

16.1 Introduzione

Sia dato il sistema di Figura 16.1 nel quale sono note, perchè misurate, le misure $y(t)$ e gli ingressi $u(t)$.

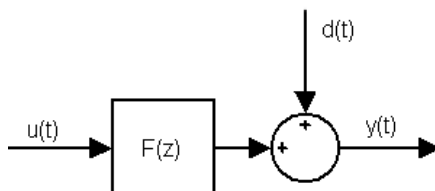


Figura 16.1. Schema generale del processo da identificare

Si desidera ricavare lo spettro del sistema. Si noti come lo schema assegnato sia del tutto generale: non è specificata la presenza del feedback, nulla si dice a riguardo della linearità. Il problema che viene risolto in questi appunti è in realtà un sottoproblema più semplificato (Figura 16.2): si assume come ingresso il rumore bianco $n(t)$, che non viene misurato, e si vuole trovare la funzione di trasferimento del sistema, ovvero trovare lo spettro $S_y(z) = W(z)W^T(z^{-1})$.

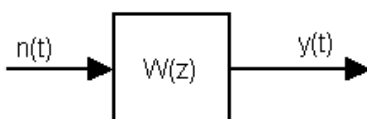


Figura 16.2. Schema semplificato del processo da identificare

L'approccio classico della teoria dell'identificazione prevede di agire secondo i seguenti passi:

1. si parametrizza la funzione di trasferimento $W(z)$ ovvero il predittore di un passo del processo y , $\hat{y}_\theta(t | t - 1)$;
2. si minimizza una funzione costo dove i parametri θ possono anche essere non lineari. Ad esempio il metodo PEM minimizza il valore atteso del costo di predizione:

$$\hat{\theta} = \arg \min_{\theta} \mathbb{E}[(y(t) - \hat{y}_\theta(t | t - 1))^2] \quad (16.1)$$

Nel caso multivariabile tale metodo è molto complicato e la soluzione non è globale (gli indici di Kronecker non sono ben definiti, si parla di *forme di overlapping*).

In generale la funzione da minimizzare risulta non convessa e dunque la procedura fin qui descritta è di difficile applicazione. Ciò giustifica l'introduzione di una tecnica diversa, nota come "algoritmi a sottospazi". Tale strumento salta i due punti precedentemente descritti evitando così la parametrizzazione e tutti i problemi che ne conseguono.

16.2 Algoritmi a sottospazi (solo per serie temporali)

Viene introdotta una rappresentazione in spazio di stato per la funzione di trasferimento $W(z)$:

$$\begin{cases} x(t+1) = Ax(t) + w(t) \\ y(t) = Cx(t) + v(t) \end{cases} \quad (16.2)$$

Vengono raccolte le misure $y(t)$, $t \in [1, \dots, T]$ e si desidera costruire le matrici A e C e ricavare qualcosa a riguardo delle statistiche sui rumori bianchi, ad esempio le covarianze. Costruiamo dunque le seguenti matrici:

$$Y^- = \begin{bmatrix} y(0) & y(1) & \dots & y(N-1) \\ y(1) & y(2) & \dots & y(N) \\ \vdots & & & \vdots \\ y(k-1) & y(k) & \dots & y(k+N-2) \end{bmatrix} \quad Y^+ = \begin{bmatrix} y(k) & \dots & y(k+N-1) \\ \vdots & & \vdots \\ y(2k-1) & \dots & y(2k+N-2) \end{bmatrix} \quad (16.3)$$

Nelle matrici l'ordine di grandezza di k è piccolo rispetto a N e le righe generano uno spazio k -dimensionale. Le matrici (16.3) servono per comprendere meglio i passaggi seguenti: utilizzando infatti le singole righe, che sono del tipo

$$Y_t = [y(t) \quad y(t+1) \quad \dots \quad y(t+N-1)]$$

si può riscrivere il modello in spazio di stato (16.2) come

$$\begin{cases} X_{t+1} = AX_t + W_t \\ Y_t = CX_t + V_t \end{cases} \quad (16.4)$$

Ricordando che i rumori w e v sono scorrelati dallo stato x , supponendo di disporre del vettore X_t , è facile ricavare le matrici A e C , infatti manipolando il sistema (16.4) si ottiene:

$$\frac{X_{t+1}X_t^T}{N} = \frac{AX_tX_t^T}{N} + \frac{W_tX_t^T}{N}$$

$$\frac{Y_tX_t^T}{N} = \frac{CX_tX_t^T}{N} + \frac{V_tX_t^T}{N}$$

Gli ultimi due termini di entrambe le equazioni tendono a zero in quanto la loro aspettazione è nulla, mentre la matrice $\frac{X_t X_t^T}{N}$ dev'essere definita positiva per N grande, altrimenti non ci sarebbe raggiungibilità; essa è dunque invertibile e si sono così ricavate le espressioni per le matrici incognite. La stima è stata fatta in una base ignota e di cui non interessa la natura. Si ricorda che tutto ciò vale, ovviamente, se si conosce lo stato.

Ritorniamo ora al modello il spazio di stato (16.2). Applicando l'operatore varianza a tale modello, definendo $\text{Var } x(t) = P$ e $\text{Var } y(t) = \Lambda(0)$ si ottengono le seguenti equazioni:

$$\begin{aligned} P &= APA^T + Q \\ \Lambda(0) &= CPC^T + R \end{aligned}$$

Si definisce $\Lambda(\tau) = \mathbb{E}[y(t + \tau)y^T(t)]$. Viene ora calcolato come esempio istruttivo $\Lambda(1)$, lasciando al lettore il compito di ricavare l'espressione generale per $\Lambda(t)$, comunque fornita.

$$\begin{aligned} \Lambda(1) &= \mathbb{E}[y(t+1)y^T(t)] = \mathbb{E}[(Cx(t+1) + v(t+1))(Cx(t) + v(t))^T] \\ &= C\mathbb{E}[x(t+1)x^T(t)]C^T + C\mathbb{E}[x(t+1)v^T(t)] \\ &= CA\mathbb{E}[x(t)x^T(t)]C^T + CA\mathbb{E}[w(t)x^T(t)] + CA\mathbb{E}[x(t)v^T(t)] + C\mathbb{E}[w(t)v^T(t)] \\ &= CAPC^T + CS = C(APC^T + S) = C\bar{C}^T \end{aligned}$$

dove è stata sfruttata la scorrelazione tra alcuni termini. In generale per le covarianze varrà dunque la relazione $\Lambda(k) = CA^{k-1}\bar{C}^T$. Dunque, si definiscono i vettori del passato e del futuro dei dati rispetto all'istante presente k come segue:

$$\mathcal{Y}_+ = \begin{bmatrix} y(k) \\ y(k+1) \\ \vdots \\ y(2k-1) \end{bmatrix} \quad \mathcal{Y}_- = \begin{bmatrix} y(k-1) \\ y(k-2) \\ \vdots \\ y(0) \end{bmatrix} \quad (16.5)$$

Si noti che gli elementi di tali vettori sono variabili aleatorie. Si possono dunque scrivere le covarianze a partire da passato e futuro, in un'unica matrice:

$$\begin{aligned} \mathbb{E}[\mathcal{Y}_+ \mathcal{Y}_-^T] &= \begin{bmatrix} \Lambda(1) & \Lambda(2) & \dots & \Lambda(k) \\ \Lambda(2) & \Lambda(3) & \dots & \Lambda(k+1) \\ \vdots & & & \vdots \\ \Lambda(k) & \Lambda(k+1) & \dots & \Lambda(2k-1) \end{bmatrix} = \begin{bmatrix} C\bar{C}^T & \dots & CA^{k-1}\bar{C}^T \\ \vdots & & \vdots \\ CA^{k-1}\bar{C}^T & \dots & CA^{2k-2}\bar{C}^T \end{bmatrix} = \\ &= \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{k-1} \end{bmatrix} [\bar{C}^T \quad A\bar{C}^T \quad \dots \quad A^{k-1}\bar{C}^T] \end{aligned} \quad (16.6)$$

Le matrici scritte nell'ultimo passaggio hanno rango n . Fattorizzando la matrice, si riescono a ricavare le matrici incognite A e C . In realtà il procedimento qui svolto non è propriamente

quello eseguito dall'algoritmo, ma sono stati svolti questi passaggi per aiutare a comprenderne il funzionamento. Esso si basa sulla decomposizione ai valori singolari (SVD) ricavando da essa la matrice di osservabilità del sistema da identificare, che altro non è che la matrice ottenuta nell'ultimo passaggio di (16.6). I passaggi sono i seguenti:

$$[\Lambda(k)] = \Sigma_{+-} = USV^T = \mathcal{O}_k \bar{\mathcal{O}}_k^T \quad (16.7)$$

Ragionando sulle dimensioni delle matrici, ricordando che U è “alta” e V “lunga”, si deduce che si può tranquillamente porre

$$\mathcal{O}_k = US^{\frac{1}{2}} \quad \bar{\mathcal{O}}_k^T = S^{\frac{1}{2}}V^T \quad (16.8)$$

Sembrerebbe di aver trovato un buon risultato, tuttavia la matrice Σ_{+-} non è normalizzata e dunque risulta difficile capire quali valori sono grandi e quali piccoli. Si rende dunque necessario normalizzare la matrice. Una volta che ciò sarà stato fatto, supponendo che i vettori aleatori \mathcal{Y}_+ e \mathcal{Y}_- siano gaussiani, si potrà calcolare il valore atteso dei valori futuri condizionato con le misure passate:

$$\begin{aligned} \mathbb{E}[\mathcal{Y}_+ | \mathcal{Y}_-^T] &= \Sigma_{+-} \Sigma_{--}^{-1} \mathcal{Y}_- = \\ &= \mathcal{O}_k \bar{\mathcal{O}}_k^T (\Sigma_{--}^{-\frac{1}{2}})^T \Sigma_{--}^{-\frac{1}{2}} \mathcal{Y}_- = \mathcal{O}_k \hat{x}(k) \end{aligned} \quad (16.9)$$

L'ultimo passaggio è dovuto al fatto che i termini che moltiplicano \mathcal{O}_k dipendono linearmente da \mathcal{Y}_- e dunque sono lo stato del sistema predetto secondo Kalman.

Esercizio Ricavare le stesse equazioni partendo da $\mathcal{Y}_+ = \mathcal{O}_k x(k) + \text{rumori}$. (Suggerimento: utilizzare l'aspettazione condizionata.)

Nella prossima sezione verrà affrontato il problema della normalizzazione, tramite un preciso strumento matematico.

16.2.1 Analisi di correlazione canonica (CCA)

L'analisi di correlazione canonica tra due spazi vettoriali¹ \mathcal{Y}_+ e \mathcal{Y}_- consiste nel trovare due vettori, uno per spazio, il più correlati possibile (ovvero tali che l'angolo tra essi compreso sia il più piccolo possibile). Successivamente ci si muove nel complemento ortogonale dei due vettori scelti e si procede allo stesso modo, trovando via via i vettori che predicano meglio lo stato. Complessivamente si trova dunque la base migliore possibile per predire quanto c'è nell'altro spazio vettoriale. Lo strumento per realizzare tale “algoritmo” è la SVD. Date due generiche matrici A e B contenenti i vettori scelti, si ha:

$$(AA^T)^{-\frac{1}{2}} AB^T (BB^T)^{-\frac{T}{2}} = USV^T \quad (16.10)$$

¹Si adotta qui un leggero abuso di notazione, indicando col medesimo simbolo i vettori di variabili aleatorie e gli spazi vettoriali da essi generati.

Si pone dunque

$$\begin{aligned}\bar{A} &= U^T(AA^T)^{-\frac{1}{2}}A \\ \bar{B} &= V^T(BB^T)^{-\frac{1}{2}}B\end{aligned}\tag{16.11}$$

Valgono le seguenti proprietà:

1. $\bar{A}\bar{B}^T = S$ i valori singolari sono i coseni degli angoli tra i vettori (prodotto scalare tra vettori);
2. $\bar{A}\bar{A}^T = I$ è base ortonormale per lo spazio;
3. $\bar{B}\bar{B}^T = I$ è base ortonormale per lo spazio.

Per il problema in questione, per normalizzare la covarianza Σ_{+-} si imposta la seguente relazione:

$$\Sigma_{++}^{-\frac{1}{2}}\Sigma_{+-}(\Sigma_{--}^{-\frac{1}{2}})^T = USV^T\tag{16.12}$$