UNIVERSITÀ DEGLI STUDI DI BERGAMO
Dipartimento di Ingegneria Gestionale, dell'Informazione e della Produzione
Control Automation Lab

IDENTIFICAZIONE DEI MODELLI E ANALISI DEI DATI (IMAD)

Corso di Laurea Magistrale in INGEGNERIA INFORMATICA

Lezione 10: Predizione

SPEAKER
Prof. Mirko Mazzoleni

PLACE
Università degli Studi di Bergamo

Syllabus

Parte II: sistemi dinamici

8. **Processi stocastici**

    8.1 Processi stocastici stazionari (pss)

    8.3 Rappresentazione spettrale di un pss

    8.4 Stimatori campionari media\covarianza

    8.5 Densità spettrale campionaria

9. **Famiglie di modelli a spettro razionale**

    9.1 Modelli per serie temporali (MA, AR, ARMA)

    9.2 Modelli per sistemi input/output (ARX, ARMAX)

10. **Predizione**

    10.1 Filtro passa-tutto

    10.2 Forma canonica

    10.3 Teorema della fattorizzazione spettrale

    10.4 Soluzione al problema della predizione

11. **Identificazione**

    11.3 Identificazione di modelli ARX

    11.4 Identificazione di modelli ARMAX

    11.5 Metodo di Newton

12. **Identificazione: analisi e complementi**

    12.1 Analisi asintotica metodi PEM

    12.2 Identificabilità dei modelli

    12.3 Valutazione dell'incertezza di stima

13. **Identificazione: valutazione**

IMAD

Parte I: sistemi statici

Stima parametrica $\hat{\theta}$

- ***θ* deterministico**
  - ***NO assunzioni su ddp dei dati***
    - ✓ Stima parametri popolazione
    - ✓ Stima modello lineare: minimi quadrati
  - ***SI assunzioni su ddp dei dati***
    - ✓ Stima massima verosimiglianza parametri popolazione
    - ✓ Stima modello lineare: massima verosimiglianza
    - ✓ Regressione logistica
- ***θ* variabile casuale**
  - ***SI assunzioni su ddp dei dati***
    - ✓ Stima Bayesiana

Machine learning

Parte II: sistemi dinamici

Stima parametrica $\hat{\theta}$

- ***θ* deterministico**
  - ***NO assunzioni su ddp dei dati***
    - ✓ Modelli lineari di pss
    - ✓ Predizione
    - ✓ Identificazione
    - ✓ Persistente eccitazione
    - ✓ Analisi asintotica metodi PEM
    - ✓ Analisi incertezza stima (numero dati finito)
    - ✓ Valutazione del modello

UNIVERSITÀ DEGLI STUDI DI BERGAMO
Dipartimento di Ingegneria Gestionale, dell'Informazione e della Produzione

Outline

1. Predizione, filtraggio e smoothing
2. Scomposizione di Wold
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. Predittore ottimo per processi MA
6. Predittore ottimo per processi ARMA
7. Predittore ottimo per processi ARMAX
8. Predittore ottimo ad un passo per sistemi ingresso\uscita
9. Confronto con il predittore di Kalman

Outline

1. **Predizione, filtraggio e smoothing**
2. Scomposizione di Wold
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. Predittore ottimo per processi MA
6. Predittore ottimo per processi ARMA
7. Predittore ottimo per processi ARMAX
8. Predittore ottimo ad un passo per sistemi ingresso\uscita
9. Confronto con il predittore di Kalman

Gli step per la risoluzione del problema

Seguiremo tre fasi per risolvere il problema della **modellazione di sistemi dinamici**:

Definizione delle **classi di modelli** $\mathcal{M}$ di sistemi dinamici

Ci concentreremo su modelli di **sistemi dinamici lineari**, espressi da **funzioni di trasferimento razionali fratte**. I parametri ignoti sono i coefficienti dei polinomi al numeratore e denominatore

Predizione

Data una particolare classe di modello, supponendo di conoscerne il valore dei parametri, qual è il **predittore ottimo**? Quanto vale la predizione ottima?

Identificazione

Come **stimo il valore dei parametri** del modello scelto per la modellazione dei dati?

Predizione, filtraggio e smoothing

Siano $y(\cdot)$ e $x(\cdot)$ due processi stocastici stazionari con $y(\cdot)$ osservabile. Un problema interessante è quello di **ottenere una stima** di $x(t)$ nei seguenti casi:

- $y(t) = x(t)$ **Misuro il processo** $x(t)$ che mi interessa stimare
- $y(t) = x(t) + e(t),\ e(t) \sim \mathrm{WN}(0, \lambda^2)$ **Misuro una «versione rumorosa»** di $x(t)$

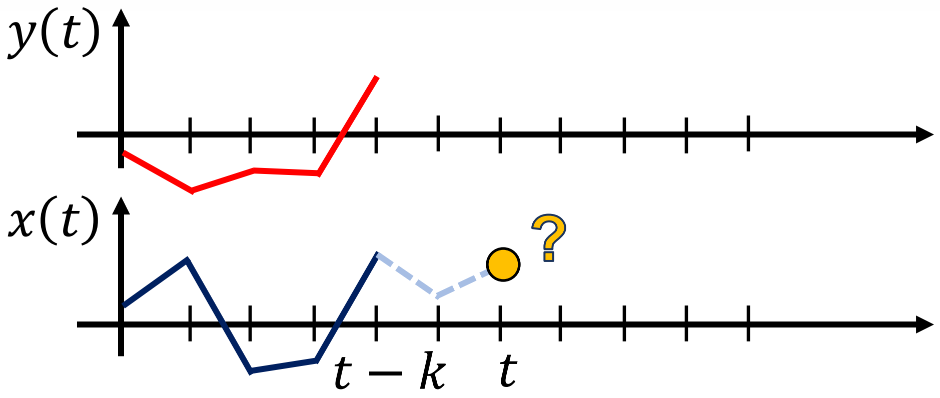
Vogliamo ottenere una stima $x(t|t_{\text{info}})$, basata sulla conoscenza di $y(s)$, per valori di:

- $s < t_{\text{info}}$ **predizione**
- $s = t_{\text{info}}$ **filtraggio**
- $s > t_{\text{info}}$ **smoothing**

Predizione a k passi

- $\hat{x}(t|t-k)$

Obiettivo: stimare il valore di $x(t)$ **a istanti futuri**

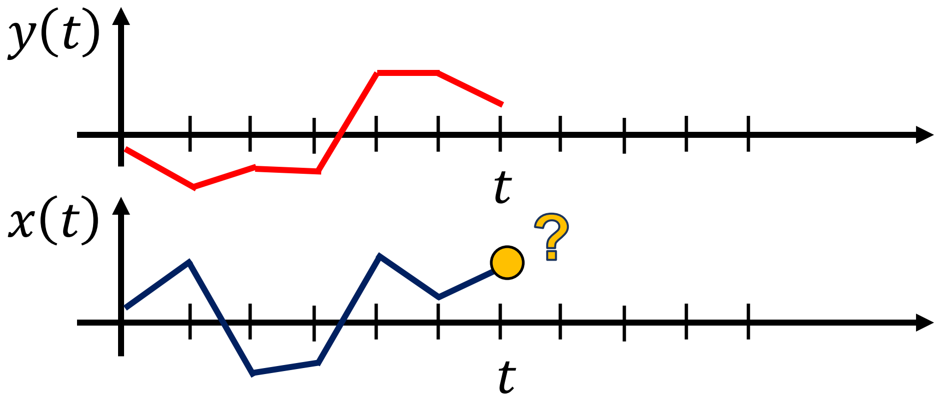


Filtraggio

ha senso solo se $x(t) \neq y(t)$

- $\hat{x}(t|t)$

Obiettivo: ottenere una stima di $x(t)$ **all'istante corrente** e «pulita dal rumore» (es. filtro di Kalman)

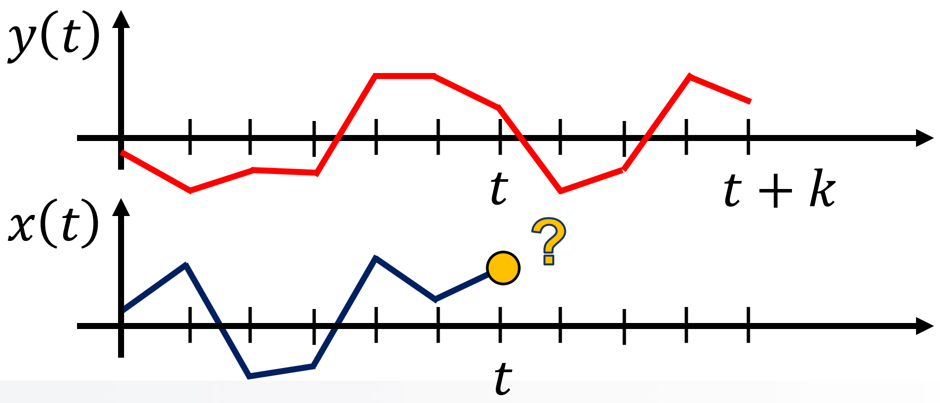


Smoothing

ha senso solo se $x(t) \neq y(t)$

- $\hat{x}(t|t+k)$

Obiettivo: ottenere una stima di $x(t)$ **all'istante passato** (es. ricostruzione traiettorie missili NASA)



Predizione ottima

Noi studieremo il **problema della predizione** per $x(t) = y(t)$, ovvero ci interesserà trovare una stima $\hat{y}(t|t-k)$ di $y(t)$ al tempo t, avendo a disposizione i dati fino al tempo $t-k$

Dato che il **predittore** $\hat{y}(t|t-k)$ si basa su valori passati di $y(t)$, sarà anch'esso un processo stocastico. L'**errore di predizione** $\varepsilon_k(t)$ è un processo stocastico definito come

$$\boxed{\varepsilon_k(t) = y(t) - \hat{y}(t|t-k)}$$

Vogliamo predittori (stimatori) **lineari ottimi**, i.e. con errore di predizione a **MSE minimo**

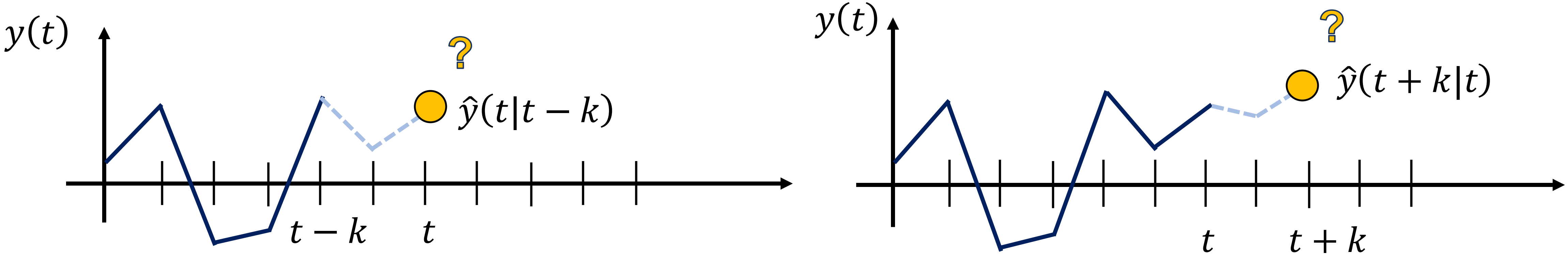
Approccio predittivo: un modello è **buono** se è in grado di **predire bene** i dati.
Più avanti, stimeremo i parametri di un modello dinamico *minimizzando la varianza dell'errore di predizione*

Predizione, filtraggio e smoothing

Note

- Studieremo il predittore ottimo per pss a spettro razionale delle famiglie **ARMA** e **ARMAX**. I predittori ottimi per le altre famiglie FIR, OE, BJ non verranno investigati

- Dato che lavoriamo con pss, le scritture $\hat{y}(t|t-k)$ e $\hat{y}(t+k|t)$ sono **equivalenti**, nel senso che la **forma del predittore ottimo è la stessa** (le predizioni potranno essere diverse)



Outline

1. Predizione, filtraggio e smoothing
2. **Scomposizione di Wold**
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. Predittore ottimo per processi MA
6. Predittore ottimo per processi ARMA
7. Predittore ottimo per processi ARMAX
8. Predittore ottimo ad un passo per sistemi ingresso\uscita
9. Confronto con il predittore di Kalman

Processi stocastici completamente predicibili

Definizione: Un processo stazionario $y(t)$ si dice **completamente predicibile** se esistono coefficienti $a_i, i = 1,2, \dots$ tali che

$$\boxed{y(t) = \sum_{i=1}^{+\infty} a_i y(t-i)}$$

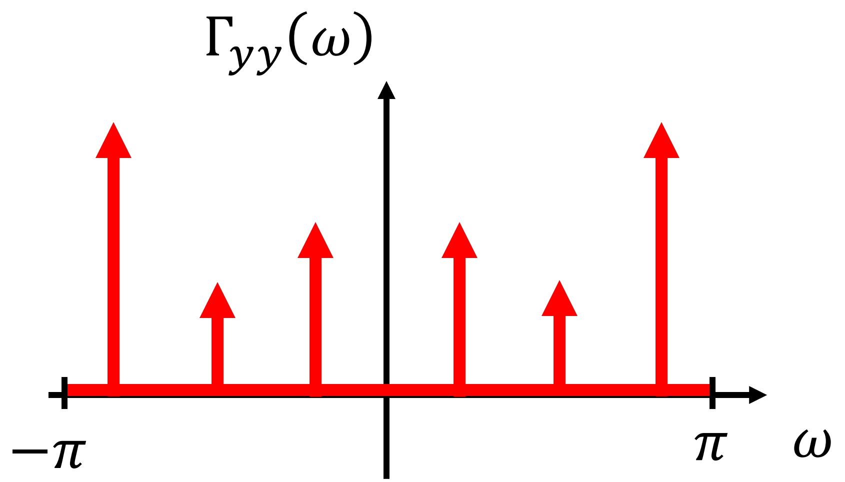
Se conosco i coefficienti a_i posso **prevedere senza errore** i valori futuri di $y(\quad)$ senza errori, a **partire dai valori passati**

Tali processi sono «l'opposto» del **rumore bianco**, che è **completamente impredicibile**

Processi stocastici completamente predicibili

Proprietà: Un processo stazionario $y(t)$ è completamente predicibile **se e solo se**

$$\boxed{\Gamma_{yy}(\omega) = \sum_i \alpha_i \delta(\omega - \omega_i)}$$



La densità spettrale di potenza di un **processo completamente predicibile** è una combinazione lineare di **delta di Dirac**

Il **rumore bianco**, in contrasto, ha una densità spettrale di potenza **costante**

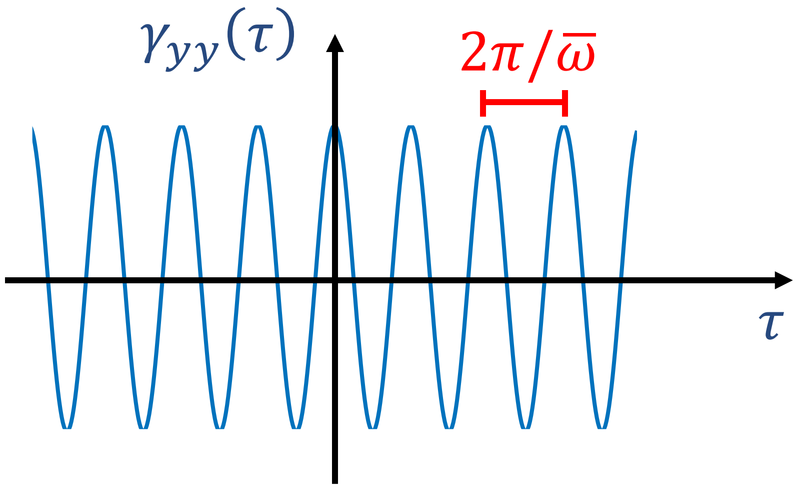
Processi stocastici completamente predicibili

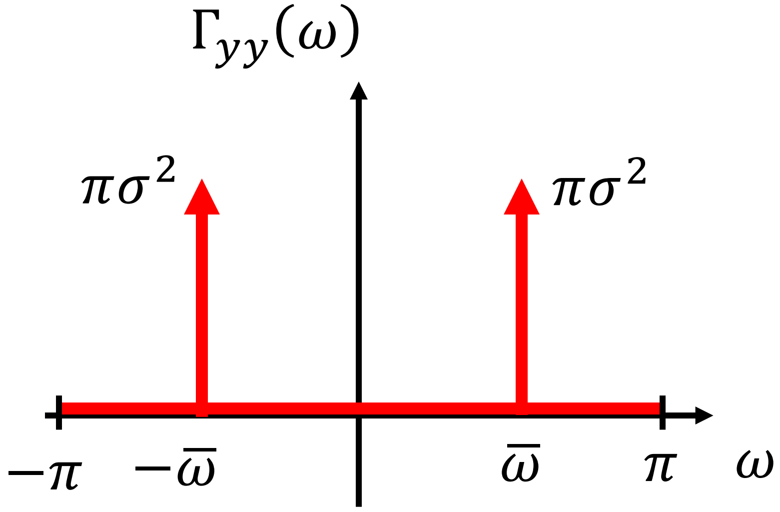
Esempio: Consideriamo il seguente processo stocastico

$$y(t) = v_1 \cos(\bar{\omega} t) + v_2 \sin(\bar{\omega} t)$$

Dove v_1 e v_2 sono variabili casuali **incorrelate** con $\mathbb{E}[v_1] = \mathbb{E}[v_2]$, e $\mathrm{Var}[v_1] = \mathrm{Var}[v_2] = \sigma^2$

Si verifica che questo è un processo stazionario (non ergodico) con $\gamma_{yy}(\tau)$ **cosinusoidale**





Una volta che ho capito l'andamento sinusoidale del processo, **sono in grado di prevederlo** da lì all'infinito

Processi stocastici completamente predicibili

Osservazione

In generale, quando $\Gamma_{yy}(\omega)$ è «a righe», il processo $y(t)$ è una **combinazione lineare di seni e coseni** (che è perfettamente predicibile dai valori passati)

In altre parole, le realizzazioni del processo sono **periodiche**

Scomposizione di Wold

Ogni **processo stocastico stazionario** $y(t)$ può essere scritto come

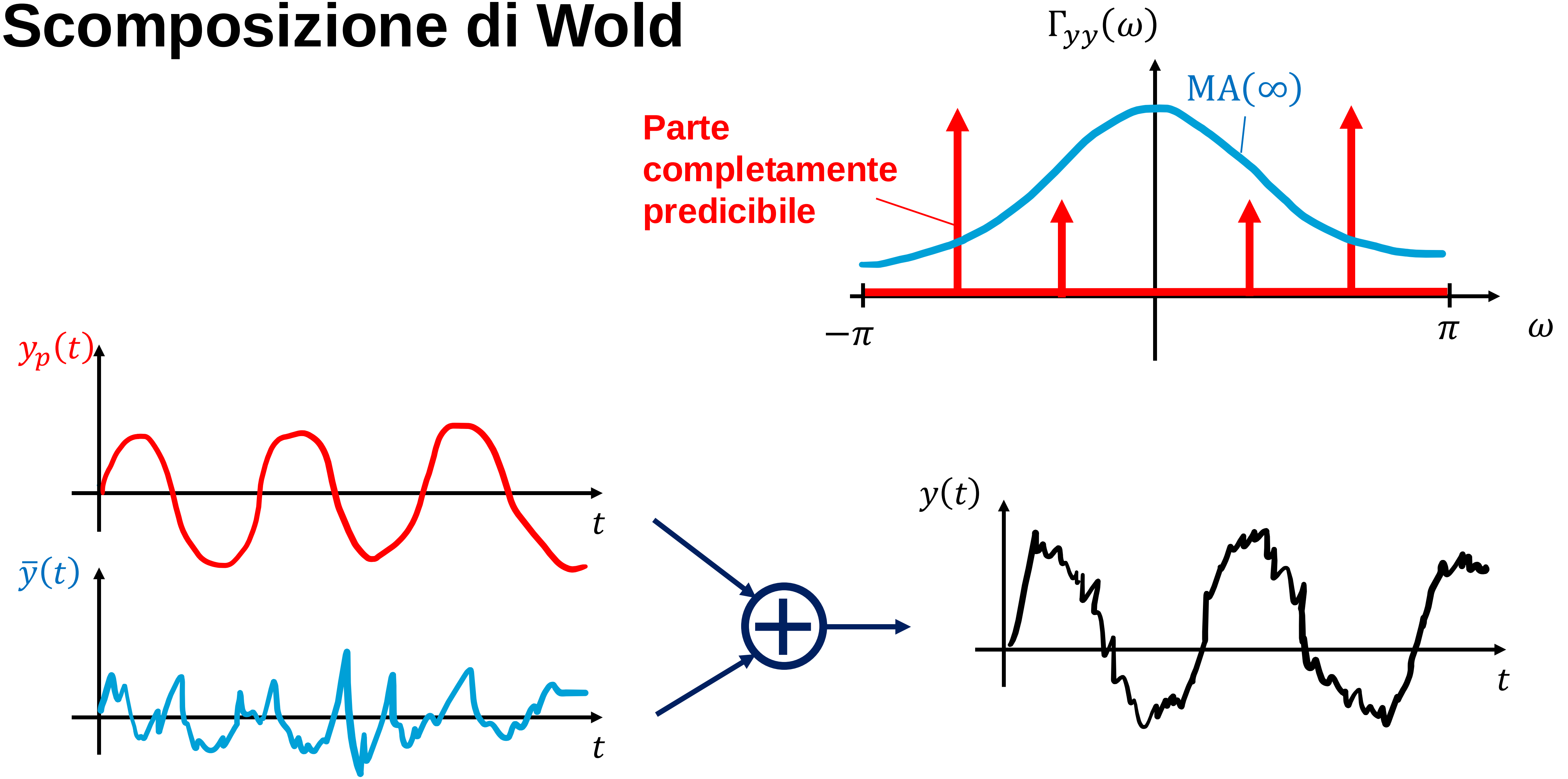
$$\boxed{y(t) = \bar{y}(t) + y_p(t)}$$

- $y_p(t)$: processo stocastico stazionario **completamente predicibile**

- $\bar{y}(t)$: parte **puramente stocastica**, tale che $\bar{y}(t) = \sum_{i=0}^{+\infty} c_i e(t-i)$, con $e(t) \sim \mathrm{WN}(0, \lambda^2)$, $\sum_{i=0}^{+\infty} c_i^2 < \infty$

- $\bar{y}(t)$ e $y_p(t)$ sono **incorrelati**

Scomposizione di Wold



Scomposizione di Wold nella pratica

Nella pratica, questo risultato ci fornisce una «linea guida» per **stimare modelli di serie temporali**

- Faccio una **stima spettrale** (es. periodogramma) per riconoscere eventuali «righe»
- **Stimo le componenti sinusoidali** $y_p(t)$ (minimi quadrati o «a mano»), ottenendo $\hat{y}_p(t)$
- Ottengo la **componente puramente stocastica** come $\bar{y}(t) = y(t) - \hat{y}_p(t)$
- Risolvo il problema della **predizione** per la parte stocastica $\bar{y}(t)$, ottenendo $\hat{\bar{y}}(t|t-k)$
- **Ottengo la predizione finale** come $\hat{y}(t|t-k) = \hat{y}_p(t) + \hat{\bar{y}}(t|t-k)$

Scomposizione di Wold nella pratica

Osservazioni

Anche eventuali **componenti di non stazionarietà** come **stagionalità** o **trend** devono essere stimate e rimosse dai dati per ottenere solo la parte stocastica del processo

- Esistono modelli di serie temporali più complessi (e.g. SARIMA) che cercano di **modellare la stagionalità**, anziché stimarla prima per poi sottrarla dai dati

- Un **trend** può anche essere un valore costante, e.g. il **valore atteso del processo.** Tale valore può essere visto come la **«componente a frequenza zero»**, che viene rimossa dal processo con la procedura vista precedentemente (per esempio stimando il valore atteso con una media temporale)

Scomposizione di Wold nella pratica

Ipotesi di lavoro

- Facciamo **l'ipotesi** che la parte puramente stocastica $\bar{y}(t)$, ovvero il processo $\mathrm{MA}(\infty)$, possa essere ben approssimato da **processi a spettro razionale**. Abbiamo già visto che ciò è possibile usando un ARMA

- L'ipotesi non è restrittiva anche perché i coefficienti c_i diventano più piccoli col tempo, e quindi potrei usare anche un $\mathrm{MA}(n_c)$

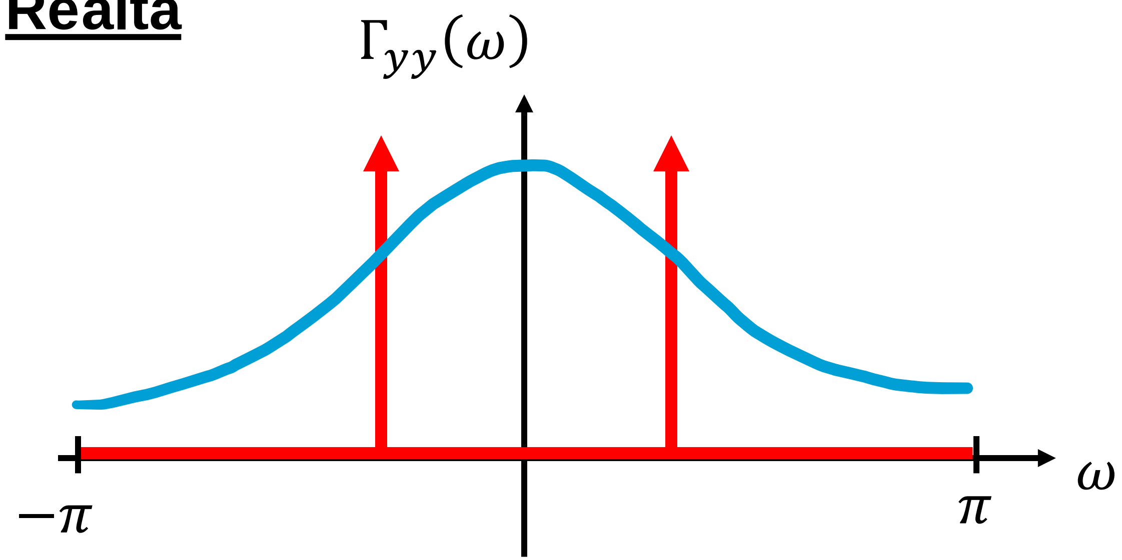
Nel seguito, **supporremo di lavorare con processi depurati dalle componenti non stazionarie e completamente predicibili** (da cui deriva che avranno media nulla)

Scomposizione di Wold nella pratica

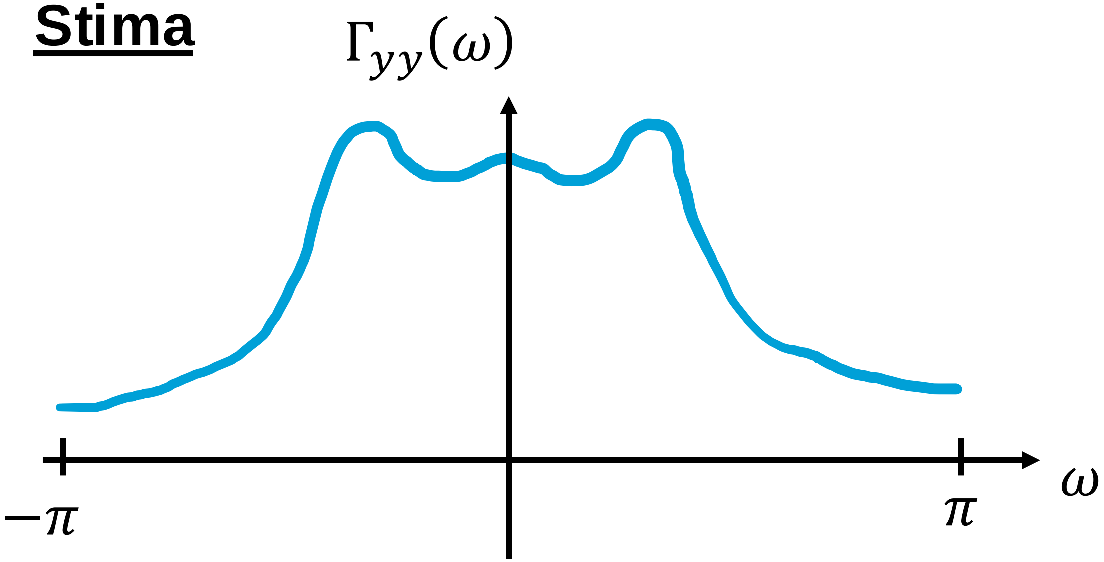
«Estrarre le righe» dal periodogramma **non è così facile**, per due motivi:

1. gli **stimatori** basati sul periodogramma **non sono molto buoni**
2. **«risonanze»** nella densità spettrale di potenza potrebbero essere dovute non solo alla presenza di delta di Dirac stimate male, ma anche a **poli**

Realtà



Stima



Outline

Filtro passa-tutto

Il **filtro passa-tutto** è un filtro di ordine 1 definito come

$$T(z) = \frac{1}{a} \cdot \frac{z + a}{z + \frac{1}{a}}, \qquad a \neq 0, a \in \mathbb{R}$$

- Lo zero è il reciproco del polo
- Il fattore moltiplicativo è come il polo

Proviamo a calcolare la **densità spettrale di potenza** $\Gamma_{yy}(\omega)$ di un processo $y(t)$ in uscita dal passa-tutto $T(z)$, alimentato da un **generico processo stazionario** in ingresso $v(t)$

$$\Gamma_{yy}(\omega) = \left|T\left(e^{j\omega}\right)\right|^2 \cdot \Gamma_{vv}(\omega)$$

Filtro passa-tutto

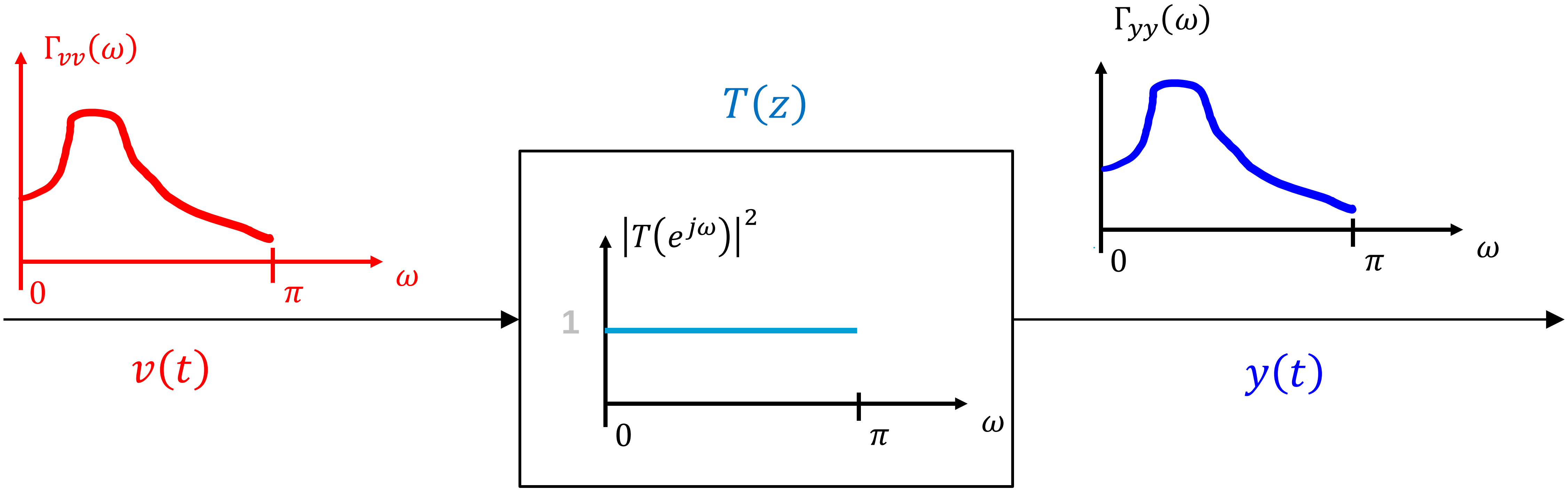
$$\left|T\left(e^{j\omega}\right)\right|^2 = \left(\frac{1}{a}\cdot\frac{e^{j\omega}+a}{e^{j\omega}+\frac{1}{a}}\right)\cdot\left(\frac{1}{a}\cdot\frac{e^{-j\omega}+a}{e^{-j\omega}+\frac{1}{a}}\right) = \frac{1}{a^2}\cdot\frac{\left(e^{j\omega}+a\right)\cdot\left(e^{-j\omega}+a\right)}{\left(e^{j\omega}+\frac{1}{a}\right)\cdot\left(e^{-j\omega}+\frac{1}{a}\right)}$$

$$= \frac{1}{a^2}\cdot\frac{1+a^2+a\left(e^{j\omega}+e^{-j\omega}\right)}{1+\frac{1}{a^2}+\frac{1}{a}\left(e^{j\omega}+e^{-j\omega}\right)} = \frac{1}{a^2}\cdot\frac{1+a^2+2a\cos\omega}{\frac{a^2+1+2a\cos\omega}{a^2}} \boxed{= 1}$$

Il filtro passa-tutto **non modifica il modulo delle frequenze** nella densità spettrale di potenza dell'ingresso. Quindi, si ha che

$$\Gamma_{yy}(\omega) = \Gamma_{vv}(\omega)$$

Filtro passa-tutto

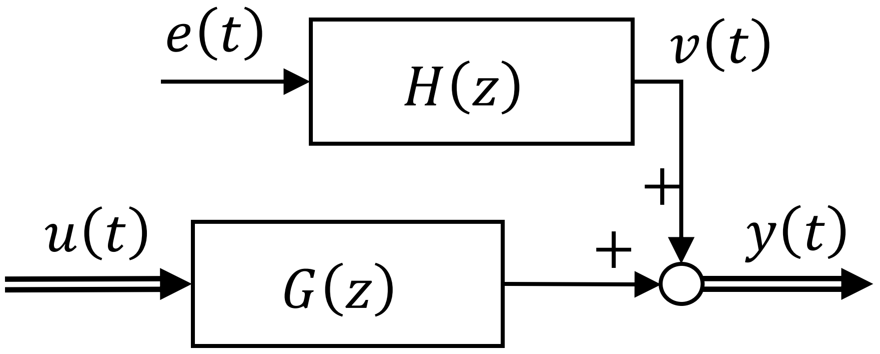


$$\Gamma_{yy}(\omega) = \Gamma_{vv}(\omega)$$

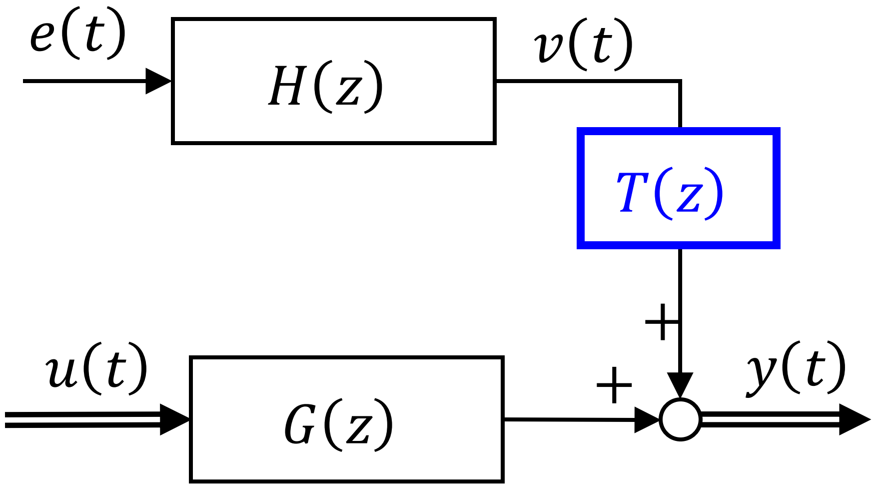
Filtro passa-tutto

Il processo $y(t)$ in uscita al passa-tutto è **spettralmente equivalente** al processo $v(t)$ in ingresso al passatutto

I due processi $y(t)$ e $v(t)$ **non sono identici** poichè il passa-tutto introduce uno **sfasamento** (come tutti i filtri causali)

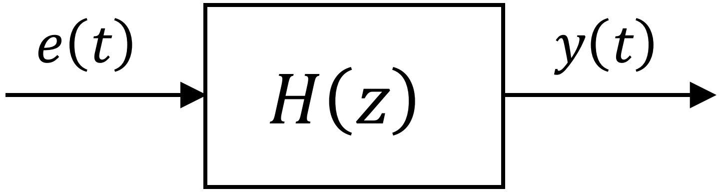


Spettralmente equivalente a



Fattorizzazione spettrale

Abbiamo detto che vogliamo risolvere il problema della predizione per **processi a spettro razionale**, ovvero processi $y(t)$ generati in uscita da un sistema dinamico lineare asintoticamente stabile con funzione di trasferimento $H(z)$ razionale fratta, alimentato da rumore bianco $e(t) \sim \mathrm{WN}(0, \lambda^2)$



$H(z) = C(z)/A(z)$ **fdt razionale fratta**

Il problema della **fattorizzazione spettrale** consiste nel trovare tutte le coppie $\{H(z), \lambda^2\}$ tali che

$$\Phi_{yy}(z) = \lambda^2 \cdot H(z)H(z^{-1})$$

Per **processi a spettro razionale**, esistono **infiniti fattori spettrali** $\{H(z), \lambda^2\}$. Ai fini della predizione ottima, ci servirà un fattore spettrale particolare, detto **canonico**

Forma canonica

Consideriamo questi 5 processi ARMA:

1) $y_1(t) = \dfrac{z+\frac{1}{2}}{z-\frac{1}{3}} e(t), \qquad e(t) \sim \mathrm{WN}(0,1)$

2) $y_2(t) = \dfrac{z+\frac{1}{2}}{z-\frac{1}{3}} e(t-2), \qquad e(t) \sim \mathrm{WN}(0,1)$

3) $y_3(t) = \dfrac{z^2-\frac{1}{4}}{z^2-\frac{5}{6}z+\frac{1}{6}} e(t), \qquad e(t) \sim \mathrm{WN}(0,1)$

4) $y_4(t) = \dfrac{2z+1}{z-\frac{1}{3}} e(t), \qquad e(t) \sim \mathrm{WN}\left(0,\frac{1}{4}\right)$

5) $y_5(t) = \dfrac{z+2}{z-\frac{1}{3}} e(t), \qquad e(t) \sim \mathrm{WN}\left(0,\frac{1}{4}\right)$

Forma canonica

Calcoliamo le densità spettrali di potenza dei processi $y_i(t)$

1) $$\Gamma_{y_1y_1}(\omega) = \left| \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}} \right|^2 \cdot 1$$

2) $$\Gamma_{y_2y_2}(\omega) = \left| \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}} \cdot e^{-2j\omega} \right|^2 \cdot 1 = \Gamma_{y_1y_1}(\omega) \cdot \left| e^{-2j\omega} \right|^2 = \Gamma_{y_1y_1}(\omega)$$

3) $$\Gamma_{y_3y_3}(\omega) = \left| \frac{e^{2j\omega} - \frac{1}{4}}{e^{2j\omega} - \frac{5}{6} e^{j\omega} + \frac{1}{6}} \right|^2 \cdot 1 = \left| \frac{\cancel{\left(e^{j\omega} - \frac{1}{2}\right)} \cdot \left(e^{j\omega} + \frac{1}{2}\right)}{\cancel{\left(e^{j\omega} - \frac{1}{2}\right)} \cdot \left(e^{j\omega} - \frac{1}{3}\right)} \right|^2 = \Gamma_{y_1y_1}(\omega)$$

Forma canonica

Calcoliamo le densità spettrali di potenza dei processi $y_i(t)$

4) $$\Gamma_{y_4 y_4}(\omega) = \left|2 \cdot \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}}\right|^2 \cdot \frac{1}{4} = 4 \cdot \left|\frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}}\right|^2 \cdot \frac{1}{4} = \Gamma_{y_1 y_1}(\omega)$$

5) $$\Gamma_{y_5 y_5}(\omega) = \left|\frac{z+2}{z - \frac{1}{3}}\right|^2_{z=e^{j\omega}} \cdot \frac{1}{4} = \left|\frac{\cancel{z+2}}{z - \frac{1}{3}} \cdot \boxed{2\frac{z + \frac{1}{2}}{\cancel{z+2}}}\right|^2_{z=e^{j\omega}} \cdot \frac{1}{4} = \left|2\frac{z + \frac{1}{2}}{z - \frac{1}{3}}\right|^2_{z=e^{j\omega}} \cdot \frac{1}{4}$$

$$= \left|2 \cdot \frac{e^{j\omega} + \frac{1}{2}}{e^{j\omega} - \frac{1}{3}}\right|^2 \cdot \frac{1}{4} = \Gamma_{y_1 y_1}(\omega)$$

Uso il **filtro passa-tutto** per cancellare lo **zero fuori dal cerchio**

Forma canonica

Tutti e 5 i processi visti sono **equivalenti** (hanno la stessa densità spettrale di potenza). Le **cause di non univocità** sono:

1. **Ritardi** puri (processo **2**)
2. **Fattori moltiplicativi** che si cancellano (processo **3**)
3. **Coefficienti moltiplicativi** che si compensano tra funzione di trasferimento e spettro dell'ingresso (processo **4**)
4. **Poli\zeri reciproci** (processo **5**)

La **forma canonica** di un processo stocastico stazionario a spettro razionale è univoca e ci permetterà di risolvere il problema della predizione

Teorema della fattorizzazione spettrale

Teorema Dato un processo stocastico stazionario a **spettro razionale**, esiste **un solo fattore spettrale** $\{\tilde{H}(z), \tilde{\lambda}^2\}$, detto **fattore spettrale canonico,** dove $\tilde{H}(z) = C(z)/A(z)$, tale che

1. $C(z)$ e $A(z)$ hanno lo **stesso grado** (grado relativo nullo)
2. $C(z)$ e $A(z)$ sono **coprimi** (non ci son fattori in comune)
3. $C(z)$ e $A(z)$ sono **monici** (il coefficiente del termine di grado massimo è 1)
4. $C(z)$ e $A(z)$ hanno **radici interne al cerchio unitario**

Esempio: calcolo della forma canonica

Consideriamo il seguente processo $\mathrm{ARMA}(1,1)$

$$y(t) = \frac{z+2}{z-\frac{1}{3}} e(t-2), \qquad e(t) \sim \mathrm{WN}(0,1)$$

$\eta(t) \sim \mathrm{WN}(0,4)$

$$y(t) = \frac{z+2}{z-\frac{1}{3}} \cdot 2\,\frac{z+\frac{1}{2}}{z+2} \cdot e(t-2) \quad = \frac{z+\frac{1}{2}}{z-\frac{1}{3}} \cdot \boxed{2e(t-2)}$$

$$\Rightarrow \quad \boxed{y(t) = \frac{1+\frac{1}{2}z^{-1}}{1-\frac{1}{3}z^{-1}}\eta(t), \qquad \eta(t) \sim \mathrm{WN}(0,4)}$$

Outline

Predittore ottimo

Predizione: stimare il dato al tempo t avendo a disposizione dati fino al tempo $t-k$. Equivalentemente, stimare il dato al tempo $t+k$ avendo a disposizione dati fino al tempo t.

Indichiamo il **predittore** come $\hat{y}(t|t-k)$ o $\hat{y}(t+k|t)$

Informazioni disponibili

- **Dati** $y(1), y(2), \dots, y(N)$
- Vecchie **predizioni** $\hat{y}(t-1|t-k-1), \hat{y}(t-2|t-k-2), \dots$
- **Modello** della parte stocastica del processo $C(z)/A(z)$

Ipotesi di lavoro

- Supponiamo $y(t)$ un pss **puramente stocastico**, depurato da componenti predicibili
- Modello $C(z)/A(z)$ **in forma canonica**

Predittore ottimo

Esistono diversi modi per definire un **predittore**, per esempio potremmo usare:

$$\hat{y}(t+1|t) = \frac{y(t) + y(t-1) + y(t-2)}{3}$$

Media di alcuni valori passati

$$\hat{y}(t+1|t) = \frac{2y(t) + \frac{1}{2}y(t-1) + \frac{1}{2}y(t-2)}{3}$$

Diamo più peso a valori più recenti

Vogliamo però trovare il **predittore lineare ottimo** dai dati, ovvero quello che minimizza il seguente criterio **Mean Squared Error (MSE)**

$$\boxed{\mathbb{E}[\varepsilon_k(t)^2] = \mathbb{E}\left[\left(y(t) - \hat{y}(t|t-k)\right)^2\right]}$$

Predittore ottimo

Definizione: Un predittore (lineare) è **ottimo** se

1. $\mathbb{E}[\varepsilon_k(t)] = \mathbb{E}[y(t) - \hat{y}(t|t-k)] = 0$, i.e. il **valore atteso** dell'errore di predizione è **nullo**

   Significa che il processo $y(t)$ e il predittore $\hat{y}(t|t-k)$ hanno lo **stesso valore atteso**

2. $\mathbb{E}[\hat{y}(t|t-k) \cdot \varepsilon_k(t)] = 0$, i.e. il predittore e l'errore di predizione sono **incorrelati**

   Significa che **il predittore ha utilizzato tutta l'informazione disponibile**. Se $\hat{y}(t|t-k)$ e $\varepsilon_k(t)$ fossero correlati, significa che «c'è qualcosa» in $\varepsilon_k(t)$ che c'è anche in $\hat{y}(t|t-k)$. Ma allora questo qualcosa avrebbe dovuto stare in $\hat{y}(t|t-k)$ per «aiutarlo» nella previsione

3. $\mathrm{Var}[\varepsilon_k(t)]$ **minima**

Predittore ottimo

Avendo definito l'**errore di predizione** come

$$\varepsilon_k(t) = y(t) - \hat{y}(t|t-k)$$

possiamo **scomporre il processo** $y(t)$ come

$$\boxed{y(t) = \hat{y}(t|t-k) + \varepsilon_k(t)}$$

dove:

- $\hat{y}(t|t-k)$ è la **parte predicibile** al tempo $t-k$
- $\varepsilon_k(t)$ è la **parte impredicibile** al tempo $t-k$

Outline

1. Predizione, filtraggio e smoothing
2. Scomposizione di Wold
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. **Predittore ottimo per processi MA**
6. Predittore ottimo per processi ARMA
7. Predittore ottimo per processi ARMAX
8. Predittore ottimo ad un passo per sistemi ingresso\uscita
9. Confronto con il predittore di Kalman

Predittore ottimo per processi MA

PREDITTORE AD UN PASSO

Consideriamo un processo $\mathrm{MA}(n_c)$ in **forma canonica**

$$y(t) = \underbrace{e(t)}_{\text{Parte impredicibile al tempo } t-1} + \underbrace{c_1 e(t-1) + c_2 e(t-2) + \cdots + c_{n_c} e(t-n_c)}_{\text{Parte predicibile al tempo } t-1}, \qquad e(t) \sim \mathrm{WN}(0, \lambda^2)$$

Un possibile predittore potrebbe quindi essere dato dalla parte predicibile al tempo $t-1$

$$\hat{y}(t|t-1) = c_1 e(t-1) + c_2 e(t-2) + \cdots + c_{n_c} e(t-n_c)$$

Predittore ottimo per processi MA

$$\hat{y}(t|t-1) = c_1 e(t-1) + c_2 e(t-2) + \cdots + c_{n_c} e(t-n_c)$$

Osservazioni

- $\hat{y}(t|t-1)$ è corretto, infatti $\mathbb{E}[y(t)] = \mathbb{E}[\hat{y}(t|t-1)] = 0$
- $\hat{y}(t|t-1)$ dipende dal WN fino al tempo $t-1$
- $\mathbb{E}[\hat{y}(t|t-1) \cdot \varepsilon_1(t)] = 0$, infatti $\varepsilon_1(t) = y(t) - \hat{y}(t|t-1) = e(t)$ e quindi

$$\mathbb{E}[\hat{y}(t|t-1) \cdot \varepsilon_1(t)] = \mathbb{E}\left[c_1 e(t-1) + c_2 e(t-2) + \cdots + c_{n_c} e(t-n_c) \cdot e(t)\right] = 0$$

- Non è possibile trovare un predittore con $\mathrm{Var}[\varepsilon_1(t)] < \mathrm{Var}[e(t)]$

Ne consegue che $\hat{y}(t|t-1)$ è il **predittore lineare ottimo**

Predittore ottimo per processi MA

Tuttavia, l'espressione di $\hat{y}(t|t-1)$ **dipende dal rumore** $e(t)$, e non **dai dati** $y(t)$. Troviamo il «predittore dai dati» osservando che

$$y(t) = \left[1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}\right] \cdot e(t) \;\Rightarrow\; e(t) = \underbrace{\frac{1}{1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}}}_{\text{FILTRO SBIANCANTE}} y(t)$$

FILTRO SBIANCANTE
(è stabile grazie alla forma canonica)

$$\hat{y}(t|t-1) = c_1 e(t-1) + c_2 e(t-2) + \cdots + c_{n_c} e(t-n_c)$$

$$= \left[c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}\right] \cdot e(t)$$

$$\Rightarrow \quad \boxed{\hat{y}(t|t-1) = \frac{c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}}{1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}} y(t)}$$

Predittore ottimo dai dati ad un passo per processi $\mathrm{MA}(n_c)$

Predittore ottimo per processi MA

Passando in **forma ricorsiva** si ottiene

$$\hat{y}(t|t-1)\left[1+c_1 z^{-1}+\cdots+c_{n_c} z^{-n_c}\right]=\left[c_1 z^{-1}+c_2 z^{-2}+\cdots+c_{n_c} z^{-n_c}\right]y(t)$$

$$\hat{y}(t|t-1) = -c_1\hat{y}(t-1|t-2)-\cdots-c_{n_c}\hat{y}(t-n_c|t-1-n_c)+$$ **Predizioni passate**

$$+c_1 y(t-1)+\cdots+c_{n_c} y(t-n_c)$$ **Dati passati**

Osservazione

Quando il processo ha una componente MA, il predittore **è dinamico**. C'è bisogno di definire il valore della **condizione iniziale** $\hat{y}(1|0)$. Di solito si usa la media del processo (i.e. zero). Se il predittore è asintoticamente stabile, **l'effetto dell'inizializzazione svanisce** col tempo

Predittore ottimo per processi MA

PREDITTORE A k PASSI

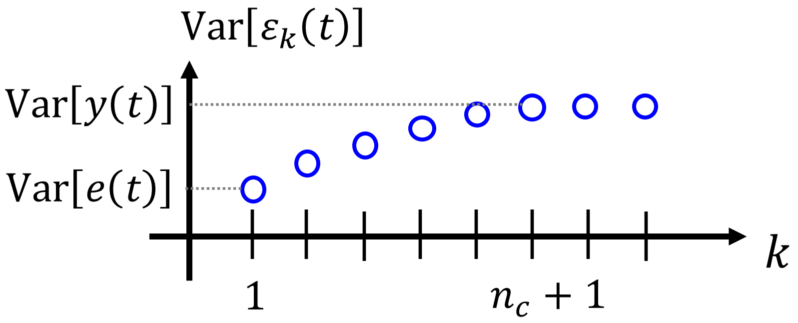
Consideriamo un processo $\mathrm{MA}(n_c)$ in **forma canonica**, con $e(t) \sim \mathrm{WN}(0, \lambda^2)$

$$y(t) = \underbrace{e(t) + c_1 e(t-1) + \cdots + c_{k-1} e(t-k+1)}_{\textbf{Parte impredicibile} \text{ al tempo } t-k} + \underbrace{c_k e(t-k) + \cdots + c_{n_c} e(t-n_c)}_{\textbf{Parte predicibile} \text{ al tempo } t-k}$$

Si dimostra che il **predittore ottimo dal rumore** è dato dalla **parte predicibile**, ovvero

$$\boxed{\hat{y}(t|t-k) = c_k e(t-k) + \cdots + c_{n_c} e(t-n_c)}$$

Predittore ottimo per processi MA



Osservazione

$$\varepsilon_1(t) = y(t) - \hat{y}(t|t-1) \quad \Rightarrow \quad \mathrm{Var}[\varepsilon_1(t)] = \mathrm{Var}[e(t)] = \lambda^2$$

$$\varepsilon_2(t) = y(t) - \hat{y}(t|t-2) \quad \Rightarrow \quad \mathrm{Var}[\varepsilon_2(t)] = \mathrm{Var}[e(t) + c_1 e(t-1)] = (1 + c_1^2)\lambda^2 > \lambda^2$$

$$\vdots$$

$$\varepsilon_{n_c+1}(t) = y(t) - \hat{y}(t|t-n_c-1) \quad \Rightarrow \quad \mathrm{Var}\left[\varepsilon_{n_c+1}(t)\right] = \mathrm{Var}\left[e(t) + c_1 e(t-1) + \cdots + c_{n_c} e(t-n_c)\right]$$

$$= \mathrm{Var}[y(t)]$$

La **varianza** di $\varepsilon_k(t)$ **aumenta con l'orizzonte di predizione**, fino a diventare uguale alla varianza del processo $y(t)$. Il predittore $\hat{y}(t|t-n_c-1)$ sarà il **predittore banale**, di solito la media del processo

Outline

1. Predizione, filtraggio e smoothing
2. Scomposizione di Wold
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. Predittore ottimo per processi MA
6. **Predittore ottimo per processi ARMA**
7. Predittore ottimo per processi ARMAX
8. Predittore ottimo ad un passo per sistemi ingresso\uscita
9. Confronto con il predittore di Kalman

Predittore ottimo per processi ARMA

Sia dato un processo $\mathrm{ARMA}(n_a, n_c)$ in **forma canonica**

$$y(t) = \frac{C(z)}{A(z)} e(t) \qquad\qquad e(t) \sim WN(0, \lambda^2)$$

- $C(z) = 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}$
- $A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_{n_a} z^{-n_a}$

Problema: In questo caso, **non è immediatamente chiaro come scomporre** la parte impredicibile da quella predicibile, poiché $y(t)$ dipende anche da $y(t-1), y(t-2), \ldots$ i quali dipendono da $e(t), e(t-1), \ldots$

Predittore ottimo per processi ARMA

Idea: si esprime $C(z)/A(z)$ come un **quoziente** $E(z)$ più un **resto** $R(z) = z^{-k}\tilde{R}(z)$ effettuando una **lunga divisione** tra polinomi

Quoziente **Resto**

$$C(z) = E(z)\,A(z) + R(z) \quad \Rightarrow \quad \frac{C(z)}{A(z)} = E(z) + \frac{R(z)}{A(z)} = E(z) + \frac{z^{-k}\tilde{R}(z)}{A(z)}$$

Didivendo **Divisore**

Sostituendo l'espressione di $C(z)$ in $y(t) = \frac{C(z)}{A(z)}e(t)$ otteniamo

$$y(t) = \underbrace{E(z)e(t)}_{\text{Parte impredicibile al tempo } t-k} + \underbrace{\frac{\tilde{R}(z)}{A(z)}e(t-k)}_{\text{Parte predicibile al tempo } t-k}$$

Parte impredicibile al tempo $t-k$

Parte predicibile al tempo $t-k$

Esempio: lunga divisione

Consideriamo il processo $\mathrm{ARMA}(n_a = 1, n_c = 1)$ e facciamo $k = 2$ passi di lunga divisione

$$y(t) = \frac{1 + \frac{1}{2} z^{-1}}{1 + \frac{1}{3} z^{-1}} e(t) \qquad e(t) \sim \mathrm{WN}(0, \lambda^2)$$

$$\begin{array}{r|l}
C(z):\ \boxed{1 + \frac{1}{2} z^{-1}} & \boxed{1 + \frac{1}{3} z^{-1}}\ A(z) \\
\cline{2-2}
-1 - \frac{1}{3} z^{-1} & \boxed{1 + \frac{1}{6} z^{-1}}\ E(z) \\
\hline
\frac{1}{6} z^{-1} & \\
-\frac{1}{6} z^{-1} - \frac{1}{18} z^{-2} & \\
\hline
-\frac{1}{18} z^{-2} &
\end{array}$$

$$R(z) = z^{-k} \tilde{R}(z) = z^{-2} \left(-\frac{1}{18} \right)$$

- $E(z)e(t) = e(t) + \frac{1}{6} e(t-1)$ è **impredicibile** al tempo $t-2$

- $\frac{\tilde{R}(z)}{A(z)} e(t-k) = -\frac{\frac{1}{18}}{A(z)} e(t-2)$ è **predicibile** al tempo $t-2$

Predittore ottimo per processi ARMA

Il **predittore ottimo dal rumore** è

$$\boxed{\hat{y}(t|t-k) = \frac{\tilde{R}(z)}{A(z)} e(t-k)}$$

Calcoliamo il **predittore ottimo dai dati** tramite il filtro sbiancante

$$y(t) = \frac{C(z)}{A(z)} e(t) \quad \Rightarrow \quad \underbrace{e(t) = \frac{A(z)}{C(z)} y(t)}$$

FILTRO SBIANCANTE

$$\hat{y}(t|t-k) = \frac{\tilde{R}(z)}{A(z)} e(t-k) \quad = \frac{\tilde{R}(z)}{A(z)} z^{-k} \cdot e(t) \quad = \frac{\tilde{R}(z)}{A(z)} z^{-k} \cdot \frac{A(z)}{C(z)} y(t) \quad = \frac{\tilde{R}(z)}{C(z)} y(t-k)$$

Predittore ottimo per processi ARMA

Il **predittore ottimo dai dati** è

$$\boxed{\hat{y}(t|t-k) = \frac{\tilde{R}(z)}{C(z)} y(t-k)}$$

L'**errore di predizione** corrispondente è

$$\boxed{\varepsilon_k(t) = y(t) - \hat{y}(t|t-k) = E(z)e(t)}$$

Predittore ottimo per processi ARMA

Caso particolare: predizione ad un passo $k = 1$

$$
\begin{array}{r|l}
C(z) & A(z) \\ \cline{2-2}
-A(z) & \boxed{1} \; E(z) \\ \hline
\boxed{C(z) - A(z)} & \\
R(z) &
\end{array}
$$

$E(z) = 1 \quad \Rightarrow \quad \varepsilon_1(t) = E(z)e(t) = e(t)$

$R(z) = C(z) - A(z)$

Predittore dai dati a un passo

$$\boxed{\hat{y}(t|t-1) = \frac{C(z) - A(z)}{C(z)} y(t)}$$

Errore di predizione

$$\boxed{\varepsilon_1(t) = E(z)e(t) = e(t)}$$

Predittore ottimo per processi ARMA

Osservazioni

- $\hat{y}(t|t-k)$ è corretto, infatti $\mathbb{E}[y(t)] = \mathbb{E}[\hat{y}(t|t-k)] = 0$
- $\hat{y}(t|t-k)$ dipende dal WN fino al tempo $t-k$
- $\mathbb{E}[\hat{y}(t|t-k) \cdot \varepsilon_k(t)] = 0$, infatti $\mathbb{E}[\hat{y}(t|t-k) \cdot \varepsilon_k(t)] = \mathbb{E}\left[\left(\frac{\tilde{R}(z)}{A(z)} e(t-k)\right) \cdot \left(E(z)e(t)\right)\right] = 0$
- Si dimostra che non è possible trovare un predittore con $\mathrm{Var}[\varepsilon_k(t)]$ minore

Ne consegue che $\hat{y}(t|t-k)$ è il **predittore lineare ottimo**

Qualità del predittore

Possiamo valutare la **qualità del predittore** mettendo a confronto la **varianza dell'errore di predizione** ottenuto con la varianza dell'errore di predizione di un **predittore banale** (che predice sempre la media processo, cioè sempre zero)

$$\mathrm{ESR} = \frac{\mathrm{Var}[y(t) - \hat{y}(t|t-k)]}{\mathrm{Var}[y(t) - 0]} = \frac{\mathrm{Var}[\varepsilon_k(t)]}{\mathrm{Var}[y(t)]}$$

- Il valore $1 - \mathrm{ESR}$ ci fornisce la **percentuale di varianza del processo che è stata «catturata» dal predittore**

- L'ESR varia tra 0 e 1. Un valore di ESR **inferiore** indica un predittore **migliore**

Outline

1. Predizione, filtraggio e smoothing
2. Scomposizione di Wold
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. Predittore ottimo per processi MA
6. Predittore ottimo per processi ARMA
7. **Predittore ottimo per processi ARMAX**
8. Predittore ottimo ad un passo per sistemi ingresso\uscita
9. Confronto con il predittore di Kalman

Predittore ottimo per processi ARMAX

Sia dato un processo $\mathrm{ARMAX}(n_a, n_c, n_b, k)$, con $C(z)/A(z)$ in **forma canonica**

$$y(t) = \frac{B(z)}{A(z)} u(t-k) + \frac{C(z)}{A(z)} e(t) \qquad e(t) \sim WN(0, \lambda^2)$$

- $C(z) = 1 + c_1 z^{-1} + c_2 z^{-2} + \cdots + c_{n_c} z^{-n_c}$
- $A(z) = 1 - a_1 z^{-1} - a_2 z^{-2} - \cdots - a_{n_a} z^{-n_a}$
- $B(z) = b_0 + b_1 z^{-1} + \cdots + b_{n_b} z^{-n_b}$

In questo caso, è **sensato fare una previsione a k passi**, in modo che l'ingresso riesca ad influenzare l'uscita. Quindi, «confondiamo» i k passi di previsione con i k passi di ritardo puro tra ingresso e uscita

Predittore ottimo per processi ARMAX

Applichiamo k passi di lunga divisione per scomporre $C(z)/A(z)$

$$y(t) = \underbrace{\frac{B(z)}{A(z)}u(t-k) + \frac{\tilde{R}(z)}{A(z)}e(t-k)}_{\text{Parte predicibile al tempo } t-k} + \underbrace{E(z)e(t)}_{\text{Parte impredicibile al tempo } t-k}$$

Parte predicibile al tempo $t-k$

Parte impredicibile al tempo $t-k$

Il **predittore ottimo dal rumore** è

$$\boxed{\hat{y}(t|t-k) = \frac{B(z)}{A(z)}u(t-k) + \frac{\tilde{R}(z)}{A(z)}e(t-k)}$$

Predittore ottimo per processi ARMAX

Calcoliamo il **predittore ottimo dai dati**

$$y(t) = \frac{B(z)}{A(z)}u(t-k) + \frac{C(z)}{A(z)}e(t) \quad \Rightarrow \quad \underbrace{e(t) = \frac{A(z)}{C(z)}y(t) - \frac{B(z)}{C(z)}u(t-k)}_{\text{FILTRO SBIANCANTE}}$$

$$\hat{y}(t|t-k) = \frac{B(z)}{A(z)}u(t-k) + \frac{\tilde{R}(z)}{A(z)}e(t-k) \quad \Rightarrow$$

$$\hat{y}(t|t-k) = \frac{\tilde{R}(z)}{C(z)}y(t-k) + \frac{B(z)E(z)}{C(z)}u(t-k)$$

Predittore ottimo per processi ARMAX

Il **predittore ottimo dai dati** è

$$\hat{y}(t|t-k) = \frac{\tilde{R}(z)}{C(z)} y(t-k) + \frac{B(z)E(z)}{C(z)} u(t-k)$$

L'**errore di predizione** corrispondente è

$$\varepsilon_k(t) = E(z)e(t)$$

Predittore ottimo per processi ARMAX

Caso particolare: predizione ad un passo $k = 1$

$$E(z) = 1 \quad \Rightarrow \quad \varepsilon_1(t) = E(z)e(t) = e(t)$$

$$R(z) = C(z) - A(z)$$

Predittore dai dati a un passo

$$\boxed{\hat{y}(t|t-1) = \frac{C(z) - A(z)}{C(z)} y(t) + \frac{B(z)}{C(z)} u(t-1)}$$

Errore di predizione a un passo

$$\boxed{\varepsilon_1(t) = E(z)e(t) = e(t)}$$

Predittore ottimo per processi ARMAX

Osservazioni

- $\hat{y}(t|t-k)$ è corretto, infatti $\mathbb{E}[y(t)] = \mathbb{E}[\hat{y}(t|t-k)] = 0$
- $\hat{y}(t|t-k)$ dipende dal WN e da $u(t)$ fino al tempo $t-k$
- $\mathbb{E}[\hat{y}(t|t-k) \cdot \varepsilon_k(t)] = 0$. Nel caso in cui $u(t)$ fosse un processo stocastico, si assume ragionevolmente che $u(t) \perp e(t)$
- Non è possible trovare un predittore con $\mathrm{Var}[\varepsilon_k(t)]$ minore

Ne consegue che $\hat{y}(t|t-k)$ è il **predittore lineare ottimo**

Predittore ottimo per processi ARMAX

Osservazione

Con il predittore ARMAX, la varianza di $\varepsilon_k(t)$ è data solo dalla parte ARMA, in quanto unica parte stocastica del modello. La **bontà del predittore** si può calcolare come

$$\mathrm{ESR} = \frac{\mathrm{Var}[\varepsilon_k(t)]}{\mathrm{Var}\left[\frac{C(z)}{A(z)}e(t)\right]}$$

Esempio: forma canonica e calcolo del predittore

Sia dato il processo $y(t)$. Calcolare il predittore dai dati e la varianza dell'errore di predizione

$$y(t) = (2 + 6z^{-1})u(t-2) + \frac{2}{3 + \frac{3}{2}z^{-1}} e(t-1), \qquad e(t) \sim \mathrm{WN}(0,1)$$

Il ritardo puro è $k = 2$. Ha quindi senso calcolare un predittore per $k = 2$ passi in avanti

Il processo è in forma canonica?

$$y(t) = (2 + 6z^{-1})u(t-2) + \frac{1}{1 + \frac{1}{2}z^{-1}} \cdot \boxed{\frac{2}{3}e(t-1)}$$

$$\frac{2}{3}e(t-1) \equiv \eta(t) \sim \mathrm{WN}\left(0, \frac{4}{9}\right)$$

Esempio: forma canonica e calcolo del predittore

$$y(t) = (2 + 6z^{-1})u(t-2) + \frac{1}{1 + \frac{1}{2}z^{-1}} \cdot \eta(t)$$

Per calcolare il predittore, la parte esogena e quella stocastica devono avere il medesimo polinomio $A(z)$ al denominatore

$$y(t) = \frac{\overbrace{(2 + 6z^{-1})\left(1 + \frac{1}{2}z^{-1}\right)}^{B(z)}}{\underbrace{1 + \frac{1}{2}z^{-1}}_{A(z)}} u(t-2) + \frac{\overbrace{1}^{C(z)}}{\underbrace{1 + \frac{1}{2}z^{-1}}_{A(z)}} \cdot \eta(t) \qquad \eta(t) \sim \mathrm{WN}\left(0, \frac{4}{9}\right)$$

Esempio: forma canonica e calcolo del predittore

Notiamo che il denominatore comune è stato fatto dopo la canonizzazione di $C(z)/A(z)$. La fdt $B(z)/A(z)$ non ha bisogno di essere in forma canonica, in quanto non è la parte stocastica. Calcoliamo il **predittore** a $k=2$ passi in avanti

$$
\begin{array}{l|l}
\boxed{1} \; C(z) & \boxed{1+\frac{1}{2}z^{-1}} \; A(z) \\
-1-\frac{1}{2}z^{-1} & \\
\hline
-\frac{1}{2}z^{-1} & \boxed{1-\frac{1}{2}z^{-1}} \; E(z) \\
\frac{1}{2}z^{-1}+\frac{1}{4}z^{-2} & \\
\hline
R(z) \; \boxed{+\frac{1}{4}z^{-2}} &
\end{array}
$$

$$E(z) = 1-\frac{1}{2}z^{-1} \qquad R(z) = z^{-k}\tilde{R}(z) = \frac{1}{4}z^{-2}$$

$$\hat{y}(t|t-k) = \frac{\tilde{R}(z)}{C(z)}y(t-k) + \frac{B(z)E(z)}{C(z)}u(t-k)$$

$$\boxed{\hat{y}(t|t-2) = \frac{\frac{1}{4}}{1}y(t-2) + \frac{2(1+3z^{-1})\left(1+\frac{1}{2}z^{-1}\right)\left(1-\frac{1}{2}z^{-1}\right)}{1}u(t-2)}$$

Esempio: forma canonica e calcolo del predittore

Esprimiamo il **predittore in forma ricorsiva**

$$\hat{y}(t|t-2) = \frac{1}{4}y(t-2) + 2(1+3z^{-1})\left(1+\frac{1}{2}z^{-1}\right)\left(1-\frac{1}{2}z^{-1}\right)u(t-2)$$

$$= \frac{1}{4}y(t-2) + 2(1+3z^{-1})\left(1-\frac{1}{4}z^{-2}\right)u(t-2)$$

$$\boxed{\hat{y}(t|t-2) = \frac{1}{4}y(t-2) + 2u(t-2) + 6u(t-3) - \frac{1}{2}u(t-4) - \frac{3}{2}u(t-5)}$$

Notiamo che nell'espressione del predittore non vi sono termini «prima» di $t-2$

Esempio: forma canonica e calcolo del predittore

Calcoliamo la **varianza dell'errore di predizione**

$$\mathrm{Var}[\varepsilon_2(t)] = \mathbb{E}[\varepsilon_2(t)^2] = \mathbb{E}\left[\left(E(z)\eta(t)\right)^2\right] = \mathbb{E}\left[\left(\left(1-\frac{1}{2}z^{-1}\right)\eta(t)\right)^2\right] = \left[1+\frac{1}{4}\right]\mathrm{Var}[\eta(t)]$$

$$= \frac{5}{4}\cdot\frac{4}{9} = \boxed{\frac{5}{9}}$$

Calcoliamo la **bontà del predittore ottimo** rispetto a quella del **predittore banale** $\hat{y}(t|t-2) = \mathbb{E}[y(t)] = 0$

$$\mathrm{ESR} = \frac{\mathrm{Var}[\varepsilon_2(t)]}{\mathrm{Var}\left[\frac{C(z)}{A(z)}\eta(t)\right]}$$

Esempio: forma canonica e calcolo del predittore

$$\mathrm{Var}\left[\frac{C(z)}{A(z)}\eta(t)\right] = \mathrm{Var}\left[\frac{1}{1+\frac{1}{2}z^{-1}}\eta(t)\right] = \mathrm{Var}[v(t)]$$

con $v(t) = \dfrac{1}{1+\frac{1}{2}z^{-1}}\eta(t)$

$$\Rightarrow \quad v(t) = -\frac{1}{2}v(t-1) + \eta(t)$$

$$\Rightarrow \quad \mathrm{Var}[v(t)] = \mathrm{Var}\left[-\frac{1}{2}v(t-1) + \eta(t)\right] = \mathbb{E}\left[\left(-\frac{1}{2}v(t-1) + \eta(t)\right)^2\right]$$

$$= \mathbb{E}\left[\frac{1}{4}v(t-1)^2 + \eta(t)^2 - v(t-1)\eta(t)\right] = \frac{1}{4}\mathrm{Var}[v(t)] + \mathrm{Var}[\eta(t)]$$

$$\frac{3}{4}\mathrm{Var}[v(t)] = \mathrm{Var}[\eta(t)] \quad \Rightarrow \quad \boxed{\mathrm{Var}[v(t)] = \frac{4}{3}\cdot\frac{4}{9} = \frac{16}{27}}$$

Esempio: forma canonica e calcolo del predittore

$$\mathrm{ESR} = \frac{\mathrm{Var}[\varepsilon_2(t)]}{\mathrm{Var}\left[\frac{C(z)}{A(z)}\eta(t)\right]} = \frac{\frac{5}{9}}{\frac{16}{27}} = \frac{5}{9}\cdot\frac{27}{16} = \frac{15}{16} = \boxed{0.9375}$$

Il predittore ottimo **ha ridotto l'incertezza con cui prevediamo «due passi in avanti»** circa del 7%.

Nota: «Predirre il 7%» di un processo, anche se sembra poco, non significa sia inutile: dipende dal contesto applicativo

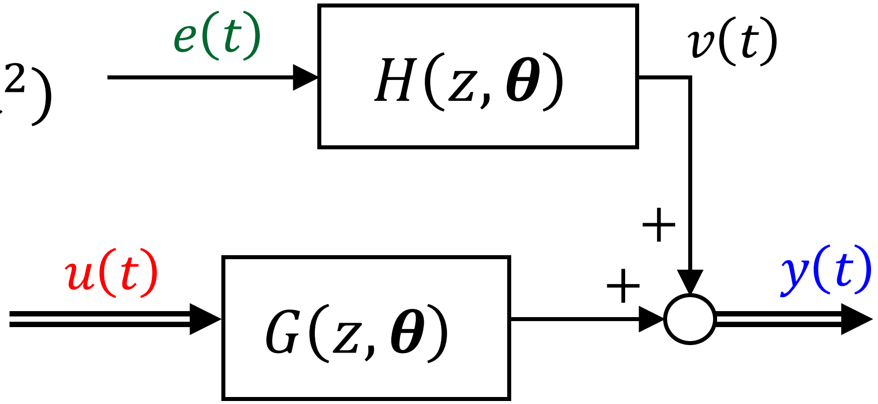
Outline

1. Predizione, filtraggio e smoothing
2. Scomposizione di Wold
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. Predittore ottimo per processi MA
6. Predittore ottimo per processi ARMA
7. Predittore ottimo per processi ARMAX
8. **Predittore ottimo ad un passo per sistemi ingresso\uscita**
9. Confronto con il predittore di Kalman

Predittore ottimo ad un passo per sistemi I\O

Abbiamo visto come, nel caso di sistemi dinamici **LTI SISO ingresso\uscita**, usiamo un modello $\mathcal{M}(\boldsymbol{\theta})$ della forma seguente

$$\mathcal{M}(\boldsymbol{\theta}): y(t) = G(z,\boldsymbol{\theta})u(t) + H(z,\boldsymbol{\theta})e(t), \quad e(t) \sim \mathrm{WN}(0,\lambda^2)$$



$$G(z,\boldsymbol{\theta}) = \sum_{i=1}^{\infty} g_i(\boldsymbol{\theta}) \cdot z^{-i} \qquad H(z) = 1 + \sum_{i=1}^{\infty} h_i(\boldsymbol{\theta}) \cdot z^{-i}$$

Si suppone $H(z)$ **in forma canonica**

Notiamo che il **filtro sbiancante** si ottiene come

$$e(t) = H^{-1}(z,\boldsymbol{\theta})[y(t) - G(z,\boldsymbol{\theta})u(t)]$$

Predittore ottimo ad un passo per sistemi I\O

$y(t) = G(z,\boldsymbol{\theta})u(t) + H(z,\boldsymbol{\theta})e(t)$ sommo e tolgo $e(t)$

$$= G(z,\boldsymbol{\theta})u(t) + [H(z,\boldsymbol{\theta}) - 1]e(t) + e(t)$$

Sostituendo l'espressione del **filtro sbiancante** che produce $e(t)$ nel secondo termine

$$y(t) = G(z,\boldsymbol{\theta})u(t) + [H(z,\boldsymbol{\theta}) - 1]H^{-1}(z,\boldsymbol{\theta})[y(t) - G(z,\boldsymbol{\theta})u(t)] + e(t)$$

$$= H^{-1}(z,\boldsymbol{\theta})G(z,\boldsymbol{\theta})u(t) + [1 - H^{-1}(z,\boldsymbol{\theta})]y(t) + e(t)$$

Dato che $H(z,\boldsymbol{\theta})$ è in **forma canonica**, anche $H^{-1}(z,\boldsymbol{\theta})$ è in forma canonica, per cui

$$\frac{1}{H(z,\boldsymbol{\theta})} = 1 + d_1 z^{-1} + d_2 z^{-2} + \cdots$$

Predittore ottimo ad un passo per sistemi I\O

Supponendo che $G(z, \boldsymbol{\theta})$ sia **strettamente propria** (i.e. almeno un passo di ritardo tra ingresso e uscita), si ha che la quantità

$$H^{-1}(z, \boldsymbol{\theta})G(z, \boldsymbol{\theta})u(t) + [1 - H^{-1}(z, \boldsymbol{\theta})]y(t)$$

dipende solo da $H(z, \boldsymbol{\theta}), G(z, \boldsymbol{\theta})$ e dai dati $u(t-1), u(t-2), \dots$ e $y(t-1), y(t-2), \dots$

Questa quantità è quindi **completamente predicibile** al tempo $t - 1$

Predittore ottimo ad un passo per sistemi I\O

Il **predittore ottimo ad un passo** $\widehat{\mathcal{M}}(\boldsymbol{\theta})$ per la **classe di modelli** $\mathcal{M}(\boldsymbol{\theta})$ è dato da

$$\boxed{\widehat{\mathcal{M}}(\boldsymbol{\theta}):\ \hat{y}(t|t-1;\boldsymbol{\theta}) = H^{-1}(z,\boldsymbol{\theta})G(z,\boldsymbol{\theta})u(t) + [1 - H^{-1}(z,\boldsymbol{\theta})]y(t)}$$

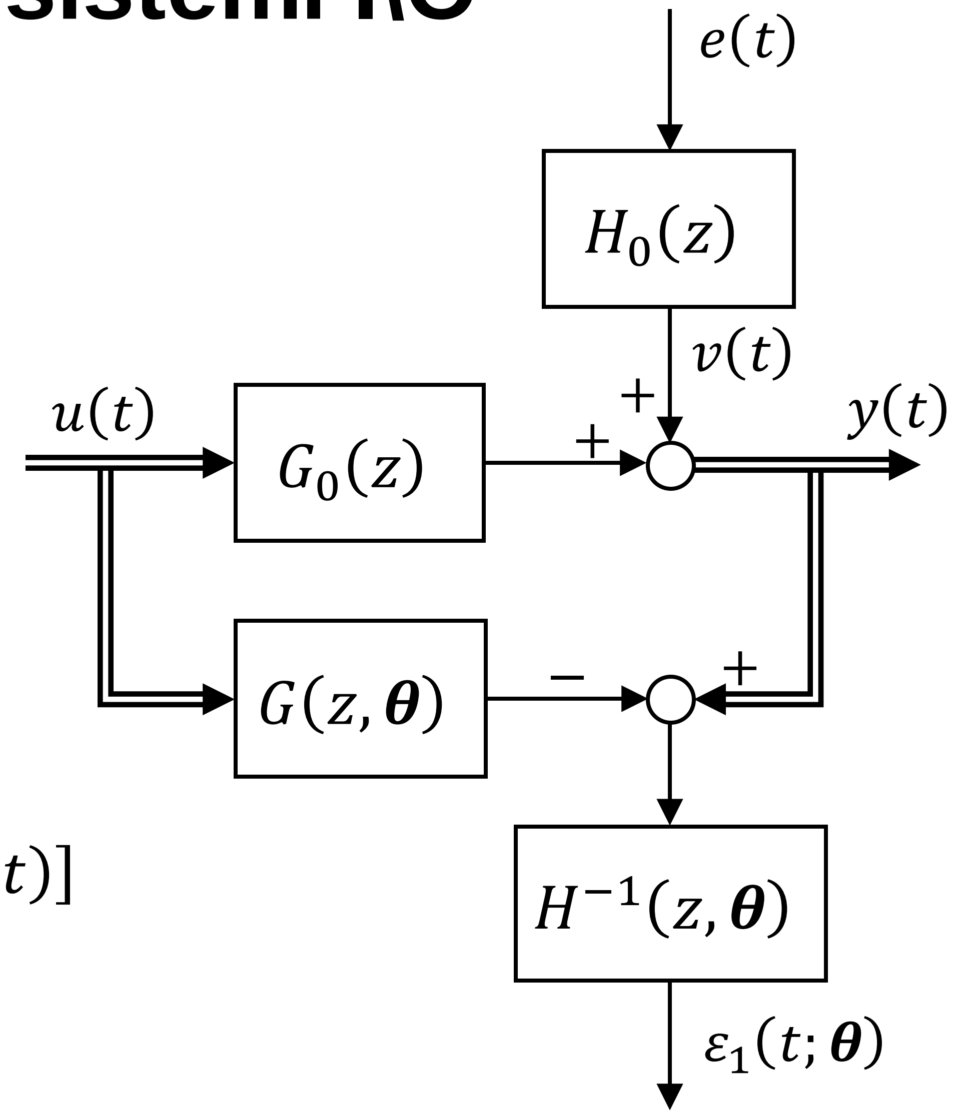
L'**errore di predizione ad un passo** $\varepsilon_1(t;\boldsymbol{\theta}) = y(t) - \hat{y}(t|t-1;\boldsymbol{\theta})$ è

$$\boxed{\varepsilon_1(t;\boldsymbol{\theta}) = H^{-1}(z,\boldsymbol{\theta})[y(t) - G(z,\boldsymbol{\theta})u(t)]}$$

Predittore ottimo ad un passo per sistemi I\O

Osservazione

Sostituendo l'equazione del **sistema che genera i dati** $y(t) = G_0(z)u(t) + H_0(z)e(t)$ all'interno dell'**errore di predizione a un passo** $\varepsilon_1(t)$ si ha



$$\varepsilon_1(t;\boldsymbol{\theta}) = H^{-1}(z,\boldsymbol{\theta})[y(t) - G(z,\boldsymbol{\theta})u(t)]$$

$$= H^{-1}(z,\boldsymbol{\theta})[G_0(z)u(t) + H_0(z)e(t) - G(z,\boldsymbol{\theta})u(t)]$$

Predittore ottimo ad un passo per sistemi I\O

Se $\exists \boldsymbol{\theta} = \boldsymbol{\theta}^0$ t.c. $G_0(z) = G(z, \boldsymbol{\theta}^0)$ e $H_0(z) = H(z, \boldsymbol{\theta}^0)$, ovvero, **se il sistema vero appartiene alla famiglia di modelli scelta**, otteniamo che

$$\varepsilon_1(t; \boldsymbol{\theta}^0) = e(t)$$

Quindi, il valore $\boldsymbol{\theta}^0$:

1. È **l'unico valore** che rende $\varepsilon_1(t; \boldsymbol{\theta}^0) = e(t)$
2. **Minimizza la varianza** dell'errore di predizione a un passo

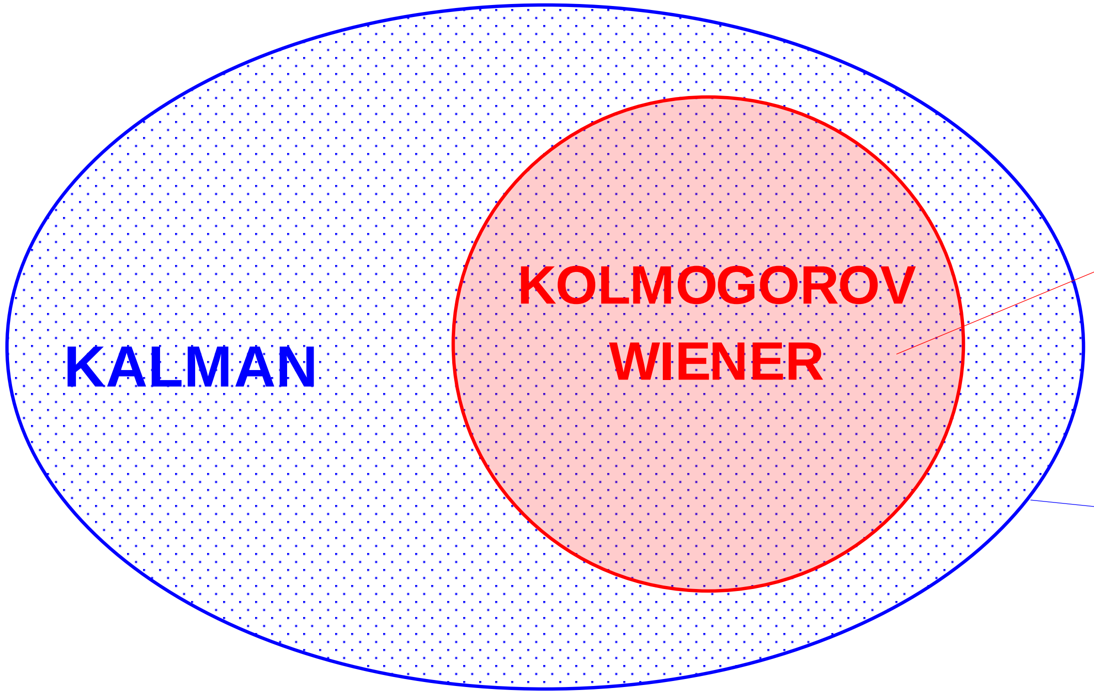
Ne consegue che $\varepsilon_1(t)$ è un buon indicatore della **bontà di un modello dinamico**. Useremo questa proprietà per trovare un **criterio di identificazione** dei modelli dinamici

Outline

1. Predizione, filtraggio e smoothing
2. Scomposizione di Wold
3. Filtro passa-tutto e forma canonica
4. Predittore ottimo
5. Predittore ottimo per processi MA
6. Predittore ottimo per processi ARMA
7. Predittore ottimo per processi ARMAX
8. Predittore ottimo ad un passo per sistemi ingresso\uscita
9. **Confronto con il predittore di Kalman**

Confronto con il predittore di Kalman

La teoria della predizione che abbiamo visto è anche nota come **predizione alla Kolmogorov-Wiener**. È interessante confrontarla con **la teoria della predizione di Kalman** per sistemi dinamici espressi in **spazio di stato**



Sistemi:
- Lineari
- Tempo Invarianti
- Asintoticamente stabili
- A regime

Sistemi:
- Non lineari
- Tempo varianti
- Non asintoticamente stabili
- In transitorio

Confronto con il predittore di Kalman

La teoria di Kalman è quindi **più generale** della teoria di Kolmogorov-Wiener. Però, la teoria KW ci fornisce la base per lo **sviluppo di metodi di identificazione** intuitivi ed efficaci

Nella prossima lezione, ci baseremo sulla teoria Kolmogorov-Wiener per definire metodi di identificazione dei modelli basati sulla **minimizzazione della varianza dell'errore di predizione**

UNIVERSITÀ
DEGLI STUDI
DI BERGAMO

Dipartimento
di Ingegneria Gestionale,
dell'Informazione e della Produzione