

TEST DI BIANCHEZZA

Contenuti:

- Motivazione
- Test sui cambi di segno
- Test Portmanteau
- Conclusioni

MOTIVAZIONE

Fatto: Dato un P.C. stazionario $y(t)$, la sequenza degli errori di predizione commessi dal predittore ottimo ad un passo $\hat{y}(t+1/t)$ è un rumore bianco.

Idea: Per validare un modello identificato controllo la sequenza degli errori di predizione: se non è un *WN* potrei aver sbagliato l'ordine o addirittura il modello.

Problema: Come faccio a capire se un P.C. è un rumore bianco?



Test di bianchezza

TEST SUI CAMBI DI SEGNO

Ipotesi:

- $w(t) \sim WN$
- $P(w(t) < 0) = P(w(t) > 0) = 0.5$

Nota: Non faccio ipotesi sulla d.d.p. di $w(t)$.

$$x_N = \sum_{i=1}^{N-1} s_i$$

$$s_i = \begin{cases} 1 & \text{se } w(i)w(i+1) \leq 0 \\ 0 & \text{se } w(i)w(i+1) > 0 \end{cases}$$

$x_N = n^\circ$ di cambi di segno nella sequenza $w(1), w(2), \dots, w(N)$

In media, mi aspetto che $w(t)$ cambi di segno una volta su due (se cambiasse di segno ad ogni passo o, all'opposto, non cambiasse mai, $w(t)$ non sarebbe più indep. da $w(t-1)$).

Problema: come capire se $w(t)$ cambia di segno troppe o troppo poche volte?

Teorema: Per $N \rightarrow \infty$, la V.C. x_N è distribuita gaussianamente con $E[x_N] = N/2$, $Var[x_N] = N/4$.

Conseguenza: La V.C.

$$\frac{x_N - \frac{N}{2}}{\frac{\sqrt{N}}{2}}$$

ha la stessa d.d.p. della V.C. gaussiana standard Z (tale che $E[Z] = 0$, $Var[Z] = 1$).

Test dei cambi di segno: Fissato il livello di significatività α (tipicamente $\alpha = 0.05$) cerco sulla tabella della V.C. gaussiana standard Z un valore z_α tale che $P(|Z| \leq z_\alpha) = 0.95$. Poi adotto la regola:

$$\bullet \left| \frac{x_N - \frac{N}{2}}{\frac{\sqrt{N}}{2}} \right| \leq z_\alpha \Rightarrow w(t) \text{ è "bianco"}$$

$$\bullet \left| \frac{x_N - \frac{N}{2}}{\frac{\sqrt{N}}{2}} \right| > z_\alpha \Rightarrow w(t) \text{ non è "bianco"}$$

Esempio: Se $\alpha = 0.05$, risulta $z_\alpha = 1.96$. Allora la sequenza $w(t)$ passa il test di bianchezza con livello di significatività del 5% se

$$\left| \frac{x_N - \frac{N}{2}}{\frac{\sqrt{N}}{2}} \right| \leq 1.96$$

Osservazioni:

- Con $\alpha = 0.05$, anche se $w(t)$ è un WN , nel 5% dei casi non passa il test di bianchezza.
- Viceversa, se $w(t)$ non è un WN può ciononostante passare il test (basta prendere un segnale periodico che cambia di segno $N/2$ volte).

TEST "PORTMANTEAU"

Ipotesi:

- $w(t) \sim WN$
- $E[w(t)] = 0$

Idea: Considero $\hat{r}(\tau) = \frac{\hat{R}(\tau)}{\hat{R}(0)}$, dove

$$\hat{R}(\tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} w(t+\tau)w(t)$$

Sotto le ipotesi fatte $\hat{r}(\tau)$ è uno stimatore di $\rho(\tau)$. Poiché per un WN , $\rho(\tau) = 0$, $\tau \neq 0$, mi aspetto che $\hat{r}(\tau) \cong 0$, $\tau \neq 0$.

Come dare rigore statistico a questa considerazione qualitativa?

Teorema: Se $w(t)$ è un WN, per $N \rightarrow \infty$

- $\sqrt{N} \hat{r}(\tau) \rightarrow G(0,1)$

(la f.d.d. di $\sqrt{N} \hat{r}(\tau)$ converge ad un gaussiana standard)

- $N \sum_{\tau=1}^m \hat{r}(\tau)^2 \rightarrow \chi^2(m)$

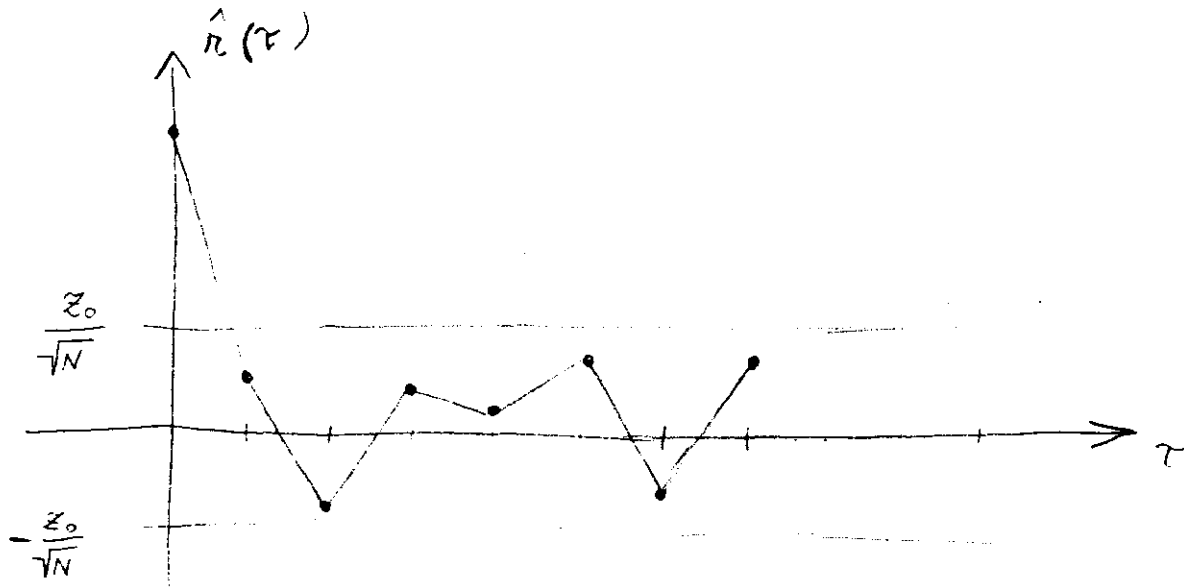
Test di bianchezza: Fissato il livello di significatività α (tipicamente, $\alpha = 0.05$) cerco sulle tabelle x_α tale che $P(\chi^2(m) < x_\alpha) = 0.95$. Poi adotto la seguente regola:

- $N \sum_{\tau=1}^m \hat{r}(\tau)^2 \leq x_\alpha \Rightarrow w(t)$ è "bianco"

- $N \sum_{\tau=1}^m \hat{r}(\tau)^2 > x_\alpha \Rightarrow w(t)$ non è "bianco"

(valori tipici di m : da 5 fino a $N/4$)

Rappresentazione grafica:



Nota: In base alla prima parte del teorema, se $w(t)$ è un WN risulta $P(\sqrt{N} \hat{r}(\tau) < 1.96) = 0.95 \Rightarrow$ mi aspetto che $\hat{r}(\tau)$ esca dalla "banda" una volta ogni 20 campioni (test di Anderson).

CONCLUSIONI

- Un modello sottoparametrizzato non riesce a spiegare bene i dati \Rightarrow i residui (= errori di predizione ad un passo) non sono bianchi
- Aumentando l'ordine del modello i residui diventano bianchi
- Strategia "semplice" per la scelta dell'ordine del modello: usare il modello più semplice che produce dei residui bianchi.
- I test di bianchezza sono "soggettivi" (devo scegliere il livello di significatività α). Come in tutti i test c'è una certa probabilità di respingere l'ipotesi vera e di accettare l'ipotesi sbagliata. Il compromesso tra i due tipi di errore dipende da α .
- Alternativa: analisi spettrale dei residui.

MINIMI QUADRATI RICORSIVI

Contenuti:

- Motivazione
- Un esempio semplice
- Algoritmo RLS
- Identificazione adattativa

MOTIVAZIONE

Problema: Identificazione in tempo reale. Mano a mano che arrivano i dati bisogna aggiornare il modello stimato, cioè $\hat{\theta}$. Se memorizzo tutti i dati e ricalcolo ad ogni passo la stima LS , l'occupazione di memoria e i calcoli diventano esplosivi.

Applicazioni:

- Stima spettrale in tempo reale
- Controllo adattativo
- Fault detection
- ...

Idea: Calcolare $\hat{\theta}(t)$ e $S(t)$ a partire da $y(t)$, $\hat{\theta}(t-1)$, $S(t-1)$ senza dover tenere in memoria $y(t-1)$, $y(t-2)$, ...

UN ESEMPIO SEMPLICE

Stima ricorsiva di una costante

$$y(k) = \theta + e(k), \quad e(k) \sim WN(0, \sigma^2), \quad 1 \leq k \leq t$$

Stimatore LS = media campionaria:

$$\begin{aligned}\hat{\theta}(t) &= \frac{1}{t} \sum_{i=1}^t y(i) \\ &= \frac{1}{t} [(t-1) \hat{\theta}(t-1) + y(t)] \\ &= \frac{1}{t} [t \hat{\theta}(t-1) + y(t) - \hat{\theta}(t-1)] \\ &= \hat{\theta}(t-1) + \frac{1}{t} [y(t) - \hat{\theta}(t-1)]\end{aligned}$$

L'informazione passata, condensata in $\hat{\theta}(t-1)$, viene corretta in base alla differenza tra osservazione presente $y(t)$ e valore "predetto" $\hat{\theta}(t-1)$

Nota: Dato che l'informazione passata diventa sempre più affidabile, il peso da dare all'errore decresce ($1/t$)

ALGORITMO RLS

Algoritmo LS "a lotti" (batch)

$$\hat{\theta}(t) = S(t)^{-1} \sum_{i=1}^t \phi(i)y(i)$$

$$S(t) = \sum_{i=1}^t \phi(i)\phi(i)'$$

Algoritmo RLS (Recursive Least Squares)

$$\beta(t-1) = I + \phi(t)'P(t-1)\phi(t)$$

$$K(t) = \beta(t-1)^{-1}P(t-1)\phi(t)$$

$$\varepsilon(t) = y(t) - \phi(t)\hat{\theta}(t-1)$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t) \varepsilon(t)$$

$$P(t) = P(t-1) - \beta(t-1)^{-1}P(t-1)\phi(t)\phi(t)'P(t-1)$$

Osservazioni:

- Non c'è bisogno di invertire matrici ($\beta(t)$ è scalare)
- $K(t)$: "guadagno" dell'algoritmo
- $P(t) = S(t)^{-1} = \sigma^2 \text{Var}[\hat{\theta}(t)]$. Se $\phi(t)$ è persistentemente eccitante, $S(t) \rightarrow \infty \Rightarrow P(t) \rightarrow 0 \Rightarrow K(t) \rightarrow 0$.
- In pratica, per migliorare la stabilità numerica si usano algoritmi "in forma fattorizzata". Per esempio, scompongo $P(t)$ come $P(t) = V(t)V(t)'$ e, invece di calcolare $P(t)$ a partire da $P(t-1)$, calcolo il "fattore" $V(t)$ a partire da $V(t-1)$.

- *Inizializzazione rigorosa.*

Essendo t_0 tale che $\det(S(t_0)) \neq 0$, si ponga

$$P(t_0) = \left(\sum_{i=1}^{t_0} \phi(i)\phi(i)' \right)^{-1}$$

$$\hat{\theta}(t_0) = P(t_0) \sum_{i=1}^{t_0} \phi(i)y(i)$$

- *Inizializzazione convenzionale (più comoda)*

$$\hat{\theta}(0) = 0$$

$$P(0) = \alpha I \quad (\Leftrightarrow \text{Var}[\hat{\theta}(0)] \propto \alpha I)$$

IDENTIFICAZIONE ADATTATIVA

Problema: Processi da identificare la cui dinamica è lentamente tempo variante (esempio: usura, cambiamento del punto di lavoro, ...)

Idea: Descrivo il processo con un modello lineare tempo invariante il cui vettore dei parametri è tempo-variante: $\theta = \theta(t)$.

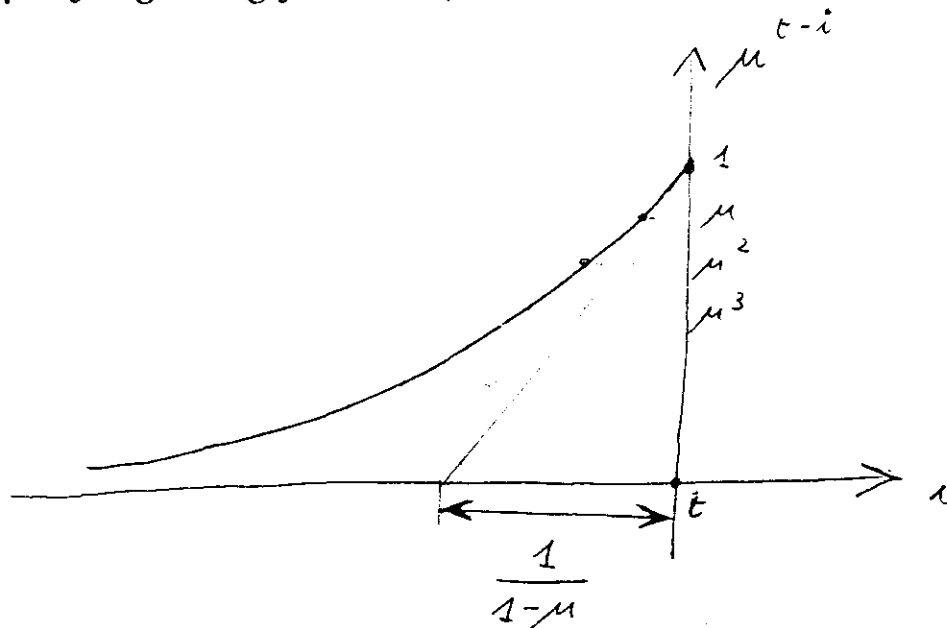
Soluzione sbagliata: Usare *RLS*. Se i dati sono persistentemente eccitanti $K(t) \rightarrow 0 \Rightarrow$ l'algoritmo diventa insensibile ai nuovi dati e si fossilizza sulla sua stima (anche se nel frattempo il processo reale è cambiato).

Ci vuole un algoritmo ricorsivo che continui a rimanere "reattivo"

Nuova cifra di merito:

$$J_t(\theta) = \frac{1}{t} \sum_{i=1}^t \mu^{t-i} \varepsilon(i)^2, \quad 0 < \mu \leq 1$$

μ : "forgetting factor" (coefficiente di oblio)



Interpretazione: Viene dato peso via via decrescente alla capacità di spiegare i dati più lontani nel tempo. I dati passati vengono "dimenticati" esponenzialmente.

La stima di θ al tempo t dipende essenzialmente dai dati $y(t), y(t-1), \dots, y(t-T)$, dove $T = 1/(1-\mu)$.

Algoritmo RLS con oblio esponenziale

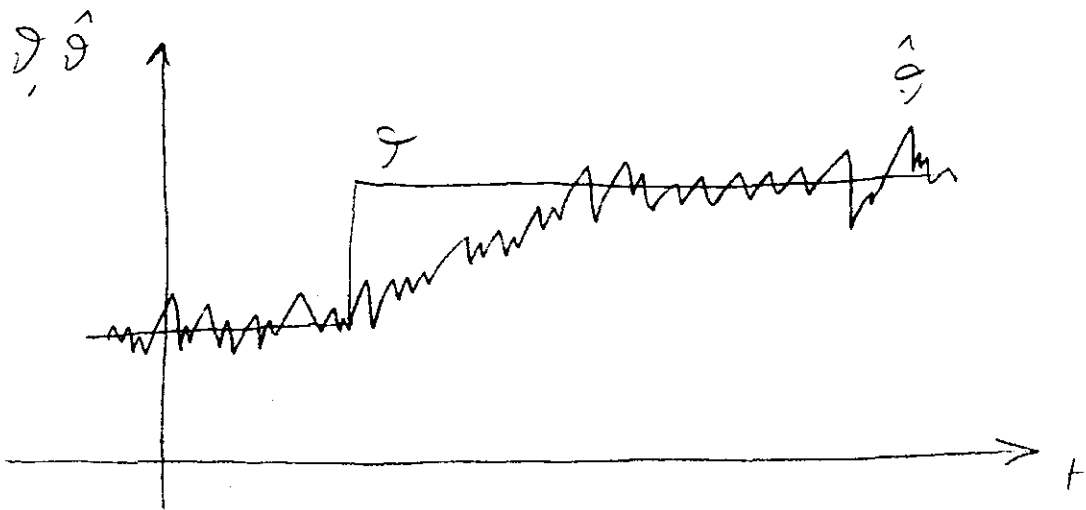
$$\beta(t-1) = \mu + \phi(t)'P(t-1)\phi(t)$$

$$K(t) = \beta(t-1)^{-1}P(t-1)\phi(t)$$

$$\varepsilon(t) = y(t) - \phi(t)\hat{\theta}(t-1)$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + K(t) \varepsilon(t)$$

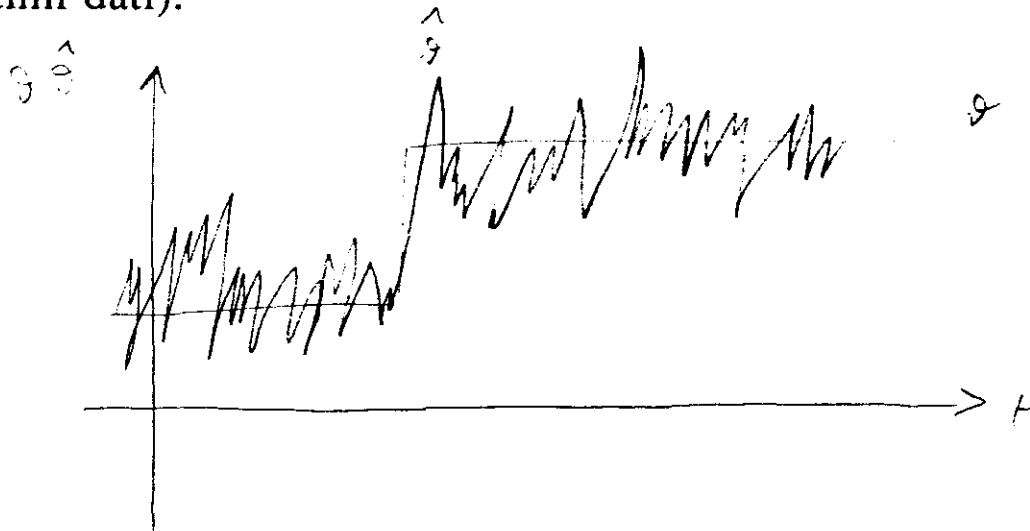
$$P(t) = \mu^{-1} [P(t-1) - \beta(t-1)^{-1}P(t-1)\phi(t)\phi(t)'P(t-1)]$$



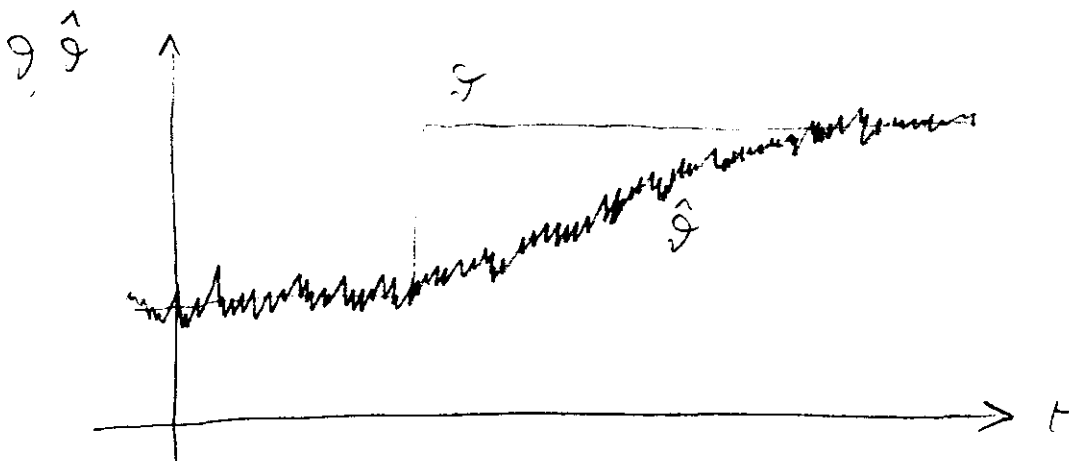
Osservazioni:

- $\mu = 1 \Rightarrow$ RLS
- $\mu < 1 \Rightarrow K(t) \not\rightarrow 0$ (purché $\phi(t)$ sia limitato) \Rightarrow l'identificatore rimane sempre "reattivo".

- μ "piccolo" \Rightarrow stime "nervose" (la stima è basata solo sugli ultimi dati).



- μ "grande" \Rightarrow scarsa reattività (se θ varia bruscamente, l' algoritmo impiega molto tempo a inseguire).

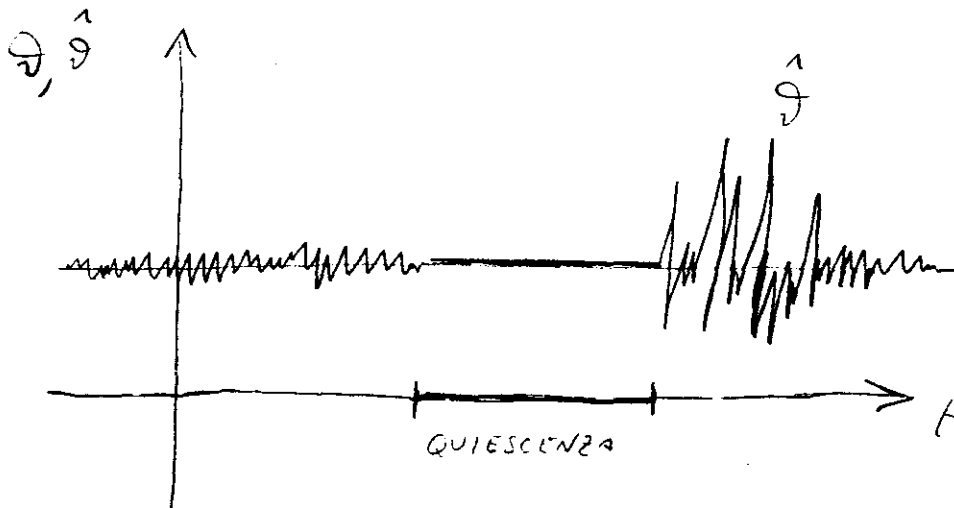


- Inizializzazione e algoritmi fattorizzati: come *RLS*.

Il fenomeno del "blow-up"

Fatto: Supponiamo che $\phi(t) = 0$ per un lungo periodo; allora $P(t) \rightarrow \infty$ e $K(t) = 0$.

Conseguenza: Quando $\phi(t)$ torna ad essere $\neq 0$, $K(t)$ diventa molto grande \Rightarrow oscillazioni violente nella stima dei parametri:



Rimedi:

- $\mu = \mu(t)$ in maniera "furba"
- Impedire che $P(t)$ cresca troppo
- ...

CONCLUSIONI

- E' possibile seguire in tempo reale le variazioni di comportamento di un processo fisico.
- Possibile uso di RLS per adattare gli algoritmi di controllo alle variazioni della dinamica del processo.
- Cambiamenti significativi nei valori dei parametri possono essere sintomi di guasti \Rightarrow diagnostica.

MODELLI NARX

- In molti casi un modello lineare non è sufficiente. Per descrivere il comportamento del sistema è necessario utilizzare modelli non lineari nelle variabili di ingresso
- Si possono determinare modelli di tipo polinomiale, reti di percettroni, reti neurali a base radiale, modelli fuzzy...
- L'applicazione al problema della modellistica del motore ha fornito ottimi risultati

In precedenza sono stati introdotte e analizzate alcune famiglie di modelli lineari (nelle misure e nei parametri), tra cui ad esempio i modelli ARX

$$y(t) = -a_1y(t-1) - a_2y(t-2) - \dots - a_ny(t-n) + b_1u(t-1) + b_2u(t-2) + \dots + b_nu(t-n) + e(t)$$

Una naturale estensione consiste nel considerare modelli in cui l'uscita $y(t)$ dipende in modo non lineare dai passati valori $y(t-i)$, $u(t-j)$

$$y(t) = f(y(t-i), u(t-j))$$

Il problema è però quello di **determinare la forma della funzione** e quindi **stimare i suoi parametri incogniti**

La scelta preliminare da fare è quella relativa alla struttura della funzione f . In base alla teoria sin qui sviluppata è possibile seguire tre strade

- modelli polinomiali (lineari nei parametri)
- reti neurali a perceptron (non lineari nei parametri)
- reti neurali a base radiale (lineari nei parametri purché i centri siano stati fissati)

Per la loro relativa semplicità, e per il fatto che le considerazioni fatte in precedenza per le reti neurali si possono applicare anche a questo caso, nel seguito si analizzeranno più in dettaglio soltanto i **modelli polinomiali**

MODELLI NARX POLINOMIALI

Supponiamo che y all'istante t dipenda da se stessa e dall'ingresso u all'istante precedente $t-1$ e che la funzione f sia composta da monomi al più di grado 2 (che chiameremo "grado di non linearità"). Il modello più generale, non lineare nelle osservazioni e lineare nei parametri ϑ_i , è

$$y(t) = \vartheta_0 + \vartheta_1 y(t-1) + \vartheta_2 y(t-1)^2 + \vartheta_3 u(t-1) + \vartheta_4 u(t-1)^2 + \vartheta_5 u(t-1)y(t-1) + e(t)$$

La stima dei parametri è semplice e può essere effettuata con i Minimi Quadrati lineari

E' però evidente che se si hanno più ingressi, se si considera un polinomio con grado di non linearità elevato e se si assumono ritardi superiori a 1, **il numero di parametri da stimare tende ad esplodere**. E' quindi necessario avere dei criteri per selezionare una struttura adeguata e "parsimoniosa"

A questo proposito si possono ancora utilizzare tutti i criteri visti in precedenza: criteri soggettivi, criteri oggettivi, **stepwise regression**

Quest'ultima, con qualche variante, è stata impiegata con successo nella determinazione di un modello dinamico della pressione nel collettore e della dinamica rotazionale

STEPWISE REGRESSION PER MODELLI NARX

- 0) Si considera il modello $y(t) = \sum_{i=0}^r g_i x_i(t)$ dove $r+1$ è il numero di potenziali regressori e x_i è il potenziale regressore. Si dividono i dati in **un insieme di identificazione e uno di validazione**

- 1) Si identificano con i Minimi Quadrati lineari (usando i dati di identificazione) $r+1$ modelli, ciascuno con un solo regressore. Si seleziona quello che minimizza la somma dei quadrati dei residui (SSR) **sui dati di validazione**. Questo corrisponde ad impiegare un modello a errore d'equazione per la stima dei parametri (per avere una soluzione in forma chiusa) e un criterio basato su un modello a errore d'uscita per la selezione del modello (in quanto interessa la capacità di simulare il sistema, più che la predizione ad un passo)

- 2) Si estende temporaneamente il modello selezionato aggiungendo uno alla volta i regressori rimanenti. Se esiste almeno un regressore che riduce il valore di SSR (sempre sui dati di validazione) rispetto al valore del passo precedente, si inserisce il regressore che fornisce il valore minimo di SSR e si ripete questo punto finché il valore di SSR diminuisce. Altrimenti la procedura termina

Naturalmente questo è un procedimento computazionalmente oneroso, ma consente di determinare un modello che può essere considerato un buon simulatore del sistema

MODELLI NARX DEI MOTORI A COMBUSTIONE INTERNA

I NARX sono stati utilizzati per la stima di modelli della pressione nel collettore e del numero di giri. Si è seguita questa via per varie ragioni

- Un modello a scatola grigia della pressione richiede una vasta campagna di prove per stimare il rendimento volumetrico e la portata aria. Con l'uso di modelli a scatola nera di tipo NARX si perde il significato fisico, ma si ha il vantaggio di ottenere un modello dai dati misurati direttamente dalla centralina montata su vettura
- Come già detto, il tentativo di determinare un modello "fisico" del numero di giri è estremamente arduo e costoso
- L'impiego di dati misurati su vettura consente di tener conto della strumentazione e dell'elettronica effettivamente utilizzate sulle vetture commerciali

I DATI DI IDENTIFICAZIONE E VALIDAZIONE

I dati impiegati per la stima sono stati preliminarmente normalizzati

$$n \rightarrow n/1000$$

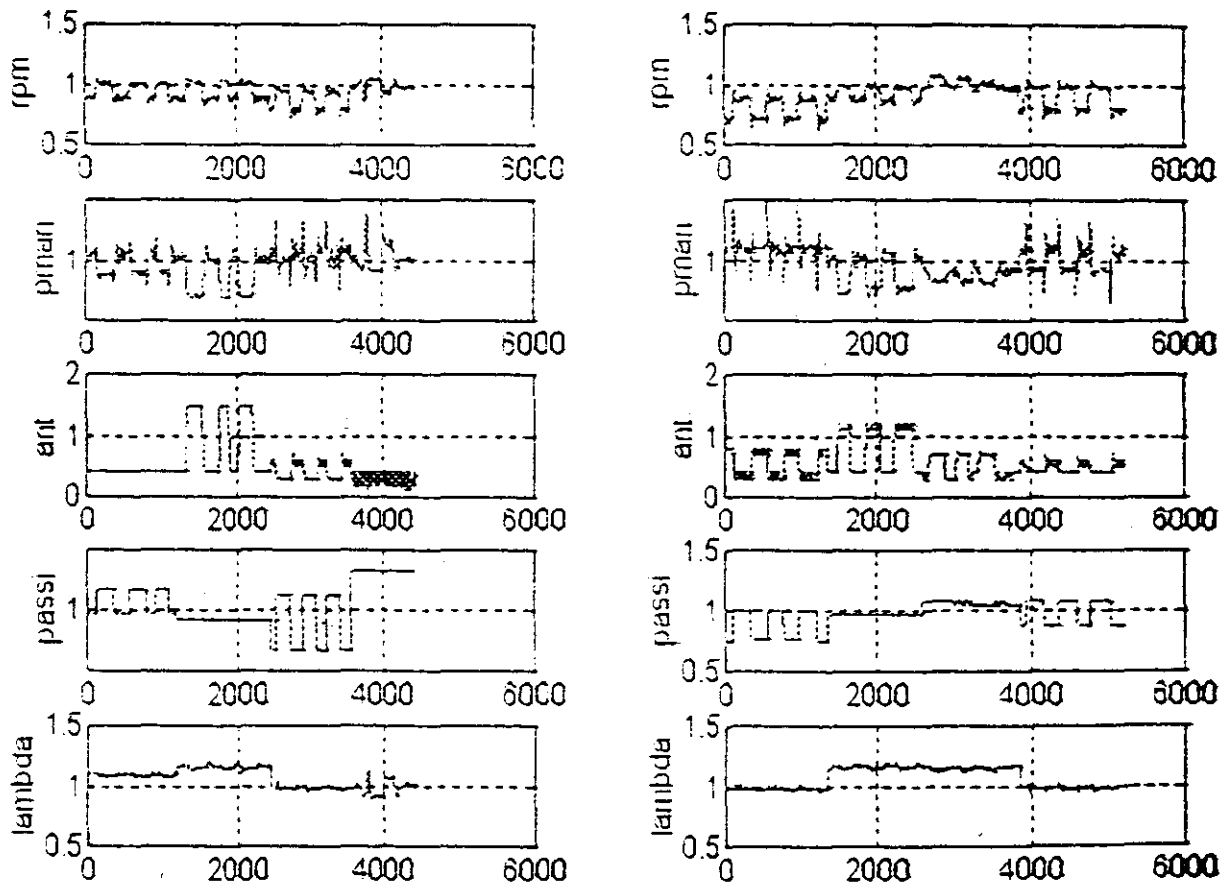
$$p \rightarrow p/300$$

$$\text{passi} \rightarrow \text{passi}/60$$

$$\theta \rightarrow \theta/10$$

$$\lambda \rightarrow \lambda/10$$

successivamente si è passati dal dominio del tempo a quello dell'angolo motore, visto che le strategie di controllo normalmente si basano su questo ultimo



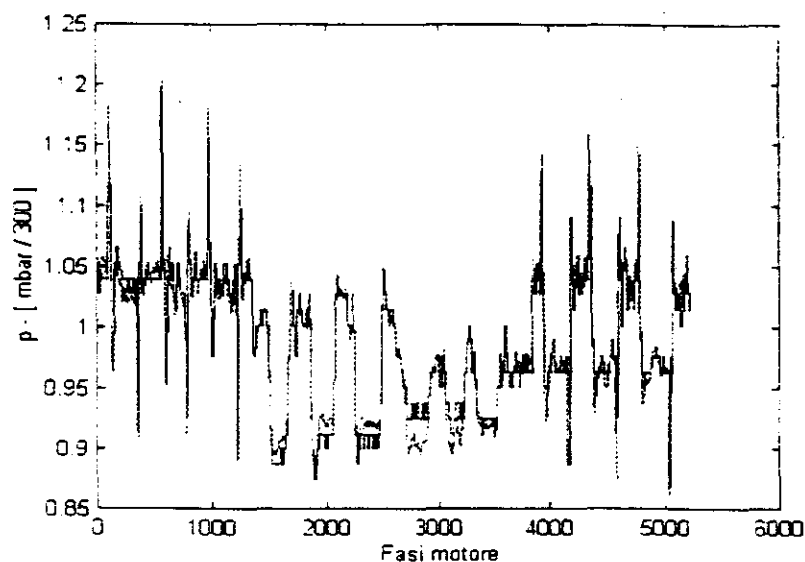
MODELLO DELLA PRESSIONE

In questo caso è facile determinare i potenziali regressori da considerare, in quanto dagli studi precedenti è nota la forma della dipendenza del rendimento volumetrico da n , p , così come è facilmente individuabile la dipendenza (essenzialmente di tipo lineare o quadratico) della portata d'aria dal numero di passi di apertura dello stepper motor. E' quindi sufficiente considerare circa 15 regressori potenziali

Il procedimento illustrato ha portato alla stima del seguente modello

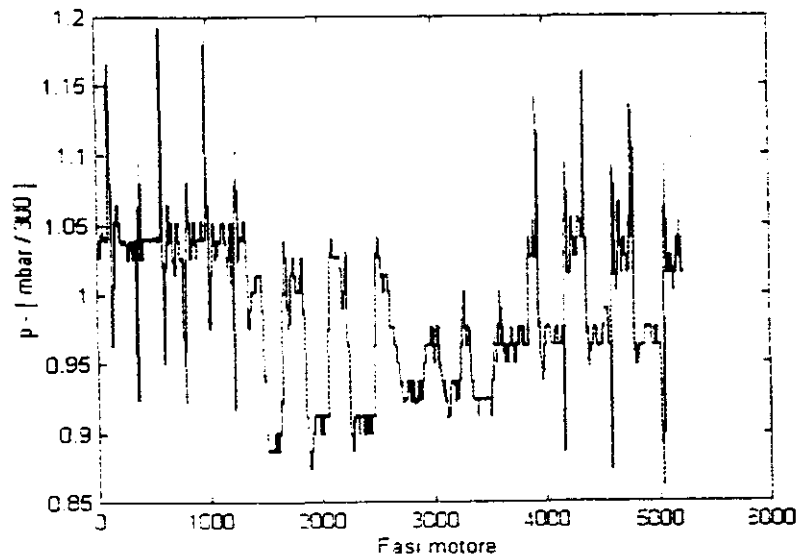
$$\hat{p}(k) = 0.0743 \text{passi}(k-1)n(k-1)^{-1} + 0.031 \ln(k-1)^{-1} + \\ - 0.8874 \hat{p}(k-1) - 0.0032 \hat{p}^2(k-1)$$

L'andamento reale e stimato sui dati di validazione è il seguente

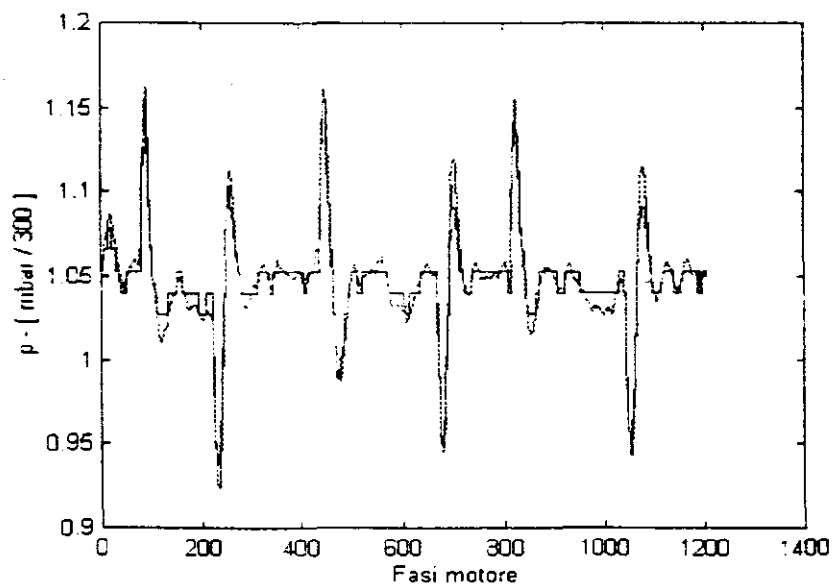


Si noti che nella simulazione si utilizzano i passati valori predetti della pressione (criterio "a errore d'uscita") e i valori misurati di n

Se invece si utilizzano anche i passati valori misurati della pressione (criterio "a errore d'equazione") i transitori sono in pratica coincidenti



Se il modello viene utilizzato su dati diversi da quelli di identificazione e di validazione e secondo un criterio a errore d'uscita si ottengono i transitori seguenti



MODELLO DEL NUMERO DI GIRI

- Il modello identificato della pressione ha la forma

$$p = f_p(p, n, passi)$$

Con questo modello si può determinare la pressione stimata \hat{p}

- Nella fase di identificazione del modello di n

$$n = f_n(p, n, passi, \theta, \lambda)$$

si utilizza la stima \hat{p} anziché la pressione misurata. Inoltre in fase di validazione l'errore viene calcolato a partire da

$$\hat{n} = f_n(f_p(\hat{p}, \hat{n}, passi), \hat{n}, passi, \theta, \lambda)$$

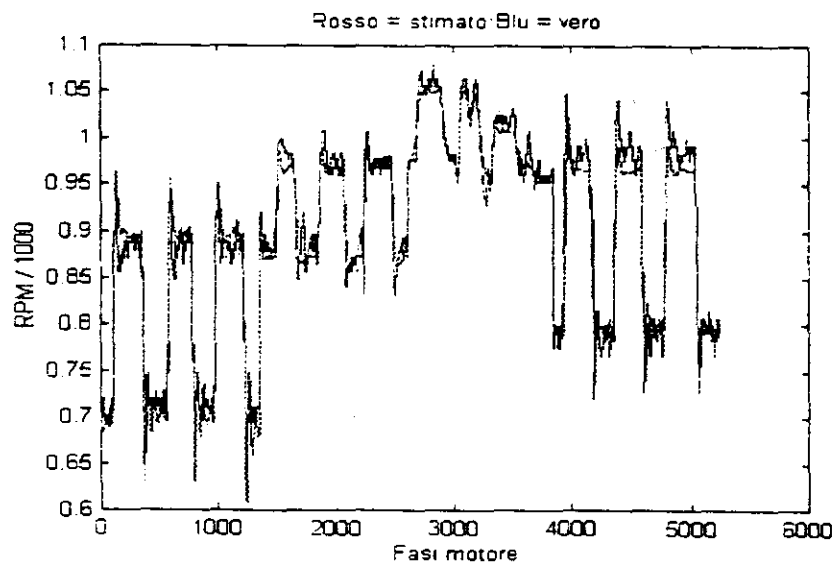
Questo è di fatto l'errore che si commette utilizzando come **simulatore in anello aperto** il modello complessivo identificato, composto dai sottomodelli di p e n .

- Non potendo basarsi su considerazioni fisiche significative, è necessario considerare un elevato numero di potenziali regressori (circa 250) e la stima richiede tempo. I risultati ottenuti possono consentire nel futuro di restringere significativamente l'insieme dei regressori candidati

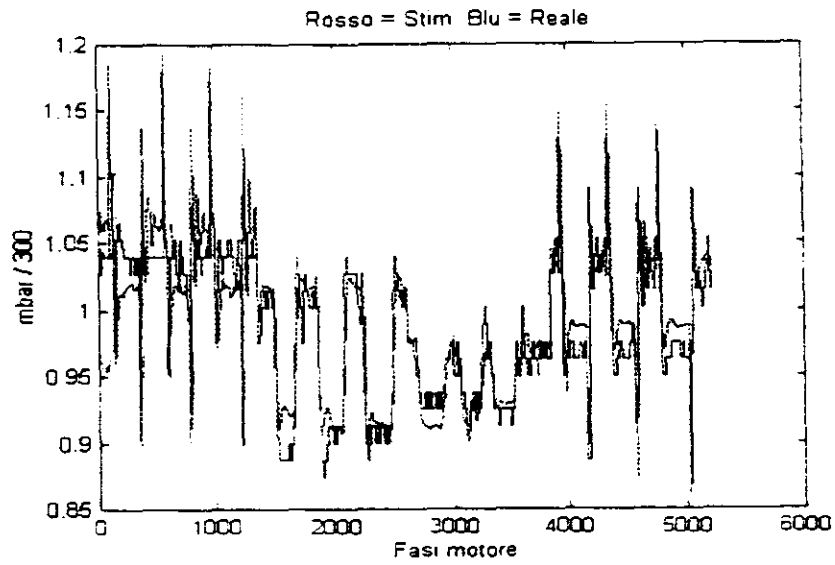
Il modello stimato è il seguente

$$\begin{aligned}
 \hat{n}(k) = & -0.0797 p(k-1) passi(k-1) \lambda(k-1) + \\
 & -0.0268 \hat{n}^{-2}(k-1) p^2(k-1) passi^2(k-1) \lambda(k-1) \vartheta(k-1) - \\
 & -0.9879 \hat{n}(k-1) - 0.1007 \hat{n}^2(k-1) p^2(k-1) \\
 & - 0.0171 \lambda^{-1}(k-1) \vartheta(k-1) - 0.0167 \hat{n}^2(k-1) \lambda(k-1) \\
 & - 0.2462 \hat{n}(k-1) p^2(k-1) - 0.0989 - \\
 & - 0.0852 passi(k-1) \lambda(k-1) - \\
 & - 0.0443 \hat{n}^{-1}(k-1) p(k-1) passi(k-1) \vartheta(k-1) \\
 & - 0.0081 \lambda^2(k-1) \vartheta(k-1) \\
 & - 0.0137 \hat{n}^{-1}(k-1) p^2(k-1) passi^2(k-1) \lambda^{-1}(k-1) \\
 & - 0.0043 \hat{n}^2(k-1) \lambda^2(k-1) \vartheta^2(k-1) \\
 & - 0.0053 p^2(k-1) passi^2(k-1) \vartheta^2(k-1) \\
 & - 0.416 \hat{n}(k-1) p(k-1) \lambda(k-1) \\
 & - 0.0133 \hat{n}^{-1}(k-1) p^2(k-1) passi(k-1) \lambda(k-1) \vartheta(k-1)
 \end{aligned}$$

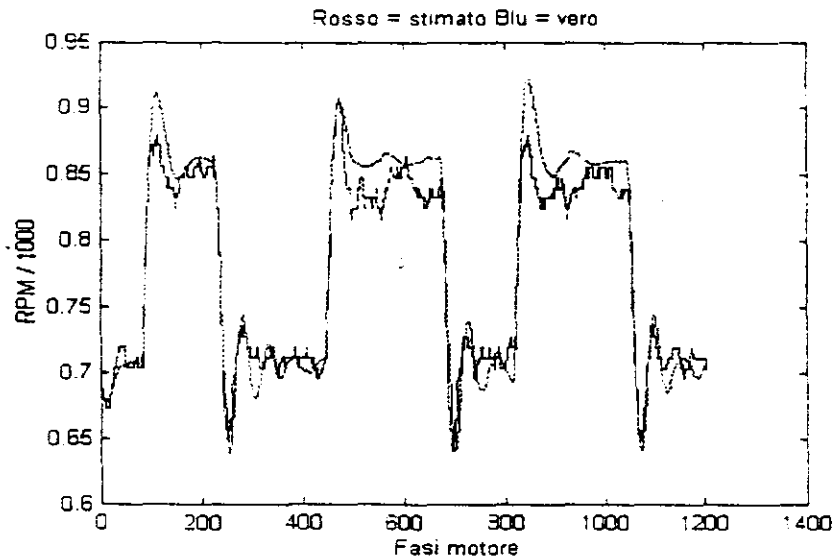
Simulando congiuntamente il modello della pressione e quello del numero dei giri con i dati di validazione e confrontando l'andamento del numero di giri stimato con quello reale si ottengono i seguenti transitori



Nelle stesse condizioni, l'andamento reale e simulato della pressione è



Infine, sempre nelle stesse condizioni, il numero di giri relativo a un diverso insieme di dati è



CONCLUSIONI

- I modelli NARX (polinomiali e non) consentono una grande flessibilità e la possibilità di determinare modelli anche di fenomeni dinamici molto complessi (numero di giri)
- Si possono estendere molte considerazioni fatte per i modelli lineari relative alla scelta della struttura, all'identificazione, alle problematiche connesse con la scelta del segnale di ingresso

ESTENSIONI E BIBLIOGRAFIA

Possibili estensioni e approfondimenti:

- Modelli NARX neurali
- Identificazione non parametrica (splines/Bayes)
- Filtro di Kalman
- "Subspace Identification Methods"

Bibliografia:

S. Bittanti, "Teoria della predizione e del filtraggio", Pitagora, Bologna, 1992.

S. Bittanti, "Identificazione dei modelli e controllo adattativo", Pitagora, Bologna, 1992

L. Ljung, "System Identification - Theory for the User", Prentice-Hall, 1987

T. Söderstrom, P. Stoica, "System Identification", Prentice-Hall, 1989.

S. Haykin, "Neural Networks - A comprehensive foundation", Macmillan, 1994