



UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Dipartimento
di Ingegneria Gestionale,
dell'Informazione e della Produzione



IDENTIFICAZIONE DEI MODELLI E ANALISI DEI DATI (IMAD)

Lezione 12: Identificazione – analisi e complementi

Corso di Laurea Magistrale in
INGEGNERIA INFORMATICA

SPEAKER

Prof. Mirko Mazzoleni

PLACE

Università degli Studi di
Bergamo

Syllabus

Parte II: sistemi dinamici

8. Processi stocastici

- 8.1 Processi stocastici stazionari (pss)
- 8.3 Rappresentazione spettrale di un pss
- 8.4 Stimatori campionari media\covarianza
- 8.5 Densità spettrale campionaria

9. Famiglie di modelli a spettro razionale

- 9.1 Modelli per serie temporali (MA, AR, ARMA)
- 9.2 Modelli per sistemi input/output (ARX, ARMAX)

10. Predizione

- 10.1 Filtro passa-tutto

10.2 Forma canonica

10.3 Teorema della fattorizzazione spettrale

10.4 Soluzione al problema della predizione

11. Identificazione

11.3 Identificazione di modelli ARX

11.4 Identificazione di modelli ARMAX

11.5 Metodo di Newton

12. Identificazione: analisi e complementi

12.1 Analisi asintotica metodi PEM

12.2 Identificabilità dei modelli

12.3 Valutazione dell'incertezza di stima

13. Identificazione: valutazione



IMAD

Parte I: sistemi statici

Parte II: sistemi dinamici

Stima parametrica $\hat{\theta}$

- θ deterministico

- ***NO assunzioni su ddp dei dati***

- ✓ Stima parametri popolazione
- ✓ Stima modello lineare: minimi quadrati

- ***SI assunzioni su ddp dei dati***

- ✓ Stima massima verosimiglianza parametri popolazione
- ✓ Stima modello lineare: massima verosimiglianza
- ✓ Regressione logistica

- θ variabile casuale

- ***SI assunzioni su ddp dei dati***

- ✓ Stima Bayesiana

Machine learning

Stima parametrica $\hat{\theta}$

- θ deterministico

- ***NO assunzioni su ddp dei dati***

- ✓ Modelli lineari di pss
- ✓ Predizione

- ✓ Identificazione

- ✓ Persistente eccitazione
- ✓ Analisi asintotica metodi PEM
- ✓ Analisi incertezza stima (numero dati finito)

- ✓ Valutazione del modello



Outline

1. Analisi asintotica dei metodi PEM
2. Identificabilità dei modelli e persistente eccitazione
3. Valutazione dell'incertezza della stima PEM
4. Robustezza dei metodi PEM e prefiltraggio
5. Empirical Transfer Function Estimate (ETFE)

Outline

- 1. Analisi asintotica dei metodi PEM**
2. Identificabilità dei modelli e persistente eccitazione
3. Valutazione dell'incertezza della stima PEM
4. Robustezza dei metodi PEM e prefiltraggio
5. Empirical Transfer Function Estimate (ETFE)



Analisi asintotica dei metodi PEM

Nella lezione precedente, abbiamo visto come ottenere una stima $\hat{\theta}_N$ data una **singola sequenza** di N dati $\{u(1), \dots, u(N)\}, \{y(1, \bar{s}), \dots, y(N, \bar{s})\}$, utilizzando **l'approccio PEM**. Abbiamo esplicitato l'esito $s = \bar{s}$ per indicare che lavoriamo con delle **sequenze di numeri**

L'idea dell'approccio predittivo è la seguente: data una **famiglia di modelli** $\{\mathcal{M}(\theta) | \theta \in \Theta\}$, la stima $\hat{\theta}_N$ è ottenuta **minimizzando la funzione di costo**

$$J_N(\theta) = \frac{1}{N} \sum_{t=1}^N \varepsilon_1(t; \theta)^2$$

dove $\varepsilon_1(t; \theta) = y(t) - \hat{y}(t|t-1; \theta)$ è l'**errore di predizione ad un passo**

Analisi asintotica dei metodi PEM

Problema: la stima $\hat{\theta}_N$, calcolata in questo modo, **ci fornisce un «buon» modello?**

Per poter rispondere a questa domanda, non possiamo limitarci ad una **specific**a realizzazione dei dati corrispondente ad un particolare esito \bar{s} : dobbiamo studiare quello che succede «in generale», ovvero considerando **tutte le possibili realizzazioni di sequenze di dati**

IPOTESI DI LAVORO

- Sia **l'ingresso** $u(t, s)$ sia **l'uscita** $y(t, s)$ sono processi stocastici **stazionari** ed **ergodici** (il che implica che tutte le funzioni di trasferimento sono **asintoticamente stabili**)

Analisi asintotica dei metodi PEM

Di conseguenza, i **dati misurati** saranno una realizzazione dei processi $u(t, s)$ e $y(t, s)$ in corrispondenza di un particolare esito \bar{s}

$$\{u(1, \bar{s}), \dots, u(N, \bar{s})\} \quad \{y(1, \bar{s}), \dots, y(N, \bar{s})\}$$

La **funzione di costo** dipende anch'essa dall'esito \bar{s} poiché utilizza i dati misurati

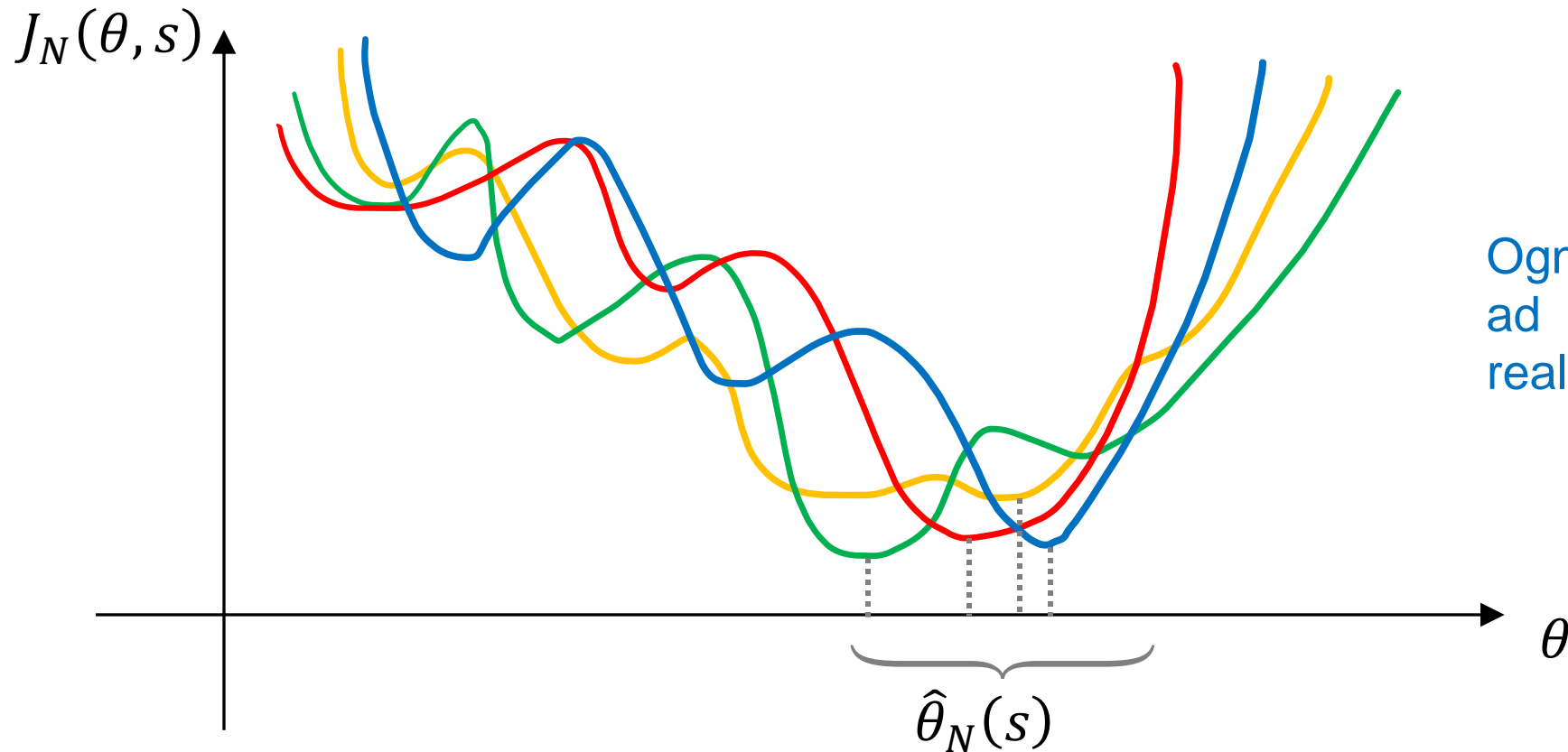
$$J_N(\boldsymbol{\theta}, \bar{s}) = \frac{1}{N} \sum_{t=1}^N \varepsilon_1(t; \boldsymbol{\theta}, \bar{s})^2$$

da cui otterrò la stima $\hat{\boldsymbol{\theta}}_N(\bar{s})$

Ne consegue che, in generale, la stima $\hat{\boldsymbol{\theta}}_N(s)$ è una **variabile casuale** perché il suo valore dipende dai dati, i quali dipendono dall'esito s

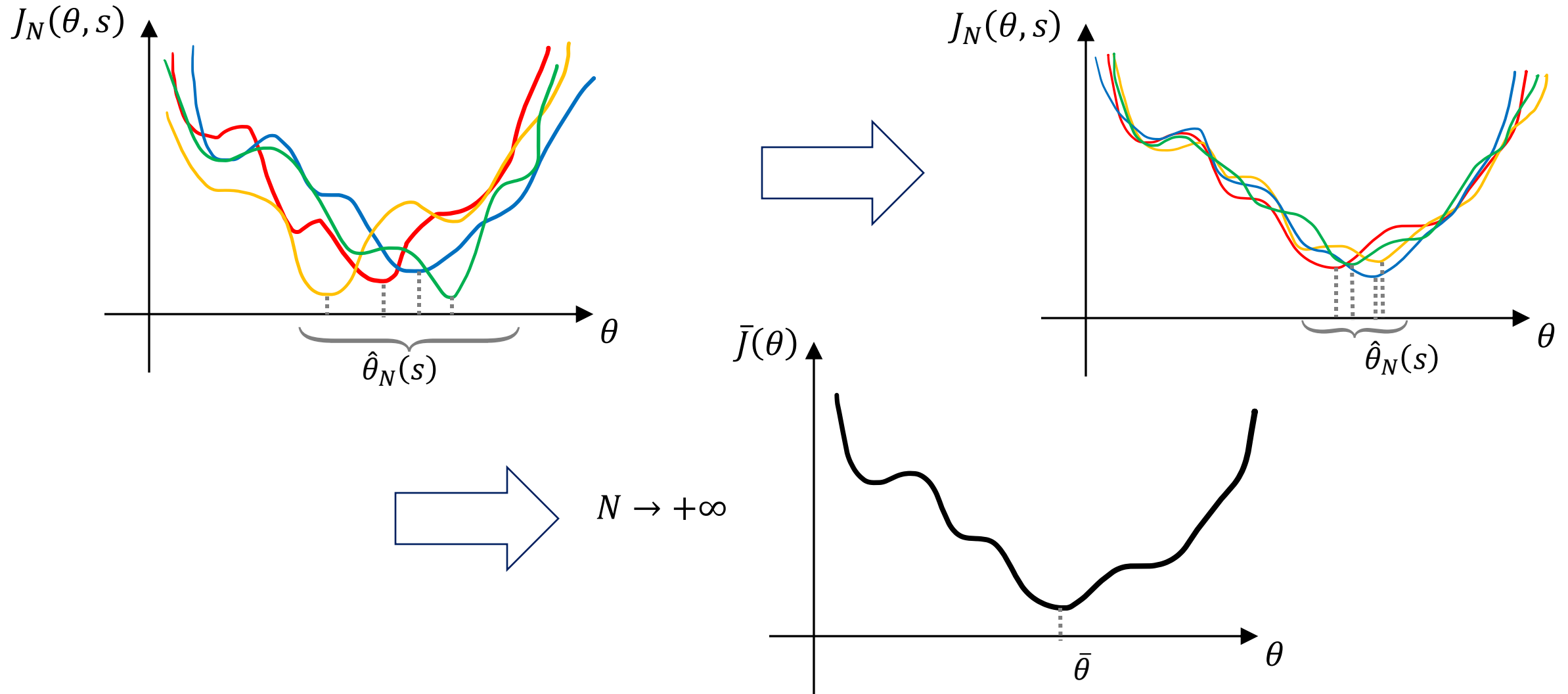
Analisi asintotica dei metodi PEM

La funzione di costo $J_N(\theta, s)$ dovrebbe essere interpretata come un **insieme di curve**, e la stima $\hat{\theta}_N(s)$ come un **insieme di punti** → difficile da descrivere per N finito!



Ogni curva corrisponde ad una diversa realizzazione di N dati

Analisi asintotica dei metodi PEM



Analisi asintotica dei metodi PEM

Grazie all'**ipotesi di ergodicità** di $u(t,s)$ e $y(t,s)$, abbiamo che i momenti temporali convergono ai rispettivi momenti di insieme. Di conseguenza:

$$J_N(\boldsymbol{\theta}, s) = \frac{1}{N} \sum_{t=1}^N \varepsilon_1(t; \boldsymbol{\theta}, s)^2 \quad \xrightarrow{N \rightarrow +\infty} \quad \bar{J}(\boldsymbol{\theta}) \equiv \mathbb{E}_s[\varepsilon_1(t, \boldsymbol{\theta})^2]$$

cioè, le curve $J_N(\boldsymbol{\theta}, s)$ convergono ad un'**unica (deterministica) curva** $\bar{J}(\boldsymbol{\theta})$

Definiamo **l'insieme dei punti di minimo globale** di $\bar{J}(\boldsymbol{\theta})$ come

$$\Delta_{\boldsymbol{\theta}} = \{\bar{\boldsymbol{\theta}} \mid \bar{J}(\boldsymbol{\theta}) \geq \bar{J}(\bar{\boldsymbol{\theta}}), \quad \forall \boldsymbol{\theta}\}$$

Analisi asintotica dei metodi PEM

Caso particolare: $\Delta_\theta = \bar{\theta}$, ovvero $\bar{J}(\theta)$ ha un **unico minimo globale**

Teorema

Sotto le ipotesi correnti, man mano che il numero di dati N tende all'infinito, si ha che

$$J_N(\theta, s) \xrightarrow{N \rightarrow +\infty} \bar{J}(\theta)$$

$$\hat{\theta}_N \xrightarrow{N \rightarrow +\infty} \Delta_\theta$$

Ne segue che se $\Delta_\theta = \bar{\theta}$, allora $\hat{\theta}_N \xrightarrow{N \rightarrow +\infty} \bar{\theta}$

Analisi asintotica dei metodi PEM

Osservazioni

- Il teorema dice che **il risultato dell'identificazione PEM è lo stesso, indipendentemente dalle realizzazioni misurate** dei processi $u(t), y(t)$ purché il numero di dati N sia **abbastanza grande**

Analizzare la singola curva $\bar{J}(\theta)$ è molto più facile che analizzare un insieme di curve!

Idea: per studiare le proprietà della stima, studiamo le sue **caratteristiche asintotiche**, ovvero studiamo il **modello stimato asintotico** $\mathcal{M}(\bar{\theta})$ oppure **l'insieme di modelli stimati asintotici** $\{\mathcal{M}(\theta) | \theta \in \Delta_{\theta}\}$

Se le proprietà di $\mathcal{M}(\bar{\theta})$ sono buone, posso pensare che lo siano anche quelle di $\mathcal{M}(\hat{\theta}_N)$, fintanto che N è grande

Analisi asintotica dei metodi PEM

IPOTESI DI LAVORO AGGIUNTIVA

Assumiamo che $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, ovvero che esista $\boldsymbol{\theta}^0 \in \Theta$ tale che $\mathcal{S} = \mathcal{M}(\boldsymbol{\theta}^0)$

Domanda: il vettore «vero» dei parametri $\boldsymbol{\theta}^0$ appartiene all'insieme Δ_{θ} dei minimi globali della cifra di costo $\bar{J}(\boldsymbol{\theta})$? Ciò è equivalente a chiedersi se $\hat{\boldsymbol{\theta}}_N$ tende asintoticamente a $\boldsymbol{\theta}^0$

Se ciò fosse vero, vorrebbe dire che **i metodi PEM sono in grado di trovare la parametrizzazione «vera» del modello**

Dimostriamo che, sotto le ipotesi fatte, $\boldsymbol{\theta}^0$ **appartiene sempre** a Δ_{θ}

Analisi asintotica dei metodi PEM

Dimostrazione

Supponiamo che i dati siano **generati dal sistema** \mathcal{S} , tale che

$$y(t) = \hat{y}(t|t-1; \boldsymbol{\theta}^0) + e(t), \quad e(t) \sim \text{WN}(0, \lambda^2)$$

Consideriamo un **generico modello** $\mathcal{M}(\boldsymbol{\theta})$, per il quale

$$y(t) = \hat{y}(t|t-1; \boldsymbol{\theta}) + \varepsilon_1(t; \boldsymbol{\theta})$$

Non è detto che $\varepsilon_1(t; \boldsymbol{\theta})$
sia bianco...

L'errore di predizione ad un passo commesso dal modello $\mathcal{M}(\boldsymbol{\theta})$ è dunque

$$\varepsilon_1(t; \boldsymbol{\theta}) = y(t) - \hat{y}(t|t-1; \boldsymbol{\theta})$$

Analisi asintotica dei metodi PEM

Aggiungiamo e togliamo $\hat{y}(t|t-1; \theta^0)$, ovvero il **predittore del sistema** \mathcal{S} che genera i dati

$$\varepsilon_1(t; \theta) = \underbrace{y(t) - \hat{y}(t|t-1; \theta^0)}_{\text{Errore di predizione «ottimo»}} + \hat{y}(t|t-1; \theta^0) - \hat{y}(t|t-1; \theta)$$

Errore di predizione «ottimo»

$$\varepsilon_1(t; \theta^0) = e(t)$$

Pertanto

$$\varepsilon_1(t; \theta) = e(t) + \hat{y}(t|t-1; \theta^0) - \hat{y}(t|t-1; \theta)$$

Calcoliamo la **varianza dell'errore** di predizione (che è a media nulla poiché il predittore è corretto):

$$\mathbb{E}[\varepsilon_1(t; \theta)^2] = \mathbb{E} \left[\left(e(t) + \hat{y}(t|t-1; \theta^0) - \hat{y}(t|t-1; \theta) \right)^2 \right]$$

Analisi asintotica dei metodi PEM

$$\mathbb{E}[\varepsilon_1(t; \boldsymbol{\theta})^2] = \mathbb{E} \left[\left(e(t) + \hat{y}(t|t-1; \boldsymbol{\theta}^0) - \hat{y}(t|t-1; \boldsymbol{\theta}) \right)^2 \right]$$

$$\begin{aligned} \Rightarrow \bar{J}(\boldsymbol{\theta}) &= \mathbb{E}[e(t)^2] + \mathbb{E} \left[\left(\hat{y}(t|t-1; \boldsymbol{\theta}^0) - \hat{y}(t|t-1; \boldsymbol{\theta}) \right)^2 \right] \\ &\quad + 2\mathbb{E} \left[e(t) \cdot \left(\hat{y}(t|t-1; \boldsymbol{\theta}^0) - \hat{y}(t|t-1; \boldsymbol{\theta}) \right) \right] \end{aligned}$$

Le quantità $\hat{y}(t|t-1; \boldsymbol{\theta}^0)$ e $\hat{y}(t|t-1; \boldsymbol{\theta})$ sono predittori, e quindi dipendono solo dai dati (e dal rumore bianco) a tempi passati. Per cui, sono **incorrelati** con $e(t)$

Analisi asintotica dei metodi PEM

$$\bar{J}(\boldsymbol{\theta}) = \mathbb{E}[e(t)^2] + \mathbb{E} \left[\left(\hat{y}(t|t-1; \boldsymbol{\theta}^0) - \hat{y}(t|t-1; \boldsymbol{\theta}) \right)^2 \right]$$

$$\Rightarrow \bar{J}(\boldsymbol{\theta}) = \lambda^2 + \underbrace{\mathbb{E} \left[\left(\hat{y}(t|t-1; \boldsymbol{\theta}^0) - \hat{y}(t|t-1; \boldsymbol{\theta}) \right)^2 \right]}$$

È una varianza, quindi una quantità ≥ 0 . In particolare, si **annulla** solo per $\boldsymbol{\theta} = \boldsymbol{\theta}^0$

$$\Rightarrow \bar{J}(\boldsymbol{\theta}) \geq \lambda^2 = \bar{J}(\boldsymbol{\theta}^0), \quad \forall \boldsymbol{\theta}$$

$$\bar{J}(\boldsymbol{\theta}) \geq \bar{J}(\boldsymbol{\theta}^0), \quad \forall \boldsymbol{\theta}$$

$\boldsymbol{\theta}^0$ è un minimo di $\bar{J}(\boldsymbol{\theta})$

Analisi asintotica dei metodi PEM

Conclusione (fondamentale)

Se $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$ e $u(t), y(t)$ sono pss ergodici, allora, per $N \rightarrow +\infty$, **un metodo PEM garantisce che il modello stimato è quello «vero»**

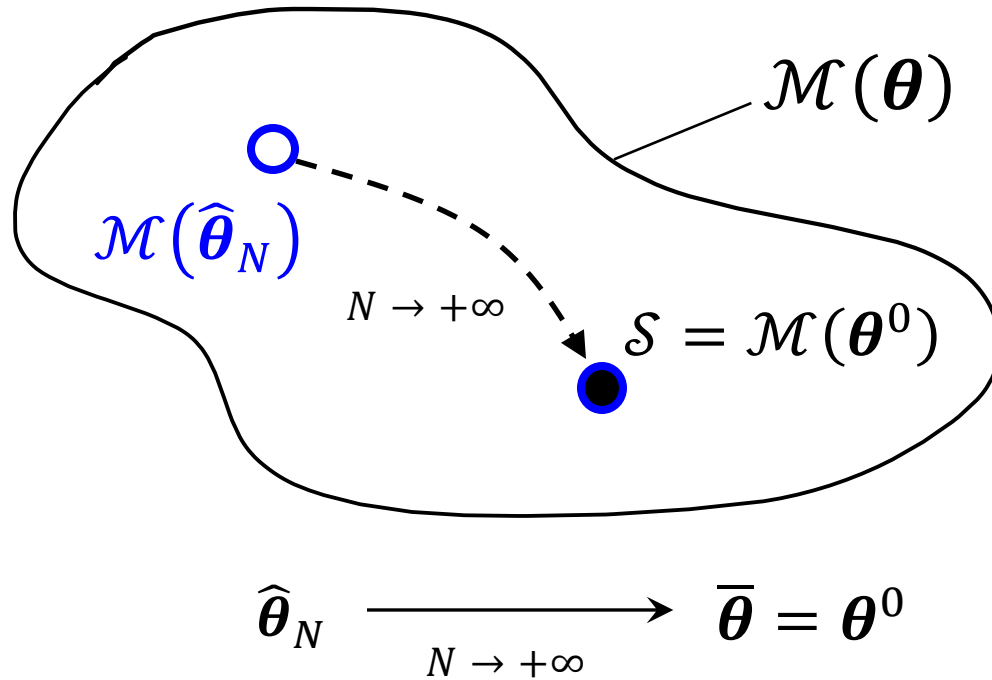
Osservazioni

- Se $\mathcal{S} \notin \mathcal{M}(\boldsymbol{\theta})$, allora i metodi PEM **non garantiscono** di stimare correttamente **TUTTE** le componenti del sistema \mathcal{S} (i.e. modello I/O e modello del rumore)
- Se $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, allora in corrispondenza di $\boldsymbol{\theta}^0$ si ha che $\varepsilon_1(t; \boldsymbol{\theta}^0) = e(t) \sim \text{WN}$

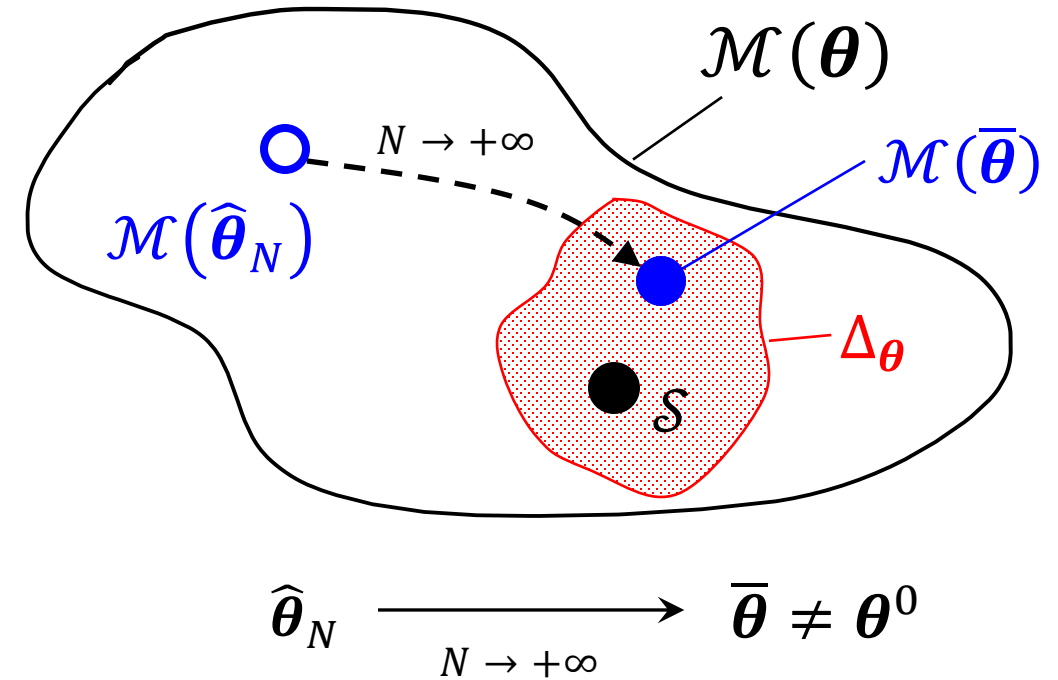
Quindi, possiamo **verificare a posteriori** se il modello identificato è quello vero facendo un **test di bianchezza sui residui** $\varepsilon_1(t; \hat{\boldsymbol{\theta}}_N)$

Quando identifichiamo un modello $\mathcal{M}(\theta)$, possono capitarci quattro casi possibili:

1) $\mathcal{S} \in \mathcal{M}(\theta)$ e $\Delta_\theta = \bar{\theta}$, allora $\bar{\theta} = \theta^0$

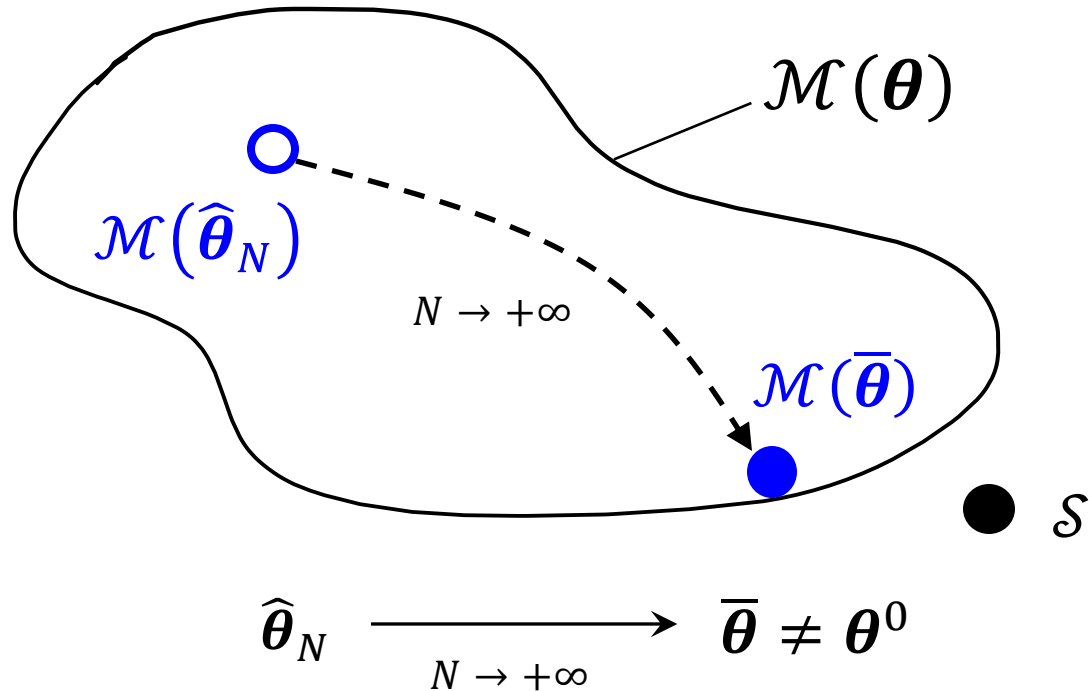


2) $\mathcal{S} \in \mathcal{M}(\theta)$ e Δ_θ contiene più valori



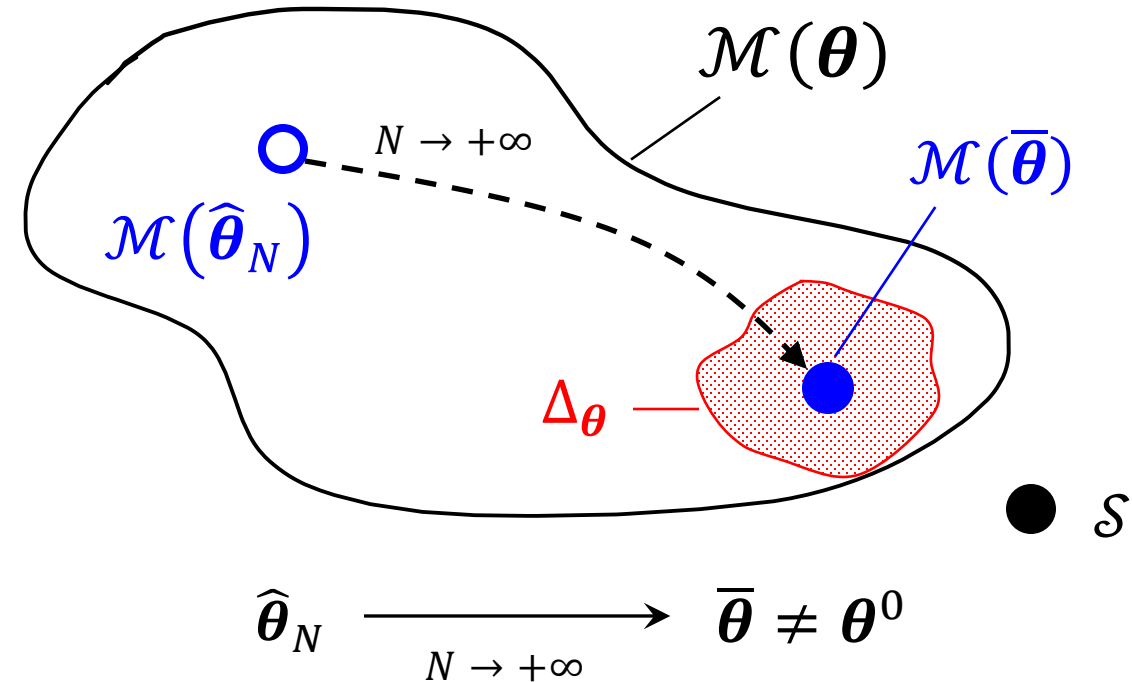
Ma $\mathcal{M}(\bar{\theta})$ ha la **stessa capacità** di $\mathcal{M}(\theta^0)$ nello spiegare i dati

3) $\mathcal{S} \notin \mathcal{M}(\theta)$ e $\Delta_\theta = \bar{\theta}$, allora $\bar{\theta} \neq \theta^0$



$\mathcal{M}(\bar{\theta})$ è la **miglior approssimazione** di \mathcal{S} nella famiglia di modelli $\mathcal{M}(\theta)$

4) $\mathcal{S} \notin \mathcal{M}(\theta)$ e Δ_θ contiene più valori



$\mathcal{M}(\bar{\theta})$ con $\bar{\theta} \in \Delta_\theta$ sono i **migliori approssimanti (equivalenti)** di \mathcal{S} nella famiglia di modelli $\mathcal{M}(\theta)$

Outline

1. Analisi asintotica dei metodi PEM
- 2. Identificabilità dei modelli e persistente eccitazione**
3. Valutazione dell'incertezza della stima PEM
4. Robustezza dei metodi PEM e prefiltraggio
5. Empirical Transfer Function Estimate (ETFE)



Identificabilità dei modelli

L' **analisi asintotica** vista precedentemente ci dice che, se $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, allora i metodi PEM **stimano asintoticamente il modello vero** $\mathcal{S} = \mathcal{M}(\boldsymbol{\theta}^0)$ o un **insieme equivalente di modelli** $\{\mathcal{M}(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \Delta_{\boldsymbol{\theta}}\}$

Questa seconda situazione, in cui troviamo un modello all'interno di $\{\mathcal{M}(\boldsymbol{\theta}) | \boldsymbol{\theta} \in \Delta_{\boldsymbol{\theta}}\}$, porta ad un legittima domanda:

In quali condizioni il sistema \mathcal{S} può essere **identificato univocamente** dai dati?

Identificabilità dei modelli

Affinché un modello sia **univocamente identificabile** è necessario avere:

1. **Identificabilità «strutturale»:** il modello $\mathcal{M}(\theta)$ non deve essere **sovra-parametrizzato** rispetto al sistema \mathcal{S}
2. **Identificabilità «sperimentale»:** i dati $\{u(t), y(t)\}_{t=1}^N$ devono contenere **sufficiente informazione**

Il problema di non identificabilità **più critico è quello sperimentale**: se non abbiamo sufficiente informazione nei dati, non possiamo fare nulla (se non ripetere l'esperimento)

La non identificabilità strutturale è, invece, facilmente risolvibile **riducendo l'ordine** del modello

Esempio: problema di identificabilità strutturale

Supponiamo che i dati siano generati da un **sistema** (deterministico) del tipo

$$\mathcal{S}: G_0(z) = \frac{z}{(z + 0.5)(z + 0.8)}$$

Per l'identificazione, usiamo un **modello** del tipo

$$\mathcal{M}(\boldsymbol{\theta}): G(z) = \frac{z(b_1 z + b_2)}{(z + a_1)(z + a_2)(z + a_3)}$$

$$\boldsymbol{\theta} = [b_1 \ b_2 \ a_1 \ a_2 \ a_3]^\top$$

Notiamo che $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$. Infatti, posso ottenere $G_0(z)$ imponendo, per esempio, che $(b_1 z + b_2) = (z + a_3)$, in modo che il modello diventi

$$\mathcal{M}(\boldsymbol{\theta}): G(z) = \frac{z}{(z + a_1)(z + a_2)}$$

Esempio: problema di identificabilità strutturale

Nonostante riusciamo a identificare un modello che è simile al sistema che genera i dati, tale modello identificato **non è unico in quanto** la condizione $(b_1z + b_2) = (z + a_3)$ è soddisfatta da un **numero infinito** di triplette di valori (b_1, b_2, a_3)

Questo accade perché **la struttura del modello è sovra-parametrizzata** rispetto alla struttura del sistema vero

Esempio: problema di identificabilità sperimentale

Supponiamo che i dati siano generati da un **sistema** del tipo

$$\mathcal{S}: \quad y(t) = \underbrace{\frac{b_0 z^{-1}}{1 + f_0 z^{-1}} u(t)}_{G_0(z)} + \underbrace{\frac{1}{1 + d_0 z^{-1}} e(t)}_{H_0(z)}$$

Consideriamo un ingresso $u(t) = 0 \quad \forall t$, cioè un **ingresso costante a zero**, e un modello

$$\mathcal{M}(\boldsymbol{\theta}): \quad y(t) = \underbrace{\frac{b z^{-1}}{1 + f z^{-1}} u(t)}_{G(z; \boldsymbol{\theta})} + \underbrace{\frac{1}{1 + d z^{-1}} \eta(t)}_{H(z; \boldsymbol{\theta})} \quad \boldsymbol{\theta} = [b \quad f \quad d]^\top$$

Notiamo che $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$

Esempio: problema di identificabilità sperimentale

Calcoliamo l'errore di predizione a un passo

$$\varepsilon_1(t; \boldsymbol{\theta}) = H^{-1}(z, \boldsymbol{\theta})[y(t) - G(z, \boldsymbol{\theta})u(t)] = H^{-1}(z, \boldsymbol{\theta})[G_0(z)u(t) + H_0(z)e(t) - G(z, \boldsymbol{\theta})u(t)]$$

$$= \underbrace{\frac{G_0(z) - G(z, \boldsymbol{\theta})}{H(z, \boldsymbol{\theta})} u(t)}_{= 0} + \frac{H_0(z)}{H(z, \boldsymbol{\theta})} e(t) = \frac{H_0(z)}{H(z, \boldsymbol{\theta})} e(t) \quad \Rightarrow \quad \boxed{\varepsilon_1(t; \boldsymbol{\theta}) = \frac{1 + dz^{-1}}{1 + d_0 z^{-1}} e(t)}$$

La quantità $\mathbb{E}[\varepsilon_1(t; \boldsymbol{\theta})^2]$ è **minimizzata** quando $\varepsilon_1(t; \bar{\boldsymbol{\theta}}) = e(t)$. Questo accade quando $H_0(z) = H(z, \bar{\boldsymbol{\theta}})$, ovvero quando

$$\bar{\boldsymbol{\theta}} = [b \ d_0 \ f]^\top, \quad \forall b, f \in \mathbb{R}$$

Abbiamo stimato il vero parametro d_0 !

Esempio: problema di identificabilità sperimentale

Misurando l'uscita $y(t)$ quando l'ingresso $u(t) = 0$ ci consente di «misurare l'effetto prodotto dal rumore $e(t)$ sull'uscita». Ciò permette di **identificare correttamente il modello dell'errore** (sotto l'ipotesi che $H_0(z) \in H(z, \theta)$, implicito nel fatto che $\mathcal{S} \in \mathcal{M}(\theta)$)

Però, questi dati **non ci consentono di identificare** i parametri di $G_0(z)$! Infatti, la funzione di costo è minimizzata indipendentemente da b ed f , che possono **assumere qualsiasi valore**.

Ancora una volta, il modello non viene stimato in modo univoco

Diciamo quindi che i dati **non sono abbastanza informativi** per stimare in modo completo ed univoco tutto il sistema (cioè, sia $G_0(z)$ che $H_0(z)$)

Identificabilità sperimentale dei modelli ARX

Investighiamo il problema di **identificabilità sperimentale**, considerando per semplicità l'identificazione di un modello $\text{ARX}(n_a, n_b, 1)$, avendo N dati $\{u(1), \dots, u(N)\}, \{y(1), \dots, y(N)\}$

$$\mathcal{M}(\boldsymbol{\theta}): y(t) = \frac{B(z, \boldsymbol{\theta})}{A(z, \boldsymbol{\theta})} u(t-1) + \frac{1}{A(z, \boldsymbol{\theta})} e(t), \quad e(t) \sim \text{WN}(0, \lambda^2)$$

- $B(z) = b_0 + b_1 z^{-1} + \dots + b_{n_b} z^{-n_b}$
- $A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_{n_a} z^{-n_a}$

Sappiamo che la stima può essere ottenuta tramite il metodo dei minimi quadrati

$$\hat{\boldsymbol{\theta}}_N = \left[\sum_{t=1}^N \underset{d \times 1}{\boldsymbol{\varphi}(t)} \underset{1 \times d}{\boldsymbol{\varphi}^\top(t)} \right]^{-1} \cdot \left[\sum_{t=1}^N \underset{d \times 1}{\boldsymbol{\varphi}(t)} y(t) \right]$$

Identificabilità sperimentale dei modelli ARX

Problema di identificabilità: quando $\hat{\theta}_N$ **esiste** ed è **unico**?



quando $\sum_{t=1}^N \underset{d \times 1}{\boldsymbol{\varphi}(t)} \underset{1 \times d}{\boldsymbol{\varphi}^\top(t)}$ è **invertibile?**

Definiamo:

$$\underset{d \times d}{S(N)} = \sum_{t=1}^N \underset{d \times 1}{\boldsymbol{\varphi}(t)} \underset{1 \times d}{\boldsymbol{\varphi}^\top(t)} \quad \Rightarrow \quad \underset{d \times 1}{\hat{\boldsymbol{\theta}}_N} = \underset{d \times d}{S(N)}^{-1} \cdot \left[\sum_{t=1}^N \underset{d \times 1}{\boldsymbol{\varphi}(t)} y(t) \right]$$

$$\underset{d \times d}{R(N)} = \frac{1}{N} S(N) \quad \Rightarrow \quad \underset{d \times 1}{\hat{\boldsymbol{\theta}}_N} = \underset{d \times d}{R(N)}^{-1} \cdot \left[\frac{1}{N} \sum_{t=1}^N \underset{d \times 1}{\boldsymbol{\varphi}(t)} y(t) \right]$$

Identificabilità sperimentale dei modelli ARX

Le matrici $S(N)$ e $R(N)$ sono **semidefinite positive** in quanto prodotto di un vettore per sé stesso. Affinché $\hat{\theta}_N$ **esista** ed sia **unico**, è però necessario che $S(N) \succ 0$ o $R(N) \succ 0$, cioè che

$$\det_{d \times d}(R(N)) > 0$$

Analizziamo la matrice $R(N)$ per $N \rightarrow +\infty$. Consideriamo come punto di partenza un modello ARX(1, 0, 1), ovvero con $n_a = 1, n_b = 0, k = 1$

$$y(t) = a_1 y(t-1) + b_0 u(t-1) + e(t), \quad e(t) \sim \text{WN}(0, \lambda^2)$$

dove $y(t)$ e $u(t)$ sono **pss ergodici a media nulla**

Identificabilità sperimentale dei modelli ARX

$$y(t) = a_1 y(t-1) + b_0 u(t-1) + e(t), \quad e(t) \sim \text{WN}(0, \lambda^2) \quad \boldsymbol{\theta} = \begin{bmatrix} a \\ b \end{bmatrix}_{2 \times 1} \quad \boldsymbol{\varphi}(t) = \begin{bmatrix} y(t-1) \\ u(t-1) \end{bmatrix}_{2 \times 1}$$

$$\boldsymbol{\varphi}(t) \boldsymbol{\varphi}^T(t) = \begin{bmatrix} y(t-1)^2 & y(t-1)u(t-1) \\ u(t-1)y(t-1) & u(t-1)^2 \end{bmatrix} \quad S(N) = \sum_{t=1}^N \boldsymbol{\varphi}(t) \boldsymbol{\varphi}^T(t)_{2 \times 2}$$

$$R(N) = \frac{S(N)}{N} = \begin{bmatrix} \frac{1}{N} \sum_{t=1}^N y(t-1)^2 & \frac{1}{N} \sum_{t=1}^N y(t-1)u(t-1) \\ \frac{1}{N} \sum_{t=1}^N u(t-1)y(t-1) & \frac{1}{N} \sum_{t=1}^N u(t-1)^2 \end{bmatrix}$$

Notiamo che $R(N)$ contiene «somme temporali»

Identificabilità sperimentale dei modelli ARX

Grazie all'**ipotesi di ergodicità**, abbiamo che $R(N) \xrightarrow[N \rightarrow +\infty]{} \bar{R}$ Anche perché $y(\cdot)$ e $u(\cdot)$ hanno media nulla

$$\bar{R} = \begin{bmatrix} \gamma_{yy}(0) & \gamma_{yu}(0) \\ \gamma_{uy}(0) & \gamma_{uu}(0) \end{bmatrix}$$

2×2

La matrice \bar{R} è la **matrice di autocovarianze** del processo congiunto $\{y(t), u(t)\}$

Idea: trovare le condizioni per cui \bar{R} è **invertibile**. Quando queste condizioni valgono, allora possiamo supporre con ragionevole certezza che, per N **grande**, anche $R(N)$ è **invertibile**

Identificabilità sperimentale dei modelli ARX

In generale, per un generico modello $ARX(n_a, n_b, 1)$ abbiamo che la matrice \bar{R} può essere divisa in quattro sotto-matrici

$$\bar{R} = \begin{bmatrix} \bar{R}_{yy} & \bar{R}_{yu} \\ \bar{R}_{uy} & \bar{R}_{uu} \end{bmatrix}$$

$(n_a + n_b + 1) \times (n_a + n_b + 1)$

$n_a \times n_a$ $n_a \times (n_b + 1)$
 $(n_b + 1) \times n_a$ $(n_b + 1) \times (n_b + 1)$

$$\bar{R}_{yy} = \begin{bmatrix} \gamma_{yy}(0) & \gamma_{yy}(1) & \gamma_{yy}(2) & \cdots & \cdots & \gamma_{yy}(n_a - 1) \\ \gamma_{yy}(1) & \gamma_{yy}(0) & \gamma_{yy}(1) & \gamma_{yy}(2) & \cdots & \gamma_{yy}(n_a - 2) \\ \gamma_{yy}(2) & \gamma_{yy}(1) & \gamma_{yy}(0) & \gamma_{yy}(1) & \cdots & \gamma_{yy}(n_a - 3) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{yy}(n_a - 1) & \cdots & \cdots & \cdots & \cdots & \gamma_{yy}(0) \end{bmatrix}$$

$n_a \times n_a$

- Matrice autocovarianza di ordine $n_a - 1$ di $y(t)$
- Struttura Toeplitz
- Dimensioni $n_a \times n_a$

Identificabilità sperimentale dei modelli ARX

$$\bar{R}_{uu} = \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) & \gamma_{uu}(2) & \cdots & \cdots & \gamma_{uu}(n_b) \\ \gamma_{uu}(1) & \gamma_{uu}(0) & \gamma_{uu}(1) & \gamma_{uu}(2) & \cdots & \gamma_{uu}(n_b - 1) \\ \gamma_{uu}(2) & \gamma_{uu}(1) & \gamma_{uu}(0) & \gamma_{uu}(1) & \cdots & \gamma_{uu}(n_b - 2) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{uu}(n_b) & \cdots & \cdots & \cdots & \cdots & \gamma_{uu}(0) \end{bmatrix}$$

$(n_b + 1) \times (n_b + 1)$

- Matrice autocovarianza di ordine n_b di $u(t)$
- Struttura Toeplitz
- Dimensioni $(n_b + 1) \times (n_b + 1)$

$$\bar{R}_{yu} = \begin{bmatrix} \gamma_{yu}(0) & \gamma_{yu}(1) & \gamma_{yu}(2) & \cdots & \cdots & \gamma_{yu}(n_a - 1) \\ \gamma_{yu}(1) & \gamma_{yu}(0) & \gamma_{yu}(1) & \gamma_{yu}(2) & \cdots & \gamma_{yu}(n_a - 2) \\ \gamma_{yu}(2) & \gamma_{yu}(1) & \gamma_{yu}(0) & \gamma_{yu}(1) & \cdots & \gamma_{yu}(n_a - 3) \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \gamma_{yu}(n_a - 1) & \cdots & \cdots & \cdots & \cdots & \gamma_{yu}(0) \end{bmatrix}$$

$n_a \times (n_b + 1)$

- Matrice autocovarianza tra $y(\cdot)$ e $u(\cdot)$
- $\bar{R}_{yu} = \bar{R}_{uy}^T$
- Dimensioni $n_a \times (n_b + 1)$

Cerchiamo una **condizione per l'invertibilità** di \bar{R}

Identificabilità sperimentale dei modelli ARX

Lemma di Schur

Data una matrice M nella forma $M = \left[\begin{array}{c|c} F & K \\ \hline K^\top & H \end{array} \right]$, con F e K simmetriche. Condizione **necessaria e sufficiente** per l'invertibilità di M è che valgano:

- $H > 0$
- $F - KH^{-1}K^\top > 0$

Ricordando che

$$\bar{R} = \left[\begin{array}{c|c} \bar{R}_{yy} & \bar{R}_{yu} \\ \hline \bar{R}_{uy} & \bar{R}_{uu} \end{array} \right]$$



Condizione **necessaria** per l'invertibilità di \bar{R} è che $\bar{R}_{uu} > 0$

Identificabilità sperimentale dei modelli ARX

La condizione (**solo necessaria**) sulla matrice \bar{R}_{uu} è interessante perché riguarda **solo il segnale di ingresso** $u(t)$, che tipicamente **progettiamo noi!**

Possiamo quindi tenere conto di questa condizione in fase di progettazione dell'esperimento, e **scegliere il segnale di eccitazione** più opportuno al fine di ottenere **dati informativi**

Persistente eccitazione

Definizione (Persistente eccitazione)

Definiamo la matrice $\bar{R}_{uu}^{(i)}$ di autocovarianza di $u(t)$ di ordine i come

$$\bar{R}_{uu}^{(i)} = \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) & \cdots & \gamma_{uu}(i-1) \\ \gamma_{uu}(1) & \gamma_{uu}(0) & \cdots & \gamma_{uu}(i-2) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{uu}(i-1) & \gamma_{uu}(i-2) & \cdots & \gamma_{uu}(0) \end{bmatrix}$$

$i \times i$

Il segnale $u(t)$ è detto **persistentemente eccitante di ordine n** se:

- $\bar{R}_{uu}^{(1)} > 0, \bar{R}_{uu}^{(2)} > 0, \dots, \bar{R}_{uu}^{(n)} > 0$
- $\bar{R}_{uu}^{(n+1)} \geq 0, \bar{R}_{uu}^{(n+2)} \geq 0, \dots \geq 0$

Ovvero se n è il massimo ordine per cui $\bar{R}_{uu}^{(i)}$ è invertibile

Persistente eccitazione

Possiamo quindi dire che **condizione necessaria** per l'identificabilità sperimentale di un modello $ARX(n_a, n_b, 1)$ è che il segnale $u(t)$, usato per produrre i dati, sia **«persistentemente eccitante» di ordine pari ad almeno $n_b + 1$** (infatti, \bar{R}_{uu} ha dimensione $(n_b + 1) \times (n_b + 1)$)

Tale conclusione si può generalizzare nel modo seguente. Supponendo che $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, definiamo il numero di parametri di $G(z, \boldsymbol{\theta})$ come n_g . Allora, la soluzione del problema di identificazione $\bar{\boldsymbol{\theta}} = \arg \min_{\boldsymbol{\theta}} \mathbb{E}[\varepsilon_1(t, \boldsymbol{\theta})^2]$ ha un'**unica soluzione** $\bar{\boldsymbol{\theta}} = \boldsymbol{\theta}^0$ se il segnale $u(t)$ che genera i dati è **persistentemente eccitante di ordine $\geq n_g$**

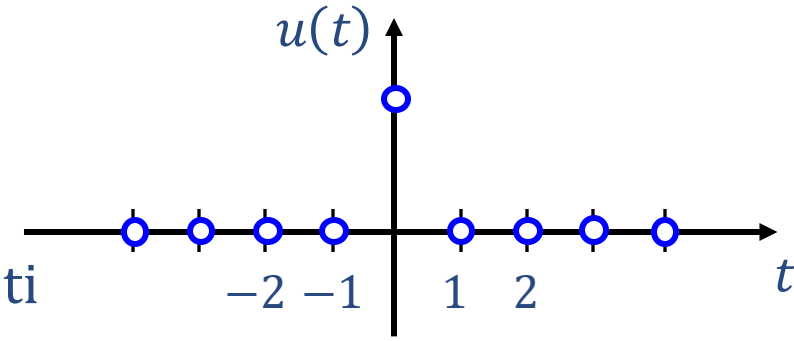
Persistente eccitazione

Osservazioni

- Se un segnale $u(t)$ è persistentemente eccitante di ordine n , allora è anche persistentemente eccitante di ordine $n - 1$
- Ribadiamo che la condizione vista è **solamente necessaria**: anche se $\bar{R}_{uu} \succ 0$, la \bar{R} potrebbe comunque **non essere invertibile** per ragioni di **non identificabilità strutturale**, per le quali il minimo di $\bar{J}(\theta)$ non è unico
- Il concetto di persistente eccitazione che abbiamo visto è stato esemplificato per la stima di modelli ARX, ma avere **un segnale eccitante è importante in ogni caso** si voglia identificare un modello dinamico

Esempio: ingresso impulsivo

Consideriamo come ingresso $u(t) = \text{imp}(t) = \begin{cases} 1 & \text{se } t = 0 \\ 0 & \text{se altrimenti} \end{cases}$



La matrice $R_{uu}(N)$ assume la forma seguente:

$$R_{uu}(N) = \frac{1}{N} \begin{bmatrix} \sum_{t=1}^N u(t-1)^2 & \sum_{t=1}^N u(t)u(t-1) & \cdots & \sum_{t=1}^N u(t)u(t-n_b) \\ \sum_{t=1}^N u(t)u(t-1) & \sum_{t=1}^N u(t-1)^2 & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \sum_{t=1}^N u(t)u(t-n_b) & \cdots & \cdots & \sum_{t=1}^N u(t-1)^2 \end{bmatrix} = \begin{bmatrix} \frac{1}{N} & 0 & \cdots & 0 \\ 0 & \frac{1}{N} & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \frac{1}{N} \end{bmatrix}$$

Esempio: ingresso impulsivo

$$\text{Dato che } R_{uu}(N) = \begin{matrix} & \begin{bmatrix} \frac{1}{N} & 0 & \cdots & 0 \\ 0 & \frac{1}{N} & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & \frac{1}{N} \end{bmatrix} \\ \begin{matrix} (n_b + 1) \times (n_b + 1) \end{matrix} & \end{matrix} \quad \text{allora} \quad \lim_{N \rightarrow \infty} R_{uu}(N) = \bar{R}_{uu} = \begin{matrix} & \begin{bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & \cdots \\ \cdots & \cdots & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \end{bmatrix} \\ \begin{matrix} (n_b + 1) \times (n_b + 1) \end{matrix} & \end{matrix}$$

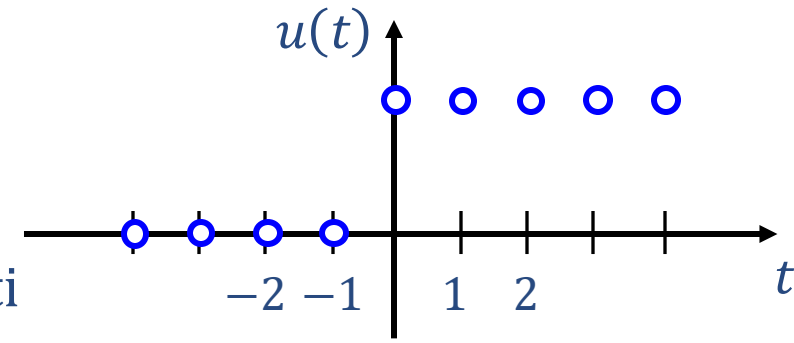
Per cui l'impulso **non è persistentemente eccitante di nessun ordine**

Questo non vuol dire che la risposta all'impulso non può essere usata per stimare modelli, anzi ci dà molte informazioni sul sistema $G_0(z)$. La nozione di persistente eccitazione serve però a garantire **stime consistenti** anche in presenza di **disturbi**

In questo caso, il problema è che «osservo solo una volta» ogni valore della risposta all'impulso, e quindi se quel valore è corrotto da rumore «non posso confrontarlo» con altri valori analoghi, e me lo devo tenere sporcato dal rumore

Esempio: ingresso scalino

Consideriamo come ingresso $u(t) = \text{sca}(t) = \begin{cases} 1 & \text{se } t \geq 0 \\ 0 & \text{se altrimenti} \end{cases}$



La matrice $R_{uu}(N)$ assume la forma seguente:

$$R_{uu}(N) = \frac{1}{N} \begin{matrix} (n_b + 1) \times (n_b + 1) & \begin{bmatrix} \sum_{t=1}^N u(t-1)^2 & \sum_{t=1}^N u(t)u(t-1) & \dots & \sum_{t=1}^N u(t)u(t-n_b) \\ \sum_{t=1}^N u(t)u(t-1) & \sum_{t=1}^N u(t-1)^2 & \dots & \dots \\ \dots & \dots & \dots & \dots \\ \sum_{t=1}^N u(t)u(t-n_b) & \dots & \dots & \sum_{t=1}^N u(t-1)^2 \end{bmatrix} \end{matrix} \rightarrow \begin{bmatrix} 1 & 1 & \dots & 1 \\ 1 & 1 & \dots & \dots \\ \dots & \dots & \dots & 1 \\ 1 & \dots & \dots & 1 \end{bmatrix}$$

**È persistentemente
eccitante di ordine 1**

Esempio: ingresso impulso e scalino

Come informazione ulteriore, notiamo che **impulso** e **scalino** **caratterizzano completamente** un sistema dinamico lineare (a meno della sua condizione iniziale), grazie anche al periodo di **transitorio** della risposta a questi segnali

È utile ricordare che il framework PEM che stiamo studiando qui **non è adatto all'utilizzo di segnali con transitori** (infatti richiede che essi siano realizzazioni di pss)

Altri metodi, come quelli di **identificazione a sottospazi**, possono utilizzare segnali con transitorio

Esempio: ingresso rumore bianco

Consideriamo ora come ingresso $u(t) = \text{WN}(0, \lambda^2)$. In questo caso, la matrice $\bar{R}_{uu}^{(i)}$ risulta essere:

$$\bar{R}_{uu}^{(i)} = \begin{matrix} i \times i \\ \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) & \cdots & \gamma_{uu}(i-1) \\ \gamma_{uu}(1) & \gamma_{uu}(0) & \cdots & \gamma_{uu}(i-2) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{uu}(i-1) & \gamma_{uu}(i-2) & \cdots & \gamma_{uu}(0) \end{bmatrix} \end{matrix} = \begin{bmatrix} \lambda^2 & 0 & 0 & \cdots & 0 \\ 0 & \lambda^2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda^2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & \lambda^2 \end{bmatrix} = \lambda^2 \cdot I_i \succ 0$$

Quindi, un **white noise** è un segnale **persistentemente eccitante di ordine** ∞ . Se usiamo un WN per eccitare il sistema, i dati generati saranno **molto informativi**. Questo perché eccitiamo **tutte le frequenze del sistema** (con la stessa «energia»)

Persistente eccitazione nel dominio delle frequenze

La nozione di **persistente eccitazione** di un segnale di ingresso $u(t)$ può essere interpretata anche nel **dominio delle frequenze**

A tale fine, consideriamo un processo stocastico stazionario generato nel modo seguente:

$$y(t) = c_0 u(t) + c_1 u(t - 1) + c_2 u(t - 2) + \cdots + c_n u(t - n)$$

dove $u(t)$ è un **processo stazionario** a **media nulla** e con funzione di autocovarianza $\gamma_{uu}(\tau)$ ($u(t)$ non è per forza un WN)

Questo processo è anche chiamato **Generalized Moving Average (GMA)** data la sua somiglianza al processo MA

Persistente eccitazione nel dominio delle frequenze

Calcoliamo la varianza di $y(t)$

$$\begin{aligned}\text{Var}[y(t)] = \mathbb{E}[y(t)^2] &= c_0^2 \mathbb{E}[u(t)^2] + c_1^2 \mathbb{E}[u(t-1)^2] + \dots + c_n^2 \mathbb{E}[u(t)^2] + \\ &+ c_0 c_1 \mathbb{E}[u(t)u(t-1)] + c_1 c_2 \mathbb{E}[u(t-1)u(t-2)] + \dots \\ &+ c_0 c_2 \mathbb{E}[u(t)u(t-2)] + c_1 c_3 \mathbb{E}[u(t-1)u(t-2)] + \dots\end{aligned}$$

Per cui

$$\text{Var}[y(t)] = (c_0^2 + c_1^2 + \dots + c_n^2)\gamma_{uu}(0) + (c_0 c_1 + c_1 c_2 + \dots)\gamma_{uu}(1) + (c_0 c_2 + c_1 c_3 + \dots)\gamma_{uu}(2) + \dots$$

Persistente eccitazione nel dominio delle frequenze

Definiamo ora la matrice $\bar{R}_{uu}^{(i)}$ di ordine $i = n + 1$

$$\bar{R}_{uu}^{(i)} = \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) & \cdots & \gamma_{uu}(i-1) \\ \gamma_{uu}(1) & \gamma_{uu}(0) & \cdots & \gamma_{uu}(i-2) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{uu}(i-1) & \cdots & \cdots & \gamma_{uu}(0) \end{bmatrix}_{i \times i} \Rightarrow \bar{R}_{uu}^{(n+1)} = \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) & \cdots & \gamma_{uu}(n) \\ \gamma_{uu}(1) & \gamma_{uu}(0) & \cdots & \gamma_{uu}(n-1) \\ \cdots & \cdots & \cdots & \cdots \\ \gamma_{uu}(n) & \gamma_{uu}(n-1) & \cdots & \gamma_{uu}(0) \end{bmatrix}_{(n+1) \times (n+1)}$$

Ponendo $\mathbf{c} = [c_0 \ c_1 \ c_2 \ \dots \ c_n]^\top \in \mathbb{R}^{(n+1) \times 1}$, possiamo scrivere che

$$\text{Var}[y(t)] = \underset{(n+1) \times (n+1)}{\mathbf{c}^\top} \cdot \underset{(n+1) \times 1}{\bar{R}_{uu}^{(n+1)}} \cdot \mathbf{c}$$

Esempio: autocovarianza di un processo GMA

Consideriamo il seguente processo GMA di ordine $n = 1$, dove $u(t)$ pss a media nulla

$$y(t) = c_0 u(t) + c_1 u(t - 1)$$

Calcoliamo la varianza di $y(t)$

$$\text{Var}[y(t)] = \mathbb{E}[y(t)^2] = c_0^2 \mathbb{E}[u(t)^2] + c_1^2 \mathbb{E}[u(t-1)^2] + c_0 c_1 \mathbb{E}[u(t)u(t-1)]$$

$$= c_0^2 \gamma_{uu}(0) + c_1^2 \gamma_{uu}(0) + c_0 c_1 \gamma(1) + c_1 c_0 \gamma(1)$$

La matrice di autocovarianze di $u(t)$ è: $\bar{R}_{uu}^{(1+1)} = \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) \\ \gamma_{uu}(1) & \gamma_{uu}(0) \end{bmatrix}$

Esempio: autocovarianza di un processo GMA

Definiamo $\mathbf{c} = [c_0 \ c_1]^\top \in \mathbb{R}^{2 \times 1}$

Per cui

$$\begin{aligned}\text{Var}[y(t)] &= \mathbf{c}^\top \cdot \bar{R}_{uu}^{(1+1)} \cdot \mathbf{c} = [c_0 \ c_1] \cdot \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) \\ \gamma_{uu}(1) & \gamma_{uu}(0) \end{bmatrix} \cdot \begin{bmatrix} c_0 \\ c_1 \end{bmatrix} = \\ &= [c_0 \ c_1] \cdot \left(c_0 \begin{bmatrix} \gamma_{uu}(0) \\ \gamma_{uu}(1) \end{bmatrix} + c_1 \begin{bmatrix} \gamma_{uu}(1) \\ \gamma_{uu}(0) \end{bmatrix} \right) = [c_0 \ c_1] \cdot \begin{bmatrix} c_0 \cdot \gamma_{uu}(0) + c_1 \cdot \gamma_{uu}(1) \\ c_0 \cdot \gamma_{uu}(1) + c_1 \cdot \gamma_{uu}(0) \end{bmatrix}\end{aligned}$$

$$= c_0^2 \gamma_{uu}(0) + c_1^2 \gamma_{uu}(0) + c_0 c_1 \gamma(1) + c_1 c_0 \gamma(1)$$

Persistente eccitazione nel dominio delle frequenze

Consideriamo quindi un generico vettore $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_n]^\top \in \mathbb{R}^{n \times 1}$, $\alpha \neq 0$, ed il segnale $\tilde{u}(t)$ generato filtrando il pss $u(t)$ a media nulla come

$$\tilde{u}(t) = \alpha_1 u(t-1) + \alpha_2 u(t-2) + \dots + \alpha_n u(t-n)$$

La funzione $H_\alpha(z)$ di trasferimento del filtro è

$$H_\alpha(z) = \frac{\alpha_1 z^{n-1} + \dots + \alpha_n}{z^n}$$

Possiamo quindi scrivere

$$\tilde{u}(t) = H_\alpha(z)u(t)$$

Persistente eccitazione nel dominio delle frequenze

Essendo $\tilde{u}(t)$ un processo GMA, la sua varianza è

$$\text{Var}[\tilde{u}(t)] = \boldsymbol{\alpha}^\top \cdot \bar{R}_{uu}^{(n)} \cdot \boldsymbol{\alpha}$$

È però possibile esprimere la varianza di $\tilde{u}(t)$ anche come

$$\text{Var}[\tilde{u}(t)] = \frac{1}{2\pi} \int_{-\pi}^{+\pi} \Gamma_{\tilde{u}\tilde{u}}(\omega) d\omega = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |H_\alpha(e^{j\omega})|^2 \cdot \Gamma_{uu}(\omega) d\omega$$

Da cui segue che

$$\boldsymbol{\alpha}^\top \cdot \bar{R}_{uu}^{(n)} \cdot \boldsymbol{\alpha} = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |H_\alpha(e^{j\omega})|^2 \cdot \Gamma_{uu}(\omega) d\omega$$

Persistente eccitazione nel dominio delle frequenze

$$\alpha^\top \cdot \bar{R}_{uu}^{(n)} \cdot \alpha = \frac{1}{2\pi} \int_{-\pi}^{+\pi} |H_\alpha(e^{j\omega})|^2 \cdot \Gamma_{uu}(\omega) d\omega$$

Il fatto che $\bar{R}_{uu}^{(n)} \succ 0$ (e quindi che $u(t)$ sia **persistentemente eccitante** di ordine n) equivale a dire che $\alpha^\top \cdot \bar{R}_{uu}^{(n)} \cdot \alpha > 0$. Quindi $\bar{R}_{uu}^{(n)}$ è **non singolare se e solo se**

$$\alpha^\top \bar{R}_{uu}^{(n)} \alpha = 0 \implies \alpha = 0$$

La precedente implicazione può essere riscritta come

$$\frac{1}{2\pi} \int_{-\pi}^{+\pi} |H_\alpha(e^{j\omega})|^2 \cdot \Gamma_{uu}(\omega) d\omega = 0 \implies H_\alpha(e^{j\omega}) = 0 \quad \forall \omega$$

Persistente eccitazione nel dominio delle frequenze

Quando succede che $\alpha^\top \cdot \bar{R}_{uu}^{(n)} \cdot \alpha = 0$ nonostante sia $\alpha \neq 0$? Notiamo che la funzione di trasferimento $H_\alpha(z)$ può avere al **massimo** $n - 1$ zeri

Quindi, $H_\alpha(z)$ può **annullarsi** se $z_k = e^{j\omega_k}$ è uno dei suoi **zeri**, con $k = 1, \dots, n - 1$

Però, se $\Gamma_{uu}(\omega) \neq 0$ per **almeno** n **pulsazioni distinte nell'intervallo** $\omega \in (-\pi, \pi]$, allora

$$\int_{-\pi}^{+\pi} |H_\alpha(e^{j\omega})| \cdot \Gamma_{uu}(\omega) d\omega \neq 0 \text{ e quindi } \alpha^\top \cdot \bar{R}_{uu}^{(n)} \cdot \alpha > 0$$

Questa interpretazione ci permette di dare una **definizione equivalente** di persistente eccitazione

Persistente eccitazione nel dominio delle frequenze

Definizione (Persistente eccitazione)

Un segnale $u(t)$, con spettro $\Gamma_{uu}(\omega)$, è **persistentemente eccitante** di ordine n se e solo se, per tutti i filtri della forma

$$H_\alpha(z) = \alpha_1 z^{-1} + a_2 z^{-2} + \dots + a_n z^{-n}$$

la relazione

$$|H_\alpha(e^{j\omega})|^2 \cdot \Gamma_{uu}(\omega) = 0 \quad \forall \omega$$

implica che $H_\alpha(e^{j\omega}) = 0 \quad \forall \omega$

Persistente eccitazione nel dominio delle frequenze

Possiamo quindi concludere che:

1. Un **pss** $u(t)$ a **media nulla** (**media non nulla**) è **persistentemente eccitante di ordine n** (**$n + 1$**) se $\Gamma_{uu}(\omega) \neq 0$ in almeno n pulsazioni $\omega \in (-\pi, \pi]$ **distinte**

Esempio: il segnale $u(t) = \sin(\omega_0 t)$ è persistentemente eccitante di ordine $n = 2$ ($\Gamma_{uu}(\omega)$ ha contributi alle pulsazioni $\pm\omega_0$)

2. Un **processo ARMA** è **persistentemente eccitante di ogni ordine**

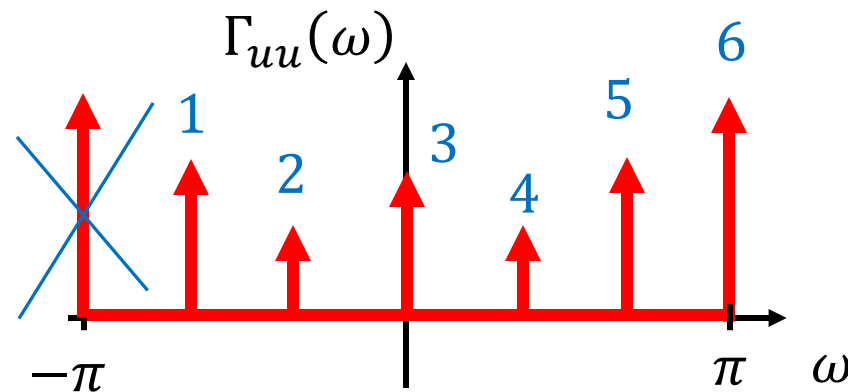
Infatti, $\Gamma_{yy}(\omega_k) = 0$ solo se $z_k = e^{j\omega_k}$ è uno degli zeri di $C(z)$, e vi sono altre infinite frequenze per cui la sua densità spettrale di potenza non si annulla

Persistente eccitazione nel dominio delle frequenze

Considerazioni ulteriori

- Un segnale $u(t)$, **periodico** di periodo T , è **al massimo** (potrebbe essere anche minore) persistentemente eccitante di ordine $n = T$

Esempio: consideriamo un pss completamente predicibile periodico di periodo T , con media non nulla



$\omega = -\pi$ e $\omega = \pi$ sono lo stesso punto e vanno contati una volta sola

Esempio: effetto dell'ingresso sulla stima

Consideriamo il seguente meccanismo di generazione dei dati \mathcal{S}

$$\mathcal{S}: y(t) = \frac{0.103 + 0.181z^{-1}}{1 - 1.991z^{-1} + 2.203z^{-2} - 1.841z^{-3} + 0.894z^{-4}} z^{-3} u(t) + e(t), \quad e(t) \sim \text{WN}(0, 0.5^2)$$

Notiamo che il sistema è del tipo **Output Error** OE(n_b, n_f, k), con $k = 3, n_b = 1, n_f = 4$

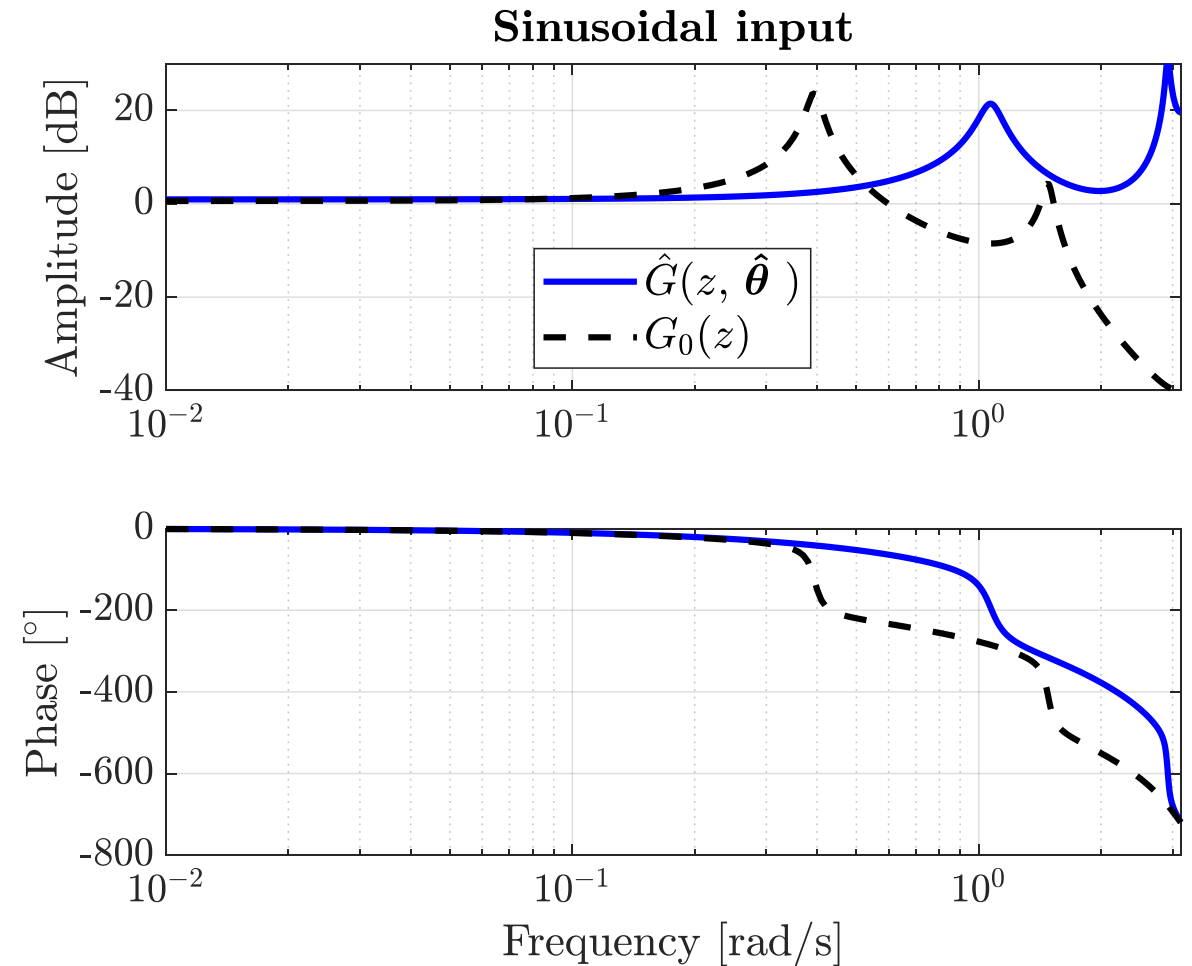
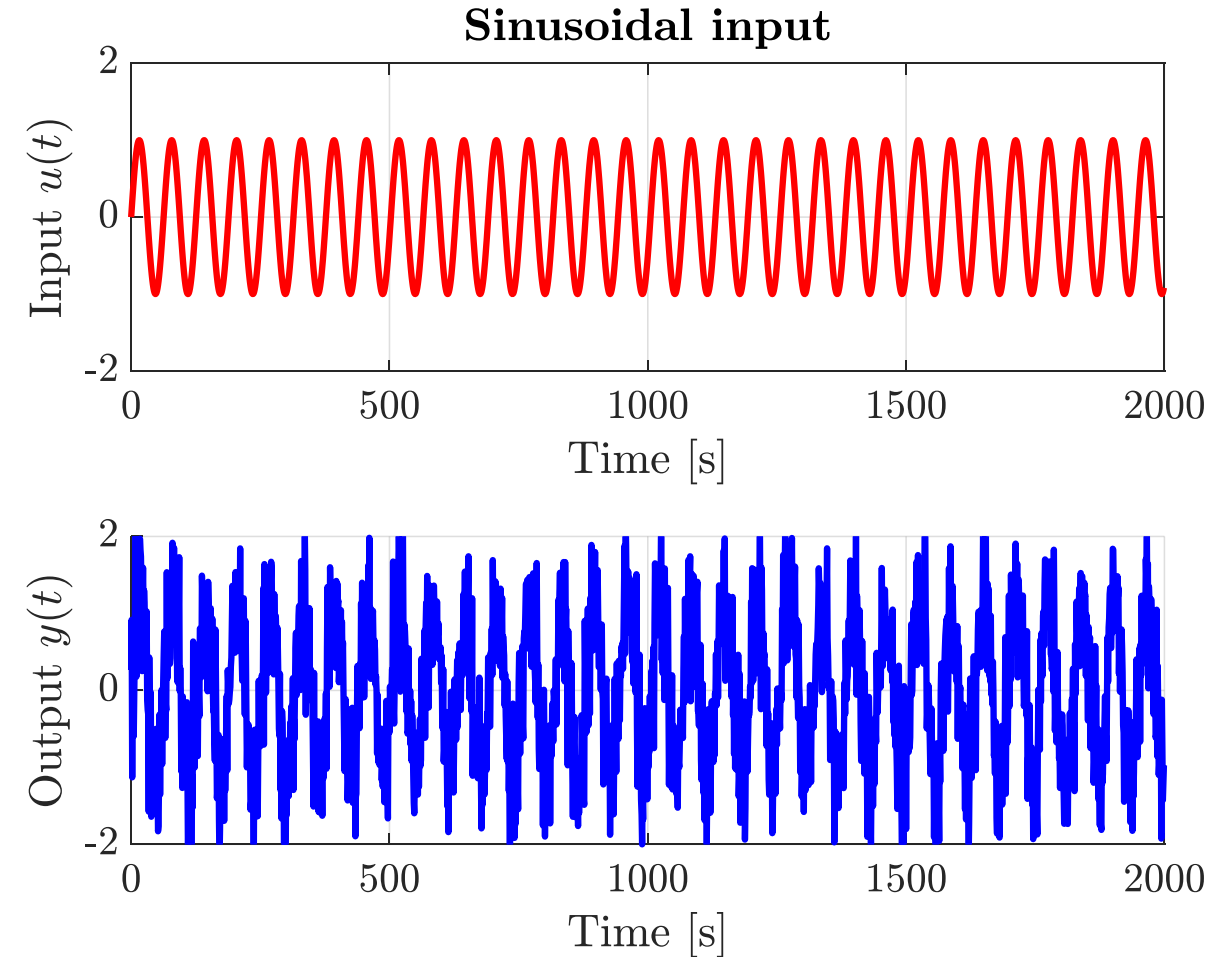
Utilizziamo una famiglia di modelli tali che $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, ovvero

$$\mathcal{M}(\boldsymbol{\theta}): \left\{ G(z, \boldsymbol{\theta}) = \frac{b_0 + b_1 z^{-1}}{1 + f_1 z^{-1} + f_2 z^{-2} + f_3 z^{-3} + f_4 z^{-4}} z^{-3}; \quad H(z, \boldsymbol{\theta}) = 1 \right\}$$

I parametri sono $\boldsymbol{\theta} = [b_0 \ b_1 \ f_1 \ f_2 \ f_3 \ f_4]^T \in \mathbb{R}^{6 \times 1}$. Il numero di parametri di $G(z, \boldsymbol{\theta})$ è $n_g = 6$

Esempio: effetto dell'ingresso sulla stima

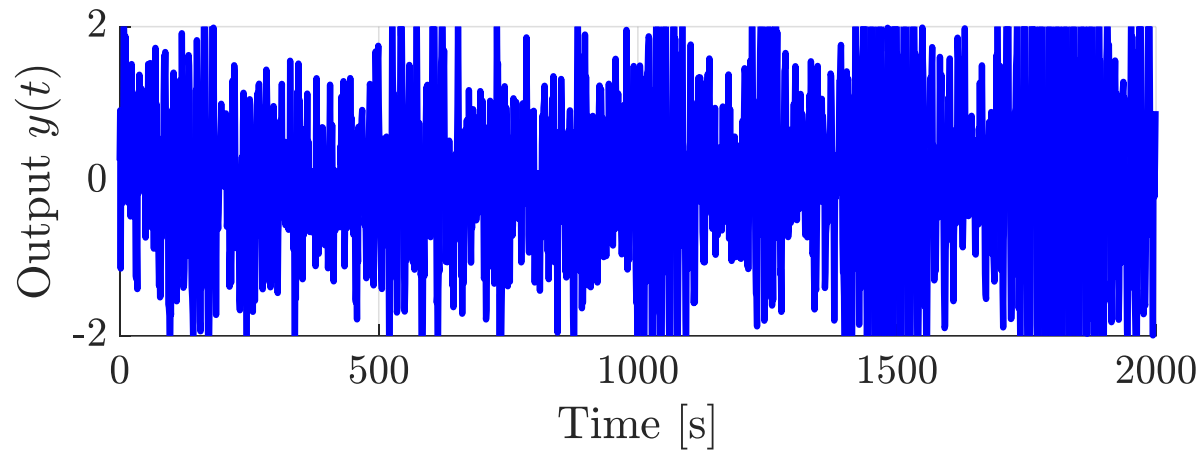
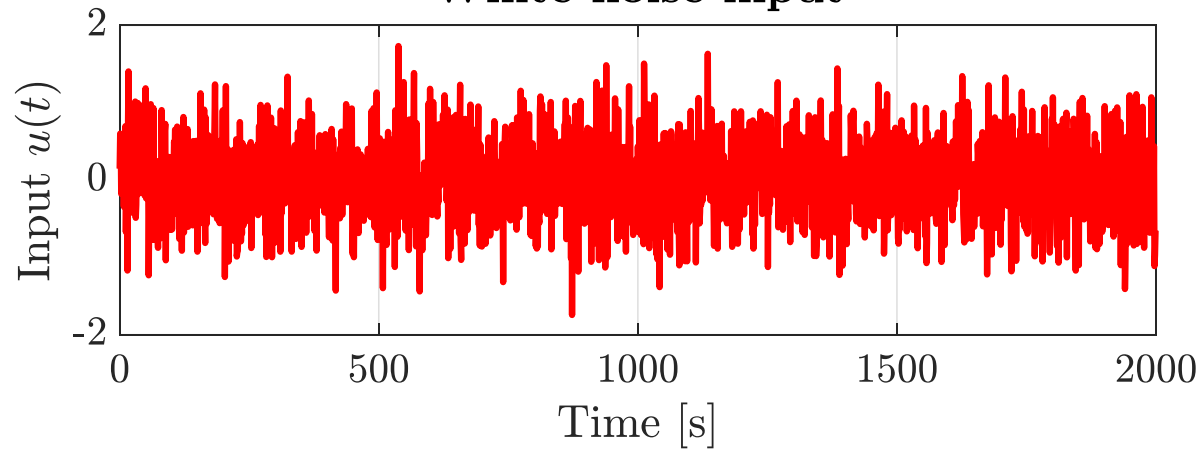
Misuriamo $N = 2000$ dati con ingresso sinusoidale $u_{\sin}(t) = \sin(0.1t)$, p.e. di ordine 2



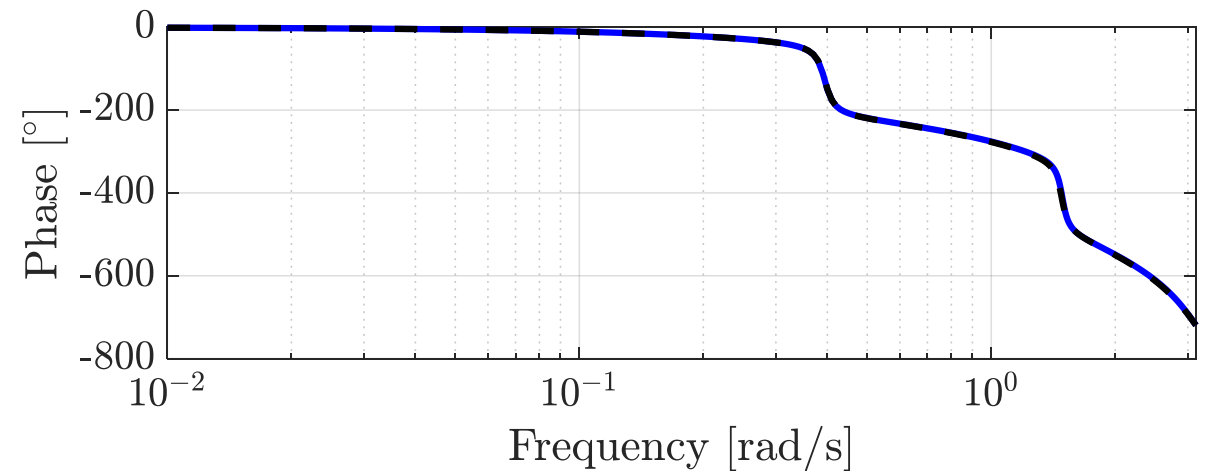
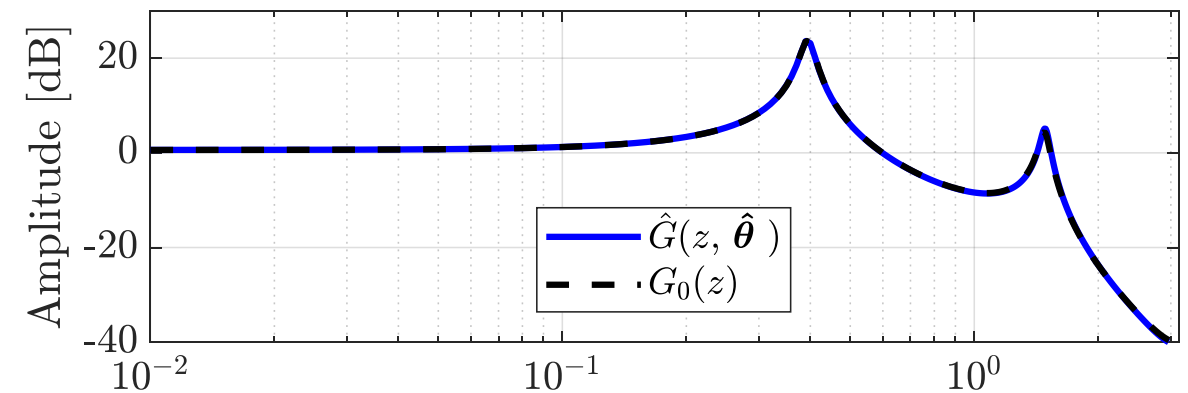
Esempio: effetto dell'ingresso sulla stima

Misuriamo $N = 2000$ dati con ingresso rumore bianco $u_{\text{wn}}(t) \sim \text{WN}(0, 0.5^2)$, p.e. di ordine ∞

White noise input



White noise input



Segnali di eccitazione ulteriori

Abbiamo visto che un **rumore bianco** è un ottimo segnale di eccitazione. Nella pratica, però, è **impossibile generare** un rumore bianco «perfetto»:

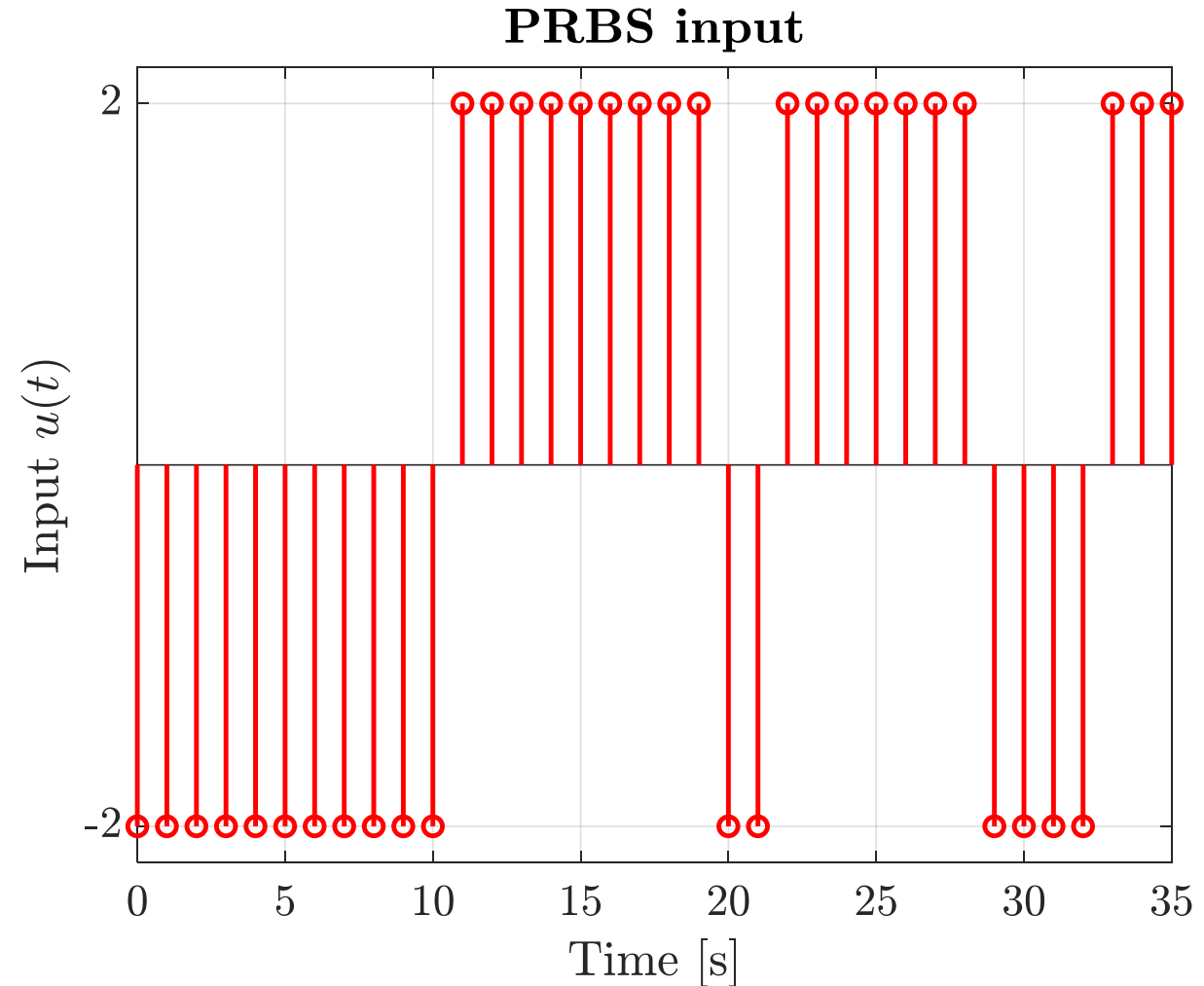
- al più, le sequenze di numeri saranno **pseudo-casuali**, e non casuali
- a causa di limiti dell'elettronica (e.g. capacità parassite), il segnale generato e trasmesso agli attuatori sarà **«filtrato passa-basso»**, per cui **non si avrà** uno spettro «perfettamente piatto». Inoltre, talvolta **non si vuole sollecitare troppo** gli attuatori ad alta frequenza per non rovinarli

Inoltre, **l'ampiezza** del rumore bianco **non è «limitata»**. Talvolta, è necessario garantire che l'attuatore **non saturi** l'ingresso, al fine di **non introdurre nonlinearietà** nell'esperimento e nei dati misurati

Pseudo-Random Binary Signal (PRBS)

Il segnale di tipo PRBS è un segnale **deterministico, periodico**, a tempo discreto, che commuta tra **due livelli**

L'utente deve definire i **due livelli** $[-\bar{u}, +\bar{u}]$, il **periodo** e l'**intervallo di clock** (il numero minimo di intervalli di tempo dopo i quali il segnale può cambiare livello)



Pseudo-Random Binary Signal (PRBS)

Di solito il **periodo** viene posto uguale al numero di dati N che si vuole collezionare, e l'**intervallo di clock** a un tempo di campionamento

I tool per la generazione dei segnali (come Matlab) generano automaticamente dei **PRBS di lunghezza massima**, ovvero dei segnali il cui periodo è $T = 2^{n_{\text{prbs}}} - 1$, dove n_{prbs} è detto **ordine** del PRBS. È poi possibile **ripetere più volte** il segnale con tale periodo

Per cui, se settiamo $T = N$ e N non è esprimibile come $2^{n_{\text{prbs}}} - 1$, il periodo effettivamente generato non sarà di N ma del valore $2^{n_{\text{prbs}}} - 1$ più vicino a N

In matlab

`idinput([N 1 3], 'prbs', Band, Range)` Periodo di lunghezza N ripetuto 3 volte



Pseudo-Random Binary Signal (PRBS)

Tali «**maximum length PRBS**» hanno proprietà interessanti. Infatti, si dimostra come la **funzione di autocovarianza** di un max. length PRBS può essere espressa come

$$\gamma_{uu}(\tau) = \frac{1}{T} \sum_{t=1}^T (u(t) - m_u) (u(t + \tau) - m_u) = \begin{cases} \bar{u}^2 \left(1 - \frac{1}{T}\right) & \text{se } \tau = 0, \pm T, \pm 2T, \dots \\ -\frac{\bar{u}^2}{T} \left(1 + \frac{1}{T}\right) & \text{se altrimenti} \end{cases}$$

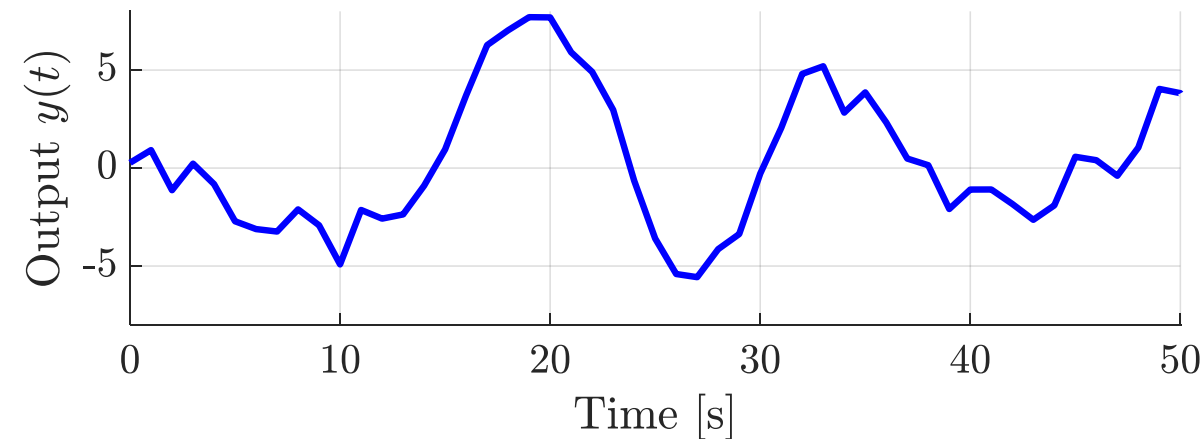
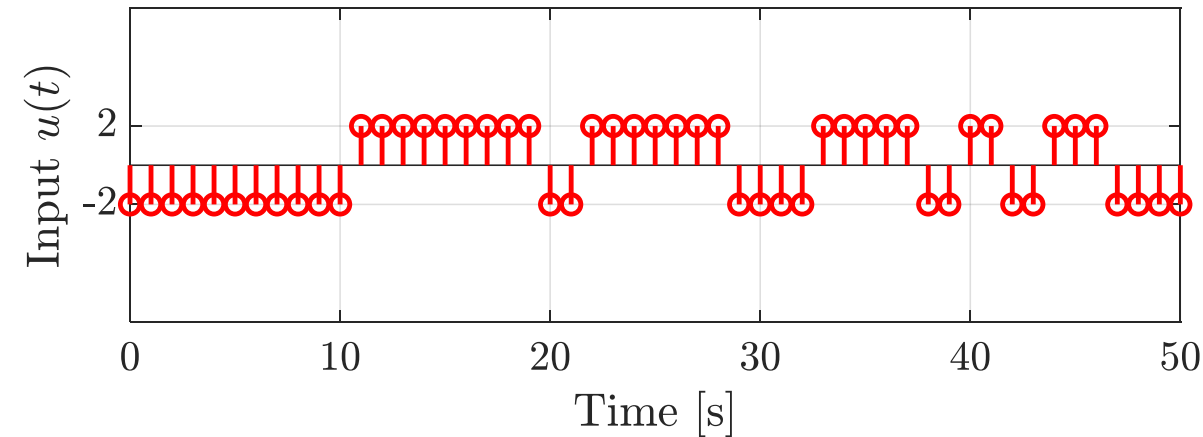
m_u è la media del PRBS, che non è esattamente zero

Notiamo che quando $T \rightarrow \infty$ il PRBS approssima un **rumore bianco**. Il PRBS è **persistentemente eccitante** di ordine T (e non può esserlo di più essendo periodico)

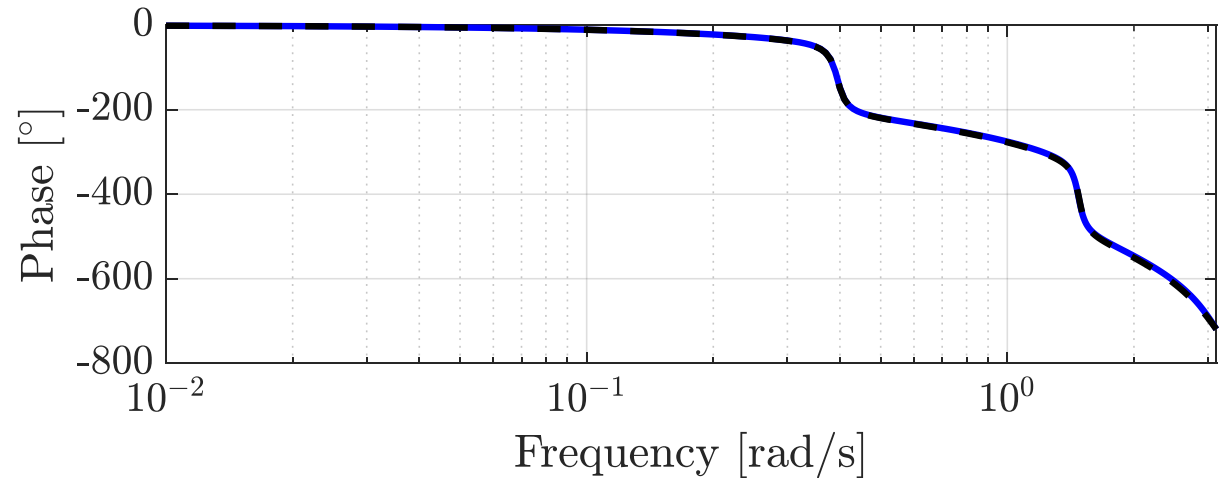
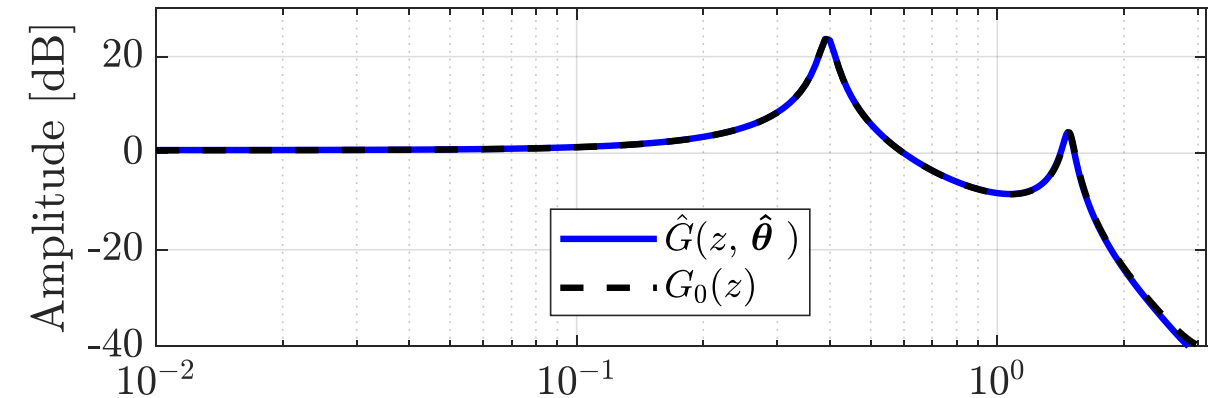
Esempio: effetto dell'ingresso sulla stima

Riprendiamo l'esempio precedente e misuriamo $N = 2000$ dati con **ingresso** PRBS($-2, +2$)

PRBS input



PRBS input



Multiseno

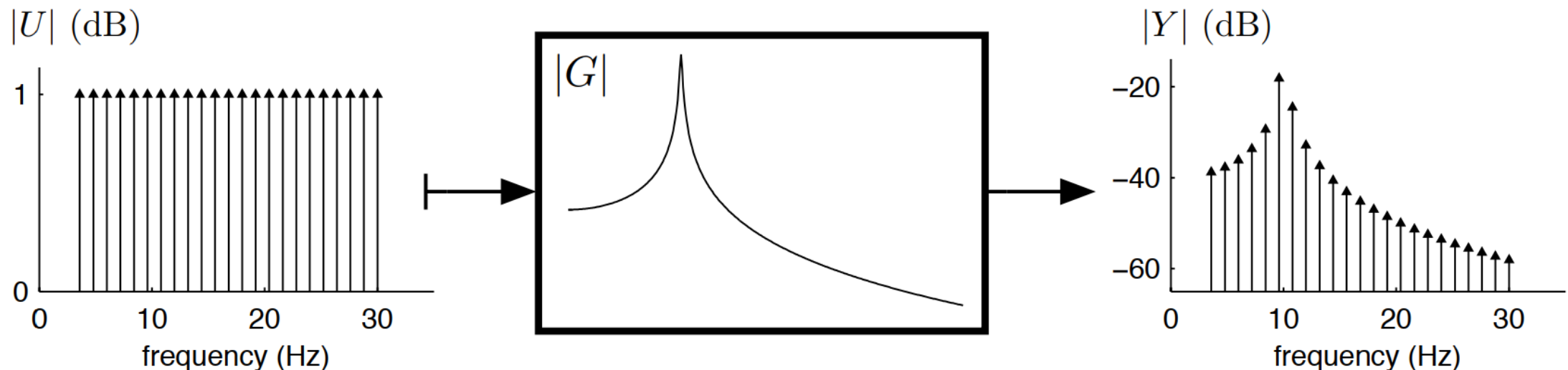
Il segnale **multiseno** è un segnale periodico, definito come una **media pesata di sinusoidi**, con frequenze multiple della risoluzione in frequenza della DFT $f_0 = f_s/N$

$$u(t) = \sum_{k=0}^F A_k \cdot \cos(2\pi \cdot k f_0 \cdot t + \phi_k)$$

- Il **numero F di componenti** in frequenza deve soddisfare il teorema del campionamento
- Gli **sfasamenti ϕ_k** sono in generale scelti in **modo casuale**, e si possono anche ottimizzare per minimizzare il valore di picco del segnale

Multiseno

Molto spesso le **ampiezze** A_k vengono scelte ad un **valore costante nella banda di frequenze di interesse**, e 0 altrove. Per esempio, se $A_k = 1 \ \forall k$, l'ampiezza dello spettro in frequenza dell'uscita $y(t)$ assume la forma dell'ampiezza della funzione di trasferimento alle frequenze eccitate



Multiseno

Quando l'ingresso $u(t)$ da un **multiseno**, anche il segnale di uscita $y(t)$ di un sistema LTI è un **multiseno (dopo un transitorio)**, come conseguenza del principio di sovrapposizione degli effetti

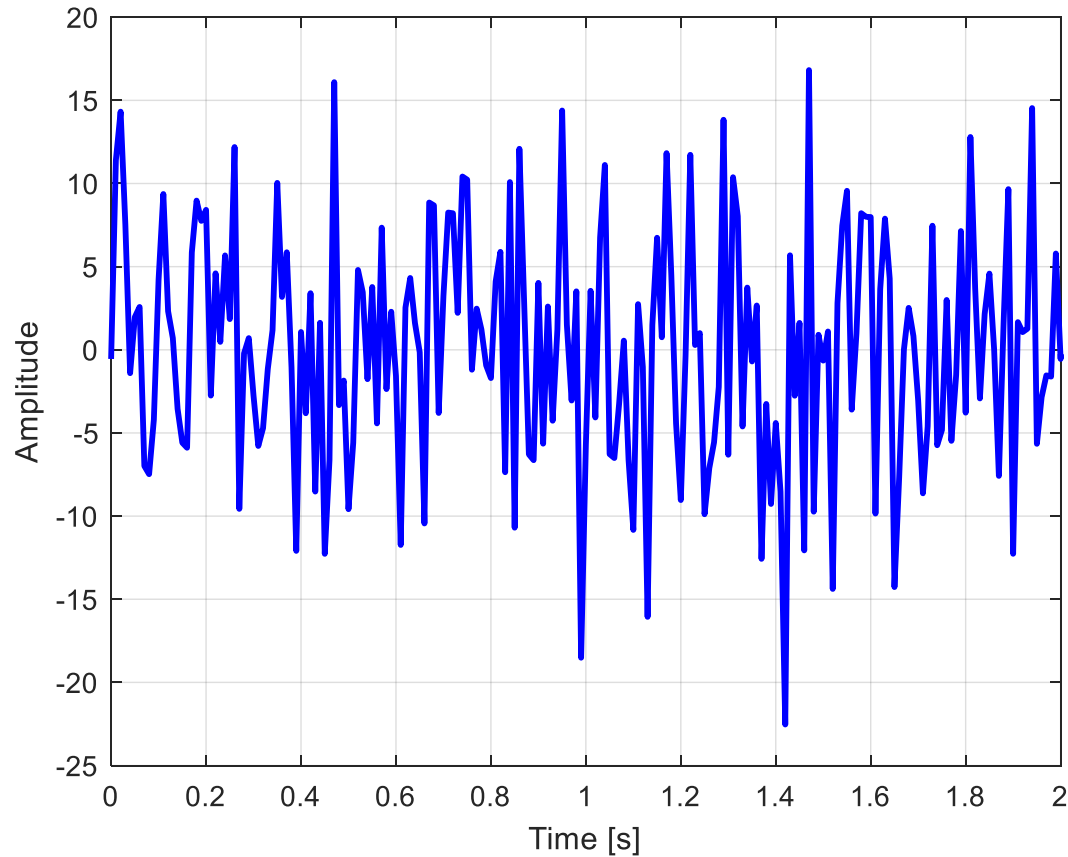
- Di solito si **scartano i primi periodi** del segnale multiseno generato

Quando si progetta un multiseno, si può **fissare la risoluzione in frequenza desiderata** e la **massima frequenza eccitata**, per calcolare automaticamente la lunghezza $N \cdot P$ del segnale, dove N = numero dati per periodo e P = numero di periodi del multiseno

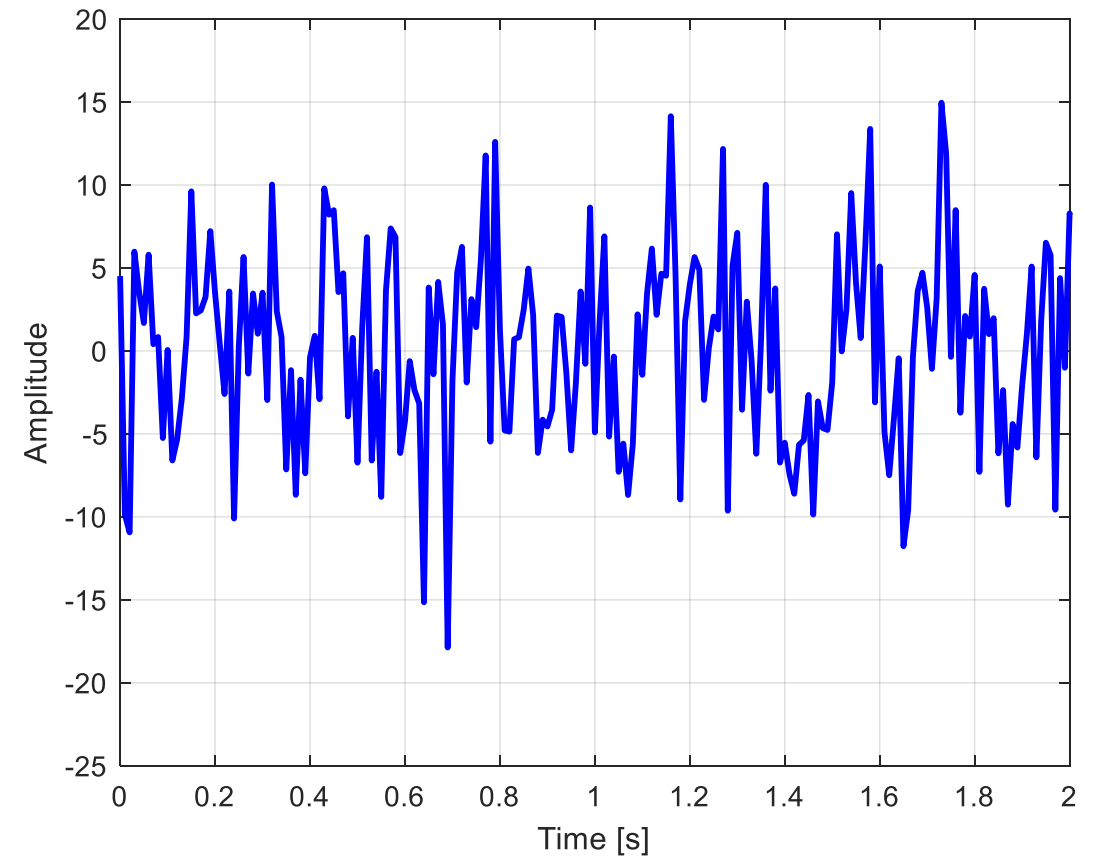
Multiseno

Visivamente, un multiseno è molto simile ad un rumore bianco

Multiseno



White noise

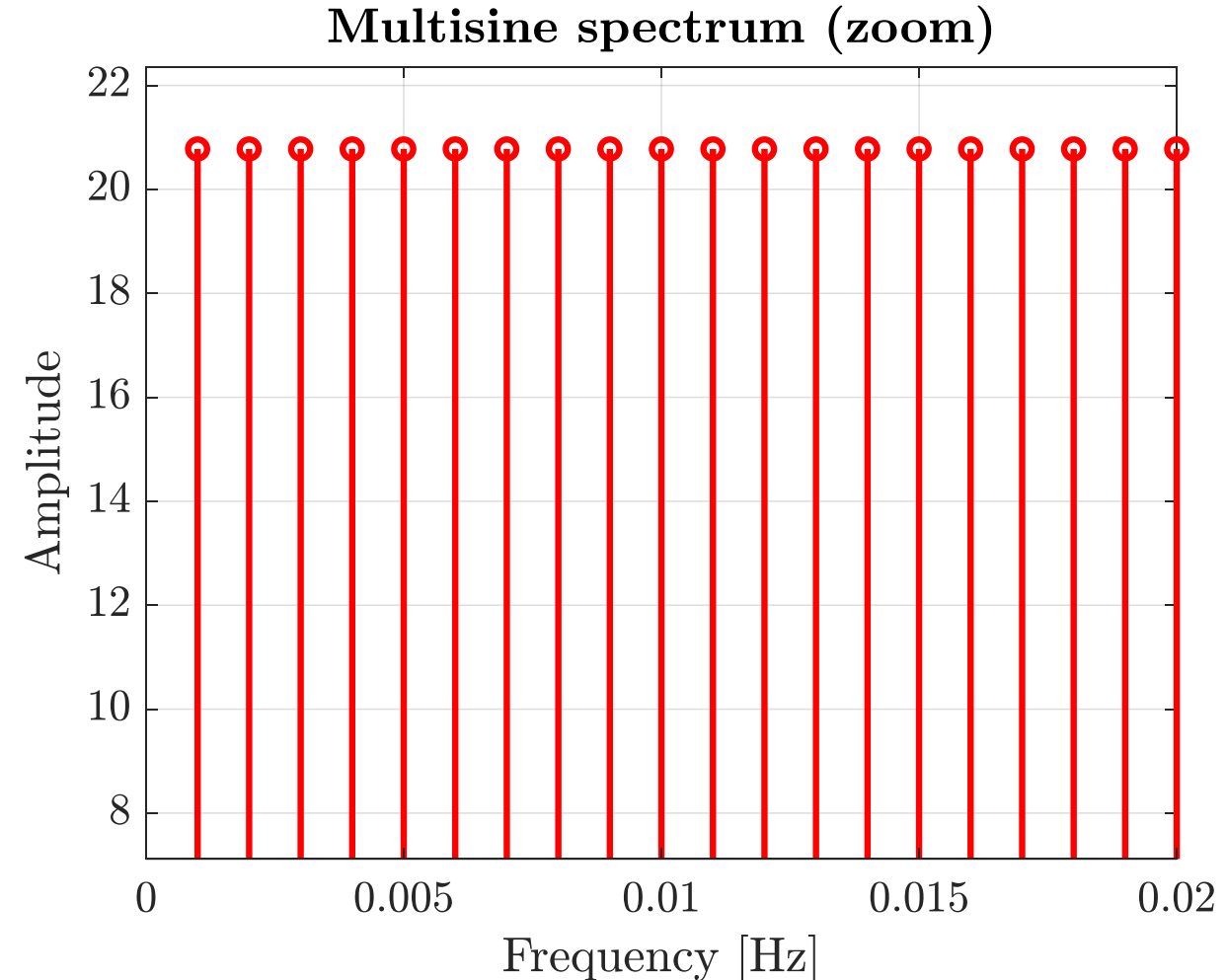


Esempio: multiseno

Progettiamo un multiseno con $N = 1000$ campioni per periodo, campionato a

$$f_s = 1 \text{ Hz}$$

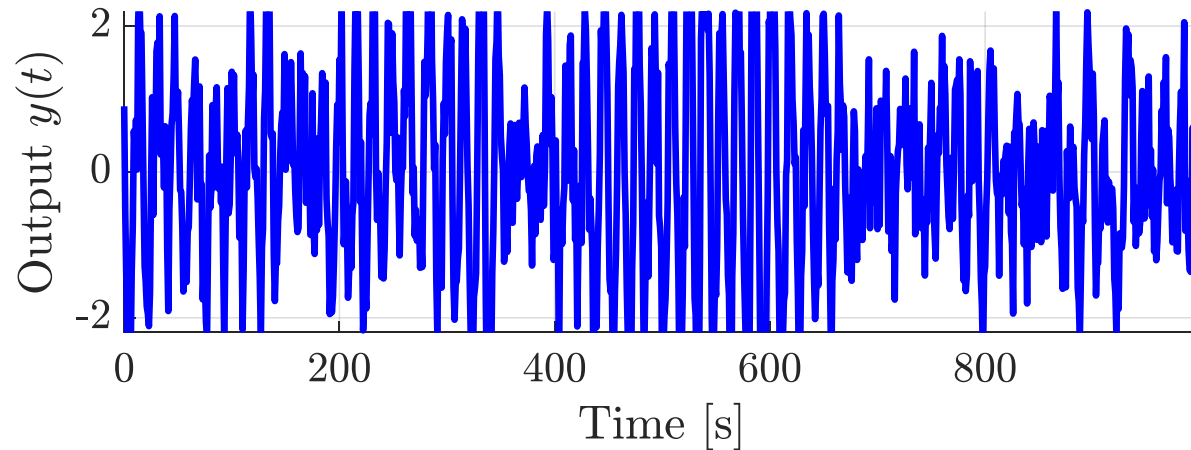
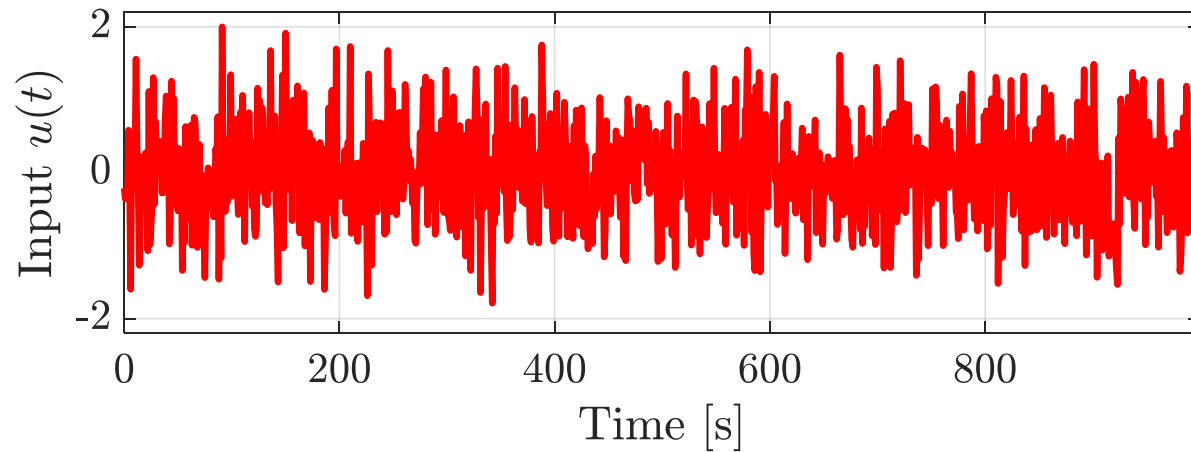
Misuriamo $P = 2$ periodi. Ogni periodo del multiseno è un segnale che eccita le frequenze da $1 \text{ bin} = f_s/N = 10^{-3} \text{ Hz}$ fino a 0.5 Hz , con una risoluzione di $1 \text{ bin} = f_s/N$ (la componente DC non è eccitata)



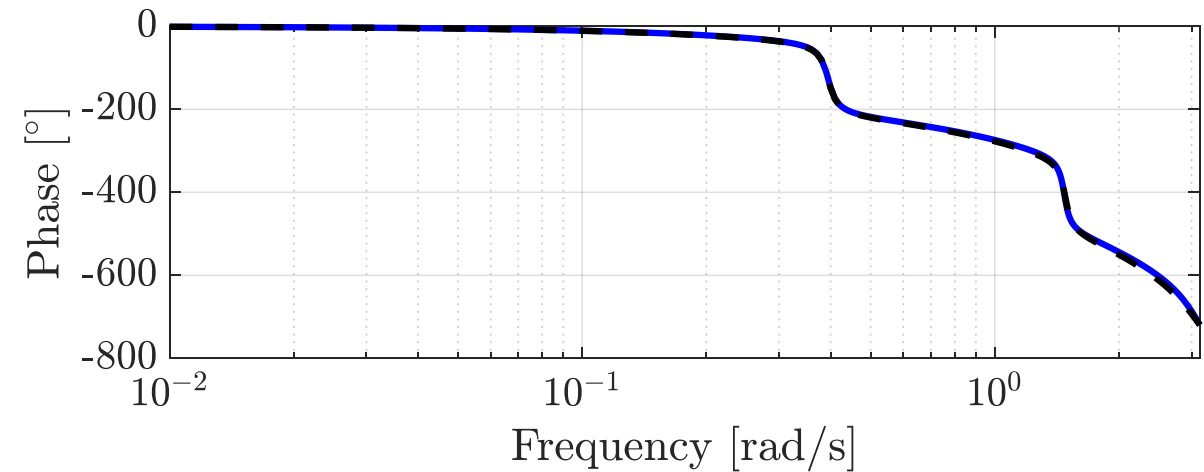
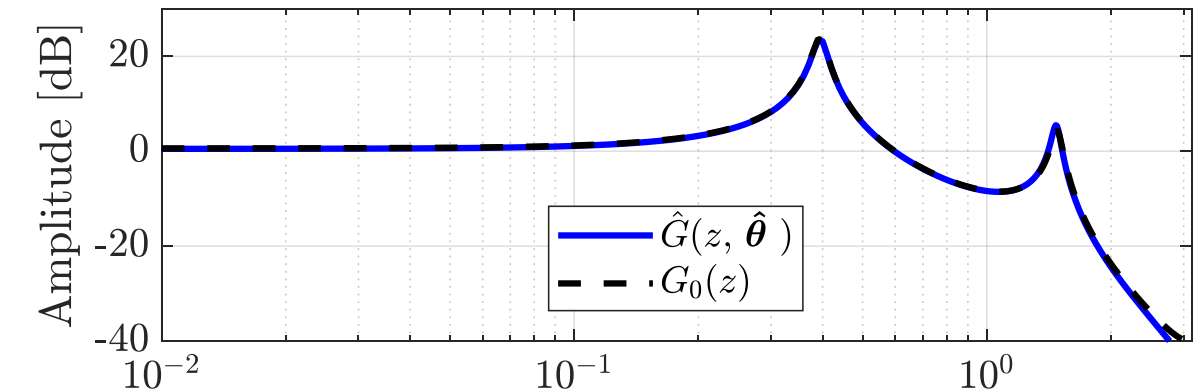
Esempio: multiseno

Riprendiamo l'esempio precedente e misuriamo $N = 2000$ dati con **ingresso multiseno**

Multisine input



Multisine input



Outline

1. Analisi asintotica dei metodi PEM
2. Identificabilità dei modelli e persistente eccitazione
- 3. Valutazione dell'incertezza della stima PEM**
4. Robustezza dei metodi PEM e prefiltraggio
5. Empirical Transfer Function Estimate (ETFE)



Varianza della stima dei parametri

Abbiamo visto che, quando $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$ e se l'ingresso è **sufficientemente eccitante** da rendere i dati informativi, i metodi PEM portano a stimare il valore vero dei parametri

Questo risultato vale però per $N \rightarrow \infty$. Cosa possiamo dire sulla **bontà della stima** PEM con un **numero finito** di dati?

IPOTESI DI LAVORO

- $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, per cui $\boldsymbol{\theta}^0 \in \Delta_{\boldsymbol{\theta}}$
- $\Delta_{\boldsymbol{\theta}} = \bar{\boldsymbol{\theta}}$, ovvero esiste un solo punto di minimo globale. Quindi $\boldsymbol{\theta}^0 = \bar{\boldsymbol{\theta}}$

Ipotizziamo di avere un **numero finito di dati** e di stimare $\hat{\boldsymbol{\theta}}_N = \arg \min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \frac{1}{N} \sum_{t=1}^N \varepsilon_1(t; \boldsymbol{\theta})^2$

Varianza della stima dei parametri

Ci ricordiamo che $\hat{\theta}_N$ è una **variabile casuale** in quanto i dati provengono da realizzazioni di processi stocastici. Dalle ipotesi precedenti ($\mathcal{S} \in \mathcal{M}(\theta)$ e $\theta^0 = \bar{\theta}$) abbiamo che $\mathbb{E}[\hat{\theta}_N] = \theta^0$

Vogliamo calcolare l'**incertezza di stima parametrica** $\text{Var}[\hat{\theta}_N] = \mathbb{E}[(\hat{\theta}_N - \theta^0) \cdot (\hat{\theta}_N - \theta^0)^\top]$

Si dimostra che:

$$\underset{d \times d}{\text{Var}[\hat{\theta}_N]} \equiv \bar{P}_\theta = \frac{1}{N} \lambda^2 \cdot \underset{d \times d}{\bar{R}_\theta^{-1}}$$

- $\bar{R}_\theta = \mathbb{E}[\psi(t; \theta^0) \psi(t; \theta^0)^\top]$

$$\underset{d \times 1}{\psi(t; \theta^0)} = - \frac{d}{d\theta} \varepsilon_1(t; \theta) \Big|_{\theta=\theta^0}$$

- $\lambda^2 = \text{Var}[e(t)]$

Varianza della stima dei parametri

Tali quantità dipendono da θ^0 . Nella pratica, si approssimano come:

$$\hat{\lambda}^2 = \frac{1}{N} \sum_{t=1}^N \left(y(t) - \hat{y}(t|t-1; \hat{\theta}_N) \right)^2 = J(\hat{\theta}_N)$$

$$\hat{R}_{\theta} = \frac{1}{N} \sum_{t=1}^N \psi(t; \hat{\theta}_N) \cdot \psi(t; \hat{\theta}_N)^{\top}$$

Interpretazione di \bar{P}_{θ}

Ricordiamo che $\bar{J}(\theta) = \mathbb{E}[\varepsilon_1(t; \theta)^2]$. Riprendendo quanto visto per la stima ARMAX:

$$\begin{aligned} \frac{d\bar{J}(\theta)}{d\theta} &= \mathbb{E} \left[2\varepsilon_1(t; \theta) \cdot \frac{d\varepsilon_1(t; \theta)}{d\theta} \right] \\ \frac{d^2\bar{J}(\theta)}{d\theta^2} &= \mathbb{E} \left[2 \frac{d\varepsilon_1(t; \theta)}{d\theta} \cdot \frac{d\varepsilon_1(t; \theta)^{\top}}{d\theta} + 2\varepsilon_1(t; \theta) \cdot \frac{d^2\varepsilon_1(t; \theta)}{d\theta^2} \right] \end{aligned}$$

Varianza della stima dei parametri

La quantità $\frac{d^2 \varepsilon_1(t; \boldsymbol{\theta})}{d\boldsymbol{\theta}^2}$ è funzione dell'errore di predizione e pertanto dipende dai valori passati $e(t-1), e(t-2), \dots$ Notiamo però che se $\boldsymbol{\theta} = \boldsymbol{\theta}^0$, allora $\varepsilon_1(t; \boldsymbol{\theta}) = e(t)$

Ne consegue che il termine $\mathbb{E} \left[2\varepsilon_1(t; \boldsymbol{\theta}) \cdot \frac{d^2 \varepsilon_1(t; \boldsymbol{\theta})}{d\boldsymbol{\theta}^2} \right]$ si **annulla**, in quanto $\varepsilon_1(t)$ e $\frac{d^2 \varepsilon_1(t; \boldsymbol{\theta})}{d\boldsymbol{\theta}^2}$ sono **incorrelati**. Quindi:

$$\left. \frac{d^2 \bar{J}(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = \mathbb{E} \left[2 \left. \frac{d\varepsilon_1(t; \boldsymbol{\theta})}{d\boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \cdot \left. \frac{d\varepsilon_1(t; \boldsymbol{\theta})^\top}{d\boldsymbol{\theta}} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} \right] = 2 \cdot \mathbb{E}[\boldsymbol{\psi}(t; \boldsymbol{\theta}^0) \boldsymbol{\psi}(t; \boldsymbol{\theta}^0)^\top] = 2 \cdot \bar{R}_{\boldsymbol{\theta}}$$



$$\bar{R}_{\boldsymbol{\theta}} = \frac{1}{2} \cdot \left. \frac{d^2 J(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$$

$\bar{R}_{\boldsymbol{\theta}}$ è la metà dell'Hessiana della funzione di costo valutata nell'ottimo

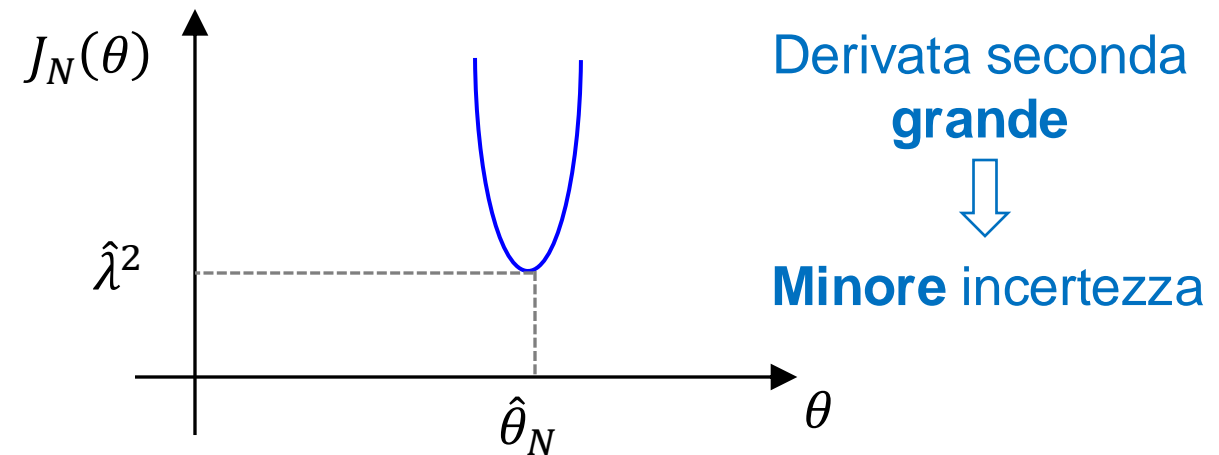
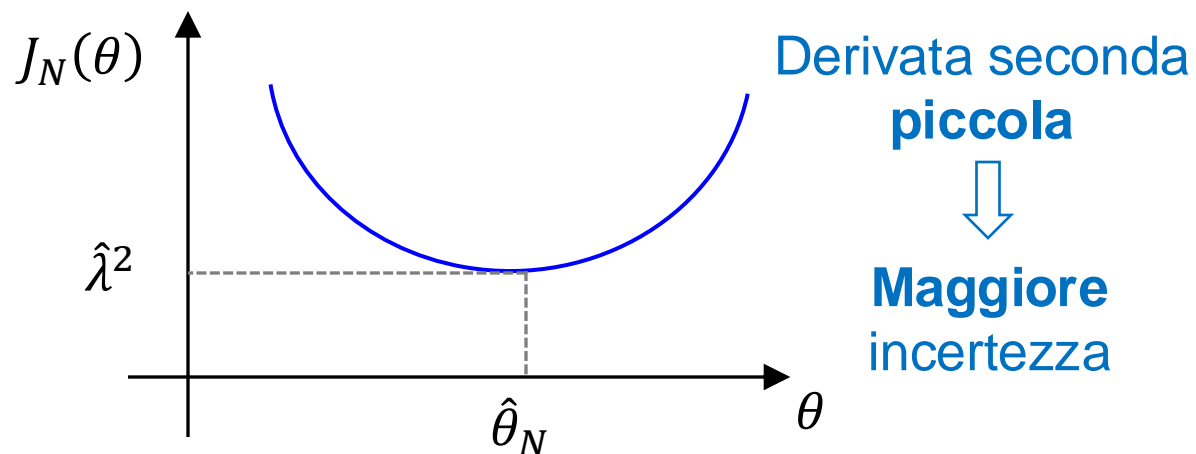
Varianza della stima dei parametri

Osservazione 1

$$\text{Var}[\hat{\theta}_N] \equiv \bar{P}_{\theta} = \frac{1}{N} \lambda^2 \cdot \bar{R}_{\theta}^{-1}$$

$$\bar{R}_{\theta} = \frac{1}{2} \cdot \frac{d^2 J(\theta)}{d\theta^2} \Big|_{\theta=\theta^0}$$

- La varianza dell'errore di stima dei parametri decresce all'aumentare di N
- La varianza dell'errore di stima dei parametri aumenta all'aumentare di λ^2
- La varianza dell'errore di stima dei parametri diminuisce all'aumentare della derivata seconda (Hessiana) della funzione di costo all'ottimo



Varianza della stima dei parametri

Osservazione 2

Più **grande** è la «potenza» del segnale di ingresso $u(t)$, più **piccola** è la matrice di varianza delle stime \bar{P}_θ

Questo perché \bar{P}_θ è proporzionale all'inverso della «potenza» del vettore di segnali $\psi(t; \theta) = -\frac{d\varepsilon_1(t; \theta)}{d\theta}$, e questo vettore di segnali è più «potente» tanto più $u(t)$ è potente. Infatti

$$\varepsilon_1(t; \theta) = H^{-1}(z; \theta)(y(t) - G(z, \theta)u(t)) = \frac{G_0(z) - G(z, \theta)}{H(z, \theta)}u(t) + \frac{H_0(z)}{H(z, \theta)}e(t)$$

Da cui si vede che $\varepsilon_1(t; \theta)$, e quindi anche $\psi(t; \theta)$, è proporzionale a $u(t)$

Varianza della stima dei parametri

Caso particolare: stima ARX

La stima è ottenuta tramite l'algoritmo dei minimi quadrati, con

$$\hat{y}(t|t-1; \boldsymbol{\theta}) = \boldsymbol{\varphi}^T(t) \boldsymbol{\theta}$$

$$\underset{d \times 1}{\boldsymbol{\varphi}(t)} = \begin{bmatrix} y(t-1) \\ y(t-2) \\ \vdots \\ y(t-n_a) \\ u(t-1) \\ u(t-2) \\ \vdots \\ u(t-n_b-1) \end{bmatrix}$$

L'errore di predizione a un passo è

$$\varepsilon_1(t; \boldsymbol{\theta}) = y(t) - \hat{y}(t|t-1; \boldsymbol{\theta}) = y(t) - \boldsymbol{\varphi}^T(t) \boldsymbol{\theta}$$

Quindi $\boldsymbol{\psi}(t; \boldsymbol{\theta}) = -\frac{d}{d\boldsymbol{\theta}} \varepsilon_1(t; \boldsymbol{\theta}) = \boldsymbol{\varphi}(t)$ e di conseguenza

$$\bar{P}_{\boldsymbol{\theta}} = \frac{1}{N} \lambda^2 \cdot \bar{R}_{\boldsymbol{\theta}}^{-1} = \frac{1}{N} \lambda^2 \cdot \mathbb{E}[\boldsymbol{\psi}(t; \boldsymbol{\theta}^0) \boldsymbol{\psi}(t; \boldsymbol{\theta}^0)^T]^{-1} = \frac{1}{N} \lambda^2 \cdot \mathbb{E}[\boldsymbol{\varphi}(t; \boldsymbol{\theta}^0) \boldsymbol{\varphi}(t; \boldsymbol{\theta}^0)^T]^{-1}$$

Varianza della stima dei parametri

Caso particolare: stima ARX

Usando la stima campionaria \hat{P}_θ di \bar{P}_θ , abbiamo che

$$\begin{aligned}\bar{P}_\theta &= \frac{1}{N} \lambda^2 \cdot \mathbb{E}[\boldsymbol{\varphi}(t; \boldsymbol{\theta}^0) \boldsymbol{\varphi}(t; \boldsymbol{\theta}^0)^\top]^{-1} \approx \frac{1}{N} \hat{\lambda}^2 \cdot \left[\frac{1}{N} \sum_{t=1}^N \boldsymbol{\varphi}(t; \hat{\boldsymbol{\theta}}_N) \cdot \boldsymbol{\varphi}(t; \hat{\boldsymbol{\theta}}_N)^\top \right]^{-1} \\ &= \hat{\lambda}^2 \cdot \left[\sum_{t=1}^N \boldsymbol{\varphi}(t; \hat{\boldsymbol{\theta}}_N) \cdot \boldsymbol{\varphi}(t; \hat{\boldsymbol{\theta}}_N)^\top \right]^{-1} = \hat{\lambda}^2 \cdot S(N)^{-1}\end{aligned}$$

Le **proprietà probabilistiche** della stima PEM di modelli ARX sono uguali a quelle della **stima a minimi quadrati** di modelli lineari «statici» ([Lezione 3 – slide 25](#))

Distribuzione asintotica delle stime dei parametri

Se $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, la distribuzione $\hat{\boldsymbol{\theta}}_N$ ottenuta tramite stima PEM converge asintoticamente ad una **Gaussiana**

$$\hat{\boldsymbol{\theta}}_N \sim \mathcal{N}(\boldsymbol{\theta}^0, \bar{P}_{\boldsymbol{\theta}})$$

$d \times 1$

Nel caso in cui $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, l'espressione di $\bar{P}_{\boldsymbol{\theta}}$ è quella ricavata precedentemente. Altrimenti, assume una forma più complicata

La relazione $\hat{\boldsymbol{\theta}}_N \sim \mathcal{N}(\boldsymbol{\theta}^0, \hat{P}_{\boldsymbol{\theta}})$, con $\hat{P}_{\boldsymbol{\theta}}$ al posto di $\bar{P}_{\boldsymbol{\theta}}$, può essere **usata nella pratica** per calcolare **intervalli di confidenza** sulla stima $\hat{\boldsymbol{\theta}}_N$, e valutare così **l'affidabilità della stima** di un certo parametro

Distribuzione asintotica delle stime dei parametri

Tali **intervalli di confidenza** ci dicono la probabilità p_θ che l'intervallo di confidenza contenga il vettore vero dei parametri

$$\mathcal{C}_\theta = \left\{ \theta \mid (\theta - \hat{\theta}_N)^\top \cdot \hat{P}_\theta^{-1} \cdot (\theta - \hat{\theta}_N) \leq \alpha \right\}$$

dove α è tale che $P(\chi^2(d) < \alpha) = p_\theta$. Di solito si usa $p_\theta = 0.95$

Il set di valori \mathcal{C}_θ è un **ellissoide** nello spazio dei parametri θ , centrato in $\hat{\theta}_N$. Notiamo che più \bar{P}_θ è grande, più grande sarà l'ellissoide

Varianza della stima delle funzioni di trasferimento

Dato che $\hat{\theta}_N$ è una **variabile casuale**, anche i **modelli identificati** $G(z, \hat{\theta}_N)$ e $H(z, \hat{\theta}_N)$ lo saranno. In molte situazioni è probabilmente più di interesse analizzare la **varianza della stima delle funzioni di trasferimento** $G(z, \hat{\theta}_N)$ e $H(z, \hat{\theta}_N)$, piuttosto che la varianza delle stime dei **parametri** $\hat{\theta}_N$

Per esempio, considerando $G(z, \theta)$, si può scrivere

$$\text{Var}[G(e^{j\omega}, \hat{\theta}_N)] \equiv \mathbb{E} \left[|G(e^{j\omega}, \hat{\theta}_N) - G(e^{j\omega}, \theta^0)|^2 \right]$$

Tale quantità può essere **stimata dai dati** usando \hat{P}_θ e $\hat{\theta}_N$

Varianza della stima delle funzioni di trasferimento

Assumendo sempre che $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$ e che $u(t) \perp e(t)$, e che inoltre $n \rightarrow +\infty$ (dove n è l'ordine di $G(z, \hat{\boldsymbol{\theta}}_N)$, cioè il numero di variabili di stato), l'espressione si può approssimare come

$$\text{Var}[G(e^{j\omega}, \hat{\boldsymbol{\theta}}_N)] \approx \frac{n}{N} \cdot \frac{\Gamma_{vv}(\omega)}{\Gamma_{uu}(\omega)}$$

in cui $\Gamma_{vv}(\omega)$ è la densità spettrale di potenza del disturbo $v(t) = H_0(z)e(t)$ e $\Gamma_{uu}(\omega)$ è la densità spettrale dell'ingresso $u(t)$

Notiamo che possiamo fare **«input shaping»** dell'ingresso $u(t)$ per **«favorire» la stima in una certa banda di frequenze** piuttosto che in altre

Varianza della stima delle funzioni di trasferimento

La varianza della stima della funzione di trasferimento del rumore $H(z, \hat{\theta}_N)$ è

$$\text{Var}[H(e^{j\omega}, \hat{\theta}_N)] \approx \frac{n}{N} \cdot \frac{\Gamma_{vv}(\omega)}{\lambda^2}$$

dove λ^2 è la varianza del rumore $e(t)$ che alimenta $H_0(z)$

Outline

1. Analisi asintotica dei metodi PEM
2. Identificabilità dei modelli e persistente eccitazione
3. Valutazione dell'incertezza della stima PEM
- 4. Robustezza dei metodi PEM e prefiltraggio**
5. Empirical Transfer Function Estimate (ETFE)



Robustezza dei metodi PEM

Un obiettivo principale dell'identificazione è valutare la **mancata corrispondenza tra il modello e realtà**. Sappiamo che quando $\mathcal{S} \notin \mathcal{M}(\boldsymbol{\theta})$ **non possiamo** raggiungere la parametrizzazione vera sia di $G_0(z)$ che di $H_0(z)$. Come possiamo **caratterizzare l'errore di stima del modello** in questa situazione?

Consideriamo il seguente meccanismo di generazione dei dati

$$\mathcal{S}: y(t) = G_0(z)u(t) + H_0(z)e(t), \quad e(t) \sim \text{WN}(0, \lambda^2)$$

E il modello ARMAX

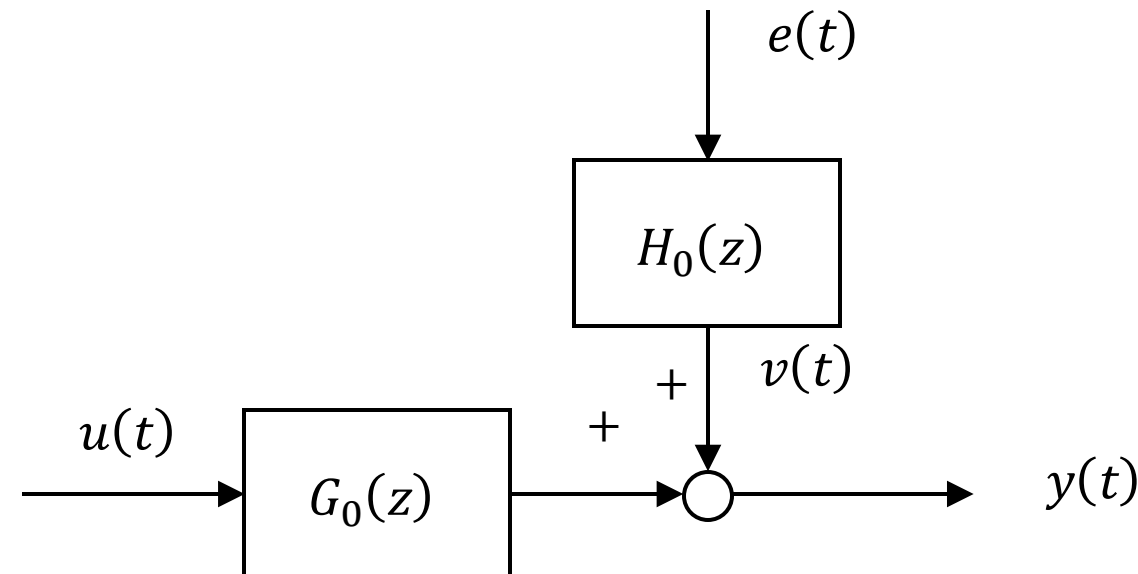
$$\mathcal{M}(\boldsymbol{\theta}): y(t) = \frac{B(z, \boldsymbol{\theta})}{A(z, \boldsymbol{\theta})} u(t-1) + \frac{C(z, \boldsymbol{\theta})}{A(z, \boldsymbol{\theta})} \eta(t), \quad \eta(t) \sim \text{WN}(0, \lambda_\eta^2)$$

Robustezza dei metodi PEM

Problema: vogliamo analizzare la **robustezza** dei metodi PEM nel caso in cui $\mathcal{S} \notin \mathcal{M}(\boldsymbol{\theta})$. In particolare, caratterizzare come **l'errore di modellazione** di $G_0(z) \notin G(z, \boldsymbol{\theta})$ dipende dall'errore di predizione a un passo

IPOTESI DI LAVORO

1. Supponiamo che il **segnale disturbo** $v(t)$ sia «**piccolo**» rispetto all'uscita del blocco $G_0(z)$
2. Vogliamo **stimare** $G_0(z)$ nel caso in cui $G_0(z) \notin G(z, \boldsymbol{\theta})$



Robustezza dei metodi PEM

Definiamo l'**errore di modellazione** come

$$\Delta G(z, \boldsymbol{\theta}) = G_0(z) - G(z, \boldsymbol{\theta})$$

Osservazione

Possiamo dividere l'errore di modellazione di $G_0(z)$ in due componenti:

$$\bullet \quad \Delta G(z, \boldsymbol{\theta}) = G_0(z) - G(z, \boldsymbol{\theta}) = \underbrace{\left(G_0(z) - G(z, \bar{\boldsymbol{\theta}}) \right)}_{\text{Bias di modello}} + \underbrace{\left(G(z, \bar{\boldsymbol{\theta}}) - G(z, \boldsymbol{\theta}) \right)}_{\text{Varianza di modello}}$$

Se $\mathcal{S} \in \mathcal{M}(\boldsymbol{\theta})$, allora $G_0(z) - G(z, \bar{\boldsymbol{\theta}}) = 0$. Il termine $G(z, \bar{\boldsymbol{\theta}}) - G(z, \boldsymbol{\theta})$ va a zero asintoticamente

Robustezza dei metodi PEM

Ricaviamo l'espressione dell'errore $\eta(t)$ del modello ARMAX

$$\mathcal{M}(\boldsymbol{\theta}): y(t) = \frac{B(z, \boldsymbol{\theta})}{A(z, \boldsymbol{\theta})} u(t-1) + \frac{C(z, \boldsymbol{\theta})}{A(z, \boldsymbol{\theta})} \eta(t), \quad \eta(t) \sim \text{WN}(0, \lambda_\eta^2)$$

$$= G(z, \boldsymbol{\theta})u(t) + H(z, \boldsymbol{\theta})\eta(t)$$



$$\eta(t) = H^{-1}(z, \boldsymbol{\theta})[y(t) - G(z, \boldsymbol{\theta})u(t)]$$

Sostituiamo l'espressione di $y(t) = G_0(z)u(t) + H_0(z)e(t)$

$$\eta(t) = H^{-1}(z, \boldsymbol{\theta})[G_0(z)u(t) - G(z, \boldsymbol{\theta})u(t) + H_0(z)e(t)]$$

$$= H^{-1}(z, \boldsymbol{\theta})[\Delta G(z, \boldsymbol{\theta})u(t) + H_0(z)e(t)]$$

Robustezza dei metodi PEM

Notiamo però che $\eta(t)$ è anche l'errore di predizione ad un passo $\varepsilon_1(t; \boldsymbol{\theta})$ (abbiamo fatto gli stessi passaggi in [Lezione 10 - slide 75](#)), per cui possiamo scrivere

$$\varepsilon_1(t; \boldsymbol{\theta}) = \frac{1}{H(z, \boldsymbol{\theta})} [\Delta G(z, \boldsymbol{\theta})u(t) + H_0(z)e(t)]$$

Questa espressione connette l'**errore di modellazione** con l'**errore di predizione**, mettendo in risalto che $\varepsilon_1(t; \boldsymbol{\theta})$ dipende linearmente da $u(t)$ e da $e(t)$.

In particolare, l'influenza di $u(t)$ su $\varepsilon_1(t; \boldsymbol{\theta})$ è dettata da $\Delta G(z, \boldsymbol{\theta})$

Robustezza dei metodi PEM

Il vettore di parametri $\boldsymbol{\theta}$ è stimato tramite un metodo PEM, ovvero minimizzando la funzione di costo $J_N(\boldsymbol{\theta})$

$$J_N(\boldsymbol{\theta}) = \frac{1}{N} \sum_{t=1}^N \varepsilon_1(t; \boldsymbol{\theta})^2$$

Asintoticamente, supponendo l'ergodicità di $u(t)$ e $y(t)$, abbiamo che

$$J_N(\boldsymbol{\theta}) \xrightarrow[N \rightarrow +\infty]{} \bar{J}(\boldsymbol{\theta}) = \mathbb{E}_s[\varepsilon_1(t, \boldsymbol{\theta})^2]$$

Possiamo **interpretare in frequenza** la funzione di costo $\bar{J}(\boldsymbol{\theta})$ come

$$\bar{J}(\boldsymbol{\theta}) = \mathbb{E}[\varepsilon_1(t, \boldsymbol{\theta})^2] = \gamma_{\varepsilon_1 \varepsilon_1}(0) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_{\varepsilon_1 \varepsilon_1}(\omega) d\omega$$

Robustezza dei metodi PEM

Ipotizzando ora che $u(t) \perp e(t)$, che non è un'ipotesi restrittiva quando il sistema è in **anello aperto**, abbiamo che $\Gamma_{\varepsilon_1 \varepsilon_1}(\omega)$ è la **somma dei contributi** dati da $\Gamma_{uu}(\omega)$ e $\Gamma_{ee}(\omega)$

$$\Gamma_{\varepsilon_1 \varepsilon_1}(\omega) = \left| \frac{\Delta G(e^{j\omega}, \boldsymbol{\theta})}{H(e^{j\omega}, \boldsymbol{\theta})} \right|^2 \Gamma_{uu}(\omega) + \left| \frac{H_0(e^{j\omega})}{H(e^{j\omega}, \boldsymbol{\theta})} \right|^2 \Gamma_{ee}(\omega)$$

Per l'**ipotesi di lavoro 1** trascuriamo il secondo termine, ottenendo

$$\Gamma_{\varepsilon_1 \varepsilon_1}(\omega) \approx \left| \frac{\Delta G(e^{j\omega}, \boldsymbol{\theta})}{H(e^{j\omega}, \boldsymbol{\theta})} \right|^2 \Gamma_{uu}(\omega)$$

Sostituendo $\Gamma_{\varepsilon_1 \varepsilon_1}(\omega) \approx \left| \frac{\Delta G(e^{j\omega}, \boldsymbol{\theta})}{H(e^{j\omega}, \boldsymbol{\theta})} \right|^2 \Gamma_{uu}(\omega)$ nell'espressione $\bar{J}(\boldsymbol{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \Gamma_{\varepsilon_1 \varepsilon_1}(\omega) d\omega$ si ha:

Robustezza dei metodi PEM

$$\bar{J}(\boldsymbol{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Delta G(e^{j\omega}, \boldsymbol{\theta})|^2 \cdot \frac{\Gamma_{uu}(\omega)}{|H(e^{j\omega}, \boldsymbol{\theta})|^2} d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Delta G(e^{j\omega}, \boldsymbol{\theta})|^2 \cdot g(\omega, \boldsymbol{\theta}) d\omega$$

- $|\Delta G(e^{j\omega}, \boldsymbol{\theta})|^2$ rappresenta **l'errore di stima della risposta in frequenza**, per ogni pulsazione ω
- $g(\omega, \boldsymbol{\theta})$ è un **peso in frequenza**. Per esempio, $g(\omega, \boldsymbol{\theta}) = \text{costante} \forall \omega$, allora gli errori alle diverse frequenze sono pesati allo stesso modo. Oppure, se $g(\omega, \boldsymbol{\theta}) = 0$ se $\omega > \bar{\omega}$, allora solo gli errori nella banda di frequenze $[0, \bar{\omega}]$ contribuiranno alla funzione di costo

Robustezza dei metodi PEM

Se consideriamo un **rumore bianco** come ingresso, abbiamo che

$$g(\omega, \boldsymbol{\theta}) = \frac{\Gamma_{uu}(\omega)}{|H(e^{j\omega}, \boldsymbol{\theta})|^2} = \frac{\lambda^2}{|H(e^{j\omega}, \boldsymbol{\theta})|^2}$$

ovvero la pesatura in frequenza dipende solo dal modello del rumore $H(e^{j\omega}, \boldsymbol{\theta})$. Se facciamo l'ulteriore ipotesi che il **modello del rumore sia una funzione fissata** $H^*(e^{j\omega})$:

$$g(\omega) = \frac{\lambda^2}{|H^*(e^{j\omega})|^2}$$

Notiamo che $g(\omega)$
non dipende più da $\boldsymbol{\theta}$

cioè **peso di più le frequenze dove l'entità del rumore è più bassa** (che vuole dire che do più importanza a dati «più affidabili»)

Robustezza dei metodi PEM

Osservazioni

- Se $G_0(z) \in G(z, \theta)$, allora $\bar{J}(\bar{\theta}) = 0$ e la funzione peso non riveste alcun ruolo
- Usando il **rumore bianco**, **non posso «focalizzare gli sforzi»** dell'identificazione in una determinata **banda di frequenze** di interesse (ad esempio, spesso per il controllo mi basta un modello buono solo fino alla ω_c)
- Potrei quindi scegliere un ingresso $u(t)$ diverso dal rumore bianco (o similari), e quindi avere un $\Gamma_{uu}(\omega)$ diverso, per pesare di più alcune frequenze a discapito di altre

Purtroppo, non è sempre possibile scegliere l'ingresso a piacere. È però comunque possibile influenzare la pesatura in frequenza tramite **prefiltraggio dei dati**

Prefiltraggio

Il **prefiltraggio** consiste nel **filtrare sia l'ingresso che l'uscita** tramite un filtro $L(z)$ **asintoticamente stabile**

$$u_F(t) = L(z)u(t)$$

$$y_F(t) = L(z)y(t)$$

Tale operazione **non modifica la relazione ingresso-uscita**, anche se modifica la densità spettrale di potenza del rumore. Infatti:

$$L(z)y(t) = G_0(z)L(z)u(t) + H_0(z)L(z)e(t)$$

Prefiltraggio

Filtrare l'ingresso $u(t)$ e l'uscita $y(t)$ equivale a filtrare l'errore di predizione $\varepsilon_1(t; \boldsymbol{\theta})$, ottenendo un **errore filtrato** $\varepsilon_F(t; \boldsymbol{\theta})$, infatti:

$$\varepsilon_F(t; \boldsymbol{\theta}) = L(z)\varepsilon_1(t; \boldsymbol{\theta}) = \frac{L(z)}{H(z, \boldsymbol{\theta})} [y(t) - G(z, \boldsymbol{\theta})u(t)] = \frac{1}{H(z, \boldsymbol{\theta})} [L(z)y(t) - G(z, \boldsymbol{\theta})L(z)u(t)]$$

La funzione di costo asintotica, in frequenza, diventa quindi:

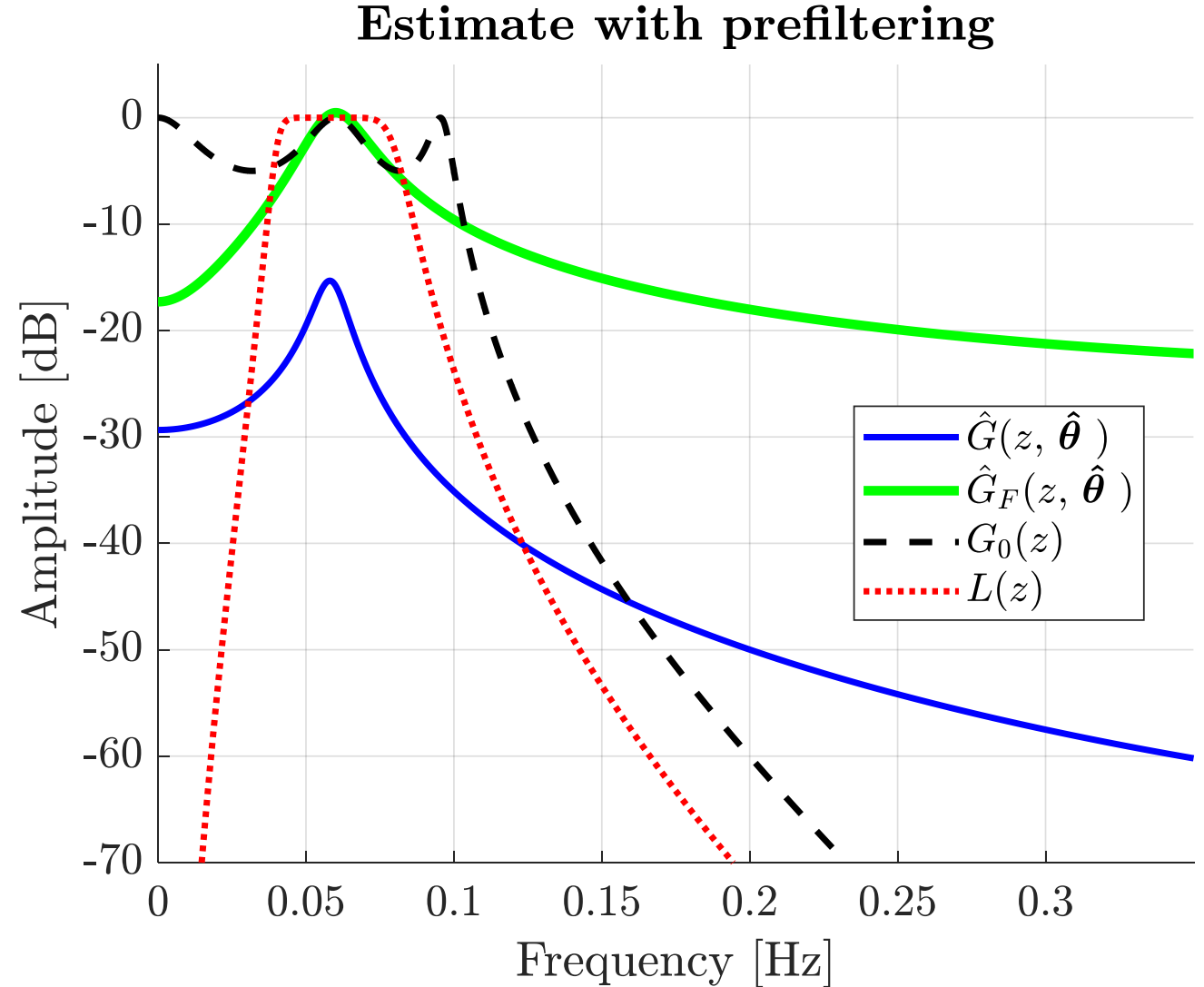
$$\bar{J}(\boldsymbol{\theta}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Delta G(e^{j\omega}, \boldsymbol{\theta})|^2 \cdot \frac{\Gamma_{uu}(\omega)}{H(e^{j\omega}, \boldsymbol{\theta})} \cdot |L(e^{j\omega})|^2 d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |\Delta G(e^{j\omega}, \boldsymbol{\theta})|^2 \cdot g_F(\omega, \boldsymbol{\theta}) d\omega$$

Esempio: prefiltraggio

Misuriamo $N = 5000$ dati da un sistema OE(5,5,0) con risposta in frequenza come in figura, e con $\lambda^2 = 0.01$

Identifichiamo un modello ARX(2,2,1), per il quale $\mathcal{S} \notin \mathcal{M}(\theta)$

Progettiamo un filtro $L(z)$ che abbia un modulo alto nella banda di frequenze $[0.04, 0.08]$ Hz



Outline

1. Analisi asintotica dei metodi PEM
2. Identificabilità dei modelli e persistente eccitazione
3. Valutazione dell'incertezza della stima PEM
4. Robustezza dei metodi PEM e prefiltraggio
- 5. Empirical Transfer Function Estimate (ETFE)**



Multiseno e stima nonparametrica

Dato che il multiseno è un segnale **periodico**, è possibile utilizzare la DFT (tramite l'algoritmo FFT) per calcolare lo spettro di uno (o più) periodi del segnale, ottenendo uno spettro **senza leakage**

Questo viene molto comodo in quanto è possibile ottenere una **stima nonparametrica** iniziale della $G(z)$, **prima di effettuare la stima parametrica** con metodi PEM

Supponiamo di dividere i segnali $u(t)$ e $y(t)$, entrambi multiseni, in sotto-sequenze $u^{[p]}, y^{[p]}$ che contengono i diversi $p = 1, \dots, P$ periodi. Definiamo:

$$\tilde{U}^{[p]} = \text{DFT}(u^{[p]}) \text{ DFT dell'ingresso}$$

$$\tilde{Y}^{[p]} = \text{DFT}(y^{[p]}) \text{ DFT dell'uscita}$$

Empirical Transfer Function Estimate (ETFE)

Possiamo ottenere una **stima nonparametrica** della $G(z)$ nelle **griglia di frequenze** definite dalla DFT $k = \text{bin} : \text{bin} : f_s/2$, come

$$\hat{U}(k) = \frac{1}{P} \sum_{p=1}^P \tilde{U}^{[p]}(k)$$



$$\hat{Y}(k) = \frac{1}{P} \sum_{p=1}^P \tilde{Y}^{[p]}(k)$$

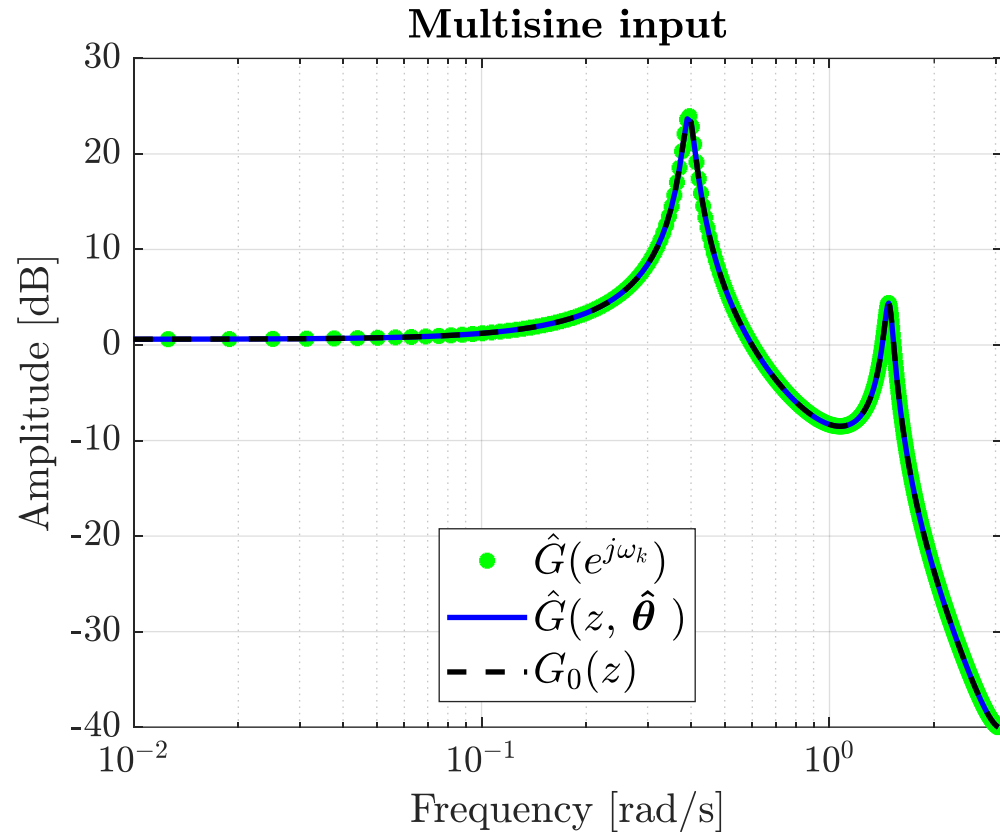
$$\hat{G}(k) = \frac{\hat{Y}(k)}{\hat{U}(k)}$$

Empirical Transfer Function Estimate (ETFE)

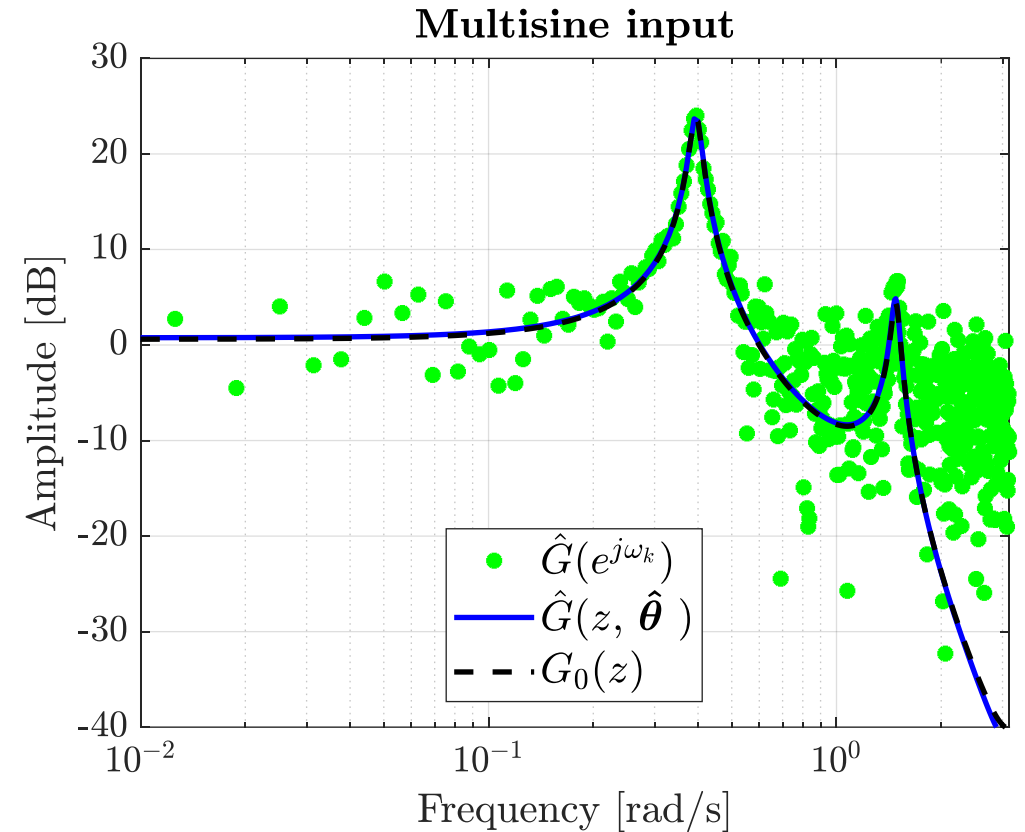
$\hat{G}(k)$ è un **numero complesso** che rappresenta la **risposta in frequenza** del sistema alla frequenza k . È poi possibile calcolare **modulo** e **fase** del numero complesso a diverse frequenze k per tracciare i **diagrammi di Bode**

Esempio: ETFE tramite multiseno

Senza rumore sull'uscita



Con rumore sull'uscita



Senza rumore, la stima è perfetta perché **non c'è leakage** indotto dalla FFT, grazie alla **periodicità** di $u(t)$ e $y(t)$

Empirical Transfer Function Estimate (ETFE)

La **stima nonparametrica ETFE** è usata per ottenere una **prima indicazione** sulle caratteristiche del sistema, in particolar modo sulla $G(z)$. **Non permette** di ottenere informazioni sullo **spettro del disturbo**

La stima ottenuta ci dà inoltre un'indicazione sulla **banda di frequenza di interesse** del sistema, sia per l'**identificazione parametrica** (es. usando poi il **prefiltraggio**) sia per l'applicazione del modello per fini di **progettazione del controllo**

La ETFE può essere effettuata anche con **segnali non periodici** (In Matlab **etfe**): in questo caso, è utile usare delle **finestre** (es. Hanning window) per ridurre il leakage (a discapito della risoluzione in frequenza), e delle procedure di **spectral averaging**

Stima nonparametrica tramite spectral analysis

In Matlab esiste anche il comando **spa** oppure **spafdr** che permette la stima della funzione di trasferimento $G_0(z)$ e $H_0(z)$ tramite **analisi spettrale**

Dato $y(t) = G_0(z)u(t) + v(t)$, con $v(t) \perp u(t)$, abbiamo che $\Gamma_{uy}(\omega) = G_0(e^{j\omega})\Gamma_{uu}(\omega)$ Slide 98
Lezione 08

E quindi una stima nonparametrica di $G_0(z)$ si può ottenere come

$$\hat{G}(e^{j\omega}) = \frac{\hat{\Gamma}_{uy}(\omega)}{\hat{\Gamma}_{uu}(\omega)}$$

Inoltre, dato che $\Gamma_{yy}(\omega) = |G_0(e^{j\omega})|^2 \cdot \Gamma_{uu}(\omega) + \Gamma_{vv}(\omega)$, una stima di $\Gamma_{vv}(\omega)$ è

$$\hat{\Gamma}_{vv}(\omega) = \hat{\Gamma}_{yy}(\omega) - \frac{|\hat{\Gamma}_{uy}(\omega)|^2}{\hat{\Gamma}_{uu}(\omega)}$$

Identificazione in frequenza

Una volta che si è ottenuta la ETFE, è anche possibile «fittare i punti in frequenza» anziché «fittare i dati nel tempo». Abbiamo quindi una vera e propria **identificazione parametrica in frequenza**, per esempio minimizzando

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \frac{1}{F} \sum_{k=1}^F |\hat{G}(e^{j\omega_k}) - G(e^{j\omega_k}; \theta)|^2 W_k$$

W_k è un peso in frequenza

L'approccio è particolarmente utile quando:

- $u(t)$ è un **multiseno** e il numero di dati nel dominio del tempo è **grande**
- vi sono **dati da diversi esperimenti**, che devono essere usati assieme
- si vuole identificare modelli a **tempo continuo**



**UNIVERSITÀ
DEGLI STUDI
DI BERGAMO**

Dipartimento
di Ingegneria Gestionale,
dell'Informazione e della Produzione