$$= \lambda_{gg} + \frac{\lambda_{gd}^2}{\lambda_{dd}^{\cancel{2}}} \cancel{\lambda_{dd}} - 2\frac{\lambda_{gd}}{\lambda_{dd}}\lambda_{gd} = \lambda_{gg} - \frac{\lambda_{gd}^2}{\lambda_{dd}}$$

Si noti che in assenza di osservazioni si ha che la varianza della stima banale ($\hat{g}=0$) è:

$$\underline{Var = \lambda_{gg}}$$

Si ha in breve la così detta STIMA A PRIORI, cioè quella stima priva di ogni misura. In questo caso le uniche informazioni disponibili sull'incognita sono il suo valore atteso e la sua varianza. Nella STIMA A POSTERIORI, invece si ha:

$$\underline{\hat{g} = \frac{\lambda_{gd}}{\lambda_{dd}} d}$$, e se λ_{dd} è molto alto, allora si ha un dato incerto e quindi $\underline{\hat{g} \to 0}$.

Quindi in questa stima sono presenti anche le osservazioni. Notiamo inoltre che:

$$\underline{E[(g-\hat{g})^2] = \lambda_{gg}\left(1 - \frac{\lambda_{gd}^2}{\lambda_{dd}\lambda_{gg}}\right) \leq \lambda_{gg}}$$

In breve $\frac{\lambda_{gd}^2}{\lambda_{gg}\lambda_{dd}}$ è il coefficiente di correlazione ρ. Tale coefficiente indica quanto è informativo il dato rispetto all'incognita. Infatti:

$$\begin{cases} \rho = 0 \Rightarrow E[(g-\hat{g})^2] = \lambda_{gg} \\ \rho = \pm 1 \Rightarrow E[(g-\hat{g})^2] = 0 \end{cases}$$

Supponiamo ora che d, g siano sempre scalari, ma che abbiamo valore atteso non nullo. Per ipotesi quindi si ha:

$$\underline{E[g] = g_m \quad , \quad E[d] = d_m}$$

Lo stimatore diventa è:

$$\underline{\hat{g} = g_m + \frac{\lambda_{gd}}{\lambda_{dd}}(d - d_m)} \qquad \text{STIMATORE OTTIMO.}$$

e la varianza è: $\underline{Var[(g-\hat{g})^2] = \lambda_{gg} - \frac{\lambda_{gd}^2}{\lambda_{dd}}}$. Se invece g e d sono vettoriali, con un generico valore atteso $E[d] = d_m$, $E[g] = g_m$, si ha la seguente matrice varianza:

$$Var\begin{bmatrix} d \\ g \end{bmatrix} = \begin{bmatrix} \Lambda_{dd} & \Lambda_{dg} \\ \Lambda_{gd} & \Lambda_{gg} \end{bmatrix}, \quad \Lambda_{gd} = \Lambda_{dg}$$

* Si noti che Λ_{dd} è una matrice quadrata delle stesse dimensioni del vettore d, Λ_{gg} è una matrice quadrata delle stesse dimensioni del vettore g.

Inoltre si ha:

$$\hat{g} = g_m + \Lambda_{gd}\Lambda_{dd}^{-1}(d - d_m)$$
$$Var[g-\hat{g}] = \Lambda_{gg} - \Lambda_{gd}\Lambda_{dd}^{-1}\Lambda_{dg}$$

La stima di Bayes ammette un'interpretazione geometrica che è molto utile. Questa richiede la definizione di SPAZIO VETTORIALE. Quest'ultimo lo indichiamo, per comodità con la lettera G. Prendiamo in esame quelle variabili casuali dove:

$$\begin{cases} E[v] = 0 \\ Var[v] = \lambda^2 \end{cases}$$

Su tale insieme di variabili aleatorie definiamo le operazioni di somma e prodotto.

Più precisamente: • $V_1 \oplus V_2 = V_1(\zeta)+V_2(\zeta)$ (SOMMA)

• $\alpha \cdot V_1 = \alpha \cdot V_1(\zeta)$ (PRODOTTO PER UNO SCALARE)

Per quanto riguarda il prodotto scalare tra le variabili aleatorie V_1 e V_2, noi lo indicheremo così:

$$(V_1,V_2)$$

Quindi lo spazio vettoriale è identificato da vettori che chiamiamo per comodità V_1 e V_2 e che godono delle operazioni di somma, di prodotto per uno scalare, e di prodotto scalare. In particolare per quest'ultimo si ha:

$$(V_1,V_2) = E[V_1 V_2]$$

ed inoltre:

- $(V_1,V_2) = (V_2,V_1)$ → proprietà COMMUTATIVA
- $(V,V) = Var[V] \geq 0$
- $(V,V) = 0 \iff V \sim G(0,0)$
- $(\mu_1 V_1 + \mu_2 V_2 , V_3) = \mu_1 (V_1,V_3) + \mu_2 (V_2,V_3)$

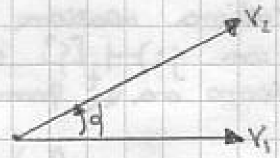
Da tutto ciò segue che la norma di una generica variabile casuale, che nello spazio vettoriale viene vista come un vettore, viene così definita:

$$\|V\| = \sqrt{(V,V)} = \sqrt{Var[V]} : \text{ deviazione standard di } V$$

Grazie al prodotto scalare è possibile definire l'angolo α tra due vettori V_1 e V_2. Quindi per definizione si ha:

$$\cos\alpha = \frac{(V_1,V_2)}{\|V_1\|\|V_2\|}$$

Ma si noti che: $\frac{(V_1,V_2)}{\|V_1\|\|V_2\|} = \frac{E[V_1V_2]}{(Var[V_1]Var[V_2])^{\frac{1}{2}}}$. Graficamente si ha:



In particolare: $\frac{E[V_1V_2]}{(Var[V_1]Var[V_2])^{\frac{1}{2}}} = \frac{E[V_1V_2]}{\underbrace{\lambda_{11}\ \lambda_{22}}_{\text{relative deviazioni standard}}} = \rho = \cos\alpha$.

* Quindi:

- se $\rho = 0$ ⇒ i due vettori sono ortogonali, e quindi le variabili aleatorie sono indipendenti.
- $\rho = 1$ ⇒ i due vettori sono allineati
- $\rho = -1$ ⇒ i due vettori sono allineati ma opposti.

Possiamo ora scrivere la formula di Bayes così:

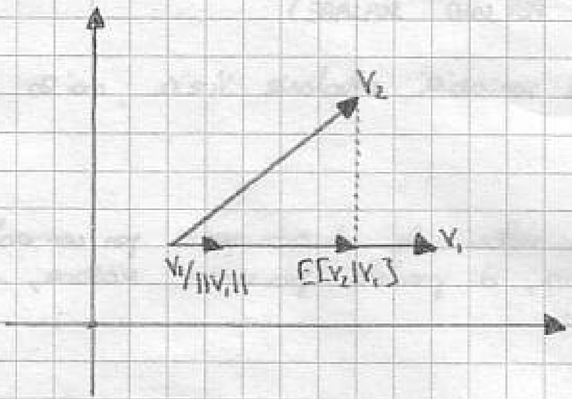
$$E[V_2|V_1] = \frac{\lambda_{12}}{\lambda_{11}} V_1 = \frac{E[V_1V_2]}{Var[V_1]} V_1$$

dove λ_{11} è la varianza di V_1 e λ_{12} è la covarianza tra le due variabili. In particolare si ha:

$$E[V_2|V_1] = \frac{(V_1,V_2)}{\|V_1\|^2} V_1 = \frac{(V_1,V_2)}{\|V_1\|\|V_2\|} \cdot \frac{V_1}{\|V_1\|} \|V_2\| =$$

$$= \|V_2\| \cos\alpha \frac{V_1}{\|V_1\|}$$

Tenendo conto del fatto che $\frac{V_1}{\|V_1\|}$ è il versore orientato come V_1, $E[V_2|V_1]$ altro non è che la proiezione di V_2 su V_1.
Quindi:

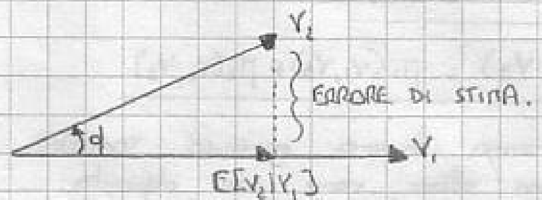


Torniamo ora sull'errore di stima. L'errore di stima è $e = V_2 - V_1$ e quindi la sua varianza può essere calcolata, noto V_1, in questo modo:

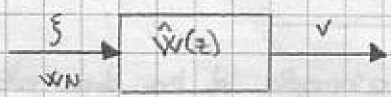
$$Var[V_2 - E[V_2|V_1]] = \|V_2\|^2 - \|E[V_2|V_1]\|^2 = \lambda_{22} - \lambda_{12}^2/\lambda_{11}$$

TEOREMA DI PITAGORA.

dove λ_{22} è la varianza di V_2. Graficamente si ha:



A questo punto, supponiamo di avere una situazione del seguente tipo:



con $\hat{W}(z)$ = FATTORE SPETTRALE CANONICO

* Si ha:

$v(t) = \hat{W}(z)\xi(t)$

Usando il concetto di risposta impulsiva si ha:

$$\hat{W}(z) = \hat{W}(0) + \hat{W}(1) z^{-1} + \dots \Rightarrow v(t) = \hat{W}(0)\xi(t) + \hat{W}(1)\xi(t-1) + \dots$$

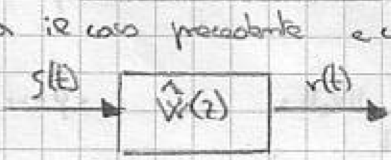
Quest'ultima relazione altro non è che una combinazione lineare in $\xi(t), \xi(t-1), \dots$. Indichiamo con $\xi: \mathcal{H}_t[\xi]$ il passato di ξ. Possiamo quindi scrivere: $v(t) \in \mathcal{H}_t[\xi]$
Riscriviamo ora la formula di $v(t)$ per $v(t-1)$, ottenendo:

$$v(t-1) = \hat{W}(0)\xi(t-1) + \hat{W}(1)\xi(t-2) + \dots = 0\cdot\xi(t) + \hat{W}(0)\xi(t-1) + \dots$$

Quindi se $v(t) \in \mathcal{H}_t[\xi]$, anche $v(t-1) \in \mathcal{H}_t[\xi]$, e quindi tutti i vettori $v(t), v(t-1), \dots v(t-n) \in \mathcal{H}_t[\xi]$.
Quindi anche una loro combinazione lineare è tale che: $\alpha v(t) + \beta v(t-1) + \dots \in \mathcal{H}_t[\xi]$.
In conclusione si ha:

$$\mathcal{H}_t[v] \subseteq \mathcal{H}_t[\xi]$$

Riconsideriamo ora il caso precedente e cioè:



* Grazie alla proprietà di $\hat{W}(z)$ di possedere il grado del numeratore uguale al grado del denominatore, posso invertire $\hat{W}(z)$ ottenendo:

$\hat{W}(z)^{-1} = \check{W}(z)$

Graficamente si ha: $v(t) \longrightarrow \boxed{\check{W}_v(z)} \xrightarrow{\xi(t)}$ * Si noti che in generale l'inversa di una funzione di trasferimento stabile non è stabile, ma per il fattore canonico, l'inversa è stabile.

Quindi filtrando $v(t)$ otteniamo il rumore bianco. Questo filtro si chiama FILTRO SBIANCANTE.

Scriviamo quindi:

$$\check{W}_v(z) = \check{w}_v(0) + \check{w}_v(1) z^{-1} + \dots$$, dove $\check{w}_v(0), \check{w}_v(1) \dots$ sono i coefficienti del filtro sbiancante.

$$\Downarrow$$

$$\xi(t) = \check{w}_v(0)\, v(t) + \check{w}_v(1)\, v(t-1) + \dots \qquad \text{con } \xi(t) \in H_t[v]$$

Analogamente:

$$\underline{\xi(t-1) \in H_t[v]}$$

Quindi: $H_t[\xi] \subseteq H_t[v]$, e quindi questi due sottospazi devono coincidere. Riprendiamo in esame il precedente schema:

$$\xi \longrightarrow \boxed{\hat{W}(z)} \xrightarrow{v}$$

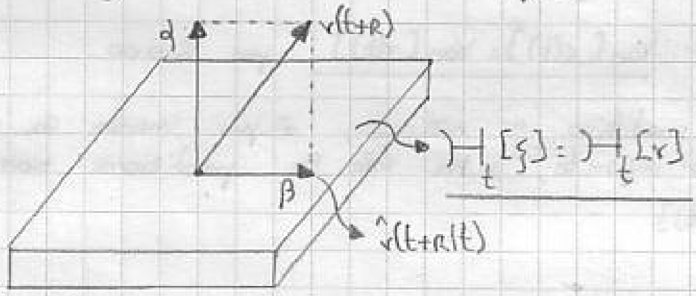
*Abbiamo: $v(t+R) = \hat{w}_v(0)\, \xi(t+R) + \dots + w_v(R-1)\, \xi(t+1) + \hat{w}_v(R)\, \xi(t) + \hat{w}_v(R+1)\, \xi(t-1) + \dots$

Possiamo scomporre $v(t+R)$ in questo modo: $\underline{v(t+R) = \alpha(t) + \beta(t)}$ NB: $\alpha(t)$ e $\beta(t)$ sono incorrelati.

con:

$$\begin{cases} \alpha(t) = \hat{w}_v(0)\, \xi(t+R) + \dots + \hat{w}_v(t-1)\, \xi(t+1) \\ \beta(t) = \hat{w}_v(R)\, \xi(t) + \hat{w}_v(R+1)\, \xi(t-1) + \dots \end{cases}$$

Possiamo rappresentare il tutto in questo modo:



Quindi per ottenere la predizione di $v(t+R)$ considerando i dati fino all'istante t, bisogna eseguire una proiezione di $v(t+R)$ sul piano dove $H_t[\xi] = H_t[v]$. In conclusione:

$$\hat{v}(t+R|t) = \hat{w}_v(R)\, \xi(t) + w_v(R+1)\, \xi(t-1) + \dots = \underbrace{\left(\hat{w}_v(R) + \hat{w}_v(R+1) z^{-1} + \dots\right)}_{\hat{W}_R(z)} \xi(t)$$

$$\xi \longrightarrow \boxed{\hat{W}_R(z)} \xrightarrow{\hat{v}(t+R|t)}$$

Si noti che è $\beta(t)$ ad essere la proiezione di $v(t+R)$, e pertanto:

$$\underline{\hat{v}(t+R|t) = \beta(t)}$$

$\alpha(t)$ invece è l'errore di predizione e quindi: $\underline{\varepsilon(t) = v(t+R) - \hat{v}(t+R|t) = \hat{w}_v(0)\, \xi(t+R) + \dots + \hat{w}_v(R-1)\, \xi(t+1)}$

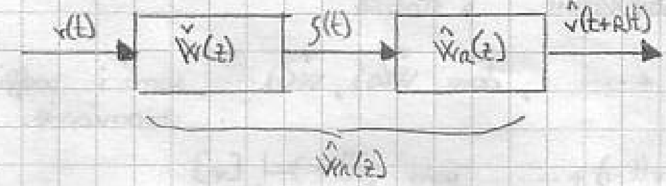
Quindi si può scrivere:

$$\underline{\hat{W}_R(z) = \hat{W}(R) + \hat{w}_v(R+1) z^{-1} + \dots}$$

La funzione di trasferimento $\hat{W}_R(z)$ si ottiene dalla funzione di trasferimento $\hat{W}(z)$ posta in forma canonica effettuando la lunga divisione per R passi, e moltiplicando la restante funzione di trasferimento per z^R. In altre parole se la nostra funzione di trasferimento in forma canonica è data da:

$$\hat{W}(z) = C(z)/A(z) \Rightarrow \hat{W}_R(z) = G_R(z)/A(z)$$, dove $G_R(z)$ è il polinomio ottenuto

moltiplicando per z^k il resto della divisione di $C(z)$ per $A(z)$ iterata per k passi. Il predittore così ottenuto ha come ingresso il rumore bianco che tra l'altro non è disponibile. Si può però fare la predizione a partire da $v(t)$. Per fare ciò basta sbiancare $v(\cdot)$ con un filtro sbiancante ottenendo $\xi(\cdot)$ in modo da poter calcolare poi $\hat{v}(t+k|t)$. Graficamente si ha:



Quindi:

$$W_k(z) = \check{W}(z)\cdot\hat{W}_k(z) = \hat{W}(z)^{-1}\cdot\hat{W}_k(z) = \frac{A(z)}{C(z)}\cdot\frac{G_k(z)}{A(z)} = G_k(z)/C(z)$$

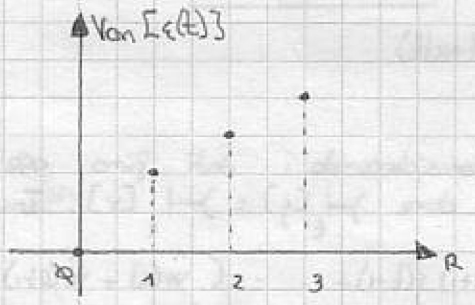
Poiché numeratore e denominatore del filtro spettrale canonico hanno singolarità interne al cerchio di raggio 1, sia il predittore $\xi(\cdot)$ che quello da $v(\cdot)$ sono stabili. La varianza dell'errore di predizione si ricava così:

$$\underline{Var[\varepsilon(t)] = (\hat{w}(0)^2 + \hat{w}(1)^2 + \dots + \hat{w}(k-1)^2)\lambda^2}$$

Questo accade perché $d(t) = \hat{w}(0)\xi(t+k) + \hat{w}(1)\xi(t+k-1) + \dots$ è un processo $MA(k)$. Quindi la varianza di $d(t)$ che non è altro che la varianza di $\varepsilon(t)$ è proprio la precedente relazione. Siccome $\hat{w}(0)=1$ per la monicità del numeratore e del denominatore del filtro spettrale canonico, si ha che la varianza di predizione ad un passo viene a coincidere con la varianza del rumore bianco $\xi(t)$. Al crescere di k cresce monotonicamente anche la varianza. In particolare:

$$\underline{Var[\varepsilon(t)] = Var[v(t)]} \quad \text{per } k \to \infty$$

Infatti se l'orizzonte predittivo si allunga, è più incerta la predizione, e quando tale orizzonte diverge all'infinito, si ha solo la predizione banale. Graficamente si ha:



Vediamo ora un esempio. Consideriamo il processo $ARMA(2,1)$ definito dalla seguente equazione:

$$v(t) + 0{,}9v(t-1) + 0{,}2v(t-2) = \xi(t) + 0{,}8\xi(t-1)$$

NB: $W(z)$ è già in forma canonica.

La funzione di trasferimento $W(z)$ è data da:

$$W(z) = \frac{v(t)}{\xi(t)} = \frac{1+0{,}8z^{-1}}{1+0{,}9z^{-1}+0{,}2z^{-2}} = \frac{z^2+0{,}8z}{z^2+0{,}9z+0{,}2} = \frac{z^2+0{,}8z}{(z+0{,}5)(z+0{,}4)}$$

Effettuiamo la lunga divisione:

$z^2+0{,}8z$	$z^2+0{,}9z+0{,}2$
$-(z^2+0{,}9z-0{,}2)$	1
$-0{,}1z-0{,}2$	

Perciò:

$$W(z) = 1 + z^{-1}\frac{(-0{,}1z-0{,}2)z}{z^2+0{,}9z+0{,}2}$$

predittore dal rumore.

Quindi il predittore a un passo è:

$$W(z) = 1 + \frac{-0{,}1z-0{,}2}{z^2+0{,}9z+0{,}2}$$

Se si vuole il predittore a un passo da $v(\cdot)$ si ha: $W(z) = -\dfrac{0,1z^2+0,2z}{z^2+0,8z}$.
Si noti che nella funzione di trasferimento del predittore entra solo la parte predicibile, ossia misurabile. La parte non predicibile non viene presa in considerazione. Quindi:

$$\left(\text{QUOZIENTE} + \frac{\text{RESTO}}{\text{DENOMINATORE } W(z)}\right) e(t+R)$$

dove: $\begin{cases} e(t+R) - \text{QUOZIENTE} = \alpha(t) \\ \beta(t) = \dfrac{\text{RESTO}}{\text{DENOMINATORE } (W(z))} \, e(t+R) \end{cases}$

* Quindi si noti che si considera soltanto $\beta(t)$. Le operazioni da svolgere per trovare il predittore ottimo a 2 passi sono le stesse. Se $W(z)$ non è in forma canonica bisogna porla sotto tale forma.

Analizziamo ora la predizione ad un passo dei processi ARMA. Si prenda in considerazione un generico processo ARMA scritto nella seguente maniera:

$$A(z)\, v(t) = C(z)\, \xi(t)$$

dove: $\begin{cases} A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_{na} z^{-na} \\ C(z) = 1 + c_1 z^{-1} + \dots + c_{nc} z^{-nc} \end{cases}$

Indichiamo con n il massimo tra na e nc, e scrivendo la funzione di trasferimento secondo le potenze positive si ottiene:

$$\frac{z^n C(z)}{z^n A(z)} = \frac{z^n - c_1 z^{n-1} - c_2 z^{n-2} - \dots}{z^n - a_1 z^{n-1} - a_2 z^{n-2} - \dots} = \frac{v(t)}{\eta(t)} = W(z)$$

Se $A(z)$ e $C(z)$ sono stabili allora $W(z)$ è posta in forma canonica. Quindi il predittore ottimo si ottiene nella seguente maniera:

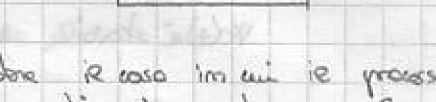
$$\frac{C(z)}{A(z)} = 1 + \frac{C(z) - A(z)}{A(z)} = 1 + z^{-1} \frac{z[C(z) - A(z)]}{A(z)}$$

Perciò il predittore ottimo ad un passo da $v(\cdot)$ è: $W_1(z) = \dfrac{z(C(z) - A(z))}{A(z)}$

$$\Downarrow$$

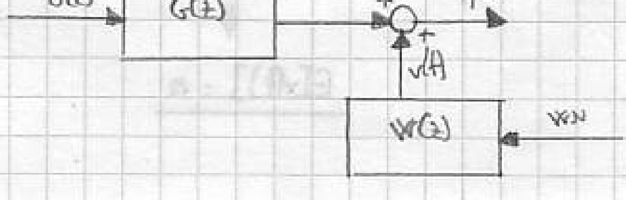
$$C(z)\, \hat{v}(t+1|t) = z(C(z) - A(z))\, v(t) = (C(z) - A(z))\, v(t+1)$$

Si osservi che la stabilità del predittore dipende esclusivamente dal polinomio $C(z)$.
Fino ad ora abbiamo sempre considerato sistemi di questo tipo:



con $W(z) = \dfrac{v(t)}{\eta(t)}$

Può anche succedere il caso in cui il processo sia influenzato da variabili esogene cioè da variabili provenienti dall'esterno. Per esempio:



con $u(t)$ = variabile esogena

In generale un modello ARMAX è un modello scritto in questo modo:

$$v(t) = a_1 v(t-1) + a_2 v(t-2) + \dots + a_{m_a} v(t-m_a) + b_1 u(t-1) + b_2 u(t-2) + \dots + b_{m_b} u(t-m_b) + \xi(t) + c_1 \xi(t-1) + \dots + c_{m_c} \xi(t-m_c)$$

Se per ipotesi $b_1 \neq 0$, il predittore ottimo si ottiene da quello relativo al modello ARMA sommando la parte X, cioè:

$$\hat{v}(t+1|t) = -c_1 \hat{v}(t|t-1) - c_2 \hat{v}(t-1|t-2) - \dots - c_{m_c} \hat{v}(t-m_c+1|t-m_c) + (a_1+c_1) v(t) + (a_2+c_2) v(t-1) + \dots + b_1 u(t) + b_2 u(t-1) + \dots + b_{m_b} u(t-m_b)$$

Oppure:

$$\underline{C(z)\hat{v}(t+1|t) = (C(z) - A(z))\, v(t+1) + B(z)\, u(t)} \qquad \text{con } B(z) = b_1 + b_2 z^{-1} + \dots + b_{m_b} z^{-m_b+1}$$

Si noti che $\underline{B(z) = G(z)}$. Vediamo ora un po' di esercizi:

1) Si consideri il modello ARMA(1,1):

$$v(t) = \tfrac{1}{4} v(t-1) + \eta(t) + 2\eta(t-1)$$

dove $\eta(t)$ è un rumore bianco $\sim(0, \tfrac{1}{4})$.

- Si dica perché la soluzione tende a un processo stazionario.
- Per il processo stazionario di regime si calcolino il valore atteso, la varianza $\gamma(0)$ e la funzione di covarianza al passo 1, $\gamma(1)$.
- Si trovi la regola di predizione ottima di $v(t+1)$ a partire da $v(t), v(t-1), v(t-2), \dots$
- Si dica quanto vale la varianza dell'errore di predizione a 1 passo.
- Si trovi la regola di predizione ottima di $v(t+2)$ a partire da $v(t), v(t-1), v(t-2), \dots$
- Si dica quanto vale la varianza dell'errore di predizione a 2 passi.

1) • Calcoliamo $W(z)$:

$$W(z) = \frac{v(t)}{\eta(t)}, \quad \text{ma: } v(t) - \tfrac{1}{4} v(t-1) = \eta(t) + 2\eta(t-1)$$

$$\Downarrow$$

$$v(t)(1 - \tfrac{1}{4} z^{-1}) = \eta(t)(1 + 2z^{-1}) \Rightarrow \frac{v(t)}{\eta(t)} = \frac{(1+2z^{-1})}{(1-\frac{1}{4}z^{-1})}$$

Quindi:

$$W(z) = \frac{z+2}{z-\frac{1}{4}} \Rightarrow \text{esiste un polo } (\tfrac{1}{4}) \text{ interno al cerchio unitario.}$$

$$\Downarrow$$

$W(z)$ è stabile $\Rightarrow$ processo stz.

- $$E[v(t)] = \tfrac{1}{4} E[v(t-1)] + \overbrace{E[\eta(t)]}^{0} + 2\overbrace{E[\eta(t-1)]}^{0} = \tfrac{1}{4} E[v(t-1)]$$

$$\Downarrow$$

$$\underline{E[v(t)] = 0}$$