Sistemi di Controllo Digitale

Identificazione di Modelli Dinamici

Contenuto delle Lezioni

- Il problema della stima e della predizione
- Teoria della stima e caratteristiche degli stimatori
- Stima ai minimi quadrati (LMS)

Il Problema della Stima e della Predizione

Il problema della stima nasce quando si vogliano determinare uno o piu` parametri incogniti a partire da osservazioni sperimentali.

Osservazioni sperimentali

$$d(t),\ t = t_1, t_2, \ldots, t_N$$

Parametro(i) incognito(i)

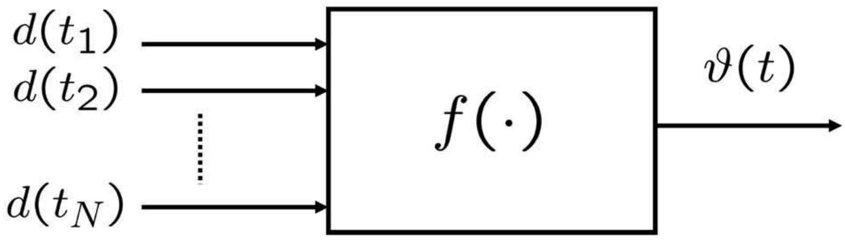
$$\vartheta(t)$$

Nella gran maggioranza dei casi i parametri incogniti sono costanti $\vartheta(t) = \vartheta$

$T = \{t_1, t_2, \ldots, t_N\}$ insieme degli istanti di osservazione

- In generale non e` detto che i t_i debbano avere cadenza regolare
- Se si ha accesso alla generazione dei t_i conviene "addensarli" dove le osservazioni sperimentali sono piu` significative

Stimatore



Lo stimatore e` quindi una funzione deterministica che a partire dai dati osservati in ingresso produce in uscita i parametri incogniti.

Stima di parametri costanti

- Se $\vartheta(t) = \bar{\vartheta} = \text{cost}$ si parla di stima o identificazione parametrica.
- La stima prodotta dallo stimatore si indica con $\widehat{\vartheta}$ o con $\widehat{\vartheta}_T$ qualora si voglia evidenziare l'insieme degli istanti di osservazione.
- Il valore "vero" del parametro incognito si indica con ϑ°

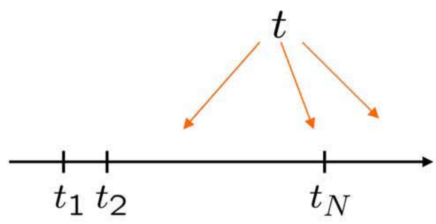
Stima di parametri varianti nel tempo

- La stima prodotta dallo stimatore si indica con $\widehat{\vartheta}(t|T)$ o semplicemente con $\widehat{\vartheta}(t|N)$ qualora si possa prendere $T = \{1, 2, \ldots, N\}$
- A seconda di come si ponga l'istante t rispetto a t_N si hanno i tre casi:

$t > t_N$: problema di predizione

$t = t_N$: problema di filtraggio

$t < t_N$: problema di regolarizzazione (smoothing)



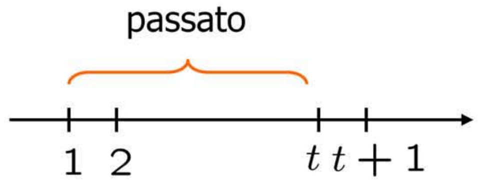
Problema della predizione

Si tratta di un problema fondamentale nell'ambito dell'identificazione dei sistemi dinamici.

- Per fissare le idee consideriamo il caso delle *serie temporali*.
- Si ha una sequenza di osservazioni $y(1), y(2), \ldots, y(N)$ di una variabile $y(\cdot)$
- Si vuole stimare $y(t+1)$
- Si vuole quindi determinare un predittore

$$\boxed{\widehat{y}(t+1|t) = f[y(t), y(t-1), \ldots, y(1)]}$$

- Il predittore esprime quindi una stima $\widehat{y}(t+1|t)$ di $y(t+1)$ in funzione di N valori assunti da $y(\cdot)$ nel passato.



- Un predittore si dice lineare se

$$\widehat{y}(t+1|t) = a_1(t)\, y(t) + \cdots + a_t(t)\, y(1)$$

- Un predittore si dice a memoria finita (ovvero utilizza una memoria limitata del passato) se

$$\widehat{y}(t+1|t) = a_1(t)\, y(t) + \cdots + a_n(t)\, y(t-n+1)$$

- Un predittore si dice lineare tempo-invariante se

$$\widehat{y}(t+1|t) = a_1\, y(t) + \cdots + a_n\, y(t-n+1)$$

dove i parametri $a_1, \ldots, a_n$ sono costanti

- Si definisce il vettore dei parametri $\vartheta^\top = [a_1 \cdots a_n]$

Determinare un buon predittore significa determinare un opportuno vettore ϑ tale che la predizione $\widehat{y}(t+1|t)$ sia il piu` possibile accurata.

Piu` precisamente:

- Si consideri un predittore lineare tempo-invariante a memoria finita

$$\widehat{y}(t+1|t) = a_1\,y(t) + \cdots + a_n\,y(t-n+1)$$

dove n e` piccolo rispetto al numero di dati misurati fino all'istante t

- La capacita` predittiva del predittore puo` quindi essere valutata rispetto ai dati gia` noti $y(i),\ i = 1, \ldots, t$:
  - si calcola

$$\widehat{y}(i+1|i) = a_1\,y(i) + \cdots + a_n\,y(i-n+1),\ \forall\, i > n$$

  - si valuta l'errore di predizione

$$\varepsilon(i+1) = y(i+1) - \widehat{y}(i+1|i),\ \forall\, i > n$$

Il vettore $\vartheta^\top = [a_1 \cdots a_n]$ e` "buono" se $\varepsilon$ e` piccolo su tutti i dati disponibili

- definiamo allora una cifra di merito:

$$J(\vartheta) = \sum_{i=n+1}^{t} \varepsilon(i)^2$$

- quindi

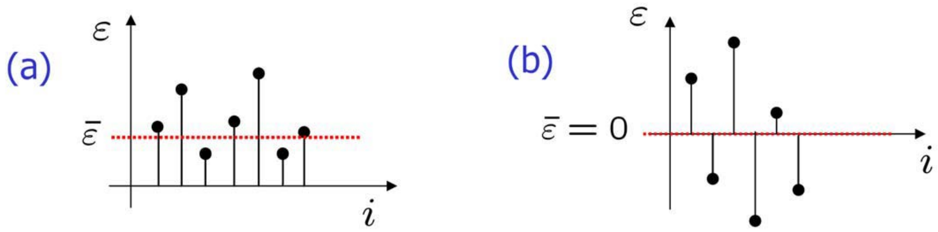
$$\vartheta^\circ = \arg\min_{\vartheta} J(\vartheta)$$

La determinazione di ϑ° e` quindi ricondotta alla soluzione di un problema di ottimizzazione.

Osservazioni

E` molto importante chiarire che cosa si intende per ε "piccolo"

la minimizzazione di $J(\vartheta)$ non e`di per se` un criterio soddisfacente



- Caso (a): non e` soddisfacente perche` l'errore medio $\bar{\varepsilon}$ e` non nullo → errore sistematico
- Caso (b): non e` soddisfacente perche` nonostante l'errore medio $\bar{\varepsilon}$ sia nullo, la sequenza e` alternata, quindi ad ogni passo si sa gia` che segno avra` l'errore al passo successivo

Il predittore non ha acquisito tutta l'informazione

Quindi:

La situazione ideale sarebbe avere un errore ε di valor medio il piu` piccolo possibile e che abbia una natura il piu` possibile impredicibile.

$$\varepsilon(\cdot) \sim WN(0, \lambda^2)$$

varianza

media

rumore bianco (white noise)

Predittore come sistema dinamico

$$\widehat{y}(t|t-1) = a_1\, y(t-1) + \cdots + a_n\, y(t-n)$$

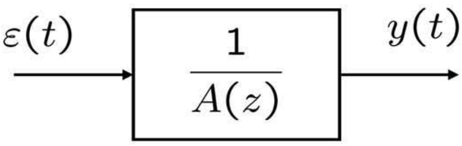
$$\varepsilon(t) = y(t) - \widehat{y}(t|t-1) \quad \Longrightarrow \quad y(t) = \varepsilon(t) + \widehat{y}(t|t-1)$$

$$y(t) = a_1\, y(t-1) + \cdots + a_n\, y(t-n) + \varepsilon(t)$$

$$y(t) = \left(a_1\, z^{-1} + \cdots + a_n\, z^{-n}\right) y(t) + \varepsilon(t)$$

$$A(z)\, y(t) = \varepsilon(t) \quad \text{con} \quad A(z) = 1 - a_1\, z^{-1} - \cdots - a_n\, z^{-n}$$

$$y(t) = \frac{1}{A(z)}\, \varepsilon(t)$$



Teoria della Stima e Caratteristiche degli Stimatori

Generalita`

- In generale abbiamo:

$$d = d(s, \vartheta^\circ)$$

dove

  - d sono i dati osservati
  - ϑ° e` la quantita` da stimare
  - s e` l'esito dell'esperimento casuale

- Lo stimatore e` una funzione:

$$\widehat{\vartheta} = f\left[d(s, \vartheta^\circ)\right]$$

Lo stimatore e` variabile aleatoria in quanto il suo valore dipende dall'esito dell'esperimento casuale s

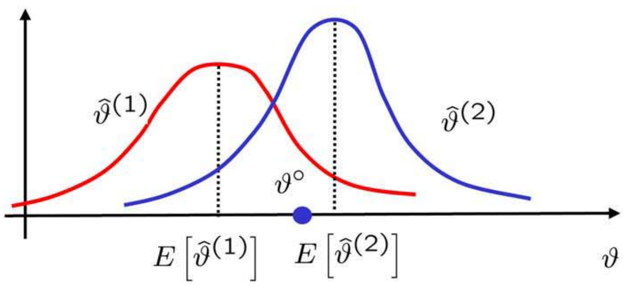
Polarizzazione

- In generale lo stimatore $\widehat{\vartheta} = f\left[d(s, \vartheta^\circ)\right]$ si dice non polarizzato se

$$E(\widehat{\vartheta}) = \vartheta^\circ$$

- La non polarizzazione di uno stimatore e` evidentemente una proprieta` che bisogna cercare di assicurare.

Nel caso in figura gli stimatori sono ambedue polarizzati ma lo stimatore $\widehat{\vartheta}^{(2)}$ ha una polarizzazione minore.



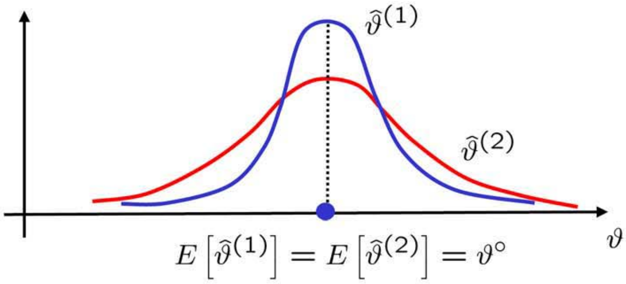
Minima varianza

- La non polarizzazione (o correttezza) non e` l'unico criterio con cui valutare la qualita` di uno stimatore.

Nel caso in figura gli stimatori sono ambedue non polarizzati.

Pero`

$$\text{var}\left[\hat{\vartheta}^{(1)}\right] << \text{var}\left[\hat{\vartheta}^{(2)}\right]$$



- Quindi lo stimatore $\hat{\vartheta}^{(1)}$ ha probabilita` maggiore di produrre valori della stima vicini al valore vero $\hat{\vartheta}^{\circ}$ rispetto allo stimatore $\hat{\vartheta}^{(2)}$

- Si vuole quindi che la varianza dello stimatore sia la piu` piccola possibile.

- In generale, a parita` di caratteristiche di polarizzazione, diremo che lo stimatore $\hat{\vartheta}^{(1)}$ e` migliore dello stimatore $\hat{\vartheta}^{(2)}$ se

$$\text{var}\left[\hat{\vartheta}^{(1)}\right] \leq \text{var}\left[\hat{\vartheta}^{(2)}\right]$$

ovvero se la matrice (ϑ puo` essere un vettore)

$$\text{var}\left[\hat{\vartheta}^{(2)}\right] - \text{var}\left[\hat{\vartheta}^{(1)}\right] \geq 0$$

- Ricordiamo che $A \geq 0 \Longrightarrow \det A \geq 0,\ \lambda_i \geq 0,\ a_{ii} \geq 0$

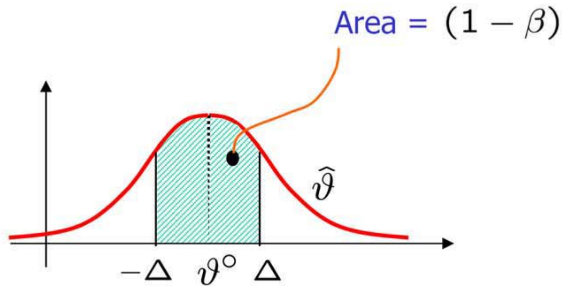
Quindi

$$\text{var}\left[\hat{\vartheta}^{(2)}\right] - \text{var}\left[\hat{\vartheta}^{(1)}\right] \geq 0 \Longrightarrow \text{var}\left[\hat{\vartheta}_i^{(2)}\right] \geq \text{var}\left[\hat{\vartheta}_i^{(1)}\right]$$

dove $\hat{\vartheta}_i^{(1)}, \hat{\vartheta}_i^{(2)}$ rappresentano le componenti i-esime dei vettori $\hat{\vartheta}^{(1)}, \hat{\vartheta}^{(2)}$

Confidenza della stima

Consideriamo uno stimatore $\hat{\vartheta}$:

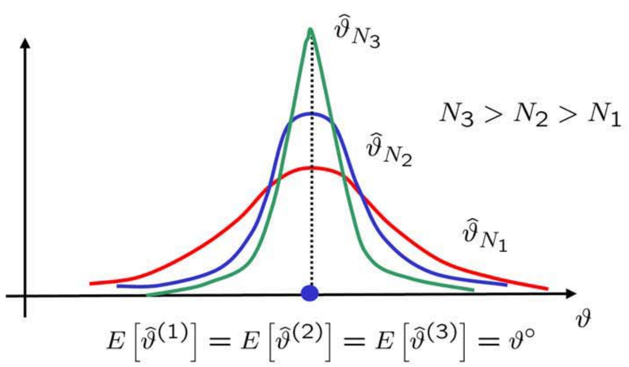


Si dice che la stima $\hat{\vartheta}$ e` entro l'intervallo $(-\Delta, \Delta)$ attorno a ϑ° con confidenza $(1-\beta)\cdot 100\%$

Caratteristiche asintotiche

- Se il numero N di dati a disposizione cresce nel tempo
  - aumenta l'informazione a disposizione per effettuare la stima
    - diminuisce l'incertezza
- Da questo punto di vista uno stimatore $\hat{\vartheta}_N$ e` buono se

$$\lim_{N\to\infty} \text{var}\left[\hat{\vartheta}_N\right] = 0$$



- Un'altra caratterizzazione della bonta` di uno stimatore $\hat{\vartheta}_N$ in cui la stima viene determinata in base ad un numero $N$ di dati crescente nel tempo e`

$$\lim_{N\to\infty} E\left[\left\|\hat{\vartheta}_N - \vartheta^\circ\right\|^2\right] = 0 \quad (\star)$$

Se vale la $(\star)$ si dice che la stima $\hat{\vartheta}_N$ converge in media quadratica a ϑ°

- Osserviamo che $\hat{\vartheta}_N$ e` un vettore casuale, $\vartheta^\circ$ e` un vettore costante e $\left\|\hat{\vartheta}_N - \vartheta^\circ\right\|$ e` una variabile aleatoria scalare per cui l'operazione "valore atteso" e` perfettamente legittima

Convergenza quasi certa

- Ricordiamo che lo stimatore basato su N dati e`

$$\hat{\vartheta}_N(s, \vartheta^\circ) = f\left[d(s, \vartheta^\circ)\right]$$

- Fissato $\bar{s} \in S$, si avra` una sequenza

$$\hat{\vartheta}_1(\bar{s}, \vartheta^\circ), \hat{\vartheta}_2(\bar{s}, \vartheta^\circ), \ldots, \hat{\vartheta}_N(\bar{s}, \vartheta^\circ), \ldots$$

- Potrebbe allora succedere che:

$$\bar{s} \in S \quad \Longrightarrow \quad \lim_{N\to\infty} \hat{\vartheta}_N(\bar{s}, \vartheta^\circ) = \vartheta^\circ$$

$$\tilde{s} \in S \quad \Longrightarrow \quad \lim_{N\to\infty} \hat{\vartheta}_N(\tilde{s}, \vartheta^\circ) \neq \vartheta^\circ$$

- Introduciamo l'insieme di esiti

$$A \subset S\,, \quad A = \left\{ s \in S : \lim_{N \to \infty} \widehat{\vartheta}_N(s, \vartheta^\circ) = \vartheta^\circ \right\}$$

- Se $A = S$ ⟹ Convergenza certa

- Se $A \subset S$ e $P(A) = 1$ ⟹ Convergenza quasi-certa

  Notiamo che se la misura dell'insieme $S \setminus A$ e` nulla cio` implica $P(A) = 1$ e quindi convergenza quasi-certa

- Evidentemente $A = S$ ⟹ $P(A) = 1$

  ↳ Convergenza certa ⟹ Convergenza quasi-certa

- Uno stimatore per cui si abbia convergenza quasi-certa si dice **consistente**

Esempio 1

- Si considerino N dati scalari $d(1), d(2), \ldots, d(N)$ tali che

$$E\,[d(i)] = \vartheta^\circ,\ i = 1, \ldots, N$$

- Si supponga che i dati siano mutuamente scorrelati, cioe`

$$E\,\{[d(i) - \vartheta^\circ]\,[d(j) - \vartheta^\circ]\} = 0\,, \quad \forall\, i \neq j$$

- Si consideri lo stimatore

$$\widehat{\vartheta}_N = \frac{1}{N} \sum_{i=1}^{N} d(i)$$

Stimatore a media campionaria

- Polarizzazione:

$$E\left[\hat{\vartheta}_N\right] = E\left\{\frac{1}{N}\sum_{i=1}^{N} d(i)\right\} = \frac{1}{N}\sum_{i=1}^{N} E\,[d(i)] = \frac{1}{N}\sum_{i=1}^{N} \vartheta^\circ = \vartheta^\circ$$

→ lo stimatore non e` polarizzato

- Varianza:

$$\text{var}\left(\hat{\vartheta}_N\right) = E\left\{\left[\hat{\vartheta}_N - E(\hat{\vartheta}_N)\right]^2\right\} = E\left\{\left[\frac{1}{N}\sum_{i=1}^{N} d(i) - \frac{1}{N}\sum_{i=1}^{N}\vartheta^\circ\right]^2\right\}$$

$$= E\left\{\frac{1}{N^2}\left[\sum_{i=1}^{N} d(i) - \sum_{i=1}^{N}\vartheta^\circ\right]^2\right\} = \frac{1}{N^2}\sum_{i=1}^{N} E\left\{[d(i)-\vartheta^\circ]^2\right\}$$

$$= \frac{1}{N^2}\sum_{i=1}^{N} \text{var}\,[d(i)]$$

(i termini "misti" sono nulli per l'ipotesi di dati mutuamente scorrelati)

Se $\text{var}\,[d(i)] \leq \sigma\,,\ 1 = 1,\ldots,N$

→ $\lim_{N\to\infty} \text{var}\left(\hat{\vartheta}_N\right) \leq \lim_{N\to\infty}\frac{\sigma}{N} = 0$ ➡ lo stimatore converge in media quadratica

Esempio 2

- Si considerino N dati scalari $d(1), d(2), \ldots, d(N)$ tali che

$$E\,[d(i)] = \vartheta^\circ,\ i = 1,\ldots,N$$

- Si supponga che i dati siano mutuamente scorrelati, cioe`

$$E\left\{[d(i)-\vartheta^\circ]\,[d(j)-\vartheta^\circ]\right\} = 0\,,\quad \forall\, i \neq j$$

- Si consideri lo stimatore

$$\hat{\vartheta}_N = \sum_{i=1}^{N} \alpha(i)\, d(i)$$

- Polarizzazione:

$$E\left[\widehat{\vartheta}_N\right] = E\left\{\sum_{i=1}^{N} \alpha(i)\, d(i)\right\} = \sum_{i=1}^{N} \alpha(i)\, E\left[d(i)\right] = \vartheta^\circ \sum_{i=1}^{N} \alpha(i)$$

lo stimatore non e` polarizzato $\Longleftrightarrow$ $\sum_{i=1}^{N} \alpha(i) = 1 \quad (\star)$

N.B. Nel caso precedente $\alpha(i) = \dfrac{1}{N}$ per cui $(\star)$ e` soddisfatta

La condizione $(\star)$ e` un vincolo da soddisfare affinche` lo stimatore non sia polarizzato e caratterizza una classe di infiniti stimatori possibili

- Determiniamo lo stimatore migliore tra quelli non polarizzati (quindi che soddisfano il vincolo $(\star)$) andando a scegliere quello di minima varianza

dati scorrelati

$$\begin{cases} \min \text{var}\left(\widehat{\vartheta}_N\right) = \min \sum_{i=1}^{N} [\alpha(i)]^2 \,\text{var}\,[d(i)] \\ 1 - \sum_{i=1}^{N} \alpha(i) = 0 \end{cases}$$

Usando il metodo dei moltiplicatori di Lagrange

$$J(\widehat{\vartheta}) = \sum_{i=1}^{N} [\alpha(i)]^2 \,\text{var}\,[d(i)] + \lambda\left(1 - \sum_{i=1}^{N} \alpha(i)\right)$$

$$\frac{\partial J}{\partial \alpha(i)} = 0 \Longleftrightarrow 2\,\alpha(i)\text{var}\,[d(i)] - \lambda = 0 \Longleftrightarrow \alpha(i) = \frac{\lambda}{2\text{var}\,[d(i)]}$$

- Imponendo il vincolo di non polarizzazione $(\star)$

$$\sum_{i=1}^{N} \alpha(i) = 1 \Longleftrightarrow \frac{\lambda}{2} \sum_{i=1}^{N} \frac{1}{\text{var}\,[d(i)]} = 1 \Longleftrightarrow \lambda = \frac{2}{\sum_{i=1}^{N} \frac{1}{\text{var}\,[d(i)]}}$$

$$\alpha(i) = \frac{1}{\text{var}\,[d(i)]}\,\alpha \quad \text{con} \quad \alpha = \frac{1}{\sum_{i=1}^{N} \frac{1}{\text{var}\,[d(i)]}}$$

Quindi $\alpha(i)$ viene scelto inversamente proporzionale alla varianza del dato $\text{var}\,[d(i)]$: piu\` e\` grande e meno peso gli viene associato.

- Calcoliamo ora la varianza dello stimatore:

$$\text{var}\left(\hat{\vartheta}_N\right) = E\left\{\left[\hat{\vartheta}_N - E(\hat{\vartheta}_N)\right]^2\right\} = E\left\{\left[\sum_{i=1}^{N} \alpha(i)\,d(i) - \vartheta^\circ \sum_{i=1}^{N} \alpha(i)\right]^2\right\}$$

$$= E\left\{\left[\sum_{i=1}^{N} \alpha(i)\,[d(i) - \vartheta^\circ]\right]^2\right\} = \sum_{i=1}^{N} \alpha(i)^2 E\left\{[d(i) - \vartheta^\circ]^2\right\}$$

$$= \sum_{i=1}^{N} \alpha(i)^2 \text{var}\,[d(i)] = \alpha^2 \sum_{i=1}^{N} \frac{1}{\text{var}\,[d(i)]} = \frac{1}{\sum_{i=1}^{N} \frac{1}{\text{var}\,[d(i)]}}$$

Se $\text{var}\,[d(i)] \leq \sigma\,,\ 1 = 1, \ldots, N$

$$\lim_{N\to\infty} \text{var}\left(\hat{\vartheta}_N\right) \leq \lim_{N\to\infty} \frac{\sigma}{N} = 0 \Longrightarrow$$ lo stimatore converge in media quadratica

Generalizzazione

- Nel caso in cui le quantita` da stimare siano varianti nel tempo e` necessario modificare gli indici di bonta` degli stimatori
- Si indichi con $\widehat{\vartheta}(t|t-1)$ la stima di $\vartheta^{\circ}(t)$ in base ai dati fino all'istante $t-1$
- Evidentemente, siccome $\vartheta^{\circ}(t)$ varia nel tempo, non e` sensato parlare di convergenza asintotica rispetto alla disponibilita` di dati in quanto i dati nel passato possono non essere piu` significativi.
- Un criterio tipico e`

$$E\left[\left\|\widehat{\vartheta}(t|t-1) - \vartheta^{\circ}(t)\right\|^2\right] \leq c$$

dove c e` una costante ragionevolmente piccola

- Non si chiede convergenza ma si chiede "non divergenza"

Stima ai Minimi Quadrati (LSM)

Regressione lineare

- Si tratta del tipico contesto in cui si usa lo stimatore ai minimi quadrati (MQ)
- Si hanno $q+1$ variabili $y(t), u_1(t), \ldots, u_q(t)$ sull'arco temporale $t = 1, 2, \ldots, N$
- Si vogliono calcolare (se possibile) q parametri $\vartheta_1, \vartheta_2, \ldots, \vartheta_q$ tali che

$$y(t) = \vartheta_1 \, u_1(t) + \cdots \vartheta_q \, u_q(t)\,, \quad t = 1, \ldots, N \quad (\star)$$

$(\star)$ viene definita come la regressione lineare della variabile $y(t)$ sulle variabili $u_1(t), \ldots, u_q(t)$

- Il problema puo` essere formulato in maniera vettoriale

$$\vartheta = \begin{bmatrix} \vartheta_1 \\ \vdots \\ \vartheta_q \end{bmatrix} \qquad \varphi(t) = \begin{bmatrix} u_1(t) \\ \vdots \\ u_q(t) \end{bmatrix}$$

$$y(t) = \varphi(t)^\top \vartheta$$

- Nei problemi reali in realta` si avra` un certo errore

$$\varepsilon(t) = y(t) - \varphi(t)^\top \vartheta$$

- Obiettivo del problema delle regressione lineare sara` quindi quello di minimizzare l'errore $\varepsilon(t)$ determinando un vettore ϑ° per cui questo minimo venga raggiunto

- Si definisce la funzione di costo quadratica

$$J(\vartheta) = \sum_{t=1}^{N} [\varepsilon(t)]^2 = \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta\right]^2$$

$$\vartheta^\circ = \arg \min_{\vartheta} J(\vartheta)$$

Stimatore ai minimi quadrati

- Indicando con ϑ_i la componente i-esima del vettore ϑ

$$\frac{\partial J}{\partial \vartheta_i} = \frac{\partial}{\partial \vartheta_i} \left\{ \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta\right]^2 \right\}$$

$$= -2 \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta\right] u_i(t), \quad i = 1, 2, \dots, q$$

ed osservando che $\frac{\partial J}{\partial \vartheta} = \left[\frac{\partial J}{\partial \vartheta_1} \; \frac{\partial J}{\partial \vartheta_2} \; \cdots \; \frac{\partial J}{\partial \vartheta_q}\right]$

$$\frac{\partial J}{\partial \vartheta} = -2 \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta\right] \varphi(t)^\top$$

- Imponendo $\dfrac{\partial J}{\partial \vartheta} = [0\ 0\ \cdots\ 0]$

$$\Rightarrow \quad -2 \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta \right] \varphi(t)^\top = [0\ 0\ \cdots\ 0]$$

$$\Rightarrow \quad \sum_{t=1}^{N} y(t)\, \varphi(t)^\top = \sum_{t=1}^{N} \varphi(t)^\top \vartheta\, \varphi(t)^\top$$

convertendo l'uguaglianza tra vettori riga in un'uguaglianza tra vettori colonna si puo` mettere in evidenza il vettore ϑ :

$$\boxed{\sum_{t=1}^{N} \varphi(t)\, y(t) = \left[\sum_{t=1}^{N} \varphi(t)\, \varphi(t)^\top \right] \vartheta}$$

Equazioni normali dei minimi quadrati (q eqz. in q incognite)

- Se $\sum_{t=1}^{N} \varphi(t)\, \varphi(t)^\top$ e` non singolare

Formula dei minimi quadrati

$$\Rightarrow \quad \boxed{\hat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\, \varphi(t)^\top \right]^{-1} \sum_{t=1}^{N} \varphi(t)\, y(t)}$$

Interpretazione geometrica delle eq. normali

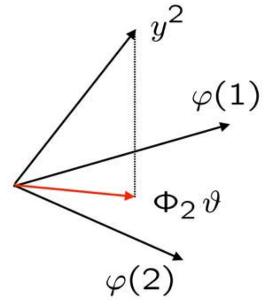
Poniamo

$$\varepsilon_\vartheta^N = \begin{bmatrix} \varepsilon_\vartheta(1) \\ \vdots \\ \varepsilon_\vartheta(N) \end{bmatrix} \qquad y^N = \begin{bmatrix} y(1) \\ \vdots \\ y(N) \end{bmatrix} \qquad \Phi_N = \begin{bmatrix} \varphi(1)^\top \\ \vdots \\ \varphi(N)^\top \end{bmatrix}$$

Scriviamo quindi

$$J(\vartheta) = \sum_{t=1}^{N} \left[y(t) - \varphi(t)^\top \vartheta \right]^2 = \left\| y^N - \Phi_N\, \vartheta \right\|^2$$

Evidentemente $\left\| y^N - \Phi_N\, \vartheta \right\|$ e` minima quando $y^N - \Phi_N\, \vartheta$ e` ortogonale a $\Phi_N\, \vartheta$



- Verifichiamo che $\widehat{\vartheta}_N$ sia un minimo valutando la definitezza della matrice simmetrica

$$\left[\frac{d^2 J}{d\vartheta^2}\right]_{i,j} = \frac{\partial^2 J}{\partial \vartheta_i \partial \vartheta_j}, \quad i,j = 1,\ldots,q$$

si ha

$$\left(\frac{\partial J}{\partial \vartheta}\right)^\top = 2\left\{\left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]\vartheta - \sum_{t=1}^{N}\varphi(t)\,y(t)\right\}$$

$$\Rightarrow \quad \frac{d^2 J}{d\vartheta^2} = 2\left[\sum_{t=1}^{N}\varphi(t)\,\varphi(t)^\top\right]$$

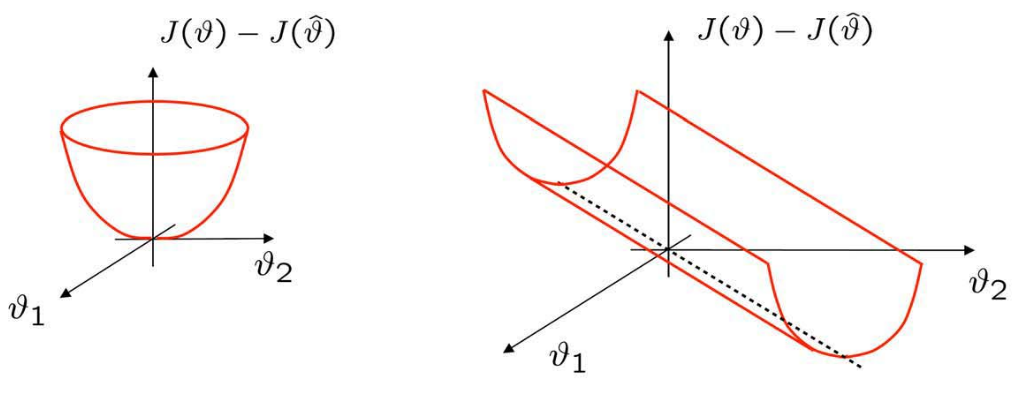
evidentemente si tratta di una matrice simmetrica e semidefinita positiva

$\Rightarrow$ $\widehat{\vartheta}_N$ e` un minimo locale di $J(\vartheta)$

- Considerando quindi la forma quadratica

$$J(\vartheta) - J(\widehat{\vartheta}) = \frac{1}{2}(\vartheta - \widehat{\vartheta})^\top \left.\frac{d^2 J}{d\vartheta^2}\right|_{\widehat{\vartheta}} (\vartheta - \widehat{\vartheta})$$

si hanno i due casi possibili:



$\det\left[\sum_{t=1}^{N}\varphi(t)\,\varphi(t)^\top\right] \neq 0$ $\qquad$ $\det\left[\sum_{t=1}^{N}\varphi(t)\,\varphi(t)^\top\right] = 0$

- Quindi:

  - Se $\det \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top \right] \neq 0$ ⟹ $\widehat{\vartheta}_N$ unico minimo globale
  - Se $\det \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top \right] = 0$ ⟹ $\widehat{\vartheta}_N$ e` uno degli infiniti minimi globali

- La condizione $\det \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top \right] \neq 0$ e` detta

  e` detta condizione di identificabilita`

Caratteristiche probabilistiche dello stimatore MQ

- Supponiamo che la condizione di identificabilita` sia verificata:

$$\det \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top \right] \neq 0$$

per cui

$$\widehat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top \right]^{-1} \sum_{t=1}^{N} \varphi(t)\,y(t)$$

- Ipotesi: $y(t) = \varphi(t)^\top \vartheta^\circ + \xi(t)$

  dove il processo e` scorrelato da $u(\cdot)$ e $E[\xi(t)] = 0$

  Si sta quindi ipotizzando che il legame vero tra $y(t)$ e $u_1(t), \dots, u_q(t)$ sia proprio lineare + rumore scorrelato e a media nulla

Polarizzazione

$$\widehat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,y(t)$$

$$= \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,\left[\varphi(t)^\top \vartheta^\circ + \xi(t)\right]$$

$$= \vartheta^\circ + \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,\xi(t)$$

$$\widehat{\vartheta}_N - \vartheta^\circ = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,\xi(t)$$

$$E\left(\widehat{\vartheta}_N - \vartheta^\circ\right) = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,E[\xi(t)] = 0$$

$E\left(\widehat{\vartheta}_N\right) = \vartheta^\circ$ **Stimatore MQ e` non polarizzato**

OSSERVAZIONE IMPORTANTE

Nell'analisi di polarizzazione dello stimatore ai MQ abbiamo considerato il vettore dei regressori $\varphi(t)$ come noto e fissato. Se invece l'analisi fosse stata effettuata considerando il vettore $\varphi(s,t)$ come vettore casuale (e quindi funzione dell'esito s di un esperimento casuale) allora si otterrebbe che lo stimatore ai MQ e` polarizzato per qualunque valore finito di

$$E\left[\hat{\vartheta}_N - \vartheta^o\right] = E\left\{\left[\sum_{t=1}^{N} \varphi(t)\varphi(t)^T\right]^{-1} \sum_{t=1}^{N} \varphi(t)\xi(t)\right\}$$

Varianza

Ulteriore ipotesi: $\xi(t) \sim WN(0, \lambda^2)$

Introduciamo la matrice simmetrica $S(N) = \sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top$

$$\text{var}\left(\hat{\vartheta}_N\right) = E\left[\left(\hat{\vartheta}_N - \vartheta^\circ\right)\left(\hat{\vartheta}_N - \vartheta^\circ\right)^\top\right]$$

$$= E\left\{\left[S(N)^{-1}\sum_{t=1}^{N}\varphi(t)\,\xi(t)\right]\left[S(N)^{-1}\sum_{s=1}^{N}\varphi(s)\,\xi(s)\right]^\top\right\}$$

$S(N)$ simmetrica

$$= E\left\{\left[S(N)^{-1}\sum_{t=1}^{N}\varphi(t)\,\xi(t)\right]\left[\sum_{s=1}^{N}\xi(s)\,\varphi(s)^\top S(N)^{-1}\right]\right\}$$

$$= S(N)^{-1}\,E\left[\sum_{t=1}^{N}\varphi(t)\,\xi(t)\sum_{s=1}^{N}\xi(s)\,\varphi(s)^\top\right]S(N)^{-1}$$

Nel prodotto $\sum_{t=1}^{N}\varphi(t)\,\xi(t)\sum_{s=1}^{N}\xi(s)\,\varphi(s)^\top$ si hanno due tipi di termini:

- $\varphi(t)\,\xi(t)^2\,\varphi(t)^\top$ se $t = s$
- $\varphi(t)\,\xi(t)\,\xi(s)\,\varphi(s)^\top$ se $t \neq s$

Ma $\xi(t) \sim WN(0, \lambda^2)$ $\Rightarrow$ $E\left[\xi(t)\xi(s)\right] = \begin{cases} \lambda^2 & \text{se } t = s \\ 0 & \text{se } t \neq s \end{cases}$

$$E\left[\sum_{t=1}^{N}\varphi(t)\,\xi(t)\sum_{s=1}^{N}\xi(s)\,\varphi(s)^\top\right] = \sum_{t=1}^{N}\lambda^2\,\varphi(t)\,\varphi(t)^\top = \lambda^2\,S(N)$$

$$\text{var}\left(\hat{\vartheta}_N\right) = S(N)^{-1}\,\lambda^2\,S(N)\,S(N)^{-1} = \lambda^2\,S(N)^{-1}$$

Interpretazione

Supponiamo che ϑ° sia scalare per cui anche $\varphi(t)$ e` scalare

$$y(t) = \varphi(t)\,\vartheta^\circ + \xi(t) = u(t)\,\vartheta^\circ + \xi(t)$$

da cui

$$\hat{\vartheta}_N = \left[\sum_{t=1}^{N} \varphi(t)\,\varphi(t)^\top\right]^{-1} \sum_{t=1}^{N} \varphi(t)\,y(t) = \frac{\frac{1}{N}\sum_{t=1}^{N} u(t)\,y(t)}{\frac{1}{N}\sum_{t=1}^{N} u(t)^2}$$

Ma:

- $\frac{1}{N}\sum_{t=1}^{N} u(t)\,y(t)$ stima campionaria della cross-correlazione $E\left[u(t)\,y(t)\right]$
- $\frac{1}{N}\sum_{t=1}^{N} u(t)^2$ stima campionaria di $E\left[u(t)^2\right]$ (varianza se $E\,(u) = 0$)

Inoltre

$$\text{var}\left(\hat{\vartheta}_N\right) = \lambda^2\,S(N)^{-1} = \frac{1}{N}\,\frac{\lambda^2}{\frac{1}{N}\sum_{t=1}^{N} u(t)^2}$$

Pertanto:

- var $\left(\hat{\vartheta}_N\right)$ cresce al crescere di λ^2 ➡ L'incertezza sulla stima aumenta all'aumentare dell'incertezza sui dati
- a parita` di $N$ e $\lambda^2$ , var $\left(\hat{\vartheta}_N\right)$ diminuisce all'aumentare della varianza campionaria di $u$ il che e` intuitivamente corretto in quanto il rumore viene "sopraffatto" dal segnale contenente informazione utile

- $\dfrac{\lambda^2}{\dfrac{1}{N}\sum_{t=1}^{N} u(t)^2}$ e` una sorta di rapporto rumore/segnale

- Se la varianza di u e` limitata ➡ $\lim_{N\to\infty} \text{var}\left(\hat{\vartheta}_N\right) = 0$

e siccome lo stimatore non e` polarizzato si ha:

$$\lim_{N\to\infty} E\left(\left\|\hat{\vartheta}_N - \vartheta^\circ\right\|^2\right) = 0$$

ovvero lo stimatore MQ converge in media quadratica

Riferimenti Bibliografici

- C. Bonivento, C. Melchiorri, R. Zanasi. *Sistemi di controllo digitale*, Esculapio ed., Bologna, 1999
- Gene F. Franklin, J. David Powell, Abbas Emami-Naeini, *CONTROLLO A RETROAZIONE DI SISTEMI DINAMICI*, volume II. EdiSES s.r.l. 2005. ISBN: 88 7959 308 0
- S. Bittanti. *Teoria della Predizione e Filtraggio*. Pitagora, Bologna, 1992
- S. Bittanti. *Identificazione dei Modelli e Controllo adattativo*. Pitagora, Bologna, 1992