

# Data Model Identification and Intelligent Control

Stefano Di Lena

2025

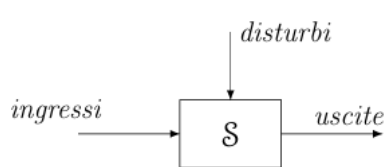


# Indice

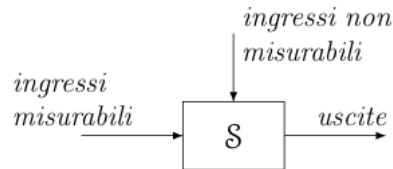
<b>1</b>	<b>Identificazione dei Sistemi</b>	<b>1</b>
1.1	Modelli matematici . . . . .	2
1.1.1	Predizione . . . . .	2
1.1.2	Simulazione . . . . .	2
1.1.3	Ottimizzazione . . . . .	3
1.1.4	Controllo . . . . .	3
1.1.5	Analisi . . . . .	3
1.1.6	Fault Detection . . . . .	3
<b>2</b>	<b>Famiglie di Modelli</b>	<b>4</b>
2.1	Box-Jenkins . . . . .	5
2.1.1	OE (Output Error) . . . . .	5
2.1.2	IE (Input Error) . . . . .	5
2.1.3	ARX (AutoRegressive with eXogenous variables) . . . . .	6
2.1.4	ARMAX . . . . .	6
2.1.5	ARARX . . . . .	7
<b>3</b>	<b>Criterio di Scelta</b>	<b>7</b>
3.1	Estrazione del predittore . . . . .	7
3.1.1	Predittore ad un passo . . . . .	7
3.1.2	ARX . . . . .	9
3.1.3	ARMAX . . . . .	9
<b>4</b>	<b>Validazione</b>	<b>9</b>
4.1	Analisi Asintotica . . . . .	9
4.2	Analisi di Bianchezza . . . . .	11
<b>5</b>	<b>Ottimizzazione ai Minimi Quadrati</b>	<b>11</b>
5.1	Analisi Probabilistica . . . . .	13
5.1.1	Persistente eccitazione . . . . .	14
5.2	Identificabilità . . . . .	16
5.3	Ordine Ottimo di un Modello . . . . .	16
5.4	Analisi Spettrale . . . . .	18
5.4.1	Pre-filtraggio . . . . .	19
5.5	Operazioni di Pre-processing sui Dati . . . . .	19

# 1 Identificazione dei Sistemi

*Identificare* significa trovare un modello matematico basandosi su **dati sperimentali**. In generale, un sistema è un insieme di componenti interagenti che produce una o più uscite, influenzate da ingressi e disturbi. Nel modello che lo descrive si distinguono, quindi, cause ed effetti.



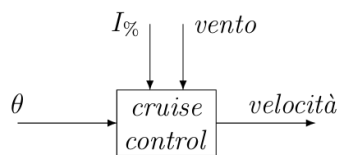
(a) modello per il controllo



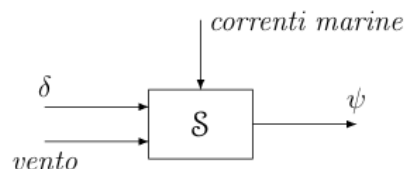
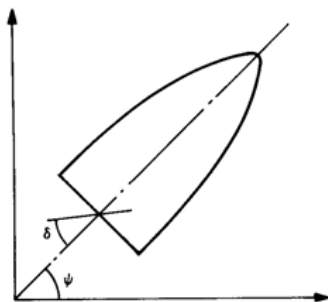
(b) modello per l'identificazione

I *disturbi* sono spesso cause non manipolabili (non sempre sono misurabili).

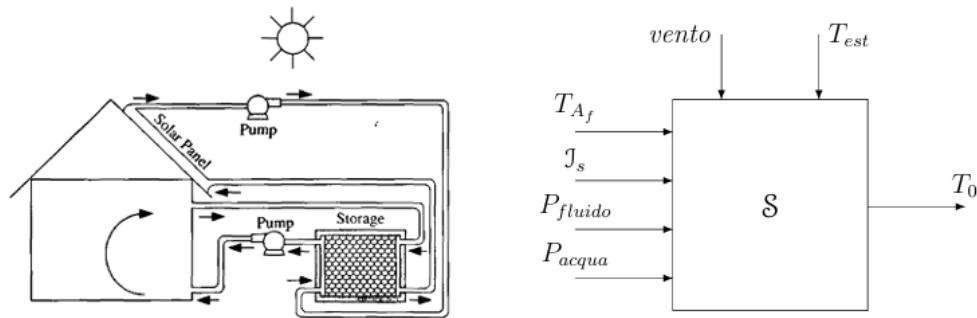
**Esempio 1: cruise control** Possiamo controllare la velocità  $v$  di un veicolo, tramite un ingresso  $\theta$  (angolo di apertura dal pedale acceleratore). Durante la marcia il sistema è sollecitato da una serie di disturbi: la pendenza della strada ( $I\%$ ), la presenza di vento, etc.



**Esempio 2: rotta veicolo marittimo** Avendo una barca a vela, gli ingressi misurabili possono essere  $\delta$  (angolo del timone, misurato rispetto alla direzione della rotta desiderata) ed il vento (considerato come spinta riesco ad ottenere). Le correnti marine sono perturbazioni del sistema e le considero come ingressi non misurabili. L'angolo  $\psi$  sarà l'uscita del sistema e fornisce la rotta finale.



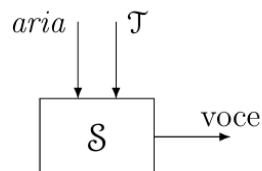
**Esempio 3: solare-termico** Un sistema per il controllo di accumulo dell'acqua calda, basato su energia solare.



La temperatura dell'acqua è l'uscita del sistema (valore che si vuole controllare).

Gli ingressi misurabili sono: l'irraggiamento solare  $J_s$ , la temperatura dell'acqua fredda  $T_A$ , la portata per il fluido di scambio e la portata in uscita dell'acqua calda. Un ingresso non misurabile potrebbe essere la temperatura dell'ambiente esterno.

**Esempio 4: voce** La voce umana è determinata dalla tensione delle corde vocali  $\mathcal{T}$  e l'aria. Questi sono però ingressi non misurabili, quindi il modello matematico prende il nome di *serie temporale* (perché non ha ingressi misurabili).



## 1.1 Modelli matematici

Un *buon* modello è un modello che ha una buona precisione e raggiunge l'obiettivo prefissato.

### 1.1.1 Predizione

Fornisce una predizione dell'uscita basandosi sulla storia precedente del sistema.

### 1.1.2 Simulazione

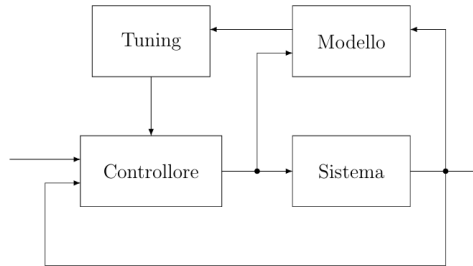
Uso la storia passata dell'ingresso per simulare l'uscita. Dato che non si hanno informazioni sull'uscita passata, i valori in uscita da questo modello non è detto coincidano con quelli veri.

### 1.1.3 Ottimizzazione

Si vuole ottenere un miglioramento dei risultati, controllando e reiterando eventualmente il processo.

### 1.1.4 Controllo

Si applicano delle leggi di controllo che riescano ad adattare i parametri alle modifiche del sistema avvenute durante lo svolgimento.

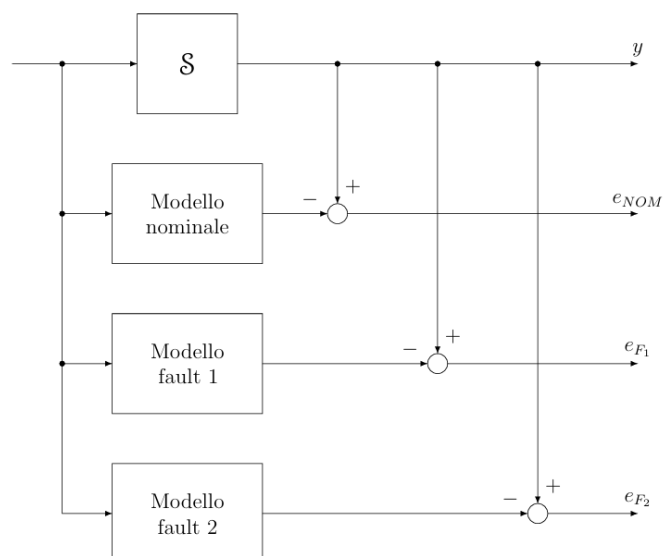


### 1.1.5 Analisi

Il sistema pone delle domande alle quali il modello fornisce delle potenziali risposte.

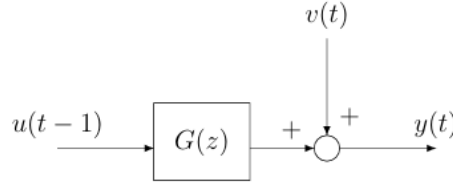
### 1.1.6 Fault Detection

Per effettuare la diagnosi dei guasti è usato questo tipo di modello. Vengono calcolati gli errori fra l'uscita del sistema e le uscite di ciascun modello, riuscendo ad ottenere una misura numerica di quanto si è vicini ad una condizione di guasto.



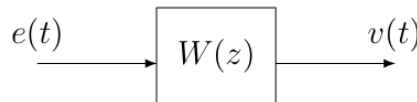
## 2 Famiglie di Modelli

Analizziamo sistemi LTI (Lineari Tempo Invarianti), SISO (Single-Input Single-Output), TD (Tempo Discreto), con funzione di trasferimento strettamente propria (grado del numeratore minore del grado del denominatore).



Nello schema in teoria dovrebbero stare  $Z^{-1}\{u(t-1)\}$  e  $Z^{-1}\{y(t)\}$ . È stato fatto un abuso di notazione per alleggerire la trattazione degli schemi.

In figura  $u(t)$  è detto *variabile esogena* (che proviene dall'esterno),  $v(t)$  *variabile di rumore* (ed è un segnale canonico in grado di descrivere i disturbi sconosciuti e non misurabili).



L'ingresso  $e(t)$  è un rumore bianco, segnale casuale per eccellenza. Esso ha uno spettro uniforme a tutte le frequenze ed è un esempio di segnale probabilistico **i.i.d.** (identicamente distribuito).

Se la f.d.t.  $W(z)$  è asintoticamente stabile, lo spettro del segnale in uscita è pari al prodotto tra il modulo al quadrato della funzione di risposta armonica discreta e la densità spettrale del segnale di ingresso:

$$\phi_v(\omega) = |W(e^{j\omega})|^2 \cdot \phi_e(\omega)$$

Possiamo esprimere una funzione di trasferimento in modo parametrico:

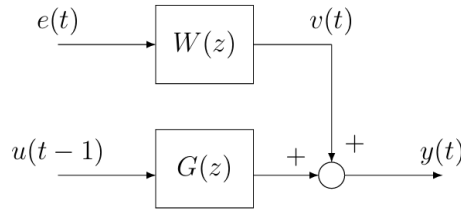
$$G(z) = k \frac{z^m + b_1 z^{m-1} + \dots + b_m}{z^n + a_1 z^{n-1} + \dots + a_n}$$

e non:

$$y(t) = \sum_{i=1}^{\infty} g(t-i)u(i)$$

Identificare un modello in forma *parametrica* consiste nell'identificazione dei coefficienti del numeratore e del denominatore della f.d.t.

## 2.1 Box-Jenkins



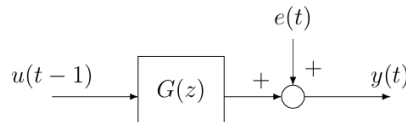
Il modello di Box-Jenkins rispetta le seguenti ipotesi:

1.  $W(z)$  deve essere asintoticamente stabile (quindi deve avere tutte le radici del polinomio al denominatore all'interno del cerchio di raggio unitario). I polinomi al numeratore e denominatore di questa f.d.t. devono essere monici (coefficiente di grado massimo uguale ad 1), co-primi (ovvero primi tra loro, il loro massimo comune divisore deve essere una costante non nulla), inoltre la funzione deve essere propria (num e den di ugual grado).
2.  $G(z)$  deve essere asintoticamente stabile, strettamente propria (grado del polinomio al denominatore deve essere maggiore di quello del numeratore).

Le equazioni del modello sono:

$$\begin{cases} y(t) = G(z) \cdot u(t-1) + W(z) \cdot e(t) \\ v(t) = W(z) \cdot e(t) \end{cases}$$

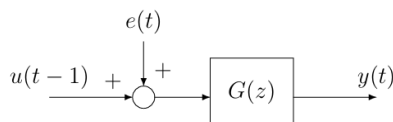
### 2.1.1 OE (Output Error)



Si assume una  $W(z)$  unitaria (uguale ad 1).

$$y(t) = G(z) \cdot u(t-1) + e(t)$$

### 2.1.2 IE (Input Error)

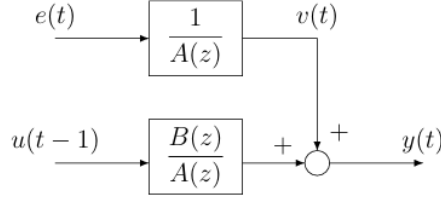


Si ha come fonte di incertezza l'errore sulla misura di ingresso (modello meno comune).

$$y(t) = G(z) \cdot [u(t-1) + e(t)]$$



### 2.1.3 ARX (AutoRegressive with eXogenous variables)



Sono modelli a due gradi di libertà la cui uscita è costituita dalla somma di tre componenti: un contributo *autoregressivo* sull'uscita (i cui campioni sono calcolati come combinazione lineare dei valori precedenti mediando con dei pesi), uno *esogeno* (è una media mobile, perché calcolo la media pesata dei campioni) ed il contributo puro della variabile d'errore (rumore bianco con media nulla e varianza  $\lambda^2$ ).

$$y(t) = \underbrace{a_1 y(t-1) + \dots + a_{n_a} y(t-n_a)}_{AR} + \underbrace{b_1 u(t-1) + \dots + b_{n_b} u(t-n_b)}_X + e(t)$$

Da cui, per la proprietà di traslazione nel tempo della Z-trasformata:

$$\underbrace{1 - a_1 z^{-1} - \dots - a_{n_a} z^{-n_a}}_{A(z)} \cdot y(t) = \underbrace{b_1 + b_2 z^{-1} + \dots + b_{n_b} z^{-n_b+1}}_{B(z)} \cdot u(t-1) + e(t)$$

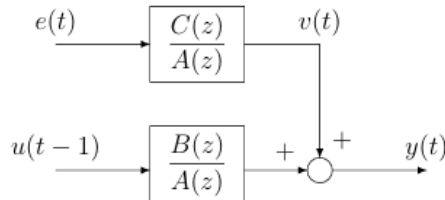
Quindi:

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

### 2.1.4 ARMAX

AutoRegressive Moving Average eXogenous model, hanno tre gradi di libertà, oltre ai contributi precedenti si aggiunge una media mobile per pesare l'errore.

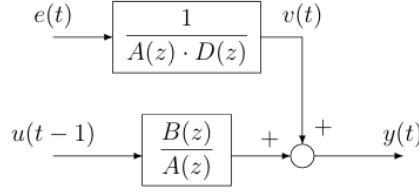
$$C(z) = 1 + c_1 z^{-1} + \dots + c_{n_c} z^{-n_c}$$



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

### 2.1.5 ARARX

Anche detti ARXAR. Oltre al contributo autoregressivo sull'uscita né è presente un'ulteriore sulla variabile d'errore.



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

## 3 Criterio di Scelta

Devo capire qual'è il miglior modello.

Se ho una sequenza di dati osservata sperimentalmente, con N campioni:

$$y_{t_0}^{t_N} = \{y(t_0), y(t_1), \dots, y(t_N)\}$$

$$u_{t_0}^{t_N} = \{u(t_0), u(t_1), \dots, u(t_N)\}$$

La *famiglia di modelli* è rappresentata dall'insieme:

$$\mathbb{M} = \{\mathcal{M}(\theta) | \theta \in \Theta\}$$

con  $\mathcal{M}$  indichiamo il modello, funzione del suo vettore di parametri  $\theta$ , individuati nello spazio vettoriale  $\Theta$ .

**Esempio** Se utilizziamo la notazione: ARX(3,3); indichiamo un modello ARX avente 3 coefficienti da identificare per il polinomio  $A(z)$  e 3 per  $B(z)$ , da cui si evince che i polinomi saranno di secondo grado e  $\Theta = \mathbb{R}^6$ .

### 3.1 Estrazione del predittore

#### 3.1.1 Predittore ad un passo

È indicato dalle seguenti forme

$$\hat{y}(t) = \hat{y}(t|t-1) = \hat{y}(t|t-1, \theta)$$

e presenta un errore di predizione:

$$\epsilon(t) = y(t) - \hat{y}(t)$$

Per una buona identificazione, vogliamo minimizziamo questo errore; calcolando un indice di qualità del modello:

$$J(\theta) = \frac{1}{N} \sum_{i=1}^N |y(t) - \hat{y}(t)|^2$$

Partendo dal modello di Box-Jenkins:

$$y(t) = G(z) \cdot u(t-1) + W(z) \cdot e(t)$$

moltiplicando ambo i membri per l'inverso di W:

$$W^{-1}(z) \cdot y(t) = W^{-1}(z) \cdot G(z) \cdot u(t-1) + W(z) \cdot e(t)$$

e sommando ad ambo i membri  $y(t)$ :

$$y(t) + W^{-1}(z) \cdot y(t) = W^{-1}(z) \cdot G(z) \cdot u(t-1) + W(z) \cdot e(t) + y(t)$$

si ottiene:

$$y(t) = W^{-1}(z) \cdot G(z) \cdot u(t-1) + W(z) \cdot e(t) + [1 - W^{-1}(z)] \cdot y(t)$$

Dove:

$$W(z) = \frac{z^m + \alpha_1 z^{m-1} + \dots}{z^m + \beta_1 z^{m-1} + \dots}$$

e

$$W^{-1}(z) = \frac{z^m + \beta_1 z^{m-1} + \dots}{z^m + \alpha_1 z^{m-1} + \dots}$$

Eseguendo la divisione (lunga) tra polinomi:

$$\begin{array}{r|l} z^m + \beta_1 z^{m-1} + \beta_2 z^{m-2} + \dots & z^m + \alpha_1 z^{m-1} + \alpha_2 z^{m-2} + \dots \\ z^m + \alpha_1 z^{m-1} + \alpha_2 z^{m-2} + \dots & 1 + (\beta_1 - \alpha_1) z^{-1} + \gamma z^{-2} + \dots \\ \hline // + (\beta_1 - \alpha_1) z^{m-1} + (\beta_2 - \alpha_2) z^{m-2} + \dots & \\ // + (\beta_1 - \alpha_1) z^{m-1} + \alpha_1 (\beta_1 - \alpha_1) z^{m-2} + \dots & \\ \hline & // + \gamma z^{m-2} + \dots \end{array}$$

possiamo scrivere il polinomio  $W^{-1}(z)$  in maniera diversa, riuscendo ad ottenere nella formula scritta in precedenza il termine di ritardo moltiplicato per  $y(t)$ , il che va ad indicare che nella parte destra dell'equazione stiamo considerando i valori passati di  $y(t)$ .

Adesso per ottenere la formula finale del predittore devo condensare la funzione  $e(t)$  in unico valore (scelgo di considerare il valore atteso, che nel caso di rumore bianco è nullo):

$$\hat{y}(t) = W^{-1}(z) \cdot G(z) \cdot u(t-1) + [1 - W^{-1}(z)] \cdot y(t)$$

### 3.1.2 ARX

Sapendo che:

$$G(z) = \frac{B(z)}{A(z)}, \quad W(z) = \frac{1}{A(z)}$$

sostituendole nell'equazione del predittore ad un passo, ottengo:

$$\hat{y}(t) = B(z) \cdot u(t-1) + [1 - A(z)] \cdot y(t)$$

### 3.1.3 ARMAX

Sapendo che:

$$G(z) = \frac{B(z)}{A(z)}, \quad W(z) = \frac{1}{A(z)}$$

sostituendole nell'equazione del predittore ad un passo, ottengo:

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

## 4 Validazione

Se  $\Theta = \mathbb{R}^n$ ,  $n$  è il numero complessivo di parametri del modello.

Ci si chiede quale legame sussiste tra  $N$  ed  $n$ . Si intuisce che  $N \gg n$ .

### 4.1 Analisi Asintotica

Se un sistema è affetto da incertezza avremo una non ripetibilità dell'acquisizione dei dati sperimentali (quindi i dati osservati in due esperimenti, condotti nelle stesse condizioni, non saranno mai esattamente gli stessi); quindi nell'identificazione non otteniamo un predittore caratterizzato sempre gli stessi parametri.

Coma abbiamo già visto, sommando tutti i quadrati degli errori di predizione si ottiene un indice di qualità:  $J_N(\theta)$ , che misura lo scostamento quadratico della predizione dall'osservazione sperimentale. Il miglior modello è quello che minimizza questo indice.

In una situazione ideale (fenomeni stocastici caratterizzati), si può definire un indice che rappresenta la descrizione completa teorica di un dato fenomeno stocastico:

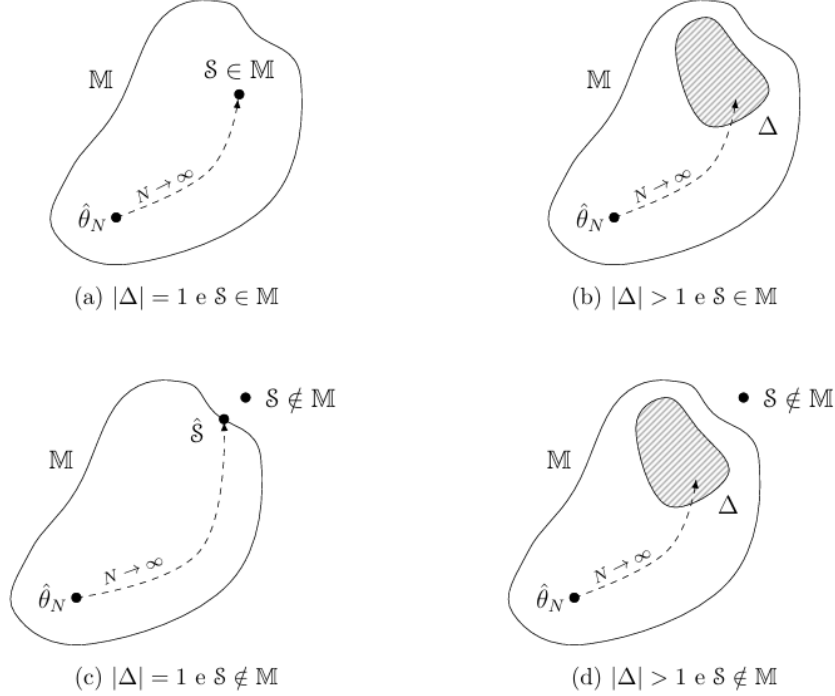
$$\bar{J}(\theta) = \{\epsilon^2(t)\}$$

L'indice  $J_N(\theta)$  però è ottenuto con  $N$  dati, quindi abbiamo una *stima campionaria*. All'aumentare di  $N$  la caratterizzazione del fenomeno stocastico migliora. Se  $N \rightarrow \infty$ , allora  $J_N(\theta) \rightarrow \bar{J}(\theta)$ .

L'insieme dei punti di minimo di  $J(\theta)$  nello spazio dei parametri è:

$$\Delta = \{\bar{\theta} \in \Theta | J(\bar{\theta}) \leq J(\theta), \forall \theta \in \Theta\}$$

Se  $\mathcal{S}$  è il **MGD** (*Meccanismo di Generazione dei Dati*), ovvero il modello da identificare; ad un aumento della quantità dei dati si tende ad un modello migliore. Si distinguono però 4 casi, a seconda della cardinalità.



- $|\Delta| = 1; \mathcal{S} \in \mathbb{M}$ . Ho un solo punto di minimo ed il modello identificato appartiene nella famiglia de modelli scelta. Allora  $\hat{\theta}_N$  converge verso  $\mathcal{S}$  nell'insieme  $\mathbb{M}$ .
- $|\Delta| = 1; \mathcal{S} \notin \mathbb{M}$ . Allora  $\hat{\theta}_N$  converge verso un modello  $\hat{\mathcal{S}}$ , che è la migliore stima del MGD all'interno della famiglia di modelli scelta e giacerebbe lungo la frontiera dell'insieme  $\mathbb{M}$ .
- $|\Delta| > 1; \mathcal{S} \in \mathbb{M}$ . Può accadere che  $\hat{\theta}_N$  converga verso un elemento di  $\Delta$  oppure che esso oscilli fra i vari punti di esso senza convergere.
- $|\Delta| > 1; \mathcal{S} \notin \mathbb{M}$ . Può accadere che  $\hat{\theta}_N$  converga verso un elemento di  $\Delta$  oppure che esso oscilli fra i vari punti di esso senza convergere.

**dim. Caso 1** Sia  $\mathcal{S} \in \mathbb{M}$  e siano le uscite del meccanismo di generazione dei dati:  $y(t|\theta^0)$ , con  $\theta^0$  vettore dei parametri veri. L'errore di predizione è:

$$\epsilon(t) = y(t) - \hat{y}(t|\theta)$$

Aggiungendo e sottraendo una quantità:

$$\epsilon(t) = y(t) - \hat{y}(t|\theta) + \hat{y}(t|\theta^0) - \hat{y}(t|\theta^0)$$

questo dovrebbe essere pari a zero; ma non lo è, perché la predizione è ottenuta sbarazzandosi dei valori casuali e sostituendoli con il valor medio.

In particolare quindi distinguiamo<sup>1</sup> :  $e(t) = y(t) - \hat{y}(t|\theta^0)$   
 Considerando un valor medio nullo:

$$Var\{\epsilon(t)\} = Var\{e(t)\} + Var\{\hat{y}(t|\theta^0) - \hat{y}(t|\theta)\} = \lambda^2 + Var\{\hat{y}(t|\theta^0) - \hat{y}(t|\theta)\}$$

Sapendo che  $e(t)$  è correlato solo a se stesso,  $\hat{y}(t|\theta^0) - \hat{y}(t|\theta)$  è correlato ai valori passati.  $\theta^0$  è un punto di minimo, quindi appartiene a  $\Delta$ . Ma non è detto però che  $\theta^0$  sia l'unico punto di minimo, inoltre ottengo  $J = 0$  solo se stimo un modello perfetto in assenza di incertezza.

## 4.2 Analisi di Bianchezza

In una situazione di identificazione ideale, tutto ciò che sta da dedurre dai dati sperimentali è stato spiegato dal modello identificato; perciò il residuo della predizione deve essere il più possibile un segnale casuale, dal momento che non deve contenere nessun *offset*<sup>2</sup>.

Il *rumore bianco* è il segnale casuale per eccellenza. Per valutare quanto siano casuali gli errori di predizione residui si effettua un'**analisi di bianchezza**, ovvero un test di somiglianza degli errori di predizione residui al rumore bianco. Per il confronto è considerata la *funzione di autocorrelazione*<sup>3</sup>. Quindi una stima campionaria della funzione di autocorrelazione dei residui della predizione deve essere quanto più possibile ad un impulso di Dirac (non sarà mai perfettamente coincidente perché non disponiamo di infiniti campioni). Un altro confronto è basato sullo *spettro*, deve risultare costante (caratteristica del rumore bianco).

## 5 Ottimizzazione ai Minimi Quadrati

Metodo di minimizzazione della funzione obiettivo molto comune, usato in tutti i casi in cui i parametri del modello hanno una dipendenza lineare dall'uscita (come gli ARX).

Supponendo di avere un'uscita sconosciuta  $y(t)$ , con  $N$  osservazioni sperimentali note di  $y$ ; si cerca una funzione che l'approssimi:

$$\hat{y}(t) = \theta_1 x_1(t) + \theta_2 x_2(t) + \dots + \theta_n x_n(t) = \theta^T \cdot \varphi(t) = \theta \cdot \varphi^T(t)$$

Definiamo il *vettore delle uscite* ( $N \times 1$ ):

$$\mathbb{Y} = \begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_N) \end{bmatrix}$$

<sup>1</sup>l'incertezza per il quale un modello non sarà mai perfetto.

<sup>2</sup>esiste un'operazione di rimozione degli offset che prende il nome di *detrend*

<sup>3</sup>il rumore bianco ha una funzione di autocorrelazione che è un impulso di Dirac (ciascun campione è correlato solamente a se stesso)

e rispettivamente abbiamo il *vettore dei regressori* (noto) e quello *dei parametri* [entrambi  $(n \times 1)$ ]:

$$\varphi(t) = \begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_n(t) \end{bmatrix}; \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_n \end{bmatrix}$$

É definita inoltre la *matrice dei regressori*  $(n \times N)$ :

$$\phi = \begin{bmatrix} \varphi^T(t_1) \\ \varphi^T(t_2) \\ \vdots \\ \varphi^T(t_N) \end{bmatrix}$$

Per risolvere il problema dei minimi quadrati:

$$\begin{aligned} J(\theta) &= \frac{1}{N} \sum_{i=1}^N \epsilon^2(t) = \frac{1}{N} (Y - \phi\theta)^T \cdot (Y - \phi\theta) = \\ &= \frac{1}{N} [(Y^T - \theta^T \phi^T)(Y - \phi\theta)] = \\ &= \frac{1}{N} [Y^T Y - \theta^T \phi^T \phi \theta - \theta^T \phi^T Y - Y^T \theta \phi] = \\ &= \frac{1}{N} [Y^T Y - \theta^T \phi^T \phi \theta - 2\theta^T \phi^T Y] \end{aligned}$$

Per minimizzare l'indice, bisogna imporre l'annullamento del suo gradiente:

$$\nabla J = \left[ \frac{\partial J}{\partial \theta_1} \cdots \frac{\partial J}{\partial \theta_n} \right] = 2\phi^T \phi \theta - 2\phi^T Y = 0$$

Da cui si ricava:

$$\phi^T \phi \theta = \phi^T Y$$

ottenendo così la soluzione:

$$\theta_{MQ} = (\phi^T \phi)^{-1} \phi^T Y = \arg \min_{\theta \in \Theta} J(\theta) = \hat{\theta}_N$$

Se la matrice  $\phi^T \phi$  non è invertibile allora non esiste la soluzione al problema. Se esiste la soluzione si ottiene in un passo (perché è di tipo algebrico).

Se dobbiamo applicare l'algoritmo dei minimi quadrati per l'identificazione di un modello ARX avente  $n = n_a + n_b$  parametri, a partire da  $N$  osservazioni sperimentali. La soluzione al problema è:

$$\theta_{MQ} = \left[ \sum_{i=1}^N \varphi(t_i) \varphi^T(t_i) \right]^{-1} \left[ \sum_{i=1}^N y(t_i) \varphi(t_i) \right]$$

Il vettore delle uscite è il seguente:

$$Y = \begin{bmatrix} y(t_1) \\ y(t_2) \\ \vdots \\ y(t_N) \end{bmatrix}$$

I vettori dei regressori e dei parametri saranno invece:

$$\varphi(t) = \begin{bmatrix} y(t_i - 1) \\ \vdots \\ y(t_i - n_a) \\ u(t_i - 1) \\ \vdots \\ u(t_i - n_b) \end{bmatrix}; \quad \theta = \begin{bmatrix} a_1 \\ \vdots \\ a_{n_a} \\ b_1 \\ \vdots \\ b_{n_b} \end{bmatrix}$$

Si avrà:

$$\hat{Y}(t_i) = \theta^T \varphi(t_i)$$

La matrice dei regressori sarà strutturata nel seguente modo:

$$\phi = \begin{bmatrix} y(t_i) & 0 & 0 & 0 & \cdots & u(t_i) & 0 & 0 & 0 & \cdots \\ y(t_2) & y(t_2) & 0 & 0 & \cdots & u(t_2) & u(t_1) & 0 & 0 & \cdots \\ \vdots & \vdots & \ddots & \cdots & \vdots & \vdots & \vdots & \ddots & \cdots & \vdots \end{bmatrix}$$

Note queste matrici, posso calcolare direttamente  $\theta_{MQ}$ .

## 5.1 Analisi Probabilistica

La linearità nei parametri dei modelli ARX è una caratteristica favorevole per svolgere l'analisi asintotica e studiare il comportamento del modello al crescere del numero N di campioni.

Considerando un caso in cui il meccanismo di generazione dei dati è un ARX:

$$S = m(\theta^0)$$

quindi, l'equazione del MGD è:

$$y(t) = \varphi^T(t) \cdot \theta^0 + e(t)$$

dove  $e(t)$ ,  $\varphi(t)$  rappresentano rispettivamente l'incertezza ed il vettore dei regressori.

L'equazione del predittore è:

$$\hat{y}(t) = \varphi^T(t) \cdot \theta$$

quindi l'errore di predizione vale:

$$\epsilon(t) = \varphi^T(t) \cdot (\theta^0 - \theta) + e(t)$$



Il termine  $(\theta^0 - \theta)$  prende il nome di *errore di stima parametrica*.

L'indice da minimizzare è:

$$\begin{aligned}\bar{J}(\theta) &= E\{e^2(t)\} = E\{(\theta^0 - \theta)^T \varphi(t) \varphi^T(t) (\theta^0 - \theta) + e^2(t) + 2\varphi^T(t) (\theta^0 - \theta) e(t)\} = \\ &= (\theta^0 - \theta)^T E\{\varphi(t) \varphi^T(t)\} (\theta^0 - \theta) + E\{e^2(t)\} + E\{2\varphi^T(t) (\theta^0 - \theta) e(t)\}\end{aligned}$$

Se  $e(t)$  è il rumore bianco, è correlato solamente a se stesso, allora:

$$\bar{J}(\theta) = (\theta^0 - \theta)^T E\{\varphi(t) \varphi^T(t)\} (\theta^0 - \theta) + E\{e^2(t)\} + E\{2\varphi^T(t)\} (\theta^0 - \theta) E\{e(t)\}$$

inoltre  $E\{e^2(t)\} = \lambda^2$  e  $E\{e(t)\} = 0$ , quindi:

$$\bar{J}(\theta) = (\theta^0 - \theta)^T E\{\varphi(t) \varphi^T(t)\} (\theta^0 - \theta) + \lambda^2$$

Otengo quindi una forma quadratica, legata fortemente alla matrice dei minimi quadrati:

$$S(N) = \sum_{t=1}^N \varphi(t) \varphi^T(t)$$

la quale per definizione è semidefinita positiva. Affinché si abbia una soluzione però il suo determinante deve essere diverso da zero (la matrice deve essere invertibile). Una buona identificazione si ha quando si può manipolare  $S(N)$  per renderla "il più possibile" non singolare.

### 5.1.1 Persistente eccitazione

Per valutare la singolarità della matrice  $S(N)$  consideriamo un modello ARX(1,1).

$$y(t) = a_1 y(t-1) + b_1 u(t-1)$$

$$\varphi(t) = [y(t-1) \quad u(t-1)]^T$$

$$\varphi(t) \varphi^T(t) = \begin{bmatrix} y(t-1) \\ u(t-1) \end{bmatrix} \cdot [y(t-1) \quad u(t-1)] = \begin{bmatrix} y^2(t-1) & y(t-1)u(t-1) \\ u(t-1)y(t-1) & u^2(t-1) \end{bmatrix}$$

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

La matrice dei minimi quadrati tende a divergere al crescere di  $N$ , per questo la si *normalizza*:

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

Poiché i segnali  $u$  ed  $y$  sono stazionari, i termini della matrice sono le stime campionarie delle rispettive funzioni di correlazione per  $\tau = 0$ , che indicheremo con  $\gamma$ :

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

Considerando invece un modello  $ARX(n_a, n_b)$ :

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

dove:

- $\Gamma_{yy}$  è il blocco relativo all'autocorrelazione del segnale in uscita. Non è manipolabile e la sua dimensione è  $(n_a \times n_a)$ .
- $\Gamma_{yu} = \Gamma_{yu}$  sono i blocchi relativi alla cross-correlazione fra il segnale in uscita e quello in ingresso. Non sono manipolabili e le loro dimensioni sono  $(n_a \times n_b)$  e  $(n_b \times n_a)$ .
- $\Gamma_{uu}$  è il blocco relativo all'autocorrelazione del segnale di ingresso. È manipolabile e la sua dimensione è  $(n_b \times n_b)$ .

Avendo una matrice partizionata nel modo seguente:

$$U = \begin{bmatrix} F & K \\ K^T & G \end{bmatrix}$$

per il **lemma di Schur**, condizione necessaria e sufficiente affinché  $U$  sia definita positiva:

$$\begin{cases} G > 0 \\ F - K^T G^{-1} K > 0 \end{cases}$$

La struttura della matrice  $R(N)$  è simile a quella della matrice  $U$ . Il blocco  $G$  corrisponde a  $\Gamma_{uu}$ , e quindi lo posso manipolare, mentre i blocchi  $K = \Gamma_{yu}$ ,  $F = \Gamma_{yy}$  no, in quanto dipendenti dall'uscita.

Nella pratica si cerca di scegliere una matrice  $(n_b \times n_b)$  definita positiva come di seguito:

$$\Gamma_{uu} = \begin{bmatrix} \gamma_{uu}(0) & \gamma_{uu}(1) & \gamma_{uu}(2) & \cdots & \gamma_{uu}(n_b - 1) \\ \gamma_{uu}(1) & \gamma_{uu}(0) & \gamma_{uu}(1) & \cdots & \gamma_{uu}(n_b - 2) \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ \gamma_{uu}(n_b - 1) & \gamma_{uu}(n_b - 2) & \gamma_{uu}(n_b - 3) & \cdots & \gamma_{uu}(0) \end{bmatrix}$$

La struttura di questa matrice (elementi coincidenti lungo le varie diagonali) la rende una matrice di **Toeplitz**.

Sulla diagonale principale si trova la varianza del segnale di ingresso, sulle diagonali al di sopra ed al di sotto di quella principale si trovano le autocorrelazioni in un passo (e così via fino ad arrivare alle funzioni di autocorrelazione in  $n_b - 1$  passi).

La scelta di una matrice  $\Gamma_{uu}$  definita positiva è una condizione necessaria ma non sufficiente (almeno sono convinto di non essere io a sbagliare l'ingresso, ma non posso garantire di avere una situazione identificabile).

Un segnale è detto **persistentemente eccitante di ordine  $N$**  se la sua matrice  $\Gamma_{uu}$  è non singolare (quindi è invertibile).

Se un segnale è persistentemente eccitante di ordine  $N$ , lo è anche per qualsiasi altro ordine minori di  $N$ .

Alcuni esempi di segnali di questo tipo:

- **rumore bianco**: può essere di ordine qualsiasi;
- **rumore colorato**: può essere di ordine qualsiasi;

- **PRBS** (*Pseudo Random Binary Signals*): è di ordine  $\ell$ , con  $\ell$  numero di registri a scorrimento impiegati per generarlo.
- **somma di sinusoidi equispaziate in frequenza**: è di ordine  $m$ , con  $m$  numero di sinusoidi (il rumore bianco rientra in questa categoria, con  $m \rightarrow \infty$ ).

## 5.2 Identificabilità

Studia la possibilità di trovare effettivamente un modello. È suddivisibile in due tipologie:

- *strutturale* (errore nella scelta del modello che causa una scarsa identificabilità);
- *sperimentale* (errore nella scelta degli ingressi).

## 5.3 Ordine Ottimo di un Modello

Uno dei problemi centrali dell'identificazione è la scelta della complessità dei modelli. Un modello risulta efficace quando può spiegare oltre ai dati disponibili nell'identificazione anche altri dati non considerati nel training.

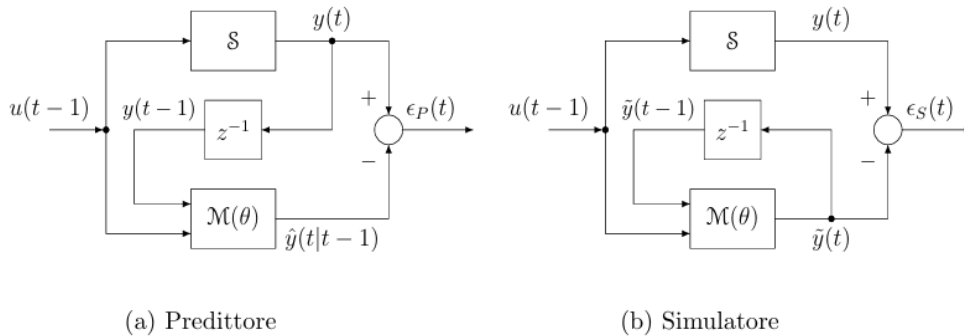
Utilizzando il criterio:

$$\mathcal{J}(\theta) = \frac{1}{N} \sum_{t=1}^N \epsilon^2(t)$$

con:

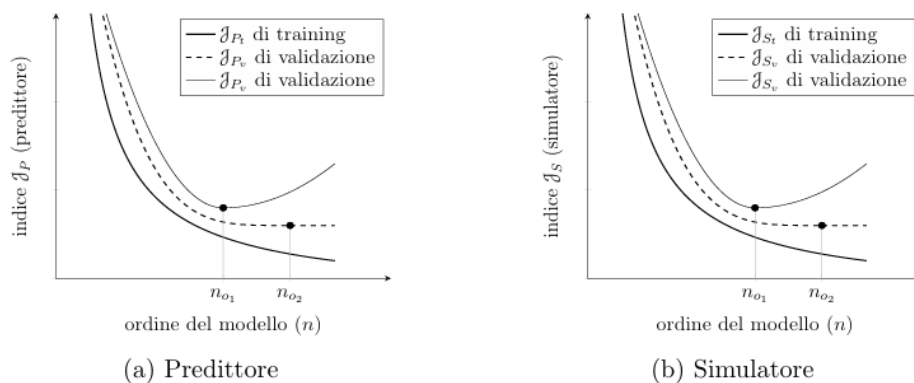
- $n$ : numero di parametri
- $\theta$ : parametri
- $\epsilon = \begin{cases} \epsilon_p(t) = y(t) - \hat{y}(t|t-1) : \text{errore di predizione} \\ \epsilon_s(t) = y(t) - \hat{y}(t) : \text{errore di simulazione} \end{cases}$

L'utilizzo di un simulatore in genere permette di ottenere risultati migliori (ma costruirne uno è più complicato).



Se il numero di dati da identificare è elevato, possiamo usare una parte di essi per effettuare il training di un modello e la restante parte per la validazione dello stesso. Al crescere dell'ordine del modello l'indice diminuisce. L'ordine ottimo si ottiene però in corrispondenza di un punto di minimo dell'indice. Perché per ordini troppo elevati l'andamento di  $\mathcal{J}$  per i dati di validazione tende a peggiorare, perché tende a seguire troppo i dati di training (anche il rumore presente), andando in una situazione di *overfitting*.

Può capitare che l'andamento dell'indice diventi costante dopo un certo ordine, in tal caso si hanno infiniti minimi. È semplice dedurre che verrà scelto il numero minimo di parametri da introdurre che portino comunque al modello ottimo.



Se non si dispone di un numero sufficiente di dati per effettuare training, l'ordine ottimo deve essere stimato con approcci differenti. Si ricorre a, l'utilizzo di indici di ottimalità (i primi due tendono a sovrastimare l'ordine corretto del modello, il terzo da una stima leggermente più precisa).

### 1. Final Prediction Error:

$$FPE = \frac{N+n}{N-n} \mathcal{J}(\hat{\theta}_N)$$

### 2. Akaike Information Criterion:

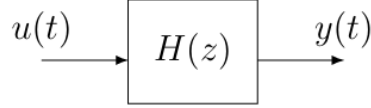
$$AIC = \frac{2n}{N} + \ln(\mathcal{J}(\hat{\theta}_N))$$

### 3. Minimum Description Length:

$$MDL = \ln(N) + \frac{n}{N} + \ln(\mathcal{J}(\hat{\theta}_N))$$

## 5.4 Analisi Spettrale

Effettuando un passaggio dal dominio del tempo a quello della frequenza (tramite le trasformate di Fourier), si possono evidenziare alcune peculiarità.



Il *teorema del filtraggio* dei segnali afferma che, dato un filtro lineare asintoticamente stabile, se  $u(t)$  è stazionario (come il rumore bianco) e  $\Phi_{uu}(\omega)$  è il suo spettro, si dimostra che lo spettro dell'uscita sarà:

$$\Phi_{yy}(\omega) = |H(e^{j\omega})|^2 \Phi_{uu}(\omega)$$

il filtro, quindi, ridistribuisce la potenza del segnale nel dominio della frequenza, alterando la forma dello spettro.

Il *teorema di Wiener* afferma che, la trasformata di Fourier della funzione di autocorrelazione di un segnale stazionario coincide con la sua densità spettrale:

$$\mathcal{F}\{\gamma_{uu}(\tau)\} = \Phi_{uu}(\omega) \iff \mathcal{F}^{-1}\{\Phi_{uu}(\omega)\} = \gamma_{uu}(\tau)$$

Supponendo di avere un MGD (Box-Jenkins):

$$S : y(t) = G_0(z)u(t-1) + W_0(z)e(t)$$

ed un modello del predittore:

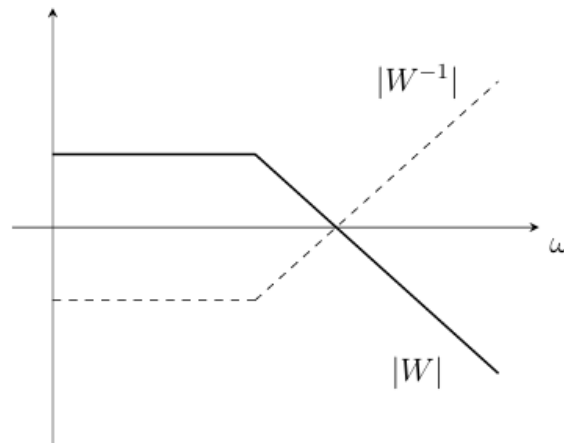
$$m : \hat{y}(t) = (1 - W^{-1}(z))y(t) + W^{-1}(z)G(z)u(t-1)$$

L'errore di predizione è:

$$\begin{aligned} \epsilon(t) &= y(t) - \hat{y}(t) = W^{-1}y(t) - W^{-1}Gu(t-1) = \\ &= W^{-1}[y(t) - Gu(t-1)] = \\ &= W^{-1}[G_0u(t-1) + W_0e(t) - Gu(t-1)] \end{aligned}$$

Definendo  $\Delta G = (G_0 - G)$ , otteniamo:

$$\epsilon(t) = W^{-1}[\Delta Gu(t-1) + W_0e(t)]$$



#### 5.4.1 Pre-filtraggio

### 5.5 Operazioni di Pre-processing sui Dati