$$K(1) \rightarrow K(2) \rightarrow \dots \rightarrow \bar{K}$$

Anche a regime la DRE deve valere e quindi:

$$\bar{P} = F \cdot \bar{P} \cdot F^T + V_1 - (F \cdot \bar{P} \cdot H^T + V_{12}) \cdot (H \cdot \bar{P} \cdot H^T + V_2)^{-1} \cdot (F \cdot \bar{P} \cdot H^T + V_{12})$$

che non è più una equazione alle differenze, ma una equazione algebrica che prende il nome di equazione algebrica di Riccati (ARE). Bisogna tener presente che la soluzione della ARE deve per forza essere definita positiva in quanto indica la varianza dell'errore di predizione dello stato.

Una volta ricavata la \bar{P} dalla ARE è facile ricavare la \bar{K} :

$$\bar{K} = (F \cdot \bar{P} \cdot H^T + V_{12}) \cdot (H \cdot \bar{P} \cdot H^T + V_2)^{-1}$$

Bisogna, però, capire in quali sistemi la varianza dell'errore di predizione converge a \bar{P} rendendo applicabile il predittore di Kalman asintotico, esistono alcuni teoremi di convergenza asintotica che fungono da condizione sufficiente e ricoprono la quasi totalità dei casi.

Teorema 1 di convergenza asintotica

Ipotesi:

- $V_{12} = 0$, ossia i rumori di processo e di misura non sono correlati
- Il sistema è stabile

Tesi:

- La ARE ha una e una sola soluzione ed è una matrice \bar{P} semidefinita positiva
- La DRE converge asintoticamente a \bar{P} , $\forall P_0$
- Il guadagno \bar{K} garantisce l'asintotica stabilità del predittore, ossia $F - \bar{K} \cdot H$ ha autovalori nel cerchio di raggio unitario

Teorema 2 di convergenza asintotica**Ipotesi:**

- $V_{12} = 0$, ossia i rumori di processo e di misura non sono correlati
- $F - G$ è raggiungibile
- $F - H$ è osservabile

Tesi:

- La ARE ha una e una sola soluzione ed è una matrice \bar{P} definita positiva
- La DRE converge asintoticamente a \bar{P} , $\forall P_0$
- Il guadagno \bar{K} garantisce l'asintotica stabilità del predittore, ossia $F - \bar{K} \cdot H$ ha autovalori nel cerchio di raggio unitario

Osservazione Questi due teoremi sono condizioni sufficienti se le loro ipotesi non sono verificate non è detto che il filtro di Kalman asintotico non sia applicabile, ma che bisogna analizzare la DRE e controllare che converga a un valore \bar{P} almeno semidefinito positivo.

24.8. Filtro di Kalman esteso(EKF)

Il filtro di Kalman esteso si applica a sistemi non lineari, ossia:

$$\begin{cases} x(t+1) = f(x(t)) + g(u(t)) + v_1(t) & \text{con } v_1(t) \sim WN(0, V_1) \\ y(t) = h(x(t)) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

dove f, h, g sono funzioni non lineari.

Il predittore di Kalman funziona con lo stesso principio, ossia si replica il sistema e si corregge la predizione in base all'errore di predizione di y , come mostrato in figura 24.3.

24. Filtro di Kalman

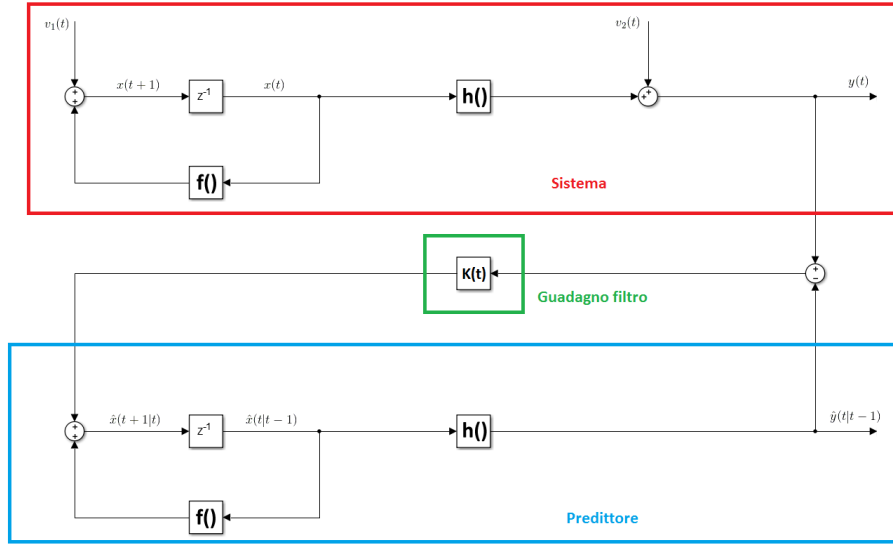


Figura 24.3.: Schema a blocchi del filtro di Kalman esteso

Per poter mantenere l'ottimalità il guadagno del filtro dovrebbe essere una funzione non lineare, ma generalizzare per tutte le funzioni non lineari è molto complicato per questo motivo, per ricavare il guadagno, si linearizza il sistema e poi si ricava come nel normale predittore.

Per linearizzare il sistema si approssima con la tangente intorno alla predizione dello stato:

$$\tilde{F} = \left. \frac{\partial f(x(t))}{\partial x(t)} \right|_{x(t)=\hat{x}(t|t-1)}$$

$$\tilde{H} = \left. \frac{\partial h(x(t))}{\partial x(t)} \right|_{x(t)=\hat{x}(t|t-1)}$$

ottenendo il sistema \tilde{S} :

$$\begin{cases} x(t+1) = \tilde{F} \cdot x(t) + g(u(t)) + v_1(t) & \text{con } v_1(t) \sim WN(0, V_1) \\ y(t) = \tilde{H} \cdot x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

da cui si può ricavare il guadagno del filtro. Si può notare che nel calcolo del guadagno e nella DRE non compare la matrice Γ e quindi è inutile linearizzare la funzione g dato che serve solo quando si replica il sistema.

Quindi:

$$\begin{cases} \hat{x}(t+1|t) = f(\hat{x}(t|t-1)) + g(u(t)) + K(t) \cdot e(t) \\ \hat{y}(t|t-1) = h(\hat{x}(t|t-1)) \\ e(t) = y(t) - \hat{y}(t|t-1) \end{cases}$$

dove:

$$K(t) = \left(\tilde{F} \cdot P(t) \cdot \tilde{H}^T + V_{12} \right) \cdot \left(\tilde{H} \cdot P(t) \cdot \tilde{H}^T + V_2 \right)^{-1}$$

e la DRE:

$$P(t+1) = \tilde{F} \cdot P(t) \cdot \tilde{F}^T + V_1 - \left(\tilde{F} \cdot P(t) \cdot \tilde{H}^T + V_{12} \right) \cdot \left(\tilde{H} \cdot P(t) \cdot \tilde{H}^T + V_2 \right)^{-1} \cdot \left(\tilde{F} \cdot P(t) \cdot \tilde{H}^T + V_{12} \right)$$

Procedura Si procede iteramente e al passo t si:

1. Si linearizzano le funzioni f e h ottenendo le matrici \tilde{F} e \tilde{H}
2. Si ricava la matrice $P(t)$ dalla DRE usando le matrici linearizzate
3. Si ricava il guadagno $K(t)$ dalla $P(t)$ appena calcolata
4. Si ricava l'errore $e(t) = y(t) - \hat{y}(t|t-1)$
5. Si ricava la stima $\hat{x}(t+1|t) = f(\hat{x}(t|t-1)) + g(u(t)) + K(t) \cdot e(t)$
6. Si ricava la stima $\hat{y}(t+1|t) = h(\hat{x}(t+1|t))$
7. Si esegue il passo $t+1$

Osservazione il filtro di Kalman esteso non è matematicamente rigoroso dato che non tutte le funzioni hanno componenti di secondo grado, e superiori, poco significative rispetto a quelle di primo grado, però la pratica ha dimostrato che questo predittore funziona bene anche se non è ottimo e la distanza dall'ottimo dipende molto da come sono fatte le funzioni f e h .

24.9. Identificazione a scatola grigia

L'identificazione a scatola grigia è una procedura che permette di identificare un sistema di cui si conosce il modello a meno di uno, o più, parametri. Quindi non è un modello a scatola bianca perchè non si conosce perfettamente il modello, ma non è neppure a scatola nera perchè lo si conosce il parte.

Si vogliono stimare i N parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_N$ di un sistema di ordine n

$$\begin{cases} x(t+1) = f(x(t), \vartheta) + v_1(t) & \text{con } v_1(t) \sim WN(0, V_1) \\ y(t) = h(x(t), \vartheta) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

in cui:

$$\vartheta_{N \times 1} = \begin{bmatrix} \vartheta_1 \\ \vartheta_2 \\ \vdots \\ \vartheta_N \end{bmatrix}$$

Dato che i parametri $\vartheta_1, \vartheta_2, \dots, \vartheta_N$ sono valori costanti si possono definire N variabili di stato con dinamica costante:

$$\vartheta(t+1) = \vartheta(t)$$

Quindi il vettore diventa un vettore di $n + N$ elementi:

$$\tilde{x}(t) = \begin{bmatrix} x(t) \\ \vartheta(t) \end{bmatrix}$$

Il problema di questo metodo è che le nuove variabili di stato non variano mai e quindi una volta imposto il valore iniziale $\vartheta(0) = \vartheta_0$ tutti gli stati successivi saranno uguali:

$$\vartheta_0 = \vartheta(0) = \vartheta(1) = \vartheta(2) = \dots = \vartheta(t)$$

quindi dato che l'obiettivo è stimare ϑ non si può avere una inizializzazione perfetta quindi bisogna introdurre un rumore alla variabile per indicare quanto sia sbagliato il valore iniziale:

$$\vartheta(t+1) = \vartheta(t) + v_0 \text{ con } v_0 \sim WN(0, V_0)$$

solitamente si sceglie V_0 in modo che:

- è una matrice diagonale, quindi i rumori dei parametri da stimare sono scorrelati
- tutti i rumori sono scorrelati con i rumori delle altre variabili di stato

ottenendo:

$$V_0 = \begin{bmatrix} \lambda_1^2 & 0 & \cdots & 0 \\ 0 & \lambda_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N^2 \end{bmatrix}$$

dove $\lambda_1^2, \lambda_2^2, \dots, \lambda_N^2$ sono le varianze del rumore e indicano l'incertezza dell'inizializzazione dei vari parametri da stimare. Inoltre si ha che:

$$\tilde{V}_1 = \begin{bmatrix} V_1 & 0 \\ 0 & V_0 \end{bmatrix}$$

ottenendo il sistema:

$$\begin{cases} \tilde{x}(t+1) = f(\tilde{x}(t)) + v_1(t) & \text{con } v_1(t) \sim WN(0, \tilde{V}_1) \\ y(t) = h(\tilde{x}(t)) + v_2(t) & \text{con } v_2(t) \sim WN(0, V_2) \end{cases}$$

Osservazione

- Per poter utilizzare questo metodo non basta conoscere una stima per inizializzare il parametro, ma bisogna anche conoscere quanto sia effettivamente buona la stima. Se per esempio si ha $\vartheta = 100$ e si inizializza con $\vartheta_0 = 0$, bisogna impostare la varianza in modo che rappresenti quanto non buona sia la nostra stima, infatti se si ponesse $V_0 = 2$ l'identificazione non sarà per nulla buona dato che con quella varianza è praticamente impossibile che ϑ arrivi a 100.
- Nella quasi totalità dei casi il parametro da stimare ϑ moltiplica una variabile di stato, come in tutti i sistemi lineari, quindi il sistema che si usa per la stima non è quasi mai lineare e bisogna utilizzare il filtro di Kalman esteso.

Esempio

Si consideri un sistema massa molla con:

- coefficiente d'elasticità k
- massa M
- coefficiente di attrito c

Facendo il bilancio delle forze si ottiene:

$$\begin{aligned} M \cdot \ddot{x} + c \cdot \dot{x} + k \cdot x &= 0 \\ \ddot{x} &= -\frac{c}{M} \cdot \dot{x} - \frac{k}{M} \cdot x \end{aligned}$$

ponendo:

$$\begin{aligned} x_1 &= x \\ x_2 &= \dot{x} \end{aligned}$$

e supponendo di misurare la posizione della massa:

$$y(t) = x(t) = x_1(t)$$

si ottiene il sistema:

$$\begin{cases} \dot{x}_1 = x_2 \\ \dot{x}_2 = -\frac{c}{M} \cdot x_2 - \frac{k}{M} \cdot x_1 \\ y = x_1 \end{cases}$$

Si può discretizzare il sistema approssimando la derivata con il rapporto incrementale:

$$\dot{x}_1 = \frac{x_1(t+1) - x_1(t)}{\Delta t}$$

$$\dot{x}_2 = \frac{x_2(t+1) - x_2(t)}{\Delta t}$$

dove Δt è il tempo di campionamento. Da cui si ottiene:

$$\begin{cases} x_1(t+1) = x_1(t) + x_2(t) \cdot \Delta t + v_{11} & \text{con } v_{11} \sim WN(0, V_{11}) \\ x_2(t+1) = -\frac{k \cdot \Delta t}{M} \cdot x_1(t) + \left(1 - \frac{c \cdot \Delta t}{M}\right) \cdot x_2(t) + v_{12} & \text{con } v_{12} \sim WN(0, V_{12}) \\ y(t) = x_1(t) \end{cases}$$

Dove v_{11} e v_{12} indicano l'incertezza del modello.

Supponendo di voler stimare il parametro c si inserisce la variabile di stato:

$$c(t+1) = c(t) + V_{13}(t) \quad \text{con } v_{13} \sim WN(0, V_{13})$$

ottenendo il sistema non lineare:

$$\begin{cases} x_1(t+1) = x_1(t) + x_2(t) \cdot \Delta t + v_{11} & \text{con } v_{11} \sim WN(0, V_{11}) \\ x_2(t+1) = -\frac{k \cdot \Delta t}{M} \cdot x_1(t) + \left(1 - \frac{c(t) \cdot \Delta t}{M}\right) \cdot x_2(t) + v_{12} & \text{con } v_{12} \sim WN(0, V_{12}) \\ c(t+1) = c(t) + V_{13}(t) & \text{con } v_{13} \sim WN(0, V_{13}) \\ y(t) = x_1(t) \end{cases}$$

a cui si può applicare il filtro di Kalman esteso per ottenere lo stimatore ottimo $\hat{c}(t|t)$ e quindi il parametro.

Parte VI.

Esercitazioni

25. Esercitazione del 15/10

Esercizio 1

Testo

Si consideri il seguente processo:

$$\begin{aligned}y(t) &= e(t) + \frac{1}{5} \cdot e(t-1) \\ \text{con} \quad e(t) &\sim WN(0, 2)\end{aligned}$$

Definire:

1. Tipo di processo
2. Se il processo $y(t)$ è un processo stocastico stazionario
3. I valori di m_y e di $\gamma_y(\tau)$
4. Lo spettro $\Gamma_y(\omega)$
5. Tutte le precedenti nel caso $e(t) \sim WN(1, 2)$

Soluzione

1. Si tratta di un processo di tipo $MA(1)$; la teoria riguardante questo tipo di processo si trova nel capitolo 4.1
2. Dato il fatto che $y(t)$ è un $MA(1)$ sappiamo già che è stazionario, ma possiamo verificare comunque.
Per essere stazionario, $e(t)$ deve essere stazionario e la funzione di trasferimento ricavata dal processo deve essere asintoticamente stabile. Sappiamo che $e(t)$, essendo un rumore bianco, è stazionario; controlliamo quindi la funzione di trasferimento:

$$\begin{aligned}y(t) &= e(t) + \frac{1}{5} \cdot e(t-1) \\ y(t) &= e(t) + \frac{1}{5} \cdot z^{-1} \cdot e(t) \\ \frac{y(t)}{e(t)} &= 1 + \frac{1}{5} \cdot z^{-1} \\ H(z) &= \frac{z + \frac{1}{5}}{z}\end{aligned}$$

Si può vedere che la funzione di trasferimento ha un solo polo in 0, quindi è as. stabile. Questo significa che $y(t)$ è un processo stocastico stazionario.

3. Si deve calcolare m_y e $\gamma_y(\tau)$:

a) Calcoliamo m_y :

$$\begin{aligned}
 m_y &= \mathbb{E} \left[e(t) + \frac{1}{5} \cdot e(t-1) \right] = \\
 &= \mathbb{E}[e(t)] + \frac{1}{5} \cdot \mathbb{E}[e(t-1)] = \\
 &= 0 + \frac{1}{5} \cdot 0 = \\
 &= 0
 \end{aligned}$$

b) Calcoliamo $\gamma_y(\tau)$:

$$\begin{aligned}
 \gamma_y(0) &= \mathbb{E}[(y(t) - \cancel{m_y})^2] = \\
 &= \mathbb{E} \left[e(t)^2 + \frac{1}{25} \cdot e(t-1)^2 + \frac{2}{5} \cdot e(t) \cdot e(t-1) \right] = \\
 &= \underbrace{\mathbb{E}[e(t)^2]}_{\lambda^2=2} + \frac{1}{25} \cdot \underbrace{\mathbb{E}[e(t-1)^2]}_{\lambda^2=2} + \frac{2}{5} \cdot \underbrace{\mathbb{E}[e(t) \cdot e(t-1)]}_0 = \\
 &= 2 + \frac{2}{25} = \\
 &= \frac{52}{25}
 \end{aligned}$$

$$\begin{aligned}
 \gamma_y(1) &= \mathbb{E}[(y(t) - \cancel{m_y}) \cdot (y(t-1) - \cancel{m_y})] = \\
 &= \mathbb{E}[y(t) \cdot y(t-1)] = \\
 &= \mathbb{E} \left[\left(e(t) + \frac{1}{5} \cdot e(t-1) \right) \cdot \left(e(t-1) + \frac{1}{5} \cdot e(t-2) \right) \right] = \\
 &= \frac{1}{5} \cdot \underbrace{\mathbb{E}[e(t-1) \cdot e(t-1)]}_{\lambda^2=2} + \frac{1}{25} \cdot \underbrace{\mathbb{E}[e(t-1) \cdot e(t-2)]}_0 + \\
 &\quad + \frac{1}{5} \cdot \underbrace{\mathbb{E}[e(t) \cdot e(t-2)]}_0 + \underbrace{\mathbb{E}[e(t) \cdot e(t-1)]}_0 = \\
 &= \frac{2}{5}
 \end{aligned}$$

Essendo un $MA(1)$, sappiamo che $\gamma_y(\tau) = 0, \forall \tau \geq 2$. Controlliamo comunque il valore di $\gamma_y(2)$:

$$\begin{aligned}
 \gamma_y(2) &= \mathbb{E}[y(t) \cdot y(t-2)] = \\
 &= \mathbb{E} \left[\left(e(t) + \frac{1}{5} \cdot e(t-1) \right) \cdot y(t-2) \right] = \\
 &= \underbrace{\mathbb{E}[e(t) \cdot y(t-2)]}_0 + \frac{1}{5} \cdot \underbrace{\mathbb{E}[e(t-1) \cdot y(t-2)]}_0 = \\
 &= 0
 \end{aligned}$$

4. Calcoliamo ora lo spettro $\Gamma_y(\omega)$. Possiamo usare due metodi per arrivare allo stesso risultato:

a) Usando la definizione, vista nel capitolo 3.1:

$$\begin{aligned}\Gamma_y(\omega) &= \sum_{\tau=-\infty}^{+\infty} \gamma_y(\tau) \cdot e^{-j\omega\tau} \\ &= \gamma_y(0) \cdot e^{-j\omega \cdot 0} + \gamma_y(1) \cdot e^{-j\omega \cdot 1} + \gamma_y(-1) \cdot e^{-j\omega \cdot (-1)} = \\ &= \gamma_y(0) + \gamma_y(1) \cdot \underbrace{(e^{-j\omega} + e^{j\omega})}_{2 \cdot \cos \omega} = \\ &= \frac{52}{25} + \frac{2}{5} \cdot 2 \cdot \cos \omega = \\ &= \frac{52}{25} + \frac{4}{5} \cdot \cos \omega\end{aligned}$$

b) Usando il teorema della fattorizzazione spettrale, visto nel capitolo 5:

$$\begin{aligned}\Gamma_y(\omega) &= |H(z)|^2 \cdot \Gamma_e(\omega) \\ \text{con } z &= e^{j\omega} \\ e \quad H(z) &= \frac{z + \frac{1}{5}}{z}\end{aligned}$$

$$\begin{aligned}\Gamma_y(\omega) &= \frac{e^{j\omega} + \frac{1}{5}}{e^{j\omega}} \cdot \frac{e^{-j\omega} + \frac{1}{5}}{e^{-j\omega}} \cdot 2 = \\ &= \left(1 + \frac{1}{5} \cdot e^{j\omega} + \frac{1}{5} \cdot e^{-j\omega} + \frac{1}{25}\right) \cdot 2 = \\ &= \frac{52}{25} + \frac{1}{5} \cdot \underbrace{(e^{j\omega} + e^{-j\omega})}_{2 \cdot \cos \omega} = \\ &= \frac{52}{25} + \frac{4}{5} \cdot \cos \omega\end{aligned}$$

Solitamente basta lo spettro interpolando i valori di $\omega = 0, \frac{\pi}{2}$ e π . Una rappresentazione più precisa è presente nella figura 25.1.

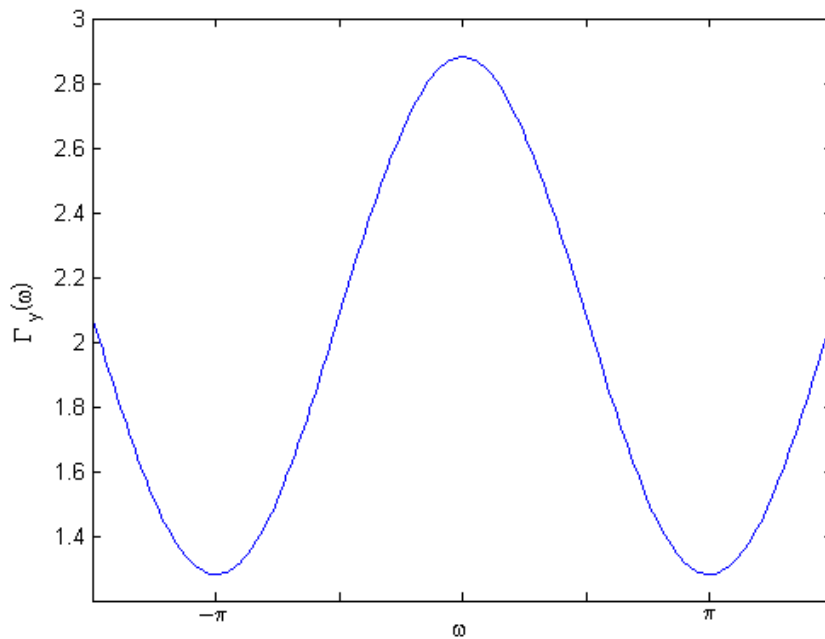


Figura 25.1.: Esercizio 1 - Rappresentazione dello spettro $\Gamma_y(\omega) = \frac{52}{25} + \frac{4}{5} \cdot \cos \omega$

5. Ora si ha che $e(t) \sim WN(1, 2)$. Possiamo saltare le prime due domande, dato che sappiamo che il processo è ancora un $MA(1)$ e che è ancora un processo stocastico stazionario. Calcoliamo quindi $m_y, \gamma_y(\tau)$ e $\Gamma_y(\omega)$.

$$\begin{aligned}
 m_y &= \mathbb{E} \left[e(t) + \frac{1}{5} \cdot e(t-1) \right] = \\
 &= \mathbb{E}[e(t)] + \frac{1}{5} \cdot \mathbb{E}[e(t-1)] = \\
 &= 1 + \frac{1}{5} = \\
 &= \frac{6}{5} \\
 \gamma_y(0) &= \mathbb{E} \left[(y(t) - m_y)^2 \right] = \\
 &= \mathbb{E} \left[\left(e(t) + \frac{1}{5} \cdot e(t-1) - \frac{6}{5} \right)^2 \right]
 \end{aligned}$$

Questo calcolo porta ad avere molti più termini di quanti una persona normale abbia lo sbatti di calcolare, quindi usiamo la tecnica della depolarizzazione, della quale si è discusso nel capitolo 3.3:

$$\begin{cases} m_y = \frac{6}{5} \\ m_e = 1 \end{cases}$$

$$\begin{cases} \tilde{y}(t) = y(t) - \frac{6}{5} \\ \tilde{e}(t) = e(t) - 1 \end{cases}$$

$$\begin{cases} y(t) = \tilde{y}(t) + \frac{6}{5} \\ e(t) = \tilde{e}(t) + 1 \end{cases}$$

$$\tilde{y}(t) = \tilde{e}(t) + \frac{1}{5} \cdot \tilde{e}(t-1)$$

Che è esattamente uguale al caso iniziale. Avremo quindi covarianza e spettro esattamente identici a prima. Questo per mostrare che variando la media del processo stesso, non cambieremo la sua covarianza e di conseguenza nemmeno lo spettro.

Esercizio 2

Testo

Determinare se i seguenti valori di funzioni di covarianza possono appartenere ad un processo stocastico stazionario, ed in caso affermativo, calcolare dei possibili valori per i coefficienti del processo stesso.

1. $\gamma_y(0) = -1, \gamma_y(1) = \frac{1}{2}, \gamma_y(-1) = \frac{1}{2}$ e $\gamma_y(\tau) = 0 \forall \tau \geq 2$
2. $\gamma_y(0) = 1, \gamma_y(1) = \frac{1}{2}, \gamma_y(-1) = -\frac{1}{2}$ e $\gamma_y(\tau) = 0 \forall \tau \geq 2$
3. $\gamma_y(0) = 2, \gamma_y(1) = -\frac{1}{5}, \gamma_y(-1) = -\frac{1}{5}$ e $\gamma_y(\tau) = 0 \forall \tau \geq 2$

Soluzione

1. Non può essere la funzione di covarianza di un processo stocastico stazionario. Come si è visto nel capitolo 3, $\gamma_y(0) \geq 0$.
2. Nemmeno questa può essere una funzione di covarianza di un processo stocastico stazionario. Sempre facendo riferimento al capitolo 3, sappiamo che $\gamma_y(\tau) = \gamma_y(-\tau)$.
3. Questa è la funzione di covarianza di un $MA(1)$, dato che $\gamma_y(\tau) = 0 \forall \tau \geq 2$. Calcoliamo dei possibili valori per i coefficienti di questo processo.

$$y(t) = c_0 \cdot e(t) + c_1 \cdot e(t-1)$$

$$\begin{cases} \gamma_y(0) = \lambda^2 \cdot (c_0^2 + c_1^2) \\ \gamma_y(1) = \lambda^2 \cdot c_0 \cdot c_1 \end{cases}$$

Questo sistema ha 2 equazioni in 3 incognite, quindi ha infinite soluzioni. Supponiamo $c_0 = 1$ per trovarne una:

$$\begin{cases} 2 = \lambda^2 \cdot (1 + c_1^2) \\ -\frac{1}{5} = \lambda^2 \cdot c_1 \end{cases}$$

$$\begin{cases} c_1 = -5 + 2 \cdot \sqrt{6} \\ \lambda^2 = \frac{5 + 2 \cdot \sqrt{6}}{5} \end{cases}$$

Esercizio 3

Testo

Dato il processo:

$$\begin{aligned} y(t) &= \Theta \cdot y(t-1) + e(t) \\ \text{con } e(t) &\sim WN(0, 1) \end{aligned}$$

1. Di che tipo di processo si tratta?
2. Determinare i valori di Θ per cui è un processo stocastico stazionario
3. Considerando $\Theta = -\frac{1}{4}$, calcolare m_y , $\gamma_y(\tau)$ e $\Gamma_y(\omega)$

Soluzione

1. Si tratta di un $AR(1)$. Abbiamo parlato di questo tipo di processi nel capitolo 4.2.
2. Per essere un processo stocastico stazionario, il sistema deve essere asintoticamente stabile. Calcoliamo la funzione di trasferimento:

$$\begin{aligned} y(t) &= \Theta \cdot z^{-1} \cdot y(t) + e(t) \\ y(t) \cdot (1 - \Theta \cdot z^{-1}) &= e(t) \\ H(z) &= \frac{1}{1 - \Theta \cdot z^{-1}} \\ H(z) &= \frac{z}{z - \Theta} \end{aligned}$$

Il che significa che il processo è stazionario per $|\Theta| < 1$.

3. Calcoliamo m_y , $\gamma_y(\tau)$ e $\Gamma_y(\omega)$:

$$\begin{aligned} m_y &= \mathbb{E} \left[-\frac{1}{4} \cdot y(t-1) + e(t) \right] \\ m_y &= -\frac{1}{4} \cdot \mathbb{E}[y(t-1)] + \underbrace{\mathbb{E}[e(t)]}_0 \\ m_y &= -\frac{1}{4} \cdot m_y + 0 \\ m_y \cdot \left(1 + \frac{1}{4} \right) &= 0 \\ m_y &= 0 \end{aligned}$$

Per quanto riguarda la funzione di covarianza, sappiamo che per gli $AR(1)$ valgono:

$$\begin{aligned} \gamma(0) &= \frac{\lambda^2}{1 - a^2} \\ \gamma(\tau) &= a \cdot \gamma(\tau-1) \text{ per } \tau > 0 \end{aligned}$$

Per lo spettro non possiamo usare la definizione, in quanto sarebbe una somma infinita di termini. Dobbiamo per forza usare il teorema della fattorizzazione spettrale:

$$\begin{aligned}
 \Gamma_y(\omega) &= |H(z)|^2 \cdot \underbrace{\Gamma_e(\omega)}_1 = \\
 &= \frac{e^{j\omega}}{e^{j\omega} + \frac{1}{4}} \cdot \frac{e^{-j\omega}}{e^{-j\omega} + \frac{1}{4}} = \\
 &= \frac{1}{1 + \frac{1}{4} \cdot (e^{j\omega} + e^{-j\omega}) + \frac{1}{16}} = \\
 &= \frac{1}{\frac{17}{16} + \frac{1}{4} \cdot 2 \cdot \cos \omega} = \\
 &= \frac{16}{17 + 8 \cdot \cos \omega}
 \end{aligned}$$

La rappresentazione dello spettro appena calcolata può essere vista in figura 25.2

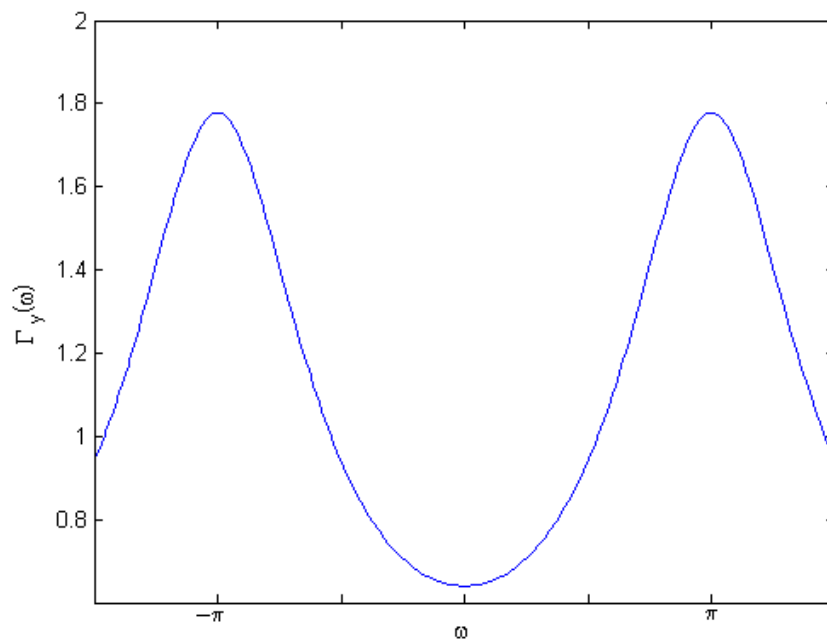


Figura 25.2.: Esercizio 3 - Rappresentazione dello spettro $\Gamma_y(\omega) = \frac{16}{17 + 8 \cdot \cos \omega}$

Esercizio 4

Testo

Associare gli spettri presenti nella prima riga di figura 25.3 con le loro possibili realizzazioni temporali, presenti nella seconda riga della stessa immagine.

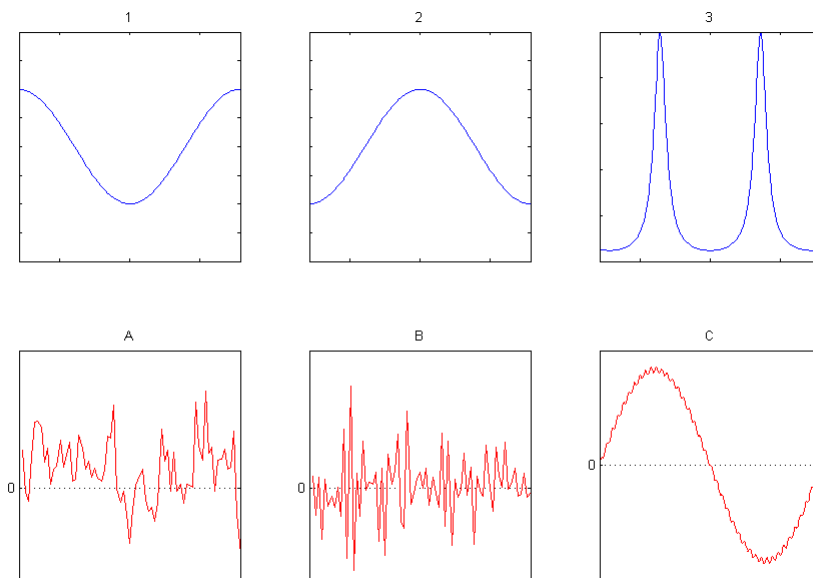


Figura 25.3.: Esercizio 4

Soluzione

Le associazioni sono:

- $1B$: Si nota dallo spettro che le componenti più marcate sono quelle a frequenza più alta, perciò la realizzazione temporale deve rispecchiare questa proprietà.
- $2A$: In questo spettro invece sono più marcate le componenti a bassa frequenza, quindi la realizzazione temporale avrà cambiamenti più "lenti"
- $3C$: Nell'ultimo spettro si può vedere che c'è una componente molto marcata, mentre le altre assumono poca rilevanza. Questo significa che nella realizzazione temporale ci sarà una sinusoide marcata a cui viene sommato un pò di rumore

Esercizio 5

Testo

Dato il sistema in figura 25.4

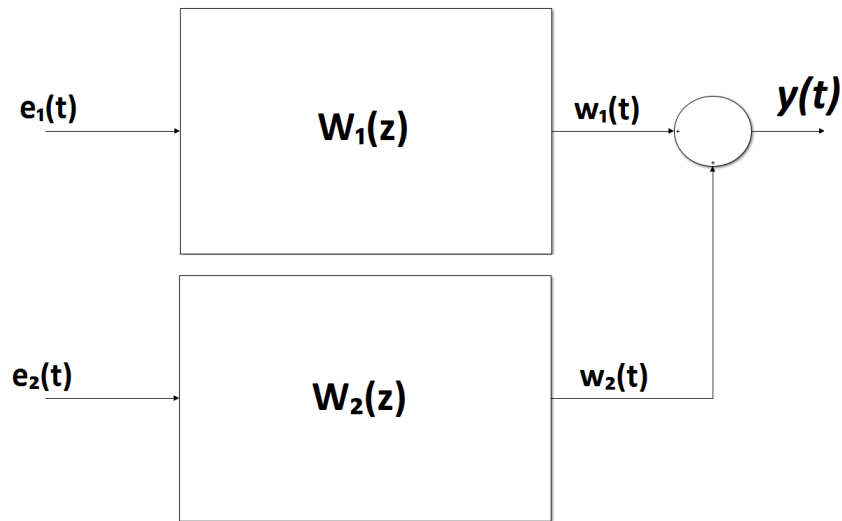


Figura 25.4.: Esercizio 5

e considerando i seguenti dati:

$$\begin{cases} e_1(t) \sim WN(0, \lambda^2) \\ e_2(t) \sim WN(0, \lambda^2) \\ e_1(t) \perp e_2(t) & (\text{scorrelati}) \\ W_1(z), W_2(z) & \text{asintoticamente stabili} \end{cases}$$

Calcolare:

1. Funzione di covarianza
2. Spettro

Soluzione

1. Calcoliamo la funzione di covarianza:

$$\begin{aligned} \gamma_y(\tau) &= \mathbb{E}[(w_1(t) + w_2(t)) \cdot (w_1(t - \tau) + w_2(t - \tau))] = \\ &= \underbrace{\mathbb{E}[w_1(t) \cdot w_1(t - \tau)]}_{\gamma_{w_1}(\tau)} + \mathbb{E}[w_1(t) \cdot w_2(t - \tau)] + \mathbb{E}[w_2(t) \cdot w_1(t - \tau)] + \underbrace{\mathbb{E}[w_2(t) \cdot w_2(t - \tau)]}_{\gamma_{w_2}(\tau)} = \\ &= \gamma_{w_1}(\tau) + \gamma_{w_2}(\tau) + \mathbb{E}[w_1(t) \cdot w_2(t - \tau)] + \mathbb{E}[w_2(t) \cdot w_1(t - \tau)] \end{aligned}$$

Ora proviamo a dimostrare che $e_1(t) \perp e_2(t) \iff w_1(t) \perp w_2(t)$:

Ipotizziamo $w_1(t)$ e $w_2(t)$ come $MA(\infty)$; la spiegazione sul perchè ci è possibile fare questo si trova nel capitolo 4.5

$$\begin{aligned} w_1(t) &= c_{01} \cdot e_1(t) + c_{11} \cdot e_1(t) + c_{21} \cdot e_1(t) + \dots \\ w_2(t) &= c_{02} \cdot e_2(t) + c_{12} \cdot e_2(t) + c_{22} \cdot e_2(t) + \dots \end{aligned}$$

$$\begin{aligned}\mathbb{E}[w_1(t) \cdot w_2(t - \tau)] &= \mathbb{E}[(c_{01} \cdot e_1(t) + c_{11} \cdot e_1(t) + c_{21} \cdot e_1(t) + \dots) \cdot \\ &\quad \cdot (c_{02} \cdot e_2(t) + c_{12} \cdot e_2(t) + c_{22} \cdot e_2(t) + \dots)] \\ &= 0\end{aligned}$$

Siccome sono solo prodotti tra due white noise incorrelati.

Ne deduciamo quindi che:

$$\gamma_y(\tau) = \gamma_{w_1}(\tau) + \gamma_{w_2}(\tau)$$

2. Calcoliamo lo spettro usando la definizione:

$$\begin{aligned}\Gamma_y(\omega) &= \sum_{\tau=-\infty}^{+\infty} \gamma_y(\tau) \cdot e^{-j\omega\tau} = \\ &= \sum_{\tau=-\infty}^{+\infty} (\gamma_{w_1}(\tau) + \gamma_{w_2}(\tau)) \cdot e^{-j\omega\tau} = \\ &= \Gamma_{w_1}(\omega) + \Gamma_{w_2}(\omega)\end{aligned}$$

Un risultato simile esce anche usando il teorema della fattorizzazione spettrale:

$$\begin{aligned}\Gamma_y(\omega) &= |W_1(z)| \cdot \Gamma_{e_1}(\omega) + |W_2(z)| \cdot \Gamma_{e_2}(\omega) = \\ &= |W_1(z)| \cdot \lambda^2 + |W_2(z)| \cdot \lambda^2 = \\ &= (|W_1(z)| + |W_2(z)|) \cdot \lambda^2\end{aligned}$$

Esercizio 6

Testo

Dato il sistema in figura 25.5, calcolare lo spettro $\Gamma_y(\omega)$ in entrambi i seguenti casi:

1. $d(t) \perp \eta(t) \perp e(t)$ e sapendo che:

$$\begin{cases} d(t) \sim WN(0, 1) \\ \eta(t) \sim WN(0, 1) \\ e(t) \sim WN(0, 2) \end{cases}$$

2. $\eta(t) \perp e(t)$ e sapendo che:

$$\begin{cases} d(t) = -e(t - 1) \\ \eta(t) \sim WN(0, 1) \\ e(t) \sim WN(0, 2) \end{cases}$$

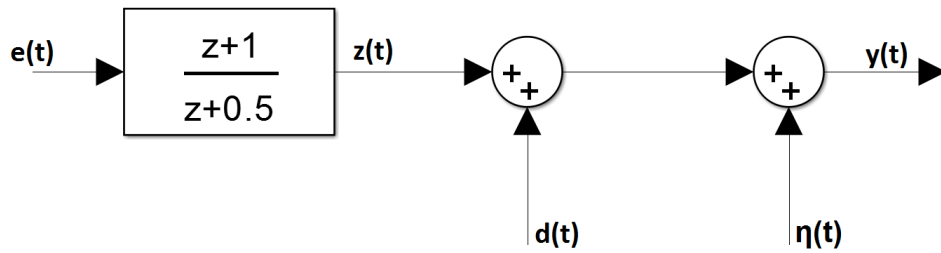


Figura 25.5.: Esercizio 6

Soluzione

1. Calcoliamo lo spettro

$$\Gamma_y(\omega) = \Gamma_z(\omega) + \underbrace{\Gamma_d(\omega)}_1 + \underbrace{\Gamma_\eta(\omega)}_1$$

L'unico spettro che dobbiamo calcolare è $\Gamma_z(\omega)$:

$$\begin{aligned} \Gamma_z(\omega) &= \left| \frac{z+1}{z+\frac{1}{2}} \right|^2 \cdot \underbrace{\Gamma_e(\omega)}_2 = \\ &= \frac{e^{j\omega} + 1}{e^{j\omega} + \frac{1}{2}} \cdot \frac{e^{-j\omega} + 1}{e^{-j\omega} + \frac{1}{2}} \cdot 2 = \\ &= \frac{1 + e^{j\omega} + e^{-j\omega} + 1}{1 + \frac{1}{2} \cdot (e^{j\omega} + e^{-j\omega}) + \frac{1}{4}} \cdot 2 = \\ &= \frac{4 + 4 \cdot \cos \omega}{\frac{5}{4} + \cos \omega} \end{aligned}$$

Ora possiamo calcolare $\Gamma_y(\omega)$:

$$\begin{aligned} \Gamma_y(\omega) &= \frac{4 + 4 \cdot \cos \omega}{\frac{5}{4} + \cos \omega} + 2 = \\ &= \frac{4 + 4 \cdot \cos \omega + \frac{5}{2} + 2 \cdot \cos \omega}{\frac{5}{4} + \cos \omega} = \\ &= \frac{\frac{13}{2} + 6 \cdot \cos \omega}{\frac{5}{4} + \cos \omega} \end{aligned}$$

Questo spettro è rappresentato in figura 25.6.

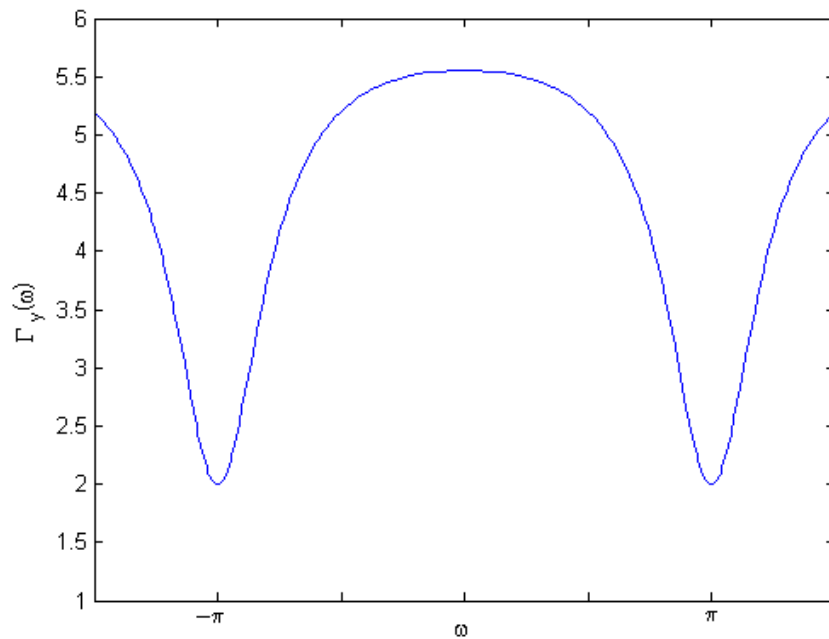


Figura 25.6.: Esercizio 6 - Rappresentazione dello spettro $\Gamma_y(\omega) = \frac{\frac{13}{2} + 6 \cdot \cos \omega}{\frac{5}{4} + \cos \omega}$

2. Per prima cosa rappresentiamo graficamente questo cambiamento nel sistema: porre $d(t) = -e(t-1)$ significa prendere $e(t)$, ritardarlo di un campione e poi sommarlo a $z(t)$. Il sistema è rappresentato in figura 25.7

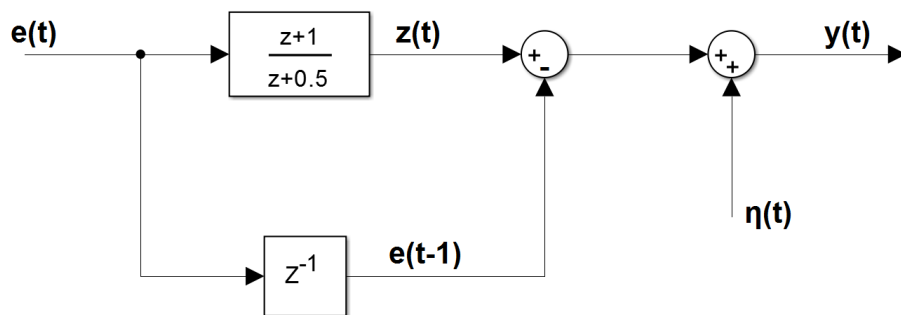


Figura 25.7.: Esercizio 6 - Rappresentazione del sistema nel caso $d(t) = e(t-1)$

25. Esercitazione del 15/10

Possiamo calcolare la funzione di trasferimento complessiva, data dalla differenza delle due:

$$\begin{aligned}
 F(z) &= \frac{z+1}{z+\frac{1}{2}} - \frac{1}{z} = \\
 &= \frac{z^2+z-z-\frac{1}{2}}{z \cdot \left(z+\frac{1}{2}\right)} = \\
 &= \frac{z^2-\frac{1}{2}}{z \cdot \left(z+\frac{1}{2}\right)}
 \end{aligned}$$

Il sistema ha un polo in 0 e uno in $-\frac{1}{2}$, perciò è asintoticamente stabile. Rappresentiamo il nuovo sistema in figura 25.8.

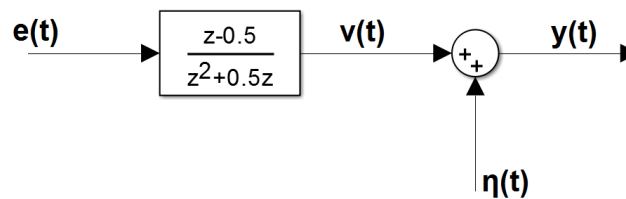


Figura 25.8.: Esercizio 6 - Rappresentazione compatta del sistema

Ora possiamo calcolare lo spettro $\Gamma_y(\omega)$. Ricordiamo che $e(t) \perp \eta(t)$.

$$\begin{aligned}
 \Gamma_y(\omega) &= \Gamma_v(\omega) + \underbrace{\Gamma_\eta(\omega)}_1 = \\
 &= \left| \frac{z^2 - \frac{1}{2}}{z \cdot \left(z + \frac{1}{2}\right)} \right|^2 \cdot \Gamma_e(\omega) + 1 = \\
 &= \frac{e^{j \cdot 2 \cdot \omega} - \frac{1}{2}}{e^{j \cdot \omega} \cdot \left(e^{j \cdot \omega} + \frac{1}{2}\right)} \cdot \frac{e^{-j \cdot 2 \cdot \omega} - \frac{1}{2}}{e^{-j \cdot \omega} \cdot \left(e^{-j \cdot \omega} + \frac{1}{2}\right)} \cdot 2 + 1 = \\
 &= \frac{1 - \frac{1}{2} \cdot (e^{j \cdot 2 \cdot \omega} + e^{-j \cdot 2 \cdot \omega}) + \frac{1}{4}}{1 + \frac{1}{2} \cdot (e^{j \cdot \omega} + e^{-j \cdot \omega}) + \frac{1}{4}} \cdot 2 + 1 = \\
 &= \frac{\frac{5}{4} - \frac{1}{2} \cdot 2 \cdot \cos(2 \cdot \omega)}{\frac{5}{4} + \frac{1}{2} \cdot 2 \cdot \cos \omega} \cdot 2 + 1 = \\
 &= \frac{\frac{5}{2} - 2 \cdot \cos(2 \cdot \omega)}{\frac{5}{4} + \cos \omega} + 1 = \\
 &= \frac{\frac{7}{2} - 2 \cdot \cos(2 \cdot \omega) + \cos \omega}{\frac{5}{4} + \cos \omega}
 \end{aligned}$$

Rappresentiamo infine questo spettro in figura 25.9.

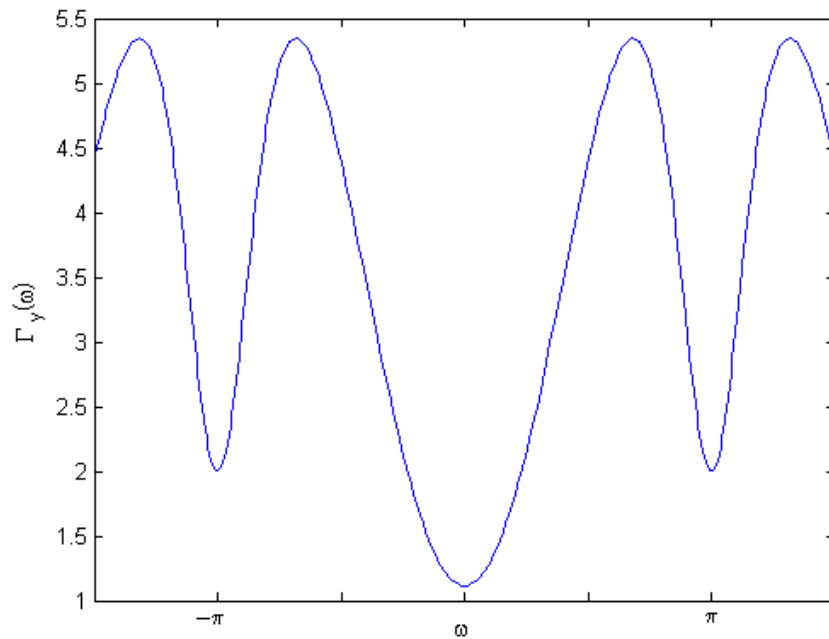


Figura 25.9.: Esercizio 3 - Rappresentazione dello spettro $\Gamma_y(\omega) = \frac{\frac{7}{2} - 2 \cdot \cos(2 \cdot \omega) + \cos \omega}{\frac{5}{4} + \cos \omega}$

Esercizio 7

Testo

Considerando il sistema rappresentato in figura 25.10 e i seguenti dati:

$$\eta(t) = 2 \quad \forall t$$

$$e(t) \sim WN(1, 1)$$

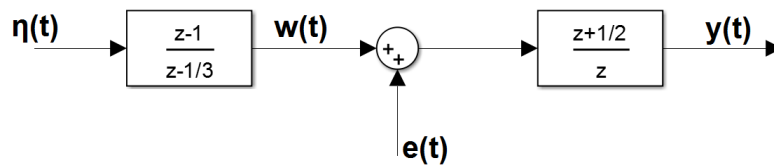


Figura 25.10.: Esercizio 7

1. Di che tipo di processo si tratta?
2. Il processo è un processo stocastico stazionario?
3. Calcolare media m_y e funzione di covarianza $\gamma_y(\tau)$

Soluzione

1. Consideriamo la prima funzione di trasferimento:

$$H_1(z) = \frac{z-1}{z-\frac{1}{3}}$$

Questa ha solo un polo in $\frac{1}{3}$, quindi è asintoticamente stabile. Per il teorema della risposta in frequenza, se l'ingresso del sistema è una sinusoide, l'uscita sarà:

$$\text{Ingresso} \quad \eta(t) = A \cdot \sin(\omega \cdot t + \varphi) = 2 \cdot \sin\left(0 \cdot t + \frac{\pi}{2}\right)$$

$$\text{Uscita} \quad w(t) = |H_1(z)| \cdot \sin(\omega \cdot t + \varphi + \angle H_1(z)) = \left| \frac{e^{j \cdot 0} - 1}{e^{j \cdot 0} - \frac{1}{3}} \right| \cdot 2 \cdot \sin\left(0 \cdot t + \frac{\pi}{2}\right) = 0$$

Il sistema diventa quindi come quello rappresentato in figura :

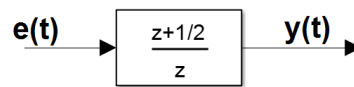


Figura 25.11.: Esercizio 7 - Rappresentazione compatta del sistema

Abbiamo quindi che:

$$\begin{aligned} y(t) &= \frac{z + \frac{1}{2}}{z} \cdot e(t) \\ y(t) &= e(t) + \frac{1}{2} \cdot z^{-1} \cdot e(t) \\ y(t) &= e(t) + \frac{1}{2} \cdot e(t-1) \end{aligned}$$

Che è un processo di tipo $MA(1)$.

2. Essendo un $MA(1)$ con un rumore bianco in ingresso, è stazionario.
 3. Calcoliamo la media:

$$\begin{aligned} m_y &= \mathbb{E}[e(t)] + \frac{1}{2} \cdot \mathbb{E}[e(t-1)] \\ &= 1 + \frac{1}{2} = \\ &= \frac{3}{2} \end{aligned}$$

Siccome è a media non nulla, prima di calcolare la funzione di covarianza, depolarizziamo:

$$\begin{aligned}\tilde{y}(t) &= y(t) - \frac{3}{2} \\ \tilde{e}(t) &= e(t) - 1\end{aligned}$$

$$\begin{aligned}\tilde{y}(t) &= \tilde{e}(t) + \frac{1}{2} \cdot \tilde{e}(t-1) \\ \tilde{y}(t) &= \tilde{e}(t) + \frac{1}{2} \cdot \tilde{e}(t-1)\end{aligned}$$

$$\begin{aligned}\gamma_{\tilde{y}}(0) &= \mathbb{E}[\tilde{y}(t)^2] = \\ &= \mathbb{E}\left[\left(\tilde{e}(t) + \frac{1}{2} \cdot \tilde{e}(t-1)\right)^2\right] = \\ &= \underbrace{\mathbb{E}[\tilde{e}(t)^2]}_{\lambda^2=1} + \frac{1}{4} \cdot \underbrace{\mathbb{E}[\tilde{e}(t-1)^2]}_{\lambda^2=1} + \underbrace{\mathbb{E}[\tilde{e}(t) \cdot \tilde{e}(t-1)]}_0 = \\ &= 1 + \frac{1}{4} = \\ &= \frac{5}{4}\end{aligned}$$

$$\begin{aligned}\gamma_{\tilde{y}}(1) = \gamma_{\tilde{y}}(-1) &= \mathbb{E}[\tilde{y}(t) \cdot \tilde{y}(t-1)] = \\ &= \mathbb{E}\left[\left(\tilde{e}(t) + \frac{1}{2} \cdot \tilde{e}(t-1)\right) \cdot \left(\tilde{e}(t-1) + \frac{1}{2} \cdot \tilde{e}(t-2)\right)\right] = \\ &= \underbrace{\mathbb{E}[\tilde{e}(t) \cdot \tilde{e}(t-1)]}_0 + \frac{1}{2} \cdot \underbrace{\mathbb{E}[\tilde{e}(t) \cdot \tilde{e}(t-2)]}_0 + \frac{1}{2} \cdot \underbrace{\mathbb{E}[\tilde{e}(t-1)^2]}_{\lambda^2=1} + \frac{1}{4} \cdot \underbrace{\mathbb{E}[\tilde{e}(t-1) \cdot \tilde{e}(t-2)]}_0 = \\ &= \frac{1}{2}\end{aligned}$$

Siccome il processo è un $MA(1)$, sappiamo che tutti i $\gamma_y(\tau)$ con $|\tau| > 1$ valgono 0.

4. Calcoliamo lo spettro usando il teorema della fattorizzazione spettrale:

$$\begin{aligned}\Gamma_y(\omega) &= \left| \frac{z + \frac{1}{2}}{z} \right| \cdot \Gamma_e(\omega) = \\ &= \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega}} \cdot \frac{e^{-j\omega} + \frac{1}{2}}{e^{-j\omega}} \cdot 1 = \\ &= 1 + \frac{1}{2} \cdot (e^{j\omega} + e^{-j\omega}) + \frac{1}{4} = \\ &= \frac{5}{4} + \cos \omega\end{aligned}$$

Che è rappresentato in figura 25.12.

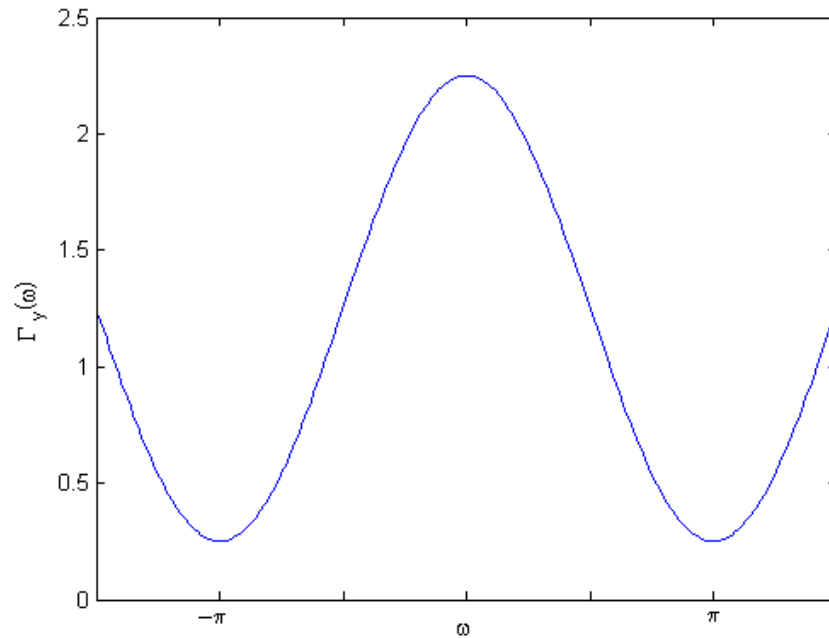


Figura 25.12.: Esercizio 7 - Rappresentazione dello spettro $\Gamma_y(\omega) = \frac{5}{4} + \cos \omega$

Esercizio 8

Testo

Dato il sistema in figura 25.13 e i seguenti dati:

$$W(z) = 1 + \alpha \cdot z^{-1} \quad \text{con} \quad -1 < \alpha < 1$$

$$e(t) \sim WN(0, 1)$$

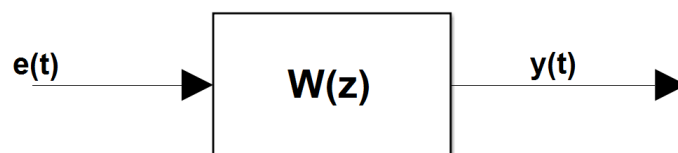


Figura 25.13.: Esercizio 8

Si calcoli α tale che minimizzi $\gamma_y(0)$

Soluzione

Conoscendo la funzione di trasferimento, possiamo calcolare lo spettro con il teorema della fattorizzazione spettrale:

$$\begin{aligned}
 \Gamma_y(\omega) &= |1 + \alpha \cdot z^{-1}| \cdot \Gamma_e(\omega) = \\
 &= (1 + \alpha \cdot e^{-j \cdot \omega}) \cdot (1 + \alpha \cdot e^{j \cdot \omega}) \cdot 1 = \\
 &= 1 + \alpha \cdot (e^{-j \cdot \omega} + e^{j \cdot \omega}) + \alpha^2 = \\
 &= 1 + 2 \cdot \alpha \cdot \cos \omega + \alpha^2
 \end{aligned}$$

Ricordiamo ora che la funzione di covarianza $\gamma_y(\tau)$ si può ricavare antitrasformando lo spettro:

$$\gamma_y(\tau) = \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \Gamma(\omega) \cdot e^{j \cdot \omega \cdot \tau} d\omega$$

da cui possiamo ricavare che:

$$\begin{aligned}
 \gamma_y(0) &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \Gamma(\omega) \cdot e^{j \cdot \omega \cdot 0} d\omega = \\
 &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} \Gamma(\omega) d\omega = \\
 &= \frac{1}{2 \cdot \pi} \cdot \int_{-\pi}^{\pi} 1 + 2 \cdot \alpha \cdot \cos \omega + \alpha^2 d\omega = \\
 &= \frac{1}{2 \cdot \pi} \cdot [\omega + 2 \cdot \alpha \cdot \sin \omega + \alpha^2 \cdot \omega]_{-\pi}^{+\pi} = \\
 &= \frac{2 \cdot \pi + 2 \cdot \pi \cdot \alpha^2}{2 \cdot \pi} = \\
 &= 1 + \alpha^2
 \end{aligned}$$

Derivando:

$$\begin{aligned}
 \frac{\delta \gamma_y(0)}{\delta \alpha} &= 0 \\
 2 \cdot \alpha &= 0 \\
 \alpha &= 0
 \end{aligned}$$

26. Esercitazione del 26/11

Esercizio 1

Testo

Si consideri il sistema seguente:

$$\begin{cases} x(t+1) = \frac{1}{2} \cdot x(t) + v_1(t) & \text{con } v_1(t) \sim WN\left(0, \frac{19}{20}\right) \\ y(t) = 2 \cdot x(t) + v_2(t) & \text{con } v_2(t) \sim WN(0, 1) \end{cases}$$

sapendo che $v_1 \perp v_2 \perp x_0$, dove x_0 è l'inizializzazione dello stato $x(t)$

Trovare:

- $\hat{x}(t|t-1)$ a regime
- $\hat{x}(t|t)$ a regime

Soluzione

Il sistema si trova già nella forma

$$\begin{cases} x(t+1) = F \cdot x(t) + v_1(t) \\ y(t) = H \cdot x(t) + v_2(t) \end{cases}$$

da cui possiamo dedurre che $F = \frac{1}{2}$, $H = 2$, $V_1 = \frac{19}{20}$ (varianza dell'errore di misura), $V_2 = 1$ (varianza dell'errore di modello), $V_{12} = 0$ (covarianza tra errore di misura ed errore di modello, che sappiamo essere a 0 siccome $v_1 \perp v_2$).

Calcolando anche le matrici di osservabilità $O = 2$ e di raggiungibilità $R = \frac{19}{20}$, si può vedere che hanno entrambe rango 1, uguale all'ordine del sistema, che risulta quindi essere osservabile e raggiungibile.

È possibile notare che sia il primo che il secondo teorema di convergenza asintotica sono verificati:

- Primo teorema di convergenza asintotica:
 - Ipotesi:
 - $V_{12} = 0$
 - Sistema as. stabile
 - Tesi:
 - La ARE ha una e una sola soluzione ed è una matrice \bar{P} semidefinita positiva
 - La DRE converge asintoticamente a \bar{P} , $\forall P_0$
 - Il guadagno \bar{K} garantisce l'asintotica stabilità del predittore, ossia $F - \bar{K} \cdot H$ ha autovalori nel cerchio di raggio unitario
- Secondo teorema di convergenza asintotica:

o Ipotesi:

- $V_{12} = 0$
- Il sistema è raggiungibile
- Il sistema è osservabile

o Tesi:

- La ARE ha una e una sola soluzione ed è una matrice \bar{P} definita positiva
- La DRE converge asintoticamente a \bar{P} , $\forall P_0$
- Il guadagno \bar{K} garantisce l'asintotica stabilità del predittore, ossia $F - \bar{K} \cdot H$ ha autovalori nel cerchio di raggio unitario

Quindi la DRE converge ad una matrice \bar{P} definita positiva (in questo specifico caso sarà solamente uno scalare positivo in quanto l'ordine del sistema è 1). Calcolando la DRE :

$$\begin{aligned}
 P(t+1) &= F \cdot P(t) \cdot F^T + V_1 - (F \cdot P(t) \cdot H^T + V_{12}) \cdot (H \cdot P(t) \cdot H^T + V_2)^{-1} \cdot (F \cdot P(t) \cdot H^T + V_{12}) = \\
 &= \frac{1}{2} \cdot P(t) \cdot \frac{1}{2} + \frac{19}{20} - \left(\frac{1}{2} \cdot P(t) \cdot 2 + 0 \right) \cdot (2 \cdot P(t) \cdot 2 + 1)^{-1} \cdot \left(\frac{1}{2} \cdot P(t) \cdot 2 + 0 \right) = \\
 &= \frac{1}{4} \cdot P(t) + \frac{19}{20} - \frac{P(t)^2}{4 \cdot P(t) + 1} = \\
 &= \frac{20 \cdot P(t)^2 + 5 \cdot P(t) + 76 \cdot P(t) + 19 - 20 \cdot P(t)^2}{80 \cdot P(t) + 20} = \\
 &= \frac{81 \cdot P(t) + 19}{80 \cdot P(t) + 20}
 \end{aligned}$$

Che convergerà alla soluzione della ARE :

$$\begin{aligned}
 P(t+1) = P(t) = \bar{P} &= \frac{81 \cdot \bar{P} + 19}{80 \cdot \bar{P} + 20} \\
 \bar{P} &= \frac{81 \cdot \bar{P} + 19}{80 \cdot \bar{P} + 20} \\
 \bar{P} \cdot (80 \cdot \bar{P} + 20) &= 81 \cdot \bar{P} + 19 \\
 80 \cdot \bar{P}^2 - 61 \cdot \bar{P} - 19 &= 0 \\
 \bar{P} &= \begin{cases} 1 \\ -\frac{19}{80} \end{cases}
 \end{aligned}$$

Si può scartare la soluzione $-\frac{19}{80}$ ricordando che $P(t)$ rappresenta la varianza dell'errore di predizione dello stato ad un passo, ed è quindi semidefinita positiva.

Dalla soluzione $\bar{P} = 1$, è possibile calcolare il valore del guadagno del filtro a regime \bar{K} :

$$\begin{aligned}
 \bar{K} &= (F \cdot \bar{P} \cdot H^T + V_{12}) \cdot (H \cdot \bar{P} \cdot H^T + V_2)^{-1} = \\
 &= \frac{\frac{1}{2} \cdot 1 \cdot 2 + 0}{2 \cdot 1 \cdot 2 + 1} = \frac{1}{5}
 \end{aligned}$$

Disegnando ora lo schema a blocchi del filtro di Kalman (rappresentato in figura 26.1)

Figura 26.1.: Esercizio 1 - Schema a blocchi del filtro di Kalman

e concentrandosi solamente sulla funzione di trasferimento da $y(t)$ a $\hat{x}(t|t-1)$, come mostrato in figura 26.2:

Figura 26.2.: Esercizio 1 - Schema a blocchi del predittore di Kalman

L'espressione del predittore di Kalman sarà quindi:

$$\begin{aligned}\hat{x}(t|t-1) &= \frac{\frac{1}{5} \cdot \frac{z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}}}{1 + \frac{1}{5} \cdot \frac{z^{-1}}{1 - \frac{1}{2} \cdot z^{-1}} \cdot 2} \cdot y(t) = \\ &= \frac{\frac{1}{5} \cdot z^{-1}}{1 - \frac{1}{10} \cdot z^{-1}} \cdot y(t)\end{aligned}$$

Da cui l'espressione nel tempo sarà:

$$\hat{x}(t|t-1) = \frac{1}{10} \cdot \hat{x}(t-1|t-2) + \frac{1}{5} \cdot y(t-1)$$

Ricordando infine la relazione che lega predittore e filtro di Kalman:

$$\hat{x}(t|t) = F^{-1} \cdot \hat{x}(t|t-1)$$

e considerando che F è invertibile (in quanto scalare), si ha che:

$$\begin{aligned}\hat{x}(t|t) &= 2 \cdot \hat{x}(t|t-1) = \\ &= \frac{1}{5} \cdot \hat{x}(t-1|t-2) + \frac{2}{5} \cdot y(t-1)\end{aligned}$$

Esercizio 2

Testo

Si consideri il sistema:

$$y(t) = \frac{1}{2} \cdot y(t-1) + e(t) \quad \text{con } e(t) \sim WN(0,1)$$

Si calcoli:

- $\hat{y}(t|t-1)$ secondo la teoria di Kalman
- $\hat{y}(t|t-1)$ secondo la teoria di Kolmogorov-Wiener

Indice analitico

- coefficiente di correlazione, 7
- covarianza, 7
- covarianza campionaria, 43
- Cross-validazione, 104
- densità di probabilità, 5
- densità di probabilità congiunta, 7
- densità spettrale, 12
- depolarizzazione, 15
- distribuzione di probabilità, 5
- error to signal ratio, 56
- ESR, 56
- Filtro passa-tutto, 32
- Forma canonica, 34
- formule di Yule-Walker, 21
- media campionaria, 42
- Modello di Box-Jenkins, 29
- osservabilità, 113
- processi $AR(n)$, 19
- processi $ARMA(m, n)$, 23
- processi $ARMAX(m, n, k + p)$, 24
- processi autoregressivi, 19
- Processi autoregressivi a media mobile, 23
- Processi autoregressivi a media mobile con input esogeni, 24
- processi $MA(n)$, 17
- processi moving average, 17
- processo stazionario in senso debole, 10
- processo stazionario in senso forte, 10
- processo stocastici: equivalenza, 11
- processo stocastico, 9
- processo stocastico: covarianza, 9
- processo stocastico: valor medio, 9
- proprietà della gaussiana di linearità, 6
- proprietà della gaussiana di linearità in distribuzione, 6
- raggiungibilità dal rumore, 114
- raggiungibilità dall'ingresso, 114
- rumore bianco, 11
- segnale persistentemente eccitante, 98
- stimatori, 40
- stimatori: asintotica correttezza, 40
- stimatori: consistenza, 40
- stimatori: correttezza, 40
- stimatori: errore di stima, 40
- stimatori: Ottimalità, 41
- teorema della fattorizzazione spettrale, 30
- valore atteso, 5
- variabile casuale, 5
- variabile casuale gaussiana, 6
- variabile casuale normale, 6
- variabile esogena, 24
- varianza, 5
- vettore di variabili casuali, 6