Quindi: $\hat{\mathcal{M}}(\vartheta)$: $C(z)\hat{y}(t) = [C(z) - A(z)]y(t) + B(z)u(t-1)$. Qui la stabilità dipende da $C(z)$.

3) MODELLO AR: $\mathcal{M}(\vartheta)$: $y(t) = a_1 y(t-1) + \dots + a_{na} y(t-na)$.

Il corrispondente predittore è:

$$\hat{\mathcal{M}}(\vartheta): \hat{y}(t) = [1 - A(z)]y(t)$$

4) MODELLO MA: $\mathcal{M}(\vartheta)$: $y(t) = \xi(t) + c_1\xi(t-1) + \dots + c_{nc}\xi(t-nc)$

Predittore: $\hat{\mathcal{M}}(\vartheta): \hat{y}(t)C(z) = [C(z)-1]y(t)$

Riconsideriamo ora la cifra di merito $J_N = (1/N)\sum_{1}^{N} \varepsilon(t)^2$. Indichiamo con $\hat{\vartheta}$ o $\hat{\vartheta}_N$ il punto di minimo di questo criterio di ottimalità. Ipotizziamo che il modello di predizione sia stabile. Inoltre ipotizziamo che $u(\cdot)$ e $y(\cdot)$ siano processi stazionari. Quindi: $\varepsilon(t) = y(t) - \hat{y}(t)$ sarà anch'esso un processo stazionario visto che $\hat{y}(\cdot)$ è anch'esso. Perciò anche $\varepsilon(t)^2$ è un processo stazionario, e conseguentemente per $N \to \infty$ $\varepsilon(\cdot)^2$ tenderà al suo valore atteso $E[\varepsilon(\cdot)^2]$. Poniamo:

$$\bar{J} = E[\varepsilon^2(t)]$$

Quindi lo stimatore ottimo con N dati, che indicheremo con $\hat{\vartheta}_N$, dipenderà dal vettore ϑ, e si ottiene dal minimo della cifra di merito:

$$\min J_N = (1/N)\sum \varepsilon(t)^2$$

Questo criterio di trovare lo stimatore ottimo prende il nome di CRITERIO QUADRATICO. Sia ora Δ l'insieme dei punti di minimo di $\bar{J}$, ci si aspetta che:

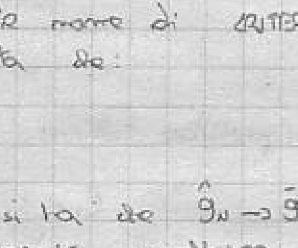
$$\hat{\vartheta}_N \to \Delta$$

Ovviamente se Δ si riduce ad un singolo punto, si ha che $\hat{\vartheta}_N \to \bar{\vartheta}$ con $\bar{\vartheta}$ il punto appena citato. Quindi per ogni $\vartheta \in \Theta$ se la convergenza, per $N \to \infty$, di J_N a $\bar{J}$ avviene in modo regolare, si ha che il punto di minimo di J_N tende al punto di minimo di $\bar{J}$. Quindi si hanno i seguenti casi:

1) Sia S un sistema tale che: $S \in \mathcal{M}$, e Δ è costituito da un punto. Sia poi:

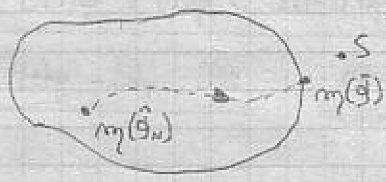
ϑ° = VETTORE DEI PARAMETRI DEL SISTEMA VERO.

Si ha che $\hat{\vartheta}_N \to \vartheta^\circ$. Graficamente si ha:



2) $S \in \mathcal{M}$, ma Δ ha più punti. Δ contiene quindi oltre a ϑ°, altri punti. Il metodo di identificazione a minimizzazione dell'errore di predizione tende ad un punto di Δ, non necessariamente ϑ°. Graficamente:

3) $S \notin \mathcal{M}$, e Δ ha un solo punto. In questo caso, nella famiglia di modelli si è un solo modello ottimo $\mathcal{M}(\bar\vartheta)$. Tale sistema non coincide con il sistema vero, ma è solo un'approssimazione. Graficamente si ha:



4) $S \notin \mathcal{M}$ e Δ è formato da più punti. Graficamente si ha:

Consideriamo ora il seguente ERRORE DI STIMA: $E[(\hat\vartheta_N-\vartheta^o)^2]$. Possiamo riscrivere tale errore nel seguente modo:

$$E[(\hat\vartheta_N-\vartheta^o)^2]=E[\|\hat\vartheta_N-\vartheta^o\|^2]=E[(\hat\vartheta_N-\vartheta^o)(\hat\vartheta_N-\vartheta^o)^T]$$

Per N elevato esiste una formula che permette di calcolare di calcolare la varianza di tale errore. Sia:

$\varepsilon(t)=y(t)-\hat y_\vartheta(t|t-1)$ e sia $\psi(t,\vartheta)=\left[\frac{\partial\varepsilon}{\partial\vartheta_1},\frac{\partial\varepsilon}{\partial\vartheta_2},\dots\right]^T=-\frac{d}{d\vartheta}\varepsilon(t,\vartheta)^T$

ψ prende il nome di GRADIENTE ed ha un numero di elementi pari al numero di elementi di ϑ. Anche il gradiente è ovviamente un vettore. Siccome la derivata di uno scalare come lo è $\varepsilon(\cdot)$ rispetto ad un vettore è per convenzione un vettore riga, ψ^T è un vettore colonna. Si consideri poi la matrice quadrata $\bar R$ così definita:

$$\bar R=E[\psi(t,\vartheta)\psi(t,\vartheta)^T]\Big|_{\vartheta=\vartheta^o}$$

NB: ψ è un vettore "stazionario".

Per N elevato si ha che la varianza dello stimatore $\hat\vartheta_N$ trovato, è: $\frac{1}{N}\,Var[\varepsilon(t,\vartheta^o)]\bar R^{-1}$. Per esattezza quest'ultima formula non va bene. Sostituendo però ϑ^o con $\hat\vartheta_N$ le cose si sistemano. Quindi:

$$\frac{1}{N}\,Var[\varepsilon(t,\hat\vartheta_N)]\bar R^{-1}$$

dove $\varepsilon(t)$ è l'errore di predizione del modello stimato.

La matrice $\bar R$ nella sua "veste" campionaria è data da:

$$(1/N)\sum_{t=1}^{N}\psi(t,\hat\vartheta_N)\psi(t,\hat\vartheta_N)^T$$

Chiusa questa parentesi sul gradiente che verrà ripreso tra breve, vediamo il METODO DEI MINIMI QUADRATI (LS = Least Squares). Si usa questo metodo nel contesto dei modelli della famiglia ARX. Abbiamo quindi il seguente modello:

$$y(t)=a_1y(t-1)+\dots+a_{na}y(t-na)+b_1u(t-1)+\dots+b_{nb}u(t-nb)+\xi(t).$$

Posto: $\vartheta=[a_1\dots a_{na}\ b_1\dots b_{nb}]^T$ e $\varphi(t)=[y(t-1)\dots y(t-na)\ u(t-1)\dots u(t-nb)]^T$

possiamo anche scrivere:

$$m(\vartheta): \; y(t) = \varphi(t)^T \vartheta + \varepsilon(t) = \vartheta^T \varphi(t) + \xi(t)$$

$$\hat{m}(\vartheta): \; \hat{y}(t|t-1) = \vartheta^T \varphi(t) = \varphi(t)^T \vartheta$$

Quindi l'errore di predizione vale: $\varepsilon(t) = y(t) - \varphi(t)^T \vartheta$. Si noti che $y(t)$ è l'uscita del sistema da identificare. Usiamo la solita cifra di merito:

$$J = (1/N) \sum_{t=1}^{N} \varepsilon(t)^2 \Rightarrow J = (1/N) \sum_{t=1}^{N} (y(t) - \varphi(t)^T \vartheta)^2.$$

Facendo la derivata di J rispetto a ϑ si ha:

$$\frac{dJ}{d\vartheta} = \frac{1}{N^2} \sum_{t=1}^{N} (y(t) - \varphi(t)^T \vartheta)(-\varphi(t))^T$$

→ vettore riga.

Ponendo $\frac{dJ}{d\vartheta} = 0$ si ha:

$$\sum_{t=1}^{N} (-\varphi(t)^T y(t) + \varphi(t) \vartheta \varphi(t)^T) = 0 \Rightarrow \sum_t y(t) \varphi(t)^T = \sum_t \varphi(t) \vartheta \varphi(t)^T$$

→ EQUAZIONI NORMALI.

Si può anche scrivere:

$$\left(\sum_t \varphi(t) \varphi(t)^T \right) \vartheta = \sum_t y(t) \varphi(t)^T$$

Se la matrice $\sum_t \varphi(t)\varphi(t)^T$ è singolare allora le equazioni normali hanno infinite soluzioni. Se tale matrice è invece invertibile, si ha:

$$\hat{\vartheta}_N = \left(\sum_t \varphi(t)\varphi(t)^T \right)^{-1} \sum_t y(t) \varphi(t)^T$$

→ FORMULA DEI MINIMI QUADRATI.

Come si può notare, il metodo dei minimi quadrati è un metodo di minimizzazione dell'errore. Ora ci chiediamo però se ϑ trovato è davvero un punto di minimo globale della cifra di merito $\bar{J}(\vartheta)$. Svolgiamo la derivata seconda e notiamo che:

$$d^2J/d\vartheta^2 = \frac{2}{N} \sum_{t=1}^{N} \varphi(t)\varphi(t)^T \geq 0$$

→ Quindi si ha un minimo globale.

Si noti inoltre che la cifra di merito è una funzione quadratica in ϑ. Analizziamo ora la matrice:

$$\sum_t \varphi(t)\varphi(t)^T$$

In breve vogliamo analizzare $\bar{R}$. Tale matrice è semidefinita positivamente e quindi ci chiediamo se è anche definita positivamente.

$$\varphi(t) = \begin{bmatrix} y(t-1) \\ \vdots \\ u(t-1) \\ \vdots \end{bmatrix} \Rightarrow \varphi(t)\varphi(t)^T = \begin{bmatrix} y(t-1) \\ \vdots \\ u(t-1) \end{bmatrix} \begin{bmatrix} y(t-1) \dots u(t-1) \end{bmatrix} = \begin{bmatrix} y(t-1)^2 & \overset{*}{y(t-1)u(t-1)} \\ \underset{**}{u(t-1)y(t-1)} & u(t-1)^2 \end{bmatrix}$$

Quindi $\bar{R}$ è:

$$\bar{R} = \begin{bmatrix} E[y(t-1)^2] & E(*) \\ E[**] & E[u(t-1)^2] \end{bmatrix}$$

. Siccome $R_N = (1/N) \sum \varphi(t)\varphi(t)^T$, si ha:

(45)

$$\hat{R}_N = \begin{bmatrix} \frac{1}{N}\sum y(t-1)^2 & \frac{1}{N}\sum * \\ \frac{1}{N}\sum ** & \frac{1}{N}\sum u(t-1)^2 \end{bmatrix}$$

. Se n_a, n_b sono qualsiasi si ottiene:

$$\varphi(t)\varphi(t)^T = \begin{bmatrix} y(t-1)^2 & y(t-1)y(t-2) & \cdots & y(t-1)u(t-1) & \cdots \\ \vdots & & \vdots & \vdots & \\ u(t-1)y(t-1) & \cdots & \cdots & u(t-1)^2 & \cdots \\ \cdots & & & & \end{bmatrix}$$

Utilizzando il valor medio si ha:

$$E[\varphi(t)\varphi(t)^T] = \begin{array}{c} \\ F \rightarrow \\ K^T \rightarrow \end{array} \left[\begin{array}{cc|cc} \gamma_{yy}(0) & \gamma_{yy}(1) \cdots & \gamma_{yu}(0) & \gamma_{yu}(1) \cdots \\ \gamma_{yy}(1) & \gamma_{yy}(0) \cdots & \gamma_{yu}(1) & \gamma_{yu}(0) \\ & \vdots & & \vdots \\ \hline & & \gamma_{uu}(0) & \gamma_{uu}(1) \cdots \\ & & & \vdots \end{array}\right] \begin{array}{l} \}\, n_a \\ \\ \}\, n_b \end{array}$$

K, H

NB: $\bar{R} = \left[\begin{array}{c|c} \gamma_{yy}(0) & \gamma_{uy}(0) \\ \hline \gamma_{yu}(0) & \gamma_{uu}(0) \end{array}\right]$ (F, K, K^T, H)

Quindi si ha:

$$\left[\begin{array}{c|c} F & K \\ \hline K^T & H \end{array}\right]$$

$H \rightarrow$ sottomatrice

condizione necessaria affinché la matrice sia definita positivamente è che H sia anch'essa definita positivamente.

$\Downarrow$

La stima LS è unica.

Perciò l'ingresso $u(\cdot)$ si dice PERSISTENTEMENTE ECCITATO di ordine n se la matrice H è definita positivamente e quindi è invertibile. (det > 0). Il rumore bianco è un esempio di segnale persistentemente eccitato. Vediamo ora il metodo della MASSIMA VEROSIMIGLIANZA. Anche questo metodo è un metodo di identificazione dell'errore di predizione. Tale metodo si applica sui modelli ARMAX. Il modello in forma predittiva di un ARMAX, è il seguente:

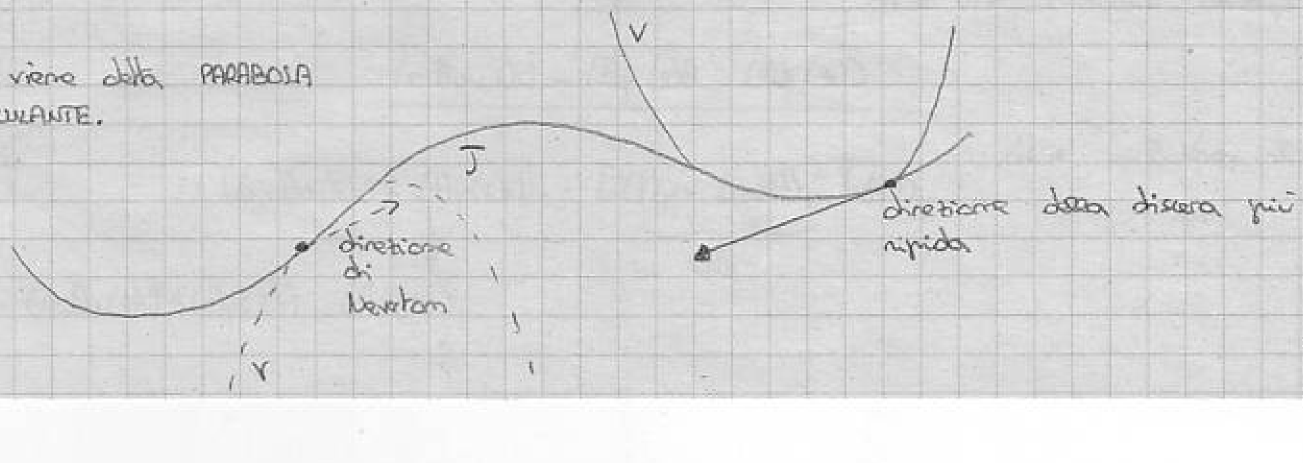
$$\hat{m}(\vartheta) : C(z)\hat{y}(t) = [C(z) - A(z)]\,y(t) + B(z)u(t-1).$$

Consideriamo poi la solita cifra di merito:

$$\bar{J}(\vartheta) = \left(\frac{1}{N}\right)\sum_{t=1}^{N} \varepsilon(t)^2$$

Si noti che qui il predittore $\hat{m}(\vartheta)$ non è più lineare nei parametri, e quindi la cifra di merito non è più quadratica nei parametri. A questo punto la cosa si complica. Non è quindi possibile dare un'espressione esplicita del punto di minimo. Bisogna invece trovare un algoritmo iterativo di ricerca del minimo. Un procedimento molto usato è quello di approssimare la funzione $J(\vartheta)$ intorno ad un certo punto con una funzione quadratica che chiamiamo per comodità $V(\vartheta)$, applichiamo la formula del minimo di un paraboloide per trovare il punto di minimo di $V(\vartheta)$, e si ripete tale procedura sul punto di minimo così trovato. Si ripete il procedimento fino alla convergenza. Graficamente si ha:

NB: V viene detta PARABOLA OSCULANTE.



In breve si indica con $g^{(k)}$ la parametrizzazione alla iterazione k-esima. Si impone che:

- $V(g)$ valutata in $g^{(k)}$ coincide con $J(g^{(k)})$
- Il gradiente di $V(g)$ in $g^{(k)}$ coincida con il gradiente di $J(g)$ in $g^{(k)}$.
- L'Hessiano di $V(g)$ in $g^{(k)}$ coincida con l'Hessiano di $J(g)$ in $g^{(k)}$.

La funzione approssimante è perciò:

$$V(g) = J(g^{(k)}) + \frac{dJ(g^{(k)})}{dg}(g-g^{(k)}) + \frac{1}{2}(g-g^{(k)})^T \frac{d^2J(g^{(k)})}{dg^2}(g-g^{(k)})$$

il cui punto di stazionarietà è:

$$g^{(k+1)} = g^{(k)} - \left(\frac{d^2J(g^{(k)})}{dg^2}\right)^{-1}\left(\frac{dJ(g^{(k)})}{dg}\right)^T$$

Il metodo di Newton può facilmente sbagliare. Analizziamo tale metodo nel problema della stima. Sia:

$$J(g) = \left(\tfrac{1}{N}\right)\sum_{1}^{N}{}_t\, \varepsilon(t)^2$$

$$\frac{dJ(g)^{(k)}}{dg} = \left(\tfrac{2}{N}\right)\sum_{1}^{N}{}_t\, \varepsilon(t)\psi(t)^T$$

$$\frac{d^2J(g^{(k)})}{dg^2} = \left(\tfrac{2}{N}\right)\sum_{1}^{N}{}_t\, \psi(t)\psi(t)^T + \sum_{1}^{N}{}_t\, \varepsilon(t)\frac{d^2\varepsilon(t)}{dg^2}$$

dove $\psi(t) = -\dfrac{d\varepsilon(t)^T}{dg}$ ed è un vettore colonna con un numero di elementi pari al numero di parametri del modello.

A volte l'addendo: $\sum_t \varepsilon(t)\dfrac{d^2\varepsilon(t)}{dg^2}$ che compare nell'Hessiano viene fatto sparire, in modo che l'Hessiano risulti essere una matrice semidefinita positiva.

Quindi in definitiva, il punto di stazionarietà è:

$$g^{(k+1)} = g^{(k)} + \left(\sum_{1}^{N}{}_t\, \psi(t)\psi(t)^T\right)^{-1}\left(\sum_{1}^{N}{}_t\, \varepsilon(t)\psi(t)\right)$$

L'algoritmo è specificato completamente se si dice come calcolare $\varepsilon(t)$ e $\psi(t)$. Per quanto riguarda $\varepsilon(t)$ si ha:

$$C(z)\varepsilon(t) = A(z)y(t) - B(z)u(t-1)$$

In particolare si ha:

$$C(z)\underbrace{[\hat{y}(t|t-1) - y(t)]}_{-\varepsilon(t)} = -(A(z)y(t) + B(z)u(t-1))$$

$$\Downarrow$$

$$C(z)\varepsilon(t) = A(z)y(t) + B(z)u(t-1)$$

Noi conosciamo però $\vartheta^{(k)}$ e quindi: $\vartheta^{(k)} = [\underbrace{a_1^{(k)} \dots a_{na}^{(k)}}_{A(z)^{(k)}} \; \underbrace{b_1^{(k)} \dots b_{nb}^{(k)}}_{B(z)^{(k)}} \; \underbrace{c_1^{(k)} \dots c_{nc}^{(k)}}_{C(z)^{(k)}}]$,
e perciò dati $u(\cdot), y(\cdot)$ si ha:

$$\underline{C(z)^{(k)} \varepsilon(t) = A(z)^{(k)} y(t) - B(z)^{(k)} u(t-1)}$$

Calcoliamoci $\psi(t) = -\dfrac{d\varepsilon}{d\vartheta}$ con $\vartheta = \begin{bmatrix} a_1 \\ \vdots \\ a_{na} \\ b_1 \\ \vdots \\ b_{nb} \\ c_1 \\ \vdots \\ c_{nc} \end{bmatrix}$, $\psi(t) = \begin{bmatrix} \alpha_1(t) \\ \vdots \\ \alpha_{na}(t) \\ \beta(t) \\ \vdots \\ \beta_{nb}(t) \\ \gamma_1(t) \\ \vdots \\ \gamma_{nc}(t) \end{bmatrix}$

Vogliamo trovare tutti gli elementi del vettore colonna $\psi(t)$
Quindi:

$$\alpha_1(t) = -\frac{d\varepsilon}{da_1}$$

$$\alpha_2(t) = -\frac{d\varepsilon}{da_2}$$

$$\vdots$$

Dobbiamo derivare:

$$C(z)\,\varepsilon(t) = A(z)y(t) - B(z)\,u(t-1)$$

rispetto ai parametri. Quindi:

$$A(z)\,y(t) = y(t) - a_1 y(t-1) - \dots \quad \text{e} \quad C(z)\frac{d\varepsilon}{da_1} = y(t-1)$$

$$\Downarrow$$

$$C(z)\alpha_1(t) = y(t-1)$$

Analogamente:

$$C(z)\frac{d\varepsilon(t)}{da_2} = -y(t-2) \Rightarrow \underline{C(z)\alpha_2(t) = -y(t-2)}$$

Il procedimento si ripete fino all'indice na. Definiamo:

Perciò:

$$\underline{C(z)\alpha(t) = y(t)}$$

. È evidente che: $\alpha_1(t) = \alpha(t-1)$, $\alpha_2(t) = \alpha(t-2)$, $\vdots$

$$\psi(t) = \begin{bmatrix} \alpha(t-1) \\ \alpha(t-2) \\ \vdots \\ \alpha(t-na) \\ \beta(t-1) \\ \vdots \\ \beta(t-nb) \\ \gamma(t-1) \\ \vdots \\ \gamma(t-nc) \end{bmatrix}$$

Infatti: $C(z)\dfrac{d\varepsilon(t)}{db_1} = -u(t-1)$ e $-B(z)u(t-1) = -b_1 u(t-1) - b_2 u(t-2) - \dots$
Quindi:

$$\left.\begin{aligned} C(z)\beta_1(t) &= u(t-1) \\ C(z)\beta_2(t) &= u(t-2) \\ &\vdots \end{aligned}\right\} \text{Questa volta si filtra } u.$$

Analogamente

$$C(z)\,\varepsilon(t) = \varepsilon(t) + c_1\varepsilon(t-1) + \dots$$

$$\Downarrow$$

$$C(z)\frac{d\varepsilon}{dc_1} + \underbrace{\frac{dC(z)}{dc_1}}_{z^{-1}}\varepsilon(t) = 0$$

Quindi: $\gamma(t) \rightarrow C(z)\gamma(t) = \varepsilon(t)$
$C(z)\gamma_1(t) = \varepsilon(t-1)$
$\vdots$

con $\underline{\gamma(t) = \gamma(t-1)}$, $\underline{C(z)\dfrac{d\varepsilon}{dc_1} = -\varepsilon(t-1)}$

Quindi ricapitolando l'algoritmo segue i seguenti passi:

1) Al passo i si dispone della stima ϑ cioè $\vartheta^{(k)}$. Corrispondentemente si avranno a disposizione $A(z)^{(k)}$, $B(z)^{(k)}$, $C(z)^{(k)}$. Si filtrano $u(\cdot)$ e $y(\cdot)$ in accordo con l'equazione:

$$\underline{C(z)^{(k)} \varepsilon(t) = A(z)^{(k)} y(t) - B(z)^{(k)} u(t-1)}$$

Si ricava così $\varepsilon(\cdot)^{(k)}$

2) Si filtrano i dati $y(\cdot)$ così: $C(z)^{(k)}\alpha(t) = y(t)$. Si ricava $\alpha(\cdot)^{(k)}$.

3) Si filtrano i dati $u(\cdot)$ così: $C(z)^{(k)}\beta(t) = u(t)$. Si ricava $\beta(\cdot)^{(k)}$

4) Si filtrano $\varepsilon(\cdot)^{(k)}$ con l'equazione: $C(z)^{(k)}\gamma(t) = \varepsilon(t)^{(k)}$ e si ricava $\gamma^{(k)}(\cdot)$.

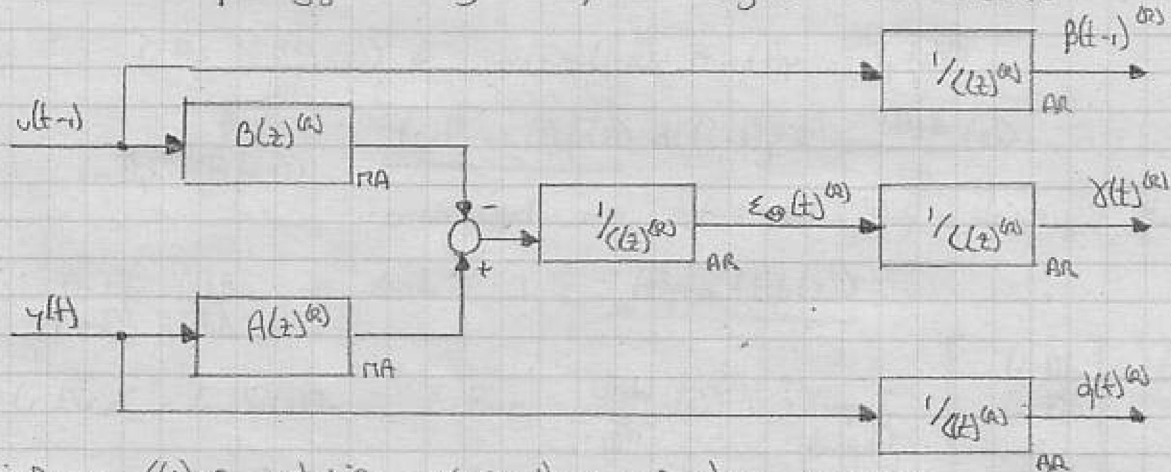
5) Mediante $\alpha(\cdot)^{(k)}$, $\beta(\cdot)^{(k)}$, $\gamma(\cdot)^{(k)}$ si costruisce il vettore $\varphi(\cdot)^{(k)}$ dato da:

$$\varphi(t) = [\,\alpha(t-1) \dots \alpha(t-n_a)\ \beta(t-1) \dots \beta(t-n_b)\ \gamma(t-1) \dots \gamma(t-n_c)\,]$$

6) Utilizzando $\varepsilon(\cdot)^{(k)}$ e $\varphi(\cdot)^{(k)}$ si ricava il nuovo vettore di parametri con la formula:

$$\vartheta^{(k+1)} = \vartheta^{(k)} + \left(\sum_{1}^{N}{}_t\, \varphi(t)^{(k)}\varphi(t)^{(k)T}\right)^{-1}\left(\sum_{1}^{N} \varepsilon_t(t)^{(k)}\varphi(t)^{(k)}\right)$$

Le operazioni di filtraggio vengono qui di seguito schematizzate:



Si noti che se $C(z)$ è instabile, l'algoritmo può dare problemi. Per questo si usa un TEST sulla stabilità. Se qualche singolarità è esterna al cerchio di raggio unitario, non si fa nemmeno il filtraggio. Quindi prima di effettuare il filtraggio, è bene assicurarsi della stabilità di $C(z)$. Vediamo un po' di esercizi.

1) Si consideri il processo ARMA

$$v(t) = a^\circ v(t-1) + \xi(t) + c^\circ \xi(t-1)$$

con $|a^\circ|<1$, $|c^\circ|<1$ e con $\xi(t)$ rumore bianco a valore atteso nullo e varianza unitaria, $\xi \sim WN(0,1)$.
Si effettua l'identificazione con un metodo a minimizzazione dell'errore di predizione, per un numero elevato di dati, con un modello AR(1):

$$v(t) = a v(t-1) + e(t) \qquad \text{con } e \sim WN(0, \delta^2).$$

- Si trovi l'espressione della stima $\hat{a}$ del parametro a che si ottiene dai dati $v(1), v(2), \dots v(N)$, per $N \to \infty$.

- Si ricavi l'espressione della funzione di trasferimento dal rumore $\xi(t)$ all'errore di predizione $\varepsilon(t)$.