

# Sistemi di Controllo Digitale

## Identificazione di Modelli Dinamici

# Contenuto delle Lezioni

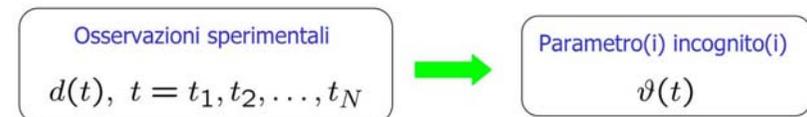
- Il problema della stima e della predizione
- Teoria della stima e caratteristiche degli stimatori
- Stima ai minimi quadrati (LMS)

08/05/2014

2

# Il Problema della Stima e della Predizione

Il problema della stima nasce quando si vogliono determinare uno o più **parametri incogniti a partire da osservazioni sperimentali**.



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

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

- In generale non è 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 più significative

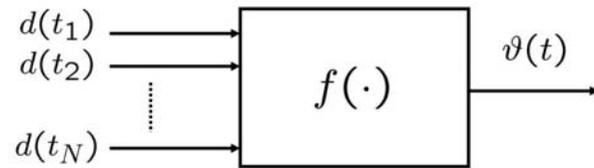
08/05/2014

3

08/05/2014

4

# 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

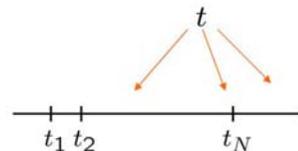
- Se  $v(t) = \bar{v} = \text{cost}$  si parla di **stima o identificazione parametrica**.
- La stima prodotta dallo stimatore si indica con  $\hat{v}$  o con  $\hat{v}_T$  qualora si voglia evidenziare l'insieme degli istanti di osservazione.
- Il valore "vero" del parametro incognito si indica con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

## Stima di parametri varianti nel tempo

- La stima prodotta dallo stimatore si indica con  $\hat{v}(t|T)$  o semplicemente con  $\hat{v}(t|N)$  qualora si possa prendere  $T = \{1, 2, \dots, N\}$

- 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)**



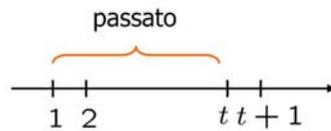
## Problema della predizione

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  $y(1), y(2), \dots, y(N)$  di una variabile  $y(\cdot)$
- Si vuole stimare  $y(t + 1)$
- Si vuole quindi determinare un **predittore**

$$\hat{y}(t + 1|t) = f[y(t), y(t - 1), \dots, y(1)]$$

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



- Un predittore si dice **lineare** se

$$\hat{y}(t+1|t) = a_1(t)y(t) + \dots + a_t(t)y(1)$$

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

$$\hat{y}(t+1|t) = a_1(t)y(t) + \dots + a_n(t)y(t-n+1)$$

08/05/2014

9

- Un predittore si dice **lineare tempo-invariante** se

$$\hat{y}(t+1|t) = a_1 y(t) + \dots + a_n y(t-n+1)$$

dove i parametri  $a_1, \dots, a_n$  sono **costanti**

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

Determinare un buon predittore significa determinare un opportuno vettore  $\vartheta$  tale che la predizione  $\hat{y}(t+1|t)$  sia il piu' possibile accurata.

08/05/2014

10

**Piu' precisamente:**

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

$$\hat{y}(t+1|t) = a_1 y(t) + \dots + a_n y(t-n+1)$$

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  $y(i), i = 1, \dots, t$ :

- si calcola

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

- si valuta **l'errore di predizione**

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

08/05/2014

11

Il vettore  $\vartheta^\top = [a_1 \dots a_n]$  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  $\vartheta^\circ$  e' quindi ricondotta alla soluzione di un problema di ottimizzazione.

08/05/2014

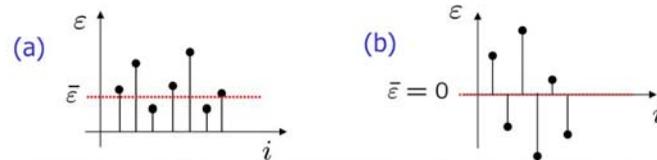
12

## Osservazioni

E' molto importante chiarire che cosa si intende per  $\varepsilon$  "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**

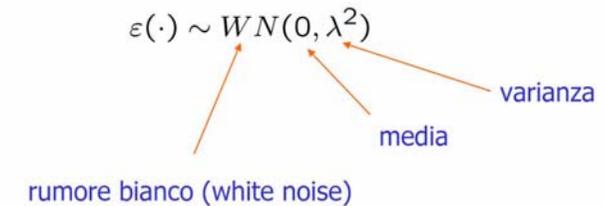


08/05/2014

13

## Quindi:

La situazione ideale sarebbe avere un errore  $\varepsilon$  di valor medio il piu' piccolo possibile e che abbia una natura il piu' possibile **impredicibile**.



08/05/2014

14

## Predittore come sistema dinamico

$$\hat{y}(t|t-1) = a_1 y(t-1) + \dots + a_n y(t-n)$$

$$\varepsilon(t) = y(t) - \hat{y}(t|t-1) \quad \rightarrow \quad y(t) = \varepsilon(t) + \hat{y}(t|t-1)$$

$$\rightarrow y(t) = a_1 y(t-1) + \dots + a_n y(t-n) + \varepsilon(t)$$

$$\rightarrow y(t) = (a_1 z^{-1} + \dots + a_n z^{-n}) y(t) + \varepsilon(t)$$

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

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

08/05/2014

15

## Teoria della Stima e Caratteristiche degli Stimatori

08/05/2014

16

## Generalita`

- In generale abbiamo:

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

dove

- $d$  sono i dati osservati
- $\vartheta^\circ$  e` la quantita` da stimare
- $s$  e` l'esito dell'esperimento casuale

- Lo stimatore e` una funzione:

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



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

08/05/2014

17

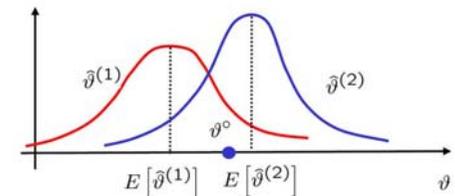
## Polarizzazione

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

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

- 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  $\hat{\vartheta}^{(2)}$  ha una polarizzazione minore.



08/05/2014

18

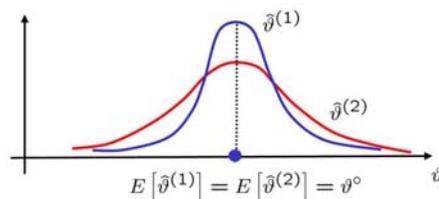
## Minima varianza

- 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`

$$\text{var} [\hat{\vartheta}^{(1)}] \ll \text{var} [\hat{\vartheta}^{(2)}]$$



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

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

08/05/2014

19

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

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

ovvero se la matrice (  $\vartheta$  puo` essere un vettore)

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

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

Quindi

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

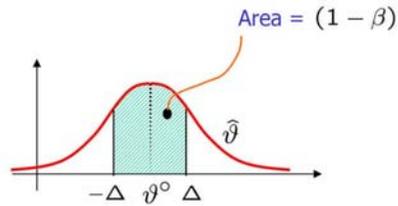
dove  $\hat{\vartheta}_i^{(1)}, \hat{\vartheta}_i^{(2)}$  rappresentano le componenti  $i$ -esime dei vettori  $\hat{\vartheta}^{(1)}, \hat{\vartheta}^{(2)}$

08/05/2014

20

## Confidenza della stima

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



Si dice che la stima  $\hat{\vartheta}$  è entro l'intervallo  $(-\Delta, \Delta)$  attorno a  $\vartheta^0$  con confidenza  $(1 - \beta) \cdot 100\%$

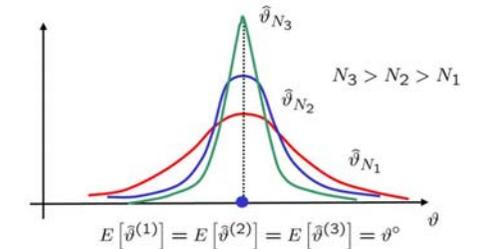
08/05/2014

21

## 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  $\hat{\vartheta}_N$  è buono se

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



08/05/2014

22

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

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

Se vale la (\*) si dice che la stima  $\hat{\vartheta}_N$  converge in media quadratica a  $\vartheta^0$

- Osserviamo che  $\hat{\vartheta}_N$  è un vettore casuale,  $\vartheta^0$  è un vettore costante e  $\|\hat{\vartheta}_N - \vartheta^0\|^2$  è una variabile aleatoria scalare per cui l'operazione "valore atteso" è perfettamente legittima

08/05/2014

23

## Convergenza quasi certa

- Ricordiamo che lo stimatore basato su  $N$  dati è

$$\hat{\vartheta}_N(s, \vartheta^0) = f[d(s, \vartheta^0)]$$

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

$$\hat{\vartheta}_1(\bar{s}, \vartheta^0), \hat{\vartheta}_2(\bar{s}, \vartheta^0), \dots, \hat{\vartheta}_N(\bar{s}, \vartheta^0), \dots$$

- Potrebbe allora succedere che:

$$\bar{s} \in S \quad \longrightarrow \quad \lim_{N \rightarrow \infty} \hat{\vartheta}_N(\bar{s}, \vartheta^0) = \vartheta^0$$

$$\bar{s} \in S \quad \longrightarrow \quad \lim_{N \rightarrow \infty} \hat{\vartheta}_N(\bar{s}, \vartheta^0) \neq \vartheta^0$$

08/05/2014

24

- Introduciamo l'insieme di esiti

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

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

Notiamo che se la misura dell'insieme  $S \setminus A$  è nulla cioè 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**

08/05/2014

25

## Esempio 1

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

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

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

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

- Si consideri lo stimatore

$$\hat{\vartheta}_N = \frac{1}{N} \sum_{i=1}^N d(i) \quad \text{Stimatore a media campionaria}$$

08/05/2014

26

- Polarizzazione:

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

 lo stimatore non è polarizzato

- Varianza:

$$\begin{aligned} \text{var}(\hat{\vartheta}_N) &= E\left\{ [\hat{\vartheta}_N - E(\hat{\vartheta}_N)]^2 \right\} = E\left\{ \left[ \frac{1}{N} \sum_{i=1}^N d(i) - \frac{1}{N} \sum_{i=1}^N \vartheta^\circ \right]^2 \right\} \\ &= E\left\{ \frac{1}{N^2} \left[ \sum_{i=1}^N d(i) - \sum_{i=1}^N \vartheta^\circ \right]^2 \right\} = \frac{1}{N^2} \sum_{i=1}^N E\{[d(i) - \vartheta^\circ]^2\} \\ &= \frac{1}{N^2} \sum_{i=1}^N \text{var}[d(i)] \quad (\text{i termini "misti" sono nulli per l'ipotesi di dati mutuamente scorrelati}) \end{aligned}$$

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

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

08/05/2014

27

## Esempio 2

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

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

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

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

- Si consideri lo stimatore

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

08/05/2014

28

- Polarizzazione:

$$E[\hat{\vartheta}_N] = E\left\{\sum_{i=1}^N \alpha(i) d(i)\right\} = \sum_{i=1}^N \alpha(i) E[d(i)] = \vartheta^0 \sum_{i=1}^N \alpha(i)$$

↳ lo stimatore non è polarizzato ↔  $\sum_{i=1}^N \alpha(i) = 1$  (\*)

N.B. Nel caso precedente  $\alpha(i) = \frac{1}{N}$  per cui (\*) è soddisfatta

La condizione (\*) è un vincolo da soddisfare affinché lo stimatore non sia polarizzato e caratterizza una classe di infiniti stimatori possibili

08/05/2014

29

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

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

dati scorrelati

Usando il metodo dei moltiplicatori di Lagrange

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

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

08/05/2014

30

- Imponendo il vincolo di non polarizzazione (\*)

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

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

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

08/05/2014

31

- Calcoliamo ora la varianza dello stimatore:

$$\begin{aligned} \text{var}(\hat{\vartheta}_N) &= E\left\{[\hat{\vartheta}_N - E(\hat{\vartheta}_N)]^2\right\} = E\left\{\left[\sum_{i=1}^N \alpha(i) d(i) - \vartheta^0 \sum_{i=1}^N \alpha(i)\right]^2\right\} \\ &= E\left\{\left[\sum_{i=1}^N \alpha(i) [d(i) - \vartheta^0]\right]^2\right\} = \sum_{i=1}^N \alpha(i)^2 E\{[d(i) - \vartheta^0]^2\} \\ &= \sum_{i=1}^N \alpha(i)^2 \text{var}[d(i)] = \alpha^2 \sum_{i=1}^N \frac{1}{\text{var}[d(i)]} = \frac{1}{\sum_{i=1}^N \frac{1}{\text{var}[d(i)]}} \end{aligned}$$

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

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

08/05/2014

32

## Generalizzazione

- Nel caso in cui le quantità da stimare siano **varianti nel tempo** e` necessario modificare gli indici di bontà degli stimatori
- Si indichi con  $\hat{\vartheta}(t|t-1)$  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`

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

dove  $c$  e` una costante ragionevolmente piccola

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

08/05/2014

33

## Stima ai Minimi Quadrati (LSM)

08/05/2014

34

## Regressione lineare

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

$$y(t) = \vartheta_1 u_1(t) + \dots + \vartheta_q u_q(t), \quad t = 1, \dots, N \quad (*)$$

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

08/05/2014

35

- Il problema puo` essere formulato in maniera vettoriale

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

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

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

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

08/05/2014

36

- Obiettivo del problema delle regressione lineare sara` quindi quello di minimizzare l'errore  $\varepsilon(t)$  determinando un vettore  $\vartheta^o$  per cui questo minimo venga raggiunto
- Si definisce la **funzione di costo quadratica**

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

↳  $\vartheta^o = \arg \min_{\vartheta} J(\vartheta)$  **Stimatore ai minimi quadrati**

08/05/2014

37

- Indicando con  $\vartheta_i$  la componente i-esima del vettore  $\vartheta$

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

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

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

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

08/05/2014

38

- Imponendo  $\frac{\partial J}{\partial \vartheta} = [0 \ 0 \ \dots \ 0]$

↳ 
$$-2 \sum_{t=1}^N [y(t) - \varphi(t)^\top \vartheta] \varphi(t)^\top = [0 \ 0 \ \dots \ 0]$$

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

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

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

**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**

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

08/05/2014

39

## Interpretazione geometrica delle eq. normali

Poniamo

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

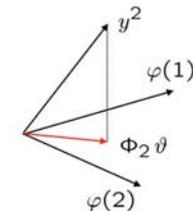
Scriviamo quindi

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

Evidentemente  $\|y^N - \Phi_N \vartheta\|$  e`

minima quando  $y^N - \Phi_N \vartheta$  e`

ortogonale a  $\Phi_N \vartheta$



08/05/2014

40

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

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

si ha

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

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

evidentemente si tratta di una matrice simmetrica e semidefinita positiva

$\rightarrow \hat{\vartheta}_N$  e' un **minimo locale** di  $J(\vartheta)$

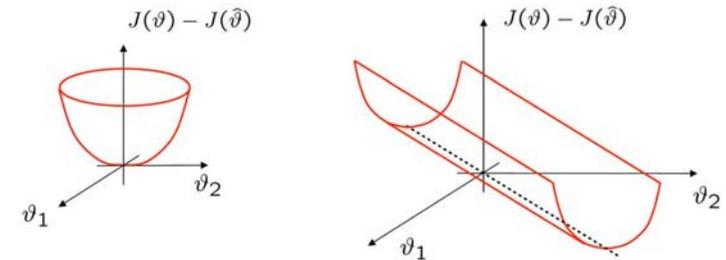
08/05/2014

41

- Considerando quindi la forma quadratica

$$J(\vartheta) - J(\hat{\vartheta}) = \frac{1}{2} (\vartheta - \hat{\vartheta})^\top \frac{d^2 J}{d\vartheta^2} \Big|_{\hat{\vartheta}} (\vartheta - \hat{\vartheta})$$

si hanno i due casi possibili:



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

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

08/05/2014

42

- Quindi:

○ Se  $\det \left[ \sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] \neq 0 \rightarrow \hat{\vartheta}_N$  **unico minimo globale**

○ Se  $\det \left[ \sum_{t=1}^N \varphi(t) \varphi(t)^\top \right] = 0 \rightarrow \hat{\vartheta}_N$  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'**

08/05/2014

43

### 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

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

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

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

08/05/2014

44

## Polarizzazione

$$\begin{aligned}\hat{\vartheta}_N &= \left[ \sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) y(t) \\ &= \left[ \sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) [\varphi(t)^\top \vartheta^0 + \xi(t)] \\ &= \vartheta^0 + \left[ \sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) \xi(t)\end{aligned}$$

$$\rightarrow \hat{\vartheta}_N - \vartheta^0 = \left[ \sum_{t=1}^N \varphi(t) \varphi(t)^\top \right]^{-1} \sum_{t=1}^N \varphi(t) \xi(t)$$

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

$$\rightarrow E(\hat{\vartheta}_N) = \vartheta^0 \quad \text{Stimatore MQ e' non polarizzato}$$

08/05/2014

45

## 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

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

08/05/2014

46

## Varianza

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

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

$$\begin{aligned}\rightarrow \text{var}(\hat{\vartheta}_N) &= E \left[ (\hat{\vartheta}_N - \vartheta^0) (\hat{\vartheta}_N - \vartheta^0)^\top \right] \\ &= E \left\{ \left[ S(N)^{-1} \sum_{t=1}^N \varphi(t) \xi(t) \right] \left[ S(N)^{-1} \sum_{s=1}^N \varphi(s) \xi(s) \right]^\top \right\} \\ &= E \left\{ \left[ S(N)^{-1} \sum_{t=1}^N \varphi(t) \xi(t) \right] \left[ \sum_{s=1}^N \xi(s) \varphi(s)^\top S(N)^{-1} \right] \right\} \\ &= S(N)^{-1} E \left[ \sum_{t=1}^N \varphi(t) \xi(t) \sum_{s=1}^N \xi(s) \varphi(s)^\top \right] S(N)^{-1}\end{aligned}$$

08/05/2014

47

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

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

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

08/05/2014

48

## Interpretazione

Supponiamo che  $\vartheta^\circ$  sia scalare per cui anche  $\varphi(t)$  e  $\xi(t)$  e scalare

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

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

08/05/2014

49

Inoltre

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

Pertanto:

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

08/05/2014

50

- $\frac{\lambda^2}{\frac{1}{N} \sum_{t=1}^N u(t)^2}$  e' una sorta di rapporto rumore/segnale

- Se la varianza di  $u$  e' limitata  $\rightarrow \lim_{N \rightarrow \infty} \text{var}(\hat{\vartheta}_N) = 0$   
e siccome lo stimatore non e' polarizzato si ha:

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

ovvero lo stimatore MQ converge in media quadratica

08/05/2014

51

## 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

08/05/2014

52