

Capitolo 9

Identificazione dei modelli

per gli studenti del corso di
Stima e identificazione

Luigi Chisci, 20 Maggio 2020

L'identificazione è la disciplina che si occupa della costruzione di modelli matematici di sistemi o segnali reali di interesse a partire da dati osservati relativi a tali sistemi/segnali. I modelli matematici stanno acquisendo un ruolo sempre più rilevante nella scienza e nell'ingegneria moderne. In particolare, tali modelli sono richiesti per una varietà di scopi specifici quali simulazione, previsione, progettazione di sistemi di controllo, elaborazione di segnali, etc.. È quindi importante saper padroneggiare varie tecniche per la costruzione di modelli. A tale proposito, ci sono essenzialmente due vie contrapposte alla modellazione di sistemi dinamici. Se si conoscono le leggi fisico-matematiche che governano l'evoluzione del sistema reale di interesse, conviene sfruttare questa conoscenza a priori per costruire il modello; si parla, in questo caso, di *modellistica basata sui principi fondamentali* o più semplicemente *modellistica*. Se, viceversa, non si hanno dettagli sulla fisica del sistema ma si dispone di dati misurati per tale sistema, si possono utilizzare questi ultimi per la costruzione del modello; si parla, in questo caso, di *modellistica basata sui dati* o, più semplicemente, *identificazione*. Nella pratica, si hanno spesso informazioni solo parziali sulla fisica del sistema per cui conviene procedere alla costruzione del modello utilizzando tali informazioni a priori in combinazione con dati sperimentali sul sistema.

Con questo capitolo ed il successivo si intende presentare una breve introduzione all'identificazione volta a fornire le conoscenze teoriche di base nonché ad illustrare le principali tipologie di modelli e tecniche di elaborazione dati impiegate per costruire tali modelli dai dati.

9.1 Procedura di identificazione

Schematicamente, vedi Fig. 1, l'identificazione è un processo di elaborazione dati che ha in ingresso un insieme di dati (*dataset*), esito dell'esperimento sul sistema o segnale reale di interesse, e produce in uscita un modello matematico di tale sistema o segnale, sfruttando anche l'informazione a priori disponibile.

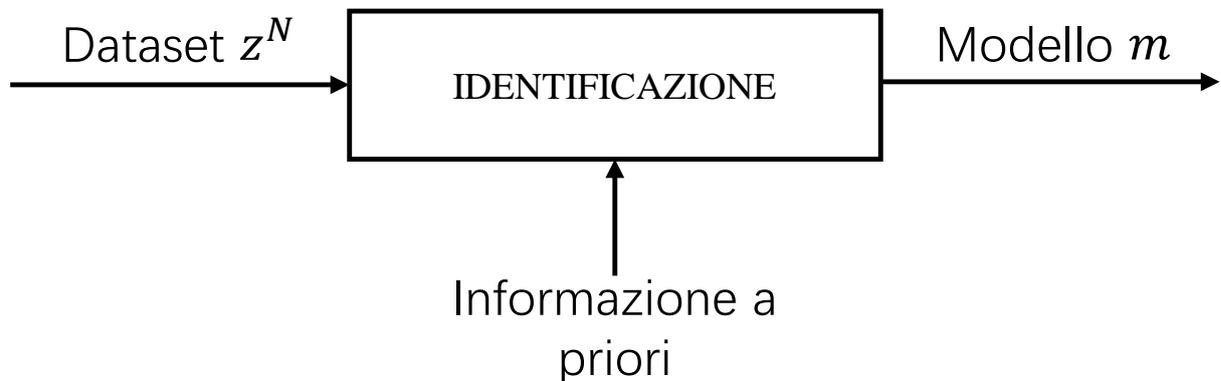


Figura 9.1: Schematizzazione ingresso-uscita del processo di identificazione: dai dati al modello, sfruttando l'informazione a priori.

Dataset

Per quanto riguarda il *dataset*

$$z^N = z_{1:N} = \{z_k\}_{k=1}^N = \{z_1, z_2, \dots, z_N\} \quad (9.1.1)$$

si suppone che esso consista di N campioni

$$z_k = \begin{cases} \begin{bmatrix} u_k \\ y_k \end{bmatrix} & \text{nel caso di identificazione di un sistema dinamico} \\ & \text{con ingresso } u \text{ ed uscita } y \\ y_k & \text{nel caso di identificazione di un segnale } y, \end{cases} \quad (9.1.2)$$

acquisiti ad istanti diversi t_k , con $t_1 < t_2 < \dots < t_N$. Di norma, i campioni z_k del *dataset* vengono acquisiti ad intervalli regolari, i.e. $t_k = kT$ dove $T > 0$ è un intervallo di campionamento costante. Se, viceversa, si adotta un campionamento irregolare occorre ridefinire il *dataset* come

$$z^N = \{(t_k, z_k)\}_{k=1}^N \quad (9.1.3)$$

includendo anche gli istanti di campionamento t_k .

Modello

L'obiettivo dell'identificazione è la determinazione di un modello che spieghi al meglio il comportamento del sistema (o segnale) reale di interesse sulla base del *dataset* disponibile. A tale proposito, si possono adottare diverse tipologie di modelli (e.g., a *tempo-continuo/tempo-discreto*, *lineari/non lineari*, *deterministici/stocastici*, *tempo-invarianti/tempo-varianti*, a *parametri concentrati/distribuiti*, a *scatola nera/grigia*, *parametrici/non parametrici*, etc.) che saranno illustrate e discusse nel seguito di questo capitolo.

Informazione a priori

La determinazione del modello m a partire dal *dataset* z^N si avvale anche di eventuale informazione a priori disponibile. A seconda del contesto, tale informazione può variare dalla completa ignoranza (*modello a scatola nera*) fino alla situazione opposta in cui siano note le equazioni differenziali che descrivono il sistema e debbano essere determinati solamente i valori numerici di alcuni parametri (*modello a scatola grigia*).

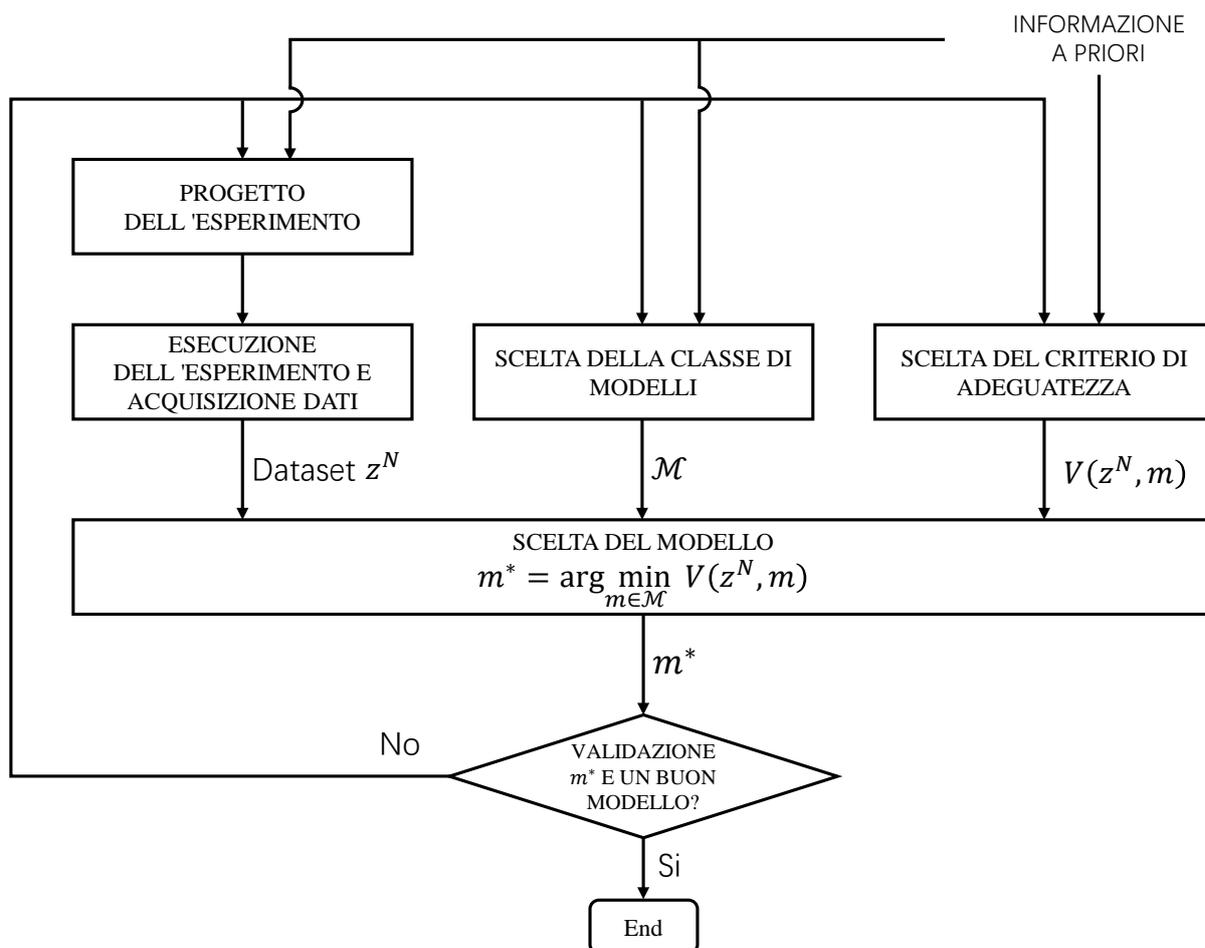


Figura 9.2: Diagramma di flusso del processo di identificazione.

Fasi della procedura di identificazione

Il compito dell'identificazione viene di norma assolto con il procedimento iterativo schematizzato nel diagramma di flusso di Fig. 2, che illustra come arrivare ad un modello

soddisfacente m^* del sistema o segnale reale di interesse attraverso l'interazione di varie fasi:

- progetto dell'esperimento, sua esecuzione e conseguente acquisizione del *dataset* z^N ;
- scelta della classe di modelli \mathcal{M} ;
- scelta del criterio di adeguatezza $V(z^N, m)$ con il quale giudicare la bontà di un modello m per un fissato *dataset* z^N ;
- scelta del modello ottimo m^* che minimizza, fra tutti i modelli della classe \mathcal{M} , il criterio di adeguatezza $V(z^N, m)$ per l'assegnato *dataset* z^N ;
- validazione del modello m^* per decidere se risulta adeguato agli scopi preposti (termine della procedura con il modello m^*) oppure no (revisione di qualcuna delle scelte fatte su esperimento, classe di modelli, criterio di adeguatezza e reiterazione della procedura).

Ciascuno dei blocchi di Fig. 2 richiede commenti specifici.

Progetto dell'esperimento - Il progetto di un esperimento di identificazione comporta la scelta di quali segnali misurare e quando misurarli (intervallo di campionamento nel caso di campionamento regolare o, viceversa, istanti di campionamento) nonché la scelta di eventuali ingressi manipolabili e del loro andamento temporale. Dal momento che, ai fini dell'elaborazione dati, la loro acquisizione viene effettuata in digitale, i segnali misurati devono essere campionati. Pertanto, un'ulteriore scelta pratica da effettuare è quella dei filtri analogici *antialiasing* di pre-campionamento. Infine, occorre scegliere la durata dell'esperimento di identificazione e, conseguentemente, il numero di campioni N risultante dall'esperimento che deve essere abbastanza elevato da garantire statistiche sufficientemente accurate dei segnali coinvolti nonché da permettere di riservare una parte dei dati per la validazione. L'aspetto cruciale per il successo dell'identificazione di un sistema dinamico, con ingresso u ed uscita y , è certamente la scelta di un ingresso u che ecciti opportunamente il sistema in modo da produrre un *dataset* sufficientemente informativo, anche in relazione alla classe di modelli adottata. Si fa notare come la scelta dell'esperimento, fra le varie scelte del procedimento di identificazione, è sicuramente quella più critica nel senso che l'effettuazione di un esperimento comporta spesso costi elevati sia per la necessità di sottrarre il sistema reale da identificare (e.g. processo industriale, mezzo di trasporto) alle normali mansioni a cui esso è adibito, che per la possibile usura degli attuatori del sistema conseguente all'esigenza di dover eccitare adeguatamente il sistema. Pertanto, nell'ipotesi in cui il modello selezionato m^* non passi il test di validazione, è opportuno modificare, senza alcun costo aggiuntivo, la scelta della classe di modelli e/o del criterio di adeguatezza piuttosto che modificare e, quindi,

dover ri-effettuare l'esperimento. Per questo motivo, l'esperimento di identificazione deve essere oculatamente progettato.

Esecuzione dell'esperimento e acquisizione dati - Questa fase prevede l'esecuzione dell'esperimento di identificazione progettato con conseguente acquisizione digitale dei dati sperimentali da opportuni sensori e memorizzazione nel *dataset* z^N . A seguito dell'acquisizione dati, si possono effettuare eventuali pre-elaborazioni dei dati per eliminare disturbi a bassa- ed alta-frequenza tramite opportuni filtri digitali, individuare e rimuovere dati anomali (*outlier*), ricostruire dati mancanti, etc..

Scelta della classe di modelli - Un passo fondamentale della procedura di identificazione è quello di scegliere una classe di modelli \mathcal{M} all'interno della quale ricercare il modello candidato m^* . È opportuno premettere che un modello, per quanto complesso ed accurato, non potrà mai coincidere con il sistema o segnale reale di interesse ma dovrà catturarne solamente le caratteristiche salienti per poter servire efficacemente allo scopo preposto (e.g. simulazione, previsione, progetto del controllo, etc.). Nel successivo paragrafo 9.2, si procederà ad una classificazione dei modelli sulla base di diversi aspetti. Una importante classificazione è quella fra modelli parametrici che sono caratterizzati da un vettore di parametri a dimensione finita, e modelli non parametrici che, viceversa, non impiegano esplicitamente un vettore di parametri nella ricerca del miglior modello. Per limiti di tempo, l'attenzione in questo corso verrà esclusivamente rivolta a modelli *lineari tempo-invarianti*, sia parametrici che non parametrici.

Scelta del criterio di adeguatezza - All'interno della classe di modelli \mathcal{M} prescelta vi sono infiniti modelli candidati per cui è opportuno introdurre un criterio di adeguatezza, ovvero una funzione $V(z^N, m)$ del *dataset* z^N e del modello $m \in \mathcal{M}$, per quantificare la bontà del modello nel senso che un modello m_1 è migliore (più adeguato) di m_2 per il fissato *dataset* z^N se $V(z^N, m_1) < V(z^N, m_2)$. Si fa notare come il criterio di adeguatezza $V(\cdot, \cdot)$ sia funzione anche del *dataset* oltre che del modello, anche se il confronto fra modelli diversi deve essere ovviamente effettuato per un fissato *dataset*. La scelta della funzione $V(\cdot, \cdot)$ deve essere chiaramente legata alla finalità del modello e, a tale proposito, sono possibili diverse scelte che pongono diversa enfasi sull'errore di simulazione oppure sull'errore di predizione oppure sull'errore di modello. In questo corso, per limiti di tempo, si farà esclusivamente ricorso al criterio dell'*errore di predizione* per modelli parametrici mentre si utilizzeranno metodi di *analisi di correlazione* e di *analisi spettrale* per i modelli non parametrici.

Scelta del modello - Una volta scelti la classe di modelli \mathcal{M} ed il criterio di adeguatezza $V(z^N, m)$, il problema di identificazione si riduce, nella maggior parte dei casi, alla

soluzione del seguente problema di ottimizzazione:

$$m^* = \arg \min_{m \in \mathcal{M}} V(z^N, m). \quad (9.1.4)$$

Tale problema di ottimizzazione risulta computazionalmente trattabile se la classe di modelli \mathcal{M} ha un numero finito di elementi oppure se i suoi elementi, seppur in numero infinito, sono parametrizzati da un vettore $\theta \in \mathbb{R}^d$ finito-dimensionale, i.e. $m = m(\theta)$ e $\mathcal{M} = \text{Im}(m) \triangleq \{m(\theta), \theta \in \mathbb{R}^d\}$, per cui il problema di identificazione parametrica si riduce a:

$$\begin{cases} \theta^* = \arg \min_{\theta \in \mathbb{R}^d} V(z^N, m(\theta)) \\ m^* = m(\theta^*). \end{cases} \quad (9.1.5)$$

Viceversa, nel caso di modelli non parametrici, l'approccio di ottimizzazione (9.1.4) diventa intrattabile per cui si procede alla determinazione sperimentale del modello parametrico m^* sulla base del *dataset* z^N con altri metodi.

Validazione - Una volta determinato il modello candidato m^* , ci si pone il problema di valutare se tale modello risulta adeguato o meno agli scopi preposti; tale problema prende il nome di *validazione del modello*. La validazione ha un esito binario: se il modello viene ritenuto valido, la procedura di identificazione viene arrestata con il modello m^* ; se, viceversa, il modello non viene accettato si procede ad una nuova iterazione, non prima di avere modificato qualcuna delle scelte fatte in precedenza (l'esperimento e/o la classe di modelli e/o il criterio di adeguatezza) in modo da ottenere un nuovo modello m^* . Ricordando quanto detto in precedenza, si dovrebbe evitare per quanto possibile di ripetere l'esperimento di identificazione e piuttosto utilizzare lo stesso *dataset* z^N con altre classi di modelli e/o criteri di adeguatezza. Ingenuamente, si potrebbe pensare che un buon criterio di validazione è quello di considerare l'adattamento (*fitting*) del modello ai dati utilizzati nell'identificazione. Questo, in realtà, è un comune errore da non commettere: la bontà di un modello non può essere giudicata sulla base degli stessi dati impiegati per l'identificazione. Viceversa, un possibile criterio è l'adattamento del modello con nuovi dati sperimentali, detti di validazione, diversi da quelli di identificazione (*cross-validazione*) da cui l'esigenza di riservare parte dei dati sperimentali per la validazione. Oltre alla *cross-validazione*, esistono altri criteri e/o strumenti per la validazione alcuni dei quali sono di uso generale mentre altri sono legati alla finalità specifica a cui è destinato il modello.

Come evidenziato in fig. 2, le varie scelte (in particolare quelle iniziali) coinvolte nel processo di identificazione sono guidate dall'informazione a priori disponibile. Fig. 2 mostra anche come il processo di identificazione sia complesso ed articolato, con diverse fasi

fra loro interconnesse, per ciascuna delle quali sono richieste basi teoriche nonché algoritmi di elaborazione dati e strumenti software per la loro implementazione. Al giorno d'oggi, l'utente dell'identificazione si può avvalere di pacchetti software interattivi (quali ad esempio Matlab [®] System Identification Toolbox) che lo assistono in tutte le fasi del procedimento. In questo contesto, un buon libro di testo o corso di identificazione deve fornire le conoscenze di base, sia teoriche che pratiche, per utilizzare con cognizione di causa un tale software interattivo. A tale proposito, il libro di testo [1] - a cura dell'autore, Prof. Lennart Ljung dell'Università di Linköping in Svezia, del *System Identification Toolbox* - fornisce un'ottima, onnicomprensiva, guida all'identificazione assistita da calcolatore. In quest'ottica, anche questi due capitoli (9 e 10) intendono offrire una panoramica, seppur incompleta a causa delle limitazioni di tempo di questo corso, delle classi di modelli e delle tecniche di stima di modelli più comuni, con particolare riferimento all'identificazione di modelli lineari tempo-invarianti.

9.2 Classificazione dei modelli

Questo paragrafo fornisce una panoramica della grande varietà di modelli esistenti per l'identificazione, suddivisi in accordo a vari criteri di classificazione.

Modelli a tempo-continuo (TC)/a tempo-discreto (TD)

Una prima importante distinzione è fra modelli a tempo-continuo (TC) descritti da equazioni differenziali sull'insieme dei tempi reale $\mathcal{T} = \mathbb{R}$, e modelli a tempo-discreto (TD) descritti da equazioni alle differenze sull'insieme dei tempi intero $\mathcal{T} = \mathbb{Z}$. I modelli TC rappresentano la soluzione naturale per la modellistica basata sui principi fondamentali di sistemi fisici. Viceversa, i modelli TD sono più convenientemente impiegati per la modellistica basata sui dati, proprio perché questi ultimi riguardano segnali campionati, quindi a tempo-discreto.

Modelli di segnale/di sistema

Una ulteriore distinzione è legata all'obiettivo dell'identificazione, ovvero se si mira al modello di un segnale y ovvero di un sistema dinamico con ingresso u ed uscita y . Esempio classico di modello di segnale è il modello ARMA trattato in dettaglio nel capitolo 4. Esempi di modelli di sistema sono il modello di stato

$$\begin{aligned}x_{k+1} &= f_k(x_k, u_k) \\ y_k &= h_k(x_k, u_k)\end{aligned}\tag{9.2.1}$$

oppure il modello ingresso-uscita

$$y_k = g_k(y_{k-1}, \dots, y_{k-n}, u_k, u_{k-1}, \dots, u_{k-m}).\tag{9.2.2}$$

Modelli deterministici/stocastici

I modelli *deterministici*, come (9.2.1) e (9.2.2), non tengono conto di eventuali disturbi che agiscono sul sistema e questo, in molti casi, risulta irrealistico. Negli sviluppi successivi si farà esclusivo riferimento a modelli *stocastici* come, ad esempio, il modello di stato

$$\begin{aligned}x_{k+1} &= f_k(x_k, u_k, w_k) \\ y_k &= h_k(x_k, u_k, v_k)\end{aligned}\tag{9.2.3}$$

dove gli ingresso stocastici w_k (disturbo di processo) e v_k (rumore di misura) tengono conto delle incertezze di modello e, rispettivamente, di misura. In modo analogo, si può avere la seguente controparte stocastica del modello ingresso-uscita (9.2.2)

$$y_k = g_k(y_{k-1}, \dots, y_{k-n}, u_k, u_{k-1}, \dots, u_{k-m}, e_k, e_{k-1}, \dots, e_{k-l}).\tag{9.2.4}$$

con un unico segnale stocastico e_k che tiene conto dell'incertezza.

Modelli lineari/non-lineari

Una importante classificazione è fra modelli lineari, caratterizzati dalla linearità delle equazioni differenziali o alle differenze che li descrivono, e modelli che, viceversa, presentano non-linearità in tali equazioni. Sebbene ogni sistema reale presenti qualche non-linearità, l'attenzione sarà limitata a modelli lineari che comunque, in molte situazioni, forniscono una ragionevole descrizione del sistema reale di interesse a fini di simulazione, previsione, sintesi del controllo, etc..

Modelli tempo-invarianti/tempo-varianti

Un ulteriore criterio di classificazione riguarda la tempo-varianza del modello, ovvero la dipendenza temporale delle funzioni g_k e/o h_k in (9.2.3), g_k in (9.2.4), dal tempo k . Tale dipendenza è solitamente legata alla variazione temporale di alcuni parametri caratteristici del sistema. Assumendo tali variazioni contenute e/o lente, la successiva trattazione sarà limitata a modelli tempo-varianti. Riassumendo, negli sviluppi successivi si farà esclusivamente riferimento a modelli *lineari tempo-invarianti* (LTI) *stocastici*, sia nello spazio di stato

$$\begin{aligned}x_{k+1} &= Ax_k + Bu_k + w_k \\ y_k &= Cx_k + Du_k + v_k\end{aligned}\tag{9.2.5}$$

che ingresso-uscita

$$y_k = G(z)u_k + v_k\tag{9.2.6}$$

dove w_k e v_k sono processi stocastici stazionari a media nulla.

Modelli scalari/multivariabili

Un'altra importante classificazione dei sistemi e, quindi, dei modelli si basa sul numero di ingressi ed uscite (nel caso di sistemi dinamici) oppure sul numero di componenti (canali) del segnale y di interesse. Indicata con m la dimensione di u e con p la dimensione di y , un modello di sistema dicesi:

SISO	(Single Input Single Output)	se	$m = 1$	e	$p = 1$;
MIMO	(Multiple Input Multiple Output)	se	$m > 1$	e	$p > 1$;
SIMO	(Single Input Multiple Output)	se	$m = 1$	e	$p > 1$;
MISO	(Multiple Input Single Output)	se	$m > 1$	e	$p = 1$.

Nel caso di modelli di segnale, si parla di modello:

a canale singolo/scalare	se	$p = 1$;
multicanale/multivariabile	se	$p > 1$.

Sebbene per tutti i modelli, sia scalari che multivariabili, si possano indifferentemente usare sia rappresentazioni di stato che ingresso-uscita vale la pena osservare come per modelli multivariabili sia di gran lunga preferibile ricorrere a rappresentazioni di stato, evitando le complicazioni delle rappresentazioni ingresso-uscita matriciali. Per modelli scalari, viceversa, soprattutto nel caso LTI, le rappresentazioni ingresso-uscita (e.g., funzione di trasferimento, risposta impulsiva, risposta in frequenza) risultano particolarmente convenienti.

Modelli a parametri concentrati/distribuiti

I modelli a *parametri distribuiti* descrivono il sistema mediante *equazioni differenziali alle derivate parziali* (PDE = *Partial Differential Equations*), diversamente dai modelli a *parametri concentrati* che utilizzano *equazioni differenziali ordinarie* (ODE = *Ordinary Differential Equations*). Poiché, mediante opportune tecniche di discretizzazione spaziale alle *differenze finite* oppure agli *elementi finiti*, è sempre possibile approssimare, con arbitrario livello di precisione, un modello a parametri distribuiti con un modello a parametri concentrati di ordine sufficientemente elevato, la trattazione sarà limitata a modelli a parametri concentrati.

Modelli parametrici/non parametrici

Un modello parametrico è completamente caratterizzato da un vettore di parametri $\theta \in \mathbb{R}^d$, $d < \infty$, finito-dimensionale. Un esempio classico di modello parametrico è il modello ARMA(n_a, n_c)

$$\begin{cases} y_k + a_1 y_{k-1} + \dots + a_{n_a} y_{k-n_a} = e_k + c_1 e_{k-1} + \dots + c_{n_c} e_{k-n_c} \\ e_k = swn(\bar{e}, r) \end{cases} \quad (9.2.7)$$

caratterizzato dal vettore di parametri $\theta = [a_1, \dots, a_{n_a}, c_1, \dots, c_{n_c}, \bar{e}, r]^T \in \mathbb{R}^d$, con $d = n_a + n_c + 2$. In realtà i parametri del modello ARMA sono vincolati a soddisfare alcune condizioni, i.e. $r > 0$ inoltre i polinomi $A(z) = 1 + a_1 z^{-1} + \dots + a_n z^{-n_a}$ e $C(z) = 1 + c_1 z^{-1} + \dots + c_{n_c} z^{-n_c}$ devono avere tutte le radici in $|z| < 1$. Pertanto

$$\theta \in \Theta = \{\theta : A(z) \text{ e } C(z) \text{ hanno tutte le radici in } |z| < 1, r > 0\} \subset \mathbb{R}^{n_a+n_c+2}.$$

Si noti come l'identificazione del modello ARMA (9.2.7) coincida con la stima parametrica del vettore $\theta \in \Theta$. Si noti anche come la definizione di una classe di modelli parametrici comporti una scelta a priori dell'ordine massimo del modello, nel caso specifico definito dagli indici strutturali n_a e n_c .

Viceversa, un modello non parametrico non è individuato da un vettore di parametri finito-dimensionale e, di conseguenza, non presuppone alcuna assunzione a priori sull'ordine massimo del modello. Con riferimento al modello LTI (9.2.6) un classico modello non parametrico nel dominio del tempo è costituito dalla *risposta impulsiva* $\{g_k\}_{k=0}^{\infty}$ corrispondente alla funzione di trasferimento $G(z) = \sum_{k=0}^{\infty} g_k z^{-k}$ e dalla *funzione di auto-covarianza* $\{R_v(k)\}_{k=0}^{\infty}$ del disturbo stocastico v_k . Riassumendo, un modello non parametrico LTI nel dominio del tempo è definito da:

$$m = \{\{g_k\}_{k=0}^{\infty}, \{R_v(k)\}_{k=0}^{\infty}\}.$$

Si può anche considerare un modello non parametrico LTI nel dominio della frequenza costituito dalla *risposta in frequenza* $\{G(e^{j\omega})\}_{\omega \in [0, \pi]}$ e dalla densità spettrale $\{\varphi_v(\omega)\}_{\omega \in [0, \pi]}$ del disturbo v_k , i.e.

$$m = \left\{ \left\{ G(e^{j\omega}) \right\}_{\omega \in [0, \pi]}, \left\{ \varphi_v(\omega) \right\}_{\omega \in [0, \pi]} \right\}.$$

Modelli a scatola nera/grigia

In molte situazioni pratiche non si ha a disposizione alcuna informazione a priori sul sistema o segnale reale di interesse. In questo caso, si deve quindi ricorrere ad un modello cosiddetto *a scatola nera* (BB = *black box*) che non presuppone alcuna considerazione fisica sulla struttura del sistema. La terminologia utilizzata, infatti, fa riferimento al fatto che la modellazione del sistema viene effettuata *alla cieca* come se il sistema fosse racchiuso all'interno di una scatola nera all'interno della quale non è possibile vedere niente e, quindi, dedurre informazioni di alcun tipo sulla struttura interna del sistema. Un esempio classico di modello BB relativo ad un segnale è il modello parametrico ARMA (9.2.7) derivante da considerazioni matematiche (teoremi di approssimazione e fattorizzazione spettrale) sulla rappresentazione di segnali casuali stazionari, ma non legato in alcun modo alla natura fisica del segnale considerato. Si noti come i parametri θ del modello ARMA non abbiano alcuna interpretazione fisica ma costituiscano unicamente un veicolo per aggiustare l'adattamento del modello ai dati sperimentali. Altri esempi di modelli BB sono i modelli non parametrici LTI, nel domini del tempo e rispettivamente della frequenza,

visti in precedenza. Anche in questo caso, il modello si può adattare ad una gran varietà di sistemi prescindendo dalla loro natura fisica purché il loro comportamento sia, in prima approssimazione, lineare e stazionario.

Ci sono anche situazioni nelle quali, basandosi sulle leggi fondamentali della fisica o di altre discipline, si riesce a descrivere il comportamento del sistema attraverso equazioni matematiche, tipicamente equazioni differenziali, nelle quali compaiono alcuni parametri da determinare. In questo caso, il modello risultante da queste equazioni con parametri incogniti prende il nome di modello *a scatola grigia* (GB = *grey box*) alludendo al fatto che la modellazione del sistema avviene come se questo fosse all'interno di una scatola grigia attraverso la quale è possibile intravedere qualcosa ma non tutto e, quindi, dedurre informazione parziale (ma non completa) sulla struttura interna del sistema. Si pensi all'esempio di una rete elettrica di cui si conoscano la tipologia e l'interconnessione di tutti i componenti, ma risultino non noti oppure incerti i valori numerici di alcuni parametri circuitali eventualmente parassiti. In questo caso, sfruttando le leggi di Kirchhoff e le equazioni costitutive dei vari componenti elettrici, si possono scrivere le equazioni di stato circuitali in funzione di alcuni parametri incogniti.

Risulta evidente come i modelli GB siano tipicamente

- a tempo-continuo, essendo le leggi fisiche solitamente espresse sotto forma di equazioni differenziali;
- parametrici, per la necessità di adattare alcuni parametri che compaiono nelle equazioni del modello ai dati sperimentali;
- espressi sotto forma di rappresentazioni di stato che meglio si prestano delle rappresentazioni ingresso-uscita a tener conto di informazioni sulla struttura interna del sistema.

Pertanto, un generico modello GB è descritto da equazioni di stato

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t), w(t), \theta) \\ y(t) = h(t, x(t), u(t), v(t), \theta) \end{cases} \quad (9.2.8)$$

con stato $x(t)$, ingresso deterministico $u(t)$, disturbo di processo $w(t)$, rumore di misura $v(t)$, vettore dei parametri da stimare θ . Nel caso particolare LTI, il modello GB assume la forma

$$\begin{cases} \dot{x}(t) = A_c(\theta)x(t) + B_c(\theta)u(t) + D_c(\theta)w(t) \\ y(t) = C(\theta)x(t) + v(t) \\ w(t) = \text{swn}(0, Q_c(\theta)) \\ v(t) = \text{swn}(0, R(\theta)) \end{cases} \quad (9.2.9)$$

dove si è realisticamente ipotizzata l'assenza del collegamento diretto (*feedthrough*) fra ingresso e uscita (i.e., l'ingresso $u(t)$ non compare nell'equazione che definisce $y(t)$). Si noti

che, per maggior generalità, tutte le componenti delle matrici $A_c(\theta)$, $B_c(\theta)$, $D_c(\theta)$, $Q_c(\theta)$, $C(\theta)$, $R(\theta)$ possono dipendere dai parametri θ anche se, nella maggior parte dei casi, la dipendenza dai parametri riguarda poche componenti. Il seguente modello di motore elettrico in corrente continua fornisce un esempio illustrativo di modello GB.

Esempio (Modello di motore in corrente-continua) - Si consideri il motore in corrente-continua (dc = *direct current*) schematizzato in fig. 9.3. controllo di armatura,

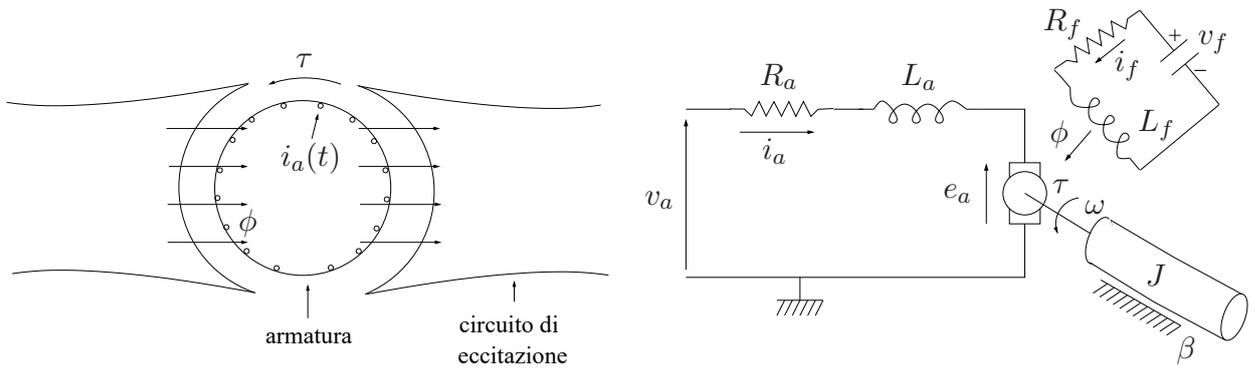


Figura 9.3: Schema fisico (a sinistra) e circuitale (a destra) del motore elettrico in corrente-continua

ovvero alimentando il circuito di campo a tensione costante, risulta che la coppia motrice $\tau(t)$ esercitata dal motore è proporzionale alla corrente di armatura $i_a(t)$ secondo la relazione

$$\tau(t) = k_T i_a(t). \quad (9.2.10)$$

Inoltre, per effetto dell'induzione elettromagnetica, viene generata nel circuito di armatura una *forza contro-elettromotrice indotta*

$$e_a(t) = k_b \omega(t) \quad (9.2.11)$$

proporzionale alla velocità angolare del rotore $\omega(t)$. La legge di Kirchhoff delle tensioni al circuito di armatura fornisce la relazione

$$L_a \frac{di_a}{dt}(t) + R_a i_a(t) + e_a(t) = v_a(t). \quad (9.2.12)$$

Inoltre, dalla legge di Newton relativa alla parte meccanica del motore si ha

$$\begin{cases} J\dot{\omega}(t) &= \tau(t) - \tau_r(t) - \beta\omega(t) \\ \dot{\psi}(t) &= \omega(t) \end{cases} \quad (9.2.13)$$

dove: β (coefficiente di attrito viscoso) e J (momento di inerzia) sono i parametri meccanici del motore; $\beta\omega(t)$ e $\tau_r(t)$ sono coppie resistenti dovute rispettivamente all'attrito

volvente viscoso ed al carico meccanico del motore; $\psi(t)$ è la posizione angolare del rotore. Tenuto conto che, di norma, la costante di tempo elettrica L_a/R_a è molto minore (di diversi ordini di grandezza) della costante di tempo meccanica J/β , si può trascurare l'induttanza di armatura, i.e. porre $L_a = 0$. Definito il vettore di stato $x = [\psi, \omega]^T$, l'ingresso manipolabile $u = v_a$, il disturbo di processo $w = \tau_r$ e l'uscita $y = \psi$, si ottengono le seguenti equazioni di stato lineari

$$\begin{cases} \dot{x}(t) = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & -\theta_1 \end{bmatrix}}_{A_c(\theta)} x(t) + \underbrace{\begin{bmatrix} 0 \\ \theta_2 \end{bmatrix}}_{B_c(\theta)} u(t) + \underbrace{\begin{bmatrix} 0 \\ -\theta_3 \end{bmatrix}}_{D_c(\theta)} w(t) \\ y(t) = \underbrace{[1, 0]}_C x(t) + v(t) \\ w(t) = \text{swn}(0, q) \\ v(t) = \text{swn}(0, r) \end{cases} \quad (9.2.14)$$

dipendenti dal vettore dei parametri $\theta = [\theta_1, \theta_2, \theta_3]^T \in \mathbb{R}^3$ con

$$\theta_1 = \frac{\beta R_a + k_T k_b}{J R_a}, \quad \theta_2 = \frac{k_T}{J R_a}, \quad \theta_3 = \frac{1}{J}.$$

In questo caso, $Q_c = q$ ed $R = r$ sono assunte indipendenti dai parametri. \square

9.3 Segnali quasi-stazionari

I segnali di uscita y risultanti dagli esperimenti di identificazione sono spesso descritti - vedi, ad esempio, modelli (9.2.5) e (9.2.6) - come processi stocastici, dovuti alla presenza di disturbi, con componenti deterministiche, dovute alla presenza dell'ingresso manipolabile u . Tali segnali non sono, in generale, stazionari; ad esempio il segnale y_k in (9.2.6) ha funzione di media

$$E[y_k] = E[G(z)u_k + v_k] = G(z)u_k + E[v_k] = G(z)u_k$$

che risulta tempo-variante, quindi non è stazionario. Di conseguenza, ai segnali in oggetto non sono direttamente applicabili definizioni nonché tecniche di analisi teorica e sperimentale proprie dei processi stocastici stazionari, quali ad esempio:

- lo spettro e la densità spettrale introdotti nel capitolo 3;
- il teorema di analisi del capitolo 4;
- il correlogramma ed il periodogramma del capitolo 4.

Nel seguito si introduce una classe di segnali, detti *quasi-stazionari*, che generalizza quella dei processi stazionari in senso lato ed, in particolare, include i segnali stocastici stazionari con componenti deterministiche generati dai modelli LTI (9.2.5) e (9.2.6).

Definizione (Segnale quasi-stazionario) - Un segnale tempo-discreto y_k dicesi *quasi-stazionario* se soddisfa quanto segue:

$$E[y_k] = \bar{y}_k, \quad \|\bar{y}_k\| < C < \infty, \forall k \quad (9.3.1)$$

$$E[y_k y_j^T] = C_Y(k, j), \quad \|C_Y(k, j)\| \leq C, \forall k, j \quad (9.3.2)$$

$$\lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N C_Y(k, k-i) = \bar{R}_Y(i), \quad \forall i \quad (9.3.3)$$

dove l'operatore di media $E[\dots]$ fa riferimento alle componenti stocastiche di y_k . \square

Si noti che:

- un segnale deterministico y_k , per il quale l'operatore di media risulta ininfluenza, è quasi-stazionario se è limitato ed esiste il limite

$$\bar{R}_Y(i) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N y_k y_{k-i}^T; \quad (9.3.4)$$

- un processo stocastico stazionario in senso lato y_k è certamente quasi-stazionario in quanto $C_Y(k, k-i) \triangleq E[y_k y_{k-i}^T] = \bar{R}_Y(i)$ per ogni k .

Si noti anche come $\bar{R}_Y(i)$ definita in (9.3.3), i.e.

$$\bar{R}_Y(i) = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N E[y_k y_{k-i}^T], \quad (9.3.5)$$

costituisca una generalizzazione della funzione di autocovarianza $R_Y(i) \triangleq E[\tilde{y}_k \tilde{y}_{k-i}^T]$, con $\tilde{y}_k \triangleq y_k - E[y_k]$, precedentemente definita per processi stocastici stazionari in senso lato. In particolare, se y_k ha media nulla risulta banalmente che $\bar{R}_Y(i) = R_Y(i)$ mentre, più in generale, $\bar{R}_Y(i) = R_Y(i) + \bar{y} \bar{y}^T$ dove $\bar{y} = E[y_k]$. Per un segnale deterministico, viceversa, $\bar{R}_Y(i)$ in (9.3.5) si riduce alla media campionaria (correlogramma) (9.3.4). Risulta conveniente, alla luce di queste considerazioni ed in vista degli sviluppi futuri, introdurre il seguente *operatore di media generalizzata* $\bar{E}[\dots]$

$$\bar{E}[f_k] \triangleq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N E[f_k]. \quad (9.3.6)$$

Si noti che:

- tale operatore \bar{E} effettua sul suo argomento una doppia media, cioè la usuale media matematica $E[\dots]$ e la media campionaria $N^{-1} \sum_{k=1}^N [\dots]$;
- la media matematica E non ha effetto su segnali deterministici;
- la media campionaria (temporale) non ha effetto su segnali stocastici stazionari;
- grazie all'operatore \bar{E} , $\bar{R}_Y(i)$ in (9.3.5) può essere riespressa come

$$\bar{R}_Y(i) = \bar{E} [y_k y_{k-i}^T] \quad (9.3.7)$$

e, quindi, interpretata come *funzione di auto-covarianza generalizzata* applicabile a tutti i segnali quasi-stazionari, ovvero sia a segnali deterministici che a segnali stocastici stazionari nonché a loro combinazioni lineari.

La funzione di auto-covarianza generalizzata consente anche di generalizzare le definizioni di spettro e densità spettrale a segnali quasi-stazionari, come segue

$$\Phi_Y(z) = \mathcal{Z} \{ \bar{R}_Y(k) \} = \sum_{k=-\infty}^{+\infty} \bar{R}_Y(k) z^{-k} \quad (9.3.8)$$

$$\varphi_Y(z) = \mathcal{F} \{ \bar{R}_Y(k) \} = \Phi(e^{j\omega}) = \sum_{k=-\infty}^{+\infty} \bar{R}_Y(k) e^{-j\omega k}. \quad (9.3.9)$$

È facile verificare (si lascia tale verifica, per esercizio, al lettore) che la generalizzazione (9.3.7)-(9.3.9) di funzione di auto-covarianza, spettro, densità spettrale permette di estendere il teorema di analisi del capitolo 4 a segnali di ingresso quasi-stazionari. Si può anche dimostrare che il *correlogramma*

$$\hat{R}_Y^N(k) = \frac{1}{N} \sum_{i=k}^N y_i y_{i-k}^T$$

ed il *periodogramma*

$$\hat{\varphi}_Y^N(\omega) = \frac{1}{N} Y_N(\omega) Y_N^T(-\omega) = \sum_{k=-(N-1)}^{N-1} \hat{R}_Y^N(k) e^{-j\omega k}, \quad Y_N(\omega) = \sum_{k=1}^N y_k e^{-j\omega k},$$

introdotti nel capitolo 4 come stimatori, asintoticamente (per $N \rightarrow \infty$) non polarizzati, della funzione di auto-covarianza e, rispettivamente, della densità spettrale di segnali stazionari preservano le loro proprietà anche per i segnali quasi-stazionari.

Negli sviluppi successivi si useranno più semplicemente $R_U(\cdot)$, $R_Y(\cdot)$, $R_{YU}(\cdot)$ anziché $\bar{R}_U(\cdot)$, $\bar{R}_Y(\cdot)$, $\bar{R}_{YU}(\cdot)$ a indicare le funzioni di covarianza relative ai segnali quasi-stazionari $\{u_k\}$ ed $\{y_k\}$, definite mediante l'operatore di media generalizzata $\bar{E}[\dots]$.

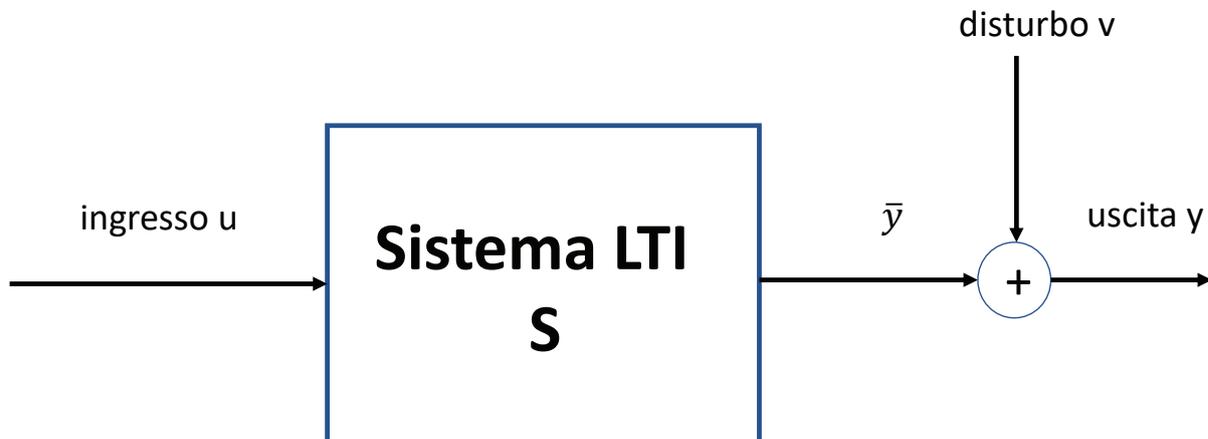


Figura 9.4: Modello LTI con disturbo stazionario

9.4 Modelli lineari tempo-invarianti

In questo corso, l'attenzione sarà rivolta esclusivamente all'identificazione di modelli LTI con disturbo stazionario come schematizzato in fig. 9.4, dove l'uscita y_k del sistema è modellata come somma di due segnali:

- un segnale deterministico \bar{y}_k generato come uscita di un sistema LTI S con ingresso deterministico u_k ;
- un disturbo stocastico stazionario v_k .

Si noti che, per quanto visto nel precedente paragrafo, il segnale y_k , risultante dalla somma di una componente deterministica \bar{y}_k e di una componente stocastica stazionaria v_k , è quasi-stazionario.

Indicando con $G(z)$ la funzione di trasferimento del sistema S , si farà l'ipotesi che tale funzione sia strettamente propria (con ritardo ingresso-uscita strettamente positivo), i.e.

$$G(z) = \sum_{k=1}^{\infty} g_k z^{-k}.$$

Il modello di fig. 9.3 è dunque descritto dalla seguente rappresentazione ingresso-uscita

$$y_k = \bar{y}_k + v_k = G(z)u_k + v_k = \sum_{k=1}^{\infty} g_i u_{k-i} + v_k. \quad (9.4.1)$$

Sostanzialmente, questo modello è caratterizzato dalla funzione di trasferimento ingresso-uscita $G(z)$ e dalle statistiche del disturbo stazionario v_k . A tale proposito, sono possibili descrizioni *non parametriche* di tale modello, sia nel *dominio del tempo* che nel

dominio della frequenza, che non presumono alcuna restrizione sull'ordine del modello. Specificamente:

- nel *dominio del tempo*, si può rappresentare la funzione di trasferimento $G(z)$ tramite la *risposta impulsiva* $\{g_k\}_{k=0}^{\infty}$ e la statistica del secondo ordine del disturbo stazionario v_k tramite la *funzione di autocovarianza* $\{R_V(k)\}_{k=0}^{\infty}$;
- nel *dominio della frequenza*, si può rappresentare la funzione di trasferimento $G(z)$ tramite la *risposta in frequenza* $\{G(e^{j\omega})\}_{\omega \in [0, \pi]}$ e la statistica del secondo ordine del disturbo stazionario v_k tramite la *densità spettrale* $\{\varphi_V(\omega)\}_{\omega \in [0, \pi]}$.

Naturalmente, sono anche possibili rappresentazioni parametriche della funzione di trasferimento $G(z)$ nonché delle statistiche del disturbo. A tale scopo si osserva che, in virtù dei risultati ben noti sull'approssimazione razionale e sulla fattorizzazione dello spettro, il disturbo stazionario v_k può essere modellato, con arbitrario grado di accuratezza, come un processo ARMA

$$\begin{cases} v_k &= H(z)e_k \\ e_k &\sim f_E(\cdot), \quad e_k \perp e_i \quad \forall k \neq i \end{cases} \quad (9.4.2)$$

per cui il modello (9.4.1) assume la forma

$$\begin{cases} y_k &= G(z)u_k + H(z)e_k = \sum_{k=1}^{\infty} g_i u_{k-i} + e_k + \sum_{k=1}^{\infty} h_i e_{k-i} \\ e_k &\sim f_E(\cdot) \end{cases} \quad (9.4.3)$$

dove $\{e_k\}$ è una sequenza di variabili aleatorie indipendenti e identicamente distribuite con PDF $f_E(\cdot)$. Si noti che il modello (9.4.3) è completamente individuato dalla terna $\{G(z), H(z), f_E(\cdot)\}$ dove:

- $G(z)$ è una funzione di trasferimento strettamente propria;
- $H(z)$, con $h_0 = I$, è una funzione di trasferimento con zeri e poli in $|z| < 1$;
- $f_E(\cdot)$ è una PDF, ovvero una funzione non-negativa con integrale $\int f_E(e) de = 1$ unitario.

Chiaramente tale terna può essere parametrizzata in molti modi diversi, i.e. $\{G(z, \theta), H(z, \theta), f_E(\cdot, \theta)\}$, in funzione di un vettore di parametri $\theta \in \mathbb{R}^{n_\theta}$, $n_\theta < \infty$, finito-dimensionale. Di solito, si assume che e_k abbia distribuzione Gaussiana con media nulla e covarianza $\Sigma = \Sigma^T > 0$, i.e. $e_k = \text{swn}(0, \Sigma)$.

9.5 Identificazione non parametrica nel dominio del tempo (analisi di correlazione)

In questo paragrafo, l'obiettivo è la determinazione, a partire dal dataset $z^N = \{u^N, y^N\}$, di un modello parametrico nel dominio del tempo costituito, con riferimento a fig. 9.4, dalla risposta impulsiva $\{g_k\}_{k=0}^\infty$ del sistema LTI e dalla funzione di auto-covarianza $\{R_V(k)\}_{k=0}^\infty$ del disturbo, i.e.

$$\text{dataset } z^N \longrightarrow \text{modello } (\{g_k\}_{k=0}^\infty, \{R_V(k)\}_{k=0}^\infty).$$

Si consideri il sistema (9.4.1) con relazione ingresso-uscita espressa mediante convoluzione fra la risposta impulsiva $\{g_k\}$ e l'ingresso $\{u_k\}$, i.e.,

$$y_k = \sum_{i=1}^{\infty} g_i u_{k-i} + v_k \quad (9.5.1)$$

e si assume che l'ingresso $\{u_k\}$ sia un segnale quasi-stazionario incorrelato con il disturbo $\{v_k\}$, i.e.,

$$\overline{E} [u_k v_j^T] = 0 \quad \forall k, j. \quad (9.5.2)$$

Si noti che l'ipotesi di incorrelazione (9.5.2), indispensabile per gli sviluppi che seguiranno, richiede che il sistema da identificare operi in anello aperto, ovvero che l'ingresso u_k non sia generato mediante retroazione dall'uscita y_k . Post-moltiplicando ambo i membri di (9.5.1) per u_{k-j}^T , $\forall j \geq 0$, ed applicando l'operatore di media generalizzato definito in (9.3.6), si ottengono le equazioni

$$\overline{E} [y_k u_{k-j}^T] = \sum_{i=1}^{\infty} g_i \overline{E} [u_{k-i} u_{k-j}^T] + \overline{E} [v_k u_{k-j}^T] \quad j = 0, 1, \dots \quad (9.5.3)$$

da cui, sfruttando (9.5.2), si ha

$$R_{YU}(j) = \sum_{i=1}^{\infty} g_i R_U(j-i) \quad \forall j \geq 0. \quad (9.5.4)$$

In modo analogo, post-moltiplicando ambo i membri di (9.5.1) per y_{k-j}^T , $\forall j \geq 0$, ed applicando l'operatore di media generalizzato definito in (9.3.6), si ottengono le equazioni

$$\overline{E} [y_k y_{k-j}^T] = \sum_{i=1}^{\infty} g_i \overline{E} [u_{k-i} y_{k-j}^T] + \overline{E} [v_k y_{k-j}^T] \quad j = 0, 1, \dots \quad (9.5.5)$$

da cui

$$R_Y(j) = \sum_{i=1}^{\infty} g_i R_{UY}(j-i) + R_{VY}(j) \quad \forall j \geq 0. \quad (9.5.6)$$

Notando che, per l'incorrelazione dei segnali $\{u_k\}$ e $\{v_k\}$, si ha

$$R_{VY}(j) = \overline{E} [v_k y_{k-j}^T] = \overline{E} \left[v_k \left(\sum_{i=1}^{\infty} g_i u_{k-j-i} + v_{k-j} \right)^T \right] = \overline{E} [v_k v_{k-j}^T] = R_V(j) \quad (9.5.7)$$

(9.5.7) diventa

$$R_V(j) = R_Y(j) - \sum_{i=1}^{\infty} g_i R_{UY}(j-i) \quad \forall j \geq 0. \quad (9.5.8)$$

Se le funzioni di covarianza $R_U(\cdot)$, $R_Y(\cdot)$, $R_{YU}(\cdot)$ fossero note, si potrebbero risolvere le equazioni lineari (9.5.4) rispetto alla risposta impulsiva $\{g_i\}$ e poi sostituire i valori trovati in (9.5.8) per ricavare la funzione di auto-covarianza del disturbo $\{R_V(j)\}$. A tale proposito, sussistono due difficoltà di seguito esposte.

1. Le funzioni di covarianza non sono note ma, a partire dal dataset $z^N = \{u^N, y^N\}$, se ne possono determinare stime non-polarizzate (per $N \rightarrow \infty$) mediante i correlogrammi

$$\begin{aligned} \widehat{R}_U^N(k) &= \frac{1}{N} \sum_{i=1}^{N-k} u_{i+k} u_i^T = \left[\widehat{R}_U^N(-k) \right]^T \\ \widehat{R}_Y^N(k) &= \frac{1}{N} \sum_{i=1}^{N-k} y_{i+k} y_i^T = \left[\widehat{R}_Y^N(-k) \right]^T \\ \widehat{R}_{UY}^N(k) &= \frac{1}{N} \sum_{i=1}^{N-k} u_{i+k} y_i^T = \left[\widehat{R}_{YU}^N(-k) \right]^T. \end{aligned} \quad (9.5.9)$$

2. (9.5.4) e (9.5.8) hanno infinite equazioni in infinite incognite che possono essere risolte mediante troncamento della risposta impulsiva, i.e. assumendo che

$$g_i = 0, \quad \forall i > M, \quad M \leq N < \infty. \quad (9.5.10)$$

Si noti che la necessità di effettuare l'esperimento di identificazione ad anello aperto per soddisfare (9.5.2) richiede la BIBO-stabilità del sistema da identificare, che equivale a $\sum_{k=1}^{\infty} \|g_k\| < \infty$ da cui $\lim_{k \rightarrow \infty} g_k = 0$. Pertanto l'ipotesi (9.5.10), che equivale ad approssimare il sistema (9.5.1) con un modello FIR (a risposta impulsiva finita) di ordine M sufficientemente elevato, risulta del tutto ragionevole.

Sostituendo in (9.5.4) e (9.5.8) le funzioni di covarianza con i corrispettivi correlogrammi ed utilizzando (9.5.10), si ottengono le seguenti equazioni lineari (per $j = 1, 2, \dots, M$):

$$\widehat{R}_{YU}^N(j) = \sum_{i=1}^M \widehat{g}_i^N \widehat{R}_U^N(j-i) \quad (9.5.11)$$

$$\widehat{R}_V^N(j) = \widehat{R}_Y^N(j) - \sum_{i=1}^M \widehat{g}_i^N \widehat{R}_{UY}^N(j-i) \quad (9.5.12)$$

da risolvere rispetto alle stime \widehat{g}_i^N e $\widehat{R}_V^N(j)$ di g_i e, rispettivamente, $R_V(j)$. Si noti che (9.5.11) può essere riscritta in forma matriciale come

$$\underbrace{\left[\widehat{R}_{YU}^N(1), \widehat{R}_{YU}^N(2), \dots, \widehat{R}_{YU}^N(M) \right]}_{\widehat{R}_{YU}^N} = \underbrace{\left[\widehat{g}_1^N, \widehat{g}_2^N, \dots, \widehat{g}_M^N \right]}_{\widehat{g}^N} \underbrace{\begin{bmatrix} \widehat{R}_U^N(0) & \widehat{R}_U^N(1) & \dots & \widehat{R}_U^N(M-1) \\ \widehat{R}_U^N(1)^T & \widehat{R}_U^N(0) & & \widehat{R}_U^N(M-2) \\ \vdots & & \ddots & \vdots \\ \widehat{R}_U^N(M-1)^T & \dots & \dots & \widehat{R}_U^N(0) \end{bmatrix}}_{\widehat{R}_U^N}$$

dove la matrice \widehat{R}_U^N (quadrata di dimensione $Mm \times Mm$ con $m = \dim u$) è, per costruzione, invertibile. Pertanto, la stima della risposta impulsiva è fornita da

$$\widehat{g}^N = [\widehat{g}_1^N, \widehat{g}_2^N, \dots, \widehat{g}_M^N] = \widehat{R}_{YU}^N \left(\widehat{R}_U^N \right)^{-1}. \quad (9.5.13)$$

Successivamente, da (9.5.12) si ricava la stima della funzione di auto-covarianza del disturbo

$$\widehat{R}_V^N(j) = \widehat{R}_Y^N(j) - \widehat{g}^N \begin{bmatrix} \widehat{R}_{UY}^N(j-1) \\ \widehat{R}_{UY}^N(j-2) \\ \vdots \\ \widehat{R}_{UY}^N(j-M) \end{bmatrix}. \quad (9.5.14)$$

Questo metodo di identificazione non parametrica nel dominio del tempo, che prende il nome di *analisi di correlazione* (in inglese: CRA = *CoRrelation Analysis*), si semplifica notevolmente se l'ingresso u_k è scelto come rumore bianco, i.e.,

$$R_U(k) = \overline{E} [u_{i+k} u_i^T] = \Sigma_U \delta_k = \begin{cases} \Sigma_U, & k = 0 \\ 0, & k \neq 0 \end{cases}. \quad (9.5.15)$$

In tal caso, infatti, si ha

$$g_i = R_{YU}(i) R_U^{-1}(0) = R_{YU}(i) \Sigma_U^{-1} \implies \widehat{g}_i^N = \widehat{R}_{YU}^N(i) \left[\widehat{R}_U^N(0) \right]^{-1} \cong \widehat{R}_{YU}^N(i) \Sigma_U^{-1}. \quad (9.5.16)$$

Nell'impossibilità di effettuare l'esperimento di identificazione con ingresso $\{u_k\}$ perfettamente bianco, si può determinare il correlogramma $\widehat{R}_U^N(\cdot)$ dell'ingresso e procedere alla stima della risposta impulsiva risolvendo le equazioni (9.5.11), i.e. tramite (9.5.13).

In alternativa, si possono pre-filtrare ingresso $\{u_k\}$ e uscita $\{y_k\}$ con lo stesso filtro digitale $F(z)$ in modo da rendere l'ingresso filtrato $\check{u}_k = F(z)u_k$ più possibile *bianco*. A

tale scopo, si può determinare una stima $\hat{\varphi}_U^N(\cdot)$ della densità spettrale di $\{u_k\}$ mediante periodogramma (o altro strumento di stima spettrale, vedi il paragrafo successivo). Successivamente si può procedere ad una approssimazione razionale $\Phi_U(z) \cong \hat{\Phi}_U^N(z)$ dello spettro corrispondente a tale stima spettrale, seguita dalla fattorizzazione spettrale

$$\Phi_U(z) = H(z) \Sigma_U H^T(z^{-1}),$$

da cui il filtro sbiancante desiderato per il segnale $\{u_k\}$ risulta, in buona approssimazione, dato da

$$F(z) = H^{-1}(z).$$

A questo punto, si può effettuare l'analisi di correlazione nel seguente modo.

1. Si pre-filtrano ingressi ed uscite per ottenere $\check{u}_k = F(z)u_k$ e $\check{y}_k = F(z)y_k$.
2. Si determinano i correlogrammi $\{\hat{R}_{\check{Y}\check{U}}^N(i)\}_{i=1}^M$, $\hat{R}_{\check{U}}^N(0)$, $\{\hat{R}_{\check{Y}}^N(i)\}_{i=1}^M$ e $\{\hat{R}_{\check{U}Y}^N(i)\}_{i=-(M-1)}^{M-1}$.
3. Si determina la stima della risposta impulsiva

$$\hat{g}_i^N = \hat{R}_{\check{Y}\check{U}}^N(i) \left[\hat{R}_{\check{U}}^N(0) \right]^{-1} \quad i = 1, 2, \dots, M \quad (9.5.17)$$

4. Si determina la stima della funzione di auto-covarianza

$$\hat{R}_{\check{Y}}^N(i) = \hat{R}_{\check{Y}}^N(i) - \sum_{j=1}^M \hat{g}_j^N \hat{R}_{\check{U}Y}^N(i-j) \quad i = 1, 2, \dots, M$$

Il metodo di analisi di correlazione per l'identificazione non parametrica nel dominio del tempo sopra esposto si basa essenzialmente sull'uso dei correlogrammi che, come visto nel capitolo 4, forniscono stime asintoticamente non polarizzate delle rispettive funzioni di covarianza. Ci si attende dunque che le stime \hat{g}_i^N della risposta impulsiva e $\hat{R}_{\check{Y}}^N(i)$ della funzione di auto-covarianza del disturbo ottenute con questo metodo siano anch'esse asintoticamente, i.e. per consistenza del dataset $N \rightarrow \infty$, non polarizzate. Infatti, facendo riferimento a (9.5.16) ovvero al caso di ingresso u_k bianco, l'errore di stima della risposta impulsiva risulta

$$\begin{aligned} g_i - \hat{g}_i^N &= g_i - \hat{R}_{\check{Y}\check{U}}^{-1}(i) \left[\hat{R}_{\check{U}}^N(0) \right]^{-1} = \left[g_i \hat{R}_{\check{U}}^N(0) - \hat{R}_{\check{Y}\check{U}}^N(i) \right] \left[\hat{R}_{\check{U}}^N(0) \right]^{-1} \\ &\cong \left[g_i \hat{R}_{\check{U}}^N(0) - \hat{R}_{\check{Y}\check{U}}^N(i) \right] R_{\check{U}}^{-1}(0). \end{aligned} \quad (9.5.18)$$

Si noti che

$$\begin{aligned} \lim_{N \rightarrow \infty} E \left[\hat{R}_{\check{Y}\check{U}}^N(i) \right] &= \lim_{N \rightarrow \infty} E \left[\frac{1}{N} \sum_{k=1}^{N-i} y_{k+i} u_k^T \right] = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^{N-i} E \left[y_{k+i} u_k^T \right] \\ &= \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{k=1}^N E \left[y_{k+i} u_k^T \right] \triangleq \bar{E} \left[y_{k+i} u_k^T \right] \triangleq R_{YU}(i). \end{aligned} \quad (9.5.19)$$

In modo del tutto analogo, si dimostra che

$$\lim_{N \rightarrow \infty} E \left[\widehat{R}_U^N(0) \right] = R_U(0). \quad (9.5.20)$$

Utilizzando (9.5.19) e (9.5.20) in (9.5.18), si ha

$$\lim_{N \rightarrow \infty} E \left[g_i - \widehat{g}_i^N \right] = [g_i R_U(0) - R_{YU}(i)] R_U^{-1}(0) = g_i - R_{YU}(i) R_U^{-1}(0) = g_i - g_i = 0. \quad (9.5.21)$$

Pertanto, la stima \widehat{g}_i^N di g_i è asintoticamente non polarizzata. Un'analisi più dettagliata (vedasi appendice A3.2 del capitolo 3 di [2]) rivela che l'errore di stima soddisfa quanto segue

$$E \left[g_i - \widehat{g}_i^N \right] = O \left(\frac{1}{N} \right) R_U^{-1}(0) \quad (9.5.22)$$

$$E \left[(g_i - \widehat{g}_i^N) (g_j - \widehat{g}_j^N)^T \right] = \frac{1}{N} \underbrace{[R_U(0) R_V^{-1}(i-j)]^{-1}}_{SNR} + O \left(\frac{1}{N} \right) \quad (9.5.23)$$

dove $O(1/N)$ indica una quantità proporzionale a $1/N$. Quindi, (9.5.22) evidenzia come l'errore di stima della risposta impulsiva converga, in media, a zero come $1/N$ per $N \rightarrow \infty$ e sia inversamente proporzionale alla varianza (energia) del segnale di ingresso. Inoltre, (9.5.23) afferma che anche la funzione di auto-covarianza di tale errore di stima tende a zero come $1/N$ per $N \rightarrow \infty$ e risulta inversamente proporzionale al *rapporto segnale-rumore* ($SNR = \text{Signal-to-Noise Ratio}$). Da (9.5.22)-(9.5.23) emerge, dunque, la convenienza di effettuare l'esperimento di identificazione con un ingresso u_k di varianza (energia) $R_U(0)$ massima compatibilmente con le limitazioni degli attuatori del sistema.

9.6 Identificazione non parametrica nel dominio della frequenza (analisi spettrale)

In questo paragrafo, l'obiettivo è la determinazione, a partire dal dataset $z^N = \{u^N, y^N\}$, di un modello parametrico nel dominio della frequenza costituito, con riferimento a fig. 9.4, dalla risposta in frequenza $\{G(e^{j\omega})\}_{\omega \in [0, \pi]}$ del sistema LTI e dalla densità spettrale $\{\varphi_V(\omega)\}_{\omega \in [0, \pi]}$ del disturbo, i.e.

$$\text{dataset } z^N \longrightarrow \text{modello } (\{G(e^{j\omega})\}_{\omega \in [0, \pi]}, \{\varphi_V(\omega)\}_{\omega \in [0, \pi]}) .$$

Applicando il teorema di analisi al sistema (9.4.1) ed ipotizzando l'incorrelazione (9.5.2) fra u_k e v_k , si ottengono le seguenti relazioni nel dominio della frequenza:

$$\varphi_{YU}(\omega) = G(e^{j\omega}) \varphi_U(\omega) \quad (9.6.1)$$

$$\varphi_Y(\omega) = G(e^{j\omega}) \varphi_U(\omega) G^T(e^{-j\omega}) + \varphi_V(\omega) \quad (9.6.2)$$

che coinvolgono le densità spettrali $\varphi_U(\cdot)$, $\varphi_{YU}(\cdot)$, $\varphi_Y(\cdot)$ e la risposta in frequenza $G(e^{j\omega})$ del sistema. Se, dunque, tali densità spettrali fossero note, si potrebbero ricavare la risposta in frequenza del sistema e la densità spettrale del disturbo tramite

$$G(e^{j\omega}) = \varphi_{YU}(\omega) \varphi_U(\omega)^{-1} \quad (9.6.3)$$

$$\varphi_V(\omega) = \varphi_Y(\omega) - G(e^{j\omega}) \varphi_U(\omega) G^T(e^{-j\omega}). \quad (9.6.4)$$

In particolare, se il sistema di interesse è SISO, le relazioni (9.6.3)-(9.6.4) si riducono più semplicemente a

$$G(e^{j\omega}) = \frac{\varphi_{YU}(\omega)}{\varphi_U(\omega)} \quad (9.6.5)$$

$$\varphi_V(\omega) = \varphi_Y(\omega) - |G(e^{j\omega})|^2 \varphi_U(\omega). \quad (9.6.6)$$

Stima empirica della risposta in frequenza

Nella pratica, le densità spettrali non sono note ma se ne possono determinare, dal dataset disponibile $z^N = \{u^N, y^N\}$, delle stime asintoticamente (i.e., per $N \rightarrow \infty$) non-polarizzate mediante i *periodogrammi*

$$\begin{aligned} \hat{\varphi}_U^N(\omega) &= \frac{1}{N} U_N(\omega) U_N^H(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{R}_U^N(k) e^{-j\omega k} \\ \hat{\varphi}_{YU}^N(\omega) &= \frac{1}{N} Y_N(\omega) U_N^H(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{R}_{YU}^N(k) e^{-j\omega k} \\ \hat{\varphi}_Y^N(\omega) &= \frac{1}{N} Y_N(\omega) Y_N^H(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{R}_Y^N(k) e^{-j\omega k} \end{aligned} \quad (9.6.7)$$

dove: $U_N^H(\omega) \triangleq \overline{U_N^T(\omega)}$; $U_N(\cdot)$ e $Y_N(\cdot)$ sono le *trasformate di Fourier discrete* di $\{u_k\}$ e, rispettivamente, $\{y_k\}$ definite come

$$\begin{aligned} U_N(\omega) &= \sum_{k=1}^N u_k e^{-j\omega k} \\ Y_N(\omega) &= \sum_{k=1}^N y_k e^{-j\omega k}; \end{aligned} \quad (9.6.8)$$

mentre $\hat{R}_U^N(\cdot)$, $\hat{R}_{YU}^N(\cdot)$, $\hat{R}_Y^N(\cdot)$ sono i *correlogrammi* definiti in (9.5.9). Sostituendo le densità spettrali in (9.6.3)-(9.6.4) con i rispettivi periodogrammi in (9.6.7), si ottengono le

seguenti stime spettrali

$$\widehat{G}_N(e^{j\omega}) = \widehat{\varphi}_{YU}^N(\omega) [\widehat{\varphi}_U^N(\omega)]^{-1} = Y_N(\omega) U_N^H(\omega) [U_N(\omega) U_N^H(\omega)]^{-1} \quad (9.6.9)$$

$$\begin{aligned} \widehat{\varphi}_V^N(\omega) &= \widehat{\varphi}_Y^N(\omega) - \widehat{G}_N(e^{j\omega}) \widehat{\varphi}_U^N(\omega) \widehat{G}_N^H(e^{j\omega}) \\ &= \widehat{\varphi}_Y^N(\omega) - \widehat{\varphi}_{YU}^N(\omega) [\widehat{\varphi}_U^N(\omega)]^{-1} \widehat{\varphi}_{UY}^N(\omega) \\ &= \frac{1}{N} \left\{ Y_N(\omega) Y_N^H(\omega) - Y_N(\omega) U_N^H(\omega) [U_N(\omega) U_N^H(\omega)]^{-1} U_N(\omega) Y_N^H(\omega) \right\}. \end{aligned} \quad (9.6.10)$$

In particolare, $\widehat{G}_N(e^{j\omega})$ in (9.6.9) prende il nome di *stima empirica della funzione id trasferimento* (ETFE = *Empirical Transfer Function Estimate*), anche se dovrebbe essere più propriamente detta *stima empirica della risposta in frequenza* $G(e^{j\omega})$. L'attributo "empirica" deriva dal fatto che, nel caso SISO (u e y scalari), tale stima si riduce a

$$\widehat{G}_N(e^{j\omega}) = \frac{\widehat{\varphi}_{YU}^N(\omega)}{\widehat{\varphi}_U^N(\omega)} = \frac{Y_N(\omega)}{U_N(\omega)} \quad (9.6.11)$$

che fornisce la stima della risposta in frequenza come rapporto tra le trasformate di Fourier dell'uscita (a numeratore) e ingresso (a denominatore), cioè come semplicistica e immediata applicazione delle tecniche di Fourier al sistema (9.4.1). Sempre nel caso SISO, (9.6.10) fornisce

$$\widehat{\varphi}_V^N(\omega) = \frac{1}{N} \left[Y_N(\omega) \overline{Y_N(\omega)} - \frac{Y_N(\omega)}{U_N(\omega)} U_N(\omega) \overline{U_N(\omega)} \frac{\overline{Y_N(\omega)}}{U_N(\omega)} \right] = 0$$

che, grossolanamente, approssima a zero la densità spettrale del disturbo. Si dimostra [1] che la stima ETFE soddisfa, nel caso SISO, le seguenti proprietà,

i) Per quanto riguarda l'errore di stima ETFE, si ha

$$E \left[\widehat{G}_N(e^{j\omega}) \right] - G(e^{j\omega}) = \frac{\varrho_1(N)}{U_N(\omega)} \text{ con } |\varrho_1(N)| \leq \frac{C_1}{\sqrt{N}} \text{ e } C_1 < \infty, \quad (9.6.12)$$

ovvero converge, in media, a zero come $1/\sqrt{N}$ per $N \rightarrow \infty$ ed è anche inversamente proporzionale a $U_N(\omega)$. In altri termini, la stima ETFE è tanto migliore quanto maggiore è la numerosità N dei dati e alle frequenze angolari ω per le quali l'ingresso ha componenti armoniche di densità spettrale, $\varphi_U(\omega) \cong |U_N(\omega)|^2/N$, più elevata.

ii) Per quanto riguarda la correlazione fra stime ETFE a frequenze angolari diverse, si ha

$$E \left[\left(\widehat{G}_N(e^{j\omega}) - G(e^{j\omega}) \right) \left(\widehat{G}_N(e^{-j\omega_0}) - G(e^{-j\omega_0}) \right) \right] \\ = \begin{cases} \frac{1}{|U_N(\omega)|^2} [\varphi_V(\omega) + \varrho_2(N)], & \omega = \omega_0 \\ \frac{\varrho_2(N)}{U_N(\omega)U_N(-\omega_0)}, & \omega \neq \omega_0 \end{cases} \quad (9.6.13)$$

con

$$|\varrho_2(N)| \leq \frac{C_2}{N}, \quad C_2 < \infty.$$

Da (9.6.12)-(9.6.13) si evince quanto segue.

1. La stima ETFE è asintoticamente non polarizzata.
2. La varianza è inversamente proporzionale al *rapporto segnale-rumore* ($\text{SNR} = \text{Signal-to-Noise-Ratio}$) $|U_N(\omega)|^2 \varphi_V^{-1}(\omega)$ e non decade a zero all'aumentare di N .
3. Stime a frequenze diverse risultano asintoticamente incorrelate.

In particolare, le ultime due costituiscono caratteristiche indesiderabili dello stimatore ETFE. Infatti, non è possibile migliorare la precisione della stima ETFE aumentando il numero di campioni N del dataset usato per l'identificazione, ma soltanto incrementando la potenza del segnale di ingresso nell'esperimento di identificazione; ovviamente tale potenza non può essere troppo elevata a causa di limitazioni degli attuatori del sistema nonché di considerazioni energetiche. Inoltre, l'incorrelazione di stime a frequenze diverse comunque vicine rende, di fatto, la risposta in frequenza stimata mediante ETFE estremamente irregolare al contrario della risposta in frequenza di effettivi sistemi reali che, viceversa, ha un andamento solitamente regolare. La stima ETFE $\widehat{G}_N(e^{j\omega})$ della risposta in frequenza risulta pertanto piuttosto grossolana e per questo motivo viene indicata con un doppio simbolo “*cappello*”.

Stima regolarizzata della risposta in frequenza

Un possibile rimedio per ovviare ai problemi (elevata varianza ed irregolarità) della stima ETFE $\widehat{G}_N(e^{j\omega})$ è di costruire da questa una stima *regolarizzata* $\widehat{G}_N(e^{j\omega_0})$ ad una certa pulsazione ω_0 effettuando una media delle stime *grezze* $\widehat{G}_N(e^{j\omega})$, $\omega \in [-\pi, \pi]$, opportunamente pesate in funzione della distanza di ω da ω_0 e del loro grado di accuratezza. La media in frequenza ha il duplice vantaggio di ridurre la varianza della stima regolarizzata $\widehat{G}_N(e^{j\omega_0})$ rispetto a quella della stime ETFE $\widehat{G}_N(e^{j\omega})$ nonché di filtrare/regolarizzare l'andamento della risposta in frequenza. L'approccio di regolarizzazione della stima ETFE

che segue si basa sul presupposto che la vera risposta in frequenza da stimare $G(e^{j\omega})$ sia una funzione regolare di ω .

Nel seguito, si consideri il caso di un sistema SISO (u e y scalari). La stima regolarizzata $\widehat{G}_N(e^{j\omega_0})$ è dunque ottenuta dalla stima ETFE $\widehat{\widehat{G}}_N(e^{j\omega})$ mediante la seguente media su tutto l'intervallo di pulsazioni $[-\pi, \pi]$:

$$\widehat{G}_N(e^{j\omega_0}) = \int_{-\pi}^{\pi} w(\omega_0, \omega) \widehat{\widehat{G}}_N(e^{j\omega}) d\omega \quad (9.6.14)$$

dove, per ogni $\omega_0 \in [-\pi, \pi]$, la funzione di peso $w(\omega_0, \omega) \geq 0$ deve soddisfare la condizione di normalizzazione

$$\int_{-\pi}^{\pi} w(\omega_0, \omega) d\omega = 1. \quad (9.6.15)$$

In accordo a quanto detto sopra, la funzione di peso può essere scelta della seguente forma

$$w(\omega_0, \omega) \propto W_\gamma(\omega_0 - \omega) \alpha(\omega) \quad (9.6.16)$$

dove:

- $W_\gamma(\omega_0 - \omega)$ è una funzione di peso, simmetrica rispetto a ω_0 , che penalizza la distanza di ω da ω_0 ;
- $\alpha(\omega)$ è una funzione di peso che tiene conto della precisione della stima ETFE alla pulsazione ω .

Tipicamente, $W_\gamma(\cdot)$ è una funzione *a campana* con picco in corrispondenza di ω_0 che decade simmetricamente a zero sulle code (i.e., per $\omega - \omega_0 \rightarrow \pm\infty$) con parametro $\gamma > 0$ che caratterizza l'ampiezza della finestra in frequenza. Alcune funzioni di peso $W_\gamma(\cdot)$ proposte in letteratura sono le seguenti

$$W_\gamma(\omega) = \begin{cases} \frac{1}{2\pi\gamma} \left(\frac{\sin \gamma\omega/2}{\sin \omega/2} \right)^2, & \text{finestra di Bartlett} \\ \frac{4(2 + \cos \omega)}{2\pi\gamma^3} \left(\frac{\sin \gamma\omega/4}{\sin \omega/2} \right)^4, & \text{finestra di Parzen} \\ \frac{1}{2}D_\gamma(\omega) + \frac{1}{4}D_\gamma(\omega - \pi/\gamma) + \frac{1}{4}D_\gamma(\omega + \pi/\gamma), & \text{finestra di Hamming} \end{cases} \quad (9.6.17)$$

dove

$$D_\gamma(\omega) = \frac{\sin(\gamma + \frac{1}{2})\omega}{2\pi \sin \omega/2}.$$

L'andamento in frequenza delle funzioni di peso (9.6.17) è riportato in fig. 9.5.

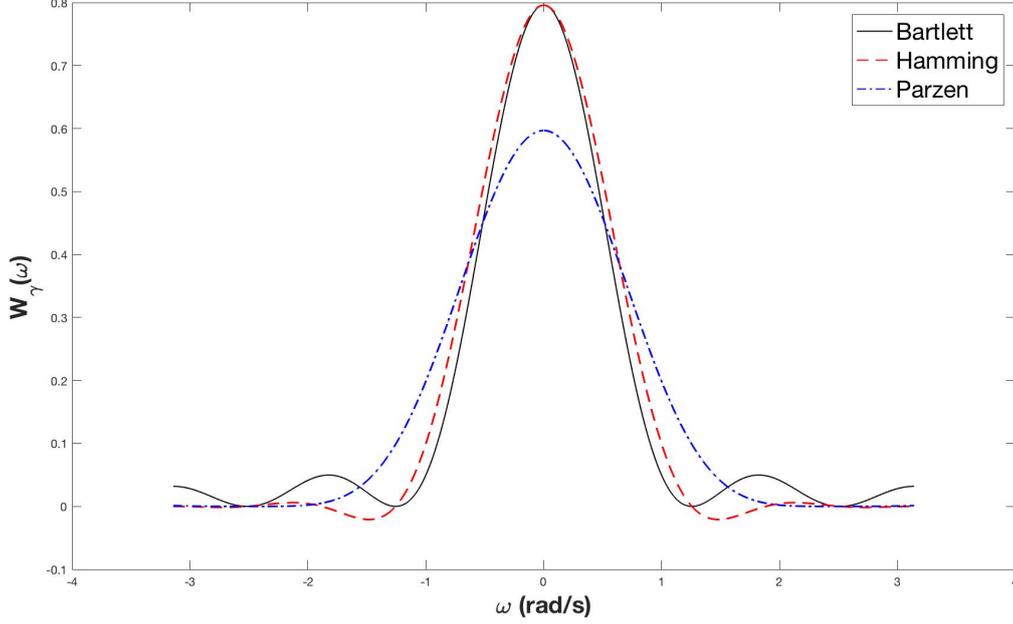


Figura 9.5: Funzioni di peso $W_\gamma(\omega)$ di alcune finestre in frequenza

Viceversa, in accordo all'analisi in (9.6.13), è naturale scegliere $\alpha(\omega)$ proporzionale al rapporto segnale-rumore, i.e.,

$$\alpha(\omega) = \frac{|U_N(\omega)|^2}{\varphi_V(\omega)} \quad (9.6.18)$$

in modo da pesare maggiormente, nella media pesata, le stime ETFE relative a frequenze per le quali il rapporto segnale-rumore è più elevato e, conseguentemente, la varianza dell'errore di stima è minore. Da (9.6.15)-(9.6.16) si ricava

$$w(\omega_0, \omega) = \frac{W_\gamma(\omega_0 - \omega)\alpha(\omega)}{\int_{-\pi}^{\pi} W_\gamma(\omega_0 - \xi)\alpha(\xi) d\xi}. \quad (9.6.19)$$

Sostituendo (9.6.18)-(9.6.19) in (9.6.14), si ottiene la seguente stima regolarizzata

$$\widehat{G}_N(e^{j\omega_0}) = \frac{\int_{-\pi}^{\pi} W_\gamma(\omega_0 - \omega) |U_N(\omega)|^2 \varphi_V^{-1}(\omega) \widehat{G}_N(e^{j\omega}) d\omega}{\int_{-\pi}^{\pi} W_\gamma(\omega_0 - \xi) |U_N(\xi)|^2 \varphi_V^{-1}(\xi) d\xi}. \quad (9.6.20)$$

Assumendo che $\varphi_V(\omega) \cong \varphi_V(\omega_0)$ per tutti gli ω tali che $W_\gamma(\omega_0 - \omega) \ll 1$, si può approssimare $\varphi_V(\omega)$ nei due integrali, a numeratore e denominatore, di (9.6.20) con $\varphi_V(\omega_0)$. Si

ottiene così la seguente approssimazione

$$\begin{aligned}
\widehat{G}_N(e^{j\omega_0}) &\cong \frac{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \omega) |U_N(\omega)|^2 \widehat{G}_N(e^{j\omega}) d\omega}{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \xi) |U_N(\xi)|^2 d\xi} \\
&= \frac{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \omega) U_N(\omega) \overline{U_N(\omega)} Y_N(\omega) U_N^{-1}(\omega) d\omega}{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \xi) U_N(\xi) \overline{U_N(\xi)} d\xi} \\
&= \frac{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \omega) Y_N(\omega) \overline{U_N(\omega)} d\omega}{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \xi) U_N(\xi) \overline{U_N(\xi)} d\xi} \\
&= \frac{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \omega) \widehat{\varphi}_{YU}^N(\omega) d\omega}{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \xi) \widehat{\varphi}_U^N(\xi) d\xi}.
\end{aligned} \tag{9.6.21}$$

Si noti come la stima ETFE (9.6.11) corrisponda ad una funzione di peso $W_{\gamma}(\cdot)$ impulsiva, i.e.

$$W_{\gamma}(\omega_0 - \omega) = \delta(\omega_0 - \omega) \Rightarrow \widehat{G}_N(e^{j\omega}) = \widehat{G}_N(e^{j\omega_0}).$$

Dai precedenti sviluppi, risulta dunque la seguente stima regolarizzata della risposta in frequenza:

$$\widehat{G}_N(e^{j\omega}) = \frac{\int_{-\pi}^{\pi} W_{\gamma}(\omega - \xi) \widehat{\varphi}_{YU}^N(\xi) d\xi}{\int_{-\pi}^{\pi} W_{\gamma}(\omega_0 - \xi) \widehat{\varphi}_U^N(\xi) d\xi} = \frac{W_{\gamma} * \widehat{\varphi}_{YU}^N(\omega)}{W_{\gamma} * \widehat{\varphi}_U^N(\omega)} \tag{9.6.22}$$

dove il denominatore è la convoluzione, in frequenza, fra la funzione di peso $W_{\gamma}(\cdot)$ ed il cross-periodogramma ingresso-uscita $\widehat{\varphi}_{YU}^N(\cdot)$, mentre il numeratore è la convoluzione fra la medesima funzione di peso ed il periodogramma dell'ingresso $\widehat{\varphi}_U^N(\cdot)$. Definendo l'antitrasformata di Fourier $w_{\gamma}(k)$ di $W_{\gamma}(\omega)$, i.e.

$$w_{\gamma}(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} W_{\gamma}(\omega) e^{j\omega k} d\omega, \tag{9.6.23}$$

ricordando che $\widehat{\varphi}_{YU}^N(\omega)$ e $\widehat{\varphi}_U^N(\omega)$ sono le trasformate di Fourier dei correlogrammi $\widehat{R}_{YU}^N(k)$ e $\widehat{R}_U^N(k)$ ed applicando il *teorema di convoluzione*

$$\mathcal{F}^{-1} [(W_{\gamma} * \widehat{\varphi}_{YU}^N)(\omega)] = w_{\gamma}(k) \widehat{R}_{YU}^N(k), \quad \mathcal{F}^{-1} [(W_{\gamma} * \widehat{\varphi}_U^N)(\omega)] = w_{\gamma}(k) \widehat{R}_U^N(k) \tag{9.6.24}$$

si può riscrivere la stima regolarizzata in (9.6.22) come

$$\widehat{G}_N(e^{j\omega}) = \frac{\sum_{k=-\infty}^{\infty} w_\gamma(k) \widehat{R}_{YU}^N(k) e^{-j\omega k}}{\sum_{k=-\infty}^{\infty} w_\gamma(k) \widehat{R}_U^N(k) e^{-j\omega k}}. \quad (9.6.25)$$

Di norma, si sceglie la *finestra in frequenza* $W_\gamma(\omega)$ in modo che la *finestra temporale* $w_\gamma(k)$ soddisfi

$$w_\gamma(k) = 0 \text{ per } |k| > \gamma \quad (9.6.26)$$

il che equivale ad avere una finestra temporale di ampiezza finita ($\gamma < \infty$). I coefficienti di Fourier $w_\gamma(k)$ corrispondenti alle funzioni di peso in frequenza $W_\gamma(\omega)$ di Bartlett, Parzen, Hamming in (9.6.17) sono riportati, per completezza, in tabella 9.1.

Finestra	$W_\gamma(\omega), \omega \in [-\pi, \pi]$	$w_\gamma(k), 0 \leq k \leq \gamma$
Bartlett	$\frac{1}{2\pi\gamma} \left(\frac{\sin \gamma\omega/2}{\sin \omega/2} \right)^2$	$1 - \frac{ k }{\gamma}$
Parzen	$\frac{4(2 + \cos \omega)}{2\pi\gamma^3} \left(\frac{\sin \gamma\omega/4}{\sin \omega/2} \right)^4$	$\begin{cases} 1 - \frac{6k^2}{\gamma^2} \left(1 - \frac{ k }{\gamma} \right), & 0 \leq k \leq \frac{\gamma}{2} \\ 2 \left(1 - \frac{ k }{\gamma} \right)^3, & \frac{\gamma}{2} \leq k \leq \gamma \end{cases}$
Hamming	$\frac{1}{2}D_\gamma(\omega) + \frac{1}{4}D_\gamma(\omega - \pi/\gamma) + \frac{1}{4}D_\gamma(\omega + \pi/\gamma)$ $D_\gamma(\omega) = \frac{\sin(\gamma + \frac{1}{2})\omega}{2\pi \sin \omega/2}$	$\frac{1}{2} \left(1 + \cos \frac{\pi k}{\gamma} \right)$

Tabella 9.1: Finestre di Bartlett, Parzen e Hamming.

Si noti che a valori piccoli di γ corrisponde una finestra temporale piccola e, di conseguenza, una finestra in frequenza grande (media della stima ETFE su un ampio intorno della frequenza di interesse che corrisponde ad un elevato livello di regolarizzazione); viceversa $\gamma \rightarrow \infty$ tende alla stima ETFE (nessuna regolarizzazione). La scelta di γ è dunque un compromesso fra polarizzazione (che tende a zero per $\gamma \rightarrow \infty$) e varianza che si riduce al diminuire di γ , $0 \leq \gamma \leq N$.

Riassumendo i precedenti sviluppi, la stima regolarizzata, con parametro di regolarizzazione γ , è fornita da

$$\widehat{G}_N(e^{j\omega}) = \frac{\sum_{k=-\gamma}^{\gamma} w_\gamma(k) \widehat{R}_{YU}^N(k) e^{-j\omega k}}{\sum_{k=-\gamma}^{\gamma} w_\gamma(k) \widehat{R}_U^N(k) e^{-j\omega k}} \quad (9.6.27)$$

che può essere ottenuta con il seguente procedimento.

1. Scelto il tipo di finestra in frequenza (Bartlett, Parzen, Hamming, etc.) e fissata l'ampiezza γ si determinano i coefficienti di Fourier $w_\gamma(k)$, $-\gamma \leq k \leq \gamma$, e.g. dalla tabella 9.1.
2. Dal *dataset* $z^N = \{u^N, y^N\}$ si determinano i correlogrammi

$$\widehat{R}_{YU}^N(k) = \frac{1}{N} \sum_{i=1}^{N-K} y_{i+k} u_i, \quad \widehat{R}_U^N(k) = \frac{1}{N} \sum_{i=1}^{N-K} u_{i+k} u_i$$

per $k = 0, 1, \dots, \gamma$.

3. Si determina la stima regolarizzata

$$\widehat{G}_N(e^{j\omega_i}) = \frac{\sum_{k=-\gamma}^{\gamma} w_\gamma(k) \widehat{R}_{YU}^N(k) e^{-j\omega_i k}}{\sum_{k=-\gamma}^{\gamma} w_\gamma(k) \widehat{R}_U^N(k) e^{-j\omega_i k}} \quad (9.6.28)$$

in corrispondenza delle pulsazioni $\omega_i = \frac{2\pi}{N}i$ per $i = 1, 2, \dots, N$. □

Esempio (Stima regolarizzata della risposta in frequenza) - A titolo esemplificativo si consideri il sistema

$$y_k = \underbrace{\frac{1 + 0.7z^{-1}}{1 - 1.4z^{-1} + 0.9z^{-2}}}_{G(z)} u_k + \underbrace{\frac{1}{1 - 0.9z^{-1}}}_{H(z)} e_k \quad (9.6.29)$$

$u_k = \text{swn}(0, 1), \quad e_k = \text{swn}(0, 9/100), \quad u \perp e.$

Nelle figg. 9.6-9.8 sono riportati i diagrammi di Bode delle stime $\widehat{G}_N(e^{j\omega})$ ottenute con un dataset di ampiezza $N = 1024$ e $\gamma = 10, 50, 100$, mentre Fig. 9.9 si riferisce alla stima ETFE $\widehat{\widehat{G}}_N(e^{j\omega})$ ($\gamma = \infty$). Si noti che all'aumentare di γ si riduce la polarizzazione ma aumenta la varianza/irregolarità della stima. □

Proprietà asintotiche della stima regolarizzata

Le proprietà asintotiche della stima (9.6.28) dipendono in modo cruciale dalla scelta del tipo di finestra in frequenza $W_\gamma(\cdot)$ e del parametro γ . A tale proposito, risulta conveniente

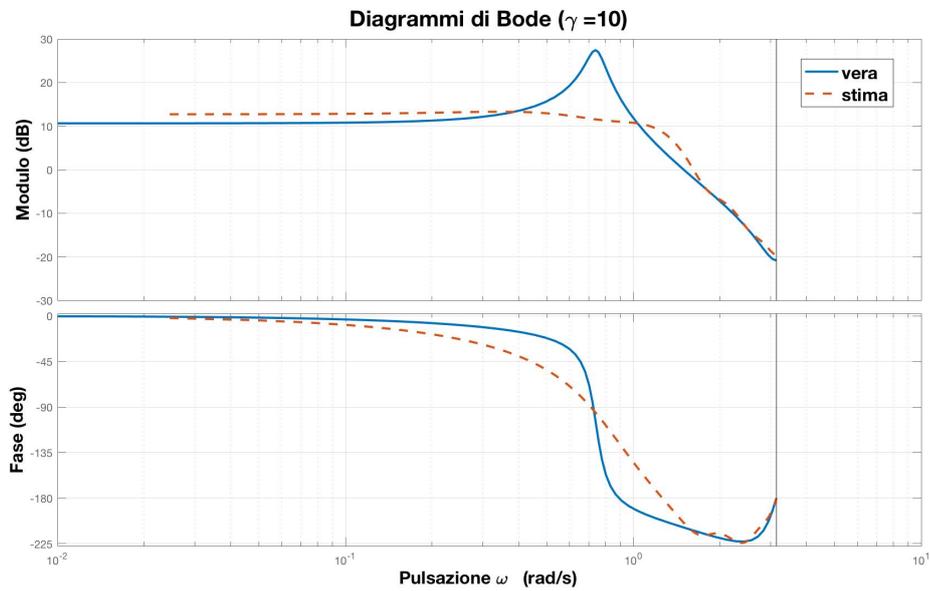


Figura 9.6: Stima della risposta in frequenza con $N = 1024$ e $\gamma = 10$

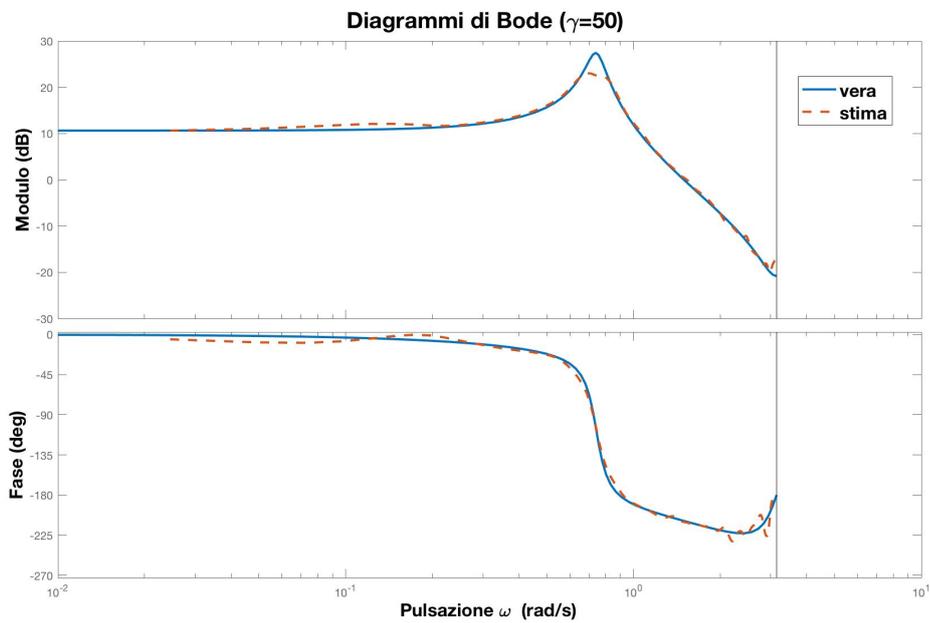


Figura 9.7: Stima della risposta in frequenza con $N = 1024$ e $\gamma = 50$

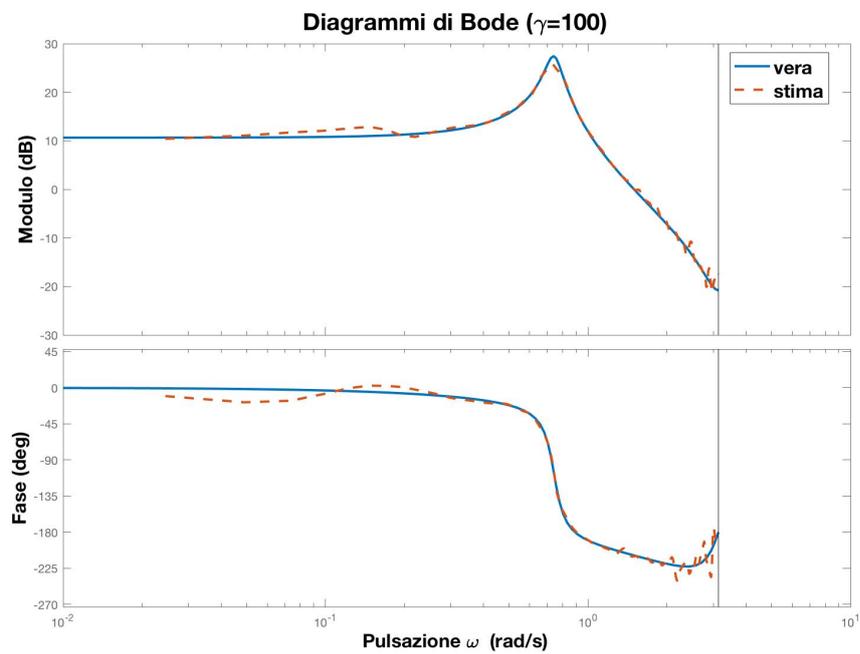


Figura 9.8: Stima della risposta in frequenza con $N = 1024$ e $\gamma = 100$

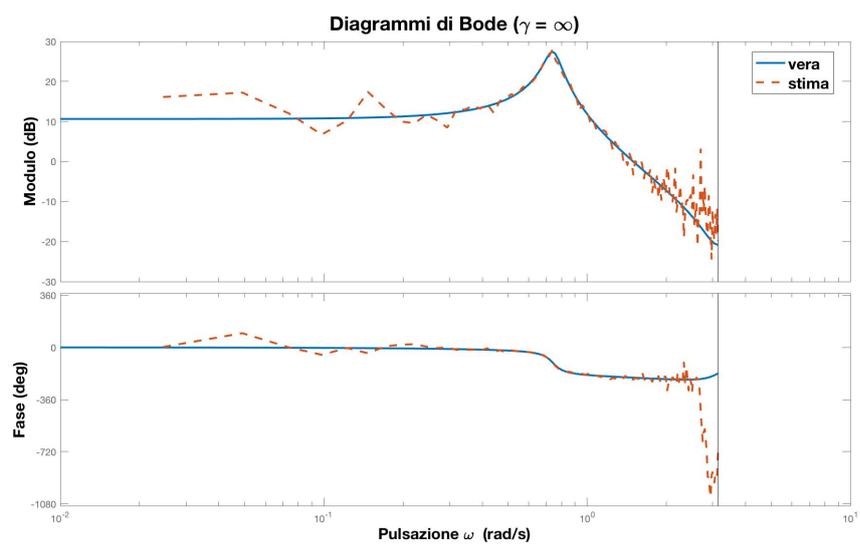


Figura 9.9: Stima ETFE della risposta in frequenza con $N = 1024$

introdurre le seguenti quantità

$$M(\gamma) = \int_{-\pi}^{\pi} \omega^2 W_{\gamma}(\omega) d\omega \quad (9.6.30)$$

$$\overline{W}(\gamma) = 2\pi \int_{-\pi}^{\pi} W_{\gamma}^2(\omega) d\omega \quad (9.6.31)$$

che sono tabulate, per i tipi di finestra considerati, in tabella 9.2.

Finestra	$M(\gamma)$	$\overline{W}(\gamma)$
Bartlett	$2.78/\gamma$	$\cong 0.67\gamma$
Parzen	$12/\gamma^2$	$\cong 0.54\gamma$
Hamming	π^2/γ^2	$\cong 0.75\gamma$

Tabella 9.2: Quantità $M(\gamma)$ e $\overline{W}(\gamma)$ relativi alle finestre di Bartlett, Parzen e Hamming.

In [1] viene dimostrato il seguente risultato, qui riportato senza dimostrazione.

Teorema (Polarizzazione, varianza ed errore quadratico medio della stima spettrale) - La stima $\widehat{G}_N(e^{j\omega})$ della risposta in frequenza fornita da (9.6.28) soddisfa le seguenti proprietà per $N \rightarrow \infty$, $\gamma \rightarrow \infty$ e $\gamma/N \rightarrow 0$:

$$\text{polarizzazione } E \left[\widehat{G}_N(e^{j\omega}) \right] - G(e^{j\omega}) = O(M(\gamma)) \quad (9.6.32)$$

$$\text{varianza } E \left[\left| \widehat{G}_N(e^{j\omega}) - E \left[\widehat{G}_N(e^{j\omega}) \right] \right|^2 \right] = O \left(\frac{\overline{W}(\gamma)}{N} \frac{\varphi_V(\omega)}{\varphi_U(\omega)} \right) \quad (9.6.33)$$

$$\text{MSE } E \left[\left| \widehat{G}_N(e^{j\omega}) - G(e^{j\omega}) \right|^2 \right] = M^2(\gamma) |R(\omega)|^2 + \frac{\overline{W}(\gamma)}{N} \frac{\varphi_V(\omega)}{\varphi_U(\omega)} \quad (9.6.34)$$

dove

$$R(\omega) = \frac{1}{2} G''(e^{j\omega}) + G'(e^{j\omega}) \frac{\varphi'_U(\omega)}{\varphi_U(\omega)} \quad (9.6.35)$$

$$G'(e^{j\omega}) = \frac{d}{d\omega} [G(e^{j\omega})], \quad G''(e^{j\omega}) = \frac{d^2}{d\omega^2} [G(e^{j\omega})], \quad \varphi'_U(\omega) = \frac{d}{d\omega} [\varphi_U(\omega)]. \quad \square$$

In altri termini, la stima $\widehat{G}_N(e^{j\omega})$ è affetta da una polarizzazione dell'ordine di $M(\gamma)$ che comunque converge a zero, come $1/\gamma$ oppure come $1/\gamma^2$ a seconda del tipo di finestra (vedi tabella 9.2), per $\gamma \rightarrow \infty$. Viceversa, tale stima presenta una varianza proporzionale a $\overline{W}(\gamma)/N \propto \gamma/N$ e inversamente proporzionale al rapporto segnale-rumore $\varphi_U(\omega)/\varphi_V(\omega)$; in particolare, tale varianza si riduce al diminuire di γ/N . Risulta evidente dal precedente teorema come il parametro γ (ampiezza della finestra temporale) debba essere scelto sulla

base di un adeguato compromesso fra polarizzazione (che si riduce all'aumentare di γ) e varianza (che, viceversa, si riduce al diminuire di γ). La scelta di γ dovrebbe dunque essere tale da rendere piccoli sia $M(\gamma)$ che $\overline{W}(\gamma)$, obiettivi che risultano conflittuali.

Si noti che la relazione (9.6.34) fornisce l'errore quadratico medio (MSE) di stima in funzione del parametro γ . Posto $M(\gamma) = M/\gamma^2$ (ipotesi valida per finestre di Parzen e Hamming) e $\overline{W}(\gamma) = \overline{W}\gamma$, si può cercare di minimizzare l'MSE in funzione di γ . A tale proposito, derivando (9.6.34) rispetto a γ , si ha

$$\frac{d}{d\gamma} \left[\frac{M^2}{\gamma^4} + \frac{\overline{W}\gamma \varphi_V(\omega)}{N \varphi_U(\omega)} \right] = -\frac{4M^2}{\gamma^5} |R(\omega)|^2 + \frac{\overline{W} \varphi_V(\omega)}{N \varphi_U(\omega)} = 0$$

da cui si ricava la scelta ottima di γ :

$$\gamma^* = \left[\frac{4M^2 |R(\omega)|^2 \varphi_U(\omega)}{\overline{W} \varphi_V(\omega)} \right]^{1/5} N^{1/5}. \quad (9.6.36)$$

Si noti come tale scelta ottima risulti irrealizzabile in quanto (9.6.36) coinvolge diverse quantità incognite (i.e. $R(\omega)$, $\varphi_U(\omega)$, $\varphi_V(\omega)$) e inoltre è dipendente dalla pulsazione ω . Si osserva, tuttavia, che tale scelta cresce come $N^{1/5}$ in funzione del numero di dati disponibili, ovvero la risposta in frequenza dovrebbe diventare più stretta al crescere del numero di dati. Inoltre, sostituendo (9.6.36) in (9.6.34) ne consegue che la scelta ottima di γ fa sì che l'MSE decada a zero come $N^{-4/5}$.

Da un punto di vista pratico, conviene considerare valori di N/γ da 5 a 100. All'aumentare di γ , i.e. al diminuire di N/γ , l'andamento della risposta in frequenza stimata $\widehat{G}_N(e^{j\omega})$ acquisisce maggiori dettagli ma anche maggiore varianza e irregolarità (picchi spuri). Conviene dunque esaminare i grafici della risposta in frequenza stimata al variare di γ , partendo da $N/\gamma = 100$ e aumentando γ fino a che non si ritiene che i dettagli emergenti siano dovuti a variazioni spurie.

Stima spettrale di Welch

È possibile regolarizzare la stima ETFE con altri tipi di medie, in alternativa alla media frequenziale precedentemente considerata. Ad esempio, si può suddividere il *dataset* originario $z^N = \{u^N, y^N\}$ di ampiezza $N = LM$ in L *sub-dataset* $z^{M,\ell} = \{u^{M,\ell}, y^{M,\ell}\}$ ($\ell = 1, 2, \dots, L$) di ampiezza M , i.e. $z^{M,\ell} = \{z_k\}_{k=(\ell-1)M+1}^{\ell M}$. Per ogni *sub-dataset* $z^{M,\ell}$ si può quindi determinare la stima ETFE $\widehat{G}_{M,\ell}(e^{j\omega})$ e successivamente effettuare su tali stime una media pesata

$$\widehat{G}_N(e^{j\omega}) = \frac{\sum_{\ell=1}^L \beta_{M,\ell}(\omega) \widehat{G}_{M,\ell}(e^{j\omega})}{\sum_{\ell=1}^L \beta_{M,\ell}(\omega)} \quad (9.6.37)$$

dove i pesi possono essere scelti in modo uniforme, i.e

$$\beta_{M,\ell}(\omega) = \frac{1}{M} \quad \ell = 1, 2, \dots, L \quad (9.6.38)$$

oppure in modo direttamente proporzionale alla densità spettrale dell'ingresso, i.e.

$$\beta_{M,\ell}(\omega) = \frac{1}{M} |U_{M,\ell}(\omega)|^2 \quad (9.6.39)$$

dove $U_{M,\ell}(\omega)$ è la DFT del *sub-dataset* di ingresso $u^{M,\ell} = \{u_k\}_{k=(\ell-1)M+1}^{\ell M}$. Questo metodo di stima spettrale è noto in letteratura come *metodo di Welch*.

Stima della densità spettrale del disturbo

Per quanto riguarda la stima della densità spettrale del disturbo, particolarizzando (9.6.4) al caso di sistema SISO ed usando le stime spettrali disponibili, si ha

$$\widehat{\varphi}_V^N(\omega) = \widehat{\varphi}_Y^N(\omega) - |\widehat{G}_N(e^{j\omega})|^2 \widehat{\varphi}_U^N(\omega) \quad (9.6.40)$$

dove $\widehat{G}_N(e^{j\omega})$ è la stima regolarizzata della risposta in frequenza mentre $\widehat{\varphi}_U^N(\omega)$ e $\widehat{\varphi}_Y^N(\omega)$ sono stime delle densità spettrali di u_k e y_k , che possono anche essere regolarizzate con il procedimento precedentemente esaminato, i.e.

$$\begin{aligned} \widehat{\varphi}_U^N(\omega) &= \int_{-\pi}^{\pi} W_{\gamma}(\omega - \xi) \left[\frac{1}{N} U_N(\xi) \overline{U_N(\xi)} \right] d\xi = \sum_{k=-\gamma}^{\gamma} w_{\gamma}(k) \widehat{R}_U^N(k) e^{-j\omega k} \\ \widehat{\varphi}_Y^N(\omega) &= \int_{-\pi}^{\pi} W_{\gamma}(\omega - \xi) \left[\frac{1}{N} Y_N(\xi) \overline{Y_N(\xi)} \right] d\xi = \sum_{k=-\gamma}^{\gamma} w_{\gamma}(k) \widehat{R}_Y^N(k) e^{-j\omega k}. \end{aligned} \quad (9.6.41)$$

La stima $\widehat{\varphi}_V^N(\omega)$ in (9.6.41) ha proprietà asintotiche del tutto analoghe alla stima $\widehat{G}_N(e^{j\omega})$ della risposta in frequenza, precisamente

$$\text{polarizzazione } E [\widehat{\varphi}_V^N(\omega)] - \varphi_V(\omega) = O(M(\gamma)) \quad (9.6.42)$$

$$\text{varianza } E \left\{ [\widehat{\varphi}_V^N(\omega) - E [\widehat{\varphi}_V^N(\omega)]]^2 \right\} = O \left(\frac{\overline{W}(\gamma)}{N} \varphi_V^2(\omega) \right). \quad (9.6.43)$$

9.7 Predittore ottimo per modelli LTI

Una caratteristica di fondamentale importanza dei modelli è la loro capacità di prevedere l'andamento futuro $\{y_{k+1}, y_{k+2}, \dots\}$ del segnale y sulla base del passato $y^k =$

$\{y_k, y_{k-1}, \dots\}$ e, se disponibile, del passato dell'ingresso $u^k \triangleq \{u_k, u_{k-1}, \dots\}$. Non a caso, uno dei criteri di adeguatezza più comunemente impiegati nell'identificazione è quello dell'errore di predizione.

A questo proposito, il presente paragrafo si propone di risolvere il problema della predizione ottima, nel senso del *minimo errore quadratico medio* (MMSE), dell'uscita y_k del sistema di Fig. 9.4. A tale scopo, si consideri la rappresentazione ingresso-uscita di tale sistema

$$y_k = G(z)u_k + H(z)e_k = \sum_{i=1}^{\infty} g_i u_{k-i} + \sum_{j=0}^{\infty} h_j e_{k-j} \quad (9.7.1)$$

dove:

- e_k è un rumore bianco stazionario, i.e. $e_k = \text{swn}(0, \Sigma)$;
- $G(z)$ è una matrice di trasferimento strettamente propria, i.e. $g_0 = 0$;
- $H(z)$ è una matrice di trasferimento tale che $h_0 = I$, BIBO-stabile e con inversa $H^{-1}(z)$ BIBO-stabile.

L'obiettivo è determinare la predizione ottima ad un passo di y_k , dati y_i e u_i per $i \leq k-1$. Tale predizione sarà di seguito indicata con $\hat{y}_{k|k-1}$. Si noti che, da $z^{k-1} = \{u^{k-1}, y^{k-1}\}$, è possibile determinare anche $v_j = y_j - G(z)u_j$ per ogni $j \leq k-1$. Da (9.7.1) risulta dunque

$$y_k = \underbrace{G(z)u_k}_{\bar{y}_k} + v_k = \sum_{i=1}^{\infty} g_i u_{k-i} + v_k \quad (9.7.2)$$

dove il termine $\bar{y}_k = G(z)u_k = \sum_{i=1}^{\infty} g_i u_{k-i}$ risulta esattamente predicibile dagli ingressi passati u_{k-i} , $i \geq 1$. Quindi,

$$\hat{y}_{k|k-1} = G(z)u_k + \hat{v}_{k|k-1} \quad (9.7.3)$$

dove $\hat{v}_{k|k-1}$ è la predizione ad un passo del disturbo

$$v_k = H(z)e_k = e_k + \underbrace{\sum_{j=1}^{\infty} h_j e_{k-j}}_{[H(z)-I]e_k} \quad (9.7.4)$$

Grazie alla BIBO-stabilità di $H^{-1}(z)$, è possibile invertire (9.7.4) e ricavare, causalmente, e_k da v_k tramite

$$e_k = H^{-1}(z)v_k = H^{-1}(z)(y_k - G(z)u_k). \quad (9.7.5)$$

Pertanto, tramite (9.7.5), le innovazioni passate e_{k-j} , $j \geq 1$, in (9.7.4) possono essere determinate causalmente da v^{k-1} , o equivalentemente da $z^{k-1} = \{u^{k-1}, y^{k-1}\}$. Viceversa

il termine e_k in (9.7.4), data la bianchezza di $\{e_k\}$, risulta completamente imprevedibile dal passato. Dai precedenti sviluppi, si ha quindi la seguente predizione ottima ad un passo del disturbo v_k :

$$\begin{aligned}\hat{v}_{k|k-1} &= [H(z) - I] e_k \\ &= [H(z) - I] H^{-1}(z) v_k \\ &= [I - H^{-1}(z)] v_k\end{aligned}\tag{9.7.6}$$

da cui si ottiene immediatamente la seguente predizione ottima ad un passo dell'uscita y_k :

$$\begin{aligned}\hat{y}_{k|k-1} &= G(z)u_k + \hat{v}_{k|k-1} \\ &= G(z)u_k + [I - H^{-1}(z)] v_k \\ &= G(z)u_k + [I - H^{-1}(z)] [y_k - G(z)u_k] \\ &= \underbrace{H^{-1}(z)G(z)}_{P_u(z)} u_k + \underbrace{[I - H^{-1}(z)]}_{P_y(z)} y_k.\end{aligned}\tag{9.7.7}$$

Il corrispondente errore di predizione ad un passo è dato da

$$\begin{aligned}\varepsilon_{k|k-1} = \tilde{y}_{k|k-1} &= y_k - \hat{y}_{k|k-1} \\ &= y_k - H^{-1}(z)G(z)u_k - [I - H^{-1}(z)] y_k \\ &= H^{-1}(z) [y_k - G(z)u_k] \\ &= e_k = \text{swn}(0, \Sigma).\end{aligned}\tag{9.7.8}$$

In altri termini, la variabile di innovazione e_k rappresenta proprio quella parte dell'uscita y_k che non può essere predetta dal passato.

Riassumendo, il predittore ottimo associato al modello LTI (9.7.1)

$$\boxed{\hat{y}_{k|k-1} = H^{-1}(z)G(z)u_k + [I - H^{-1}(z)] y_k}\tag{9.7.9}$$

merita le seguenti considerazioni.

- Esiste una corrispondenza biunivoca fra il modello LTI (9.7.1) ed il predittore (9.7.9). Infatti, definito il modello m come coppia delle funzioni di trasferimento $G(z)$ e $H(z)$, i.e. $m = (G, H)$, ed il predittore associato p come coppia delle funzioni di trasferimento $P_u(z)$ e $P_y(z)$ definite in (9.7.7), i.e. $p = (P_u, P_y)$, è possibile passare univocamente dal modello m al predittore p tramite le relazioni

$$p : \begin{cases} P_u(z) &= H^{-1}(z) G(z) \\ P_y(z) &= I - H^{-1}(z). \end{cases}\tag{9.7.10}$$

e dal predittore p al modello m tramite le seguenti relazioni inverse

$$m : \begin{cases} G(z) &= [I - P_y(z)]^{-1} P_u(z) \\ H(z) &= [I - P_y(z)]^{-1}. \end{cases} \quad (9.7.11)$$

Grazie a questa corrispondenza biunivoca, nel seguito si farà riferimento indifferentemente al modello $m = (G, H)$ o al predittore associato $p = (P_u, P_y)$.

- Negli sviluppi precedenti si è implicitamente assunto che, per la predizione di y_k , sia disponibile tutto il passato dei segnali u_i e y_i per i che va da $-\infty$ a $k-1$. Nella pratica, viceversa, il *dataset* disponibile parte dall'istante iniziale $i = 1$. L'approssimazione più semplice sarebbe quella di rimpiazzare i dati mancanti con zero. Se

$$\begin{aligned} P_u(z) &= H^{-1}(z)G(z) = \sum_{i=1}^{\infty} \beta_i z^{-i} \\ P_y(z) &= I - H^{-1}(z) = \sum_{i=1}^{\infty} \gamma_i z^{-i} \end{aligned}$$

tale approssimazione corrisponde a

$$\hat{y}_{k|k-1} = \sum_{i=1}^{k-1} \beta_i u_{k-i} + \sum_{i=1}^{k-1} \gamma_i y_{k-i}. \quad (9.7.12)$$

Dal momento che, per ipotesi, $H^{-1}(z)$ è BIBO-stabile anche $P_y(z) = I - H^{-1}(z)$ lo è. Se anche $G(z)$ è BIBO-stabile, come accade per molti sistemi di interesse, la risposte impulsive $\{\beta_i\}$ e $\{\gamma_i\}$ decadono a zero per $i \rightarrow \infty$ per cui il troncamento di tali risposte in (9.7.12) fornisce un'approssimazione soddisfacente.

- Nel caso di modelli parametrici $m(\theta) = (G(z, \theta), H(z, \theta))$, entrambe le funzioni di trasferimento del modello sono parametrizzate da un vettore finito-dimensionale $\theta \in \mathbb{R}^{n_\theta}$. Di conseguenza, anche il predittore associato $p(\theta) = (P_u(z, \theta), P_y(z, \theta))$ dipende da θ e così anche la predizione ottima

$$\hat{y}_{k|\theta} \triangleq \hat{y}_{k|k-1, \theta} = H^{-1}(z, \theta)G(z, \theta) u_k + [I - H^{-1}(z, \theta)] y_k \quad (9.7.13)$$

ed il corrispondente errore di predizione

$$\varepsilon_{k|\theta} \triangleq \varepsilon_{k|k-1, \theta} = H^{-1}(z, \theta) [y_k - G(z, \theta)u_k]. \quad (9.7.14)$$

Vale la pena notare che affinché il modello $m(\theta) = (G(z, \theta), H(z, \theta))$ abbia senso occorre che il predittore associato (9.7.13) fornisca predizioni $\hat{y}_{k|\theta}$ limitate per segnali u_k e y_k limitati, ovvero che il predittore $p(\theta) = (P_u(z, \theta), P_y(z, \theta))$ sia BIBO-stabile.

Questa considerazione motiva la seguente definizione di insieme di parametri ammissibili.

Definizione (Insieme dei parametri ammissibili) - Data la classe di modelli parametrici $\mathcal{M} = \{m(\theta) = (G(z, \theta), H(z, \theta)), \theta \in \mathbb{R}^{n_\theta}\}$, si definisce l'insieme dei parametri ammissibili $\Theta \subseteq \mathbb{R}^{n_\theta}$ come segue

$$\Theta \triangleq \left\{ \theta \in \mathbb{R}^{n_\theta} : \underbrace{G(z, \theta)H^{-1}(z, \theta)}_{P_u(z, \theta)} \ \& \ \underbrace{I - H^{-1}(z, \theta)}_{P_y(z, \theta)} \text{ sono BIBO-stabili} \right\} \quad (9.7.15)$$

Un metodo classico di identificazione parametrica, che verrà trattato in dettaglio nel prossimo capitolo, consiste nel determinare il vettore dei parametri come soluzione del seguente problema di ottimizzazione

$$\theta^* = \arg \min_{\theta} \sum_{k=1}^N \varepsilon_{k|k-1, \theta}^T W_k \varepsilon_{k|k-1, \theta} = \arg \min_{\theta} \sum_{k=1}^N (y_k - \hat{y}_{k|k-1, \theta})^T W_k (y_k - \hat{y}_{k|k-1, \theta})$$

che comporta la minimizzazione di una opportuna media pesata, con pesi $W_k = W_k^T > 0$, degli errori di predizione.

- Il precedente predittore ad un passo può essere generalizzato ad $\ell \geq 1$ passi come di seguito esposto. L'obiettivo è dunque determinare $\hat{y}_{k|k-\ell}$ avendo a disposizione $u^{k-1} \triangleq \{u_i, i \leq k-1\}$ e $y^{k-\ell} \triangleq \{y_i, i \leq k-\ell\}$. Si noti che, in analogia con (9.7.3), si ha

$$\hat{y}_{k|k-\ell} = G(z)u_k + \hat{v}_{k|k-\ell}. \quad (9.7.16)$$

Inoltre, in analogia con (9.7.4), si può decomporre v_k come segue

$$\begin{aligned} v_k &= \underbrace{[I + h_1 z^{-1} + \cdots + h_{\ell-1} z^{-(\ell-1)}]}_{H_\ell(z)} e_k + \underbrace{\sum_{i=\ell}^{\infty} h_i z^{-i}}_{H(z) - H_\ell(z)} e_k \\ &= \underbrace{e_k + h_1 e_{k-1} + \cdots + h_{\ell-1} e_{k-\ell+1}}_{\tilde{v}_{k|k-\ell}} + \underbrace{\sum_{i=\ell}^{\infty} h_i e_{k-i}}_{\hat{v}_{k|k-\ell}} \end{aligned} \quad (9.7.17)$$

dove il termine $\hat{v}_{k|k-\ell}$ è esattamente predicibile sulla base di $v^{k-\ell}$ mentre $\tilde{v}_{k|k-\ell}$ è completamente imprevedibile. Quindi,

$$\begin{aligned} \hat{v}_{k|k-\ell} &= [H(z) - H_\ell(z)] e_k \\ &= [H(z) - H_\ell(z)] H^{-1}(z) v_k \\ &= [I - H_\ell(z)H^{-1}(z)] v_k \end{aligned} \quad (9.7.18)$$

da cui si deduce la seguente predizione ottima ad ℓ passi dell'uscita y_k :

$$\begin{aligned}\hat{y}_{k|k-\ell} &= G(z)u_k + [I - H_\ell(z)H^{-1}(z)] v_k \\ &= G(z)u_k + [I - H_\ell(z)H^{-1}(z)] [y_k - G(z)u_k] \\ &= H_\ell(z)H^{-1}(z)G(z)u_k + [I - H_\ell(z)H^{-1}(z)] y_k.\end{aligned}\quad (9.7.19)$$

Il corrispondente errore di predizione ad ℓ passi è dato da

$$\varepsilon_{k|k-\ell} = H_\ell(z)H^{-1}(z) [y_k - G(z)u_k]. \quad (9.7.20)$$

Si noti che per $\ell = 1$, $H_\ell(z) = I$ e quindi le formule (9.7.18)-(9.7.20) del predittore ad ℓ passi si riducono alle formule (9.7.6)-(9.7.8) del predittore ad un passo. Si noti anche che l'errore di predizione ad ℓ passi in (9.7.20)

$$\varepsilon_{k|k-\ell} = \tilde{v}_{k|k-\ell} = H_\ell(z)e_k = e_k + h_1e_{k-1} + \cdots + h_{\ell-1}e_{k-\ell+1}$$

è un processo a media mobile (MA) di ordine $\ell - 1$ e di varianza

$$E [\varepsilon_{k|k-\ell} \varepsilon_{k|k-\ell}^T] = \Sigma + h_1 \Sigma h_1^T + \cdots + h_{\ell-1} \Sigma h_{\ell-1}^T = \sum_{i=0}^{\ell-1} h_i \Sigma h_i^T \quad (9.7.21)$$

che, nel caso di uscita scalare ($\dim y = p = 1$, $\Sigma = \sigma^2$), si riduce a

$$\sigma_{k|k-\ell}^2 = E [\varepsilon_{k|k-\ell}^2] = (1 + h_1^2 + \cdots + h_{\ell-1}^2) \sigma^2. \quad (9.7.22)$$

Da (9.7.21) o (9.7.22) risulta evidente come la qualità della predizione si deteriori all'aumentare del numero ℓ di passi. Per questo motivo, solitamente si preferisce misurare l'adeguatezza di un modello sulla base dell'errore di predizione ad un passo. \square

Se il modello considerato è descritto da equazioni di stato, si può effettuare la predizione ottima dell'uscita avvalendosi del predittore di Kalman. Si consideri, ad esempio, il modello di stato lineare

$$\left\{ \begin{array}{l} x_{k+1} = A_k x_k + B_k u_k + D_k w_k \\ y_k = C_k x_k + v_k \\ w_k = wn(0, Q_k) \\ v_k = wn(0, R_k) \\ w_k \perp v_j \end{array} \right. \quad (9.7.23)$$

per il quale la predizione ad un passo dell'uscita può essere determinata mediante il seguente *predittore di Kalman*

$$\begin{cases} \hat{y}_{k|k-1} = C_k \hat{x}_{k|k-1} \\ K_k = A_k P_{k|k-1} C_k^T (R_k + C_k P_{k|k-1} C_k^T)^{-1} \\ \hat{x}_{k+1|k} = A_k \hat{x}_{k|k-1} + B_k u_k + K_k (y_k - C_k \hat{x}_{k|k-1}) \\ P_{k+1|k} = (A_k - K_k C_k) P_{k|k-1} A_k^T + D_k Q_k D_k^T. \end{cases} \quad (9.7.24)$$

Se inoltre il modello (9.7.23) è tempo-invariante con disturbi stazionari, si può fare ricorso anche al *predittore di Kalman stazionario*

$$\begin{cases} \hat{y}_{k|k-1} = C \hat{x}_{k|k-1} \\ K = A P C^T (R + C P C^T)^{-1} \\ \hat{x}_{k+1|k} = A \hat{x}_{k|k-1} + B u_k + K (y_k - C \hat{x}_{k|k-1}) \end{cases} \quad (9.7.25)$$

dove $P = P^T > 0$ è soluzione della seguente equazione algebrica di Riccati

$$P = A P A^T - A P C^T (R + C P C^T)^{-1} C P A^T + D Q D^T.$$

9.8 Modelli polinomiali a scatola nera

In questo paragrafo si vuole presentare, relativamente a sistemi o segnali scalari, una classe di modelli LTI parametrici *a scatola nera* che considera le funzioni di trasferimento G ed H razionali e le parametrizza in modo naturale mediante i coefficienti dei polinomi a numeratore e denominatore.

Modello BB generale

Nella struttura più generale, tali modelli assumono la forma

$$\begin{cases} A(z) y_k = \frac{B(z)}{F(z)} u_{k-d} + \frac{C(z)}{D(z)} e_k \\ e_k = \text{swn}(0, \sigma^2) \end{cases} \quad (9.8.1)$$

con polinomi

$$\begin{aligned} A(z) &= 1 + a_1 z^{-1} + \dots + a_{n_a} z^{-n_a} \\ B(z) &= b_1 + \dots + b_{n_b} z^{-n_b+1} \\ C(z) &= 1 + c_1 z^{-1} + \dots + c_{n_c} z^{-n_c} \\ D(z) &= 1 + d_1 z^{-1} + \dots + d_{n_d} z^{-n_d} \\ F(z) &= 1 + f_1 z^{-1} + \dots + f_{n_f} z^{-n_f} \end{aligned} \quad (9.8.2)$$

di gradi $n_a \geq 0, n_b \geq 0, n_c \geq 0, n_d \geq 0, n_f \geq 0$ e ritardo ingresso-uscita $d \geq 1$. Un modello polinomiale della forma (9.8.1)-(9.8.2) viene denominato in seguito $BB(n_a, n_b, n_c, n_d, n_f, d)$ dove BB sta per *black-box* e $(n_a, n_b, n_c, n_d, n_f, d)$ sono i suoi indici strutturali. Tali modelli sono caratterizzati dalla due funzioni di trasferimento

$$G(z, \theta) = \frac{B(z)}{A(z)F(z)} z^{-d}, \quad H(z, \theta) = \frac{C(z)}{A(z)D(z)} \quad (9.8.3)$$

individuate dal vettore di parametri

$$\begin{aligned} \theta &= [a_1, \dots, a_{n_a}; b_1, \dots, b_{n_b}; c_1, \dots, c_{n_c}; d_1, \dots, d_{n_d}; f_1, \dots, f_{n_f}]^T \in \mathbb{R}^{n_\theta} \\ n_\theta &= n_a + n_b + n_c + n_d + n_f \end{aligned} \quad (9.8.4)$$

che contiene i coefficienti di tutti i polinomi considerati in ordine lessicografico. Si fa notare come l'utilizzo di cinque polinomi in (9.8.1) risulti ridondante. Infatti, ne basterebbero quattro (e.g. B, F, C, D) per parametrizzare indipendentemente numeratore e denominatore delle due funzioni di trasferimento $G(z, \theta)$ e $H(z, \theta)$. In realtà, considerando un denominatore comune di $G(z, \theta)$ e $H(z, \theta)$ come minimo comune multiplo dei loro denominatori, ne potrebbero bastare tre (e.g. A, B, C) per poter rappresentare un generico modello LTI $m = (G, H)$. Si fa presente, tuttavia, come la complessità del modello dipenda dal numero di parametri n_θ (coefficienti dei polinomi) e non dal numero di polinomi. A questo proposito, si vedrà in seguito come l'aumento del numero di polinomi da tre (A, B, C) a quattro (B, C, D, F), ma anche da quattro (B, C, D, F) a cinque (A, B, C, D, F) possa consentire di ridurre il numero complessivo di coefficienti di tali polinomi e, conseguentemente, la complessità del modello.

Per quanto riguarda il predittore BB, le sue funzioni di trasferimento risultano

$$P_u(z, \theta) = \frac{G(z, \theta)}{H(z, \theta)} = \frac{B(z)D(z)}{C(z)F(z)} z^{-d}, \quad P_y(z, \theta) = 1 - \frac{1}{H(z, \theta)} = \frac{C(z) - A(z)D(z)}{C(z)} \quad (9.8.5)$$

da cui

$$\begin{aligned} \hat{y}_{k|k-1, \theta} &= \frac{B(z)D(z)}{C(z)F(z)} u_{k-d} + \frac{C(z) - A(z)D(z)}{C(z)} y_k \\ \varepsilon_{k|k-1, \theta} &= \frac{A(z)D(z)}{C(z)} \left[y_k - \frac{B(z)}{A(z)F(z)} u_{k-d} \right] \\ &= \frac{A(z)D(z)}{C(z)} y_k - \frac{B(z)D(z)}{C(z)F(z)} u_{k-d}. \end{aligned} \quad (9.8.6)$$

Si noti che il predittore BB risulta BIBO-stabile se e solo se i polinomi $C(z)$ ed $F(z)$ hanno tutte le radici in $|z| < 1$. Nel seguito, si vuole mostrare come tale predittore possa essere posto nella seguente forma di *regressione pseudo-lineare*

$$\hat{y}_{k|k-1, \theta} = \varphi_{k|\theta}^T \theta \quad (9.8.7)$$

per un opportuno regressore $\varphi_{k|\theta} \in \mathbb{R}^{n_\theta}$ dipendente dal vettore di parametri $\theta \in \mathbb{R}^{n_\theta}$. Per semplicità di notazione, nel seguito si pone $\hat{y}_{k|\theta} = \hat{y}_{k|k-1,\theta}$ e $\varepsilon_{k|\theta} = \varepsilon_{k|k-1,\theta}$; inoltre si omette l'argomento z dei polinomi A, B, C, D, F . Da (9.8.6), si ha

$$C\hat{y}_{k|\theta} = D \underbrace{\frac{B}{F} u_{k-d}}_{w_{k|\theta}} + (C - DA) y_k. \quad (9.8.8)$$

Posto

$$\begin{aligned} w_{k|\theta} \triangleq \frac{B}{F} u_{k-d} &\Rightarrow Fw_{k|\theta} = Bu_{k-d} \\ &\Rightarrow w_{k|\theta} + (F - 1)w_{k|\theta} = Bu_{k-d} \\ &\Rightarrow w_{k|\theta} = Bu_{k-d} + (1 - F)w_{k|\theta}, \end{aligned} \quad (9.8.9)$$

si ottiene

$$C\hat{y}_{k|\theta} = Dw_{k|\theta} + Cy_k - DAy_k. \quad (9.8.10)$$

Negli sviluppi che seguono, si separano i campioni all'istante k dei segnali $\hat{y}_{k|\theta}, w_{k|\theta}, y_k$ da quelli passati, utilizzabili per la predizione, mediante la seguente decomposizione $C = 1 + (C - 1)$ ed analoghe per gli altri polinomi. Quindi, (9.8.10) diventa

$$\hat{y}_{k|\theta} = (1 - C)\hat{y}_{k|\theta} + w_{k|\theta} + (D - 1)w_{k|\theta} + y_k + (C - 1)y_k - Ay_k - (D - 1)Ay_k. \quad (9.8.11)$$

Posto

$$v_{k|\theta} \triangleq w_{k|\theta} - Ay_k \quad (9.8.12)$$

e sfruttando (9.8.9), (9.8.11) si trasforma in

$$\hat{y}_{k|\theta} = (1 - A)y_k + Bu_{k-d} + (C - 1) \underbrace{(y_k - \hat{y}_{k|\theta})}_{\varepsilon_{k|\theta}} + (D - 1)v_{k|\theta} + (1 - F)w_{k|\theta}. \quad (9.8.13)$$

Osservando che

$$\begin{aligned} [1 - A(z)]y_k &= -a_1y_{k-1} - \cdots - a_{n_a}y_{k-n_a} \\ B(z)u_{k-d} &= b_1u_{k-d} + \cdots + b_{n_b}u_{k-d-n_b+1} \\ [C(z) - 1]\varepsilon_{k|\theta} &= c_1\varepsilon_{k-1|\theta} + \cdots + c_{n_c}\varepsilon_{k-n_c|\theta} \\ [D(z) - 1]v_{k|\theta} &= d_1v_{k-1|\theta} + \cdots + d_{n_d}v_{k-n_d|\theta} \\ [1 - F(z)]w_{k|\theta} &= -f_1w_{k-1|\theta} - \cdots - f_{n_f}w_{k-n_f|\theta} \end{aligned} \quad (9.8.14)$$

(9.8.13) può essere posto nella forma (9.8.7) definendo:

$$\begin{aligned} \varphi_{k|\theta} &= \left[-y_{k-1}, \dots, -y_{k-n_a}; u_{k-d}, \dots, u_{k-d-n_b+1}; \varepsilon_{k-1|\theta}, \dots, \varepsilon_{k-n_c|\theta}; \right. \\ &\quad \left. v_{k-1|\theta}, \dots, v_{k-n_d|\theta}; -w_{k-1|\theta}, \dots, -w_{k-n_f|\theta} \right]^T \in \mathbb{R}^{n_\theta} \end{aligned} \quad (9.8.15)$$

dove i segnali $\varepsilon_{i|\theta}$, $w_{i|\theta}$, $v_{i|\theta}$ sono ottenuti mediante le seguenti operazioni di filtraggio lineare dipendenti dai parametri θ :

$$\begin{aligned}\varepsilon_{i|\theta} &= y_i - \hat{y}_{i|\theta}, & i &= k-1, \dots, k-n_c \\ w_{i|\theta} &= \frac{B(z)}{F(z)}u_{i-d}, & i &= k-1, \dots, k-n_f \\ v_{i|\theta} &= w_{i|\theta} - A(z)y_i, & i &= k-1, \dots, k-n_d.\end{aligned}\tag{9.8.16}$$

Si noti come il predittore (9.8.7), a causa della dipendenza del regressore $\varphi_{k|\theta}$ in (9.8.15) da θ tramite le variabili $\{\varepsilon_{k|\theta}, v_{k|\theta}, w_{k|\theta}\}$, non sia *lineare nei parametri* (LIP = Linear In Parameters) da cui deriva l'attributo *pseudo-lineare* riferito alla regressione (9.8.7) che sta ad indicare la apparente somiglianza ma effettiva diversità rispetto ad una vera regressione lineare del tipo $\hat{y}_{k|\theta} = \varphi_k^T \theta$ nel caso in cui il regressore φ_k non dipenda da θ .

La classe di modelli BB in (9.8.1) è la più generale possibile nell'ambito dei modelli BB polinomiali. Essa include vari casi particolari a seconda di quali polinomi fra A, B, C, D, F vengano effettivamente utilizzati. In particolare, se $B(z) = 0$, l'ingresso u_k non è presente in (9.8.1) per cui anche il polinomio $F(z)$ risulta inutilizzato ed il polinomio $D(z)$ diventa irrilevante (può essere inglobato in $A(z)$). In questo caso, il modello risultante

$$A(z)y_k = C(z)e_k$$

diventa un modello del segnale y_k (AR, MA o ARMA a seconda di quali polinomi fra A e C vengono impiegati).

Viceversa, se $B(z) \neq 0$, (9.8.1) costituisce un modello ingresso-uscita di sistema dinamico con ingresso u ed uscita y . Poichè ciascuno dei cinque polinomi (A, B, C, D, F) del modello può essere indipendentemente utilizzato oppure no, ci sono in teoria $2^5 - 1 = 31$ combinazioni possibili (escludendo ovviamente la combinazione banale in cui nessun polinomio viene utilizzato) di classi di modelli polinomiali di cui solo alcune (vedi tabella 9.3) sono comunemente impiegate nella pratica. Poiché i modelli di segnale (AR/MA/ARMA) sono già stati ampiamente trattati nel capitolo 4, nel seguito di questo paragrafo l'attenzione sarà rivolta ad alcuni modelli di sistema (FIR, ARX, OE, ARMAX, BJ), casi particolari del modello BB, di maggior interesse pratico. Per ciascuno di questi, come già fatto per il modello generale BB, verrà ricavato il predittore ottimo ad un passo ed espresso nella forma di regressione pseudo-lineare (9.8.7) definendo opportunamente il regressore, lineare o pseudo-lineare, $\varphi_{k|\theta}$.

Modello FIR

Il modello di sistema più semplice è il modello *FIR* (n_b, d)

$$y_k = B(z)u_{k-d} + e_k\tag{9.8.17}$$

Nome	Modello	Polinomi	$y/u \rightarrow y$	LIP/NLIP	BIBO-s
AR	$Ay_k = e_k$	A	y	LIP	
MA	$y_k = Ce_k$	C	y	NLIP	C
ARMA	$Ay_k = Ce_k$	A, C	y	NLIP	C
FIR	$y_k = Bu_{k-d} + e_k$	B	$u \rightarrow y$	LIP	
ARX	$Ay_k = Bu_{k-d} + e_k$	A, B	$u \rightarrow y$	LIP	
OE	$y_k = \frac{B}{F}u_{k-d} + e_k$	B, F	$u \rightarrow y$	NLIP	F
ARMAX	$Ay_k = Bu_{k-d} + Ce_k$	A, B, C	$u \rightarrow y$	NLIP	C
ARARX	$Ay_k = Bu_{k-d} + \frac{1}{D}e_k$	A, B, D	$u \rightarrow y$	NLIP	
ARARMAX	$Ay_k = Bu_{k-d} + \frac{C}{D}e_k$	A, B, C, D	$u \rightarrow y$	NLIP	C
BJ	$y_k = \frac{B}{F}u_{k-d} + \frac{C}{D}e_k$	B, C, D, F	$u \rightarrow y$	NLIP	C, F
BB	$Ay_k = \frac{B}{F}u_{k-d} + \frac{C}{D}e_k$	A, B, C, D, F	$u \rightarrow y$	NLIP	C, F

Tabella 9.3: Alcune classi speciali di modelli BB polinomiali. Per ciascuna classe viene indicato nella varie colonne da sinistra verso destra: (1) l'acronimo; (2) il modello; (3) i polinomi utilizzati; (4) se il modello si riferisce ad un segnale y oppure ad un sistema con ingresso u ed uscita y ; (5) se il predittore è *lineare/non lineare* nei parametri (LIP/NLIP); (6) i polinomi che devono avere radici in $|z| < 1$ per garantire BIBO-stabilità del predittore.

che fa uso del solo polinomio $B(z)$. Questo modello, con funzioni di trasferimento

$$G(z, \theta) = B(z)z^{-d}, \quad H(z, \theta) = 1, \quad (9.8.18)$$

è caratterizzato da *risposta impulsiva finita* (FIR = Finite Impulse Response), i.e.,

$$g_i = \begin{cases} b_{i+1-d}, & i = d, \dots, d + n_b - 1 \\ 0, & i < d \text{ \& } i > d + n_b - 1. \end{cases} \quad (9.8.19)$$

Pertanto, esso può descrivere in modo approssimato solo sistemi BIBO-stabili la cui risposta impulsiva decade esponenzialmente a zero. La scelta dell'ordine n_b del modello dovrebbe essere tale da garantire $g_i \cong 0$ per $i \geq d + n_b$. Poiché

$$P_u(z) = \frac{G(z, \theta)}{H(z, \theta)} = B(z)z^{-d}, \quad P_y(z) = 1 - \frac{1}{H(z, \theta)} = 0 \quad (9.8.20)$$

il predittore FIR risultante

$$\hat{y}_{k|\theta} = B(z)u_{k-d} \quad (9.8.21)$$

può essere posto nella forma (9.8.7) con vettore dei parametri θ e regressore $\varphi_{k|\theta}$ dati da

$$\begin{aligned}\theta &= [b_1, b_2, \dots, b_{n_b}]^T \in \mathbb{R}^{n_\theta} \\ \varphi_{k|\theta} = \varphi_k &= [u_{k-d}, u_{k-d-1}, \dots, u_{k-d-n_b+1}]^T \in \mathbb{R}^{n_\theta} \\ n_\theta &= n_b.\end{aligned}\tag{9.8.22}$$

Pertanto, nel caso FIR, $\varphi_{k|\theta} = \varphi_k$ non dipende dai parametri θ ed il risultante predittore è effettivamente LIP, per cui l'identificazione di un modello FIR si riduce alla soluzione di un problema ai minimi quadrati lineare che, come ben noto, ammette soluzione esatta. Si noti che il predittore FIR è sempre BIBO-stabile a prescindere dal valore dei parametri, i.e. $\Theta = \mathbb{R}^{n_\theta}$.

Modello ARX

Il modello *autoregressivo con ingresso esogeno* (ARX = AutoRegressive with eXogenous input) definito come segue ed indicato con $ARX(n_a, n_b, d)$

$$A(z)y_k = B(z)u_{k-d} + e_k\tag{9.8.23}$$

fa uso dei soli polinomi $A(z)$ e $B(z)$. Le funzioni di trasferimento del modello ARX risultano

$$G(z, \theta) = \frac{B(z)}{A(z)}z^{-d}, \quad H(z, \theta) = \frac{1}{A(z)}\tag{9.8.24}$$

e quelle del predittore associato

$$P_u(z, \theta) = \frac{G(z, \theta)}{H(z, \theta)} = B(z)z^{-d}, \quad P_y(z, \theta) = 1 - \frac{1}{H(z, \theta)} = 1 - A(z)\tag{9.8.25}$$

da cui il predittore ottimo

$$\hat{y}_{k|\theta} = B(z)u_{k-d} + [1 - A(z)]y_k\tag{9.8.26}$$

può essere posto nella forma (9.8.7) con vettore dei parametri θ e regressore $\varphi_{k|\theta}$ dati da

$$\begin{aligned}\theta &= [a_1, \dots, a_{n_a}; b_1, \dots, b_{n_b}]^T \in \mathbb{R}^{n_\theta} \\ \varphi_{k|\theta} = \varphi_k &= [-y_{k-1}, \dots, -y_{k-n_a}; u_{k-d}, \dots, u_{k-d-n_b+1}]^T \in \mathbb{R}^{n_\theta} \\ n_\theta &= n_a + n_b.\end{aligned}\tag{9.8.27}$$

Pertanto, anche il predittore ARX è LIP con possibilità di determinare i parametri mediante stima ai minimi quadrati lineare. Si fa notare, viceversa, che il modello ARX non è in grado di descrivere un generico sistema LTI con disturbo stazionario in quanto le due funzioni di trasferimento ingresso-uscita $G(z, \theta)$ e del disturbo $H(z, \theta)$ sono vincolate ad avere lo stesso denominatore $A(z)$. Inoltre, il disturbo $v_k = A^{-1}(z)e_k$ considerato nel modello ARX è un processo AR quindi non può avere spettro generico. Anche il predittore ARX risulta BIBO-stabile per ogni θ , i.e. $\Theta = \mathbb{R}^{n_\theta}$.

Modello OE

Il modello, cosiddetto a *errore di uscita* (OE = Output Error), assume la forma

$$y_k = \frac{B(z)}{F(z)} u_{k-d} + e_k \quad (9.8.28)$$

in cui vengono impiegati soltanto i polinomi $B(z)$ e $F(z)$. La sua prerogativa è quella di avere funzione di trasferimento ingresso-uscita generica ma funzione di trasferimento del disturbo unitaria, i.e.,

$$G(z, \theta) = \frac{B(z)}{F(z)} z^{-d}, \quad H(z, \theta) = H(z) = 1 \quad (9.8.29)$$

corrispondente ad un disturbo $v_k = e_k$ bianco. Il modello OE risulta particolarmente appropriato quando si può ritenere trascurabile, rispetto al rumore di misura, il disturbo di processo agente sul sistema. Infatti, se il sistema SISO da modellare è della forma

$$\begin{cases} x_{k+1} = Ax_k + Bu_k \\ y_k = Cx_k + v_k \\ v_k = e_k = wn(0, \sigma^2) \end{cases} \quad (9.8.30)$$

dove si è trascurato il termine di disturbo di processo $D_k w_k$, si ottiene la seguente rappresentazione ingresso-uscita

$$y_k = \underbrace{[C(zI - A)^{-1}B]}_{G(z, \theta) = z^{-d}B(z)/F(z)} u_k + e_k \quad (9.8.31)$$

che corrisponde proprio al modello OE con

$$B(z)z^{-d} = C \operatorname{adj}(zI - A) B, \quad F(z) = \det(zI - A).$$

Il predittore OE ha funzioni di trasferimento

$$P_u(z, \theta) = \frac{G(z, \theta)}{H(z, \theta)} = G(z, \theta) = \frac{B(z)}{F(z)} z^{-d}, \quad P_y(z, \theta) = 1 - \frac{1}{H(z, \theta)} = 0 \quad (9.8.32)$$

da cui

$$F(z) \hat{y}_{k|\theta} = B(z) u_{k-d} \quad (9.8.33)$$

ovvero

$$\hat{y}_{k|\theta} = B(z) u_{k-d} + [1 - F(z)] \hat{y}_{k|\theta} \quad (9.8.34)$$

che può essere espresso in forma di regressione pseudo-lineare (9.8.7) con vettore dei parametri θ e regressore $\varphi_{k|\theta}$ come segue:

$$\begin{aligned}\theta &= [b_1, \dots, b_{n_b}; f_1, \dots, f_{n_f}]^T \in \mathbb{R}^{n_\theta} \\ \varphi_{k|\theta} &= [u_{k-d}, \dots, u_{k-d-n_b+1}; -\hat{y}_{k-1|\theta}, \dots, -\hat{y}_{k-n_f|\theta}]^T \in \mathbb{R}^{n_\theta} \\ n_\theta &= n_b + n_f.\end{aligned}\tag{9.8.35}$$

Pertanto il predittore OE, diversamente dai predittori FIR e ARX, non è LIP e l'identificazione (stima dei parametri) di tale modello richiede dunque la soluzione di un problema ai minimi quadrati non lineare data la dipendenza di $\varphi_{k|\theta}$ da θ . Ovviamente, il modello OE non è generale in quanto è vincolato ad un disturbo bianco (i.e., con densità spettrale costante). Il predittore OE risulta BIBO-stabile se e solo se il polinomio $F(z)$ ha tutte le radici in $|z| < 1$.

Modello ARMAX

Il modello *ARMAX* (n_a, n_b, n_c, d), dove ARMAX sta per *ARMA with eXogenous input*, è della forma

$$A(z)y_k = B(z)u_{k-d} + C(z)e_k\tag{9.8.36}$$

che utilizza soltanto i tre polinomi $A(z), B(z), C(z)$. Le funzioni di trasferimento del modello ARMAX sono:

$$G(z, \theta) = \frac{B(z)}{A(z)}z^{-d}, \quad H(z, \theta) = \frac{C(z)}{A(z)}.\tag{9.8.37}$$

Di primo acchito, il modello ARMAX sembra non essere del tutto generale in quanto le due funzioni di trasferimento $G(z, \theta)$ e $H(z, \theta)$ risultano avere lo stesso denominatore $A(z)$. In realtà, come si vedrà di seguito, la famiglia dei modelli ARMAX è in grado di contenere tutti i modelli LTI con arbitrarie $G(z, \theta)$ e $H(z, \theta)$. Si consideri, infatti, il modello

$$y_k = \frac{B_G(z)}{A_G(z)}u_{k-d} + \frac{C_H(z)}{A_H(z)}e_k\tag{9.8.38}$$

con generiche funzioni di trasferimento $G = B_G/A_G$ e $H = C_H/A_H$ parametrizzate con numeratori e denominatori indipendenti. Considerando $A(z)$ come *minimo comune multiplo* di $A_G(z)$ e $A_H(z)$, i.e.

$$A(z) = mcm(A_G(z), A_H(z)),\tag{9.8.39}$$

e $\Delta(z)$ come *massimo comun divisore* dei medesimi, i.e.

$$\Delta(z) = MCD(A_G(z), A_H(z)),\tag{9.8.40}$$

il modello (9.8.38) può essere posto nella forma ARMAX (9.8.36) con $A(z)$ dato da (9.8.39) e

$$B(z) = B_G(z) \underbrace{A_H(z)/\Delta(z)}_{\tilde{A}_H(z)}, \quad C(z) = C_H(z) \underbrace{A_G(z)/\Delta(z)}_{\tilde{A}_G(z)}. \quad (9.8.41)$$

Si fa notare, tuttavia, come la trasformazione di (9.8.38) in (9.8.36) possa comportare un aumento dei coefficienti (parametri del modello). Ad esempio, se $A_G(z)$ e $A_H(z)$ sono primi i.e. $\Delta(z) = 1$, risulta che il modello (9.8.38) coinvolge un numero inferiore di parametri rispetto al modello ARMAX trasformato con polinomi A, B, C definiti in (9.8.39) e (9.8.41). Infatti,

$$\partial A_G + \partial A_H + \partial B_G + \partial C_H < \partial A + \partial B + \partial C = 2\partial A_G + 2\partial A_H + \partial B_G + \partial C_H.$$

Le funzioni di trasferimento del predittore ARMAX risultano

$$P_u(z, \theta) = \frac{G(z, \theta)}{H(z, \theta)} = \frac{B(z)}{C(z)} z^{-d}, \quad P_y(z, \theta) = 1 - \frac{1}{H(z, \theta)} = \frac{C(z) - A(z)}{C(z)} \quad (9.8.42)$$

da cui

$$C(z) \hat{y}_{k|\theta} = B(z) u_{k-d} + [C(z) - A(z)] y_k. \quad (9.8.43)$$

Pertanto, il predittore ARMAX è BIBO-stabile se e solo se il polinomio $C(z)$ ha tutte le radici in $|z| < 1$. Con semplici passaggi, da (9.8.43) si ha

$$\hat{y}_{k|\theta} = (1 - A) y_k + B u_{k-d} + (C - 1) \varepsilon_{k|\theta} \quad (9.8.44)$$

da cui risulta immediato porre il predittore in forma di regressione pseudo-lineare (9.8.7) con

$$\begin{aligned} \theta &= [a_1, \dots, a_{n_a}; b_1, \dots, b_{n_b}; c_1, \dots, c_{n_c}]^T \in \mathbb{R}^{n_\theta} \\ \varphi_{k|\theta} &= [-y_{k-1}, \dots, -y_{k-n_a}; u_{k-d}, \dots, u_{k-d-n_b+1}; \varepsilon_{k-1|\theta}, \dots, \varepsilon_{k-n_c|\theta}]^T \in \mathbb{R}^{n_\theta} \\ n_\theta &= n_a + n_b + n_c. \end{aligned} \quad (9.8.45)$$

Anche in questo caso il regressore $\varphi_{k|\theta}$ risulta dipendere da θ .

Modello BJ

Il modello $BJ(n_b, n_c, n_d, n_f, d)$, dove BJ sta per Box-Jenkins [4], assume la seguente forma:

$$y_k = \frac{B(z)}{F(z)} u_{k-d} + \frac{C(z)}{D(z)} e_k \quad (9.8.46)$$

con polinomi B, F, C, D che parametrizzano indipendentemente le funzioni di trasferimento del modello

$$G(z, \theta) = \frac{B(z)}{F(z)} z^{-d}, \quad H(z, \theta) = \frac{C(z)}{D(z)}. \quad (9.8.47)$$

Il predittore BJ ha funzioni di trasferimento

$$P_u(z, \theta) = \frac{G(z, \theta)}{H(z, \theta)} = \frac{B(z)D(z)}{C(z)F(z)}z^{-d}, \quad P_y(z, \theta) = 1 - \frac{1}{H(z)} = \frac{C(z) - D(z)}{C(z)}. \quad (9.8.48)$$

per cui risulta

$$C(z)\hat{y}_{k|\theta} = D(z)\frac{B(z)}{F(z)}u_{k-d} + [C(z) - D(z)]y_k. \quad (9.8.49)$$

Quindi, il predittore BJ è BIBO-stabile se e solo se i polinomi $C(z)$ ed $F(z)$ hanno tutte le radici in $|z| < 1$. Con semplici passaggi, di seguito riportati, si pone il predittore (9.8.49) nella forma (9.8.7):

$$\begin{aligned} C\hat{y}_{k|\theta} &= D \underbrace{\frac{B}{F}u_{k-d}}_{w_{k|\theta}} + (C - D)y_k \\ &= Dw_{k|\theta} + (C - 1)y_k + (1 - D)y_k \\ &= w_{k|\theta} + (C - 1)y_k + (D - 1)(w_{k|\theta} - y_k) \end{aligned} \quad (9.8.50)$$

da cui, posto $v_{k|\theta} \triangleq w_{k|\theta} - y_k$ e osservando che $w_{k|\theta} = Bu_{k-d} + (1 - F)w_{k|\theta}$, si ottiene

$$\hat{y}_{k|\theta} = Bu_{k-d} + (C - 1) \underbrace{(y_k - \hat{y}_{k|\theta})}_{\varepsilon_{k|\theta}} + (D - 1)v_{k|\theta} + (1 - F)w_{k|\theta}. \quad (9.8.51)$$

Il predittore (9.8.51) può essere espresso in forma di regressione pseudo-lineare (9.8.7) con regressore $\varphi_{k|\theta}$ e vettore dei parametri θ dati da

$$\begin{aligned} \theta &= [b_1, \dots, b_{n_b}; c_1, \dots, c_{n_c}; d_1, \dots, d_{n_d}; f_1, \dots, f_{n_f}]^T \in \mathbb{R}^{n_\theta} \\ \varphi_{k|\theta} &= [u_{k-d}, \dots, u_{k-d-n_b+1}; \varepsilon_{k-1|\theta}, \dots, \varepsilon_{k-n_c|\theta}; v_{k-1|\theta}, \dots, v_{k-n_d|\theta}; \\ &\quad -w_{k-1|\theta}, \dots, -w_{k-n_f|\theta}]^T \in \mathbb{R}^{n_\theta} \\ n_\theta &= n_b + n_c + n_d + n_f. \end{aligned} \quad (9.8.52)$$

Pertanto il predittore BB è anch'esso non lineare nei parametri.

È opportuno osservare che se i denominatori $F(z)$ e $D(z)$ del modello BJ non sono primi, i.e. hanno un massimo comun divisore $\Delta(z) = MCD(F(z), D(z))$ di grado $\partial\Delta(z) > 0$, allora risulta conveniente, dal punto di vista del numero di parametri n_θ , passare al modello BB aggiungendo il quinto polinomio $A(z) = \Delta(z)$. Infatti, posto $F(z) = \Delta(z)\tilde{F}(z)$ e $D(z) = \Delta(z)\tilde{D}(z)$, il modello BB così ottenuto, i.e.,

$$\Delta(z)y_k = \frac{B(z)}{\tilde{F}(z)}u_{k-d} + \frac{C(z)}{\tilde{D}(z)}e_k \quad (9.8.53)$$

ha numero di parametri inferiore a quello del modello BJ originario, i.e.

$$\partial\Delta + \partial B + \partial C + \partial\tilde{D} + \partial\tilde{F} < \partial B + \partial C + \partial D + \partial F = 2\partial\Delta + \partial B + \partial C + \partial\tilde{D} + \partial\tilde{F}.$$

9.9 Modelli di stato a scatola grigia

In questo paragrafo, si vuole esaminare la predizione dell'uscita con riferimento a modelli parametrici di stato a *scatola grigia* ($GB = Grey Box$). Si ricorda come tali modelli sono di norma a tempo-continuo, ovvero espressi sotto forma di equazioni differenziali di stato. Nel caso LTI, di maggior interesse, tali modelli assumono la forma

$$m(\theta) : \begin{cases} \dot{x}(t) = A_c(\theta)x(t) + B_c(\theta)u(t) + D_c(\theta)w(t) \\ y(t) = C(\theta)x(t) + v(t) \\ w(t) = \text{swn}(0, Q_c(\theta)) \\ v(t) = \text{swn}(0, R(\theta)) \\ w(t) \perp v(s) \end{cases} \quad (9.9.1)$$

dove le matrici $A_c(\theta)$, $B_c(\theta)$, $C(\theta)$, $D_c(\theta)$, $Q_c(\theta)$, $R(\theta)$ dipendono da un vettore di parametri finito-dimensionale $\theta \in \mathbb{R}^{n_\theta}$ che individua univocamente un modello parametrico $m(\theta)$. Si assuma che:

- le uscite y_k sono acquisite mediante campionamento regolare, i.e. $y_k = y(kT)$ dove T è l'intervallo di campionamento;
- l'ingresso $u(t)$ viene mantenuto costante in ogni intervallo di campionamento, i.e. $u_k = u(t)$ per ogni $t \in [kT, (k+1)T)$ e per ogni k .

Allora si può effettuare una discretizzazione temporale esatta del sistema dinamico stocastico (9.9.1) come mostrato nel capitolo 4. Si ottiene in questo modo il modello parametrico a tempo-discreto

$$\begin{cases} x_{k+1} = A(\theta)x_k + B(\theta)u_k + w_k \\ y_k = C(\theta)x_k + v_k \\ w_k = \text{swn}(0, Q(\theta)) \\ v_k = \text{swn}(0, R(\theta)) \\ w_k \perp v_j \end{cases} \quad (9.9.2)$$

con $x_k \triangleq x(kT)$, $x_{k+1} \triangleq x((k+1)T)$, $v_k \triangleq v(kT)$ e

$$\begin{aligned} A(\theta) &= e^{A_c(\theta)T} \\ B(\theta) &= \left[\int_0^T e^{A_c(\theta)r} dr \right] B_c(\theta) \\ Q(\theta) &= \int_0^T e^{A_c(\theta)r} D_c(\theta) Q_c(\theta) D_c^T(\theta) e^{A_c^T(\theta)r} dr. \end{aligned} \quad (9.9.3)$$

Al sistema LTI a tempo-discreto (9.9.2)-(9.9.3) si può quindi applicare il predittore di Kalman, sia quello tempo-variante che quello stazionario (a guadagno costante), per generare le predizioni ottime ad un passo $\hat{y}_{k|\theta} = \hat{y}_{k|k-1,\theta}$ nonché il corrispondente errore di predizione $\varepsilon_{k|\theta} = y_k - \hat{y}_{k|\theta}$ che coincide con l'innovazione $e_{k|\theta}$ fornita dal predittore/filtro di Kalman. Ad esempio, se si usa per semplicità il predittore di Kalman stazionario, la predizione desiderata è ottenuta con il seguente algoritmo

$$\begin{cases} \hat{y}_{k|\theta} = C(\theta) \hat{x}_{k|\theta} \\ \varepsilon_{k|\theta} = y_k - \hat{y}_{k|\theta} \\ K(\theta) = A(\theta)P(\theta)C^T(\theta) [R(\theta) + C(\theta)P(\theta)C^T(\theta)]^{-1} \\ \hat{x}_{k+1|\theta} = A(\theta) \hat{x}_{k|\theta} + B(\theta) u_k + K(\theta) \varepsilon_{k|\theta} \end{cases} \quad (9.9.4)$$

dove $P(\theta) = P^T(\theta) > 0$ è soluzione della ARE

$$\begin{aligned} P(\theta) = & A(\theta)P(\theta)A^T(\theta) + Q(\theta) \\ & - A(\theta)P(\theta)C^T(\theta) [R(\theta) + C(\theta)P(\theta)C^T(\theta)]^{-1} C(\theta)P(\theta)A^T(\theta). \end{aligned} \quad (9.9.5)$$

L'algoritmo (9.9.5) deve essere inizializzato con una opportuna condizione iniziale $\hat{x}_{1|\theta}$.

Nel caso in cui si voglia utilizzare l'effettivo predittore ottimo, con guadagno tempo-variante, per tener conto in modo ottimale dell'informazione mancante per $k < 1$, è sufficiente il guadagno costante $K(\theta)$ in (9.9.5) con il guadagno di Kalman tempo-variante $K_k(\theta)$ dipendente da $P_k(\theta)$ soluzione dell'equazione alle differenze di Riccati

$$\begin{aligned} P_{k+1}(\theta) = & A(\theta)P_k(\theta)A^T(\theta) + Q(\theta) \\ & - A(\theta)P_k(\theta)C^T(\theta) [R(\theta) + C(\theta)P_k(\theta)C^T(\theta)]^{-1} C(\theta)P_k(\theta)A^T(\theta) \end{aligned} \quad (9.9.6)$$

che, ovviamente, richiede una condizione iniziale $P_1(\theta)$ per quantificare la covarianza iniziale $E \left[(x_1 - \hat{x}_{1|\theta}) (x_1 - \hat{x}_{1|\theta})^T \right]$.

Vale la pena fare alcune considerazioni finali sulla predizione per modelli a scatola grigia.

- Si ricorda che, sotto opportune ipotesi di osservabilità sul sistema (9.9.1) che risultano genericamente soddisfatte (i.e., per quasi tutti i valori del parametro $\theta \in \mathbb{R}^{n_\theta}$), il predittore di Kalman è BIBO-stabile, vale a dire fornisce predizioni $\hat{y}_{k|\theta}$ limitate per ingressi u_k e uscite y_k limitate.
- Inoltre, vale la pena osservare come tale predittore sia fortemente non lineare nei parametri θ . Infatti, il procedimento che porta da θ alla predizione $\hat{y}_{k|\theta}$ coinvolge la discretizzazione temporale del modello a tempo-continuo (9.9.1) che, attraverso l'esponenziale di matrice, porta le matrici del sistema discretizzato (9.9.2) a dipendere in modo fortemente non lineare dai parametri θ . A questo segue il calcolo

di $\hat{y}_{k|\theta}$ e $\varepsilon_{k|\theta}$ mediante filtraggio alla Kalman che, anche a seguito della soluzione dell'equazione di Riccati, risulta anch'esso fortemente non lineare rispetto a θ . Di fatto, la complessità della mappa $\theta \rightarrow \hat{y}_{k|\theta}$ richiede un elevato onere computazionale per la valutazione di qualunque criterio di adeguatezza $V(z^N, \theta)$ che coinvolga $\varepsilon_{k|\theta}$ e, di fatto, rende impossibile la valutazione del Jacobiano e/o dell'Hessiano di tale criterio rispetto a θ che potrebbero agevolare la minimizzazione.

- Gli strumenti di discretizzazione temporale approssimata (e.g., metodo di Eulero) e di filtraggio/predizione non lineare del capitolo 8 consentono l'estensione dell'approccio sopra esposto anche a modelli parametrici GB non lineari della forma generale

$$m(\theta) : \begin{cases} \dot{x}(t) &= f(t, x(t), u(t), w(t), \theta) \\ y(t) &= h(t, x(t), u(t), v(t), \theta). \end{cases} \quad (9.9.7)$$

Si lascia al lettore l'elaborazione di un procedimento basato sul metodo di discretizzazione di Eulero e qualche algoritmo di filtraggio non lineare (e.g., EKF, UKF o filtro a particelle) per il calcolo delle predizioni $\hat{y}_{k|\theta}$ / errori di predizione $\varepsilon_{k|\theta}$.

Bibliografia

- [1] L. Ljung: *System identification: theory for the user*, Prentice Hall, 1999.
- [2] T. Söderström, P. Stoica: *System identification*, Prentice Hall, 1989.
- [3] S. Bittanti: *Identificazione dei modelli e controllo adattativo*, Pitagora, Bologna, 1997.
- [4] G. Box, G. Jenkins: *Time-series analysis: forecasting and control*, Holden Day, 1979.