

Lezione 15 — 21 Aprile

Docente: Luca Schenato

Stesore:

R. Ghirardello, F. Paggiaro, A. Zugno
K. Schmiedhofer, R. Alberton, M. Pattarello

15.1 Identificazione

In questa lezione si tratta l'identificazione di processi. Alcuni di questi sono instabili per natura e pertanto bisogna fare identificazione usando la catena chiusa ed un controllo minimo (ad esempio proporzionale), affinché il sistema deve essere almeno stabile. Si illustreranno alcune delle tecniche di identificazione utilizzate, in funzione del tipo di sistema in esame.

Classificazione delle tecniche di identificazione

1. Sistemi non lineari

In questa sezione si intende la dinamica del sistema come non lineare, diversa dalla non linearità dei parametri (ove invece la dinamica è di per sé è lineare).

Esempio 1

$$\ddot{x} = -kx^3 - b\dot{x}|\dot{x}|$$

Sistema con dinamica non lineare

$$\ddot{x} = -kx - b^2k\dot{x}$$

Sistema non lineare nei parametri,
ma con dinamica lineare.

□

Questo problema è ancora aperto e molto difficile da risolvere. In generale si cercano delle classi particolari di sistemi non-lineari per i quali sviluppare tecniche specifiche.

2. Sistemi lineari

Per questi sistemi ci sono due tipi di approccio

a. Identificazione parametrica

Nella *identificazione parametrica* si hanno un numero finito di parametri (stabilito a priori) da identificare, indipendenti dal numero di dati $\{u, y\}$. Al variare del numero di dati che si utilizzano per l'identificazione, la dimensione n del parametro θ ($\in \mathbb{R}^n$) resta costante.

Computazionalmente è difficile da risolvere perchè spesso la funzione di costo non è convessa e quindi presenta più minimi locali.

Si definisce un funzionale che dipende dai dati:

$$J(\theta; u, y) \in \mathbb{R}^+ ,$$

che si cercherà di minimizzare.

Si sceglie come parametro del modello:

$$\theta^*(u, y) = \operatorname{argmin}_{\theta} J(\theta; u, y) .$$

In generale si cerca di definire un modello statistico di y, u, θ e un funzionale di costo naturale in questo contesto e' la probabilità dei dati osservati condizionata sui parametri incogniti $P(y, u|\theta)$:

$$J(\theta; u, y) = -P(y, u|\theta) .$$

E quello che si cerca diviene quindi lo stimatore di massima verosimiglianza (MV).

$$\theta_{MV}^* = \operatorname{argmax}_{\theta} P(y, u|\theta)$$

Un'ultima subclassificazione dei metodi parametrici basata sulle informazioni a-priori disponibili è la seguente:

I. Metodo Black Box

Nel metodo di risoluzione *Black Box* non si ha nessuna informazione sulla dinamica. Sono noti solo gli ingressi, le uscite e l'ordine ipotetico del sistema in esame.

Il funzionamento di tale metodo è rappresentato in Figura 15.1.

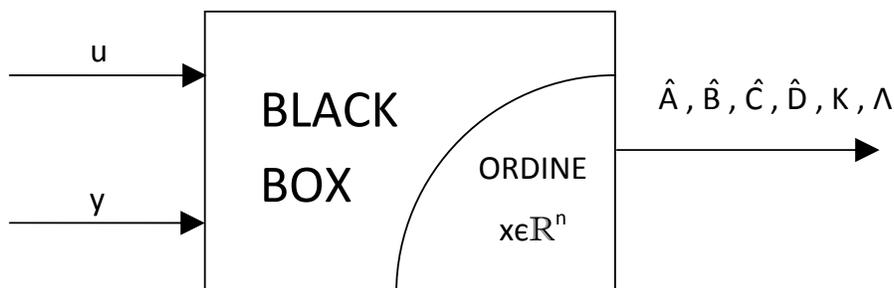


Figura 15.1. Schema di funzionamento del metodo black-box.

dove K è il guadagno di Kalman e Λ è la varianza d'errore di predizione d'uscita $e_y = (y - \hat{y})$.

Si introducono i segnali nel sistema e si osserva la risposta in uscita. Si può variare solo l'ordine del parametro da stimare. All'aumentare dell'ordine il

modello descriverà sempre meglio i dati, ma potrebbe discostarsi dalla reale natura del processo.

Per tale metodo ci sono due tecniche di risoluzione:

- **Spazio di stato** (Identificazione Toolbox di Matlab) (per MIMO)

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k & w_k \sim \mathcal{N}(0, Q) \\ y_k = Cx_k + Du_k + v_k & v_k \sim \mathcal{N}(0, R) \end{cases} \quad (15.1)$$

Qui si ha la problematica della base, poiché per un sistema dinamico lineare corrispondono infinite rappresentazioni in spazio di stato, a differenza di una sua rappresentazione tramite funzione di trasferimento, per esempio. La scelta della base può risultare fondamentale per poter ottenere una identificazione numericamente stabile.

- **Sottospazi** (per MIMO)

Questa tecnica è molto più robusta dal punto di vista numerico. Esso va a scegliere la rappresentazione in spazio di stato più opportuna. Si vedano per esempio gli appunti del Prof. Alessandro Chiuso disponibili sulla pagina del sito del corso.

II. Metodo Gray Box

Nel metodo di risoluzione *Gray Box* si ha come informazione la struttura della dinamica. Tali informazioni rappresentano delle specie di vincoli che si accostano alla minimizzazione dell'indice di costo. In linea di principio questa informazione riduce lo spazio dei parametri da identificare, ma al tempo stesso rende il funzionale di costo da ottimizzare fortemente non convesso con molti minimi e massimi locali.

Esempio 2

Abbiamo il modello in spazio di stato:

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k \\ y_k = Cx_k + Du_k + v_k \end{cases} \quad (15.2)$$

vogliamo imporre che la matrice del sistema abbia la seguente struttura:

$$A = \begin{bmatrix} 0 & 1 \\ * & * \end{bmatrix},$$

oppure, come sarà per il motore in laboratorio:

$$A = \begin{bmatrix} 0 & 1 \\ m & m^2 \end{bmatrix},$$

in cui si aggiunge vincolo in più, che è rappresentato dalla relazione fra i due elementi incogniti della seconda riga. Inoltre si può notare che il sistema considerato è non lineare nei parametri ma lineare nella dinamica.

□

Cioè ogni volta che si ha un problema di ottimizzazione, si può pensare:

$$\begin{aligned} \min_{\theta} \quad & J(\theta; u, y) \\ \text{soggetto a} \quad & f_i(\theta) = 0 . \end{aligned}$$

Ogni volta che si aggiungono vincoli il problema risulta più difficile, dato che tali vincoli possono essere non lineari o non convessi. Questo è molto sensibile e si valuta caso per caso: può convergere sotto buone condizioni iniziali, come no, perché i vincoli e la minimizzazione sono molto complessi. Inoltre con il Gray-box il problema può essere mal posto (si veda *Esempio 3*).

Esempio 3

Se la matrice del sistema (15.2) è del tipo:

$$A = \begin{bmatrix} 0 & 1 \\ mk & m^2b \end{bmatrix} ,$$

si riescono a identificare i due elementi della matrice, ma avendo tre parametri da determinare è impossibile sperare di discriminare i 3 parametri separatamente. In questo caso, quindi, si dice che il problema è mal condizionato.

Per tale motivo si tende ad usare il metodo Black-box nel caso in cui ci siano molti parametri fisici incogniti, mentre si tende ad usare il metodo Gray-box quando i parametri sono pochi o quando si hanno delle buone stime iniziali di questi parametri.

b. Identificazione non parametrica

Nella *identificazione non parametrica* si hanno un numero non predefinito di parametri da identificare che dipendono dal numero di dati $\{u, y\}$. Cioè, al variare del numero di dati che si utilizzano per l'identificazione, la dimensione n del parametro θ ($\in \mathbb{R}^n$) aumenta.

Questo comporta un aumento anche della complessità di tale metodo, anche se come risoluzione è più facile dell'identificazione parametrica perchè spesso la funzione di costo è convessa e si ha un solo minimo.

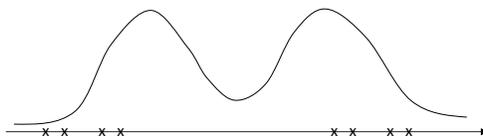
Esempio 4

Supponiamo di voler identificare una densità di probabilità e di avere solo certi campioni di questa a disposizione.

- *Identificazione parametrica*: si ipotizza ad esempio che la densità di probabilità sia la combinazione convessa di due densità di probabilità gaussiane.

Con abuso di notazione possiamo calcolare:

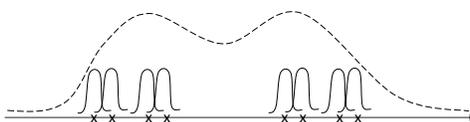
$$\hat{f} = p\mathcal{N}(\mu_1, \sigma_1) + (1 - p)\mathcal{N}(\mu_2, \sigma_2)$$



dove $p \in [0, 1]$ con un numero fisso di parametri:

$$(p, \mu_1, \mu_2, \sigma_1, \sigma_2) = \theta$$

- *Identificazione non parametrica (statistical learning)*: si associa ad ogni campione una densità di probabilità gaussiana centrata nella posizione del campione x_i e con varianza fissata σ .



Pesando tutte queste densità di probabilità in maniera uniforme si calcola, con abuso di notazione, la funzione di densità stimata come segue:

$$\hat{f} = \frac{1}{N} \sum_{k=1}^N \mathcal{N}(x_i, \sigma),$$

in cui la varianza σ è stabilita a priori, e si vede che il numero di funzioni utilizzate per la stima è pari al numero di dati.

15.2 Rappresentazione dei sistemi a tempo discreto

Vi sono diverse rappresentazioni dei sistemi a tempo discreto:

1. Equazioni alle differenze

Si parte dall'equazione alle differenze del tipo:

$$y(t) = a_1 y(t-1) + a_2 y(t-2) + \dots + b_0 u(t) + b_1 u(t-1) + \dots + c_0 e(t) + c_1 e(t-1) + \dots \quad (15.3)$$

dove con $u(\cdot)$ si identifica l'ingresso mentre con $e(\cdot)$ il rumore.

Applicando la trasformata Zeta all'equazione alle differenze (15.3), si ottiene:

$$Y(z) = a_1 z^{-1} Y(z) + a_2 z^{-2} Y(z) + \dots + b_0 U(z) + b_1 z^{-1} U(z) + \dots + c_0 E(z) + c_1 z^{-1} E(z) + \dots,$$

da cui raccogliendo si ottiene:

$$(1 - a_1 z^{-1} - a_2 z^{-2} + \dots) Y(z) = (b_0 + b_1 z^{-1} + \dots) U(z) + (c_0 + c_1 z^{-1} + \dots) E(z), \quad (15.4)$$

e quindi si trova il modello ARMAX:

$$A(z^{-1})Y(z) = B(z^{-1})U(z) + C(z^{-1})E(z) \quad (15.5)$$

Dalle semplificazioni del modello ARMAX ($A(z^{-1})$, $B(z^{-1})$, $C(z^{-1})$) nascono altre descrizioni dei sistemi a tempo discreto:

a. Il **modello (ARX?)**

In cui $C(z^{-1})=1$, che risulta interessante perchè il problema di massimizzare la verosimiglianza diventa un problema ai minimi quadrati;

b. Il **modello OE** (Output Error)

In cui $C(z^{-1})=A(z^{-1})$ e quindi: $Y(z) = \frac{B(z^{-1})}{A(z^{-1})} U(z) + E(z)$;

c. Il **modello Box-Jenkins**

$$Y(z) = F(z^{-1})U(z) + H(z^{-1})E(z)$$

con $F(z^{-1})$ e $H(z^{-1})$ funzioni razionali.

2. Spazio di stato

Un altro modo per rappresentare i sistemi discreti è lo spazio di stato:

$$\begin{cases} x_{k+1} = Ax_k + Bu_k + w_k \\ y_k = Cx_k + Du_k + v_k \end{cases} \quad (15.6)$$

La rappresentazione scelta per il sistema discreto sarà utilizzata successivamente nel Toolbox di Identificazione di Matlab che cercherà di minimizzare il funzionale di costo del metodo PEM di identificazione.

Si ha a disposizione la stima, in funzione dei parametri:

$$\hat{y}(t) = g_{\theta}(y(0), \dots, y(t-1), \dots, u(0), \dots, u(t-1)) . \quad (15.7)$$

dove g_{θ} dipende dalla classe di modelli scelta (ARX, ARMAX, OE, etc..). Indicando con $\epsilon(t) = y(t) - \hat{y}(t) \in \mathbb{R}^m$ l'errore di predizione, si considera il seguente funzionale di costo:

$$R_T(\theta) = \frac{1}{T} \sum_{t=1}^T \epsilon(t)\epsilon(t)^T \geq 0 . \quad (15.8)$$

L'obiettivo è minimizzare tale funzionale di costo, cioè calcolare:

$$\min_{\theta} \begin{cases} \text{tr}(R_T(\theta)) \\ \text{oppure} \\ \det(R_T(\theta)) \end{cases} \quad (15.9)$$

che è esattamente quello che calcola il Toolbox di Identificazione di Matlab. In particolare la minimizzazione del determinante corrisponde a massimizzare la verosimiglianza nel caso asintotico, cioè per $T \rightarrow \infty$.