

RETI NEURALI

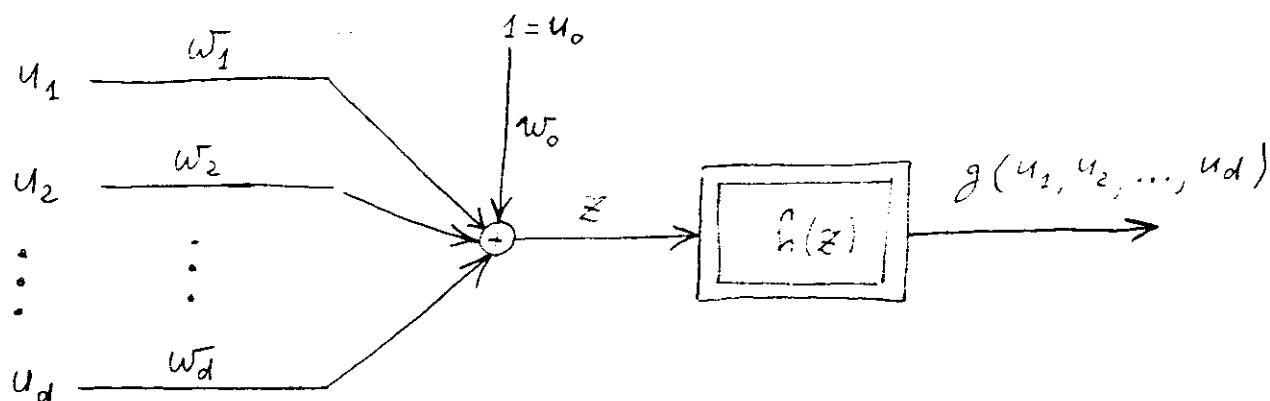
Contenuti:

- Reti di perceptroni
- Reti neurali a base radiale
- Esempio: Stima del rendimento volumetrico η
- Conclusioni

Motivazione: Modelli flessibili e modulari (= costituiti dalla interconnessione di blocchi elementari) per rappresentare relazioni del tipo $y = g(u_1, u_2, \dots, u_d)$. (identificazione a scatola nera di modelli nonlineari).

RETI DI PERCETTRONI

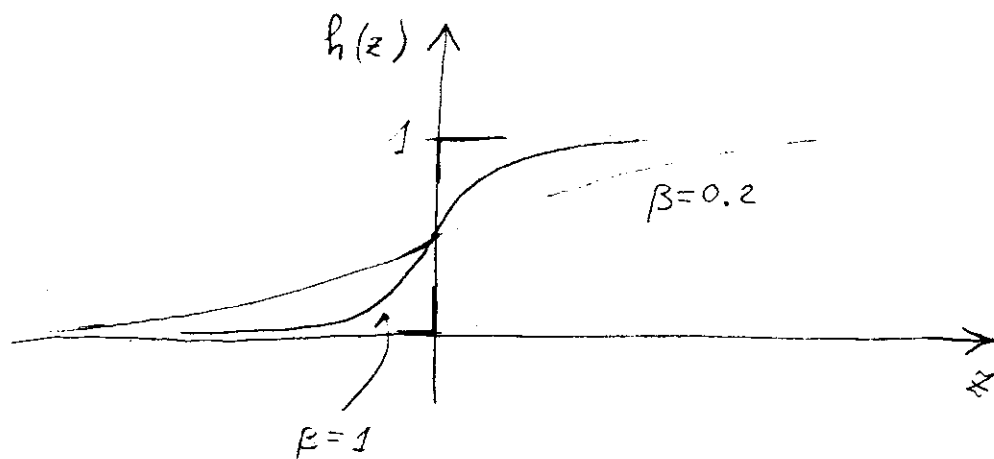
Il perceptrone (Rosenblatt 1958):



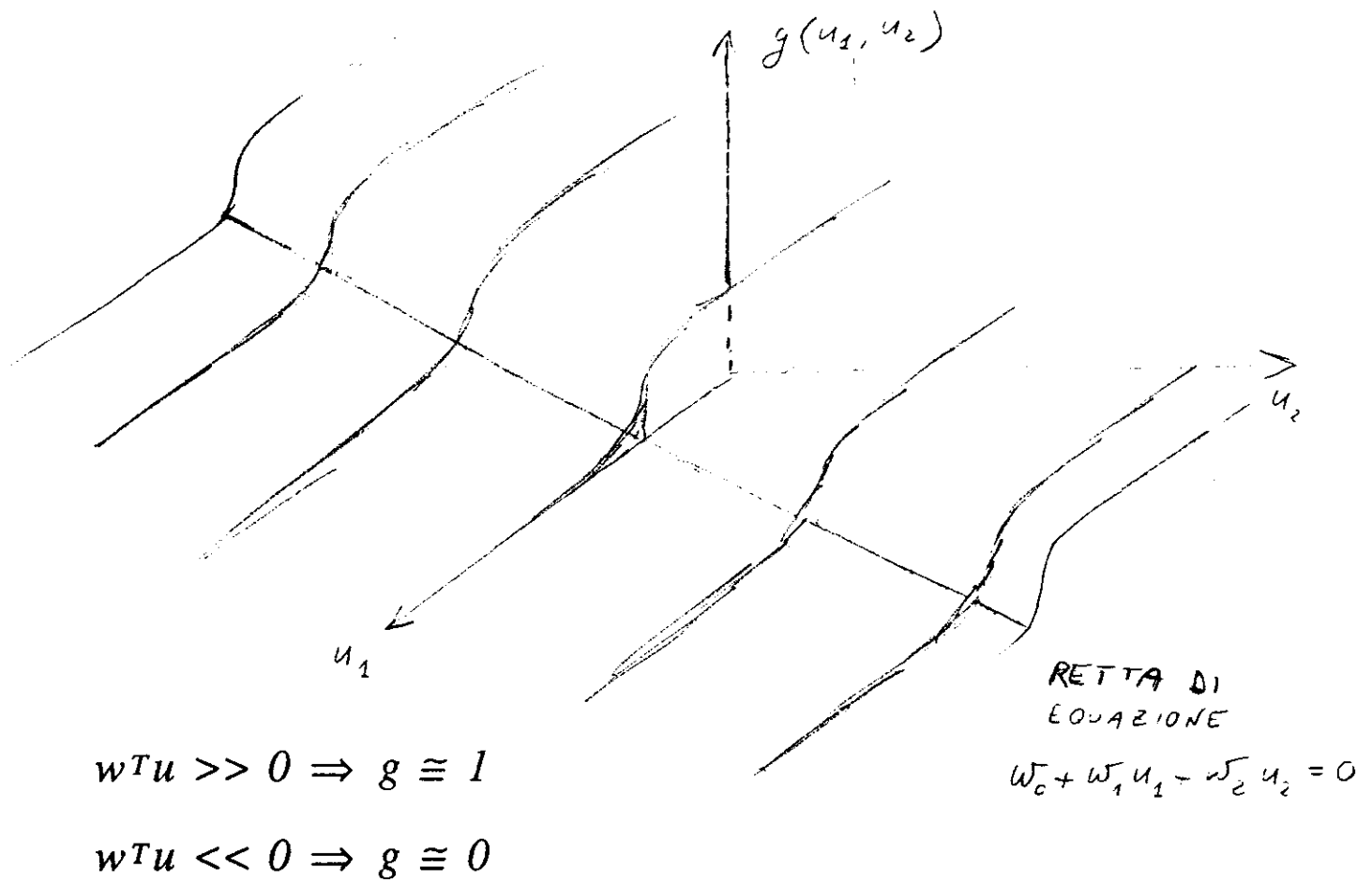
$$g(u_1, u_2, \dots, u_d) = h \left(\sum_{i=0}^d w_i u_i \right) = h(w^T u)$$

- $w = [w_0 \ w_1 \ w_2 \ \dots \ w_d]^T$: vettore dei pesi (è l'equivalente di θ)
- $h(z)$ è tipicamente una nonlinearity di tipo sigmoide.
Per esempio:

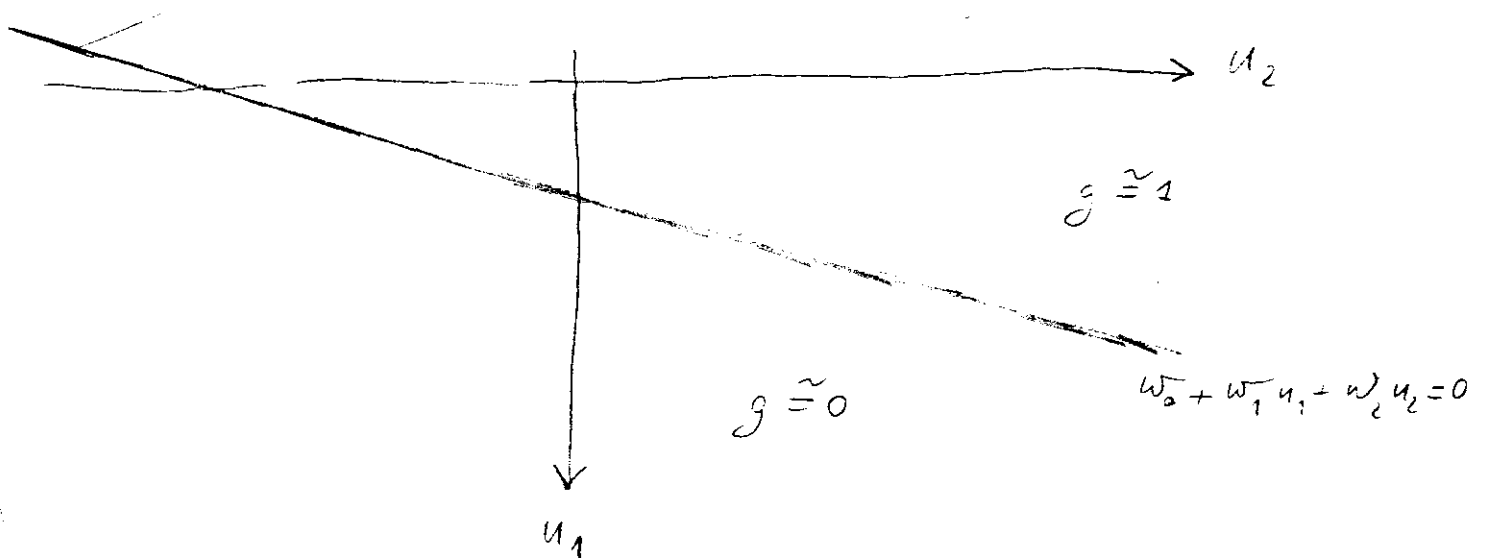
$$h(z) = (1 + \exp(-\beta z))^{-1}$$



Interpretazione grafica del perceptrone

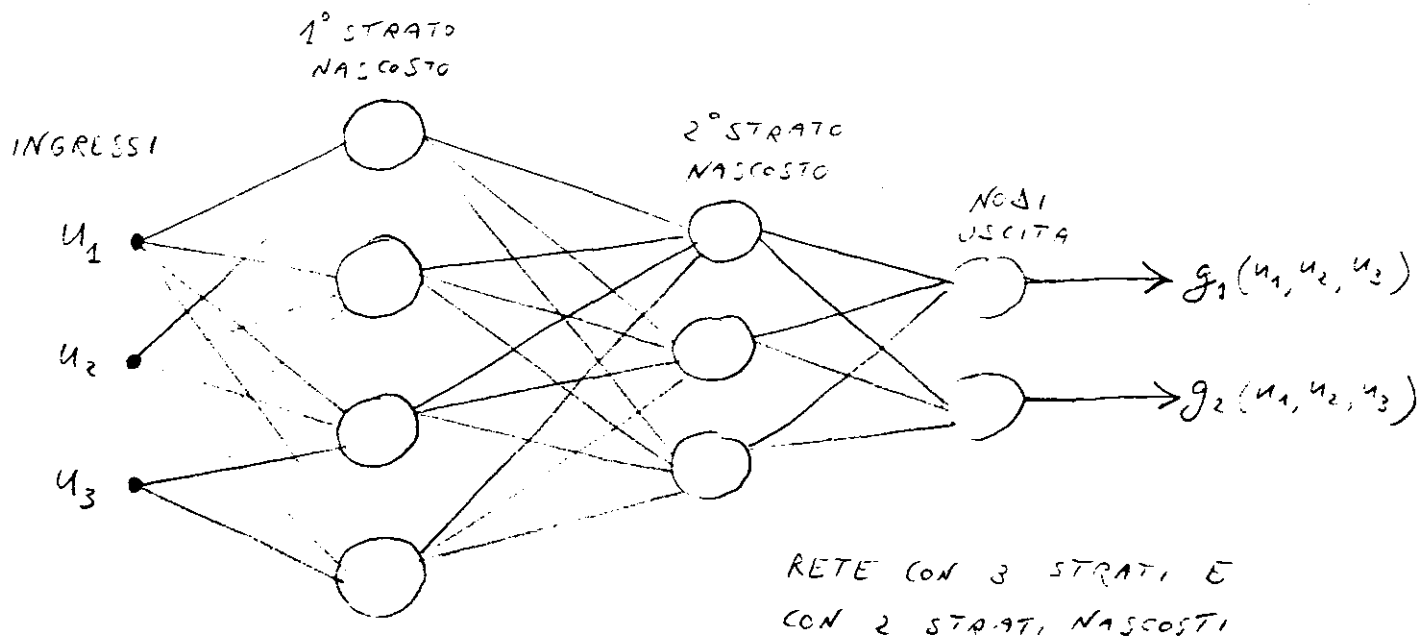


Il singolo perceptrone divide lo spazio delle variabili indipendenti in due regioni (classificatore).



Idea: combinare in cascata più perceptron

Rete neurale di perceptron a più strati
(MLP NN: Multi-Layer Perceptron Neural Network)



Teorema (Cybenko 1989): Una rete con 2 strati (= con 1 strato nascosto) è un *approssimatore universale*, può cioè approssimare in modo arbitrariamente preciso ogni funzione continua $g(u_1, u_2, \dots, u_d)$ (pur di prendere un numero sufficiente di neuroni).

Nota: il teorema vale anche se nei nodi di uscita la nonlinearietà sigmoide è sostituita da una relazione lineare.

Come stimare i parametri (= pesi w_i)?

$g(u, \theta)$ è funzione nonlineare di θ



problema di stima nonlineare

Criterio: Minimizzazione di SSR

Algoritmi: Quelli già visti, tra cui:

- Metodo del gradiente (chiamato "backpropagation")
- Gauss-Newton
- etc.

Difficoltà numeriche: Problema di ottimizzazione non lineare che può avere vari minimi locali. E' bene ripetere gli algoritmi iterativi partendo da inizializzazioni di volta in volta diverse. Richiede molto tempo.

Glossario

Teoria dell'identificazione

Reti neurali

Modello ↔ Rete neurale

Parametri ↔ Pesi

Stimare i parametri ↔ Addestrare la rete

Dati per l'identificazione ↔ Esempi, training set

Sovraparametrizzazione ↔ Overtraining

Problema fondamentale: Scelta del n° di neuroni e della struttura della rete.

Osservazione: Per valutare la bontà di un modello, tra tutti i criteri visti, il più applicabile è la crossvalidazione dividendo i dati in due gruppi.

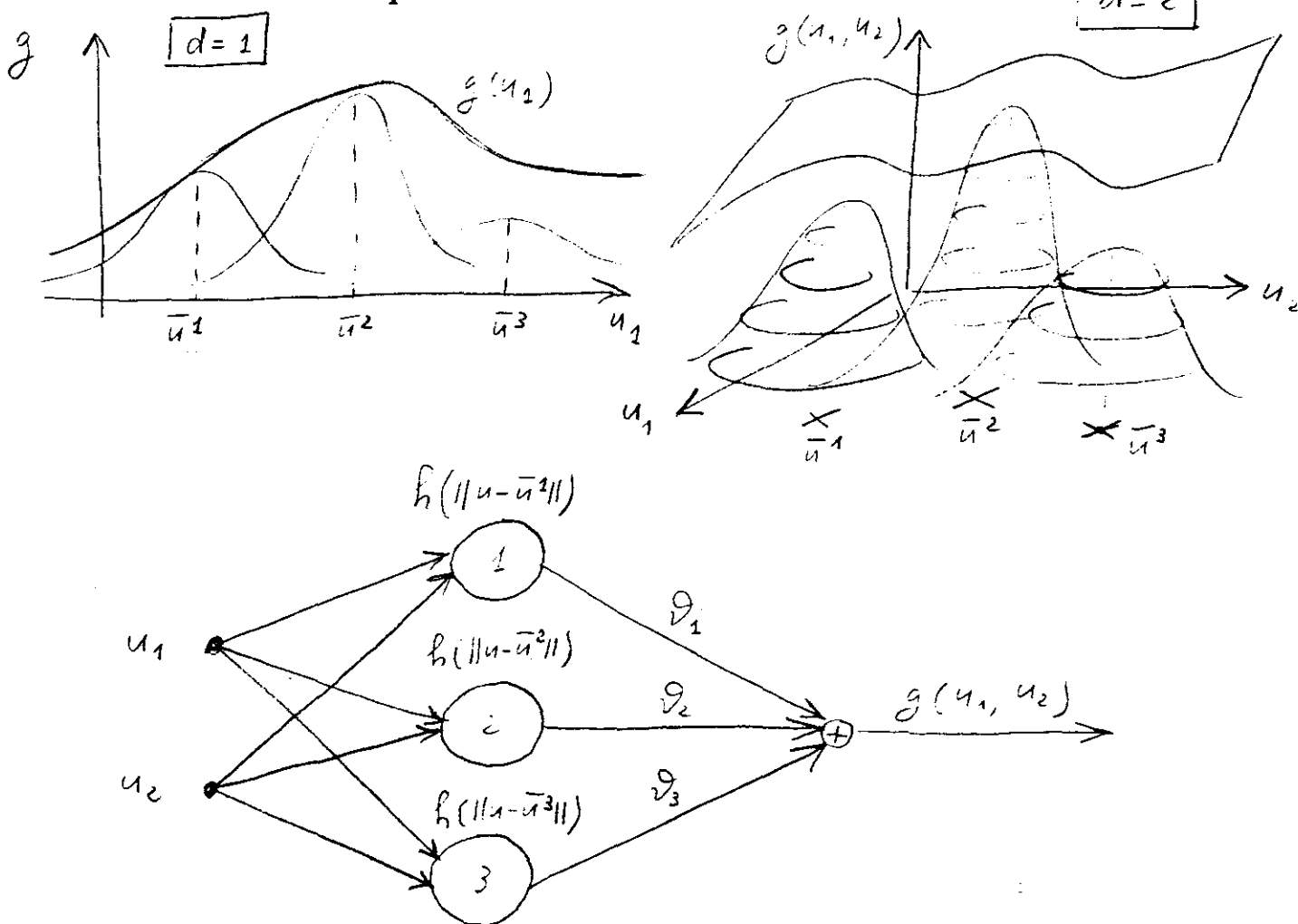
Alcune tecniche basate sulla crossvalidazione:

- Comincio con pochi neuroni e ne stimo i parametri. Poi "faccio crescere" la rete aggiungendo altri neuroni. Per decidere quando fermarmi uso la crossvalidazione.
- Optimal brain damage: uso una rete sovradimensionata e distruggo nodi e/o pesi che contano poco ("pruning") fino a quando riesco a migliorare SSR^v .
- Uso una rete sovradimensionata e la addestro con un algoritmo iterativo monitorando i progressi mediante crossvalidazione. Invece di arrivare alla convergenza, mi fermo quando le prestazioni non migliorano più.
- ...

Regola empirica: Dato che n° parametri \gg n° neuroni, è bene che il n° neuroni sia \ll del numero N dei dati.

RETI NEURALI A BASE RADIALE

Idea: Rappresento $g(u_1, u_2, \dots, u_d)$ come una combinazione lineare di funzioni "a campana".



$$g(u_1, u_2, \dots, u_d) = \sum_{k=1}^n \theta_k h(\|u - \bar{u}^k\|)$$

Se la forma $h(\cdot)$ delle campane è fissata, i parametri liberi sono i centri \bar{u}^k e le ampiezze θ_k .

Tipicamente, $h(r)$ è una funzione a simmetria radiale:

- $h(r) = e^{-\frac{r^2}{2c^2}}$
- $h(r) = \sqrt{r^2+c^2}$
- $h(r) = (r^2+c^2)^\beta$, $0 < \beta < 1$
- $h(r) = (r^2+c^2)^{-\beta}$, $\beta > 0$

c, β : parametri liberi che caratterizzano $h(r)$.

Osservazioni:

- Anche le Radial Basis Function Neural Networks (RBF NN) sono approssimatori universali.
- Parentela con i modelli fuzzy (basta interpretare $h(r)$ come una "membership function").
- Fatto interessante: una volta note le coordinate dei centri \bar{u}^k , la stima di θ è un problema lineare (risolvibile mediante LS).

Idea: Spezzo il problema dell'addestramento in due fasi:

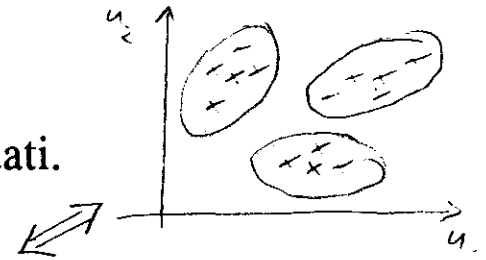
- 1) Posizionamento dei centri \bar{u}^k
- 2) Stima delle ampiezze θ_k

La forma di $h(r)$ viene di solito scelta a priori (sono comunque possibili aggiustamenti dei parametri c, β).

Tecniche per il posizionamento dei centri:

(i) n° dei centri fissato a priori

- Griglia uniforme (inutile però mettere tanti centri dove ho pochi dati).
- Scelta casuale di un sottoinsieme dei dati.
- Divido i dati $\{u^{(k)}\}$ in n "grappoli" con un opportuno algoritmo di "clustering" (K-MEANS, per esempio) e uso come centri i baricentri dei grappoli.
- Algoritmi più sofisticati: ottimizzazione simultanea di tutti i parametri (coordinate \bar{u}^k dei centri e ampiezze θ_k). Difficoltà numeriche (stima nonlineare). Ne vale la pena?



(ii) n° dei centri non fissato a priori

- Stepwise regression considerando come regressori candidati tutte le funzioni radiali centrate in corrispondenza degli N dati.

Scelta del n° di centri: E' consigliabile ricorrere alla crossvalidazione (si procede per tentativi cercando di minimizzare SSR^v).

ESEMPIO: STIMA DEL RENDIMENTO VOLUMETRICO η

Basandosi sui risultati dell'identificazione mediante modelli polinomiali sono state considerate solo n , p_{man} come variabili indipendenti

$$\eta = \eta(n, p_{man})$$

RBF

- $h(r)$: gaussiana
- Posizionamento dei centri mediante K-MEANS
- n° dei neuroni e parametro c scelti mediante crossvalidazione

MLP

- $h(z) = \tanh(z)$
- Stima dei pesi mediante backpropagation.
- Poiché il risultato della stima dipende dall'inizializzazione dei pesi, la backpropagation è stata ripetuta usando inizializzazioni casuali.

Sia per RBF che MLP è essenziale normalizzare le variabili indipendenti n , p_{man} .

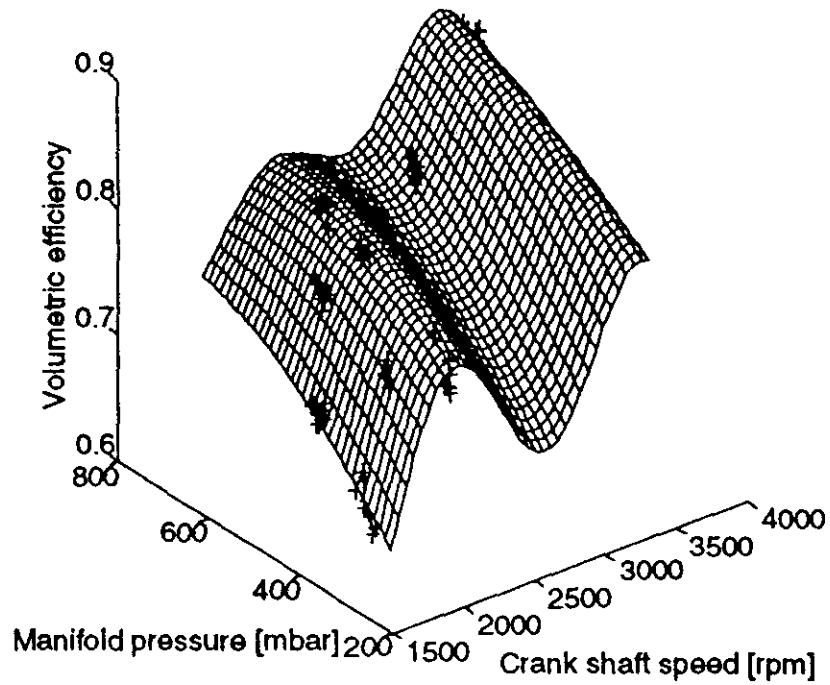


Figure 9: Radial basis function neural-network identification of η_v as a function of n , p_{man} (8 neurons). The symbol "+" indicates the validation data.

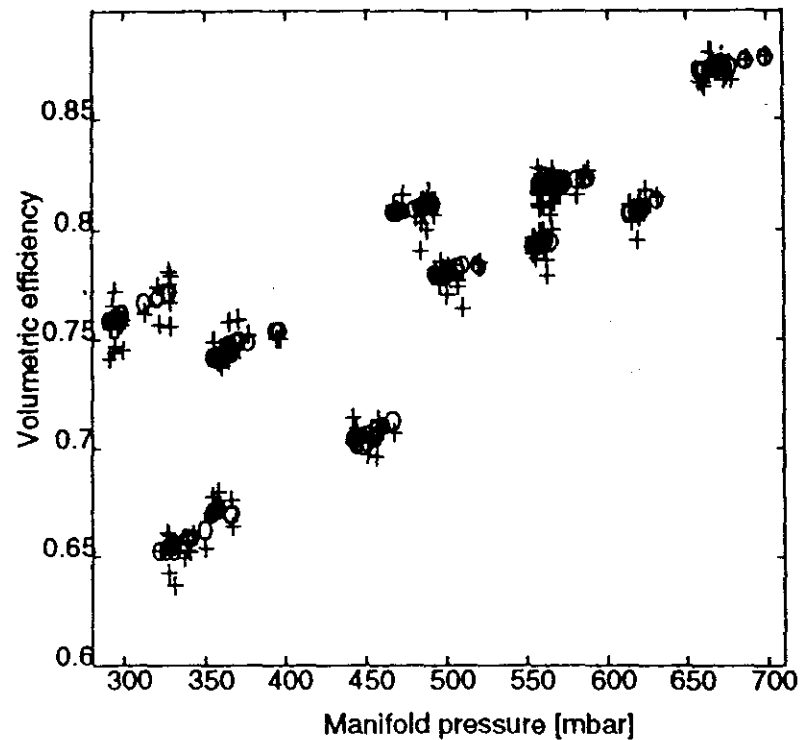


Figure 10: Radial basis function neural-network identification of η_v . Validation (+) versus predicted (o) data.

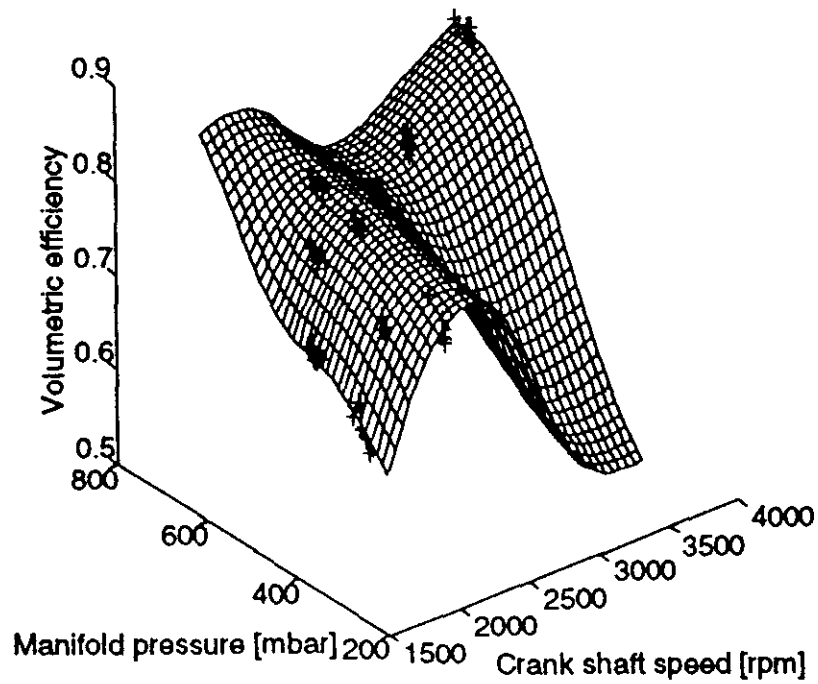


Figure 11: Multi-layer perceptron identification of η_v as a function of n , p_{man} (5 neurons). The symbol "+" indicates the validation data.

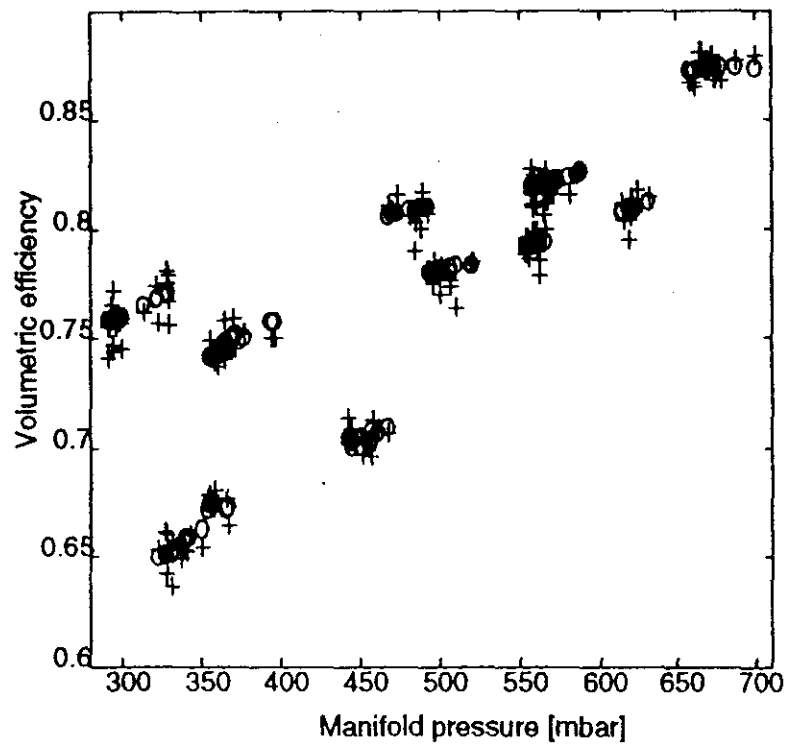


Figure 12: Multi-layer perceptron identification of η_v (5 neurons). Validation (+) versus predicted (o) data

Table 3. RBF neural network identification: sum of squared residuals (SSR) and largest standard deviation of identified b_k parameters ($M=5-9$).

Number of neurons	SSR	Largest SD
5	0.0079	2.7%
6	0.0077	31.2%
7	0.0075	28%
8	0.0071	27%
9	0.0072	11.8%

Table 4. MLP neural network identification: sum of squared residuals (SSR) ($M=3-7$).

Number of neurons	SSR
3	0.0072
4	0.0072
5	0.0069
6	0.0075
7	0.0074

Table 5. Performances of the different identification method: SSR on the validation data and SD of residuals.

Identification method	SSR	SD of residuals
<i>Additive model</i>	0.0076	0.0070
<i>Polynomial model</i>	0.0107	0.0083
<i>RBF (8 neurons)</i>	0.0071	0.0068
<i>ML (5 neurons)</i>	0.0069	0.0067

CONCLUSIONI

- Reti MLP: strumento versatile. Problemi di minimi locali e di tempo di calcolo.
- Reti RBF: una volta fissati i centri, l'addestramento si riduce ad una stima *LS*. lineare Problema: posizionamento dei centri.
- Nella stima del rendimento volumetrico le reti neurali danno risultati (leggermente) migliori dei modelli polinomiali, ma la loro identificazione è più complicata.
- Per identificare "a scatola nera" una funzione $g(u_1, u_2, \dots, u_d)$ senza conoscerne a priori la struttura funzionale, è necessario avere molti dati, soprattutto al crescere di d ("maledizione della dimensionalità"). Se, grazie alle leggi fisiche, conosco la struttura di $g(u_1, u_2, \dots, u_d, \theta)$ a meno di (pochi) parametri incogniti θ , posso ottenere risultati migliori mediante la stima parametrica. Però, se d è "piccolo" rispetto al n° di dati disponibili, l'identificazione a scatola nera mi consente di scavalcare tutti i problemi di modellistica fisica.

TRASFORMATA Z
E
TRASFORMATA DI FOURIER

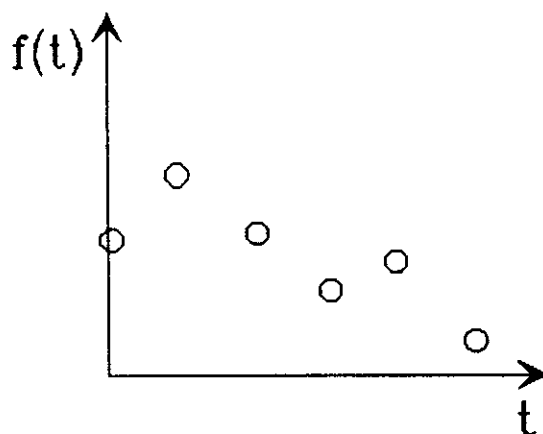
- Si introducono gli strumenti matematici necessari per lo studio dei sistemi a segnali campionati

TRASFORMATA ZETA

Opera sui segnali a tempo discreto così come la trasformata di Laplace opera sui segnali a tempo continuo

Definizione

Sia $f(t)$ un segnale a tempo discreto, reale e tale che $f(t)=0$ per $t<0$



Sia inoltre r un numero reale positivo tale che

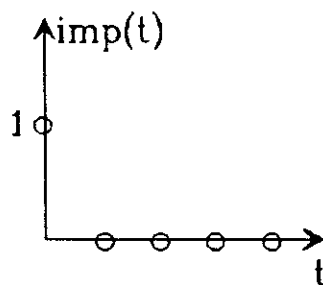
$$\sum_{t=0}^{\infty} |f(t)| r^{-t} < \infty$$

Allora per ogni numero complesso z tale che $|z|>r$ la trasformata zeta $F(z)$ di $f(t)$ è definita come

$$F(z) = \sum_{t=0}^{\infty} f(t) z^{-t} = f(0) + f(1) z^{-1} + f(2) z^{-2} + \dots$$

Essa è dunque funzione complessa della variabile complessa z , definita nel coccchio complesso e aperto $|z|>r$ e, per le proprietà delle funzioni analitiche, estendibile quasi ovunque nel piano complesso.

$$\text{Impulso discreto: } f(t) = \text{imp}(t) = \begin{cases} 1 & , t=0 \\ 0 & , t>0 \end{cases}$$

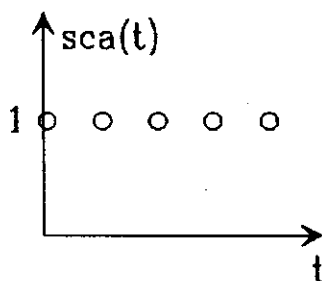


Dalla definizione segue che

$$F(z) = f(0) = 1$$

si noti l'analogia con la trasformata di Laplace dell'impulso a tempo continuo.

$$\text{Scalino discreto: } f(t) = \text{sca}(t) = 1, t \geq 0$$



Dalla definizione segue che (nell'ipotesi di convergenza della serie)

$$F(z) = 1 + z^{-1} + z^{-2} + \dots = \frac{1}{1 - z^{-1}} = \frac{z}{z - 1}$$

la trasformata ha uno zero nell'origine e un polo in $z=1$.

Antitrasformazione

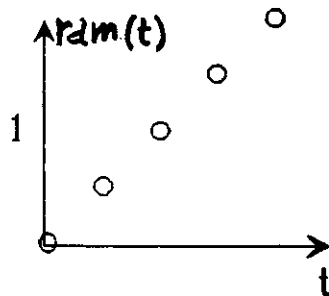
Esiste una formula analitica di scarso interesse pratico. In alternativa, nell'ipotesi che il numeratore sia di grado inferiore al denominatore, si può effettuare la divisione del primo per il secondo e ottenere delle potenze in z^{-i} $i=1,2,\dots$ i cui coefficienti sono, per definizione, i valori $f(i)$. In questo modo non si ottiene la forma analitica di f , ma i suoi valori nel tempo.

Un esempio

Si determini l'antitrasformata di

$$F(z) = \frac{z}{(z-1)^2} = z^{-1} + 2z^{-2} + 3z^{-3} + 4z^{-4} + \dots$$

si deduce che $F(z)$ è la trasformata zeta della funzione rampa $f(t) = \text{ram}(t) = t$



Infatti

z				$z^2 - 2z + 1$
$-z$	2	$-z^{-1}$		$z^{-1} + 2z^{-2} + 3z^{-3}$
	2	$-z^{-1}$		
	-2	$4z^{-1}$	$-2z^{-2}$	
		$3z^{-1}$	$-2z^{-2}$	
			...	

Proprietà

Linearità

$$Z(af(t)+bg(t))=aF(z)+bG(z)$$

Traslazione temporale

$$1) \quad Z(f(t-m))=z^{-m}F(z) \quad \Leftrightarrow \quad z^{-1}u(t)=u(t-1)$$

questa proprietà giustifica il significato operatoriale di operatore ritardo unitario attribuito a z^{-1}

$$2) \quad Z(f(t+1))=zF(z)-zf(0) \quad \Leftrightarrow \quad zu(t)=u(t+1)$$

questa proprietà giustifica il significato operatoriale di operatore anticipo unitario attribuito a z

Prodotto di convoluzione

$$Z(f(t)*g(t))=F(z)G(z)$$

dove * indica il prodotto di convoluzione

Derivazione a tempo discreto

$$Z(f(t)-f(t-1))=\frac{z-1}{z}F(z)$$

quindi $\frac{z-1}{z}$ (uno zero in $z=1$ e un polo nell'origine) rappresenta l'operatore derivazione (s nel continuo)

Integrazione a tempo discreto

$$Z\left(\sum_{h=0}^t f(h)\right)=\frac{z}{z-1}F(z)$$

quindi $\frac{z}{z-1}$ (un polo in $z=1$ e uno zero nell'origine) rappresenta l'operatore integrazione ($1/s$ nel continuo)

Teorema del valore iniziale

$$f(0) = \lim_{z \rightarrow \infty} \sum_{t=0}^{\infty} f(t)z^{-t} = \lim_{z \rightarrow \infty} F(z)$$

Teorema del valore finale

Se esiste finito il limite per $t \rightarrow \infty$ di $f(t)$, allora

$$\lim_{t \rightarrow \infty} f(t) = \lim_{z \rightarrow 1} (z-1)F(z)$$

Nel caso di segnali a trasformata razionale il limite esiste se tutti i poli sono in modulo minori di 1 o al più nel punto 1 del piano complesso.

TRASFORMATA DI FOURIER DI SEGNALI NON PERIODICI A TEMPO DISCRETO

Sia $f(t)$ una funzione reale a tempo discreto tale che

$$\sum_{t=-\infty}^{+\infty} |f(t)| < \infty$$

ciò significa che $f(t)$ non è periodica.

Si definisce la trasformata di Fourier continua di $f(t)$ come

$$F(e^{j\omega}) = \sum_{t=-\infty}^{+\infty} f(t)e^{-j\omega t} \quad , \quad \text{con } \omega \in [-\pi, \pi]$$

Data $F(e^{j\omega})$ esiste una formula di antitrasformazione.

Proprietà

Periodicità

$F(e^{j\omega})$ è periodica di periodo 2π

Linearità

$$\mathbb{F}(af(t) + bg(t)) = aF(e^{j\omega}) + bG(e^{j\omega})$$

Invarianza del modulo allo scorrimento temporale

$$\mathbb{F}(f(t-m)) = e^{-j\omega m} F(e^{j\omega})$$

Trasformata Zeta

Simmetria

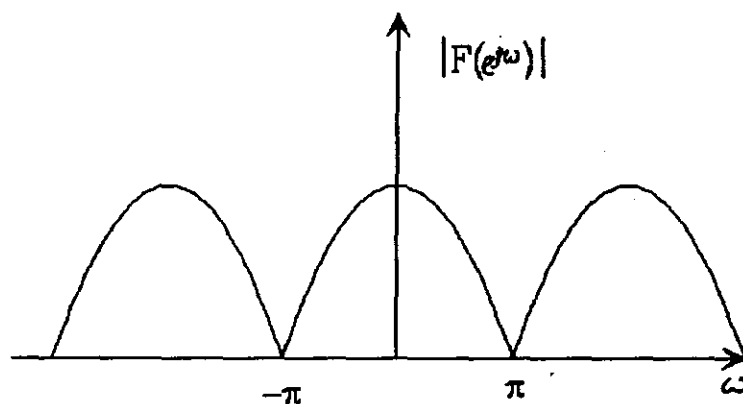
$F(e^{-j\omega}) = F^*(e^{j\omega})$ (* rappresenta il coniugato).

Questa proprietà, unitamente alla periodicità, significa che è sufficiente conoscere $F(e^{j\omega})$ soltanto per $\omega \in [0, \pi]$.

Se $f(t)$ è definita soltanto per $t \geq 0$ e si pone $f(t) = 0$ per $t < 0$, la trasformata di Fourier coincide con la trasformata Zeta ristretta alla circonferenza di raggio unitario, cioè $|z| = 1$, quindi $z = e^{j\omega}$.

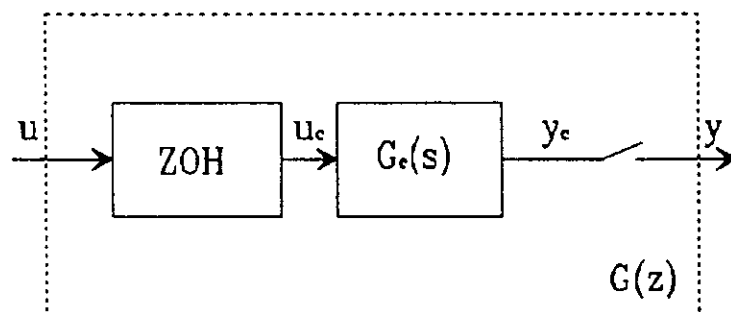
Rappresentazioni di $F(e^{j\omega})$

Spettro del modulo



analogamente si può rappresentare lo spettro della fase, o di $\arg(F(e^{j\omega}))$

SISTEMI A SEGNALI CAMPIONATI



- I modelli dinamici che normalmente vengono identificati sono a tempo discreto e rappresentano il sistema a tempo continuo, il campionatore e il mantenitore di ordine zero
- Si analizzano separatamente questi componenti per comprendere le caratteristiche del sistema che deve essere identificato
- Si discute la scelta del periodo di campionamento

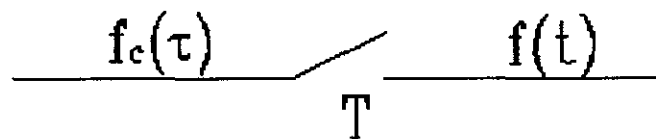
Il campionatore ideale

τ = tempo continuo

T = periodo di campionamento (campionamento uniforme)

$\omega_c = \frac{\omega}{T}$ = pulsazioni continue

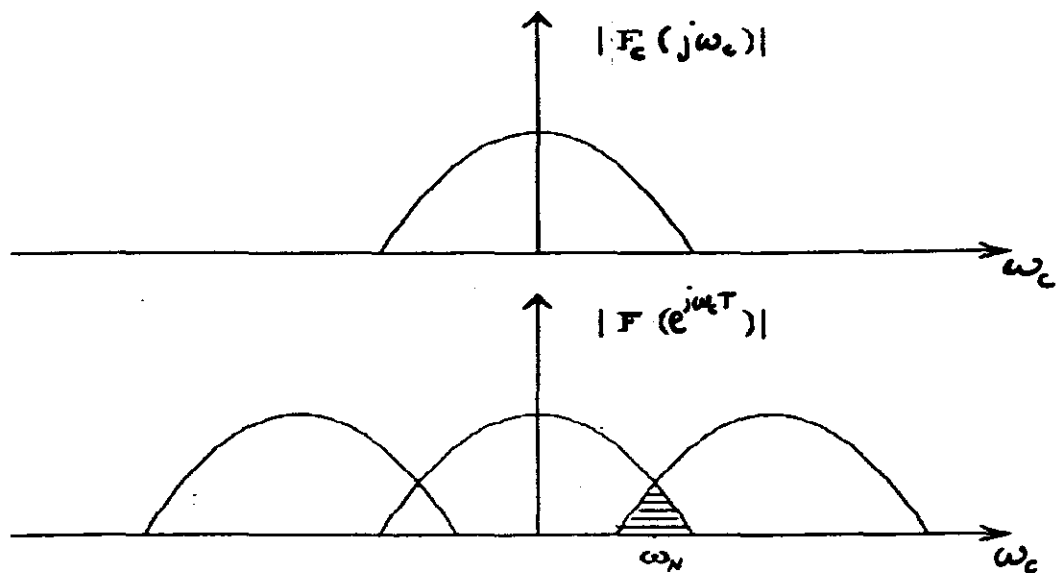
$\omega_N = \pi/T$ = pulsazione di Nyquist



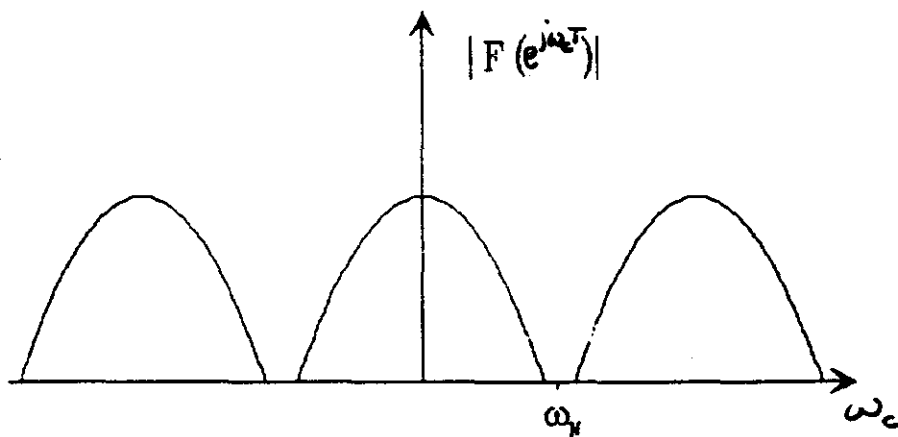
Legame tra la trasformata di Fourier $F_c(j\omega_c)$ del segnale a tempo continuo $f_c(\tau)$ e quella $F(e^{j\omega T}) = F(e^{j\omega_c T})$ del segnale campionato $f(t)$:

$$F(e^{j\omega T}) = \frac{1}{T} \sum_{k=-\infty}^{+\infty} F_c(j(\omega_c + 2k\omega_N))$$

Può essere presente il fenomeno dell'*aliasing*



per evitare l'aliasing $f_c(\tau)$ deve essere a spettro limitato, cioè $F_c(j\omega_c)=0$ per $\omega_c > \omega_s$ ed è necessario scegliere il periodo di campionamento T in modo che $\omega_s < \omega_N$. (teorema di Shannon)

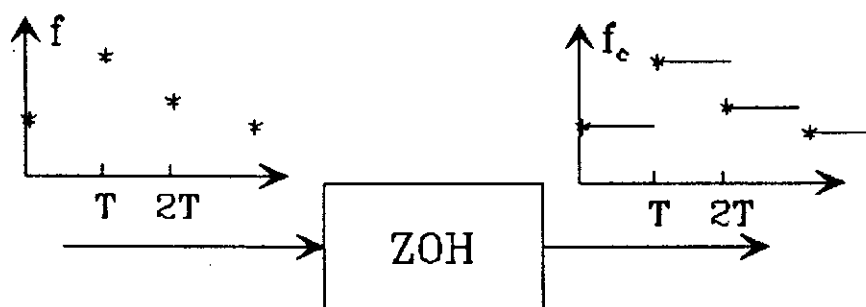


E' opportuno scegliere T piccolo, cioè allargare la pulsazione di Nyquist

Il mantentore di ordine zero

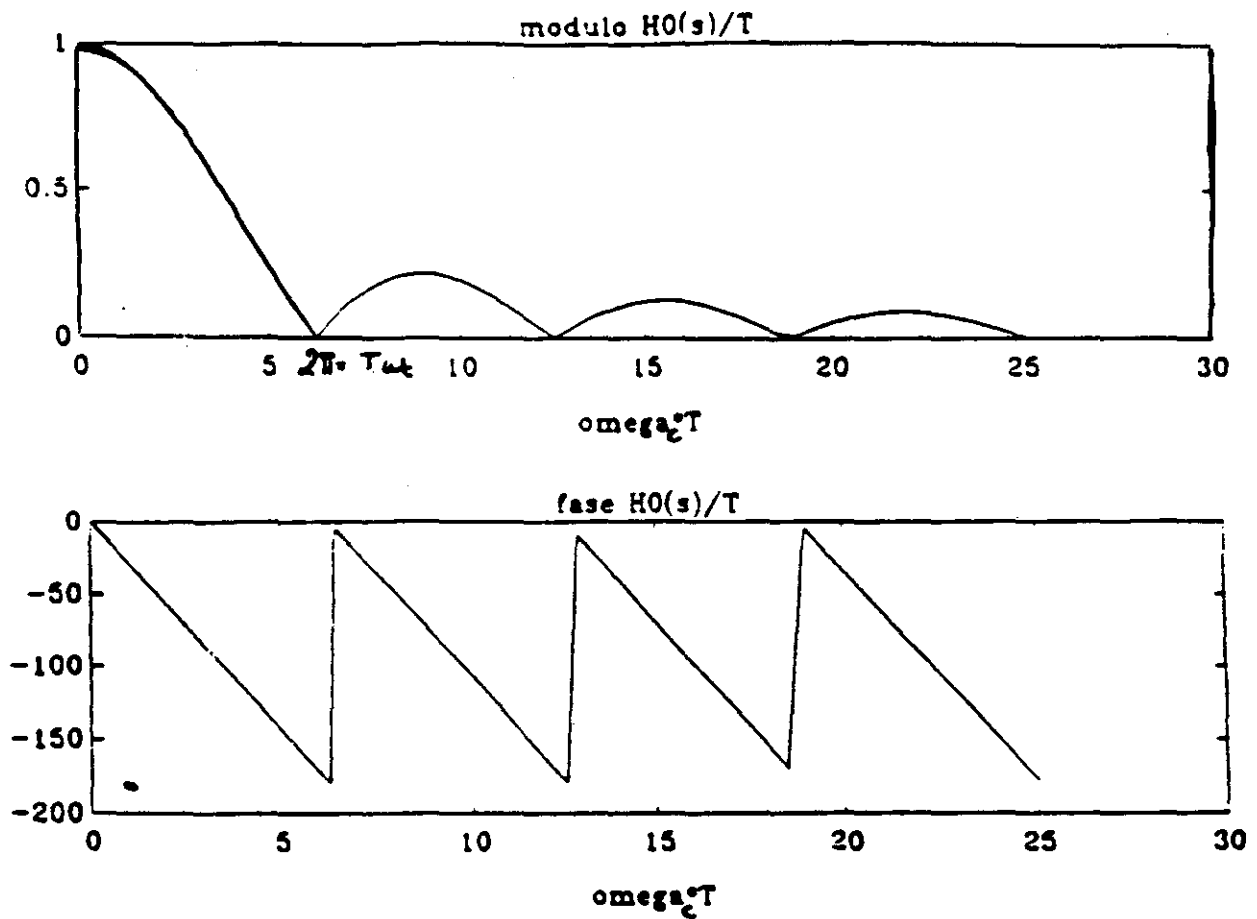
(Zero Order Hold - ZOH)

Consente di generare un segnale a tempo continuo a partire da un segnale a tempo discreto



Risposta in frequenza dello ZOH

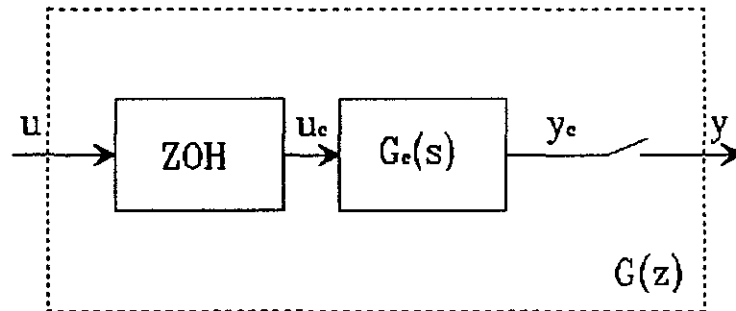
$$H(j\omega_c) = T e^{-\frac{j\omega_c T}{2}} \frac{\sin(\omega_c T / 2)}{\omega_c T / 2}$$



E' opportuno scegliere T piccolo per ridurre lo sfasamento negativo introdotto dallo ZOH

Sistemi a segnali campionati

Normalmente nell'identificazione dei sistemi dinamici si vuole determinare dai dati la funzione di trasferimento $G(z)$ di un modello a tempo discreto che rappresenti il sistema a tempo continuo con funzione di trasferimento $G_c(s)$, lo ZOH e il campionatore.



Si analizza ora il legame tra $G_c(s)$ e $G(z)$. Se $G_c(s)$ ammette la rappresentazione in variabili di stato

$$\dot{x}_c(\tau) = A_c x_c(\tau) + B_c u_c(\tau)$$

$$y_c(\tau) = C_c x_c(\tau)$$

negli istanti di campionamento, cioè nel tempo discreto t , il sistema dalla variabile discreta u alla variabile discreta y è descritto da

$$x(t+1) = Ax(t) + Bu(t)$$

$$y(t) = Cx(t)$$

dove $A = e^{A_c T}$, $B = \int_0^T e^{A_c \eta} B_c d\eta$, $C = C_c$. Da questo sistema si ricava direttamente la funzione di trasferimento a tempo discreto

$$G(z) = C(zI - A)^{-1}B$$

Legame tra i poli di $G_c(s)$ e i poli di $G(z)$

Se s è un polo di G_c , $z=e^{sT}$ (trasformazione di campionamento) è un polo di G . Per $T \rightarrow 0$ tutti i poli di G tendono al punto 1 (di fatto diventano tutti equivalenti a degli integratori). Questo in fase di identificazione può produrre gravi problemi numerici.

La trasformazione è biunivoca solamente a patto che tutti i poli di G_c abbiano parte immaginaria in modulo minore di π/T .

Effetto della quantizzazione sulla posizione dei poli

Supponiamo che la trasformazione sia biunivoca e che il polo di G_c in s si trasformi per effetto della quantizzazione in $z=e^{sT}(1+\varepsilon)$. Dalla trasformazione di campionamento inversa risulta che a z corrisponde nel continuo un polo $s'=s+\log(1+\varepsilon)/T$ che per $T \rightarrow 0$ si allontana indefinitamente dal polo originario.

Legame tra gli zeri di $G_c(s)$ e gli zeri di $G(z)$

Se G_c ha n poli e m zeri, con $m < n$, per $T \rightarrow 0$, G avrà m zeri che tendono al punto 1 e $n-m-l$ zeri che tendono a posizioni preassegnate sul semiasse reale negativo del piano complesso. Il fatto che prendendo un periodo di campionamento troppo piccolo gli zeri di G_c tendano anch'essi a 1 significa che l'identificazione corretta del modello a partire dalle misure di u e y sarà gravemente compromessa dall'addensarsi di zeri e poli nello stesso punto. Ci saranno problemi di quasi cancellazioni ed estrema sensibilità al rumore che corrompe i dati.

La struttura di G e di G_c può essere diversa:

$$G_c(s) = \frac{\mu}{s^n + \alpha_1 s^{n-1} + \dots + \alpha_n} \Rightarrow G(z) = \frac{\gamma_1 z^{n-1} + \dots + \gamma_n}{z^n + \beta_1 z^{n-1} + \dots + \beta_n}$$

Analizziamo ora le caratteristiche (stabilità, risposta allo scalino, risposta in frequenza) del sistema discreto con funzione di trasferimento $G(z)$ supponendo che il numero dei poli di G coincida con il numero di autovalori della matrice della dinamica A del sistema espresso in variabili di stato.

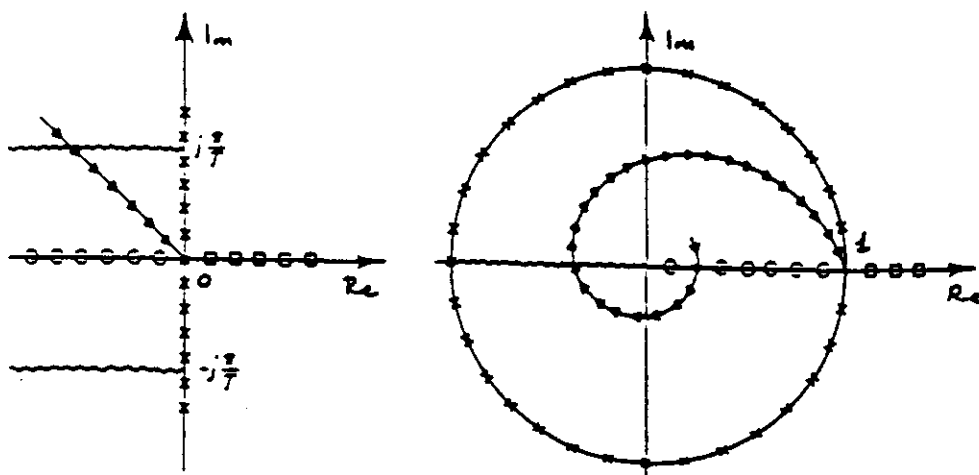
Stabilità

Il sistema è asintoticamente stabile se e solo se tutti i poli di $G(z)$ sono in modulo minori di 1, cioè se sono strettamente all'interno della circonferenza di raggio unitario centrata nell'origine del piano complesso Z .

Risposta allo scalino

Per valutare l'effetto della posizione dei poli sulla risposta allo scalino, cioè se la presenza di un polo in una data posizione produce transitori lenti o veloci, oscillanti o monotoni, conviene ricordare la trasformazione di campionamento $z=e^{sT}$ che lega il polo s nel continuo al polo z nel discreto. Questa trasformazione consente anche di trasformare particolari luoghi di punti nel piano S in altrettanti luoghi nel piano Z . Ad esempio il semiasse reale negativo in S corrisponde al segmento $(0,1)$ in Z (si ricorda che la trasformazione è biunivoca per poli con modulo della parte immaginaria minore di π/T).

PIANO S	PIANO Z
0	1
a. reali negativi	$0+1$
a. reali positivi	$1+\infty$
a. parte im. cost.	$(-1)+0$
a. smorz. cost.	spirali log



E' importante ricordare che $G(z)$ è comunque frutto di campionamento a T fissato per interpretare correttamente il concetto di "lento o veloce".

Ad esempio, si può verificare che la risposta allo scalino di un sistema con funzione di trasferimento

$$G_1(z) = \frac{0.7}{z - 0.3}$$

è monotona e supera il 90% del suo valore di regime in 3 campioni discreti, mentre la risposta del sistema con funzione di trasferimento

$$G_2(z) = \frac{0.2}{z - 0.8}$$

supera il 90% del suo valore di regime dopo 17 campioni. Può sembrare che G_2 sia più "lento" di G_1 , ma entrambe possono essere frutto del campionamento di

$$G_c(s) = \frac{1}{s + 1}$$

per G_1 si è utilizzato il periodo di campionamento $T_1 = 1.204$, mentre per G_2 si è utilizzato $T_2 = 0.223$. Evidentemente la velocità di risposta è la stessa.

Guadagno

Se $G(z)$ ha tutti i poli di modulo minore di 1, il guadagno del sistema si valuta ponendo $z=1$, cioè è dato da $G(1)$. Quindi a fronte di un ingresso costante U , il valore di regime dell'uscita è $Y=G(1)U$.

RISPOSTA IN FREQUENZA

Si definisce risposta in frequenza associata a $G(z)$ la funzione $G(e^{j\omega})$, con $\omega \in [0, \pi]$. Di essa si può dare una rappresentazione in termini di diagramma polare e/o di diagrammi di Bode.

A partire dalla risposta in frequenza si può enunciare anche nel caso discreto il Teorema della risposta in frequenza così come il criterio di Nyquist per sistemi reazionati.

Se $G(z)$ è frutto del campionamento di un sistema a tempo continuo con funzione di trasferimento $G_c(s)$ più ZOH e campionatore, la sua significatività si limita solo all'insieme di pulsazioni continue $\omega_c \in [0, \omega_N = \pi/T]$. Per confrontare la risposta in frequenza $G(e^{j\omega})$ con $G_c(j\omega_c)$ si può porre $\omega = T\omega_c$ e disegnare i diagrammi di Bode di G e G_c con $\omega_c \in [0, \omega_N]$.

Un esempio

Discretizzando il sistema a tempo continuo $G_c(s) = \frac{1}{(s+1)^2}$ con 2 diversi periodi di campionamento si ottiene

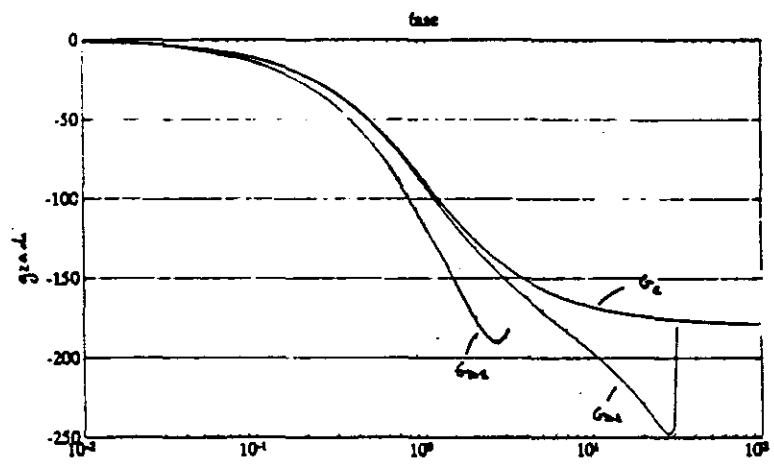
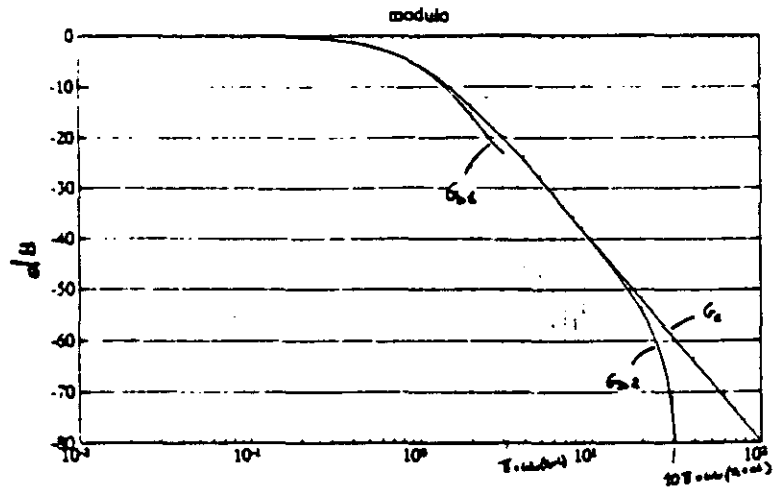
$$G_1(z) = \frac{0.264z + 0.135}{(z - 0.3679)^2} \quad \text{per } T_1 = 1$$

e

$$G_2(z) = \frac{0.0047z + 0.0044}{(z - 0.9048)^2} \quad \text{per } T_2 = 0.1$$

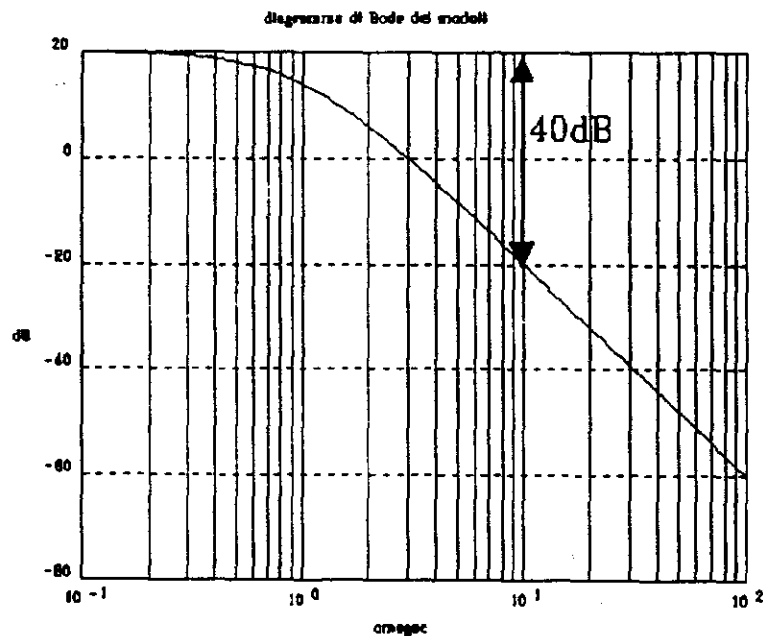
i corrispondenti diagrammi di Bode sono confrontati nel seguito. E' evidente che per $T \rightarrow 0$ risulta $\omega_N \rightarrow \infty$. Ciò significa che tutte le singolarità (poli e zeri) tendono a zero nella banda $[0, \omega_N]$.

Sistemi a segnali campionati



SCELTA DEL PERIODO DI CAMPIONAMENTO

Nella scelta di T da impiegare in fase di identificazione di $G(z)$ vi sono 2 esigenze contrastanti: ridurlo per soddisfare il Teorema del campionamento (e per poi progettare un regolatore "quasi analogico) e non prenderlo troppo piccolo per evitare che tutte le singolarità finiscano nel punto 1 con evidenti problemi numerici e di rumore. Inoltre l'identificazione può portare ad un modello valido soprattutto alle alte frequenze e poco significativo in bassa/media frequenza. D'altra parte per il progetto di un sistema di controllo è utile conoscere la risposta in frequenza non molto oltre la pulsazione di taglio (nell'ipotesi che il sistema controllato in anello chiuso abbia all'incirca la stessa velocità del sistema in anello aperto).



Un criterio è quello di scegliere ω_N in modo che si abbia un'attenuazione rispetto alle basse frequenze dai $20dB$ ai $40dB$.

Obiezione: come si fa se non si conosce il sistema?

Si può scegliere $T=S/(20\div30)$, dove S è il tempo di assestamento della risposta allo scalino del sistema da identificare. Per un sistema del I ordine questa scelta porta proprio a selezionare ω_N nell'intervallo specificato.

RAPPRESENTAZIONE DI $G(z)$ IN TERMINI DI UNA RELAZIONE INGRESSO/USCITA NEL TEMPO

Sfruttando le proprietà della trasformata Z è possibile determinare facilmente una rappresentazione ingresso/uscita nel tempo di un sistema con funzione di trasferimento $G(z)$. Infatti

$$Y(z) = G(z)U(z) = \frac{b_0 z^m + b_1 z^{m-1} + b_2 z^{m-2} + \dots + b_m}{z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_n} U(z), \quad m < n$$

è equivalente a

$$(z^n + a_1 z^{n-1} + a_2 z^{n-2} + \dots + a_n)Y(z) = (b_0 z^m + b_1 z^{m-1} + b_2 z^{m-2} + \dots + b_m)U(z)$$

cioè

$$Y(z) = -a_1 z^{-1}Y(z) - a_2 z^{-2}Y(z) - \dots - a_n z^{-n}Y(z) \\ + b_0 z^{-n+m}U(z) + b_1 z^{-n+m-1}U(z) + b_2 z^{-n+m-2}U(z) + \dots + b_m z^{-n}U(z)$$

o anche, ricordando il significato operatoriale di ritardo unitario di z^{-l}

$$y(t) = -a_1 y(t-1) - a_2 y(t-2) - \dots - a_n y(t-n) + b_0 u(t-n+m) + b_1 u(t-n+m-1) + \dots - b_m u(t-n)$$

E' altrettanto facile percorrere inversamente i passaggi per determinare la funzione di trasferimento $G(z)$ a partire da questa equazione alle differenze in u e y .

Nell'identificazione di modelli dinamici si fa riferimento all'espressione nel tempo per la stima dei coefficienti a_i e b_j , ma di fatto si stima la funzione di trasferimento

RAPPRESENTAZIONE *FIR*
(Finite Impulse Response)

Dividendo numeratore per denominatore la funzione di trasferimento $G(z)$ può essere scritta nella forma

$$G(z) = \frac{b_{n-1}z^{n-1} + b_2z^{n-2} + \dots + b_n}{z^n + a_1z^{n-1} + \dots + a_n} = g_1z^{-1} + g_2z^{-2} + g_3z^{-3} + \dots$$

Nel caso di sistemi asintoticamente stabili per $i \rightarrow \infty$ i coefficienti $g_i \rightarrow 0$ e per N sufficientemente grande si può ipotizzare

$$G(z) \cong G'(z) = g_1z^{-1} + g_2z^{-2} + g_3z^{-3} + \dots + g_Nz^{-N}$$

Quindi

$$Y(z) = G(z)U(z) \cong G'(z)U(z) = (g_1z^{-1} + g_2z^{-2} + \dots + g_Nz^{-N})U(z)$$

e per quanto visto

$$y(t) = g_1u(t-1) + g_2u(t-2) + \dots + g_Nu(t-N)$$

che giustifica il nome di sistemi con risposta all'impulso finita. Si può osservare che

$$G'(z) = \frac{g_1z^{N-1} + g_2z^{N-2} + g_3z^{N-3} + \dots + g_N}{z^N}$$

ha tutti i poli nell'origine del piano complesso, è cioè un sistema di tipo dead-beat. Si noti tuttavia che l'approssimazione di un sistema con un FIR comporta l'uso di un gran numero di parametri, cioè N deve in genere essere grande.

CONCLUSIONI

- Il periodo di campionamento T non deve essere troppo grande per ridurre il fenomeno dell'aliasing, ma neppure troppo piccolo per non avere problemi numerici
- Per la scelta di T (o della pulsazione di Nyquist $\omega_N = \pi/T$) si ricordi che se il modello da stimare deve essere utilizzato per il progetto del controllo, è bene che esso sia significativo soprattutto nell'intorno della pulsazione di taglio del sistema. In questo caso la conoscenza del sistema ad alta frequenza interessa poco. Si discuterà nel seguito come "centrare" la stima nell'intervallo di frequenze di interesse (oltre che con una buona scelta di T)
- La struttura di $G(z)$ può essere diversa da quella di $G_c(s)$ (per la presenza di zeri aggiuntivi)
- L'identificazione dei parametri di $G(z)$ viene fatta con riferimento ad un modello a tempo discreto

PROCESSI CASUALI STAZIONARI

Contenuti:

- Nozione di P.C. stazionario
- Medie del I e del II ordine
- Rumore bianco
- Spettro
- Fattorizzazione spettrale canonica
- Conclusioni

Motivazione: Nell'identificazione di modelli dinamici bisogna tener conto della presenza di segnali (disturbi) che non siamo in grado di descrivere deterministicamente.

NOZIONE DI P.C. STAZIONARIO

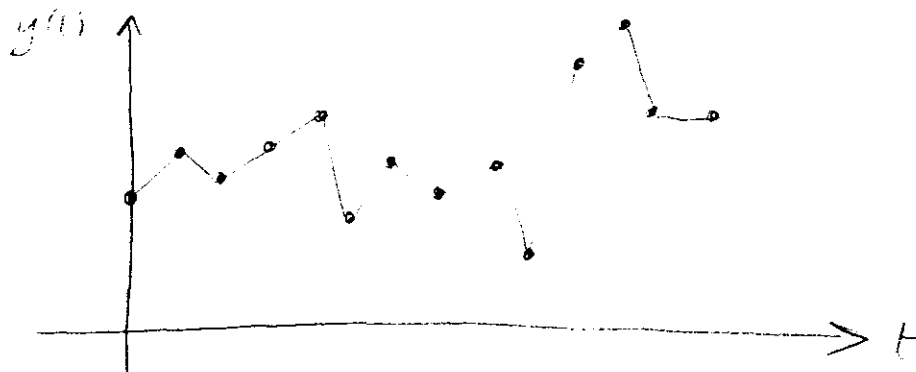
Processo Casuale (P.C.) $y(t)$: un esperimento casuale il cui esito è una funzione del tempo $y(t)$

Esempio 1: Lancio di una moneta. Se esce "testa" $y(t) = \sin(2\pi t/T)$, T intero. Se esce croce $y(t) = -\sin(2\pi t/T)$.

Esempio 2: $y(t+1) = y(t) + w(t)$, dove $w(t)$ è una V.C. indep. da $w(t+i)$, $i \neq 0$, con $P(w(t) = 1) = 0.5$, $P(w(t) = -1) = 0.5$. Il P.C. $y(t)$ è l'andamento dei beni di un giocatore che gioca sempre la stessa posta a testa o croce.

Esempio 3: Una qualsiasi serie di dati rilevata da un esperimento (temperatura in un motore, vibrazioni in un veicolo, etc.) E' sicuramente (almeno in parte) casuale perché ci sono gli errori di misura.

Nota: D'ora in poi verranno considerati P.C. a tempo discreto (l'indice temporale t è un numero intero).



Osservazioni:

- Il risultato di un esperimento casuale che produce un P.C. è una funzione del tempo $y(t)$ che prende il nome di *realizzazione*.
- Posso pensare al P.C. come ad un vettore di V.C. di lunghezza infinita.
- Se fisso un istante di tempo \bar{t} , allora $y(\bar{t})$ è una V.C.

P.C. stazionario $X(t)$: Se per un qualsiasi n , scelti comunque n istanti t_1, t_2, \dots, t_n , il vettore di V.C. $[y(t_1) y(t_2) \dots y(t_n)]^T$ ha la stessa d.d.p. di $[y(t_1 + \tau) y(t_2 + \tau) \dots y(t_n + \tau)]^T$ per ogni possibile valore intero di τ .

Interpretazione: Le caratteristiche probabilistiche del processo sono invarianti rispetto a traslazioni temporali.

Esempio 1: Per $t = kN$, $y(t) = 0 \Rightarrow$ non c'è incertezza. Poiché negli altri istanti c'è incertezza, il processo non è stazionario.

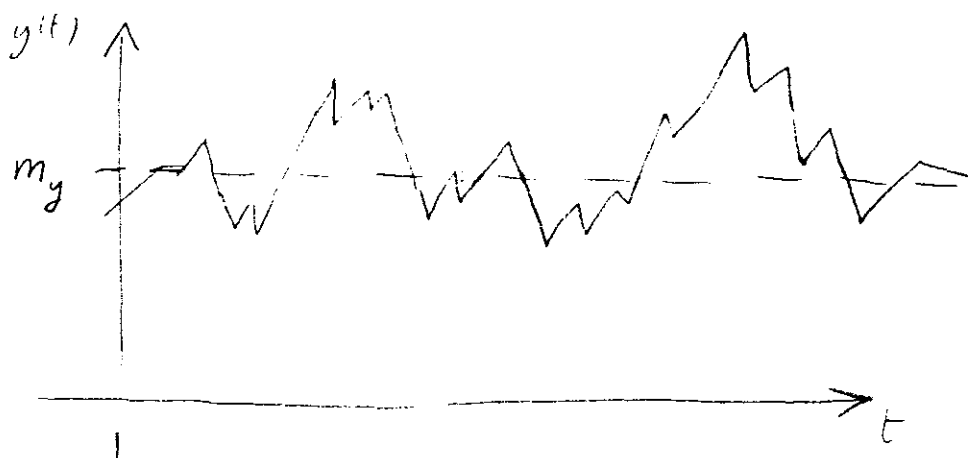
Esempio 2: $w(t)$ è una P.C. stazionario. Si può invece mostrare che $y(t)$ non lo è (la varianza di $y(t)$ cresce al crescere di t).

D'ora in poi verranno considerati solo P.C. stazionari.

MEDIE DEL I E DEL II ORDINE

Media: In generale $m_y(t) = E[y(t)]$ è una funzione di t . Se però $y(t)$ è un P.C. è stazionario, risulta $m_y(t) = m_y, \forall t$ (se $E[y(t)]$ dipendesse da t verrebbe violata l'invarianza rispetto alle traslazioni temporali).

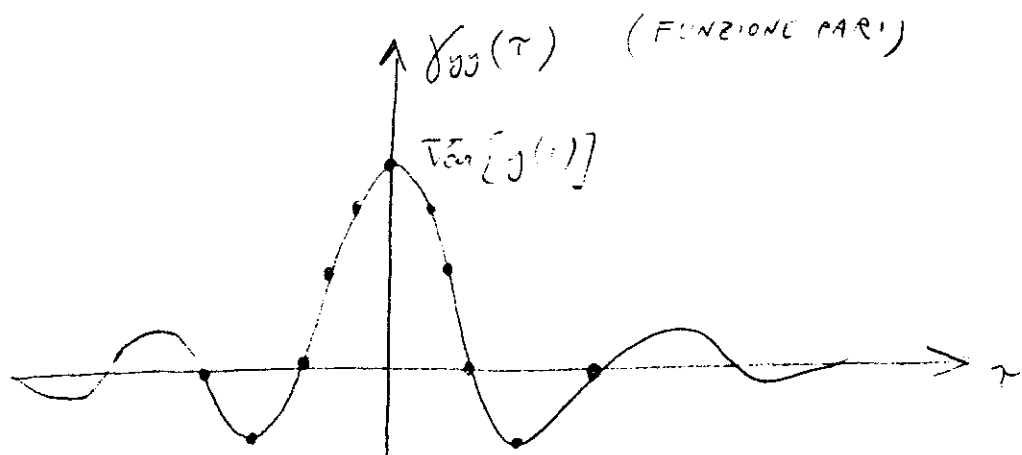
Per un P.C. stazionario è facile stimare m_y : basta calcolare la media campionaria di $y(t)$. Se considero $\tilde{y}(t) = y(t) - m_y$, vedo subito che $E[\tilde{y}(t)] = 0$. Con questo pretrattamento, ci si può sempre ricondurre a P.C. a media nulla (*nota: $y(t)$ e $\tilde{y}(t)$ hanno la stessa autocovarianza*)



Funzione di autocovarianza: $\gamma_{yy}(t_1, t_2) = \text{Cov}[y(t_1), y(t_2)]$. Se $y(t)$ è un P.C. stazionario, $\gamma_{yy}(t_1, t_2) = \gamma_{yy}(t_2 - t_1)$, $\forall t_1, t_2$.

Nota: $\gamma_{yy}(0) = \text{Var}[y(t)]$

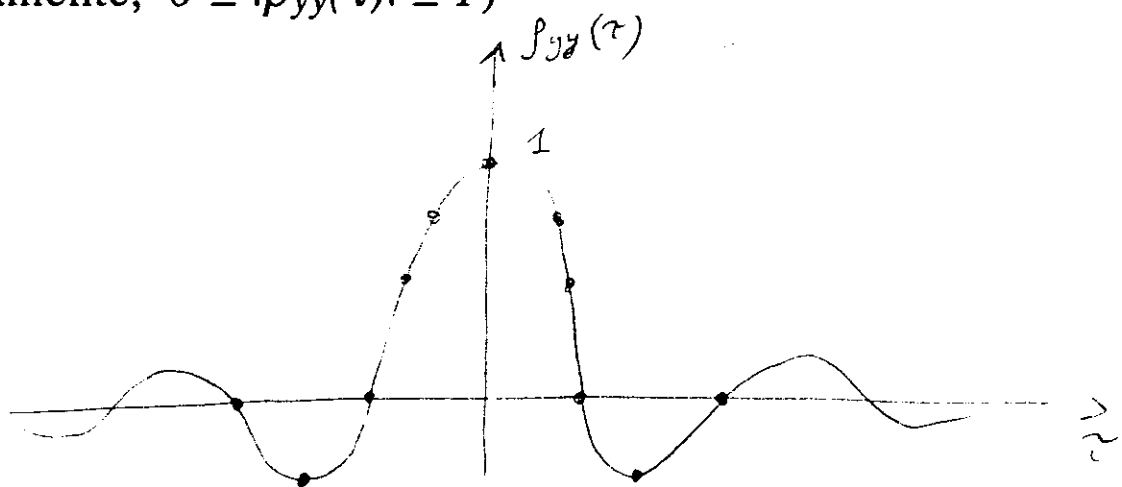
Proprietà: $|\gamma_{yy}(\tau)| \leq \gamma_{yy}(0)$, $\forall \tau$.



Funzione di autocovarianza normalizzata (per P.C. staz.):

$$\rho_{yy}(\tau) = \frac{\gamma_{yy}(\tau)}{\gamma_{yy}(0)}$$

(Ovviamente, $0 \leq |\rho_{yy}(\tau)| \leq 1$)



P.C. stazionario in senso debole: $y(t)$ è detto stazionario in senso debole quando $m_y(t) = m_y, \forall t, \gamma_{yy}(t_1, t_2) = \gamma_{yy}(t_2 - t_1), \forall t_1, t_2$.

Nota: Stazionarietà \Rightarrow stazionarietà in senso debole
Stazionarietà in senso debole $\not\Rightarrow$ stazionarietà

Processo casuale gaussiano: $y(t)$ è un P.C. gaussiano se per un qualsiasi n , scelti comunque n istanti t_1, t_2, \dots, t_n , il vettore di V.C. $[y(t_1) y(t_2) \dots y(t_n)]^T$ è distribuito gaussianamente.

Proprietà: Un P.C. gaussiano è completamente caratterizzato (in senso probabilistico) dalla conoscenza di $m_y(t)$ e $\gamma_{yy}(t_1, t_2)$.

Proprietà: Un P.C. gaussiano è stazionario se e solo se è stazionario in senso debole.

RUMORE BIANCO

Rumore bianco in senso stretto: $y(t)$ è un rumore bianco (*white noise*) se per un qualsiasi n , scelti comunque n istanti t_1, t_2, \dots, t_n , le V.C. $y(t_1), y(t_2), \dots, y(t_n)$ sono mutuamente indipendenti.

Interpretazione: E' come se $y(t)$ fosse il risultato di un esperimento casuale, il cui esito è del tutto indipendente dai valori precedenti e futuri del P.C.

Esempio: La sequenza dei numeri usciti alla roulette.

Un rumore bianco è un P.C. completamente imprevedibile. Conoscere i valori passati non mi dà nessuno spunto per prevedere quelli futuri. Migliore stima: $\hat{y}(t) = m_y$.

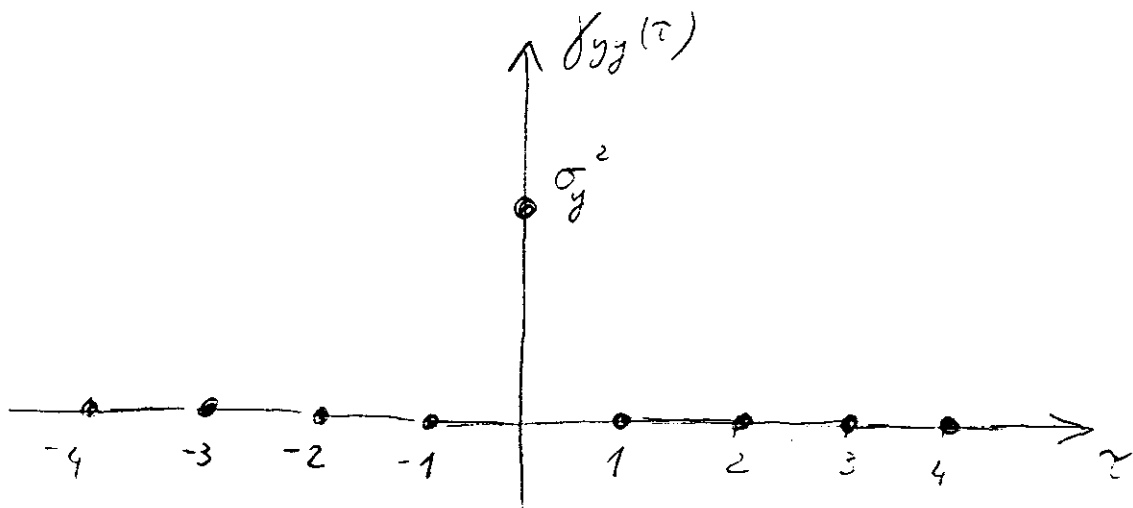
Rumore bianco (in senso lato): $y(t)$ è un rumore bianco in senso lato se $\gamma_{yy}(\tau) = 0, \tau \neq 0$ (mi accontento che $y(t_1)$ e $y(t_2)$ siano incorrelate $\forall t_1, t_2$).

Per P.C. gaussiani le nozioni di rumore bianco in senso stretto e senso lato sono equivalenti.

Notazione: $y(t) \sim WN(m_y, \sigma_y^2)$ (White Noise)
 $y(t) \sim WGN(m_y, \sigma_y^2)$ (White Gaussian Noise)

Funzione di autocovarianza del rumore bianco:

$$\gamma_{yy}(\tau) = \begin{cases} \sigma_y^2, & \tau = 0 \\ 0, & \tau \neq 0 \end{cases}$$



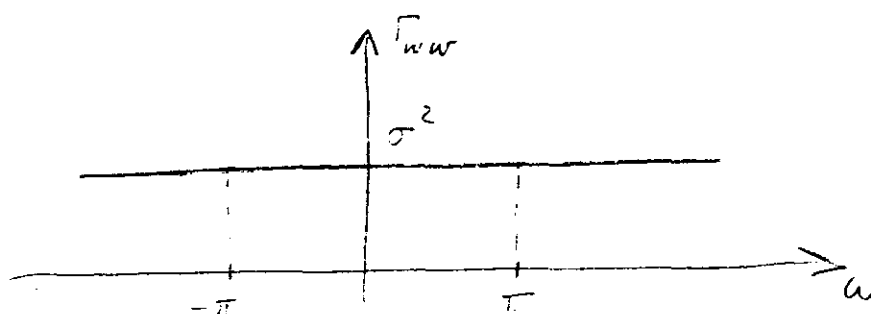
SPETTRO

Definizione: Lo spettro (o "densità spettrale di potenza") di un P.C. stazionario $y(t)$ è la trasformata di Fourier della funzione di autocovarianza

$$\Gamma_{yy}(\omega) = \mathcal{F}[\gamma_{yy}(\tau)] = \sum_{\tau=-\infty}^{+\infty} \gamma_{yy}(\tau) \exp(-j\omega\tau)$$

Esempio 1: $w(t) \sim WN(0, \sigma^2)$.

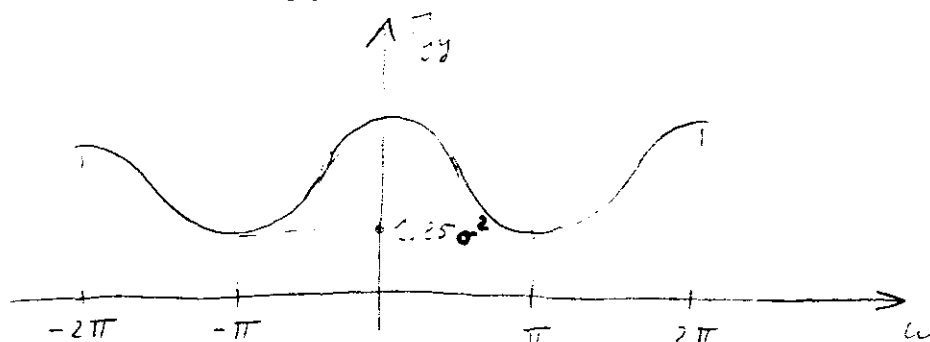
$$\Gamma_{ww}(\omega) = \mathcal{F}[\gamma_{ww}(\tau)] = \sigma^2$$



Spettro costante a tutte le frequenze: "luce bianca".

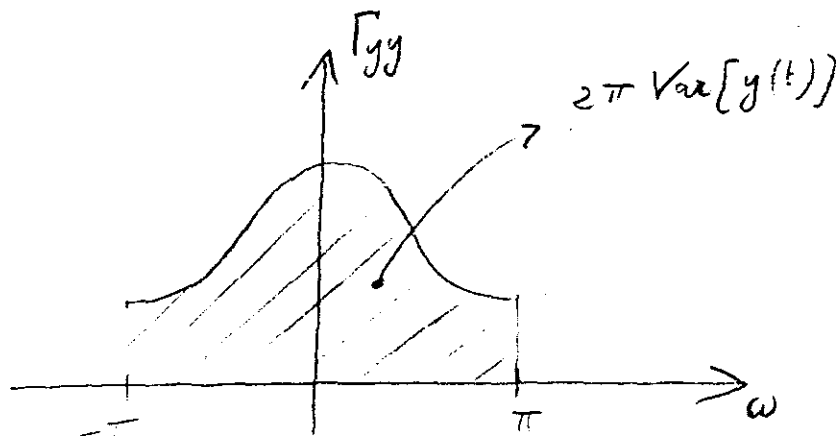
Esempio 2: $y(t) = w(t) + 0.5w(t-1)$, $w(t) \sim WN(0, \sigma^2)$.

Facendo i conti, risulta $\Gamma_{yy}(\omega) = [1.25 + \cos(\omega)]\sigma^2$



Proprietà dello spettro:

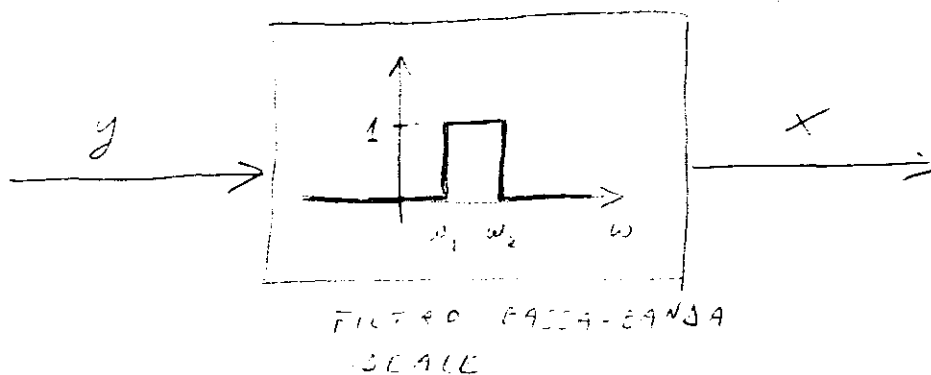
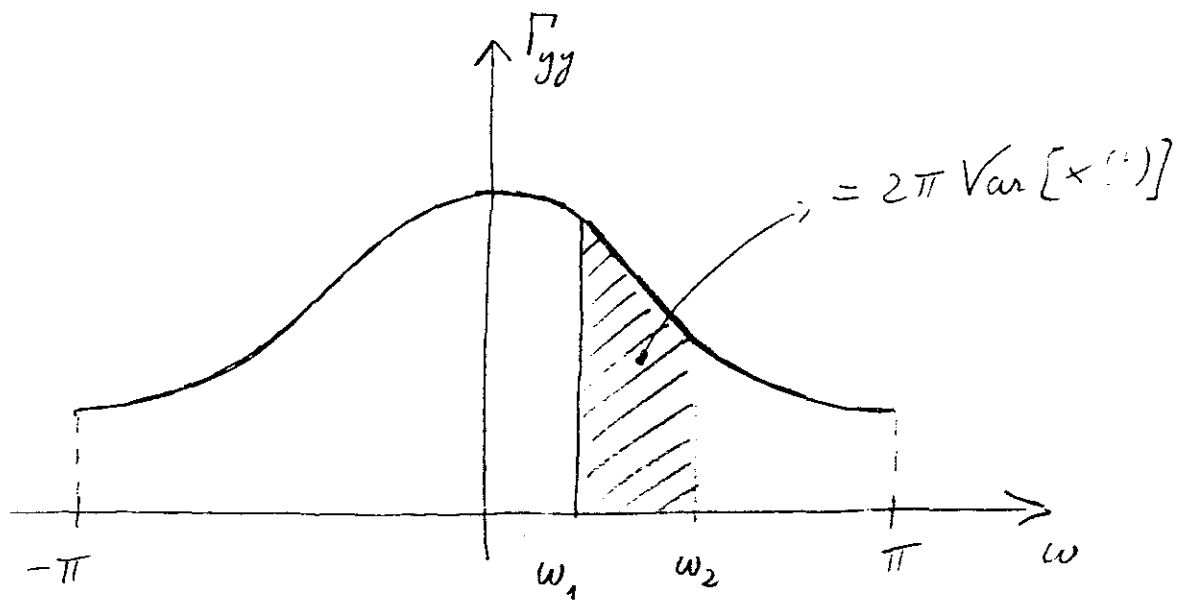
- $\Gamma_{yy}(\omega)$ è reale
- $\Gamma_{yy}(\omega) \geq 0, \forall \omega$
- $\Gamma_{yy}(-\omega) = \Gamma_{yy}(\omega)$ (funzione pari)
- $\Gamma_{yy}(\omega+2\pi) = \Gamma_{yy}(\omega)$ (2π -periodica)
- Dato che è pari e periodico basta conoscerlo per $\omega \in [0, \pi]$
- $\Gamma_{yy}(\omega) = \Phi_{yy}(ej\omega)$, dove $\Phi_{yy}(z) = Z[\gamma_{yy}(\tau)]$
- $\int_{-\pi}^{\pi} \Gamma_{yy}(\omega) d\omega = 2\pi \text{Var}[y(t)]$



Interpretazione: La quantità

$$\int_{-\omega_1}^{\omega_2} \Gamma_{yy}(\omega) d\omega$$

è proporzionale all'energia del processo concentrata nella banda di frequenze $[\omega_1, \omega_2]$.



Stima dello spettro (ipotesi: $E[y(t)] = 0$)

Idea base: Stimo $\gamma_{yy}(\tau)$ mediante l'autocovarianza campionaria e poi ne calcolo la trasformata di Fourier

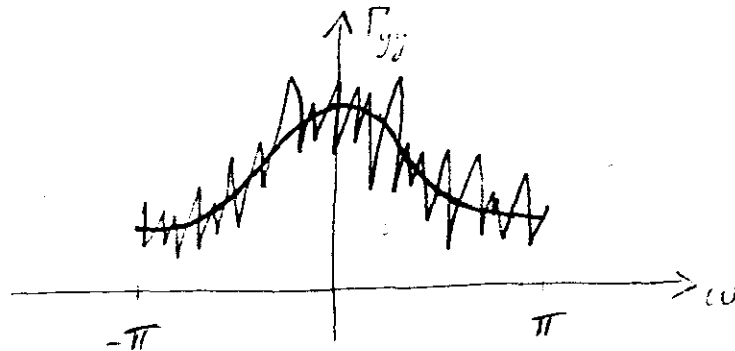


$$\text{"Periodogramma"} = \frac{1}{N} |FFT[y(t)]|^2$$

($FFT[y(t)]$): Fast Fourier Transform di $y(t)$, $t = 1, \dots, N$)

Proprietà del periodogramma:

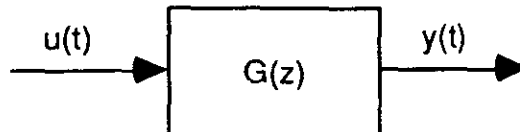
- E' uno stimatore asintoticamente ($N \rightarrow \infty$) non polarizzato dello spettro
- *Non è uno stimatore consistente.* La varianza dell'errore di stima non decresce al crescere di N (cerca di stimare N punti dello spettro avendo N dati \Rightarrow non riesce a convergere)



Metodo di Bartlett: Divido i dati in M spezzoni di lunghezza uguale, calcolo il periodogramma di ciascun spezzone ed infine uso come stimatore la media dei periodogrammi (perdo risoluzione ma diminuisce la varianza).

FATTORIZZAZIONE SPETTRALE CANONICA

Spettro dell'uscita di un sistema lineare

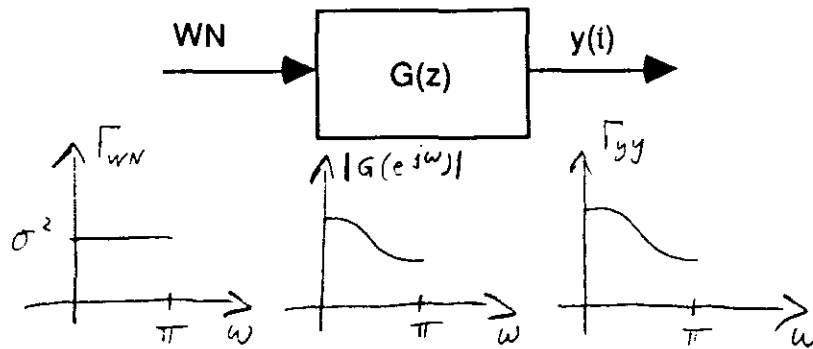


Teorema: Se $u(t)$ è un P.C. stazionario e $G(z)$ (rapporto di polinomi in z) è stabile, allora $y(t)$ converge ad un P.C. stazionario tale che:

- $E[y(t)] = G(1)E[u(t)]$
- $\Phi_{yy}(z) = G(z)G(z^{-1})\Phi_{uu}(z)$
- $\Gamma_{yy}(\omega) = |G(ej\omega)|^2 \Gamma_{uu}(\omega)$

Interpretazione: Filtrando un P.C. $u(t)$ mediante la f.d.t. $G(z)$ cambio la ripartizione dell'energia nelle varie bande di frequenza. Per esempio, se $G(z)$ è un passabasso (ideale), $y(t)$ non contiene più le componenti ad alta frequenza eventualmente presenti in $u(t)$.

Idea interessante: Dato un P.C. $y(t)$ con spettro $\Gamma_{yy}(\omega)$, se riesco a trovare $G(z)$ tale che $\Phi_{yy}(z) = \sigma^2 G(z)G(z^{-1})$ ($\Rightarrow \Gamma_{yy}(\omega) = \sigma^2 |G(e^{j\omega})|^2$) posso immaginare $y(t)$ come l'uscita di $G(z)$ alimentata da $WN(0, \sigma^2)$.



$G(z)$: "shaping filter"

Definizione: Quando $\Phi_{yy}(z) = \sigma^2 G(z)G(z^{-1})$, dove $G(z)$ è un rapporto di polinomi in z si dice che $y(t)$ è "a spettro razionale" e che $G(z)$ è un fattore spettrale.

Nota: Per simulare $y(t)$ basta simulare un WN (facile) e usarlo come ingresso di un filtro con f.d.t. $G(z)$.

Fattorizzazione spettrale canonica

Dato un P.C. stazionario $y(t)$ a spettro razionale, esistono sempre infiniti fattori spettrali \Rightarrow problemi di non unicit . Per la predizione e l'identificazione ne serve uno particolare.

Teorema: Dato un P.C. stazionario $y(t)$ a spettro razionale, esiste un unico fattore spettrale $(\hat{G}(z), \hat{\sigma}^2)$ (detto "canonico") tale che:

- il numeratore e il denominatore di $\hat{G}(z)$ hanno ugual grado;
- il numeratore e il denominatore di $\hat{G}(z)$ sono monici (il coeff. della potenza di grado massimo   = 1);
- le radici del denominatore di $\hat{G}(z)$ hanno tutte modulo < 1 ;
- le radici del numeratore di $\hat{G}(z)$ hanno tutte modulo ≤ 1 .

Esempio:

$$\Phi_{yy}(z) = \sigma^2 G(z)G(z^{-1})$$

$$G(z) = \frac{10(z+2)}{(z+0.3)(z+0.1)}, \quad \sigma^2 = 1$$

Il fattore $(G(z), \sigma^2)$ non è canonico perché

- numeratore e denominatore non sono monici;
- numeratore e denominatore non hanno ugual grado;
- il numeratore ha uno zero in $z = -2$.

Il fattore spettrale canonico è

$$\hat{G}(z) = \frac{z(z+0.5)}{(z+0.3)(z+0.1)}, \quad \hat{\sigma}^2 = 400$$

(è facile verificare che $\Phi_{yy}(z) = \hat{\sigma}^2 \hat{G}(z)\hat{G}(z^{-1})$).

CONCLUSIONI

- P.C. stazionari: caratterizzati da media e autocovarianza.
- In alternativa all'autocovarianza posso considerare lo spettro.
- Idea forte: sotto larghe ipotesi un P.C. stazionario può essere visto come l'uscita di $G(z)$ stabile alimentata da un rumore bianco



posso usare $G(z)$ (fattore spettrale) per caratterizzare il P.C.