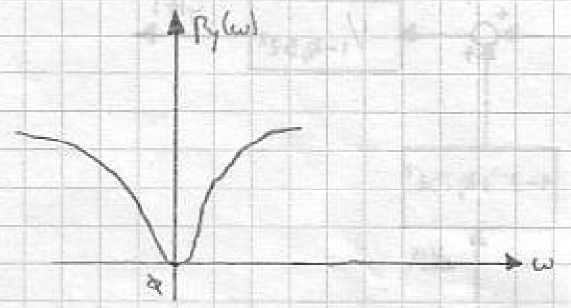
• $\frac{y(t)}{d(t)} E(z) = 1 - z^{-1} \Rightarrow \Gamma_y(\omega) = |E(e^{j\omega})|^2 \, Var(d(t)) = |1 - e^{-j\omega}|^2 = 2(1 - \cos\omega)$



* FILTRO DI KALMAN:

Il problema del filtraggio alla Kalman viene così impostato. Si consideri il seguente sistema dinamico:

$$x(t+1) = Fx(t) + v_1(t)$$
$$y(t) = Hx(t) + v_2(t)$$

Si osservi che y è un segnale misurato mentre x no. $x(t)$ denota lo stato ed è un vettore di n componenti. Anche l'uscita $y(t)$ è un vettore di p componenti. $v_1(t)$ e $v_2(t)$ sono disturbi descritti in modo probabilistico. Ipotizziamo che $v_1(t)$ e $v_2(t)$ siano rumori bianchi incorrelati tra loro e a valore atteso nullo. Le matrici varianza di $v_1(t)$ e $v_2(t)$ verranno indicate con V_1 e V_2. Quindi:

$$v_1(t) \sim WN(0, V_1) \qquad v_2(t) \sim WN(0, V_2)$$

* V_1, V_2, F, H sono comunque tutte matrici note.

Anche lo stato iniziale è un ulteriore elemento di incertezza. Se consideriamo l'istante iniziale $(t=1)$, l'insieme delle variabili incerte in $[1, N]$ può essere così organizzato:

$$\vartheta = \begin{bmatrix} x(1) \\ v_1(1) \\ \vdots \\ v_1(N) \\ v_2(1) \\ \vdots \\ v_2(N) \end{bmatrix}$$

* Quindi $y(t)$ rappresenta il dato e $x(t)$ l'incognita.

I problemi che possiamo affrontare sono i problemi di stima, di predizione, o di filtraggio. Però se scriviamo:

$$\hat{x}(t+k|t) \Rightarrow \begin{array}{l} k=0 \rightarrow \text{FILTRAGGIO} \\ k>0 \rightarrow \text{PREDIZIONE} \\ k<0 \rightarrow \text{REGOLARIZZAZIONE} \end{array}$$

Iniziamo con il problema di predizione. Siccome $v_1(t)$ e $v_2(t)$ vengono descritti in maniera probabilistica, $x(t)$ e $y(t)$ sono variabili casuali. Quindi ci siamo ricondotti ad un problema di stima da affrontare con il "metodo" di Bayes. Si ricordi che lo stimatore di Bayes è:

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

Nel nostro caso si ha:

$$\hat{x}(N+R|N) = x(N+R)_m + \Lambda_{x(N+R)d}\Lambda_{dd}^{-1}(d-d_m)$$

dove $x(N+R)_m$ è il valore atteso di $x(N+R)$. Siccome i disturbi hanno valore atteso nullo, anche $x(t)$ e $y(t)$ hanno valore atteso nullo. Quindi:

$$\hat{x}(N+R|N) = \Lambda_{x(N+R)d}\Lambda_{dd}^{-1} d \qquad \text{con } d = y.$$

Ipotizziamo ora che g sia l'incognita del problema e d l'insieme di dati. Per semplicità supponiamo che g sia scalare e d sia formato da due dati $d(1)$, $d(2)$. Inoltre supponiamo che il valore atteso di $d(1)$, $d(2)$, g sia nullo. Generalizzando si ottiene:

$$\begin{bmatrix} g \\ d(1) \\ d(2) \end{bmatrix} \sim \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \lambda_{gg} & \lambda_{g1} & \lambda_{g2} \\ \lambda_{1g} & \lambda_{11} & \lambda_{12} \\ \lambda_{2g} & \lambda_{21} & \lambda_{22} \end{bmatrix} \right) \qquad \text{dove } \lambda_{g1} = E[g\,d(1)].$$

$\vdots$

Quindi in virtù della formula di Bayes la stima di g basata sul dato $d(1)$ è:

$$E[g|d(1)] = \frac{\lambda_{g1}}{\lambda_{11}} d(1)$$

mentre la stima di g basata su $d(1)$ e $d(2)$ è:

$$E[g|d(1),d(2)] = [\lambda_{g1}, \lambda_{g2}] \cdot \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix}^{-1} \cdot \begin{bmatrix} d(1) \\ d(2) \end{bmatrix}, \qquad \lambda_{21} = \lambda_{12}$$

calcolando la matrice inversa e sviluppando i prodotti si ottiene:

$$E[g|d(1),d(2)] = \frac{1}{\lambda_{11}\lambda^2}(-\lambda_{g1}\lambda_{21} + \lambda_{g2}\lambda_{11})\,d(2) + \frac{1}{\lambda_{11}\lambda^2}(\lambda_{g1}\lambda_{22} - \lambda_{g2}\lambda_{12})\,d(1)$$

dove si è posto:

$$\lambda^2 = \lambda_{22} - \frac{\lambda_{12}^2}{\lambda_{11}}$$

Aggiungendo ora $(\lambda_{g1}/\lambda_{11})d(1)$ si ha: $E[g|d(1),d(2)] = \frac{1}{\lambda_{11}\lambda^2}(-\lambda_{g1}\lambda_{21} + \lambda_{g2}\lambda_{11})d(2) + \frac{1}{\lambda_{11}\lambda^2}(\lambda_{g1}\lambda_{22} -$

$$-\lambda_{g2}\lambda_{12})\,d(1) - \frac{\lambda_{g1}}{\lambda_{11}}d(1) + \frac{\lambda_{g1}}{\lambda_{11}}d(1)$$

Quindi:

$$E[g|d(1),d(2)] = \frac{1}{\lambda^2}\left(\lambda_{g2} - \lambda_{g1}\frac{\lambda_{12}}{\lambda_{11}}\right)d(2) + \frac{1}{\lambda^2}\left(\lambda_{g1}\frac{\lambda_{22}}{\lambda_{11}} - \lambda_{g2}\frac{\lambda_{12}}{\lambda_{11}} - \frac{\lambda_{g1}}{\lambda_{11}}\lambda^2\right)d(1) + \frac{\lambda_{g1}}{\lambda_{11}}d(1)$$

$$\Downarrow$$

Sostituendo al posto di λ^2 l'espressione prima trovata si ottiene: $\lambda_{g1}\lambda_{22}/\lambda_{11} + (-)\lambda_{g2}\frac{\lambda_{12}}{\lambda_{11}} - \frac{\lambda_{g1}}{\lambda_{11}}\lambda^2 = \frac{\lambda_{12}}{\lambda_{11}}\left(-\lambda_{g2} + \lambda_{g1}\frac{\lambda_{12}}{\lambda_{11}}\right.$

Si ottiene alla fine:

$$E[g \mid d(1), d(2)] = \frac{\lambda_{g1}}{\lambda_{11}} d(1) + \frac{1}{\lambda^2}\left(\lambda_{g2} - \lambda_{g1}\frac{\lambda_{12}}{\lambda_{11}}\right)\left[d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1)\right]$$

Si dice INNOVAZIONE di $d(2)$ rispetto a $d(1)$ la seguente quantità:

$$e = d(2) - E[d(2) \mid d(1)] = d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1)$$

Si noti innanzitutto che l'innovazione è una variabile gaussiana perché deriva dalla somma di due variabili gaussiane $(d(1), d(2))$. Le principali proprietà dell'innovazione sono:

1) $E[e] = 0$

2) $\lambda_{ee} = Var[e] = E\left[\left(d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1)\right)^2\right] = \lambda_{22} + \frac{\lambda_{12}^2}{\lambda_{11}^2}\lambda_{11} - 2\frac{\lambda_{12}^2}{\lambda_{11}} = \lambda^2$

3) $\lambda_{ge} = E\left[g\left(d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1)\right)\right] = \lambda_{g2} - \lambda_{g1}\frac{\lambda_{12}}{\lambda_{11}}$.

4) $\lambda_{1e} = E\left[d(1)\left(d(2) - \frac{\lambda_{12}}{\lambda_{11}} d(1)\right)\right] = \lambda_{12} - \frac{\lambda_{12}}{\lambda_{11}}\lambda_{11} = 0$

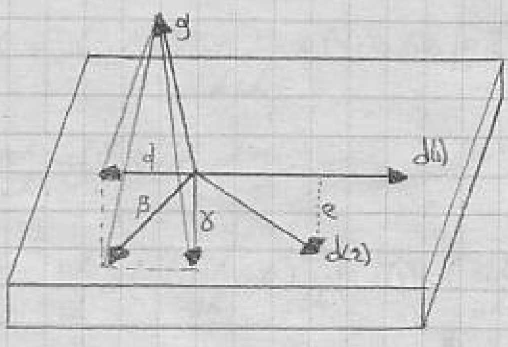
Quindi $E[g \mid d(1), d(2)]$ può essere così riscritta: $E[g \mid d(1), d(2)] = \frac{\lambda_{g1}}{\lambda_{11}} d(1) + \frac{\lambda_{ge}}{\lambda_{ee}} e$.
Si può perciò concludere affermando che:

$$E[g \mid d(1), d(2)] = E[g \mid d(1)] + E[g \mid e]$$

Si noti che non vale però la seguente relazione: $E[g \mid d(1), d(2)] = E[g \mid d(1)] + E[g \mid d(2)]$. L'unica volta in cui tale relazione è vera si ha quando:

$$E[d(2) \mid d(1)] = 0 \Rightarrow d(2) = e.$$

Vediamo ora la interpretazione geometrica di quanto appena visto. Siano g, $d(1)$, $d(2)$ dei vettori:



$$\begin{cases} \alpha = E[g \mid d(1)] \\ \beta = E[g \mid d(1), d(2)] \\ \gamma = E[g \mid e] \end{cases}$$

Quindi la stima di g conoscendo $d(1)$ e $d(2)$ è data dalla proiezione di g sul piano.

La stima di g conoscendo solo $d(1)$ è data invece dalla proiezione di g su $d(1)$. Vediamo ora la naturale estensione al caso vettoriale. Consideriamo:

$$\begin{bmatrix} g \\ d(1) \\ d(2) \end{bmatrix} \sim G\left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \Lambda_{gg} & \Lambda_{g1} & \Lambda_{g2} \\ \Lambda_{1g} & \Lambda_{11} & \Lambda_{12} \\ \Lambda_{2g} & \Lambda_{21} & \Lambda_{22} \end{bmatrix} \right)$$

dove $\Lambda_{g1} = \Lambda_{1g}^T$; $\Lambda_{2g}^T = \Lambda_{g2}$; $\Lambda_{12} = \Lambda_{21}^T$. Definiamo quindi anche l'innovazione nel caso vettoriale nel seguente modo:

$$e = d(2) - E[d(2)|d(1)] = d(2) - \Lambda_{21}\Lambda_{11}^{-1} d(1)$$

Si ha quindi:

$$E[g|d(1), d(2)] = \Lambda_{g1}\Lambda_{11}^{-1} d(1) + \Lambda_{ge}\Lambda_{ee}^{-1} e = E[g|d(1)] + E[g|e].$$

Così si ottiene l'interpretazione geometrica rappresentata in precedenza. Predittiamo ora il predittore di Kalman. Nel problema della predizione ad un passo i dati sono costituiti dalle osservazioni dell'uscita y, dall'istante 1 all'istante N. Indichiamo con y^N tale insieme di dati. Quindi:

$$y^N = [y(N)', y(N-1)', \dots, y(1)']$$

Indicheremo poi con $\mathcal{H}[y^N]$ il SOTTOSPAZIO DEL PASSATO. L'innovazione del dato $N+1$ è data da:

$$e(N+1) = y(N+1) - E[y(N+1)|y^N]$$

Ma come abbiamo già visto, $E[y(N+1)|y^N]$ altro non è che la proiezione di $y(N+1)$ su $\mathcal{H}[y^N]$. Si vede che l'innovazione è ortogonale al passato. Consideriamo ora un'altra variabile, che nella teoria di Kalman riveste una notevole importanza. Tale variabile prende il nome di ERRORE DI PREDIZIONE DELLO STATO. La indicheremo con $v(N+1)$. Essa è così definita:

$$v(N+1) = x(N+1) - E[x(N+1)|y^N)]$$

L'errore di predizione dello stato ha n componenti, tante quante lo stato del sistema. Si può notare geometricamente che anche $v(N+1)$ è ortogonale al passato. La predizione ottima ad un passo dell'uscita è data da:

$$\hat{y}(N+1|N) = E[y(N+1)|y^N]$$

$$\Downarrow$$

$$\hat{y}(N+1|N) = E[Hx(N+1) + v_2(N+1|y^N)] = HE[x(N+1)|y^N] + \overbrace{E[v_2(N+1)|y^N]}^{\text{STIMA A PRIORI}}$$

Mostriamo che il secondo addendo è nullo, mentre osservando la seguente equazione:

$$x(N+1) = Fx(N) + v_1(N) \Rightarrow$$ $x(N)$ dipende dal disturbo $v_1(\cdot)$ fino al tempo $N-1$, oltre che dalle condizioni iniziali.

Quindi: $x(N) = p(v_1^{N-1}, x(1))$. Perciò:

$$y(N) = p(v_1^{N-1}, x(1), v_2(N))$$

Concludendo si può notare che y^N è indipendente da $V_2(N+1)$ e quindi $E[V_2(N+1)|y^N]$ coincide con quella non condizionata $E[V_2(N+1)]$. Quindi:

$$\underline{\hat{y}(N+1|N) = H\hat{x}(N+1|N)}$$

Questo si ha perché $E[V_2(N+1)] = 0$. Vogliamo ora la predizione ottima di x. Vogliamo cioè:

$$\hat{x}(N+1|N) = E[x(N+1)|y^N] = E[x(N+1)|y^{N-1}, y(N)] = \dots$$

Usiamo la formula di Bayes ricorsiva:

$$E[x(N+1)|y^{N-1}] + E[x(N+1)|e(N)]$$

$$\Downarrow$$

$$E[x(N+1)|e(N)] = \Lambda_{x(N+1)e(N)} \Lambda^{-1}_{e(N)e(N)} e(N) \quad \text{con} \quad e(N) = y(N) - \hat{y}(N|N-1) =$$

$$= Hx(N) + V_2(N) - H\hat{x}(N|N-1) =$$

$$= H(x(N) - \hat{x}(N|N-1)) + V_2(N) =$$

$$= Hv(N) + V_2(N)$$

Quindi:

$$\Lambda_{e(N)e(N)} = E[e(N)e(N)'] =$$

$$= E[Hv(N)v(N)'H'] + E[V_2(N)V_2(N)'] +$$

$$+ \cancel{E[Hv(N)V_2(N)'] + \dots}^{\,0} = HE[v(N)v(N)']H' + \underbrace{E[V_2(N)V_2(N)']}_{V_2} =$$

$$= HE[v(N)v(N)']H' + V_2$$

NB: $' = T$

Chiamiamo:

$P(N) = E[v(N)v(N)']$ → matrice VARIANZA DELL'ERRORE DI PREDIZIONE DI STATO.

Si ha: $\underline{\Lambda_{e(N)e(N)} = HP(N)H' + V_2}$

Analogamente si ha: $\Lambda_{x(N+1)e(N)} = FP(N)H'$. In conclusione si ha:

$$\underline{E[x(N+1)|e(N)] = \underbrace{(FP(N)H')(HP(N)H'+V_2)^{-1}}_{k(N) = \text{GUADAGNO DI KALMAN}} e(N)}$$

Ma noi siamo partiti con:

$$\hat{x}(N+1|N) = E[x(N+1)|y^{N-1}] + E[x(N+1)|e(N)]$$

$$\Downarrow$$

$$\underline{F\hat{x}(N+1|N) + k(N)e(N)}$$

Concludendo:

$$S: \; x(t+1) = Fx(t) + \overset{\text{rumore}}{V_1(t)}$$

$$\Downarrow$$

$$P: \; \hat{x}(N+1|N) = F\hat{x}(N|N-1) + k(N)e(N) \quad \text{con} \begin{cases} e(N) = y(N) - \hat{y}(N|N-1) \\ \hat{y}(N|N-1) = H\hat{x}(N|N-1) \\ k(N) = (FP(N)H')(HP(N)H'+V_2)^{-1} \end{cases}$$

Si noti che l'equazione del predittore di Kalman può essere così inizializzata:

$$\begin{cases} \hat{x}(1|0) = ? \\ P(1) = ? \end{cases} \quad \text{, ma dati } y(1), y(2), \dots, y(N)$$

$$\Downarrow$$

$$\hat{x}(1|0) = E[\, x(1) \,|\, y(0), y(-1), \dots] \quad \text{, ma } E[x(t)] = 0$$

Inoltre: $Var[x(1)] = P(1)$ e quindi:

$$\nu(1) = x(1) - \hat{x}(1|0) = \dots$$

In generale si noti che:

$$\hat{x}(2|1) = E[x(2)|y(1)] = \Lambda_{x(2)y(1)} \Lambda^{-1}_{y(1)y(1)} y(1) = F P_1 H^T (H P_1 H^T + V_2)^{-1} y(1).$$

Tenendo conto che: $e(N) = y(N) - \hat{y}(N|N-1) = y(N) - H\hat{x}(N|N-1)$; $\hat{x}(N+1|N) = F\hat{x}(N|N-1) + K(N)e(N)$

$$\Downarrow \quad \text{(ho svolto il confronto).}$$

$$\underline{\hat{x}(1|0) = 0}$$

Quindi in assenza di informazioni la stima più ragionevole di $x(1)$ è quella del valore atteso, che è nullo. Analizziamo ora il problema del filtraggio alla Kalman. Proviamo prima ad estendere la teoria vista precedentemente sul predittore ad un passo di Kalman, al caso di predittore a più passi. A tale scopo consideriamo:

$$x(t+1) = Fx(t) + v_1(t)$$
$$y(t) = Hx(t) + v_2(t)$$

dove $v_1(\cdot)$ e $v_2(\cdot)$ sono rumori bianchi a valore atteso nullo, eventualmente correlati tra loro. Quindi:

$$\underline{E[v_i(t_1) v_j(t_2)] = V_{ij}\, \delta(t_2 - t_1)} \quad \text{per } i = 1,2 \text{ e } j = 1,2.$$

Il predittore ottimo ad R passi con $R > 1$ ha il compito di stimare la variabile casuale $x(N+R)$ a partire dai dati fino a N e cioè dalle osservazioni di $y(N), y(N-1), \dots$. Quindi il predittore ottimo $\hat{x}(N+R|N)$ sarà:

$$\underline{\hat{x}(N+R|N) = E[\, x(N+R) \,|\, y^N]}$$

Ricordandosi che:

$$x(N+R) = Fx(N+R-1) + v_1(N+R-1)$$

e che per ogni $R > 1$, il campione $v_1(N+R-1)$ è incorrelato con i dati fino ad N, si ha:

$$\underline{\hat{x}(N+R|N) = F\hat{x}(N+R-1|N)}$$

Iterando si avrà:

$$\underline{\hat{y}(N+R|N) = F^{R-1}\hat{x}(N+R|N)}$$

Per quanto riguarda l'uscita si ha che per $R > 1$, $v_2(N+R)$ risulta incorrelato con i dati fino a N. Quindi la predizione ottima sarà:

$$\underline{\hat{x}(N+R|N) = H\hat{x}(N+R|N)}$$

Quindi il predittore ottimo a più passi dello stato si ricava a partire da quello ad un passo. Facciamo ora un piccolo passo all'indietro. Precedentemente si è usata la varianza dell'errore di stima dello stato $\nu(N)$. Tale varianza è una matrice $n \times n$.

Tale matrice l'abbiamo indicata con $P(N)$. Poiché: $v(N+1) = x(N+1) - \hat{x}(N+1|N)$ e sapendo che $x(t+1) = Fx(t) + V_1(t)$ ed $\hat{x}(N+1|N) = F\hat{x}(N|N-1) + K(N)e(N)$ si ha:

$$v(N+1) = [Fx(N) + V_1(N) - F\hat{x}(N|N-1) + K(N)e(N)]$$

$$\Downarrow$$

$$v(N+1) = [Fv(N) + V_1(N) + K(N)e(N)]$$

Siccome $e(N) = y(N) - \hat{y}(N|N-1)$ si ha:

$$v(N+1) = [F v(N) + V_1(N) + K(N)(y(N) - \hat{y}(N+1|N))] =$$

$$= [Fv(N) + V_1(N) + K(N)(Hx(N) + V_2(N) - H\hat{x}(N|N-1))] =$$

$$= [Fv(N) + V_1(N) + K(N)(Hv(N) + V_2(N))] = \underline{[F - K(N)H]v(N) + V_1(N) - K(N)V_2(N)}$$

Quindi da quest'ultima relazione si può calcolare la matrice varianza dell'errore di predizione dello stato in questo modo:

$$P(N+1) = Var[v(N+1)] = E[v(N+1)v(N+1)^T] = E[(F - K(N)H)v(N)v(N)^T(F - K(N)H)^T] +$$

$$+ E[V_1(N)V_1(N)^T] + E[K(N)V_2(N)V_2(N)^T K(N)^T] + E[(F - K(N)H)v(N)V_1(N)^T] +$$

$$- E[(F - K(N)H)v(N)V_2(N)^T]K(N)^T - E[V_1(N)V_2(N)^T]K(N)^T + E[V_1(N)v(N)^T$$

$$(F - K(N)H)^T] - E[K(N)V_2(N)v(N)^T(F - K(N)H)^T]$$

$$\Downarrow$$

$$P(N+1) = (F - K(N)H)P(N)(F - K(N)H)^T + V_1 + K(N)V_2K(N)^T$$

Sviluppando il prodotto del primo addendo si ha: $P(N+1) = FP(N)F^T - K(N)HP(N)F^T - FP(N)H^TK(N)^T + K(N)HP(N)H^TK(N)^T + V_1 + K(N)V_2K(N)^T$.

Quest'ultima espressione può essere così riscritta:

$$\underline{K(N)HP(N)H^TK(N)^T + K(N)V_2K(N)^T = K(N)(HP(N)H^T + V_2)K(N)^T}$$

Siccome però $K(N) = FP(N)H^T[HP(N)H^T + V_2]^{-1}$ si ha:

$$K(N)HP(N)H^TK(N)^T + K(N)V_2(K(N))^T = FP(N)H^T\cancel{[HP(N)H^T + V_2]^{-1}(HP(N)H^T + V_2)}K(N)^T$$

$$\Downarrow$$

$$\underline{K(N)HP(N)H^TK(N)^T + K(N)V_2(K(N))^T = FP(N)H^TK(N)^T}$$

Perciò:

$$P(N+1) = FP(N)F^T - K(N)HP(N)F^T + V_1$$

Ma $K(N) = FP(N)H^T[HP(N)H^T + V_2] \Rightarrow P(N+1) = FP(N)F^T + V_1 - FP(N)H^T(HP(N)H^T + V_2)^{-1}HP(N)F^T$

$$\Downarrow$$

$$\underline{P(N+1) = FP(N)F^T + V_1 - K(N)(HP(N)H^T + V_2)K(N)^T}$$

Quest'ultima equazione prende il nome di EQUAZIONE DI RICCATI ALLE DIFFERENZE (<u>DRE</u>).