UNIVERSITÀ DEGLI STUDI DI BERGAMO | Dipartimento di Ingegneria Gestionale, dell'Informazione e della Produzione

Control Automation Lab

# IDENTIFICAZIONE DEI MODELLI E ANALISI DEI DATI (IMAD)

Corso di Laurea Magistrale in INGEGNERIA INFORMATICA

## Lezione 11: Identificazione – concetti fondamentali

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

**Stima parametrica $\hat{\theta}$**

- **$\theta$ 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
- **$\theta$ variabile casuale**
  - ***SI assunzioni su ddp dei dati***
    - ✓ Stima Bayesiana

**Machine learning**

## Parte II: sistemi dinamici

**Stima parametrica $\hat{\theta}$**

- **$\theta$ 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. Introduzione all’identificazione dei modelli dinamici
2. Metodi a minimizzazione dell’errore di predizione (PEM)
3. Identificazione PEM di modelli ARX
4. Identificazione PEM di modelli ARMAX

# Outline

1. **Introduzione all’identificazione dei modelli dinamici**
2. Metodi a minimizzazione dell’errore di predizione (PEM)
3. Identificazione PEM di modelli ARX
4. Identificazione PEM di modelli ARMAX

# Gli step per la risoluzione del problema

Seguiremo tre fasi per risolvere il problema della **modellazione di sistemi dinamici**:

Definizione delle **classi di modelli** $\mathcal{M}$ di sistemi dinamici

Ci concentreremo su modelli di **sistemi dinamici lineari**, espressi da **funzioni di trasferimento razionali fratte**. I parametri ignoti sono i coefficienti dei polinomi al numeratore e denominatore

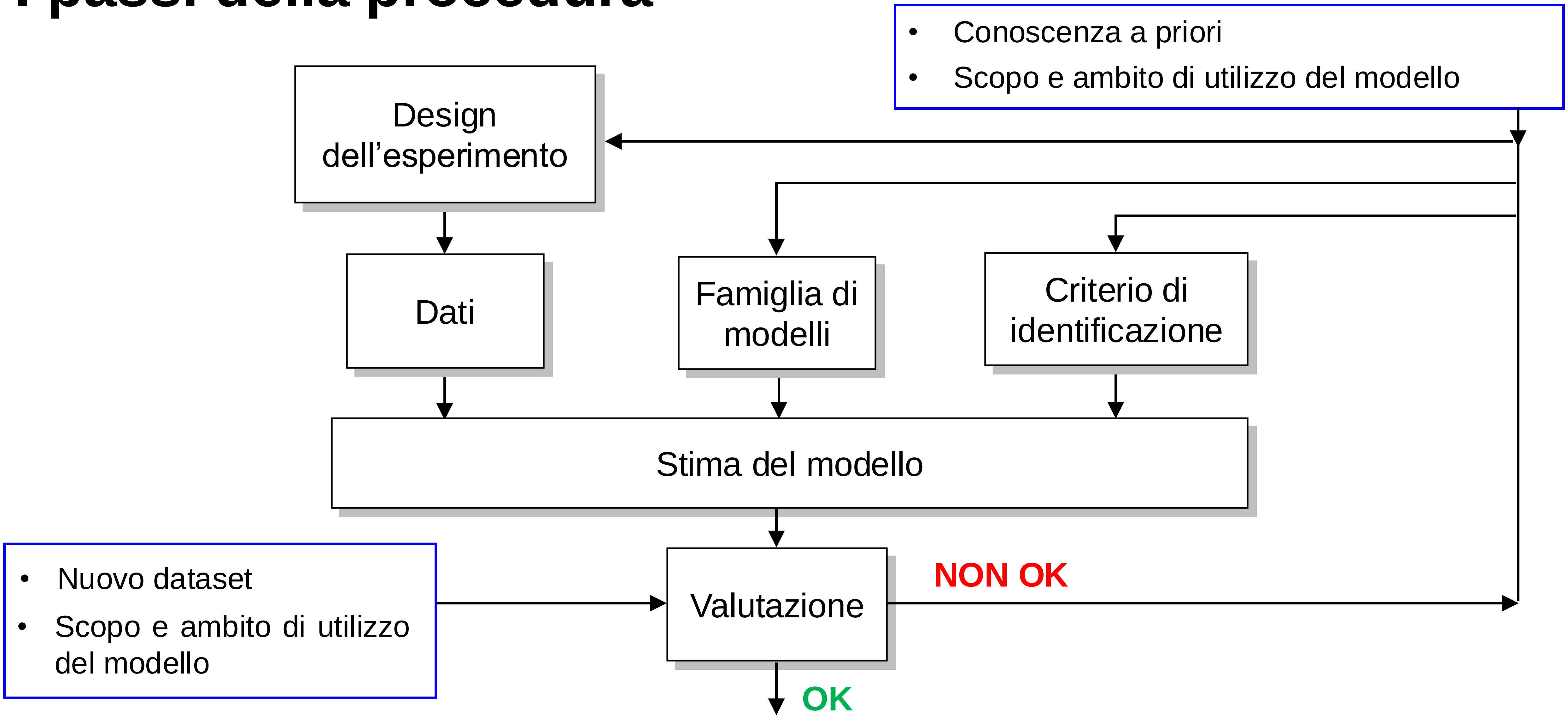
**Predizione**

Data una particolare classe di modello, supponendo di conoscerne il valore dei parametri, qual è il **predittore ottimo**? Quanto vale la predizione ottima?

**Identificazione**

Come **stimo il valore dei parametri** del modello scelto per la modellazione dei dati?

# I passi della procedura



# I passi della procedura di identificazione

**Dati:** i dati possono essere raccolti o dal funzionamento nominale del sistema, oppure tramite **esperimenti progettati ad-hoc**, in modo da ottenere **specifiche informazioni** (guadagno, risposte a scalino, risposta in frequenza,…)

**Famiglia di modelli:** è necessario scegliere quale tipo di modello usare per spiegare il fenomeno. Possibili scelte sono:

- **Lineare** \ non lineare
- **Tempo invariante** \ tempo variante
- **Discreto** \ continuo
- Altre proprietà (e.g. complessità del modello, struttura,…)

# I passi della procedura di identificazione

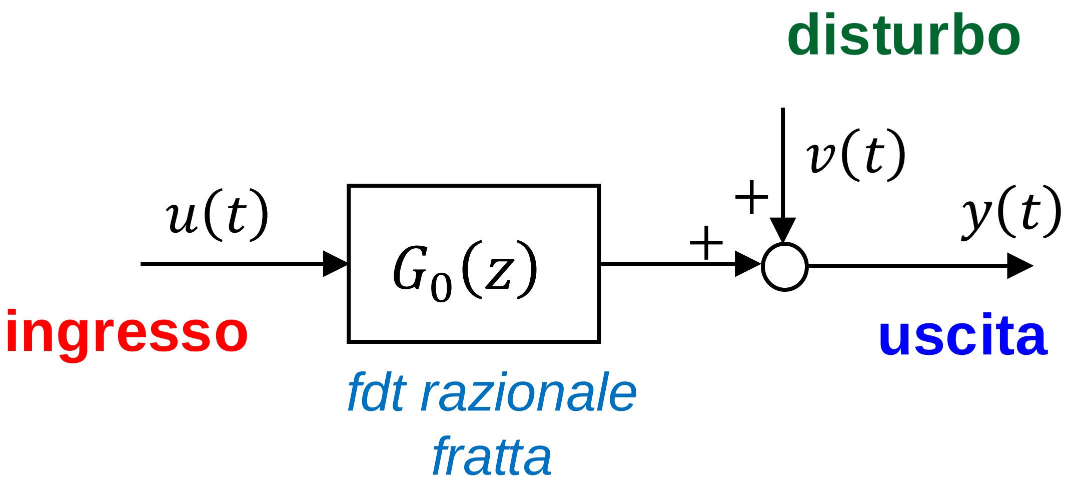
**Criterio di identificazione:** avendo a disposizione le misure e la famiglia di modelli scelta, è necessario decidere **come confrontare il modello con i dati**. Ciò si traduce spesso nella definizione di una **funzione di costo** da minimizzare

In tutti questi tre aspetti, una **conoscenza a priori** riguardo il sistema da identificare può essere di aiuto (es. la fisica mi dice che il modello deve avere un certo numero di poli\zeri)

**Validazione del modello:** in aggiunta a criteri di bontà «oggettivi», un modello potrebbe essere buono o meno a seconda dell'**applicazione** per il quale verrà usato. Per esempio, modelli «per il controllo» possono essere meno accurati di un «modello per la simulazione»

# Identificazione dei sistemi dinamici

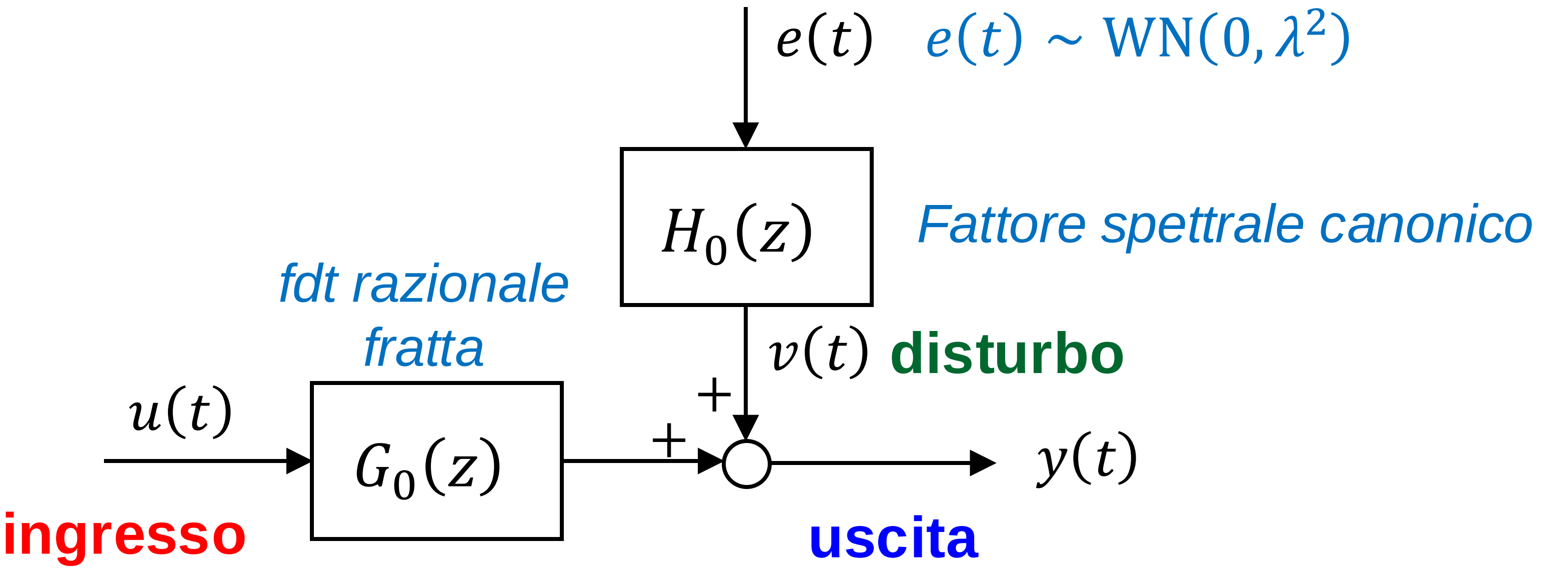
**Ipotesi di lavoro 1:** i dati sono generati da un **sistema LTI SISO con uscita rumorosa**




- I **parametri da stimare** sono i coefficienti del numeratore e del denominatore di $G_0(z)$
- Il **disturbo** $v(t)$ modella rumore di misura, ingressi non misurabili. Nel caso in cui il sistema vero non sia LTI, $v(t)$ modella anche gli scostamenti da questa assunzione

# Identificazione dei sistemi dinamici

**Ipotesi di lavoro 2:** il disturbo $v(t)$ è modellizzabile come un **processo stocastico stazionario a spettro razionale**, indipendente da $u(t)$

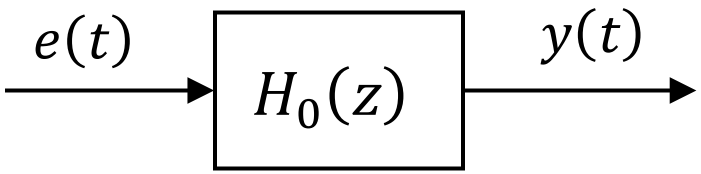



- In questo caso, vogliamo sia stimare i **coefficienti del numeratore e denominatore** di $G_0(z)$ sia quelli di $H_0(z)$ (e anche stimare $\lambda^2$)

# Identificazione dei sistemi dinamici

**Caso particolare:** non c'è l'ingresso $u(t)$, ovvero sto trattando una **serie temporale**. In pratica, misuro solo l'uscita alimentata alimentata dal rumore bianco

$e(t) \sim \mathrm{WN}(0, \lambda^2)$




*fdt razionale fratta*

Vogliamo sia stimare i **coefficienti del numeratore e denominatore** di $H_0(z)$ e $\lambda^2$. Posso poi calcolare $\Gamma_{yy}(\omega) = \left|H_0\left(e^{j\omega}\right)\right|^2 \cdot \lambda^2$

Questo approccio alla stima di $\Gamma_{yy}(\omega)$ prende il nome di **stima spettrale parametrica** (la **stima «nonparametrica»** è quella basata sul periodogramma)

# Identificazione dei sistemi dinamici

Il **modello più generale** che usiamo per stimare un sistema dinamico è dato da

$$\boxed{y(t) = G(z,\boldsymbol{\theta})u(t) + H(z,\boldsymbol{\theta})e(t), \qquad e(t) \sim \mathrm{WN}(0,\lambda^2)}$$

- $H(z,\boldsymbol{\theta})$: **fattore spettrale canonico**, ovvero numeratore e denominatore monici, coprimi, di uguale grado, poli e zeri nel cerchio unitario

- $G(z,\boldsymbol{\theta})$: **funzione di trasferimento** che descrive l'effetto dell'ingresso $u(t)$, misurabile o noto, sull'uscita $y(t)$

# Identificazione dei sistemi dinamici

## Proprietà di $G(z, \boldsymbol{\theta})$

- Spesso si ipotizza che $G(z, \boldsymbol{\theta})$ sia **strettamente propria**, ovvero che il grado del numeratore è minore del grado del denominatore
  - ✓ Questo fa si che vi sia un **ritardo puro** $k \neq 0$ tra ingresso e uscita

- $G(z, \boldsymbol{\theta})$ può avere **zeri fuori dal cerchio** o numeratore e denominatore **non monici**

- $G(z, \boldsymbol{\theta})$ rappresenta un **sistema fisico**, mentre $H(z, \boldsymbol{\theta})$ ed $e(t)$ **non esistono nella realtà** (sono solo costrutti matematici)

# Outline

1. Introduzione all’identificazione dei modelli dinamici
2. **Metodi a minimizzazione dell’errore di predizione (PEM)**
3. Identificazione PEM di modelli ARX
4. Identificazione PEM di modelli ARMAX

# L’approccio predittivo all’identificazione

Una volta definita la classe di modelli (per esempio ARMAX), potrei stimare i coefficienti $\boldsymbol{\theta}$ usando la stima a **massima verosimiglianza** o la **stima Bayesiana**

In questi casi, dovrei però fare ipotesi sulla distribuzione dei dati (per esempio assumendo che $e(t)$ sia un **processo gaussiano**)

In alternativa, una via più **semplice** ed **intuitiva**, e che **non necessita di ipotesi ulteriori**, è quella di trovare un approccio basato sulla **minimizzazione di una somma di residui al quadrato** (come per il minimi quadrati)

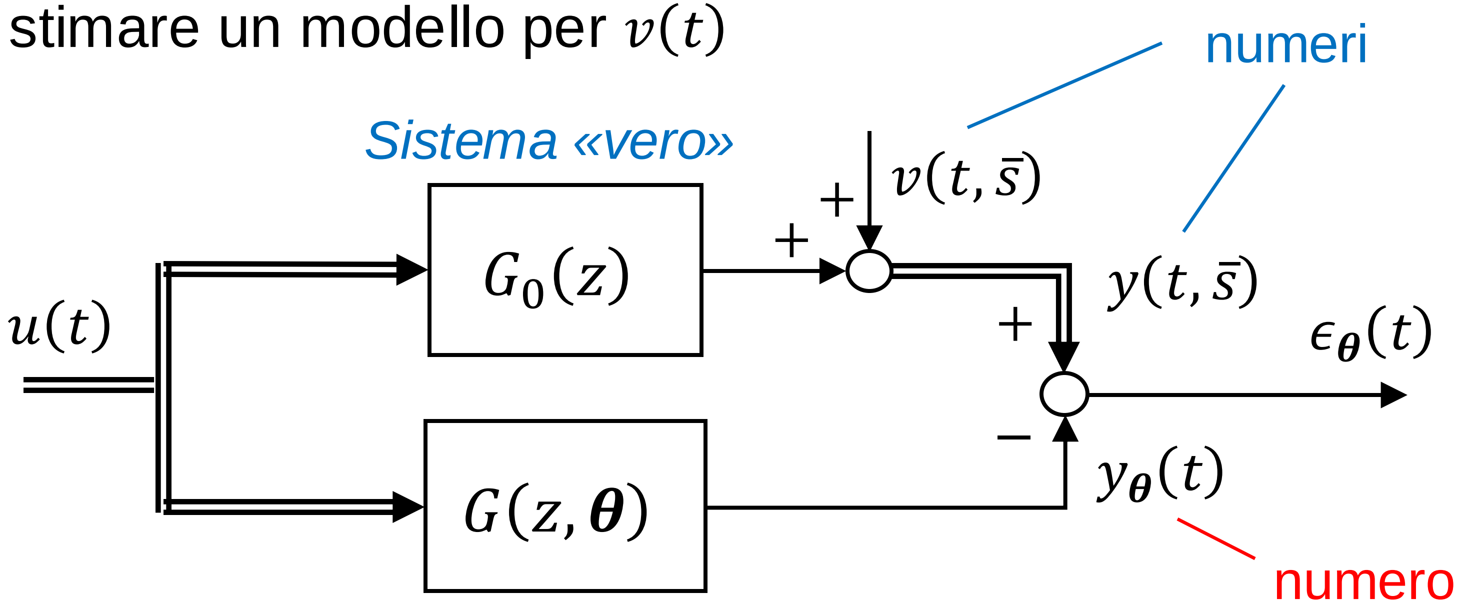
L’approccio che seguiremo si basa su questa strada

# L'approccio predittivo all'identificazione

**Caso «semplice»:** non mi interessa stimare un modello per $v(t)$

**Dati a disposizione**

- $\{u(1), \dots, u(N)\}$
- $\{y(1), \dots, y(N)\}$



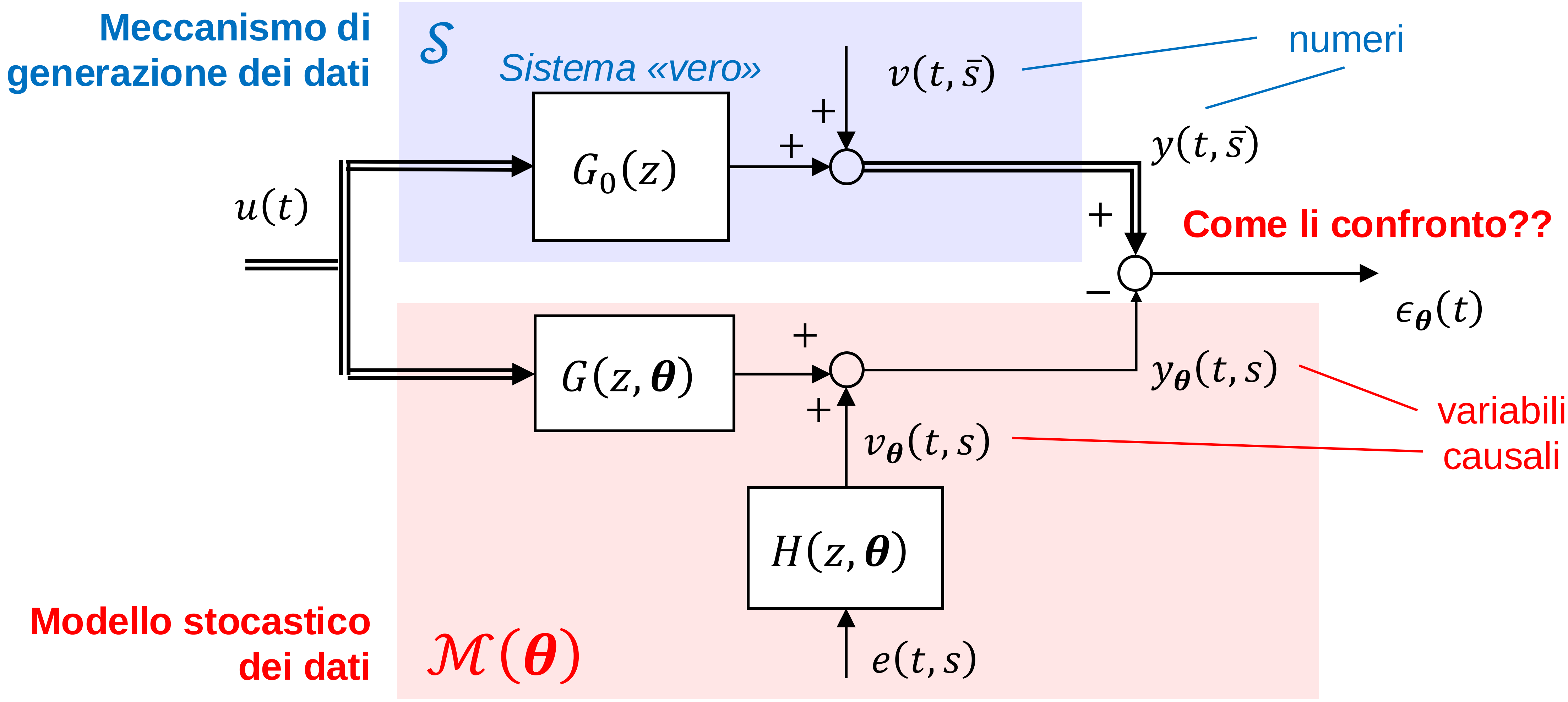

Il valore $y_{\boldsymbol{\theta}}(t)$ è la **simulazione** del modello $G(z, \boldsymbol{\theta})$ a fronte dell'ingresso $u(t)$. La stima a minimi quadrati si trova minimizzando l'**errore di simulazione** $\epsilon_{\boldsymbol{\theta}}(t)$

$$\widehat{\boldsymbol{\theta}}_{\mathrm{LS}} = \arg\min_{\boldsymbol{\theta}} \sum_{t=1}^{N} \epsilon_{\boldsymbol{\theta}}(t)^2$$

La soluzione è in forma chiusa solo se $y_{\boldsymbol{\theta}}(t)$ è lineare dei parametri

# L'approccio predittivo all'identificazione

**Caso «più difficile»:** oltre a stimare $G_0(z)$, voglio stimare anche un modello per $v(t)$



# L'approccio predittivo all'identificazione

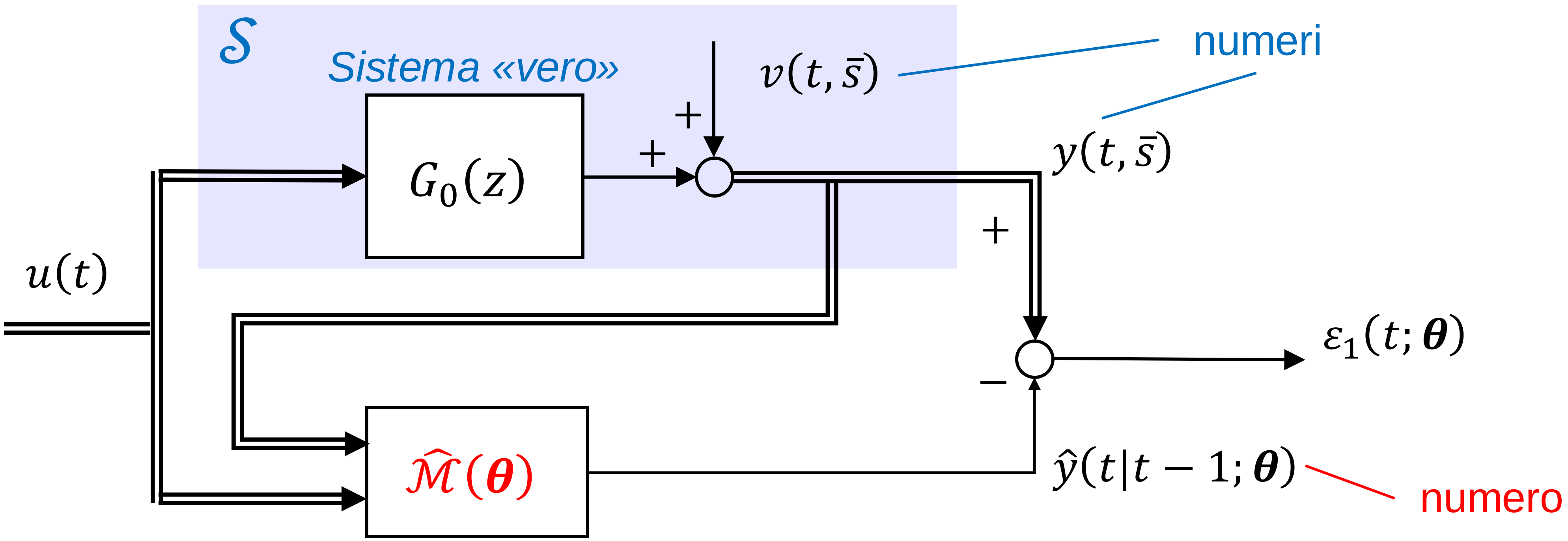
Anche se $v(t, s)$ è un processo stocastico, nella pratica **una volta che i dati sono stati collezionati**, «è già avvenuta» una «scelta» dell'esito $s = \bar{s}$ che ha **generato quei dati osservati** $y(t, s = \bar{s})$. Quindi, la quantità $y(t, s = \bar{s})$ è un vettore di **numeri** perché frutto di una particolare realizzazione $v(t, s = \bar{s})$

Il mio modello, invece, ha come uscita $y_{\boldsymbol{\theta}}(t, s)$ che, se non fisso un esito, è un **processo stocastico**. Abbiamo quindi il dilemma che **non possiamo confrontare** un vettore di numeri con un processo stocastico

Se conoscessi il valore dell'esito $s = \bar{s}$, allora **potrei simulare** l'uscita del mio modello con quell'esito, e far si che $y_{\boldsymbol{\theta}}(t, s = \bar{s})$ sia un vettore di numeri. Tuttavia, **non conosco l'esito** $\bar{s}$

# L’approccio predittivo all’identificazione

**Idea:** considero come residuo $\epsilon_{\boldsymbol{\theta}}(t)$ da minimizzare l’**errore di predizione a un passo** $\varepsilon_1(t;\boldsymbol{\theta})$




$\widehat{\mathcal{M}}(\boldsymbol{\theta})$: **predittore ottimo ad un passo** associato al modello $\mathcal{M}(\boldsymbol{\theta})$

# L'approccio predittivo all'identificazione

Definiamo quindi la stima ottenuta, considerando gli errori di predizione, come:

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} J_N(\boldsymbol{\theta}) \qquad\qquad J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_1(t;\boldsymbol{\theta})^2$$

$\boldsymbol{\Theta}$ è l'insieme dei valori ammissibili di $\boldsymbol{\theta}$

È possibile stimare la varianza $\lambda^2$ di $e(t)$ come:

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

Uno stimatore corretto richiederebbe di dividere per $N-d$, dove $d$ è il numero di parametri

# L’approccio predittivo all’identificazione

**Osservazioni**

- $\Theta$: insieme dei **valori ammissibili** per $\boldsymbol{\theta}$. Per esempio, se $\mathcal{M}(\boldsymbol{\theta}) = \mathrm{ARMAX}$, voglio solo i $\boldsymbol{\theta}$ tali per cui $C(z)/A(z)$ è canonico

- Per valori iniziali $t = 1,2, \dots$ il predittore potrebbe non avere dati disponibili. Si usa quindi un’ **inizializzazione «convenzionale»** (ipotizzo che i valori passati di $y(\quad)$ siano nulli). L’inizializzazione non è un problema in quanto il **predittore è stabile**

  ✓ In alternativa, **scarto quei dati iniziali** che non hanno una predizione associata

# L'approccio predittivo all'identificazione

- Abbiamo già in parte visto che l'errore di predizione a un passo $\varepsilon_1(t;\boldsymbol{\theta})$ gode di interessanti proprietà, che ci permettono di **distinguere** $\boldsymbol{\theta}^0$ da un $\boldsymbol{\theta}$ qualsiasi

1. Dato $\boldsymbol{\theta}$ e i dati $\{u(1),\dots,u(N)\},\{y(1),\dots,y(N)\}$, è sempre possibile calcolare $\varepsilon_1(t;\boldsymbol{\theta})$
2. Se $\exists\boldsymbol{\theta}=\boldsymbol{\theta}^0$ t.c. $G_0(z)=G(z,\boldsymbol{\theta}^0)$ e $H_0(z)=H(z,\boldsymbol{\theta}^0)$, abbiamo che $\varepsilon_1(t;\boldsymbol{\theta}^0)=e(t)$, ovvero ci permette di capire se il modello è buono
3. $\varepsilon_1(t;\boldsymbol{\theta})\neq e(t)$ per qualsiasi $\boldsymbol{\theta}\neq\boldsymbol{\theta}^0$ (supponendo di avere un *ingresso adeguato*)
4. $\boldsymbol{\theta}^0$ minimizza la varianza dell'errore di predizione a un passo

# L’approccio predittivo all’identificazione

I metodi di stima basati sulla minimizzazione dell’errore di predizione prendono il nome di **Prediction Error Methods (PEM)**

**Nota**

Se ipotizzo che $\mathcal{S} = \mathcal{M}(\boldsymbol{\theta}^0)$ e $e(t) \sim \mathrm{WN}$ Gaussiano, lo **stimatore PEM è circa uguale allo stimatore a massima verosimiglianza**

La differenza sta in come i due approcci trattano l’inizializzazione del predittore: la funzione di verosimiglianza «propriamente detta» sarebbe più difficile da trattare rispetto a quella **«condizionata»** ai valori iniziali, per la quale c’è l’equivalenza coi metodi PEM

Se i **dati sono molti, non c’è differenza**

# Outline

1. Introduzione all’identificazione dei modelli dinamici
2. Metodi a minimizzazione dell’errore di predizione (PEM)
3. **Identificazione PEM di modelli ARX**
4. Identificazione PEM di modelli ARMAX

# Identificazione PEM di modelli ARX

Consideriamo un **modello** $\mathrm{ARX}(n_a, n_b, 1)$, e di avere a disposizione $N$ dati $\{u(1), \dots, u(N)\}, \{y(1), \dots, y(N)\}$

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

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

## Osservazioni

- $C(z) = 1$ poiché non c'è la parte $\mathrm{MA}$
- Abbiamo supposto che $k = 1$. Fissando un ritardo unitario **non lediamo di generalità**. Per esempio, se il ritardo vero fosse $k = 2$, stimeremmo $b_0 = 0$

# Identificazione PEM di modelli ARX

Il modello in **forma di predittore** è dato da

$$\widehat{\mathcal{M}}(\boldsymbol{\theta}):\ \hat{y}(t|t-1;\boldsymbol{\theta}) = H^{-1}(z,\boldsymbol{\theta})G(z,\boldsymbol{\theta})u(t) + [1 - H^{-1}(z,\boldsymbol{\theta})]y(t)$$

$$= B(z,\boldsymbol{\theta})u(t-1) + [1 - A(z,\boldsymbol{\theta})]y(t)$$

$$= \left(b_0 + b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}\right)u(t-1) + \left(a_1 z^{-1} + \cdots + a_{n_a} z^{-n_a}\right)y(t)$$

$$\Downarrow$$

$$\hat{y}(t|t-1;\boldsymbol{\theta}) \ = b_o u(t-1) + b_1 u(t-2) + \cdots + b_{n_b} u(t-n_b-1) + a_1 y(t-1) + \cdots + a_{n_a} y(t-n_a)$$

# Identificazione PEM di modelli ARX

Definendo i vettori

$$\underset{\substack{(n_a+n_b+1)\times 1 \\ d\times 1}}{\boldsymbol{\theta}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \\ b_0 \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix} \qquad \underset{\substack{(n_a+n_b+1)\times 1 \\ d\times 1}}{\boldsymbol{\varphi}(t)} = \begin{bmatrix} y(t-1) \\ y(t-2) \\ \vdots \\ y(t-n_a) \\ u(t-1) \\ u(t-2) \\ \vdots \\ u(t-n_b-1) \end{bmatrix}$$

Possiamo scrivere

$$\mathcal{M}(\boldsymbol{\theta}): y(t) = \boldsymbol{\varphi}^\top(t)\boldsymbol{\theta} + e(t)$$

$$\widehat{\mathcal{M}}(\boldsymbol{\theta}): \hat{y}(t|t-1;\boldsymbol{\theta}) = \boldsymbol{\varphi}^\top(t)\boldsymbol{\theta} \quad \Rightarrow$$

Il predittore è **lineare nei parametri $\theta$!**

# Identificazione PEM di modelli ARX

Per trovare la stima PEM minimizziamo la funzione di costo

$$J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N} \varepsilon_1(t;\boldsymbol{\theta})^2 \quad = \frac{1}{N}\sum_{t=1}^{N}\big(y(t) - \hat{y}(t|t-1;\boldsymbol{\theta})\big)^2 \quad \boxed{= \frac{1}{N}\sum_{t=1}^{N}(y(t) - \underset{1\times d}{\boldsymbol{\varphi}^\top(t)}\underset{d\times 1}{\boldsymbol{\theta}})^2}$$

La soluzione è analoga alla **stima a minimi quadrati** di un modello lineare statico!

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

Se $S(N) = \sum_{t=1}^{N} \boldsymbol{\varphi}(t)\boldsymbol{\varphi}^\top(t)$ è invertibile, allora la soluzione $\hat{\theta}_N$ è **unica** in quanto la funzione di costo è convessa

# Identificazione PEM di modelli ARX

Come per la regressione lineare, posso esprimere il modello ARX in forma matriciale

$$\underset{N \times d}{\Phi} = \begin{bmatrix} \boldsymbol{\varphi}^\top(1) \\ \boldsymbol{\varphi}^\top(2) \\ \vdots \\ \boldsymbol{\varphi}^\top(N) \end{bmatrix} \qquad \underset{\substack{(n_a+n_b+1)\times 1 \\ d \times 1}}{\boldsymbol{\theta}} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n_a} \\ b_0 \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix} \qquad \underset{N \times 1}{Y} = \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(N) \end{bmatrix} \qquad \underset{N \times 1}{E} = \begin{bmatrix} e(1) \\ e(2) \\ \vdots \\ e(N) \end{bmatrix}$$

Da cui segue che

$$\underset{N \times 1}{Y} = \underset{N \times d}{\Phi} \underset{d \times 1}{\boldsymbol{\theta}} + \underset{N \times 1}{E} \quad \Rightarrow \quad \underset{d \times 1}{\hat{\boldsymbol{\theta}}_N} = \underset{d \times d}{(\Phi^\top \Phi)^{-1}} \underset{d \times N}{\Phi^\top} \underset{N \times 1}{Y}$$

# Esempio: stima di un modello AR(1)

Si supponga di avere 5 dati da una serie temporale $y(t)$, stazionaria e ergodica, a media nulla

$$y(1) = \frac{1}{2} \qquad y(2) = 0 \qquad y(3) = -1 \qquad y(4) = -\frac{1}{2} \qquad y(5) = \frac{1}{4}$$

**Identificare** un modello $\mathrm{AR}(1)$ del tipo $y(t) = ay(t-1) + e(t), \;\; e(t) \sim \mathrm{WN}(0, \lambda^2)$

Usando il modello identificato, si calcoli la **predizione** $\hat{y}(6|5)$ e la **varianza** del rumore $\hat{\lambda}^2$

**Nota**

La media campionaria $\hat{m} = \frac{1}{5}\sum_{t=1}^{5} y(i) = -0{,}15$ non è nulla. Il problema però ci dice di considerare una media nulla. In caso contrario, avremmo dovuto depolarizzare il processo al fine di avere un predittore corretto, tale che $\mathbb{E}[\varepsilon_1(t)] = 0$

# Esempio: stima di un modello AR(1)

## Calcoliamo il predittore ad un passo

Supponiamo che il modello sia in forma canonica (ovvero che $|a| < 1$)

$$y(t) = \frac{1}{1 - az^{-1}} e(t) \quad \Rightarrow \quad \hat{y}(t|t-1;a) = \frac{C(z) - A(z)}{C(z)} y(t) = \frac{1 - 1 + az^{-1}}{1} y(t) \boxed{= ay(t-1)}$$

## Calcoliamo la funzione di costo

Abbiamo due alternative: o inizializzo i valori mancanti del predittore (per esempio $\hat{y}(1|0)$) a zero, oppure parto da $t = 2$ fino a $t = 5$. Scegliamo questa seconda strada (per tanti dati il risultato non cambia)

# Esempio: stima di un modello AR(1)

$$y(1) = \frac{1}{2} \qquad y(2) = 0 \qquad y(5) = \frac{1}{4}$$

$$y(3) = -1 \qquad y(4) = -\frac{1}{2}$$

$$J_5(\boldsymbol{\theta}) = \frac{1}{5-1}\sum_{t=2}^{5}\big(y(t) - ay(t-1)\big)^2$$

$$= \frac{1}{4}\Big[\big(y(2) - ay(1)\big)^2 + \big(y(3) - ay(2)\big)^2 + \big(y(4) - ay(3)\big)^2 + \big(y(5) - ay(4)\big)^2\Big]$$

$$= \frac{1}{4}\left[\left(0 - a\frac{1}{2}\right)^2 + (-1 - a\cdot 0)^2 + \left(-\frac{1}{2} + a\cdot 1\right)^2 + \left(\frac{1}{4} + a\frac{1}{2}\right)^2\right]$$

$$= \frac{1}{4}\left[\frac{1}{4}a^2 + 1 + \frac{1}{4} + a^2 - a + \frac{1}{16} + \frac{1}{4}a^2 + \frac{1}{4}a\right]$$

# Esempio: stima di un modello AR(1)

$$= \frac{1}{4}\left[\frac{16+4+1}{16} + \frac{-4a+a}{4} + \frac{a^2+a^2+4a^2}{4}\right] = \frac{1}{4}\left[\frac{21}{16} - \frac{3}{4}a + \frac{3}{2}a^2\right]$$

**Minimizziamo la funzione di costo**

$$\frac{dJ_5(a)}{da} = 0 \quad \Rightarrow \quad 3a - \frac{3}{4} = 0 \quad \Rightarrow \quad \boxed{\hat{a}_5 = \frac{1}{4}}$$

Se avessimo ottenuto $|\hat{a}_5| > 1$, avremmo potuto usare un filtro passa-tutto per rendere il modello stabile

# Esempio: stima di un modello AR(1)

**Stimiamo la varianza del rumore**

$$\hat{\lambda}^2 = J_5(\hat{a}_5) = \frac{1}{4}\left[\frac{21}{16} - \frac{3}{4}\cdot\frac{1}{4} + \frac{3}{2}\cdot\left(\frac{1}{4}\right)^2\right] \approx \boxed{0.3}$$

**Predizione a un passo** $\hat{y}(6|5)$

Notiamo che $\lambda^2$ non è molto importante: infatti non mi serve per calcolare la predizione

$$\hat{y}(t|t-1;\hat{a}_5) = \hat{a}_5 y(t-1) \quad \Rightarrow \quad \hat{y}(6|5) = \frac{1}{4}y(5) = \boxed{\frac{1}{16}}$$

# Esempio: stima di un modello AR(1)

**Modello identificato**

$$y(t) = \frac{1}{1 - \frac{1}{4} z^{-1}} e(t), \qquad e(t) \sim \mathrm{WN}(0, 0.3)$$

# Outline

1. Introduzione all’identificazione dei modelli dinamici
2. Metodi a minimizzazione dell’errore di predizione (PEM)
3. Identificazione PEM di modelli ARX
4. **Identificazione PEM di modelli ARMAX**

# Identificazione PEM di modelli ARMAX

Consideriamo un **modello** $\mathrm{ARMAX}(n_a, n_c, n_b, 1)$ , e di avere a disposizione $N$ dati $\{u(1), \dots, u(N)\}, \{y(1), \dots, y(N)\}$

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

- $B(z) = b_0 + b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}$
- $A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_{n_a} z^{-n_a}$
- $C(z) = 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}$

Il vettore dei parametri, in questo caso, è:

$$\underset{\substack{(n_a+n_b+1+n_c)\times 1 \\ d\times 1}}{\boldsymbol{\theta}} = \begin{bmatrix} a_1 \cdots a_{n_a} & b_0\, b_1 \cdots b_{n_b} & c_1 \cdots c_{n_c} \end{bmatrix}^\top$$

# Identificazione PEM di modelli ARMAX

Calcoliamo l'espressione dell'errore di predizione ad un passo. In questo caso, si ha che $E(z) = 1$, e quindi $\varepsilon_1(t) = e(t)$. Di conseguenza, esprimendo $e(t)$ in funzione di $u(t)$ e $y(t)$,

$$\varepsilon_1(t;\boldsymbol{\theta}) = e(t) = \frac{A(z,\boldsymbol{\theta})}{C(z,\boldsymbol{\theta})}y(t) - \frac{B(z,\boldsymbol{\theta})}{C(z,\boldsymbol{\theta})}u(t-1)$$

Potevamo ottenere la stessa cosa anche con l'espressione generica di $\varepsilon_1(t)$:

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

$$= \frac{A(z)}{C(z)}\left[y(t) - \frac{B(z)}{A(z)}u(t-1)\right] \quad = \frac{A(z)}{C(z)}y(t) - \frac{B(z)}{C(z)}u(t-1)$$

# Identificazione PEM di modelli ARMAX

Utilizziamo l'approccio predittivo

$$J_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{t=1}^{N}\varepsilon_1(t;\boldsymbol{\theta})^2 \quad \boxed{= \frac{1}{N}\sum_{t=1}^{N}\left[\frac{A(z,\boldsymbol{\theta})}{C(z,\boldsymbol{\theta})}y(t) - \frac{B(z,\boldsymbol{\theta})}{C(z,\boldsymbol{\theta})}u(t-1)\right]^2}$$

**Osservazioni**

- Dato che ho $C(z,\boldsymbol{\theta})$ al denominatore, questa funzione di costo **non è più convessa**! In generale, avrò dei **minimi locali**

- Per la risoluzione del problema di minimizzazione, devo utilizzare dei **metodi iterativi** (per esempio il **metodo del gradiente**)

# Identificazione PEM di modelli ARMAX

## Come gestire i minimi locali

Una strategia semplice è la seguente. Data una inizializzazione $\hat{\boldsymbol{\theta}}^{(0)}$ all'iterazione $0$:

- Scegliamo $N_{\text{init}}$ **inizializzazioni $\hat{\boldsymbol{\theta}}^{(0)}$ diverse**, ottenendo $N_{\text{init}}$ soluzioni
- Se le $N_{\text{init}}$ soluzioni sono **uguali**, posso pensare (non sono certo) di aver raggiunto il minimo globale di $J_N(\boldsymbol{\theta})$
- Se le $N_{\text{init}}$ soluzioni sono **diverse**, considero quella che mi ha dato $J_N(\boldsymbol{\theta})$ minore

Come alternativa (più efficiente) al metodo del gradiente, vedremo il metodo di **ottimizzazione iterativo** noto come **Metodo di Newton**. Questo metodo, oltre al gradiente, sfrutta anche l'informazione data dalla **matrice Hessiana**

# Identificazione PEM di modelli ARMAX

## METODO DI NEWTON

**Idea:** sviluppo in serie di Taylor troncata al 2° ordine di $J_N(\boldsymbol{\theta})$, nell'intorno della stima all'iterazione $i$-esima $\widehat{\boldsymbol{\theta}}^{(i)}$

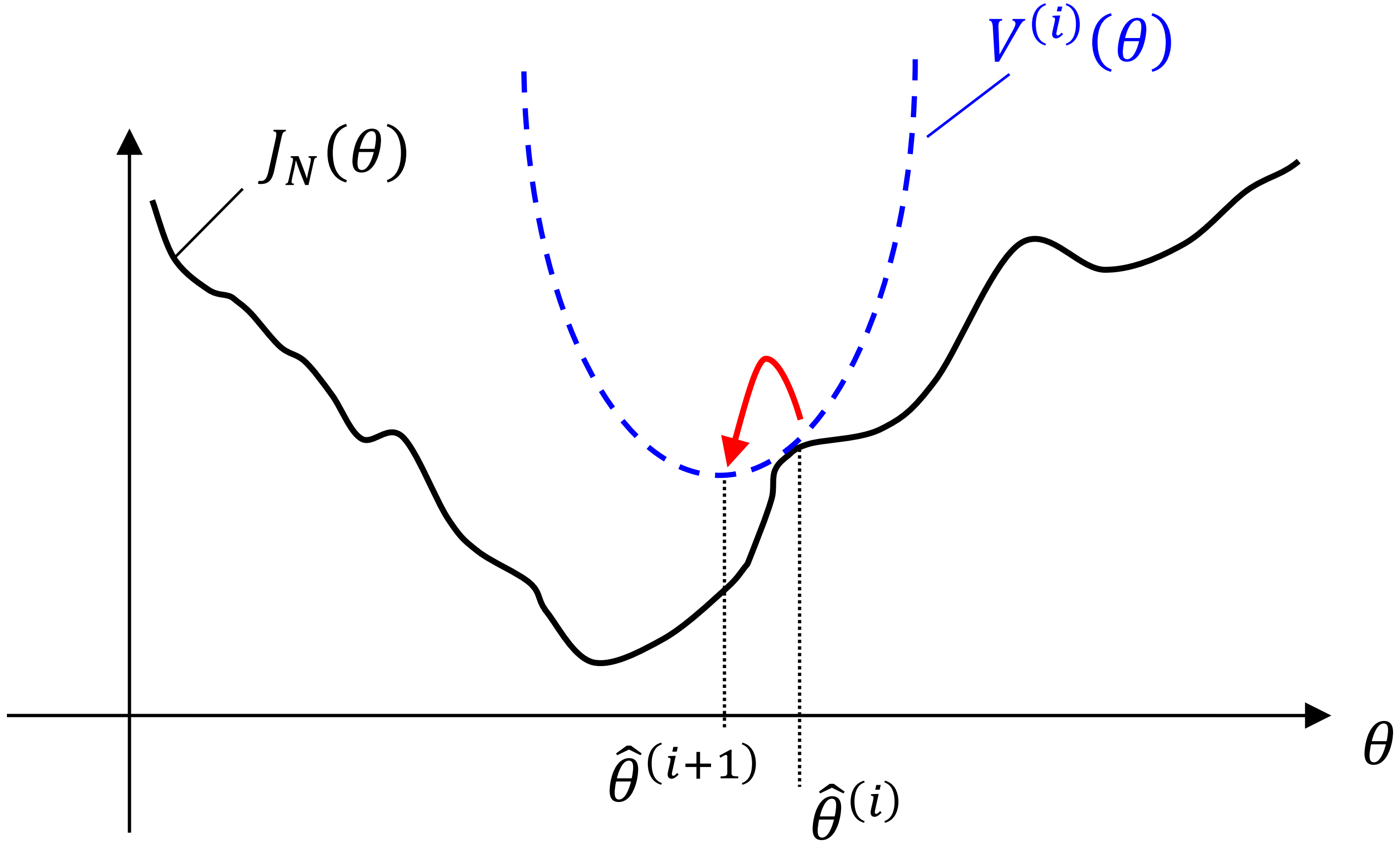
$J_N(\boldsymbol{\theta}) \approx V(\boldsymbol{\theta})$ ⟶ La funzione $V(\boldsymbol{\theta})$ è un **paraboloide** (è facile calcolarne il minimo)

$$\underset{1\times 1}{V^{(i)}(\boldsymbol{\theta})} = J_N\left(\widehat{\boldsymbol{\theta}}^{(i)}\right) + \underset{1\times d}{\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(i)}\right)^{\top}} \cdot \underbrace{\underset{d\times 1}{\left.\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}}}_{\textbf{Gradiente}} + \frac{1}{2}\underset{1\times d}{\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(i)}\right)^{\top}} \cdot \underbrace{\underset{d\times d}{\left.\frac{d^2J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}}}_{\textbf{Matrice Hessiana}} \cdot \underset{d\times 1}{\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(i)}\right)}$$

# Identificazione PEM di modelli ARMAX

Una volta ottenuta l'approssimazione $V^{(i)}(\boldsymbol{\theta})$, si calcola $\hat{\boldsymbol{\theta}}^{(i+1)}$ come il **minimo** di $V^{(i)}(\boldsymbol{\theta})$.

Consideriamo un caso scalare per semplicità



# Identificazione PEM di modelli ARMAX

$\nabla_{\boldsymbol{x}}(\boldsymbol{x}^\top \boldsymbol{b}) = \boldsymbol{b}$

$\nabla_{\boldsymbol{x}}(\boldsymbol{x}^\top A\boldsymbol{x}) = (A + A^\top)\boldsymbol{x}$

Troviamo un'espressione esplicita per $\widehat{\boldsymbol{\theta}}^{(i+1)}$ imponendo $\underset{d \times 1}{\dfrac{dV^{(i)}(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \boldsymbol{0}}$

$$\underset{d \times 1}{\frac{dV^{(i)}(\boldsymbol{\theta})}{d\boldsymbol{\theta}}} = \underset{d \times 1}{\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}} + \frac{1}{2} \cdot 2 \underset{d \times d}{\frac{d^2 J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}} \cdot \underset{d \times 1}{\left(\boldsymbol{\theta} - \widehat{\boldsymbol{\theta}}^{(i)}\right)} = \underset{d \times 1}{\boldsymbol{0}} \quad \Rightarrow$$

Ricavo il minimo e lo chiamo $\widehat{\boldsymbol{\theta}}^{(i+1)}$

**Regola di update per il metodo di Newton**

$$\Rightarrow \quad \underset{d \times 1}{\widehat{\boldsymbol{\theta}}^{(i+1)}} = \underset{d \times 1}{\widehat{\boldsymbol{\theta}}^{(i)}} - \underset{d \times d}{\left[\frac{d^2 J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}\right]^{-1}} \cdot \underset{d \times 1}{\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}}\Bigg|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}$$

È simile al gradient descent se al posto dell'Hessiana metto la learning rate $\alpha$

# Identificazione PEM di modelli ARMAX

Dobbiamo quindi calcolare queste due quantità:

$$\left.\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}$$ **Gradiente** di $J_N(\boldsymbol{\theta})$

$$\left.\frac{d^2J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}\right|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}$$ **Hessiano** di $J_N(\boldsymbol{\theta})$

**Calcolo di** $\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$

$$\underbrace{\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}}_{d\times 1} = \frac{d}{d\boldsymbol{\theta}}\left[\frac{1}{N}\sum_{t=1}^{N}\varepsilon_1(t;\boldsymbol{\theta})^2\right] = \frac{1}{N}\sum_{t=1}^{N}\frac{d}{d\boldsymbol{\theta}}\varepsilon_1(t;\boldsymbol{\theta})^2 = \boxed{\frac{2}{N}\sum_{t=1}^{N}\varepsilon_1(t;\boldsymbol{\theta})\cdot\underbrace{\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}}_{d\times 1}}$$

# Identificazione PEM di modelli ARMAX

$$\frac{d(v\mathbf{u})}{d\mathbf{x}} = v \cdot \frac{d\mathbf{u}}{d\mathbf{x}} + \frac{dv}{d\mathbf{x}} \mathbf{u}^{\top}$$

Nel nostro caso:

$v = \varepsilon_1(t;\boldsymbol{\theta})$

$\mathbf{u} = \dfrac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}$

$\mathbf{x} = \boldsymbol{\theta}$

**Calcolo di** $\dfrac{d^2 J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}$

$$\underbrace{\frac{d^2 J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}}_{d \times d} = \frac{d}{d\boldsymbol{\theta}} \frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}} = \frac{d}{d\boldsymbol{\theta}} \left[ \frac{2}{N} \sum_{t=1}^{N} \varepsilon_1(t;\boldsymbol{\theta}) \cdot \underbrace{\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}}_{d \times 1} \right]$$

Derivata del prodotto

$$= \frac{2}{N} \sum_{t=1}^{N} \underbrace{\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}} \cdot \frac{d\varepsilon_1(t;\boldsymbol{\theta})^{\top}}{d\boldsymbol{\theta}}}_{d \times d} + \frac{2}{N} \sum_{t=1}^{N} \varepsilon_1(t;\boldsymbol{\theta}) \cdot \underbrace{\frac{d^2 \varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}}_{d \times d}$$

# Identificazione PEM di modelli ARMAX

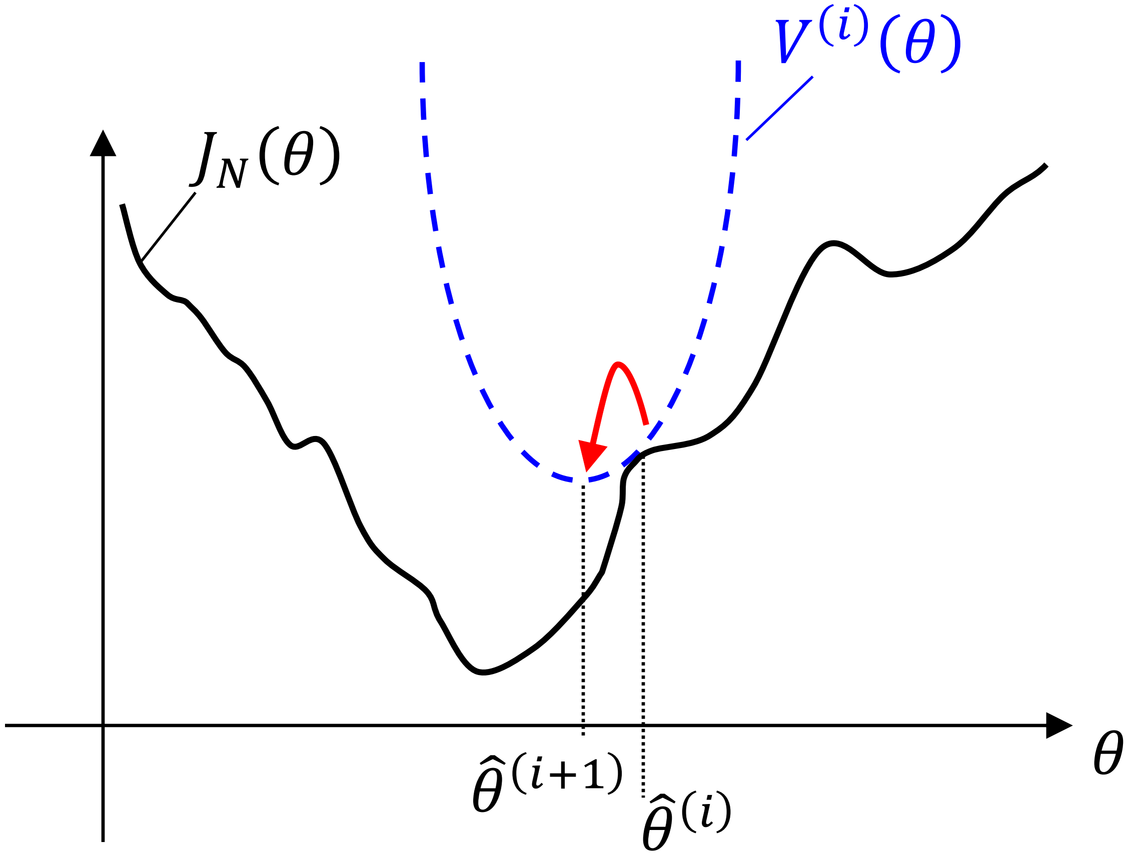
$$\underbrace{\frac{d^2 J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}}_{d \times d} = \frac{2}{N}\sum_{t=1}^{N} \underbrace{\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}} \cdot \frac{d\varepsilon_1(t;\boldsymbol{\theta})^\top}{d\boldsymbol{\theta}}}_{d \times d} + \cancel{\frac{2}{N}\sum_{t=1}^{N} \varepsilon_1(t;\boldsymbol{\theta}) \cdot \underbrace{\frac{d^2\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}}_{d \times d}}$$

**Ignoriamo** il secondo termine, **approssimando l'Hessiana**, dato che:

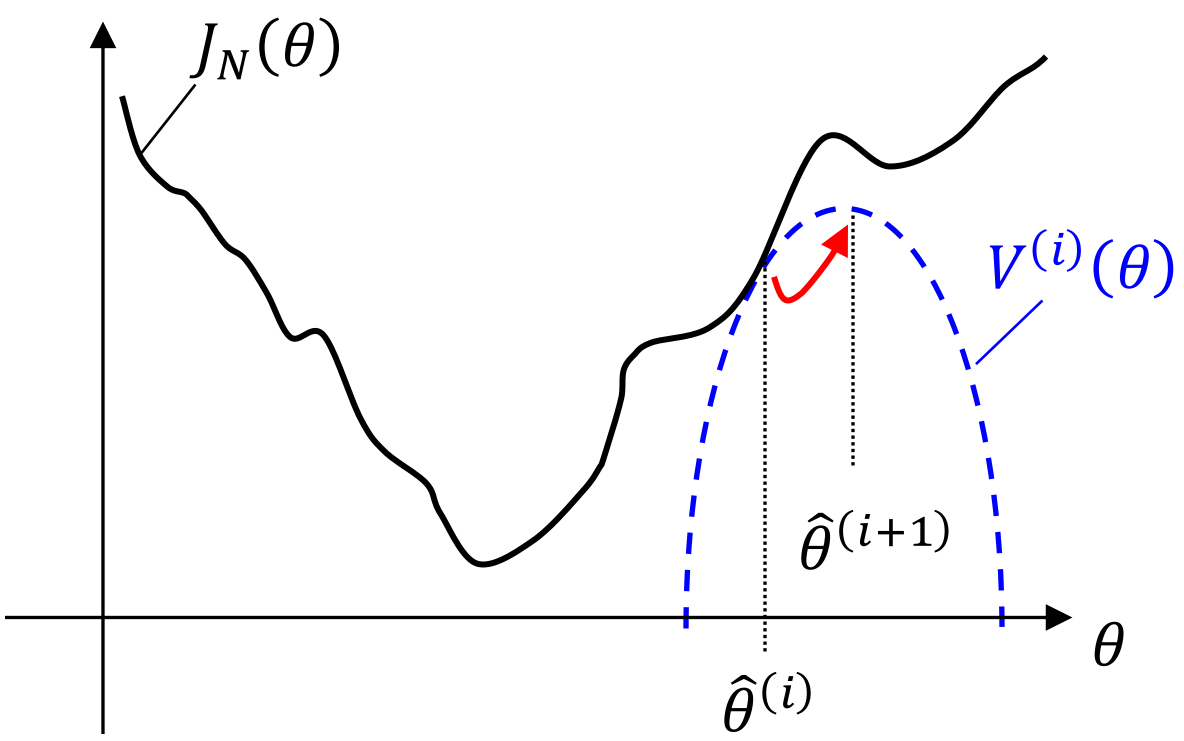
- Se siamo **vicini all'ottimo**, $\varepsilon_1(t;\boldsymbol{\theta})$ è **«piccolo»** e il termine «conta poco»
- Possiamo **evitare di calcolare** $\dfrac{d^2\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}$
- Ci assicuriamo una Hessiana **semi-definita positiva**. In questo modo, la direzione dell'algoritmo è **sicuramente discendente** (concetto simile ad avere learning rate $\geq 0$)

# Identificazione PEM di modelli ARMAX

**Hessiana $\succ 0$**




**Hessiana $\prec 0$**



# Identificazione PEM di modelli ARMAX

## Osservazione

L'aggiornamento da $\hat{\boldsymbol{\theta}}^{(i)}$ a $\hat{\boldsymbol{\theta}}^{(i+1)}$, in generale, può essere fatto con tre categorie di metodi:

1. Metodo del **gradiente**
2. Metodo di **Newton**
3. Metodi di «**quasi-Newton**»

*Metodo del gradiente*

$$\underset{d\times 1}{\hat{\boldsymbol{\theta}}^{(i+1)}} = \hat{\boldsymbol{\theta}}^{(i)} - \underset{1\times 1}{\alpha} \cdot \left[ \underset{d\times 1}{\frac{dJ_N(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}} \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(i)}} \right]$$

- **Semplice** e **robusto** (va sempre nella direzione di discesa)
- Può essere **molto lento** quando ci avviciniamo al minimo (poiché il gradiente tende a 0)

# Identificazione PEM di modelli ARMAX

*Metodo di Newton*

L'Hessiana «modula» il passo del gradiente

$$\underset{d\times 1}{\hat{\boldsymbol{\theta}}^{(i+1)}} = \underset{d\times 1}{\hat{\boldsymbol{\theta}}^{(i)}} - \underset{d\times d}{\left[\frac{d^2 J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(i)}}\right]^{-1}} \cdot \underset{d\times 1}{\frac{dJ_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(i)}}}$$

- Converge **velocemente**
- Computazionalmente più **complesso**
- Rischio di **instabilità** se l'Hessiana è definita negativa

*Metodi «quasi Newtoniani»*

$O^{-1}$ è un'approssimazione dell'Hessiana semidefinita positiva o definita positiva

$$\underset{d\times 1}{\hat{\boldsymbol{\theta}}^{(i+1)}} = \hat{\boldsymbol{\theta}}^{(i)} - \underset{1\times 1}{O^{-1}} \cdot \underset{d\times 1}{\left[\frac{dJ_N(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(i)}}\right]}$$

- **Più semplice** del metodo di Newton
- **Più veloce** del metodo del gradiente
- **Non c'è rischio** di allontanarsi dal minimo
- **Non è veloce** come il metodo di Newton

# Identificazione PEM di modelli ARMAX

I metodi «quasi Newtoniani» sono molto usati e differiscono fra loro nel **modo in cui viene fatta l'approssimazione** definita positiva dell'Hessiana

Per **garantire l'invertibilità** di $O^{-1} \succcurlyeq 0$, si aggiunge un termine positivo, molto piccolo, di **«regolarizzazione»**

$$\frac{d^2 J_N(\boldsymbol{\theta})}{d\boldsymbol{\theta}^2} \approx \frac{2}{N} \sum_{t=1}^{N} \frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}} \cdot \frac{d\varepsilon_1(t;\boldsymbol{\theta})^\top}{d\boldsymbol{\theta}} + \delta I_d$$

# Identificazione PEM di modelli ARMAX

Dopo aver introdotto l'approssimazione dell'Hessiana, la **regola di update** diventa:

$$\underset{d\times 1}{\hat{\boldsymbol{\theta}}^{(i+1)}} = \underset{d\times 1}{\hat{\boldsymbol{\theta}}^{(i)}} - \underbrace{\left[\frac{2}{N}\sum_{t=1}^{N}\left.\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(i)}} \cdot \left.\frac{d\varepsilon_1(t;\boldsymbol{\theta})^\top}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(i)}}\right]^{-1}}_{d\times d} \cdot \underbrace{\left[\frac{2}{N}\sum_{t=1}^{N}\varepsilon_1\left(t;\hat{\boldsymbol{\theta}}^{(i)}\right)\cdot\left.\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^{(i)}}\right]}_{d\times 1}$$

- Notiamo che i termini $2/N$ si possono semplificare

# Identificazione PEM di modelli ARMAX

**Calcoliamo** $\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}$

Ricordiamo che $\varepsilon_1(t;\boldsymbol{\theta}) = e(t) = \frac{A(z,\boldsymbol{\theta})}{C(z,\boldsymbol{\theta})}y(t) - \frac{B(z,\boldsymbol{\theta})}{C(z,\boldsymbol{\theta})}u(t-1)$

$$\varepsilon_1(t;\boldsymbol{\theta}) = \frac{1 - a_1 z^{-1} - \cdots - a_{n_a} z^{-n_a}}{1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}} y(t) - \frac{b_0 + b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}}{1 + c_1 z^{-1} + \cdots + c_{n_c} z^{-n_c}} u(t-1)$$

$$\underset{d\times 1}{\boldsymbol{\theta}} = \begin{bmatrix} a_1 \cdots a_{n_a} & b_0\, b_1 \cdots b_{n_b} & c_1 \cdots c_{n_c} \end{bmatrix}^{\mathrm{T}}$$

# Identificazione PEM di modelli ARMAX

**Derivate di $\varepsilon_1(t;\boldsymbol{\theta})$ rispetto a $a_1, a_2, \dots, a_{n_a}$**

$$\frac{d\varepsilon_1(t)}{da_1} = -\frac{z^{-1}}{C(z)}y(t) = \alpha(t-1)$$

$$\frac{d\varepsilon_1(t)}{da_2} = -\frac{z^{-2}}{C(z)}y(t) = \alpha(t-2)$$

$$\vdots$$

$$\frac{d\varepsilon_1(t)}{da_{n_a}} = -\frac{z^{-n_a}}{C(z)}y(t) = \alpha(t-n_a)$$

$$\boxed{\alpha(t) \equiv -\frac{1}{C(z)}y(t)}$$

# Identificazione PEM di modelli ARMAX

**Derivate di $\varepsilon_1(t;\boldsymbol{\theta})$ rispetto a $b_0, b_1, \dots, b_{n_b}$**

$$\frac{d\varepsilon_1(t)}{db_0} = -\frac{1}{C(z)}u(t-1) = \beta(t-1)$$

$$\frac{d\varepsilon_1(t)}{db_1} = -\frac{z^{-1}}{C(z)}u(t-1) = \beta(t-2)$$

$$\vdots$$

$$\frac{d\varepsilon_1(t)}{db_{n_b}} = -\frac{z^{-n_b}}{C(z)}u(t-1) = \beta(t-n_b-1)$$

$$\boxed{\beta(t) \equiv -\frac{1}{C(z)}u(t)}$$

# Identificazione PEM di modelli ARMAX

**Derivate di $\varepsilon_1(t;\boldsymbol{\theta})$ rispetto a $c_1, c_2, \dots, c_{n_c}$**

$$\varepsilon_1(t) = \frac{A(z)}{C(z)} y(t) - \frac{B(z)}{C(z)} u(t-1)$$

$$\Downarrow$$

$$\left(1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}\right)\varepsilon_1(t) = A(z)y(t) - B(z)u(t-1)$$

$$\frac{d\left[\left(1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}\right) \cdot \varepsilon_1(t)\right]}{dc_1} = \boxed{\frac{d[A(z)y(t) - B(z)u(t-1)]}{dc_1}}$$

Non dipende da $c_1$

$$\frac{d\left[\left(1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}\right) \cdot \varepsilon_1(t)\right]}{dc_1} = 0$$

Devo fare la derivata del prodotto

# Identificazione PEM di modelli ARMAX

**Derivate di $\varepsilon_1(t;\boldsymbol{\theta})$ rispetto a $c_1, c_2, \dots, c_{n_c}$**

$$\gamma(t) \equiv -\frac{1}{C(z)} \cdot \varepsilon_1(t)$$

$$\frac{d\left[\left(1 + c_1 z^{-1} + c_2 z^{-2} + \dots + c_{n_c} z^{-n_c}\right) \cdot \varepsilon_1(t)\right]}{dc_1} = 0$$

$$z^{-1} \cdot \varepsilon_1(t) + C(z)\frac{d\varepsilon_1(t)}{dc_1} = 0 \quad \Rightarrow \quad \frac{d\varepsilon_1(t)}{dc_1} = -\frac{1}{C(z)} \cdot \varepsilon_1(t-1) = \gamma(t-1)$$

$$\vdots$$

$$\frac{d\varepsilon_1(t)}{dc_{n_c}} = -\frac{1}{C(z)} \cdot \varepsilon_1(t-n_c) = \gamma(t-n_c)$$

# Identificazione PEM di modelli ARMAX

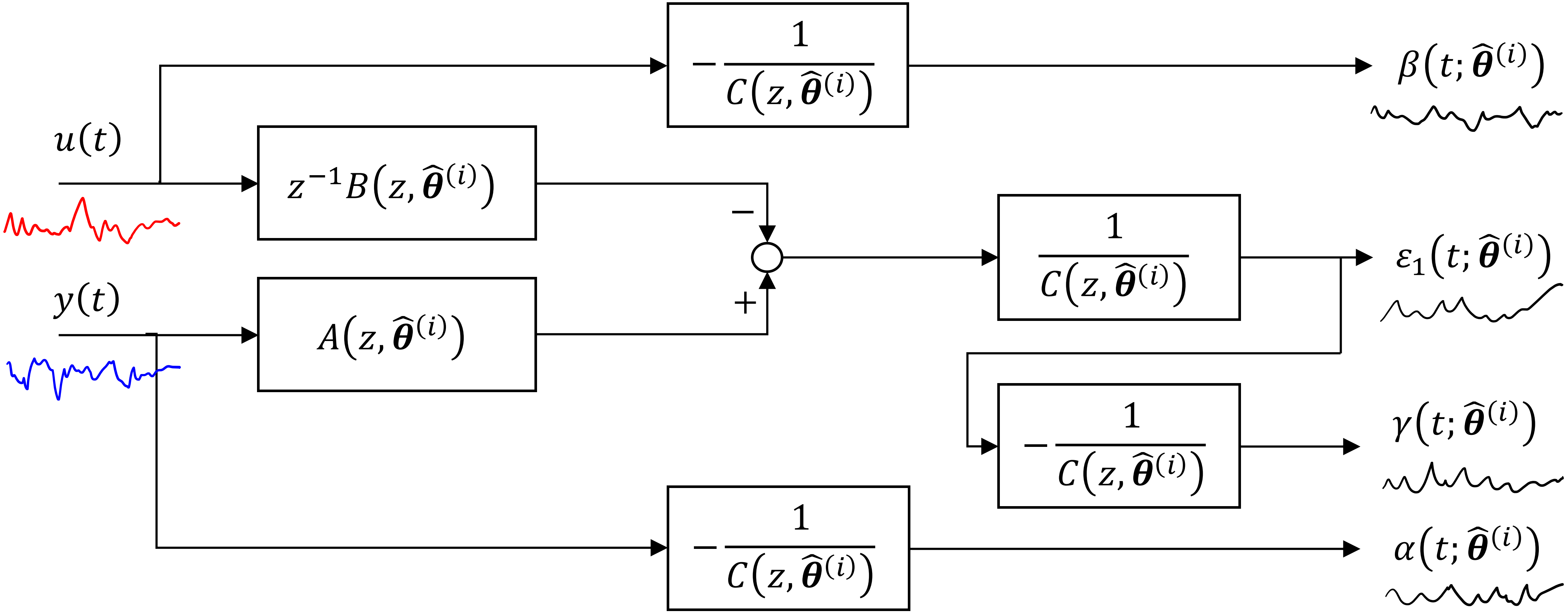
Riassumendo, il **vettore gradiente** è:

$$\underset{d \times 1}{\frac{d\varepsilon_1(t)}{d\boldsymbol{\theta}}} = \begin{bmatrix} \alpha(t-1) \\ \vdots \\ \alpha(t-n_a) \\ \beta(t-1) \\ \vdots \\ \beta(t-n_b-1) \\ \gamma(t-1) \\ \\ \gamma(t-n_c) \end{bmatrix} \qquad t = 1, \dots, N$$

È possibile definire in modo elegante il calcolo del gradiente tramite una serie di **filtraggi dei segnali di ingresso e uscita**

# Identificazione PEM di modelli ARMAX

Abbiamo il seguente **schema di filtraggio dei segnali** per trovare il gradiente



# Identificazione PEM di modelli ARMAX

## Riassunto dell'implementazione dell'algoritmo di Newton per modelli ARMAX

1. Calcolare i polinomi $A\left(z,\widehat{\boldsymbol{\theta}}^{(i)}\right), B\left(z,\widehat{\boldsymbol{\theta}}^{(i)}\right), C\left(z,\widehat{\boldsymbol{\theta}}^{(i)}\right)$ all'iterazione $i$-esima

2. Calcolare i segnali $\varepsilon_1\left(t;\widehat{\boldsymbol{\theta}}^{(i)}\right), \alpha\left(t;\widehat{\boldsymbol{\theta}}^{(i)}\right), \beta\left(t;\widehat{\boldsymbol{\theta}}^{(i)}\right), \gamma\left(t;\widehat{\boldsymbol{\theta}}^{(i)}\right)$ filtrando i dati $u, y$ disponibili

3. Costruire il vettore gradiente $\frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}$ coi segnali filtrati ricavati al passo 2.

4. Aggiornare la stima dei parametri tramite la regola di update

$$\widehat{\boldsymbol{\theta}}^{(i+1)} = \widehat{\boldsymbol{\theta}}^{(i)} - \left[\sum_{t=1}^{N} \frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}} \cdot \frac{d\varepsilon_1(t;\boldsymbol{\theta})^\top}{d\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}\right]^{-1} \cdot \left[\sum_{t=1}^{N} \varepsilon_1\left(t;\widehat{\boldsymbol{\theta}}^{(i)}\right) \cdot \frac{d\varepsilon_1(t;\boldsymbol{\theta})}{d\boldsymbol{\theta}}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}^{(i)}}\right]$$

# Identificazione PEM di modelli ARMAX

**Osservazione**

- Prima di filtrare tramite $1/C\left(z, \widehat{\boldsymbol{\theta}}^{(i)}\right)$, dobbiamo controllare che $C\left(z, \widehat{\boldsymbol{\theta}}^{(i)}\right)$ abbia radici interne al cerchio unitario. Se non è il caso, possiamo utilizzare un filtro passa-tutto per rendere $1/C\left(z, \widehat{\boldsymbol{\theta}}^{(i)}\right)$ asintoticamente stabile

Per le famiglie di modelli viste, vale che

- **ARX, FIR:** funzione di costo **convessa** (minimo globale)
- **ARMAX, BJ, OE:** funzione di costo **non convessa** (minimi locali)

UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Dipartimento
di Ingegneria Gestionale,
dell'Informazione e della Produzione