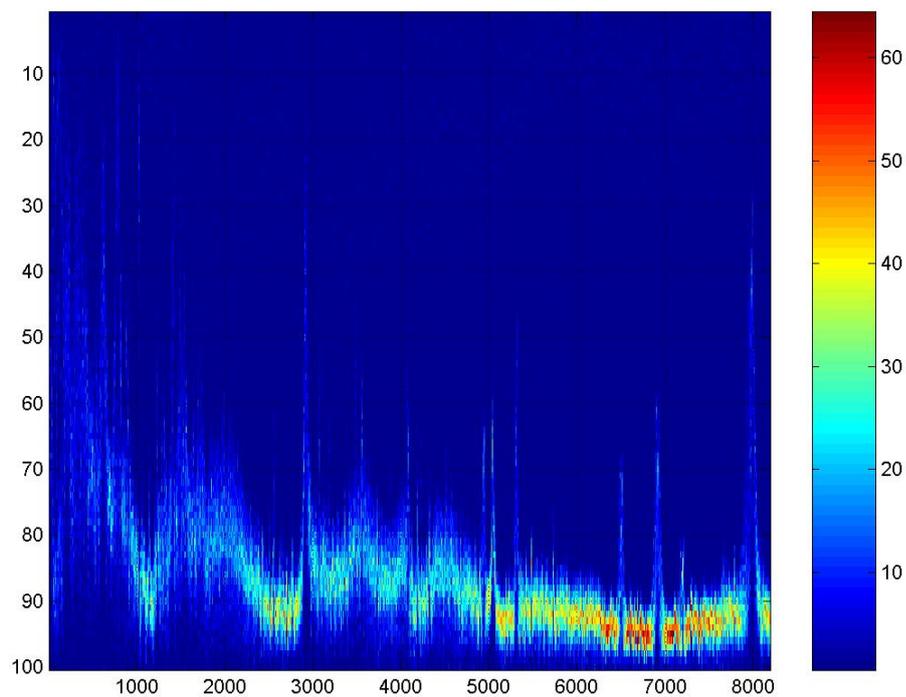


Sergio Frasca

Analisi dei Segnali

Dipartimento di Fisica
Università di Roma “La Sapienza”



Versione 13 dicembre 2011

Versione aggiornata in <http://grwavsf.roma1.infn.it/sp/sp.pdf>

Sommario

1	Introduzione: segnali e sistemi	7
1.1	Introduzione	7
1.2	Segnali.....	8
1.2.1	Classificazione dei segnali	8
1.2.2	Segnali notevoli	9
1.3	Sistemi.....	12
1.3.1	Classificazione dei sistemi	13
2	Argomenti introduttivi.....	16
2.1	Sistemi lineari continui	16
2.1.1	Trasformata di Laplace	18
2.1.2	Trasformata di Fourier	19
2.1.3	La funzione di trasferimento.....	26
2.1.4	Semplici sistemi lineari continui.....	29
2.2	Teoria delle probabilità.....	37
2.2.1	Variabili casuali	37
2.2.2	Variabili casuali discrete.....	37
2.2.3	Variabili casuali continue	39
2.2.4	Somma di variabili casuali indipendenti e funzione caratteristica.....	46
2.2.5	Distribuzione di una funzione di variabile casuale.....	48
2.2.6	Variabili casuali multiple e covarianza.....	49
2.3	Statistica.....	56
2.3.1	L'inferenza statistica - Stima dei parametri.....	56
2.3.2	Il Principio dei Minimi Quadrati (e del minimo χ^2).....	57
2.3.3	Il Principio della Massima Verosimiglianza.....	59
2.3.4	La Stima Bayesiana.....	62
2.3.5	Stima del valor medio	63
2.3.6	Stima della varianza	65
2.3.7	Esempi di stima ai minimi quadrati: stima dei parametri di una retta sperimentale ("fit lineare")	66
2.3.8	Esempi di stima ai minimi quadrati: fit lineare generale – fit polinomiale	69
2.3.9	Media pesata	71
2.3.10	Test statistici	72
2.3.11	Test di consistenza con un valore teorico	73
2.3.12	Test di consistenza tra due valori sperimentali	73
2.3.13	Test del χ^2	74
3	Sistemi discreti	76
3.1	Introduzione	76
3.2	Casi particolari	78
3.2.1	Semplice applicazione di sistemi MA e AR	79
3.3	Trasformata z	81
3.3.1	Analogia con la trasformata di Laplace	82
3.3.2	Proprietà della trasformata z	82
3.3.3	Alcune trasformate z	84
3.4	Trasformata di Fourier discreta ed FFT	85
3.4.1	La trasformata di Fourier per dati discreti (DTFT).....	85
3.4.2	La DFT (e la FFT).....	86
3.5	Equazioni alle differenze	87

3.6	Funzione di trasferimento discreta.....	88
3.6.1	Differenze e derivate.....	90
3.7	Risposta di un sistema discreto.....	92
3.7.1	Risposta impulsiva.....	92
3.7.2	Risposta forzata.....	92
3.7.3	Evoluzione libera.....	93
3.8	Stabilità.....	95
3.9	Sistemi semplici.....	96
3.9.1	Sistema di ordine 0.....	96
3.9.2	Sistema MA del primo ordine.....	96
3.9.3	Due sistemi MA del primo ordine in cascata: MA del secondo ordine.....	103
3.9.4	Sistema AR del primo ordine (reale).....	106
3.9.5	Sistema AR del primo ordine (complesso).....	113
3.9.6	Sistema AR del secondo ordine.....	116
3.9.7	Semplici sistemi ARMA.....	119
3.10	Sistemi non-lineari – i sistemi di Volterra.....	124
4	<i>Segnali transitori</i>	125
4.1	Energia, posizione e lunghezza di un segnale transitorio.....	125
4.2	Convoluzione, cross-correlazione e distanza di due impulsi.....	127
4.3	Trasformata di Fourier di un segnale ad energia finita.....	128
4.4	Autocorrelazione e spettro di energia.....	130
4.5	Segnale analitico.....	131
4.6	Segnali permanenti.....	134
5	<i>Processi stocastici</i>	135
5.1	Introduzione.....	135
5.1.1	Definizione.....	135
5.1.2	Funzioni del secondo ordine.....	136
5.1.3	Il caso di due processi.....	137
5.1.4	Stazionarietà, ergodicità.....	138
5.1.5	Esempi.....	141
5.2	Trasformazioni di processi stocastici.....	141
5.2.1	Sistema statico (senza memoria).....	142
5.2.2	Sistema lineare (tempo invariante).....	142
5.2.3	Un caso particolare: il derivatore.....	144
5.3	Processi stocastici normali.....	145
5.3.1	Proprietà fondamentali.....	145
5.3.2	Il rumore bianco.....	146
5.3.3	Processi stocastici normali e sistemi lineari.....	146
5.4	Processi discreti: processi ARMA.....	148
5.4.1	Résumé dei risultati.....	148
5.4.2	Rumore bianco discreto.....	150
5.4.3	Processi AR, MA e ARMA.....	153
5.4.4	Processo AR del primo ordine (reale).....	153
5.4.5	Processo AR del primo ordine (complesso).....	155
5.4.6	Processo AR del secondo ordine.....	158
5.5	Processo di Poisson.....	161
6	<i>Analisi statistica dei segnali</i>	163
6.1	Il campionamento.....	164
6.1.1	Teorema del campionamento.....	164
6.1.2	Aliasing e filtri anti-aliasing.....	166

6.1.3	Generalizzazione.....	167
6.1.4	Il caso dei segnali complessi.....	167
6.1.5	Rumore di quantizzazione.....	168
6.2	Caratteristiche statiche – Istogramma, momenti campionari.....	170
6.3	Autocorrelazione.....	172
6.4	Spettro di potenza.....	174
6.4.1	Stimatori spettrali non parametrici.....	175
6.4.2	Stimatori spettrali parametrici.....	187
6.5	Cross-correlazione e cross-spettro.....	189
6.6	Coerenza.....	190
7	<i>Filtraggio e trasformazione dei dati</i>	191
7.1	Segnali e rumori, rapporto segnale/rumore.....	191
7.2	Il filtro adattato.....	193
7.2.1	Caso del rumore bianco.....	193
7.2.2	Altre dimostrazioni.....	196
7.2.3	Caso generale.....	197
7.3	Teoria della rivelazione (Detection Theory).....	200
7.4	Filtro di Wiener.....	204
7.5	Realizzazione di filtri.....	206
7.5.1	Filtri FIR.....	206
7.5.2	Filtri IIR.....	206
7.5.3	Filtri a sfasamento nullo.....	206
7.5.4	Filtri in frequenza.....	207
7.6	I disturbi impulsivi - Filtro a mediana.....	209
8	<i>Cenno alla trasformata wavelet</i>	210
9	<i>Cenni al caso di dati non stazionari</i>	211
9.1	Analisi tempo-frequenza.....	212
9.2	Filtri adattivi.....	213
10	<i>Cenni ai modelli non lineari</i>	214
10.1	Filtri non lineari.....	215
10.2	Il bispettro.....	216
11	<i>Cenno all' image processing</i>	226
11.1	Immagini ed elaborazione delle immagini.....	226
11.2	Elaborazione lineare delle immagini.....	230
11.3	La compressione JPEG.....	237
12	<i>Un'applicazione: la ricerca di periodicità</i>	238
12.1	Introduzione al problema.....	238
12.2	Il periodogramma.....	240
12.3	Metodi parametrici.....	242
12.4	Il lock-in.....	243
12.5	L'analisi di fase.....	246
12.6	Il fit armonico.....	247
12.7	Il caso del campionamento non uniforme.....	248
12.8	Periodicità non coerenti.....	250
12.9	Periodicità di eventi.....	251
13	Appendici.....	252
	Esercitazioni con Matlab.....	252
	Introduzione a Matlab.....	252
	Snag.....	255
	Uso di SnagLab (versione VB).....	256

Introduzione	256
Istallazione	256
Uso	256
Tabelle.....	261
Distribuzione cumulativa normale standardizzata	261
Valori del χ^2 per un dato livello di fiducia.....	262
Bibliografia	263
Generale	263
Segnali deterministici – Filtraggio.....	263
Segnali stocastici.....	264
Aspetti matematici	265
Immagini	266
Altro	267
Indice Analitico.....	268

1 Introduzione: segnali e sistemi

1.1 Introduzione

Si supponga di osservare, con opportuni strumenti di misura, una o più grandezze relative a un certo fenomeno fisico. Queste misure sono in genere dei segnali variabili nel tempo.

L'analisi di questi segnali consiste in genere in:

- Descrivere in modo sintetico e significativo le grandezze in esame ed eventualmente la loro variazione temporale.
- Correlare¹ tra loro le varie grandezze, trovando eventuali relazioni di causalità o di concausalità.
- Indagare su parametri interni del processo generatore.
- Ridurre l'eventuale “rumore” presente nei dati di misura.

Per far ciò si utilizzano **modelli** matematici che rappresentano (soltanto) le caratteristiche di nostro interesse (cioè “semplificano” la realtà) e ci si può utilmente avvalere di una serie di **strumenti**, una specie di “cassetta degli attrezzi” (toolbox) che potranno essere utilizzati a seconda della necessità.

Per lo sviluppo di questi modelli e strumenti ci si avvale della teoria dei sistemi e della teoria dei processi stocastici.

L'analisi dei segnali ha avuto negli ultimi cinquanta anni un grande sviluppo, soprattutto per la diffusione dei calcolatori digitali. Ma i segnali, che vengono in genere prodotti in modo continuo dai processi fisici, per essere “trattati” dai calcolatori digitali devono essere discretizzati. La matematica per trattare questo tipo di dati è in genere abbastanza diversa dall'analisi classica.

Lo scopo di questo corso è presentare sinteticamente gli elementi della teoria dei sistemi discreti e la teoria dei processi stocastici che sono utili per l'analisi e l'elaborazione dei segnali. Sono quindi presentati vari strumenti sviluppati per questi scopi.

¹ Qui il termine “correlare” è usato in senso lato, nel significato di cercare relazioni e dipendenze statistiche tra di loro: non necessariamente “calcolare il coefficiente di correlazione” o la correlazione incrociata.

1.2 Segnali

Un segnale è una funzione del tempo² che rappresenta una grandezza fisica.

Un segnale è detto **a tempo continuo**, o semplicemente **continuo**, se è definito per tutti i valori di un intervallo della variabile reale tempo³ (t); per esempio una tensione elettrica

$$(1.1) \quad x(t) \quad \text{per } t_{\text{iniziale}} \leq t \leq t_{\text{finale}}$$

È detto **discreto** se è definito per un insieme discreto dei valori della variabile t ; in tal caso in genere viene rappresentato dalla successione di questi valori

$$(1.2) \quad x_1, x_2, \dots, x_N$$

$\{x_1, x_2, \dots, x_N\}$ è detta **successione o serie temporale**.

Talora gli indici non partono da 1, ma possono prendere anche valori negativi (e ovviamente 0). Spesso un segnale discreto viene ottenuto “campionando” un segnale continuo, cioè estraendo i valori del segnale continuo ad un insieme discreto di valori del tempo t_1, t_2, \dots, t_N ; considerando il segnale continuo $x(t)$, otteniamo i valori

$$(1.3) \quad x(t_1), x(t_2), \dots, x(t_N)$$

Tali valori sono detti **campioni** (samples) e possono essere indicati come nella (1.2). Tale procedura viene detta **campionamento** (sampling); particolarmente usato è il **campionamento uniforme**, in cui è costante la differenza $\Delta t = t_{i+1} - t_i$, detta “tempo di campionamento” (sampling time). Viene definita la **frequenza di campionamento** (sampling frequency) come

$$(1.4) \quad \nu_s = \frac{1}{\Delta t}$$

Nel seguito, considerando segnali discreti, faremo riferimento implicitamente sempre a segnali campionati uniformemente.

1.2.1 Classificazione dei segnali

Oltre alla classificazione tra segnali continui e discreti, esistono altre caratteristiche peculiari di un segnale:

² Si fa differenza tra un **tempo assoluto** e un **tempo relativo**, ovvero una **differenza temporale**. Un segnale può essere osservato a un preciso tempo (tempo assoluto), ma in genere viene descritto in tempo relativo, per esempio prendendo l'origine temporale all'inizio del segnale.

³ Si distingue in genere tra *tempo assoluto* (indicata in genere con t) e *distanza temporale* (in genere indicata con τ). In questo caso parliamo di tempo assoluto.

- **Reale** o **complesso**, cioè a un dato istante il segnale è un numero reale o un numero complesso.
- **Scalare** o **vettoriale**, cioè se è descritto da un singolo numero ad ogni istante, o da più numeri.
- **Analogico** o **digitale**, cioè il segnale ad un dato istante può assumere qualsiasi valore reale in un dato intervallo (segnale analogico) o solo un numero discreto di valori (segnale digitale), tipicamente multipli di un certo valore detto “quanto di conversione analogico-digitale”: è questo il caso dei segnali acquisiti da un sistema digitale tramite un convertitore ADC. In genere i segnali digitali sono anche discreti nel tempo.
- **Periodico**, se esiste un valore τ tale che, per tutti i valori di t , se il segnale è continuo (τ è un numero reale)

$$(1.5) \quad x(t) = x(t + \tau)$$

e se il segnale è discreto⁴ (τ è un numero intero)

$$(1.6) \quad x_i = x_{i+\tau}$$

- **Deterministico** o **casuale** (o **stocastico**): questa classificazione riguarda il “modello” del segnale: nel primo caso se è completamente definito a priori il suo valore, nel secondo se è definito solo statisticamente. I segnali deterministici possono godere di particolari simmetrie rispetto a un certo istante, per esempio all’istante $t=0$: sono detti pari se $x(t)=x(-t)$ e dispari se $x(t)=-x(-t)$.

1.2.2 Segnali notevoli

- **Segnale costante**

Continuo

$$x(t) = c$$

Discreto

$$x_i = c$$

quando $c = 0$, diciamo che c’è assenza di segnale.

- **Segnale a gradino**

Continuo

$$\begin{cases} t < 0 \rightarrow u(t) = 0 \\ t \geq 0 \rightarrow u(t) = 1 \end{cases}$$

Discreto

$$\begin{cases} i < 0 \rightarrow u_i = 0 \\ i \geq 0 \rightarrow u_i = 1 \end{cases}$$

⁴ In effetti questa definizione è restrittiva: se campioniamo un segnale continuo periodico, abbiamo in genere un segnale discreto che non è periodico secondo questa definizione.

- **Segnale a delta**; è la “derivata” (o la differenza nel caso discreto) del segnale a gradino. Nel caso continuo è una delta di Dirac, nel caso discreto è un segnale sempre nullo, ma che vale 1 in 0, detto anche **impulso unitario** o **funzione impulsiva discreta**:

Continuo

$$\delta(t)$$

Discreto

$$\begin{cases} i = 0 \rightarrow \delta_i = 1 \\ i \neq 0 \rightarrow \delta_i = 0 \end{cases}$$

e quindi

$$u(t) = \int_{-\infty}^t \delta(\kappa) \cdot d\kappa$$

$$u_i = \sum_{k=-\infty}^i \delta_k$$

- **Segnale sinusoidale**

Continuo

$$x(t) = A \cdot \sin(\omega \cdot t + \varphi)$$

Discreto

$$x_i = A \cdot \sin(\omega \cdot \Delta t \cdot i + \varphi)$$

Nella figura il segnale continuo è indicato con la linea continua e quello discreto con i pallini:

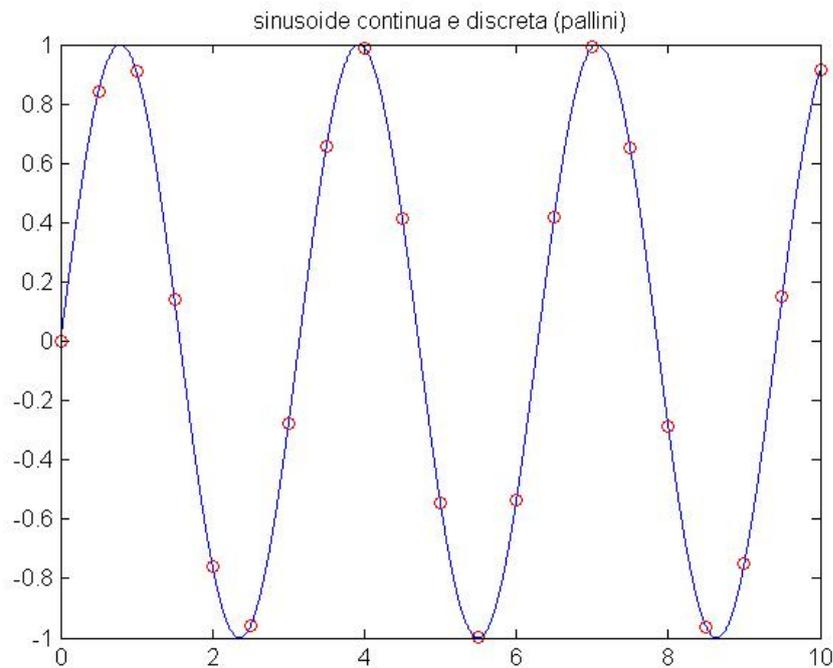


Figura 1-1

- **Segnale esponenziale complesso**⁵

Continuo

$$x(t) = A \cdot \exp\left[-\frac{t}{\tau} + j \cdot (\omega \cdot t + \varphi)\right]$$

Discreto

$$x_i = A \cdot \exp\left[-\frac{\Delta t \cdot i}{\tau} + j \cdot (\omega \cdot \Delta t \cdot i + \varphi)\right]$$

- **Rumore bianco**, il più semplice segnale stocastico, se ne parlerà nel capitolo sui processi stocastici.

⁵ Indichiamo con j l'unità immaginaria.

1.3 Sistemi

Un sistema è un modello matematico di un dispositivo o di un processo fisico o di un algoritmo, che connette uno o più segnali di ingresso (eccitazione o “input”) a uno o più segnali di uscita (risposta o “output”). **Un sistema quindi è sostanzialmente un elaboratore di segnali.**



Figura 1-2

Mentre i processi fisici sono modellizzabili più naturalmente con sistemi continui ed i segnali fisici con segnali continui, il loro trattamento con i calcolatori digitali rende fondamentale lo studio di sistemi e segnali discreti.

Più sistemi possono essere connessi insieme, e questo insieme può essere rappresentato da un nuovo sistema. Nella figura seguente è rappresentato un insieme di sistemi connessi tra di loro in vario modo.

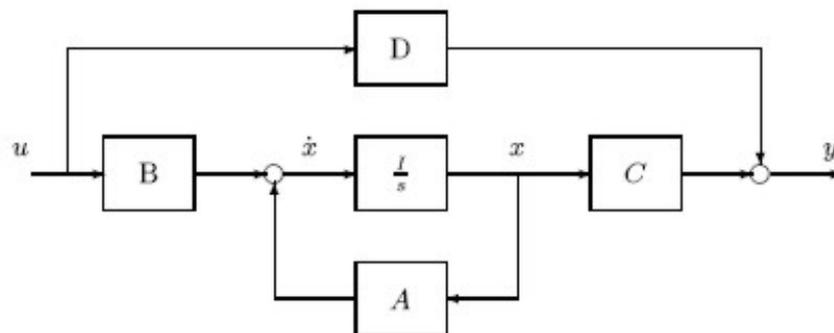


Figura 1-3

Si notino i simboli circolari che indicano i “nodi somma”, semplici sistemi con due o più ingressi e una sola uscita che eseguono la somma dei segnali di ingresso.

Modi particolarmente semplici, e spesso utili, di connettere dei sistemi sono i due seguenti:

- Connessione **in serie** (o **in cascata**)

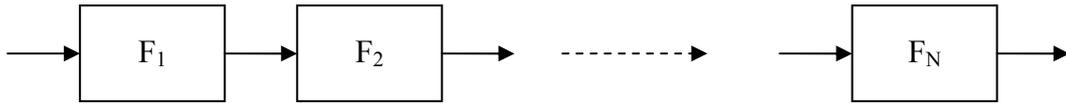


Figura 1-4

- Connessione **in parallelo**

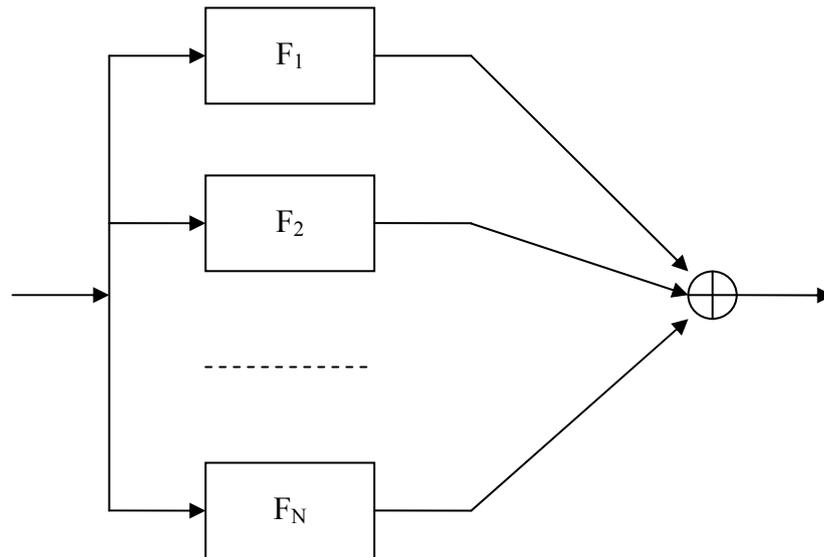


Figura 1-5

Attenzione ! Nella teoria dei sistemi il comportamento di un sistema non è modificato dalla presenza degli altri a cui è connesso. Come è noto, non è così per un normale sistema fisico (basti pensare ad un circuito RC passa-basso, la cui risposta dipende dall'impedenza d'ingresso del sistema che lo segue). Per sistemi elettrici, in cui le variabili di ingresso e di uscita sono tensioni elettriche, questa caratteristica equivale ad una impedenza d'ingresso infinita e un'impedenza di uscita nulla.

Come vedremo in questo corso, segnali e sistemi sono correlati tra di loro: i sistemi non ci serviranno solo per elaborare segnali, ma, in certi casi, per rappresentarli.

1.3.1 Classificazione dei sistemi

Oltre alla classificazione tra sistemi continui e discreti (cioè che elaborano segnali discreti o continui), esistono altre caratteristiche peculiari di un sistema:

- **Statico** o **dinamico**, detto anche senza memoria o con memoria, cioè se l'uscita ad un dato istante dipende solo dall'ingresso a quell'istante o no. Un sistema dinamico è caratterizzato dalla presenza di stati interni. In altri termini l'effetto della "memoria" del sistema viene schematizzato supponendo che l'uscita a un dato istante non dipende solo dall'ingresso a quell'istante, ma anche dal valore dello "stato". Quindi, mentre un sistema statico è completamente definito dall'equazione

$$(1.7) \quad \vec{y} = F(\vec{x})$$

dove \vec{x} e \vec{y} sono rispettivamente il vettore degli ingressi e quello delle uscite, ad un dato istante. Se il sistema è dinamico, invece, nel caso di sistemi continui, occorrono le due equazioni

$$(1.8) \quad \begin{cases} \vec{y} = F(\vec{x}, \vec{s}) \\ \dot{\vec{s}} = G(\vec{x}, \vec{s}) \end{cases}$$

dove \vec{s} è il vettore di stato in quell'istante. La prima equazione viene detta equazione di uscita (**output equation**), la seconda equazione di stato (**state equation**). Nel caso di sistemi discreti,

$$(1.9) \quad \begin{cases} \vec{y}_i = F(\vec{x}_i, \vec{s}_i) \\ \vec{s}_i = G(\vec{x}_i, \vec{s}_{i-1}) \end{cases}$$

Un esempio di sistema dinamico è un circuito RC passa-basso, in cui l'ingresso e l'uscita sia tensioni: in tal caso la variabile di stato è la carica del condensatore⁶.

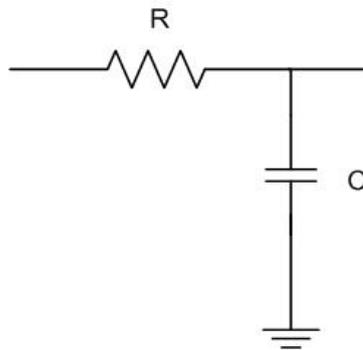


Figura 1-6

- **Causale**, se l'uscita non precede l'ingresso. Tutti i sistemi fisici sono causali, non così quelli simulati su calcolatore.
- **Lineare**, se ad una qualsiasi combinazione lineare di differenti ingressi, corrisponde in uscita la stessa combinazione lineare delle relative uscite (principio di sovrapposizione). Indichiamo con L l'operatore che descrive il sistema. Se il sistema è lineare, allora

⁶ Il circuito in figura non ha le caratteristiche ideali di un sistema: infatti il suo "comportamento" varia a seconda di cosa mettiamo alla sua uscita ed inoltre esso "carica" il generatore che mettiamo all'ingresso. Se supponiamo che all'ingresso ci sia un generatore di impedenza di uscita nulla e all'uscita un apparato di impedenza di ingresso infinita, il suo comportamento è assimilabile a un sistema ideale.

(1.10)
$$L[a_1 \cdot x_1(t) + a_2 \cdot x_2(t)] = a_1 \cdot L[x_1(t)] + a_2 \cdot L[x_2(t)]$$

Nel caso di sistemi lineari, le equazioni (1.8) sono lineari.

- **Tempo-invariante**, se il comportamento non varia nel tempo. In tal caso le funzioni F e G sono indipendenti dal tempo.
- **Stabile**, se ad un ingresso di ampiezza limitata, corrisponde un'uscita di ampiezza limitata⁷.

⁷ Nella teoria dei sistemi di controllo vengono definiti vari tipi di stabilità che non interessano la presente trattazione.

2 Argomenti introduttivi

2.1 Sistemi lineari continui

In questo capitolo ci occuperemo dei sistemi continui lineari tempo-invarianti. Inoltre faremo riferimento essenzialmente a sistemi con un solo ingresso e una sola uscita.

Un sistema lineare tempo-invariante è completamente definito una volta nota la risposta impulsiva, cioè la risposta del sistema a un input a delta di Dirac.

Per dimostrarlo, indichiamo questa risposta come

$$(2.1) \quad f(t) = \mathbf{W}[\delta(t)]$$

dove $\mathbf{W}[\]$ indica l'operazione effettuata dal sistema (l'operatore del sistema).

Ricordiamo che una proprietà importante della funzione delta è

$$(2.2) \quad x(t) = \int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau$$

data la linearità del sistema, abbiamo che la risposta a un ingresso $x(t)$ è

$$(2.3) \quad \begin{aligned} y(t) &= \mathbf{W}[x(t)] = \mathbf{W}\left[\int_{-\infty}^{\infty} x(\tau) \delta(t - \tau) d\tau\right] = \\ &= \int_{-\infty}^{\infty} x(\tau) \mathbf{W}[\delta(t - \tau)] d\tau = \int_{-\infty}^{\infty} x(\tau) f(t - \tau) d\tau \end{aligned}$$

Possiamo anche porre

$$(2.4) \quad y(t) = \int_{-\infty}^{\infty} x(t) f(t - \tau) d\tau = \int_{-\infty}^{\infty} f(\tau) x(t - \tau) d\tau$$

Questo integrale è l'**integrale di convoluzione** tra le funzioni x e f . Questo risultato può essere visto in questo modo:

- L'uscita di un sistema lineare è una media pesata del suo ingresso, il peso essendo dato dalla risposta impulsiva del sistema.

Notiamo che se il sistema è causale, $f(\tau)$ è nulla per $\tau < 0$.

L'operazione di convoluzione (2.4) è spesso indicata nella forma abbreviata

$$(2.5) \quad y = f \otimes x$$

Esso gode delle proprietà

- **commutativa** $f \otimes x = x \otimes f$
- **associativa** $a \otimes b \otimes c = (a \otimes b) \otimes c = a \otimes (b \otimes c)$
- **distributiva** $a \otimes (b + c) = (a \otimes b) + (a \otimes c)$

Inoltre si dimostra che

$$(2.6) \quad \frac{d(a \otimes b)}{dx} = \frac{da}{dx} \otimes b = a \otimes \frac{db}{dx}$$

Quindi se facciamo la convoluzione della funzione a per la derivata della b, abbiamo come risultato la derivata della convoluzione della a per la b (e l'analogo per la commutatività). Infine

$$(2.7) \quad \int_{-\infty}^{\infty} (a \otimes b) \cdot dx = \left[\int_{-\infty}^{\infty} a(x) \cdot dx \right] \cdot \left[\int_{-\infty}^{\infty} b(x) \cdot dx \right]$$

Il che significa, tra l'altro, che se facciamo la convoluzione di una qualsiasi funzione per un'altra il cui integrale da $-\infty$ e ∞ è nullo, l'integrale del risultato, da $-\infty$ e ∞ , è nullo.

Un sistema continuo lineare, con un ingresso e un'uscita, può essere definito talora da un'equazione differenziale ordinaria lineare (caso dei sistemi lineari a parametri concentrati; il caso alternativo è detto "a parametri distribuiti", il cui caso più semplice è una linea di ritardo). Se il sistema è tempo-invariante i coefficienti dell'equazione sono indipendenti dal tempo:

$$(2.8) \quad \sum_{k=0}^n A_k \frac{d^k y(t)}{dt^k} = \sum_{l=0}^m B_l \frac{d^l x(t)}{dt^l}$$

In esso x è la variabile indipendente (ovvero l'ingresso, per esempio la forza) e y la variabile dipendente (ovvero l'uscita, per esempio lo spostamento).

Un esempio di un tale sistema può essere un circuito elettrico comprendente solo elementi lineari, cioè resistenze ohmiche, induttanze e condensatori.

Un sistema continuo lineare a parametri concentrati con più ingressi e più uscite è descritto da un sistema di equazioni differenziali lineari.

Come è noto, un sistema descritto da un'equazione differenziale lineare ha la soluzione composta da due parti additive: la **risposta libera** e la **risposta forzata**, la prima dipendente da un numero di parametri pari all'ordine dell'equazione, è la risposta che si ha nel caso sia nulla l'eccitazione (l'ingresso), la seconda nel caso in cui sia nullo lo stato.

Nella teoria dei sistemi, spesso si trascura la presenza della risposta libera.

2.1.1 Trasformata di Laplace

La trasformata di Laplace di una funzione è definita da (trasformata bilaterale)

$$(2.9) \quad F(s) = \int_{-\infty}^{\infty} f(t) \cdot e^{-st} dt$$

o (trasformata unilaterale, più frequentemente usata)

$$(2.10) \quad F(s) = \int_0^{\infty} f(t) \cdot e^{-st} dt$$

dove $s = \sigma + j\omega$ (con j indichiamo l'unità immaginaria).

Essa viene anche indicata con $F(s) = \mathcal{L}\{f(t)\}$ o, se è chiaro dal contesto che stiamo trattando con trasformate di Laplace, con $f(t) \leftrightarrow F(s)$.

In genere l'integrale converge solo per un sotto-insieme del piano s , detto regione di convergenza, ovvero per un intervallo della variabile σ , parte reale di s .

La trasformata inversa viene ricavata con i metodi della teoria delle funzioni di variabile complessa, e si ha

$$(2.11) \quad f(t) = \frac{1}{2\pi j} \cdot \int_{c-j\infty}^{c+j\infty} e^{st} \cdot F(s) \cdot ds$$

dove c è scelto in modo tale che tutti i punti singolari della $F(s)$ giacciono a sinistra della retta $\text{Re}(s) = c$ del piano complesso rappresentante s .

Le proprietà più importanti della trasformata di Laplace sono appresso elencate. Sia $x(t) \leftrightarrow X(s)$, $x_1(t) \leftrightarrow X_1(s)$, $x_2(t) \leftrightarrow X_2(s)$. Si ha:

- Linearità: $a \cdot x_1(t) + b \cdot x_2(t) \leftrightarrow a \cdot X_1(s) + b \cdot X_2(s)$
- Spostamento nel tempo (ritardo t_0): $x(t - t_0) \leftrightarrow e^{-st_0} X(s)$
- Spostamento nel dominio s : $e^{s_0 t} x(t) \leftrightarrow X(s - s_0)$
- Cambiamento di scala temporale: $x(at) \leftrightarrow \frac{1}{|a|} X\left(\frac{s}{a}\right)$
- Inversione temporale: $x(-t) \leftrightarrow X(-s)$
- Differenziazione nel dominio del tempo: $\frac{d x(t)}{d t} \leftrightarrow s \cdot X(s)$

- Integrazione nel dominio del tempo: $\int_{-\infty}^t x(\tau) d\tau \leftrightarrow \frac{1}{s} X(s)$
- Differenziazione nel dominio s: $-t \cdot x(t) \leftrightarrow \frac{dX(s)}{ds}$
- Convoluzione: $\int_{-\infty}^{\infty} x_1(\tau) x_2(t - \tau) d\tau \leftrightarrow X_1(s) \cdot X_2(s)$

Come abbiamo visto, la risposta di un sistema lineare tempo-invariante può essere calcolato con l'integrale di convoluzione dell'ingresso con la risposta impulsiva del sistema. **L'uso della trasformata di Laplace riduce formalmente questa operazione a un prodotto tra la trasformata di Laplace del segnale d'ingresso e la trasformata di Laplace della risposta impulsiva del sistema, che viene detta *funzione di trasferimento*.**

2.1.2 Trasformata di Fourier

Se una funzione è periodica, cioè se $x(t) = x(t + T_0)$ per tutti i t (T_0 è il periodo, $\nu_0 = \frac{1}{T_0}$ è la frequenza e $\omega_0 = 2\pi\nu_0$ la pulsazione), come è noto possiamo svilupparla in serie di Fourier, cioè

$$(2.12) \quad x(t) = \sum_{k=-\infty}^{\infty} X_k \cdot e^{jk\omega_0 t}$$

dove i coefficienti X_k sono ricavati da

$$(2.13) \quad X_k = \frac{1}{T_0} \int_{-\frac{T_0}{2}}^{\frac{T_0}{2}} x(t) \cdot e^{-j \cdot k \cdot \frac{2\pi}{T_0} \cdot t} \cdot dt$$

La trasformata di Fourier è una generalizzazione dello sviluppo in serie di Fourier al caso in cui la funzione x(t) non sia periodica (ovvero sia di periodo infinito).

La trasformata di Fourier è simile alla trasformata di Laplace; in essa però la variabile "trasformata", coniugata del tempo, è reale e indica la pulsazione (in Inglese *angular frequency*; spesso viene indicata con ω). Ricordiamo che $\omega = 2\pi \cdot \nu$, dove ν è la frequenza. La trasformata di Fourier di una funzione x(t) è data da⁸

$$(2.14) \quad X(\omega) = \int_{-\infty}^{\infty} x(t) \cdot e^{-j\omega t} dt$$

⁸ Esistono altre definizioni, quasi equivalenti; per esempio con l'esponenziale positivo, oppure con costanti a moltiplicare differenti.

Come si vede è un caso particolare della trasformata (bilatera) di Laplace, dove $s = j\omega$. Le funzioni della base trasformata sono, come si vede, esponenziali complesse (ricordiamo che $e^{j\omega t} = \cos \omega t + j \cdot \sin \omega t$).

Nella trasformata di Fourier c'è una simmetria perfetta tra la variabile t e la variabile ω . In pratica però, molto spesso nel dominio t le funzioni sono reali mentre sono complesse (ed hermitiane) nel dominio ω .

Notiamo che

$$(2.15) \quad X(0) = \int_{-\infty}^{\infty} x(t) dt$$

La trasformata inversa si ottiene da

$$(2.16) \quad x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega) \cdot e^{j\omega t} d\omega$$

Essa viene anche indicata con $X(s) = \mathcal{F}\{x(t)\}$ o, quando è chiaro che parliamo di trasformate di Fourier, con $x(t) \leftrightarrow X(\omega)$. Se $X_L(s)$ è la trasformata di Laplace di $x(t)$ e $X_F(\omega)$ la trasformata di Fourier, allora⁹ $X_F(\omega) = X_L(j\omega)$.

Data la parentela con la trasformata di Laplace, la trasformata di Fourier gode di proprietà simili, più altre dovute alla maggiore "simmetria" tra le variabili coniugate.

Guardiamo con attenzione queste proprietà, poiché nell'analisi dei segnali si fa largo uso di questa trasformata.

Sia $x(t) \leftrightarrow X(\omega)$, $x_1(t) \leftrightarrow X_1(\omega)$, $x_2(t) \leftrightarrow X_2(\omega)$. Si ha:

- **Linearità:** $a \cdot x_1(t) + b \cdot x_2(t) \leftrightarrow a \cdot X_1(\omega) + b \cdot X_2(\omega)$

Se conosciamo la trasformata di due segnali, possiamo calcolare immediatamente quella di una combinazione di questi due segnali. Possiamo, quando è il caso, scomporre opportunamente un segnale per calcolare, o intuire, la sua trasformata.

- **Spostamento nel tempo:** $x(t - t_0) \leftrightarrow e^{-j\omega t_0} X(\omega)$

Se trasliamo un segnale nel tempo, per la trasformata abbiamo lo stesso valore assoluto e uno sfasamento che varia linearmente con ω .

Si noti che $|X(\omega)|$ è indipendente da spostamenti nel tempo.

- **Spostamento nel dominio ω :** $e^{j\omega_0 t} x(t) \leftrightarrow X(\omega - \omega_0)$

⁹ Nel caso in cui esistono entrambe.

Questa è la duale della precedente. $e^{j\omega_0 t}$ rappresenta una oscillazione complessa: moltiplicare $x(t)$ per questa oscillazione complessa provoca una traslazione della trasformata di Fourier.

Se $x(t)=1$, troviamo che

$$(2.17) \quad e^{j\omega_0 t} \leftrightarrow \delta(\omega - \omega_0)$$

Ora notiamo che $e^{j\omega_0 t} + e^{-j\omega_0 t} = 2 \cdot \cos(\omega_0 t)$ e $e^{j\omega_0 t} - e^{-j\omega_0 t} = 2j \cdot \sin(\omega_0 t)$ allora

$$(2.18) \quad x(t) \cdot \cos(\omega_0 t) \leftrightarrow \frac{X(\omega - \omega_0) + X(\omega + \omega_0)}{2}$$

$$(2.19) \quad x(t) \cdot \sin(\omega_0 t) \leftrightarrow \frac{X(\omega - \omega_0) - X(\omega + \omega_0)}{2j}$$

- **Cambiamento di scala temporale:** $x(at) \leftrightarrow \frac{1}{|a|} X\left(\frac{\omega}{a}\right)$

Anche questa è una proprietà molto importante. la trasformazione $x'(t) = x(a \cdot t)$ produce una contrazione o una dilatazione a seconda che $|a|$ sia maggiore o minore di 1. Una dilatazione nel dominio t corrisponde a una contrazione nel dominio ω e viceversa.

Questa proprietà è la ragione del cosiddetto “principio di indeterminazione di Fourier”, simile a quello di Heisenberg, che vedremo in seguito (par. 4.3).

- **Inversione temporale:** $x(-t) \leftrightarrow X(-\omega)$

Se la $x(t)$ è reale, allora $X(\omega) = X^*(-\omega)$ e quindi anche $x(-t) \leftrightarrow X^*(\omega)$

- **Differenziazione nel dominio del tempo:** $\frac{d x(t)}{d t} \leftrightarrow j\omega \cdot X(\omega)$

Questa proprietà può essere generalizzata con

$$(2.20) \quad \frac{d^n x(t)}{d t^n} \leftrightarrow (j\omega)^n \cdot X(\omega)$$

- **Integrazione nel dominio del tempo:** $\int_{-\infty}^t x(\tau) d\tau \leftrightarrow \pi X(0)\delta(\omega) + \frac{1}{j\omega} X(\omega)$

- **Differenziazione nel dominio ω :** $-j \cdot t \cdot x(t) \leftrightarrow \frac{dX(\omega)}{d\omega}$

- **Convoluzione:** $\int_{-\infty}^{\infty} x_1(\tau) x_2(t-\tau) d\tau \leftrightarrow X_1(\omega) \cdot X_2(\omega)$

Sostituire una operazione come la convoluzione, spesso complicata, se eseguita analiticamente, o computazionalmente costosa, se eseguita numericamente, con una moltiplicazione rende spesso molto utile lavorare nel dominio trasformato di Fourier.

e inoltre:

- **Moltiplicazione:** $x_1(t) \cdot x_2(t) \leftrightarrow \int_{-\infty}^{\infty} X_1(\theta) X_2(\omega-\theta) d\theta$

È la duale della proprietà della convoluzione.

- **Identità di Parseval:**

$$\int_{-\infty}^{\infty} x_1(t) \cdot x_2^*(t) \cdot dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} X_1(\omega) \cdot X_2^*(\omega) \cdot d\omega$$

$$\int_{-\infty}^{\infty} |x(t)|^2 \cdot dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)|^2 \cdot d\omega$$

Questa proprietà è un'utilissima relazione sull'“energia” totale del segnale nel dominio del tempo e di ω . Si noti che se l'integrale si esegue nella variabile ν non compare il “fastidioso” coefficiente $\frac{1}{2\pi}$.

Inoltre, se il segnale $x(t)$ è reale, la sua trasformata di Fourier è $X(\omega) = X^*(-\omega)$, cioè la sua parte reale è pari e la parte immaginaria dispari (funzione hermitiana); se $x(t)$, reale o complessa, è una funzione hermitiana, la trasformata è reale, se è anti-hermitiana ($x(t) = -x^*(-t)$), è immaginaria.

Ecco alcuni casi:

Dominio del tempo	Dominio della frequenza
Reale	Hermitiana
Reale positiva	Massimo reale in 0, hermitiana
Reale pari o hermitiana	Reale
Reale dispari o anti-hermitiana	Immaginaria

Ricordiamo qui che, se le dimensioni di $x(t)$ sono $[x]$, quelle della trasformata X sono $[xt]$.

Un risultato importante per i sistemi che si deduce dalla proprietà della trasformata di Fourier per la convoluzione, è che se l'ingresso di un sistema lineare è una sinusoide, l'uscita è anch'essa una sinusoide, con ampiezza e fase che dipendono dalla frequenza. Quindi per ogni

frequenza possiamo definire un numero complesso che abbia come modulo il “guadagno” del sistema, cioè il rapporto tra l’ampiezza della sinusoide in uscita e quella in ingresso, e la fase la differenza di fase tra di esse. La funzione di trasferimento (definita come la trasformata di Laplace o di Fourier della risposta impulsiva; la vedremo meglio in seguito) dà, per ogni frequenza, questo numero complesso.

Per quanto riguarda i segnali, la trasformata di Fourier ci dà il contenuto in energia (o, come vedremo per i segnali di durata infinita, in potenza) alle varie frequenze. Definiamo **spettro di energia** di un dato segnale, il modulo quadro della sua trasformata di Fourier.

Infine riportiamo la trasformata di Fourier di alcune significative funzioni:

- **Gaussiana:**

$$(2.21) \quad \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{t^2}{2\sigma^2}} \leftrightarrow e^{-\frac{\omega^2 \cdot \sigma^2}{2}}$$

Notiamo che la trasformata ha la stessa forma funzionale della gaussiana (ma mentre nel tempo è normalizzata a 1, in ω no). Se prendiamo come “larghezza” nel tempo la deviazione standard, in ω la larghezza è $\frac{1}{\sigma}$, quindi il prodotto delle due larghezze è sempre 1.

- **Pacchetto gaussiano:**

Un “pacchetto” è un segnale sinusoidale moltiplicato per una “finestra”. In questo caso la finestra è gaussiana. Il calcolo della trasformata è facile, ricordando le proprietà su riportate. Sono qui riportati i tre casi del pacchetto esponenziale complesso (dati complessi, un solo picco nella trasformata), del pacchetto cosinusoidale (funzione reale pari, trasformata reale) e del pacchetto sinusoidale (funzione dispari, trasformata immaginaria).

Notiamo poi che più è ”largo” il pacchetto nel tempo, più sono stretti i picchi (o il picco, nel caso complesso) nella trasformata.

$$(2.22) \quad \frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{t^2}{2\sigma^2}} \cdot e^{j\omega_0 t} \leftrightarrow e^{-\frac{(\omega-\omega_0)^2 \cdot \sigma^2}{2}}$$

$$\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{t^2}{2\sigma^2}} \cdot \cos(\omega_0 t) \leftrightarrow \frac{e^{-\frac{(\omega-\omega_0)^2 \cdot \sigma^2}{2}} + e^{-\frac{(\omega+\omega_0)^2 \cdot \sigma^2}{2}}}{2}$$

$$\frac{1}{\sqrt{2\pi} \cdot \sigma} \cdot e^{-\frac{t^2}{2\sigma^2}} \cdot \sin(\omega_0 t) \leftrightarrow j \cdot \frac{e^{-\frac{(\omega+\omega_0)^2 \cdot \sigma^2}{2}} - e^{-\frac{(\omega-\omega_0)^2 \cdot \sigma^2}{2}}}{2}$$

In figura c'è un pacchetto gaussiano cosinusoidale

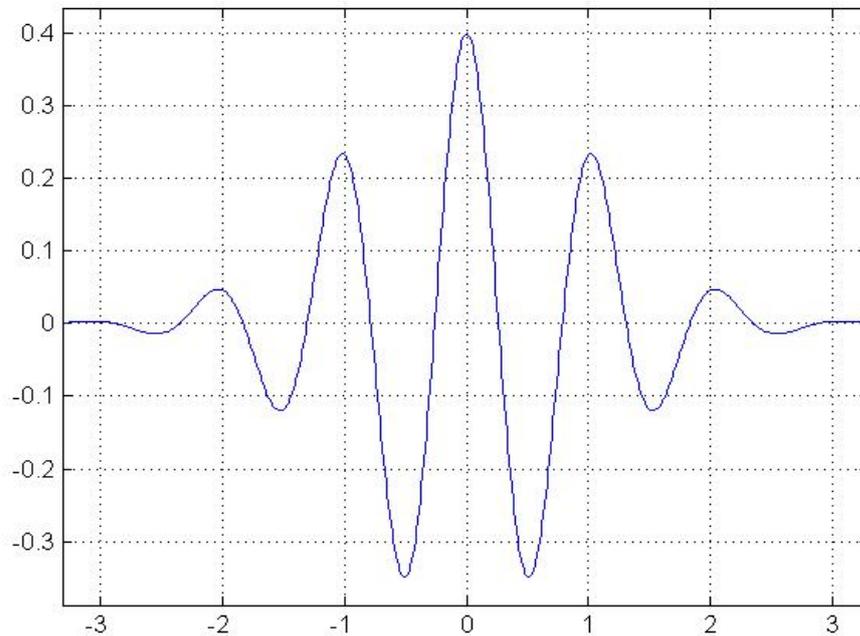


Figura 2-1

- **Rettangolo:**

$$(2.23) \quad r(t; a) = \begin{cases} 1 & \text{per } |t| < a \\ 0 & \text{per } |t| > a \end{cases} \leftrightarrow 2a \cdot \frac{\sin(a \cdot \omega)}{a \cdot \omega}$$

Se prendiamo come misura della larghezza delle due funzioni il valore più basso in cui si azzerano, abbiamo che nel dominio t è a , nel dominio ω è $\frac{\pi}{a}$.

È stata introdotta la funzione sinc(x) come

$$(2.24) \quad \text{sinc}(x) = \frac{\sin(x)}{x}$$

e la trasformata dell'impulso rettangolare può scriversi in termini di questa funzione. La sinc ha una notevole importanza nello studio del campionamento di segnali continui.

- **Pacchetto rettangolare:**

La situazione è analoga al caso del pacchetto gaussiano.

...

- **Esponenziale simmetrico \leftrightarrow Lorentziana (distribuzione di Cauchy):**

$$(2.25) \quad e^{-\frac{|t|}{\tau}} \leftrightarrow \frac{\frac{2}{\tau}}{x^2 + \left(\frac{1}{\tau}\right)^2}$$

- **Pacchetto esponenziale simmetrico:**

La situazione è analoga al caso del pacchetto gaussiano.

...

- **delta:**

$$(2.26) \quad \delta(t) \leftrightarrow 1$$

- **costante:**

$$(2.27) \quad x(t) = c \leftrightarrow 2\pi c \cdot \delta(\omega)$$

- **gradino:**

$$(2.28) \quad u(t) \leftrightarrow \frac{1}{j\omega} + \pi \cdot \delta(\omega)$$

- **esponenziale complesso:**

$$(2.29) \quad e^{j\omega_0 t} \leftrightarrow 2\pi \cdot \delta(\omega - \omega_0)$$

- **esponenziale decrescente:**

$$(2.30) \quad u(t) \cdot e^{-\frac{t}{\tau}} \leftrightarrow \frac{1}{\frac{1}{\tau} + j\omega}$$

Si noti che la stessa funzione invertita nel tempo ha trasformata $\frac{1}{\frac{1}{\tau} - j\omega}$. Sommando la

$u(t) \cdot e^{-\frac{t}{\tau}}$ e la invertita nel tempo, si ha una funzione che ha come trasformata la somma delle trasformate

$$(2.31) \quad u(t) \cdot e^{-\frac{t}{\tau}} + u(-t) \cdot e^{\frac{t}{\tau}} = e^{-\frac{|t|}{\tau}} \leftrightarrow \frac{1}{\frac{1}{\tau} + j\omega} + \frac{1}{\frac{1}{\tau} - j\omega} = \frac{2}{\frac{1}{\tau^2} + \omega^2}$$

che non è altro che la (2.25).

- “doppio polo”:

$$(2.32) \quad u(t) \cdot t \cdot e^{-\frac{t}{\tau}} \leftrightarrow \frac{1}{\left(\frac{1}{\tau} + j\omega\right)^2}$$

2.1.3 La funzione di trasferimento

Abbiamo visto che un sistema lineare tempo invariante può essere definito dalla risposta impulsiva. Infatti (equazione (2.4)) la risposta si calcola facendo la convoluzione dell'ingresso con la funzione risposta impulsiva.

La **funzione di trasferimento** $F(s)$ è la trasformata di Laplace della risposta impulsiva $f(t)$. Per la proprietà della trasformata di Laplace sulla convoluzione, essendo $X(s)$ e $Y(s)$ le trasformate di Laplace dell'ingresso e dell'uscita del sistema, abbiamo

$$(2.33) \quad Y(s) = F(s) \cdot X(s)$$

Inoltre, se abbiamo n sistemi in cascata (in serie), con funzioni di trasferimento

$F_1(s), F_2(s), \dots, F_n(s)$, la funzione di trasferimento complessiva $F(s)$ è data dal prodotto delle singole funzioni

$$(2.34) \quad F(s) = F_1(s) \cdot F_2(s) \cdot \dots \cdot F_n(s)$$

e se abbiamo n sistemi in parallelo (cioè tutti con lo stesso ingresso e con tutte le uscite che vanno a un nodo somma), con funzioni di trasferimento $F_1(s), F_2(s), \dots, F_n(s)$, la funzione di trasferimento complessiva $F(s)$ è data dalla somma delle singole funzioni

$$(2.35) \quad F(s) = F_1(s) + F_2(s) + \dots + F_n(s)$$

Se il sistema è a parametri concentrati, e quindi è descritto da un'equazione del tipo (2.8), allora la funzione di trasferimento è una funzione razionale di s , e cioè

$$(2.36) \quad F(s) = \frac{\sum_{i=0}^m B_i \cdot s^i}{\sum_{i=0}^n A_i \cdot s^i}$$

La funzione di trasferimento è completamente descritta avendo dato

- gli zeri del polinomio al numeratore (chiamati semplicemente “zeri” z_k)
- gli zeri del polinomio al denominatore (chiamati “poli” p_k)
- un fattore di guadagno che, nel caso di A_0 e B_0 diversi da 0, può essere il “guadagno in continua” $F_0 = \frac{B_0}{A_0}$

Se tutti i poli sono differenti, e se $n > m$, dalla (2.36) si può ricavare la risposta impulsiva (calcolata come trasformata inversa di $F(s)$), come un'opportuna combinazione lineare degli n termini $e^{p_i t}$

$$(2.37) \quad f(t) = \sum_{i=1}^n K_i \cdot e^{p_i t}$$

dove

$$(2.38) \quad K_i = \left[(s - p_i) \cdot F(s) \right]_{s=p_i}$$

Nel caso generale, in cui si hanno n_1 volte il polo p_1 , n_2 volte il polo p_2 , e così via, essendo N i poli distinti ed essendo

$$(2.39) \quad \sum_{i=1}^N n_i = n$$

abbiamo la soluzione generale, sempre nell'ipotesi di $n > m$ (altrimenti si avrebbe una “catastrofe ultravioletta”),

$$(2.40) \quad f(t) = \sum_{i=1}^N g_i(t) \cdot e^{p_i t}$$

dove

$$(2.41) \quad g_i(t) = \sum_{k=1}^{n_i} \frac{1}{(n_i - k)!} \frac{d^{n_i - k}}{ds^{n_i - k}} \left[(s - p_i)^{n_i} F(s) \right]_{s=p_i} \cdot \frac{t^{k-1}}{(k-1)!}$$

Perché un sistema lineare tempo invariante sia stabile, condizione necessaria e sufficiente è che la risposta impulsiva tenda a 0 per t tendente all'infinito.

Quindi dalla (2.40) si deduce immediatamente che condizione necessaria e sufficiente per avere la stabilità è che tutti i poli abbiano parte reale negativa.

Possiamo facilmente calcolare la risposta in frequenza di un sistema di funzione di trasferimento $F(s)$, come

$$(2.42) \quad F(j\omega)$$

Spesso la risposta in frequenza viene rappresentata con i diagrammi di Bode o col diagramma di Nyquist. Nel primo caso si producono due grafici, uno in cui si rappresenta il guadagno (il valore assoluto di $F(j\omega)$) in funzione della frequenza (o della pulsazione) in scala doppio-logaritmica¹⁰ e un altro in cui si rappresenta la fase in funzione della frequenza, quest'ultima in scala logaritmica. Nel secondo (diagramma di Nyquist) si grafica la parte immaginaria di $F(j\omega)$ in funzione della parte reale.

La (2.42) può essere ricavata direttamente dalla (2.4), che qui riportiamo

$$y(t) = \int_{-\infty}^{\infty} f(\tau) x(t - \tau) d\tau$$

e che definisce un generico sistema lineare ($x(t)$ è l'ingresso e $y(t)$ l'uscita). Se prendiamo

$$(2.43) \quad x(t) = e^{j\omega t}$$

abbiamo

$$(2.44) \quad y(t) = \int_{-\infty}^{\infty} f(\tau) e^{j\omega(t-\tau)} d\tau = e^{j\omega t} \cdot \int_{-\infty}^{\infty} f(\tau) e^{-j\omega\tau} d\tau = e^{j\omega t} \cdot F(j\omega)$$

L'integrale è anche la trasformata di Fourier della risposta impulsiva $f(t)$. Ciò ci porta alla seguente definizione di funzione di trasferimento $F(j\omega)$ (ponendo ω come parametro):

la $F(j\omega)$ è l'auto-valore relativo all'auto-funzione $e^{j\omega t}$ del sistema.

¹⁰ Spesso viene rappresentato il guadagno in decibel in funzione del logaritmo della frequenza.

2.1.4 Semplici sistemi lineari continui

1. Sistema del primo ordine **passa-basso**¹¹, di funzione di trasferimento

$$(2.45) \quad \frac{B_0}{s - p_1} = \frac{B_0}{s + \frac{1}{\tau}}$$

dove il polo p_1 è reale negativo. La risposta impulsiva è

$$(2.46) \quad f(t) = B_0 \cdot u(t) \cdot e^{-\frac{t}{\tau}}$$

La risposta in frequenza si calcola come

$$(2.47) \quad \frac{B_0}{j\omega + \frac{1}{\tau}}$$

e il guadagno è

$$(2.48) \quad G = \frac{B_0}{\sqrt{\left(j\omega + \frac{1}{\tau}\right)\left(-j\omega + \frac{1}{\tau}\right)}} = \frac{B_0}{\sqrt{\omega^2 + \frac{1}{\tau^2}}}$$

mentre la fase è

$$(2.49) \quad \varphi = -\arctan(\omega \cdot \tau)$$

Diamo i grafici di guadagno e di fase in funzione della frequenza (non della pulsazione ω) e di Nyquist per il caso $\tau = B_0 = 1$

¹¹ Esso può essere realizzato, per esempio, tramite l'uso di un amplificatore operazionale (e pochi altri componenti). Vedi Appendice.

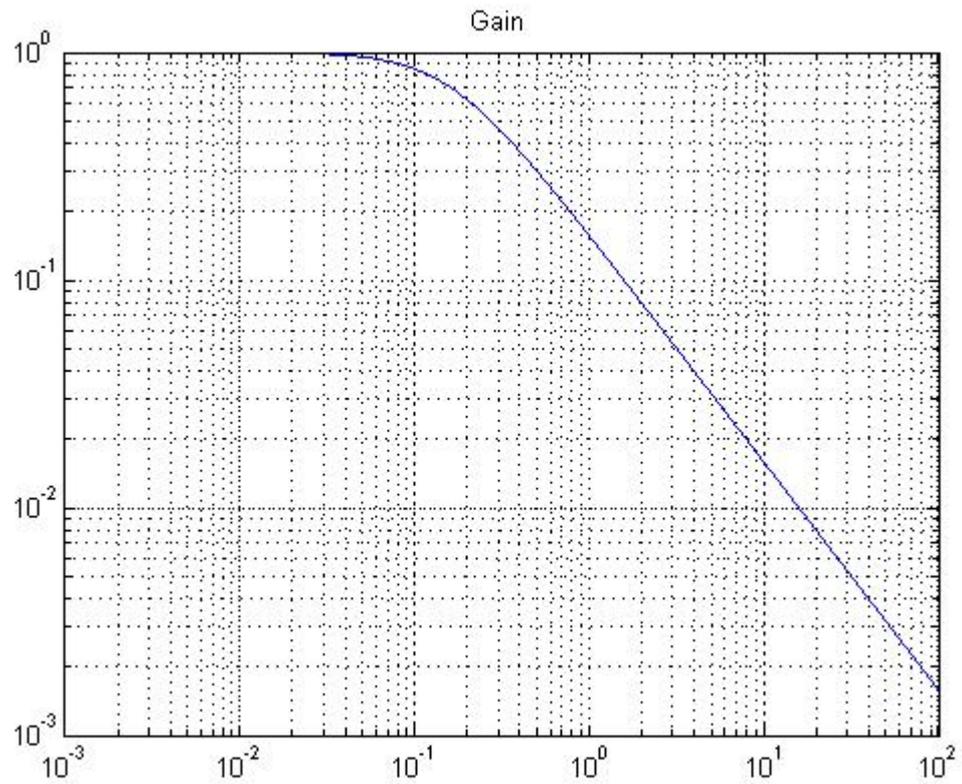


Figura 2-2

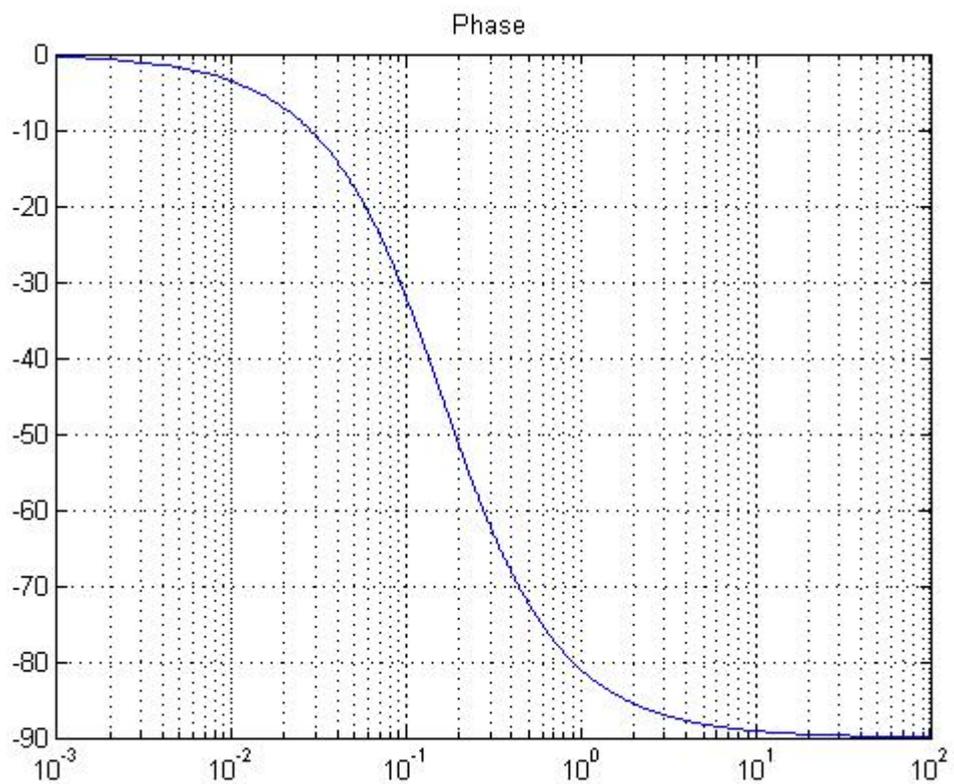


Figura 2-3

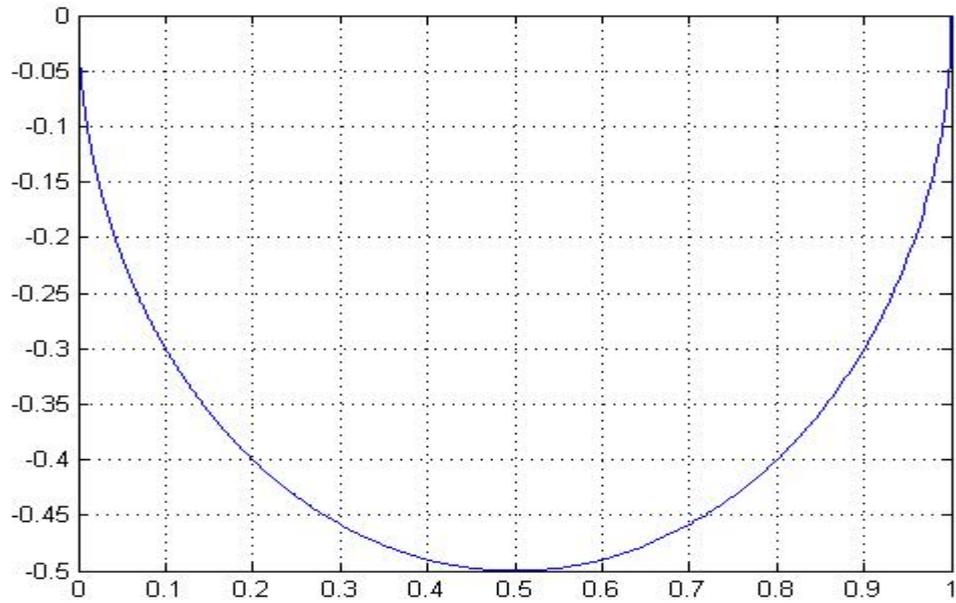


Figura 2-4

2. sistema del primo ordine **passa-alto**, di funzione di trasferimento

$$(2.50) \quad \frac{B_0 \cdot s}{s - p_1} = \frac{B_0 \cdot s}{s + \frac{1}{\tau}}$$

dove il polo p_1 è reale negativo, mentre è presente lo zero nell'origine. La risposta impulsiva è

$$(2.51) \quad f(t) = B_0 \cdot \left[\delta(t) - e^{-\frac{t}{\tau}} \cdot u(t) \right]$$

La risposta in frequenza si calcola come

$$(2.52) \quad \frac{B_0 \cdot j\omega}{j\omega + \frac{1}{\tau}}$$

e il guadagno è

$$(2.53) \quad G = \frac{B_0 \sqrt{-j\omega \cdot j\omega}}{\sqrt{\left(j\omega + \frac{1}{\tau}\right)\left(-j\omega + \frac{1}{\tau}\right)}} = \frac{B_0 \cdot \omega}{\sqrt{\omega^2 + \frac{1}{\tau^2}}}$$

mentre la fase è

$$(2.54) \quad \varphi = \arctan(\omega \cdot \tau) + \frac{\pi}{2}$$

Ecco i grafici di guadagno e di fase in funzione della frequenza (non della pulsazione ω) e di Nyquist per il caso $\tau = B_0 = 1$

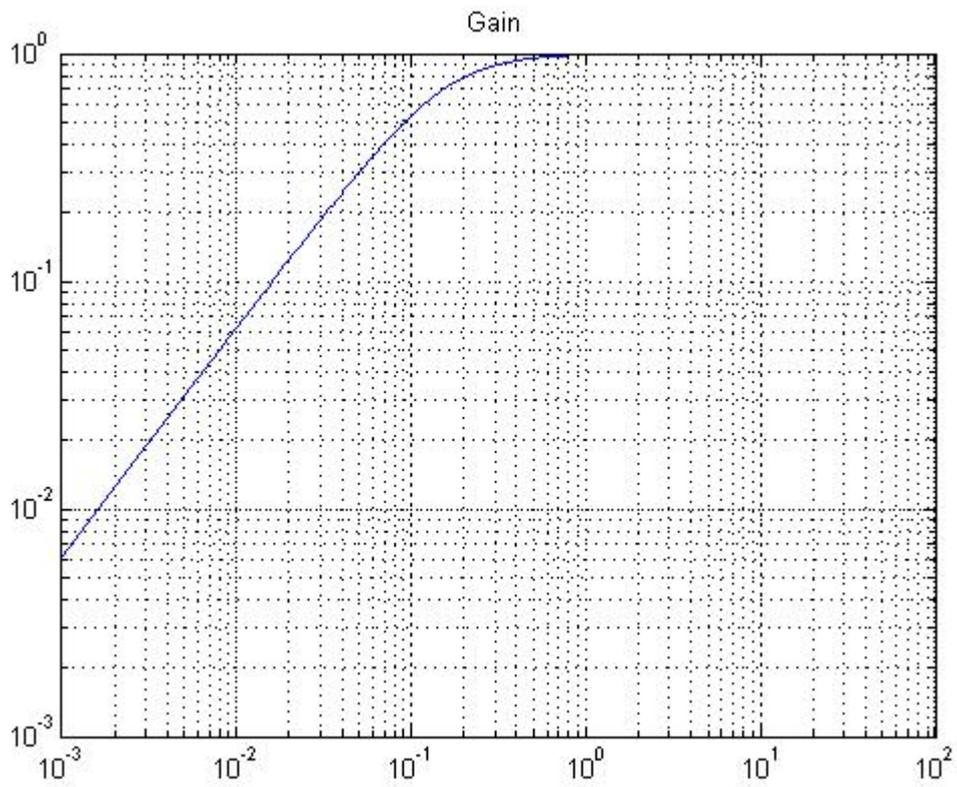


Figura 2-5

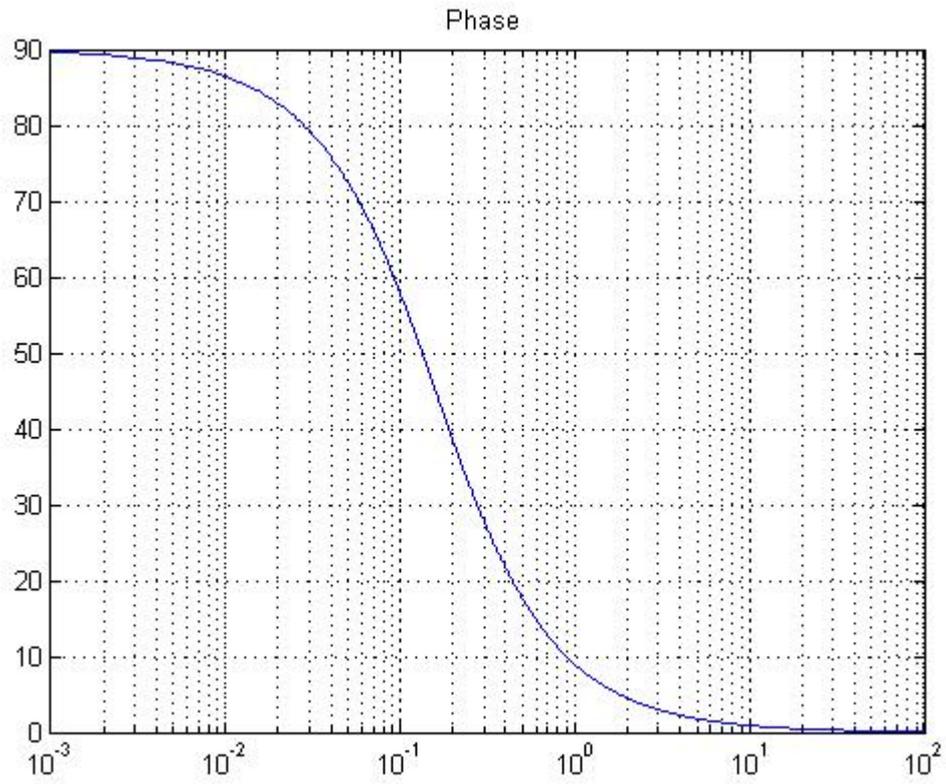


Figura 2-6

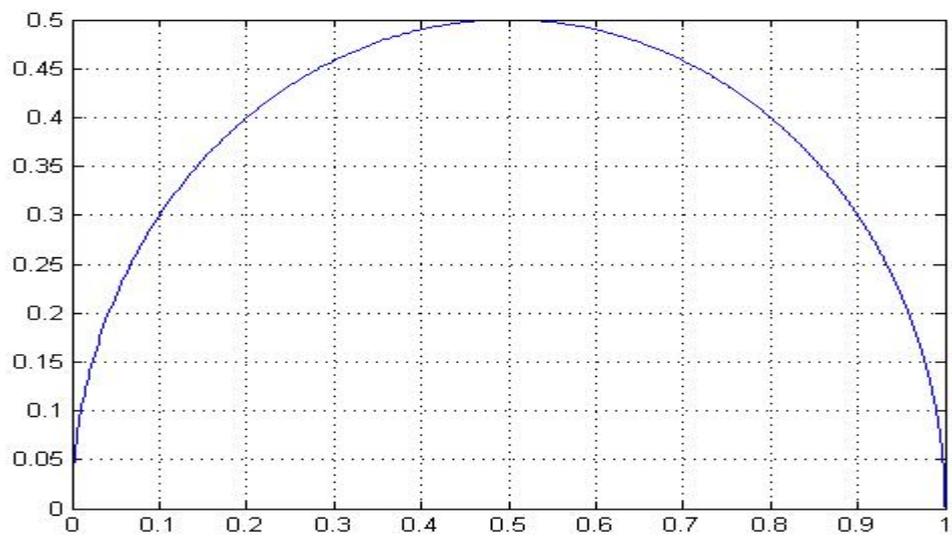


Figura 2-7

3. sistema del secondo ordine semplice (**risonanza**)

$$(2.55) \quad F(s) = \frac{\omega_0}{s^2 + \frac{2}{\tau} \cdot s + \left(\omega_0^2 + \frac{1}{\tau^2}\right)}$$

Si hanno due poli complessi coniugati di valore

$$(2.56) \quad p_{1,2} = \frac{1}{\tau} \pm j \cdot \omega_0$$

La risposta impulsiva è

$$(2.57) \quad f(t) = u(t) \cdot e^{-\frac{t}{\tau}} \cdot \sin(\omega_0 t)$$

La risposta in frequenza è

$$(2.58) \quad F(j\omega) = \frac{\omega_0}{-\omega^2 + \frac{2}{\tau} \cdot j \cdot \omega + \left(\omega_0^2 + \frac{1}{\tau^2}\right)} = \omega_0 \cdot \frac{\left(\omega_0^2 - \omega^2 + \frac{1}{\tau^2}\right) - \frac{2}{\tau} \cdot j \cdot \omega}{\frac{4}{\tau^2} \cdot \omega^2 + \left(\omega_0^2 - \omega^2 + \frac{1}{\tau^2}\right)^2}$$

da cui il guadagno è

$$(2.59) \quad |F(j\omega)| = \omega_0 \cdot \frac{1}{\sqrt{\frac{4}{\tau^2} \cdot \omega^2 + \left(\omega_0^2 - \omega^2 + \frac{1}{\tau^2}\right)^2}}$$

mentre la fase è

$$(2.60) \quad \varphi = \arctan \frac{-\frac{2\omega}{\tau}}{\omega_0^2 - \omega^2 + \frac{1}{\tau^2}}$$

Ecco i grafici di guadagno e di fase in funzione della frequenza (non della pulsazione ω) e di Nyquist per il caso $\omega_0 = 1$, $\tau = 100$

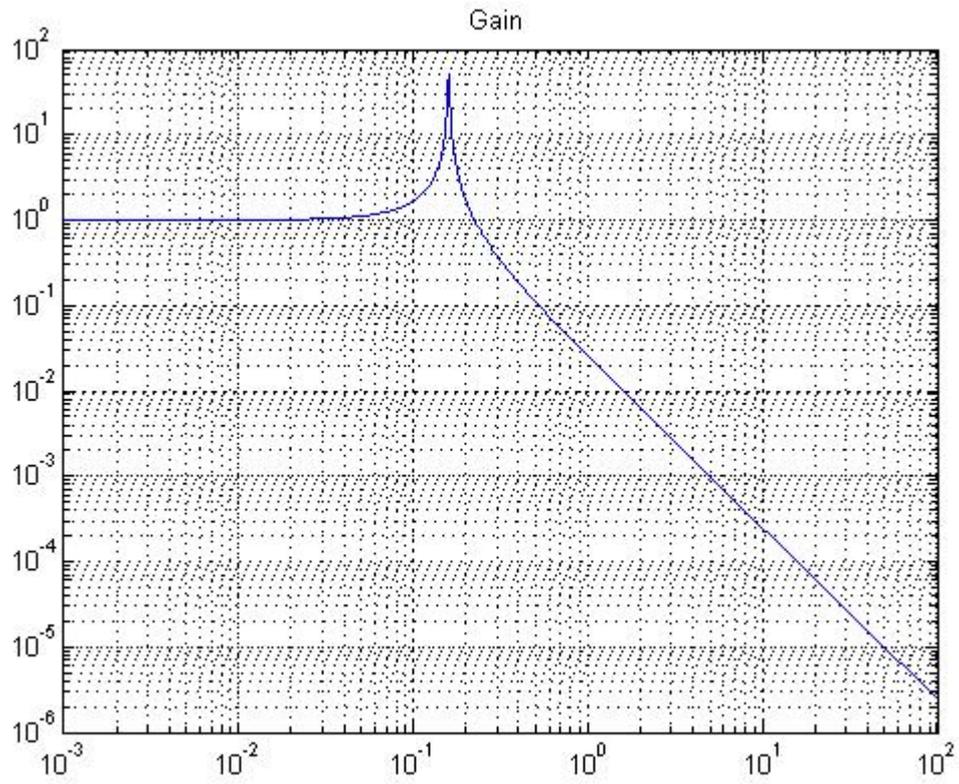


Figura 2-8

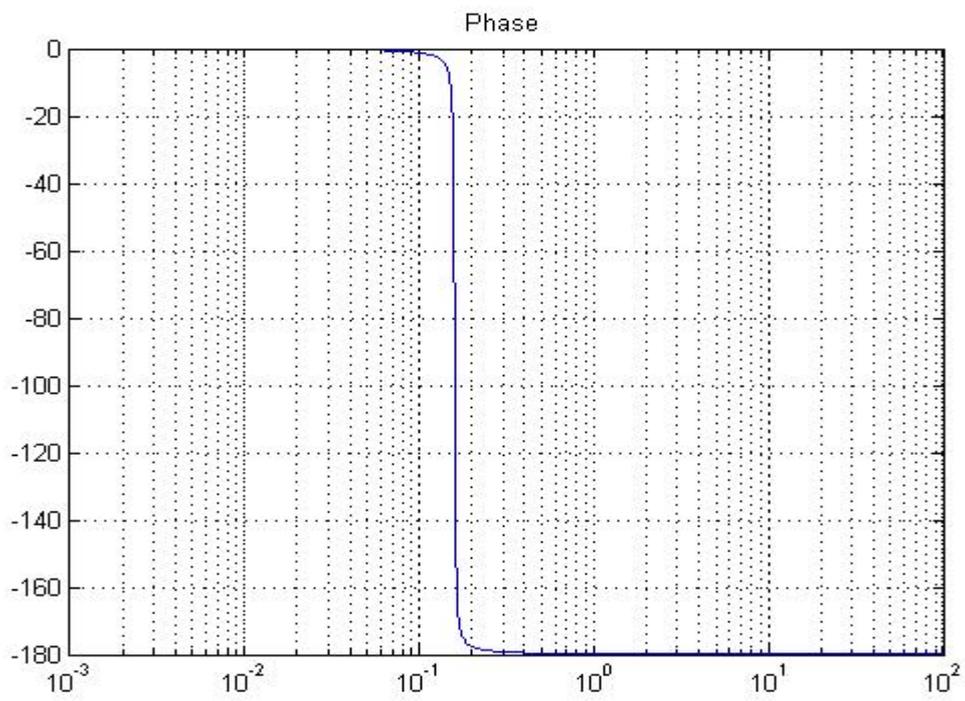


Figura 2-9

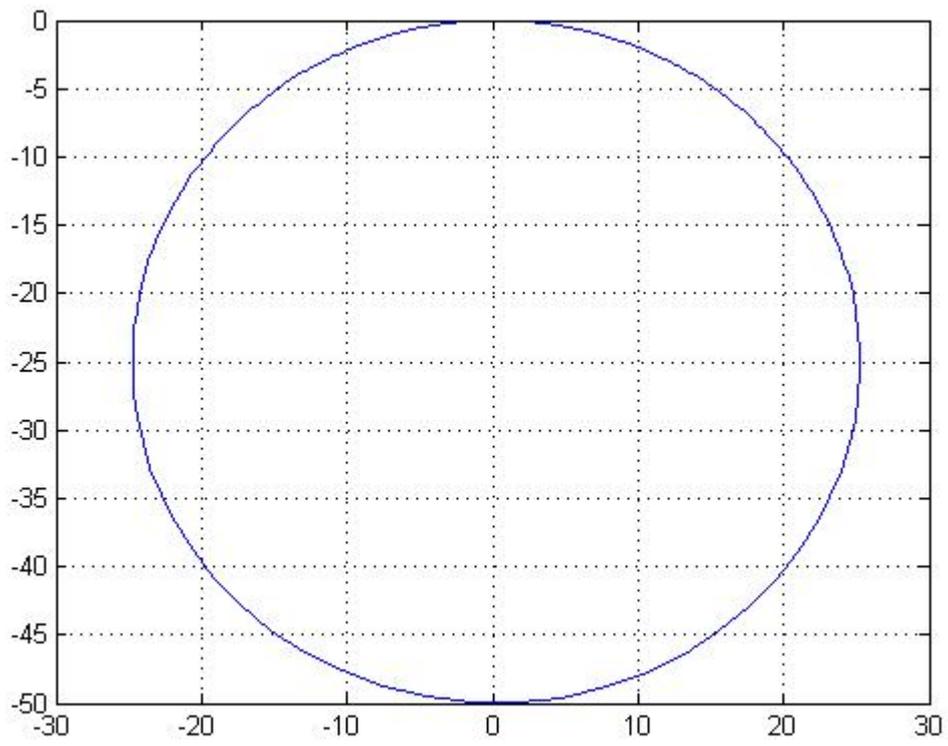


Figura 2-10

2.2 Teoria delle probabilità

Queste note si intendono come un compendio di risultati relativi alla teoria delle probabilità.

2.2.1 Variabili casuali

Una variabile casuale è una variabile (reale o complessa) al cui valore è associata una distribuzione di probabilità.

I parametri fondamentali di una distribuzione di probabilità sono il **valore medio** o **valore atteso**¹², che in genere indicheremo con μ , che è un parametro di localizzazione della distribuzione, e la **deviazione standard**, che in genere indicheremo con σ , e che è un parametro di “sparpagliamento” (o “larghezza”) della distribuzione. Spesso invece della deviazione standard si indica il suo quadrato σ^2 , detto **varianza**.

Distinguiamo tra variabili casuali discrete e continue.

2.2.2 Variabili casuali discrete

Le variabili casuali discrete sono definite da insiemi numerabili (finiti o infiniti) di valori che possono prendere, a cui è associata una distribuzione discreta di probabilità p_k , con k la variabile discreta, tale che

$$(2.61) \quad \sum_k p_k = 1$$

Per una distribuzione discreta si definisce

- il valore atteso

$$(2.62) \quad \mu = E[k] = \sum_k k \cdot p_k$$

- la varianza

$$(2.63) \quad \sigma^2 = E[(k - \mu)^2] = \sum_k (k - \mu)^2 \cdot p_k$$

La deviazione standard σ è definita come la radice quadrata della varianza.

¹² In passato veniva anche indicato come **speranza matematica**.

- i momenti centrali di ordine N

$$(2.64) \quad \mu^{(N)} = E\left[(k - \mu)^N\right] = \sum_k (k - \mu)^N \cdot p_k$$

Esempi di distribuzioni discrete sono:

- la **distribuzione uniforme discreta** (per esempio quella delle facce di un dado "onesto")

$$(2.65) \quad p_k = \frac{1}{N} \quad \text{per } 1 \leq k \leq N$$

con

$$(2.66) \quad \mu = \frac{N+1}{2}$$

e

$$(2.67) \quad \sigma = \sqrt{\frac{N^2 - 1}{12}}$$

- la **distribuzione binomiale** (che è definita per un numero finito di valori)

$$(2.68) \quad p_k = \binom{N}{k} p^k (1-p)^{N-k}$$

definita dai parametri p ed N. Per essa

$$(2.69) \quad \mu = E[k] = N \cdot p$$

$$(2.70) \quad \sigma^2 = N \cdot p \cdot (1-p)$$

- la **distribuzione di Poisson** (che è definita per un numero infinito, ma numerabile di valori)

$$(2.71) \quad P(k; \mu) = \frac{\mu^k}{k!} e^{-\mu}$$

definita dal solo parametro μ . Il valore aspettato è proprio μ e anche $\sigma^2 = \mu$.

2.2.3 Variabili casuali continue

Una variabile casuale continua è una variabile reale o complessa su cui è definita una funzione di densità di probabilità $f(x)$, che ha le seguenti proprietà, per $-\infty < x < \infty$,

- $f(x) \geq 0$
- $\int_{-\infty}^{\infty} f(x) dx = 1$
- $\text{Prob}(a \leq x \leq b) = \int_a^b f(x) dx$

Date queste proprietà, si ha che, se $f(x)$ è una densità di probabilità, allora lo è anche

$$(2.72) \quad f'(x) = A \cdot f(A \cdot x + B)$$

con $A > 0$. Questa trasformazione descrive una dilatazione (o una contrazione se $A > 1$) e una traslazione.

Vogliamo notare che le variabili discrete sono un caso particolare delle variabili continue: per una variabile discreta definita dalle $\{p_k\}$, abbiamo una densità di probabilità

$$(2.73) \quad f(x) = \sum_k p_k \cdot \delta(x - k)$$

Come per una variabile casuale discreta, anche per una variabile casuale continua può definirsi il **valore atteso** o **valor medio** della variabile casuale. In questo caso si pone

$$(2.74) \quad \mu = E[x] = \int_{-\infty}^{\infty} x \cdot f(x) dx$$

Vogliamo ricordare che esistono altri parametri di localizzazione di una variabile casuale, oltre al valor medio, e cioè:

- la **mediana**, il valore m per cui $F(m) = \frac{1}{2}$, cioè il valore della variabile casuale per cui si ha esattamente la stessa probabilità che il risultato sia maggiore o inferiore ad essa. È ovvia l'analogia alla mediana introdotta del capitolo 5 come parametro di posizione per descrivere un insieme di dati.
- la **moda**, il valore per cui la densità di probabilità ha il massimo assoluto, cioè il valore verso cui più si addensano i risultati delle ripetizioni dell'esperimento probabilistico descritto da $f(x)$.

Data una funzione $g(x)$ della variabile casuale x , possiamo calcolare il valore atteso di $g(x)$ (detto **valor medio di $g(x)$**) come

$$(2.75) \quad E[g(x)] = \int_{-\infty}^{\infty} g(x) \cdot f(x) dx$$

Un caso particolare è $g(x) = (x - \mu)^2$, il cui valore atteso definisce la **varianza** della densità $f(x)$, detta anche varianza di insieme della variabile casuale x o varianza della distribuzione¹³

$$(2.76) \quad \sigma^2 \equiv \text{Var}[x] = E[(x - \mu)^2] = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx$$

Se μ e σ sono valor medio e deviazione standard della $f(x)$, allora la $f'(x)$, ottenuta dalla trasformazione (2.72), ha sono valor medio e deviazione standard

$$(2.77) \quad \mu' = \mu - B$$

$$(2.78) \quad \sigma' = \frac{\sigma}{A}$$

La deviazione standard descrive quanto una densità di probabilità è "stretta" intorno alla media. La **disuguaglianza di Chebyshev** dà a questa definizione qualitativa un carattere probabilistico quantitativo.

Per una variabile casuale x con media μ e deviazione standard σ , si ha che

$$(2.79) \quad \text{Probabilità}(|x - \mu| \geq k \cdot \sigma) \leq \frac{1}{k^2}$$

Ovviamente questa disuguaglianza¹⁴ ha senso per $k > 1$. Come vedremo, per le più comuni distribuzioni, $\text{Probabilità}(|x - \mu| \geq k \cdot \sigma)$ è molto minore di $\frac{1}{k^2}$.

Si possono definire i momenti (rispetto all'origine) di ordine k come

$$(2.80) \quad m^{(k)} = \int_{-\infty}^{\infty} x^k f(x) dx$$

e i momenti centrali di ordine k come

$$(2.81) \quad \mu^{(k)} = \int_{-\infty}^{\infty} (x - \mu)^k f(x) dx$$

Tra i momenti rispetto a 0 e quelli centrali ci sono delle precise relazioni. Per esempio

$$(2.82) \quad \begin{aligned} \mu^{(2)} &= m^{(2)} - (m^{(1)})^2 \\ \mu^{(3)} &= m^{(3)} - 3 \cdot m^{(1)} m^{(2)} + 2(m^{(1)})^3 \\ \mu^{(4)} &= m^{(4)} - 4 \cdot m^{(1)} m^{(3)} + 6(m^{(1)})^2 m^{(2)} - 3(m^{(1)})^4 \end{aligned}$$

¹³ In statistica viene anche detta "varianza della popolazione"

¹⁴ Il caso limite è la distribuzione costituita da due delte eguali. In tal caso, per $k > 1$, la probabilità è 0.

A partire dai momenti di ordine 3 e 4 si definiscono i parametri **asimmetria** (**skewness** in inglese) e **curtosi** (**kurtosis** in inglese) come

$$(2.83) \quad \text{asimmetria} = \frac{\mu^{(3)}}{\sigma^3}$$

e

$$(2.84) \quad \text{curtosi} = \frac{\mu^{(4)}}{\sigma^4} - 3$$

Lo “strano” 3 nella definizione della curtosi permette di avere 0 nel caso della gaussiano.

Per una variabile casuale reale si definisce **distribuzione cumulativa** la funzione

$$(2.85) \quad F(x) = \int_{-\infty}^x f(\xi) d\xi = \text{Prob}[\text{variabile casuale} \leq x]$$

Esempi di distribuzioni di variabili casuali continue sono:

a) la **distribuzione uniforme** in cui la densità di probabilità è

$$(2.86) \quad f(x) = \begin{cases} \frac{1}{b-a} & \text{per } a \leq x \leq b \\ 0 & \text{altrove} \end{cases}$$

Il valor medio è

$$(2.87) \quad \mu = \frac{1}{b-a} \int_a^b x dx = \left[\frac{x^2/2}{b-a} \right]_a^b = \frac{b^2 - a^2}{2 \cdot (b-a)} = \frac{b+a}{2}$$

e i momenti centrali sono

$$(2.88) \quad \mu^{(k)} = \frac{1}{b-a} \int_a^b (x-\mu)^k dx = \left[\frac{1}{(k+1)(b-a)} \cdot (x-\mu)^{k+1} \right]_a^b = \frac{(b-a)^{k+1} - (a-b)^{k+1}}{2^{k+1}(k+1) \cdot (b-a)}$$

Si nota che per k dispari $\mu^{(k)} = 0$, mentre per k pari

$$(2.89) \quad \mu^{(k)} = \frac{(b-a)^k}{2^k \cdot (k+1)}$$

La varianza è quindi

$$(2.90) \quad \sigma^2 = \mu^{(2)} = \frac{(b-a)^2}{12}$$

b) la **distribuzione di Gauss** o **distribuzione normale**, con densità

$$(2.91) \quad f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} = N(\mu, \sigma)$$

con valor medio μ e varianza σ^2 (spesso si indica semplicemente con $N(\mu, \sigma)$).

Spesso si usano variabili gaussiane normalizzate o “standard”, in cui $\mu = 0$ e $\sigma = 1$.

La grande importanza della distribuzione normale è dovuta soprattutto al **teorema del limite centrale**. Con esso si dimostra che, **se sommiamo N variabili casuali indipendenti, con distribuzioni anche diverse, ma con varianze dello stesso ordine di grandezza, se N tende all’infinito, la distribuzione della somma tende a una distribuzione gaussiana che ha valor medio la somma dei valor medi e varianza la somma delle varianze.**

c) la distribuzione di Laplace, con densità

$$(2.92) \quad f(x; \mu, b) = \frac{1}{2b} e^{-\frac{|x-\mu|}{b}}$$

con valor medio μ e varianza $2b^2$. La curtosi è 3. È utile per modellizzare diversi processi reali con code pesanti.

d) la **distribuzione del χ^2** , che indica la distribuzione della somma di N variabili gaussiane normalizzate al quadrato. Ha densità

$$(2.93) \quad f(\chi^2; N) = \frac{1}{2^{N/2} \Gamma(N/2)} (\chi^2)^{\frac{N}{2}-1} e^{-\frac{\chi^2}{2}}$$

Il valore aspettato e la varianza sono rispettivamente \mathbf{N} e $\mathbf{2N}$; il terzo e quarto momento centrale sono $\mu_3 = 8N$ e $\mu_4 = 12 \cdot N \cdot (N + 4)$. L’asimmetria è $2\sqrt{\frac{2}{N}}$ e la curtosi è $\frac{12}{N}$.

Nella prima figura ci sono le distribuzioni del χ^2 per $N = 1, 2, 3, 4$. Nella seconda sono confrontate la distribuzione del χ^2 con $N = 100$ e la gaussiana con la stessa media e la stessa varianza (in rosso la gaussiana).

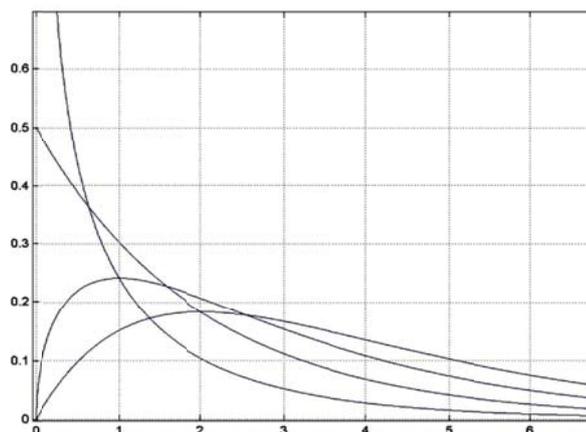


Figura 2-11

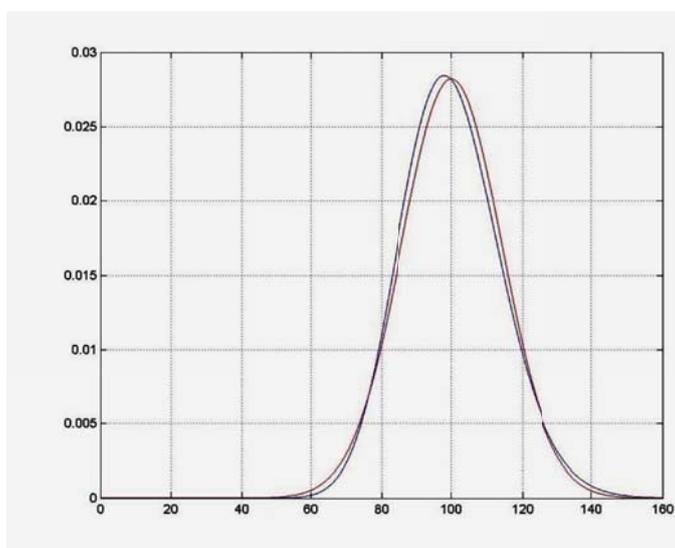


Figura 2-12

Talora si usa la distribuzione del χ^2 , ottenuta da questa dividendo la variabile per N.

e) la **distribuzione esponenziale**. Un caso particolare della distribuzione del χ^2 è la distribuzione esponenziale. Nel caso generale la distribuzione esponenziale ha densità

$$(2.94) \quad f(x) = u(x) \cdot \frac{1}{\mu} \cdot e^{-\frac{x}{\mu}}$$

e la distribuzione integrale è

$$(2.95) \quad F(x) = 1 - u(x) \cdot e^{-\frac{x}{\mu}}$$

il valore atteso è μ ed è uguale alla deviazione standard.

Tale distribuzione si ottiene se sommiamo due variabili gaussiane indipendenti di media nulla, quadrate, con la stessa varianza σ_G^2 . In tal caso si ha $\mu = 2 \cdot \sigma_G^2$.

- f) la **distribuzione di Rayleigh**. Se si fa la radice quadrata della somma dei quadrati di due variabili gaussiane indipendenti di media nulla e uguale varianza σ_G^2 , abbiamo la distribuzione di Raleigh

$$(2.96) \quad f(x) = u(x) \cdot \frac{x}{\sigma_G^2} \cdot e^{-\frac{x^2}{2\sigma_G^2}}$$

che ha valor medio $\sigma_G \cdot \sqrt{\frac{\pi}{2}}$ e varianza $\left(2 - \frac{\pi}{2}\right) \cdot \sigma_G^2$.

- g) la **distribuzione di Cauchy**, nota in Fisica anche col nome di distribuzione di distribuzione di Breit-Wigner o di Lorentz, ha la densità

$$(2.97) \quad f(x; \mu, d) = \frac{1}{\pi d} \cdot \frac{1}{1 + \left(\frac{x - \mu}{d}\right)^2}$$

È caratterizzata da code molto pesanti e dal fatto che non ha varianza (l'integrale è divergente). Il valor medio è μ (calcolata con argomenti di simmetria). Per la simmetria, i momenti centrali di ordine dispari sono nulli. Anche i momenti pari superiori non sono calcolabili.

Un esempio di esperimento i cui risultati sono distribuiti secondo la distribuzione di Cauchy è il seguente. Si faccia ruotare un disco in modo che si fermi a caso con distribuzione dell'angolo uniforme tra 0 e 360 gradi. La tangente di questo angolo è distribuita secondo Cauchy, con media 0 e parametro $d = 1$.

Un altro importante caso in cui compare la distribuzione di Cauchy è quello del rapporto tra due variabili gaussiane a media nulla. Siano x e y due variabili gaussiane con valor medio nullo, varianze σ_x^2 e σ_y^2 e coefficiente di correlazione r (per la definizione vedi (2.138)).

Essendo $z = \frac{x}{y}$, abbiamo

$$(2.98) \quad f(z) = \frac{\sqrt{1-r^2} \cdot \sigma_x \cdot \sigma_y}{\sigma_y^2 \cdot \left(z - r \cdot \frac{\sigma_x}{\sigma_y}\right)^2 + \sigma_x^2 \cdot (1-r^2)}$$

e, se le due variabili sono scorrelate,

$$(2.99) \quad f(z) = \frac{\frac{\sigma_x}{\pi \cdot \sigma_y}}{z^2 + \frac{\sigma_x^2}{\sigma_y^2}}$$

Se si cerca di fare la media di un campione estratto da una distribuzione di Cauchy, si trova che comunque si fa crescere della dimensione del campione, la media non converge al valor medio. Infatti non esistendo la varianza, la disuguaglianza di Chebyshev non funziona.

In figura è riportata la distribuzione di Cauchy (in rosso) con valor medio nullo e $d = 1$, in confronto con la normale standardizzata (in blu). Si noti la differente “pesantezza” delle code.

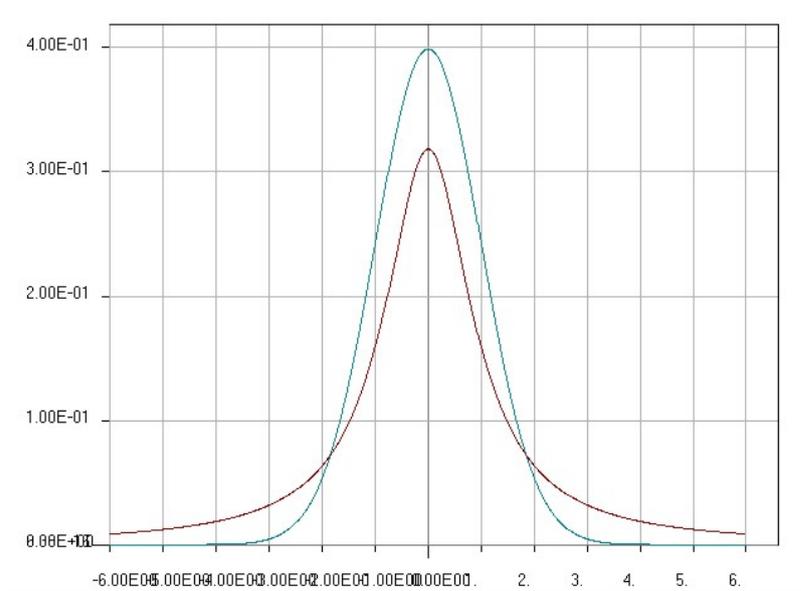


Figura 2-13

h) la **distribuzione t di Student** (statistica per piccoli campioni).

Supponiamo di avere un campione casuale di dimensione N $\{x_1, x_2, \dots, x_N\}$ estratto da una popolazione normale di media μ e deviazione standard σ (per esempio N misure indipendenti di una certa grandezza fisica, con errore casuale gaussiano).

Calcoliamo la media e la varianza campionaria (vedi paragrafo 2.3)

$$\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i \quad \text{e} \quad S^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

La variabile

$$(2.100) \quad t = \frac{\bar{x} - \mu}{S / \sqrt{N}}$$

(analoga alla variabile z) segue la distribuzione t di Student, con N-1 gradi di libertà.

Tale distribuzione è

$$(2.101) \quad f(x) = \frac{\Gamma\left(\frac{M+1}{2}\right)}{\sqrt{M\pi} \cdot \Gamma\left(\frac{M}{2}\right)} \cdot \left(1 + \frac{x^2}{M}\right)^{-\frac{M+1}{2}}$$

dove M è chiamato "numero dei gradi di libertà"); ha media nulla e deviazione standard

$$\sigma = \sqrt{\frac{M}{M-2}}.$$

Per M grande tende a una gaussiana, per M=1 è una distribuzione di Cauchy. È in genere una distribuzione a campana con code più pesanti di una gaussiana.

2.2.4 Somma di variabili casuali indipendenti e funzione caratteristica

Se abbiamo due variabili casuali indipendenti (vedremo in seguito il significato rigoroso di questa parola) \mathbf{x} e \mathbf{y} , definite da due densità di probabilità $\mathbf{f}(\mathbf{x})$ e $\mathbf{g}(\mathbf{y})$, si dimostra che la loro somma $\mathbf{z} = \mathbf{x} + \mathbf{y}$ ha densità di probabilità $\mathbf{h}(\mathbf{z})$ data dalla convoluzione delle due densità

$$(2.102) \quad h(z) = \int_{-\infty}^{\infty} f(\zeta)g(z-\zeta) d\zeta$$

Ce ne possiamo convincere considerando che, per il valore di x compreso nell'intervallo $x' \leq x \leq x'+dx$, che ha probabilità $f(x) \cdot dx$ la distribuzione della somma è $g(y-x')$.

Un risultato importante è che in questo caso si ha

$$(2.103) \quad E[z] = E[x] + E[y]$$

e

$$(2.104) \quad Var[z] = Var[x] + Var[y]$$

(la prima è vera anche se le variabili non sono indipendenti).

Quando si ha a che fare con la somma di variabili casuali indipendenti, è particolarmente comodo introdurre la **funzione caratteristica** di una distribuzione di probabilità, come

$$(2.105) \quad \Phi(\omega) = E[e^{j\omega x}]$$

che, nel caso continuo, è

$$(2.106) \quad \Phi(\omega) = \int_{-\infty}^{\infty} e^{j\omega x} f(x) dx$$

e nel caso discreto

$$(2.107) \quad \Phi(\omega) = \sum_k e^{j\omega k} p_k$$

Si noti che la funzione caratteristica è la trasformata di Fourier della densità di probabilità (con una convenzione diversa da quella che abbiamo usato noi nella (2.14); con la nostra convenzione è la complessa coniugata della trasformata di Fourier). Ricordando le proprietà della trasformata di Fourier, dalla (2.102) ricaviamo che **la funzione caratteristica della somma di due variabili casuali è data dal prodotto delle due funzioni caratteristiche delle due variabili.**

Consideriamo la funzione caratteristica associata a una variabile gaussiana:

$$(2.108) \quad \Phi(\omega) = e^{j\mu\omega - \frac{\sigma^2\omega^2}{2}} = \exp\left(j\mu\omega - \frac{\sigma^2\omega^2}{2}\right)$$

Se abbiamo la somma di n variabili gaussiane, ciascuna definita da $\{\mu_i, \sigma_i\}$, abbiamo per la somma

$$(2.109) \quad \Phi(\omega) = \exp\left(j\omega \cdot \sum_i \mu_i - \frac{\omega^2 \cdot \sum_i \sigma_i^2}{2}\right)$$

da cui deduciamo che la somma ha distribuzione normale con valor medio pari alla somma dei valor medi e varianza pari alla somma delle varianze.

Se facciamo la media di n variabili gaussiane con eguali parametri (μ, σ) , ciò equivale a

sommare n variabili con $\left(\frac{\mu}{n}, \frac{\sigma}{n}\right)$ e quindi si ottiene $\left(\mu_{media} = \mu, \sigma_{media} = \frac{\sigma}{\sqrt{n}}\right)$. Quindi

facendo la media “guadagniamo” una riduzione della deviazione standard di \sqrt{n} .

La funzione caratteristica della distribuzione di Cauchy di equazione (2.97) è

$$(2.110) \quad \Phi(\omega) = e^{j\mu\omega - d|\omega|}$$

Per n variabili casuali di Cauchy, definite ciascuna da (μ_i, d_i) , troviamo

$$(2.111) \quad \Phi(\omega) = \exp\left(j\omega \cdot \sum_{i=1}^n \mu_i - |\omega| \cdot \sum_{i=1}^n d_i\right)$$

Abbiamo cioè una nuova variabile di Cauchy, con parametri $\left(\sum_{i=1}^n \mu_i, \sum_{i=1}^n d_i\right)$.

Se facciamo la media di n variabili di Cauchy con eguali parametri (μ, d) , ciò equivale a sommare n variabili con $\left(\frac{\mu}{n}, \frac{d}{n}\right)$ e quindi si ottiene $(\mu_{media} = \mu, d_{media} = d)$. Quindi facendo la media otteniamo esattamente la stessa distribuzione.

Altra proprietà interessante della funzione caratteristica è che

$$(2.112) \quad \left. \frac{d^k \Phi}{d\omega^k} \right|_{\omega=0} = j^k \cdot m^{(k)}$$

dove $m^{(k)}$ indica il momento non centrale di ordine k (vedi (2.80)).

2.2.5 Distribuzione di una funzione di variabile casuale

Supponiamo di avere una variabile casuale x di densità $f_x(x)$ e costruiamo la nuova variabile

$$(2.113) \quad y = g(x)$$

Ci domandiamo quale sia la densità di probabilità della variabile y.

Nell'ipotesi che la g(x) sia continua e monotona, si ha che, essendo $x=g^{-1}(y)$,

$$(2.114) \quad f_y(y) = f_x(g^{-1}(y)) \cdot \left| \frac{d g^{-1}(y)}{d y} \right|$$

Se la condizione di monotonicità non è verificata, possiamo dividere la funzione in vari pezzi monotoni $y=g_1(x), y=g_2(x), \dots$ ciascuno dei quali ammette una soluzione $x=g_1^{-1}(y), x=g_2^{-1}(y), \dots$ allora

$$(2.115) \quad f_y(y) = \sum_i f_x(g_i^{-1}(y)) \cdot \left| \frac{d g_i^{-1}(y)}{d y} \right|$$

Supponiamo che $\varphi(x)$ sia una densità di probabilità; allora $F(x) = \int_{-\infty}^x \varphi(\xi) \cdot d\xi$ è monotona (non decrescente), senza pezzi costanti. Consideriamo la $g(x)=F^{-1}(x)$, anch'essa monotona crescente e siamo quindi nelle condizioni della (2.114). Inoltre $0 \leq g^{-1}(x) \leq 1$. Se la $f_x(x)$ è uniforme tra 0 e 1, è immediato verificare che $f_y(y) = \varphi(y)$. È questo un semplice modo per generare variabili con densità $\varphi(x)$ qualsiasi a partire da una uniforme (facilmente generabile con routine di pseudo-noise): basta calcolarne la distribuzione integrale e invertirla, ottenendo così la $y=g(x)$ necessaria: se per x si pongono campioni distribuiti uniformemente, si ottengono per y i campioni voluti.

2.2.6 Variabili casuali multiple e covarianza

Si parla di variabili casuali multiple quando si associa una distribuzione di probabilità a più variabili (discrete o continue).

Analogamente alle variabili casuali semplici, ad esse è associata:

- (caso discreto) una probabilità per ogni combinazione dei possibili valori delle n variabili; la somma delle probabilità di tutte le combinazioni è 1

$$(2.116) \quad \sum_{i_1=1}^{N_1} \sum_{i_2=1}^{N_2} \dots \sum_{i_n=1}^{N_n} p_{i_1 i_2 \dots i_n} = 1$$

- (caso continuo) una densità di probabilità $f(x_1, x_2, \dots, x_n)$, il cui integrale, su tutte le n variabili, è 1

$$(2.117) \quad \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = 1$$

Integrando rispetto ad alcune variabili la funzione “mista” $f(x_1, x_2, \dots, x_n)$, tra $-\infty$ e ∞ , abbiamo la densità mista delle rimanenti ("densità marginali").

Nel seguito faremo riferimento solo alle variabili casuali continue. Analogamente al caso di una singola variabile, si può definire il valore aspettato di una qualsiasi funzione g delle n variabili come

$$(2.118) \quad E[g(x_1, x_2, \dots, x_n)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(x_1, x_2, \dots, x_n) \cdot f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n$$

Indichiamo¹⁵ con $f(x_1, x_2, \dots, x_k | x_{k+1}, x_{k+2}, \dots, x_n)$ la densità di $\{x_1, x_2, \dots, x_k\}$ condizionata da $\{x_{k+1}, x_{k+2}, \dots, x_n\}$. Si ha che

$$(2.119) \quad f(x_1, x_2, \dots, x_k | x_{k+1}, x_{k+2}, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_k, \dots, x_n)}{f(x_{k+1}, x_{k+2}, \dots, x_n)}$$

Da questa si ha

$$(2.120) \quad f(x_1 | x_2, x_3, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n)}{f(x_2, x_3, \dots, x_n)}$$

e quindi

$$(2.121) \quad f(x_1, x_2, \dots, x_n) = f(x_1 | x_2, x_3, \dots, x_n) \cdot f(x_2 | x_3, \dots, x_n) \cdot \dots \cdot f(x_{n-1} | x_n) \cdot f(x_n)$$

Diciamo che due variabili x e y sono **stocasticamente indipendenti** (o **statisticamente indipendenti**, o, semplicemente, **indipendenti**) se tutta l'informazione su x è contenuta nella marginale \mathbf{f}_x e tutta l'informazione su y nella marginale \mathbf{f}_y .

Se le n variabili casuali $\{x_i\}$ sono indipendenti, si ha

$$(2.122) \quad f(x_1, x_2, \dots, x_n) = \prod_{i=1}^n f_i(x_i)$$

e per la distribuzione integrale

$$(2.123) \quad F(x_1, x_2, \dots, x_n) = \prod_{i=1}^n F_i(x_i)$$

Prendiamo il semplice caso in cui la funzione \mathbf{g} sia la combinazione lineare di tutte le variabili

$$\mathbf{x}_i, \quad y = \sum_{i=1}^n a_i x_i,$$

$$(2.124) \quad E[y] = E\left[\sum_{i=1}^n a_i x_i\right] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} \sum_{i=1}^n a_i x_i \cdot f(x_1, x_2, \dots, x_n) dx_1 dx_2 \dots dx_n = \sum_{i=1}^n a_i E[x_i]$$

quindi il valore atteso della combinazione lineare di n variabili casuali (anche non indipendenti) è pari alla stessa combinazione lineare dei valori attesi (come potevamo intuire, data la linearità dell'operatore $E[\cdot]$ "valore atteso").

¹⁵ **Attenzione alla notazione:** per semplificarla, spesso qui indichiamo con f funzioni di densità che sono in genere diverse, dipendentemente dagli argomenti. In effetti ciascuna andrebbe indicata in modo differente (per esempio con opportuni indici).

Consideriamo per semplicità il caso in cui n sia eguale a 2 e chiamiamo x e y le due variabili. Le "densità marginali" sono le funzioni

$$(2.125) \quad f_x(x) = \int_{-\infty}^{\infty} f(x, y) dy$$

e

$$(2.126) \quad f_y(y) = \int_{-\infty}^{\infty} f(x, y) dx$$

Queste due funzioni sono due densità di probabilità che possono vedersi come le densità di probabilità associate alle due variabili x e y , indipendentemente l'una dall'altra; cioè, per esempio, la prima descrive le qualità statistiche della variabile x , se non si sa niente di y .

È immediato verificare che

$$(2.127) \quad E[x] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot f(x, y) dx dy = \int_{-\infty}^{\infty} x \cdot f_x(x) dx$$

e analogamente

$$(2.128) \quad E[y] = \int_{-\infty}^{\infty} y \cdot f_y(y) dy$$

Vediamo ora il valore aspettato del prodotto. In generale

$$(2.129) \quad E[x \cdot y] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f(x, y) dx dy$$

Se x e y sono indipendenti, si ha

$$(2.130) \quad E[x \cdot y] = \int_{-\infty}^{\infty} f_x(x) dx \int_{-\infty}^{\infty} f_y(y) dy = E[x] \cdot E[y]$$

Torniamo ora alla combinazione lineare $y = \sum_{i=1}^n a_i x_i$. Siano $\mu_i = E[x_i]$ e σ_i^2 le varianze delle n variabili \mathbf{x}_i . Abbiamo visto che $\mu_y = E[y] = \sum_{i=1}^n a_i \mu_i$. Sviluppriamo ora l'espressione della varianza di y . Si ha

$$(2.131) \quad \begin{aligned} \sigma_y^2 &= E \left[\left(\sum_{i=1}^n a_i (x_i - \mu_i) \right)^2 \right] = E \left[\sum_{i=1}^n \sum_{j=1}^n a_i a_j (x_i - \mu_i)(x_j - \mu_j) \right] = \\ &= \sum_{i=1}^n \sum_{j=1}^n a_i a_j E \left[(x_i - \mu_i)(x_j - \mu_j) \right] \end{aligned}$$

Indichiamo ora col termine **covarianza** delle variabili \mathbf{x}_i e \mathbf{x}_j il valore

$$(2.132) \quad \sigma_{ij} = E[(x_i - \mu_i)(x_j - \mu_j)] = E[x_i x_j] - \mu_i \mu_j$$

Si noti che $\sigma_{ii} = \sigma_i^2$. Dalla (2.130) è immediato dimostrare che se \mathbf{x}_i e \mathbf{x}_j sono indipendenti la loro covarianza è nulla; altrimenti può essere positiva o negativa. **Se due variabili hanno covarianza nulla, non è detto in genere che siano indipendenti.**

Se quindi le n variabili \mathbf{x}_i sono indipendenti, la varianza della combinazione lineare si riduce a

$$(2.133) \quad \sigma_y^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$$

Se abbiamo n variabili indipendenti con la stessa varianza σ^2 , abbiamo che la varianza della somma è $n \cdot \sigma^2$ e la varianza della media è $\frac{\sigma^2}{n}$. La deviazione standard sulla media è

$$(2.134) \quad \sigma_y = \frac{\sigma}{\sqrt{n}}$$

Tutte le covarianze delle n variabili \mathbf{x}_i formano una matrice quadrata

$$(2.135) \quad C = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}$$

detta **matrice di covarianza**; sulla diagonale ha le varianze delle n variabili. Se le \mathbf{x}_i sono indipendenti, la matrice è diagonale.

Possiamo quindi esprimere la varianza di y come

$$(2.136) \quad \sigma_y^2 = \sum_{i=1}^n \sum_{j=1}^n a_i a_j \sigma_{ij}$$

Se le \mathbf{x}_i sono indipendenti, otteniamo il risultato, già anticipato,

$$(2.137) \quad \sigma_y^2 = \sum_{i=1}^n a_i^2 \sigma_i^2$$

Date due variabili x e y, il valore della covarianza σ_{xy} dipende anche dalle varianze di x e y. Per indicare il tipo di dipendenza evidenziata da σ_{xy} in una forma indipendente dalla varianza

di x e y (e da eventuali fattori di amplificazione), si è introdotto un parametro adimensionale, il **coefficiente di correlazione**

$$(2.138) \quad \rho = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

Ricordiamo che la **disuguaglianza di Schwartz** (o di **Cauchy-Schwartz**) è

$$(2.139) \quad \left| \int_{-\infty}^{\infty} x(t) \cdot y(t) \cdot dt \right|^2 \leq \int_{-\infty}^{\infty} |x(t)|^2 dt \cdot \int_{-\infty}^{\infty} |y(t)|^2 dt$$

e, nel discreto,

$$(2.140) \quad \left| \sum_i x_i \cdot y_i \right|^2 \leq \sum_i |x_i|^2 \cdot \sum_i |y_i|^2$$

L'uguaglianza si ha se y è proporzionale a x .

Questo risultato, che utilizzeremo più volte in questo corso, può scriversi anche

$$(2.141) \quad E[x \cdot y] \leq \sqrt{E[x^2] \cdot E[y^2]}$$

Quindi ρ , calcolato nella (2.138), è un numero compreso tra -1 e 1.

Introduciamo ora la distribuzione gaussiana per variabili casuali multiple.

Se le variabili sono indipendenti, a media nulla e con la stessa varianza, allora

$$(2.142) \quad f(x_1, x_2, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} \cdot \sigma^n} \cdot e^{-\frac{x_1^2 + x_2^2 + \dots + x_n^2}{2\sigma^2}}$$

Nel caso di due variabili abbiamo

$$(2.143) \quad f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} \exp \left[-\frac{1}{2(1-\rho^2)} \cdot \left(\frac{(x-\mu_x)^2}{\sigma_x^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2} \right) \right]$$

dove ρ indica il coefficiente di correlazione tra le due variabili.

Nel caso generale (ma con i valori aspettati delle variabili tutti nulli, situazione a cui comunque si arriva con una banale trasformazione di variabile) la densità congiunta ha la forma

$$(2.144) \quad f(x_1, x_2, \dots, x_n) = k \cdot \exp\left(-\sum_{i,j=1}^n a_{ij} \cdot x_i \cdot x_j\right)$$

ovvero, essendo \mathbf{C} la matrice della covarianza e \mathbf{C}^{-1} la sua inversa,

$$(2.145) \quad f(x_1, x_2, \dots, x_n) = \frac{\exp\left(-\frac{1}{2} \cdot \mathbf{x} \mathbf{C}^{-1} \mathbf{x}^T\right)}{\sqrt{(2\pi)^n |\mathbf{C}|}}$$

Si dimostra facilmente dalla (2.145) che, nel caso normale, se le variabili sono scorrelate (cioè se la matrice \mathbf{C} ha tutti gli elementi nulli a parte la diagonale), sono indipendenti. Infatti, se

$$(2.146) \quad \mathbf{C} = \begin{pmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \sigma_n^2 \end{pmatrix}$$

e quindi

$$(2.147) \quad \mathbf{C}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2} & 0 & \dots & 0 \\ 0 & \frac{1}{\sigma_2^2} & \dots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & \frac{1}{\sigma_n^2} \end{pmatrix}$$

e la (2.145) diventa

$$(2.148) \quad f(x_1, x_2, \dots, x_n) = \frac{\exp\left(-\frac{1}{2} \cdot \mathbf{x} \mathbf{C}^{-1} \mathbf{x}^T\right)}{\sqrt{(2\pi)^n |\mathbf{C}|}} = \prod_{i=1}^n \frac{\exp\left(-\frac{x_i^2}{2 \cdot \sigma_i^2}\right)}{\sqrt{2\pi \cdot \sigma_i}}$$

Come abbiamo già detto, nel caso generale, se è vero che due variabili indipendenti sono scorrelate, non è vero il viceversa. Un esempio è quello in cui esista una relazione deterministica tra x e y del tipo

$$(2.149) \quad y = g(|x|)$$

(e quindi le variabili non sono assolutamente indipendenti) e la distribuzione di x sia pari ($f(x) = f(-x)$). In tal caso è facile dimostrare che la correlazione è nulla.

2.3 Statistica

Il calcolo delle probabilità dà gli strumenti per descrivere teoricamente un fenomeno la cui occorrenza o alcuni suoi aspetti sono basati sul caso.

La statistica ci dà gli strumenti per acquisire una conoscenza sperimentale di tale fenomeno.

In effetti la statistica si è sviluppata per descrivere i caratteri di una **popolazione** a partire dall'osservazione di un sotto-insieme dei suoi "individui", detto **campione**.

Nelle scienze sperimentali definiamo **campione** una successione di osservazioni sperimentali (o misure) $\{x_1, x_2, \dots, x_n\}$ relative ad un dato fenomeno. Queste sono rappresentative del fenomeno, di cui si suppone esista una distribuzione teorica (che è analoga alla distribuzione della popolazione). Si fa in genere sotto l'ipotesi di stazionarietà del fenomeno in osservazione, cioè sotto l'ipotesi che il fenomeno non cambi nel tempo, a parte la presenza di **errori sperimentali** casuali.

Il campione viene quindi analizzato con tecniche statistiche, che sono classificate in **statistica descrittiva** e **statistica inferenziale**.

2.3.1 L'inferenza statistica - Stima dei parametri

L'inferenza è il processo con cui, a partire da dati sperimentali, valutiamo ("stimiamo") il valore di uno o più parametri di nostro interesse, oppure **decidiamo** sulla correttezza di certe ipotesi. Se i dati sperimentali sono estratti come campione da una popolazione o se essi sono affetti da errori casuali, parliamo di inferenza statistica.

Nelle scienze sperimentali esistono due distinte procedure sperimentali:

- gli **esperimenti controllati** (o "di laboratorio"; qualcuno li chiama "galileiani"), in cui fissiamo alcuni parametri sperimentali e ne misuriamo altri. Per esempio in un esperimento sulla caduta dei gravi, scegliamo un luogo geografico e una altezza da cui far cadere un oggetto e misurarne il tempo che ci mette a raggiungere il suolo. In questo caso usiamo l'inferenza statistica a causa delle incertezze di misura.
- gli **esperimenti osservativi**, in cui gli oggetti della sperimentazione non sono sotto il nostro controllo e quindi noi possiamo limitarci a misurarne ("osservarne") caratteristiche "campionarie" ed inferire leggi tra di esse. È questo per esempio il caso dell'astrofisica (c'è qualche difficoltà a portare in laboratorio una stella e a controllarne qualche caratteristica, per esempio la massa). In questo caso usiamo l'inferenza statistica sia perché operiamo su un campione di una popolazione, sia perché sono presenti incertezze di misura, sia perché ci può essere l'influenza di altri fattori non sotto il nostro controllo).

In generale, supponiamo di dover stimare un parametro θ da un campione sperimentale $\{x_1, x_2, \dots, x_n\}$, ottenendo la **stima** $\hat{\theta}$ (d'ora in poi indicheremo con la cuspide " $\hat{}$ " il valore "stimato" di un parametro). Realizziamo ciò tramite una funzione f , detta anche "statistica" (*statistic* in Inglese)

$$(2.150) \quad \hat{\theta} = f(x_1, x_2, \dots, x_n)$$

In linea di principio, la funzione f può essere qualsiasi: il problema è avere un buon “stimatore” (ancora meglio, lo stimatore “ottimo”).

Il valore stimato $\hat{\theta}$ può essere visto come una variabile casuale. Alcune “qualità” auspicabili di uno stimatore sono le seguenti:

- la **consistenza**, cioè se, al tendere ad infinito della dimensione n del campione, il valore della stima tende al valore del parametro. Questa è, ovviamente, una qualità fondamentale.
- la **correttezza**, o assenza di **distorsione** (bias in inglese), se $E[\hat{\theta}] = \theta$.
- l'**efficienza**, cioè quanto rapidamente $\hat{\theta}$ converge a θ al crescere di n ; è, cioè, per una data dimensione del campione, tanto maggiore quanto minore è la deviazione standard dello stimatore.
- la **robustezza**, cioè quanto le qualità di uno stimatore sono indipendenti dalla distribuzione dei dati a cui è applicato.

Notiamo che la “correttezza” è analoga all’accuratezza delle misure fisiche, mentre l’“efficienza” è analoga alla precisione

Per “costruire” buoni stimatori si utilizzano vari metodi, per esempio il **metodo dei minimi quadrati**, il **metodo della massima verosimiglianza** o la **stima bayesiana**. Vedremo brevemente in cosa consistono.

2.3.2 Il Principio dei Minimi Quadrati (e del minimo χ^2)

Questo principio per costruire stimatori fu introdotto da Gauss per risolvere il problema della stima dei parametri orbitali del pianetino Cerere (scoperto nel 1801 a Palermo da Piazzi, ma poi “perso”), a partire da dati con errori casuali.

Supponiamo di avere n osservazioni $\{x_1, x_2, \dots, x_n\}$ di una grandezza g dipendente da un insieme di parametri $\vec{\theta} = \{\theta_1, \theta_2, \dots, \theta_m\}$ e siano ciascuna di esse

$$(2.151) \quad x_i = g(i; \vec{\theta}) + \varepsilon_i$$

dove ε_i sono errori di misura casuali. Lo stimatore ai minimi quadrati di θ è quella funzione $\hat{\theta} = \hat{\theta}(x_1, x_2, \dots, x_n)$ tale che sia minima la somma

$$(2.152) \quad \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (x_i - g(i; \vec{\theta}))^2$$

detta “costo”. Ovviamente questo principio si può generalizzare a una diversa forma della **funzione costo**

$$(2.153) \quad \sum_{i=1}^n \beta(\varepsilon_i)$$

Con la forma quadratica di (2.152), i calcoli e i risultati sono particolarmente eleganti, ma in alcuni casi sono più utili forme diverse; per esempio con

$$(2.154) \quad \beta(\varepsilon_i) = |\varepsilon_i|^{n>2}$$

si fa “pagare” di più un singolo scostamento grande che tanti piccoli.

La soluzione del problema può farsi anche per via numerica, calcolando la (2.152) (o l’analogia con una differente funzione di costo) su un opportuno reticolato nello spazio del vettore dei parametri.

Si noti che questo problema può essere visto anche come la ricerca del “best fit” delle osservazioni $\{x_1, x_2, \dots, x_n\}$ con la funzione g al variare dei parametri $\{\theta_1, \theta_2, \dots, \theta_m\}$.

Un diverso modo di porre il problema, che ci sarà molto utile nel seguito, è il seguente:

Siano date $n+1$ variabili casuali $\{y, x_1, x_2, \dots, x_n\}$ e si conosca la densità di probabilità $f(y, x_1, \dots, x_n)$. Si voglia trovare lo stimatore $g(x_1, x_2, \dots, x_n)$ di y conoscendo le x , tale che sia minimo l’errore quadratico

$$(2.155) \quad \varepsilon^2 = E\left[\left(y - g(x_1, x_2, \dots, x_n)\right)^2\right]$$

Si dimostra che

$$(2.156) \quad g(x_1, x_2, \dots, x_n) = E[y | x_1, x_2, \dots, x_n] = \int_{-\infty}^{\infty} y \cdot f(y | x_1, x_2, \dots, x_n) \cdot dy$$

Il problema si semplifica notevolmente se si impone che la g sia una funzione lineare

$$(2.157) \quad g(x_1, x_2, \dots, x_n) = a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n$$

In tal caso si trova che i valori delle costanti a_i che rendono minimo l’errore quadratico ε^2 dato dalla (2.155), sono tali che $y - (a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n)$ sia ortogonale a $\{x_i\}$ (**Principio di Ortogonalità**), ovvero

$$(2.158) \quad E\left[\left(y - (a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n)\right) \cdot x_i\right] = 0$$

per tutti gli i . Infatti per rendere minimo ε^2 si deve avere, per tutti gli i ,

$$(2.159) \quad \frac{\partial \varepsilon^2}{\partial a_i} = \frac{\partial E \left[\left(y - (a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n) \right)^2 \right]}{\partial a_i} = 0$$

Per la linearità dell'operatore E, abbiamo che la derivata del valore atteso è uguale al valore atteso della derivata e quindi

$$(2.160) \quad \frac{\partial \varepsilon^2}{\partial a_i} = -2 \cdot E \left[\left(y - (a_1 \cdot x_1 + a_2 \cdot x_2 + \dots + a_n \cdot x_n) \right) \cdot x_i \right] = 0$$

da cui si deduce la (2.158).

Ora, posto

$$(2.161) \quad R_{ij} = E \left[x_i \cdot x_j \right]$$

e

$$(2.162) \quad R_{0i} = E \left[y \cdot x_i \right]$$

(ovviamente la matrice R_{ij} è simmetrica) abbiamo che il sistema (2.158) si può scrivere come

$$(2.163) \quad \begin{cases} R_{11} \cdot a_1 + R_{12} \cdot a_2 + \dots + R_{1n} \cdot a_n = R_{01} \\ R_{21} \cdot a_1 + R_{22} \cdot a_2 + \dots + R_{2n} \cdot a_n = R_{02} \\ \dots \\ R_{n1} \cdot a_1 + R_{n2} \cdot a_2 + \dots + R_{nn} \cdot a_n = R_{0n} \end{cases}$$

Questa è una conseguenza della felice scelta di una funzione di costo quadratica.

Nel caso in cui le $\{y, x_1, x_2, \dots, x_n\}$ sono normali miste, si dimostra che lo stimatore ottimo è lineare del tipo (2.157).

Il principio dei minimi quadrati si applica quando gli “errori di misura” hanno la stessa varianza per tutte le osservazioni. Quando, come è spesso, non è così, il principio si generalizza al **principio del minimo χ^2** (che deriva dal principio della massima verosimiglianza - vedi in seguito), dove il χ^2 è la somma dei quadrati degli errori normalizzati con le proprie deviazioni standard.

2.3.3 Il Principio della Massima Verosimiglianza

Supponiamo di osservare un campione $\{x_1, x_2, \dots, x_n\}$ estratto, in modo che i singoli elementi siano indipendenti, da una popolazione descritta da

- (caso discreto) una distribuzione di probabilità $\{P_i\}$
- (caso continuo) una densità di probabilità $f(x)$

Siano nei due casi rispettivamente la distribuzione di probabilità o la densità di probabilità note a parte un parametro θ che vogliamo stimare dal campione osservato.

Definiamo **funzione di verosimiglianza (likelihood)**, nel caso discreto,

$$(2.164) \quad L(\theta; x_1, x_2, \dots, x_n) = \prod_i P(x_i | \theta)$$

e nel caso continuo

$$(2.165) \quad L(\theta; x_1, x_2, \dots, x_n) = \prod_i f(x_i | \theta)$$

La $L(\theta; x_1, x_2, \dots, x_n)$, se le $\{x_1, x_2, \dots, x_n\}$ sono prese come variabili e θ è un parametro, è una distribuzione di probabilità (corretta solo se le $\{x_1, x_2, \dots, x_n\}$ sono indipendenti). Come funzione di θ ed essendo $\{x_1, x_2, \dots, x_n\}$ risultati dell'esperimento (il campione), ovviamente no.

Una molto diffusa tecnica per costruire uno stimatore del parametro θ consiste nel massimizzare l'espressione della verosimiglianza. La distribuzione col parametro θ così determinato è quella con la massima probabilità condizionata dall'aver prodotto il campione osservato e quindi il più *verosimile*.

Gli stimatori costruiti con questo metodo sono sempre consistenti, ma possono essere distorti.

Poiché la $L(\theta)$ è sempre positiva, se ne può prendere il logaritmo

$$(2.166) \quad l(\theta) = \log(L(\theta))$$

e massimizzare questa nuova funzione che spesso è più semplice, ottenendo ovviamente lo stesso risultato.

Supponiamo che il parametro θ sia il valore atteso delle x_i e quindi le $f(x_i | \theta)$ siano normali, con parametri (θ, σ_i) . La verosimiglianza è

$$(2.167) \quad L(\theta; x_1, x_2, \dots, x_n) = \prod_i \frac{1}{\sqrt{2\pi} \cdot \sigma_i} \exp\left(-\frac{(x_i - \theta)^2}{2\sigma_i^2}\right) =$$

$$= \left(\frac{1}{(2\pi)^{\frac{n}{2}}} \prod_i \frac{1}{\sigma_i}\right) \cdot \exp\left(-\sum_i \frac{(x_i - \theta)^2}{2\sigma_i^2}\right)$$

Si vede che massimizzare la verosimiglianza equivale a minimizzare $\sum_i \frac{(x_i - \theta)^2}{2\sigma_i^2}$ (la parte fuori dell'esponenziale è inessenziale per la massimizzazione), quindi, se le σ_i sono tutte eguali, il criterio della massima verosimiglianza equivale a quello dei minimi quadrati. Lo stimatore in questo caso non è altro che

$$(2.168) \quad \hat{\theta} = \frac{1}{n} \cdot \sum_{i=1}^n x_i$$

Se le σ_i non sono tutte eguali, il massimo di L nella (2.167) si ha per

$$(2.169) \quad \frac{\partial \left(\sum_i \frac{(x_i - \theta)^2}{2\sigma_i^2} \right)}{\partial \theta} = 0$$

da cui

$$(2.170) \quad \sum_i \frac{(x_i - \theta)}{\sigma_i^2} = 0$$

e quindi

$$(2.171) \quad \hat{\theta} = \frac{\sum_{i=1}^n \frac{x_i}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}$$

cioè la migliore stima di θ , secondo il criterio della massima verosimiglianza, si ha con una media pesata (o ponderata)

$$(2.172) \quad \hat{\theta} = \sum_{i=1}^n a_i \cdot x_i$$

con

$$(2.173) \quad a_i = \frac{\frac{1}{\sigma_i^2}}{\sum_{i=1}^n \frac{1}{\sigma_i^2}}$$

2.3.4 La Stima Bayesiana

Ricordiamo brevemente il teorema di Bayes.

Supponiamo di avere un evento \mathbf{A} che può essere causato da uno ed uno solo di \mathbf{N} eventi $\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_N$, mutuamente esclusivi. Supponiamo di conoscere le \mathbf{N} probabilità condizionate $\mathbf{P}(\mathbf{A} | \mathbf{B}_i)$ che capitano \mathbf{A} se è capitato \mathbf{B}_i e le \mathbf{N} probabilità $\mathbf{P}(\mathbf{B}_i)$.

Se osserviamo \mathbf{A} , possiamo chiederci quale degli \mathbf{N} eventi \mathbf{B}_i lo ha causato, o meglio quale è la probabilità che è accaduto \mathbf{B}_i avendo osservato \mathbf{A} .

Si dimostra che (Teorema di Bayes)

$$(2.174) \quad P(B_i | A) = \frac{P(B_i) \cdot P(A | B_i)}{\sum_{k=1}^N P(B_k) \cdot P(A | B_k)}$$

Si noti che, con questa relazione, si passa dalle probabilità $\mathbf{P}(\mathbf{A} | \mathbf{B}_i)$ alle $\mathbf{P}(\mathbf{B}_i | \mathbf{A})$, cioè dall'informazione sull'effetto delle \mathbf{B}_i su \mathbf{A} , tramite anche le $\mathbf{P}(\mathbf{B}_i)$, a "inferire" sulle cause dell'evento osservato \mathbf{A} .

Il teorema di Bayes, nel caso di una variabile casuale continua, diventa

$$(2.175) \quad f(x | y) = \frac{f_1(y | x) \cdot f_0(x)}{\int_{-\infty}^{\infty} f_1(y | x) \cdot f_0(x) \cdot dx}$$

dove f_0 è la distribuzione della variabile x e f_1 la distribuzione condizionale di y dato x .

Vediamo ora come si costruisce uno stimatore bayesiano.

Supponiamo che il parametro θ da stimare si possa considerare, prima di eseguire l'esperimento statistico, una variabile casuale definita da una certa **densità di probabilità a priori** $f_\theta(\theta)$. Allora $f(x | \theta)$ è la densità di probabilità condizionata di x dato θ e possiamo scrivere

$$(2.176) \quad f(x_1, x_2, \dots, x_n, \theta) = f(x_1, x_2, \dots, x_n | \theta) \cdot f_\theta(\theta)$$

essendo $\{x_1, x_2, \dots, x_n\}$ il campione, e calcolare la marginale

$$(2.177) \quad f(x_1, x_2, \dots, x_n) = \int_{\text{dom}(\theta)} f(x_1, x_2, \dots, x_n, \theta) \cdot d\theta$$

Ricaviamo quindi col teorema di Bayes

$$(2.178) \quad f(\theta | x_1, x_2, \dots, x_n) = \frac{f(x_1, x_2, \dots, x_n, \theta)}{f(x_1, x_2, \dots, x_n)} = \frac{f(x_1, x_2, \dots, x_n | \theta) \cdot f_\theta(\theta)}{f(x_1, x_2, \dots, x_n)}$$

ottenendo così lo stimatore, come

$$(2.179) \quad \hat{\theta} = E[f(\theta | x_1, x_2, \dots, x_n)]$$

Facciamo un esempio. Abbia il parametro θ da stimare una densità a priori normale

$$(2.180) \quad f_\theta(\theta) = N(\mu_0, \sigma_0)$$

.... [pag 258 Pap III]

La stima bayesiana integra informazioni *a priori* sulla grandezza da stimare e osservazioni (misure sperimentali), in modo da utilizzare al meglio tutte le informazioni disponibili.

Si noti che le informazioni a priori possono essere ricavate anche da misure precedenti; esse sono condensate nella **distribuzione a priori** sulla grandezza in esame. Da questa e dal risultato delle misure cambia lo stato di informazione, che è condensato nella **distribuzione a posteriori**. La teoria della stima bayesiana indica il modo di calcolarla.

Si può quindi utilizzare quest'ultima come distribuzione a priori per successive misure.

Se non si sa nulla "a priori", a parte vagamente il possibile intervallo della grandezza, si suppone una distribuzione a priori uniforme nell'intervallo. Ma in questo caso questa stima è meno interessante.

2.3.5 Stima del valor medio

Supponiamo di dover eseguire la misura di una grandezza fisica. Facciamo ciò eseguendo una certa procedura sperimentale e la ripetiamo più volte. Partendo dai risultati di queste operazioni ripetute, ci possiamo chiedere:

- quale è il valore vero della misura della grandezza in esame ?
- quale è l'incertezza del valore che forniamo ?
- quale è l'intervallo di fiducia, cioè l'intervallo, relativo ad una certa probabilità, che possiamo dire contenga il valore vero, con quella probabilità ?

Intuitivamente sappiamo che un buon stimatore del "valore vero" di un misurando (in linguaggio statistico, il "valore atteso"), in presenza di errori casuali (e quindi a media nulla¹⁶), è la media dei risultati di successive misure. Ora possiamo chiederci: è questo un buon stimatore ? Verifichiamone la correttezza:

¹⁶ Se il valor medio degli errori non fosse nullo, tale valore sarebbe considerato un errore sistematico.

$$(2.181) \quad E[\bar{x}] = E\left[\frac{1}{n} \sum_{i=1}^n x_i\right] = \frac{1}{n} \sum_{i=1}^n E[x_i] = \mu$$

quindi, come speravamo è corretto (e quindi anche consistente). Non discutiamo qui altre qualità (alcune delle quali dipendono dalla distribuzione degli errori casuali, che non è detto che siano sempre gaussiani).

Per l'incertezza sulla media, (vedi che, se la rappresentiamo con la deviazione standard, dall'equazione (2.134), è $\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$).

Se stabiliamo una probabilità φ , che chiamiamo il **livello di fiducia** o **livello di confidenza**, quale è, a partire dalle nostre misure, l'intervallo **(a,b)** (detto **intervallo di confidenza**) entro cui sta il valore vero?

Supponiamo che i nostri errori siano gaussiani e che conosciamo la loro deviazione standard σ_x . Allora, trovato con le tavole il valore **z** della variabile standardizzata che corrisponde ad un intervallo simmetrico di probabilità φ , diciamo che l'intervallo relativo al livello di fiducia φ è

$$(2.182) \quad (\bar{x} - z \cdot \sigma_{\bar{x}}, \bar{x} + z \cdot \sigma_{\bar{x}})$$

Nel caso che la distribuzione non sia gaussiana, se il numero di misure è abbastanza elevato, per il teorema del limite centrale questa formula è una buona approssimazione.

Se la deviazione standard non è nota, la si può stimare (si veda il prossimo paragrafo), ma in questo caso il calcolo dell'intervallo di fiducia è più complesso (occorre introdurre la distribuzione t di Student; è questa una distribuzione con un parametro detto numero dei gradi di libertà, che "parte" da quella di Cauchy e "arriva" asintoticamente a quella di Gauss); tuttavia, se il campione è abbastanza numeroso la formula che abbiamo dato è una buona approssimazione.

Una volta stimato l'intervallo di confidenza (dato un livello di fiducia di nostra scelta), possiamo decidere se i nostri risultati sperimentali sono compatibili con una aspettativa teorica (a priori). Se il valore teorico è entro l'intervallo di fiducia, allora diciamo che le nostre misure sono compatibili con esso, o anche che l'esperimento conferma la teoria, altrimenti lo dobbiamo escludere (e dobbiamo cambiare teoria o correggere errori sperimentali; c'è anche l'eventualità che siamo di fronte a una fluttuazione statistica: in questo caso se ripetiamo l'esperimento dovremmo avere risultati compatibili).

Nota: Il livello di fiducia indica quanto siamo al sicuro da inevitabili fluttuazioni. Porre il livello di fiducia **p** per esempio a **0.95** significa che questa è la probabilità di non essere tratti in errore dalle fluttuazioni: **ma con questo livello di fiducia il 5 % degli sperimentatori che operano correttamente daranno il risultato sbagliato**. Ovviamente miglioriamo le cose alzando il livello di fiducia **p**, per esempio, a **0.99**: ma in questo caso nell' **1 %** dei casi dei buoni sperimentatori daranno dei risultati errati. Se si ripete l'esperimento però, le cose migliorano: la probabilità di riavere la fluttuazione ingannevole due volte di seguito è molto bassa (in genere inferiore a $[1 - (1 - p)^2]$).

Il livello di fiducia è un parametro che qualifica probabilisticamente l'incertezza di una misura. Una misura viene quindi descritta da tre numeri: il valore stimato, l'incertezza su tale valore, che determina un intervallo, e il livello di fiducia che associa una probabilità a questo intervallo.

Attenzione ! spesso anche il valore teorico ha un'incertezza. In tal caso dobbiamo calcolare gli intervalli di confidenza sia per il valore teorico (\mathbf{v}_T) che per il valore sperimentale (\mathbf{v}_S). In effetti la cosa si semplifica se le due incertezze sono entrambe assimilabili a due deviazioni standard (di distribuzioni gaussiane). Allora possiamo costruire la differenza $\mathbf{d} = \mathbf{v}_S - \mathbf{v}_T$. Nell'ipotesi che la teoria sia in accordo con l'esperimento, ci aspetteremmo $d = 0$, con una varianza pari alla somma delle due varianze. Quindi, dato il nostro livello di fiducia, possiamo costruirci l'intervallo di fiducia (che sarà simmetrico rispetto a 0): se il d trovato è dentro questo intervallo, allora la teoria è in accordo con l'esperimento, altrimenti no.

2.3.6 Stima della varianza

Nel capitolo 5 abbiamo dato questa definizione di varianza campionaria

$$(2.183) \quad \tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

ma è questa una buon stimatore della varianza della popolazione da cui il campione $\{x_i\}$ è estratto ? Prendiamone il valore aspettato

$$(2.184) \quad E \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right] = E \left[\frac{1}{n} \sum_{i=1}^n \left\{ (x_i - \mu) - (\bar{x} - \mu) \right\}^2 \right] = \\ = \frac{1}{n} E \left[\sum_{i=1}^n (x_i - \mu)^2 - 2 \cdot \sum_{i=1}^n (x_i - \mu)(\bar{x} - \mu) + \sum_{i=1}^n (\bar{x} - \mu)^2 \right]$$

ora per il termine misto si ha

$$(2.185) \quad 2 \cdot \sum_{i=1}^n (x_i - \mu)(\bar{x} - \mu) = 2 \cdot (\bar{x} - \mu) \cdot \sum_{i=1}^n (x_i - \mu) = 2 \cdot n \cdot (\bar{x} - \mu)^2$$

e quindi, ricordando che la varianza sulla media è n volte minore della varianza sul singolo,

$$(2.186) \quad E \left[\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \right] = E \left[\frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2 - (\bar{x} - \mu)^2 \right] = \sigma^2 - \frac{\sigma^2}{n} = \frac{n-1}{n} \cdot \sigma^2$$

Come si vede il valore aspettato dello stimatore del capitolo 5 non è la varianza della popolazione. Quindi questo stimatore è distorto, anche se per grandi valori di n , la distorsione è molto piccola. Esso dà mediamente un valore un po' più piccolo di quello vero. Preferiamo perciò, come stimatore della varianza, il seguente

$$(2.187) \quad \hat{\sigma}^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$$

che è corretto (“unbiased”, in inglese).

Se conosciamo il valore atteso μ , lo stimatore

$$(2.188) \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \mu)^2$$

è corretto.

Il calcolo della varianza sulla varianza campionaria è piuttosto complicato. Diamo qui, per riferimento, il risultato

$$(2.189) \quad \sigma_{\hat{\sigma}^2}^2 = \frac{1}{n} \left(\mu^{(4)} - \frac{n-3}{n-1} \sigma^4 \right)$$

dove $\mu^{(4)}$ è il momento centrale del quarto ordine. Se la distribuzione è gaussiana, si ha

$$(2.190) \quad \sigma_{\hat{\sigma}^2}^2 = \frac{2 \cdot \sigma^4}{n-1}$$

e la deviazione standard (che può essere presa come “incertezza” nella stima della varianza)

$$(2.191) \quad \sigma_{\hat{\sigma}^2} = \sigma^2 \cdot \sqrt{\frac{2}{n-1}}$$

L’incertezza relativa sulla varianza è quindi $\sqrt{\frac{2}{n-1}}$ e, usando la solita regola per la propagazione dell’incertezza percentuale, l’incertezza relativa sulla deviazione standard è la metà della precedente, cioè $\sqrt{\frac{1}{2(n-1)}}$.

2.3.7 Esempi di stima ai minimi quadrati: stima dei parametri di una retta sperimentale (“fit lineare”)

In un grafico che rappresenta punti sperimentali che collegano due parametri con dei punti allineati affetti da errori casuali (per esempio la misura del periodo al quadrato (T^2) e la misura della massa appesa (M) nell’esperimento della molla), si vogliono stimare i parametri della retta che meglio li descrive e valutare l’incertezza su di essi.

È questo il più semplice caso di “fit” di dati sperimentali ed è detto il problema del fit lineare.

Nel caso più generale, il problema è quello di trovare i parametri di una funzione di un certo tipo in modo da approssimare al meglio i dati sperimentali. Gauss, che doveva trovare i parametri orbitali di un asteroide a partire dalle osservazioni, introdusse, per questo tipo di problemi, il **principio dei minimi quadrati**.

Limitandoci al nostro semplice caso lineare. Supponendo che tutte le misure abbiano la stessa incertezza, esso stabilisce che la migliore retta $y = m \cdot x + q$ è quella in cui la somma dei quadrati tra i valori sperimentali e quelli della retta sia minima.

Supponiamo di avere le n coppie di misure $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$, dove le misure x hanno errore trascurabile e le misure y hanno errore gaussiano con media 0 e deviazione standard σ ; per ciascuna di queste coppie scriviamo l'equazione

$$(2.192) \quad y_i = m \cdot x_i + q + \varepsilon_i$$

dove ε_i è l'errore sulla misura y_i . La somma dei quadrati degli errori è

$$(2.193) \quad n \cdot \overline{\varepsilon^2} = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - m \cdot x_i - q)^2$$

imponiamo quindi le condizioni per il minimo (annullamento della derivata prima)

$$\frac{\partial \overline{\varepsilon^2}}{\partial m} = 0 \quad \text{e} \quad \frac{\partial \overline{\varepsilon^2}}{\partial q} = 0$$

e risolviamo il sistema. Si trova per la stima di m

$$(2.194) \quad \hat{m} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\hat{\sigma}_{xy}}{\Delta}$$

dove

$$\overline{x \cdot y} = \frac{1}{n} \sum_{i=1}^n x_i y_i \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i \quad \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \overline{x^2} = \frac{1}{n} \sum_{i=1}^n x_i^2$$

Notare che il numeratore è la covarianza campionaria di x e y ; il denominatore $\Delta = \overline{x^2} - \bar{x}^2$, detto braccio, ha la forma di una varianza campionaria, anche se in effetti non lo è, perché x non è una variabile casuale (sono valori scelti da noi): essa comunque misura la "larghezza" dell'insieme dei dati x . Quindi, come è intuitivo, più è larga la base dei dati x , migliore è la stima della pendenza della retta.

Per ottenere la stima di q facciamo

$$(2.195) \quad \hat{q} = \bar{y} - \hat{m} \cdot \bar{x}$$

Per le incertezze, in forma di deviazione standard, abbiamo

$$(2.196) \quad \sigma_m = \frac{\sigma}{\sqrt{n \cdot \Delta}}$$

e

$$(2.197) \quad \sigma_q = \sigma_m \cdot \sqrt{x^2}$$

Attenzione ! queste formule sono ricavate nel caso in cui sono verificate le due ipotesi

- a) siano uguali tutte le incertezze sulle misure y
- b) siano trascurabili le incertezze sulle x

Se non è valida la prima condizione, possiamo ricavare altre espressioni per le stime e le loro incertezze, basate sul “principio del minimo χ^2 ”. Se non è valida la seconda condizione, il problema è molto più complesso (diventa non lineare) e si può risolvere o con trucchi (riportando per esempio l'incertezza sulla x sulla variabile y) o con procedure di ottimizzazione ricorsiva.

2.3.8 Esempi di stima ai minimi quadrati: fit lineare generale – fit polinomiale

Possiamo applicare il metodo del minimo χ^2 , o dei minimi quadrati che ne è un caso particolare, a un fit fatto con una combinazione lineare di funzioni $f_i(x)$. Per esempio, se vogliamo fare un fit con una funzione sinusoidale di nota frequenza, ma incognita fase e ampiezza, possiamo porre

$$(2.198) \quad y(x) = A \cdot \sin(\omega_0 x + \varphi) = a_1 \sin \omega_0 x + a_2 \cos \omega_0 x$$

quindi le funzioni base in questo caso sono $f_1(x) = \sin(\omega_0 x)$ e $f_2(x) = \cos(\omega_0 x)$. Se stimiamo quindi a_1 e a_2 possiamo ottenere

$$\hat{A} = \sqrt{\hat{a}_1^2 + \hat{a}_2^2}$$

$$\cos \hat{\varphi} = \frac{\hat{a}_1}{\hat{A}} \quad \sin \hat{\varphi} = \frac{\hat{a}_2}{\hat{A}}$$

Un altro caso notevole è quello del fit con un polinomio, in cui le funzioni base della combinazione lineare sono le $f_i(x) = x^i$, quindi

$$(2.199) \quad y(x) = a_1 + a_2 x + a_3 x^2 + \dots + a_M x^{M-1}$$

che, come è chiaro, è lineare nelle incognite a_i .

In questi casi possiamo operare analogamente al caso del fit lineare, minimizzando l'espressione

$$(2.200) \quad \chi^2(a_1, a_2, \dots, a_M) = \sum_{k=1}^n \left(\frac{y_k - \sum_{i=1}^M a_i f_i(x_k)}{\sigma_k} \right)^2$$

dove n è il numero dei punti sperimentali (o “osservazioni”) ed M il numero dei parametri liberi, x_k sono i valori dell'ascissa (noti senza incertezza) e σ_k le incertezze sulle osservazioni y_k (deviazioni standard dell'errore casuale atteso). Questa stima è particolarmente conveniente se gli errori di misura sono gaussiani.

In genere si deve avere $M \leq n$, cioè il numero di parametri del fit dovrebbe essere minore o al massimo uguale al numero di osservazioni, altrimenti il problema è indefinito.

Il minimo della (2.200) si calcola annullando le derivate rispetto agli M parametri. Diamo soltanto il risultato del calcolo.

Avendo posto il vettore \mathbf{a} il vettore le cui componenti sono gli M parametri a_i da stimare, la matrice \mathbf{A} $n \times M$ con elementi

$$(2.201) \quad A_{ki} = \frac{f_i(x_k)}{\sigma_k}$$

e il vettore \mathbf{y}' con n elementi

$$(2.202) \quad y_k' = \frac{y_k}{\sigma_k}$$

costruiamo la matrice MxM

$$(2.203) \quad \mathbf{B} = \mathbf{A}^T \mathbf{A}$$

e il vettore con M elementi

$$(2.204) \quad \mathbf{b} = \mathbf{A}^T \cdot \mathbf{y}'$$

Si ha che la stima dei parametri è data dalla soluzione del sistema di equazioni

$$(2.205) \quad \mathbf{B} \cdot \hat{\mathbf{a}} = \mathbf{b}$$

La matrice inversa

$$(2.206) \quad \mathbf{C} = \mathbf{B}^{-1}$$

è la matrice di covarianza dei parametri a_k . Gli elementi sulla diagonale sono le varianze sui parametri, e quindi le incertezze sono

$$(2.207) \quad \sigma_i = \sqrt{C_{ii}}$$

2.3.9 Media pesata

Se si hanno varie valutazioni della misura di una grandezza, con incertezze differenti, come possiamo utilizzare questi risultati per stimare al meglio il valore vero della grandezza ?

La risposta è sì. Supponiamo di avere n misure indipendenti m_i , di una grandezza M , ciascuna con incertezza (assimilata a una deviazione standard) s_i , e gli errori di misura siano normali. Come abbiamo visto nella (2.171), si ha

$$(2.208) \quad \hat{M} = \frac{\sum_{i=1}^n m_i}{\sum_{i=1}^n \frac{1}{s_i^2}}$$

Questa può essere vista come una media pesata con pesi

$$(2.209) \quad a_i = \frac{\frac{1}{s_i^2}}{\sum_{i=1}^n \frac{1}{s_i^2}}$$

ovviamente i pesi sono maggiori per le misure con incertezze minori. Si noti che il risultato non cambia se tutte le incertezze sono moltiplicate per uno stesso fattore (le a_i sono le stesse).

L'incertezza è data da

$$(2.210) \quad \Delta M = \frac{1}{\sqrt{\sum_{i=1}^n \frac{1}{s_i^2}}}$$

Se le incertezze sono tutte eguali, si ricava la solita espressione per l'incertezza delle misure ripetute.

2.3.10 Test statistici

Un problema centrale che si pone negli esperimenti scientifici è **decidere** se i risultati di un certo esperimento sono in accordo con la teoria. Per esempio, la traiettoria osservata di un corpo in caduta libera è in accordo con la legge di gravitazione? I risultati di un esperimento con il pallinometro sono in accordo con la formula teorica della distribuzione attesa delle frequenze dei vari bin?

In altri casi sperimentali, potremmo voler **testare** se ci sono differenze in due diverse procedure sperimentali, oppure **verificare** l'efficacia di un certo trattamento (per esempio l'assunzione di un farmaco per la cura di una certa malattia).

Analoghe domande si possono porre per esempio nel controllo di qualità dei prodotti (i pneumatici prodotti da un certo stabilimento hanno una durata maggiore di quelli di un altro?) o in altri campi della tecnica.

Questo tipo di problemi, che implicano una decisione (sì o no) è formalizzato nella teoria statistica dei test d'ipotesi, che è un altro aspetto dell'inferenza statistica. In questa teoria vengono prima di tutto formulate due ipotesi: l'ipotesi H_0 , detta **ipotesi nulla**, e l'ipotesi H_1 , detta **ipotesi alternativa**. Per esempio, l'ipotesi H_0 , che in genere è quella più precisamente definita, nel caso della verifica sperimentale di una teoria, potrebbe essere che ci sia perfetto accordo tra teoria ed esperimento (errore nullo); nel caso del farmaco potrebbe essere che non ci sia effetto (effetto nullo).

Lo scopo di un test di ipotesi è valutare se c'è sufficiente evidenza statistica per accettare l'ipotesi H_0 . Per far ciò calcoliamo una funzione del campione (la "statistica" del test) particolarmente utile per poter discriminare la presenza o l'assenza dell'effetto in oggetto. La distribuzione della statistica del test viene quindi utilizzata per definire l'insieme dei valori di accettazione (detto *regione di accettazione*) e il suo complementare, l'insieme dei valori di rigetto (detto *regione critica* o *di rigetto*).

Il risultato del test, cioè la nostra **decisione**, è sempre o l'accettazione o il rifiuto dell'ipotesi, a seconda che la statistica del test fornisca valori interni alla regione di accettazione o di rigetto.

Nel far ciò possiamo aver fatto la scelta giusta o compiuto un errore. Abbiamo il seguente schema:

Decisione statistica	Situazione reale	
	H_0 vera	H_0 falsa
Accettare H_0	$P = 1 - \alpha$	Errore di secondo tipo: $P = \beta$
Rifiutare H_0	Errore di primo tipo: $P = \alpha$	$P = 1 - \beta$

Vediamo che si possono commettere due tipi di errori:

- il primo, a cui è associata la probabilità α , chiamata **livello di significatività** del test, capita quando rifiutiamo un'ipotesi H_0 vera. Questa probabilità viene chiamata talvolta *probabilità di falso allarme*.
- il secondo, a cui è associata una probabilità β , capita quando accettiamo come buona una ipotesi H_0 falsa.

Notiamo che $1 - \alpha$ è equivalente al livello di fiducia φ introdotto nel capitolo precedente parlando della stima dell'intervallo di fiducia.

Possiamo decidere di diminuire la probabilità di un certo tipo di errore (cambiando la soglia di decisione), ma così facendo aumentiamo la probabilità dell'errore dell'altro tipo. La scelta della soglia può essere fatta associando a ciascun tipo di errore un **costo** (per esempio C_1 e C_2), e quindi porre la soglia di decisione in modo da minimizzare il suo valore atteso

$$(2.211) \quad E[C] = C_1 \cdot \alpha + C_2 \cdot \beta$$

La probabilità associata alla statistica osservata viene detta "livello di significatività osservato". Se questo valore è vicino al livello di significatività α , si può pensare, ove possibile, di ripetere il test.

Nella costruzione di un test statistico talora non si considerano gli errori di secondo tipo.

2.3.11 Test di consistenza con un valore teorico

Supponiamo di avere una misura sperimentale \mathbf{m} , con incertezza $\Delta\mathbf{m}$ (rappresentante la deviazione standard dell'errore casuale) e vogliamo decidere se è consistente con un valore teorico \mathbf{t} (con incertezza trascurabile). Definiamo *a priori* un livello di fiducia φ (o, equivalentemente, un livello di significatività $\alpha = 1 - \varphi$), e calcoliamo (con le tavole dell'integrale della gaussiana) il semi-intervallo di fiducia relativo z_φ .

Valutiamo quindi

$$(2.212) \quad z = \frac{|m - t|}{\Delta m}$$

e decidiamo sul superamento del test sulla base del valore di $z_\varphi - z$: se è positivo il test è superato (accettiamo l'ipotesi), altrimenti rigettiamo l'ipotesi.

2.3.12 Test di consistenza tra due valori sperimentali

Supponiamo di avere due misure \mathbf{m}_1 e \mathbf{m}_2 con incertezze relativamente $\Delta\mathbf{m}_1$ e $\Delta\mathbf{m}_2$ (o analogamente un valore sperimentale e uno teorico con incertezza non trascurabile). Ci domandiamo se sono compatibili, avendo definito un certo livello di fiducia.

Per costruire il test, costruiamo una nuova grandezza $\mathbf{d} = \mathbf{m}_1 - \mathbf{m}_2$ che ha incertezza

$\Delta d = \sqrt{\Delta m_1^2 + \Delta m_2^2}$. Se le due misure sono consistenti, il valore aspettato di \mathbf{d} è 0.

Costruiamo quindi l'intervallo di fiducia relativo al livello di fiducia e al $\Delta \mathbf{d}$, simmetricamente a \mathbf{d} , e verifichiamo se il valore 0 è interno od esterno a questo intervallo: nel primo caso accettiamo l'ipotesi, altrimenti la rigettiamo.

2.3.13 Test del χ^2

Se la verifica di un'ipotesi teorica corrisponde alla verifica della consistenza di più valori sperimentali, ognuno con la sua incertezza, con altrettanti valori teorici, è molto usato il **test del χ^2** (chi quadro).

Supponiamo per esempio di avere un grafico (x, y), in cui, in corrispondenza di n punti $\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n\}$ sulle ascisse, siano riportati n punti sperimentali $\{y_1^{(S)}, y_2^{(S)}, \dots, y_n^{(S)}\}$, con incertezze "gaussiane" $\{\sigma_1, \sigma_2, \dots, \sigma_n\}$, e si voglia testare la consistenza tra essi e gli n valori teorici $\{y_1^{(T)}, y_2^{(T)}, \dots, y_n^{(T)}\}$. Supponiamo inoltre che gli errori di misura siano indipendenti e che non siano state ricavate informazioni dai dati sperimentali per calcolare i valori teorici. In questo caso costruiamo la variabile

$$(2.213) \quad \chi^2 = \sum_{i=1}^n \frac{(y_i^{(S)} - y_i^{(T)})^2}{\sigma_i^2}$$

Notiamo che gli elementi della sommatoria sono quadrati di variabili normali standardizzate, quindi la variabile da noi costruita ha la distribuzione del χ^2 con \mathbf{n} gradi di libertà. Se quindi definiamo un livello di fiducia φ (o, equivalentemente, un livello di significatività $\alpha = 1 - \varphi$), possiamo calcolare quale sia il valore limite χ_{MAX}^2 tale che

$$(2.214) \quad \varphi = \int_0^{\chi_{MAX}^2} f_{\chi^2}^{(n)}(x) dx$$

dove $f_{\chi^2}^{(n)}(\cdot)$ è la densità di probabilità del χ^2 con n gradi di libertà. I valori di χ_{MAX}^2 per un dato valore del livello di fiducia e per un dato numero di gradi di libertà, sono tabulati su tavole come quella in appendice a questi appunti.

Se per calcolare i valori teorici abbiamo dovuto valutare \mathbf{m} parametri indipendenti di un'equazione a partire dagli \mathbf{n} numeri sperimentali, allora si abbassa il numero di gradi di libertà a $\mathbf{n} - \mathbf{m}$ e quindi dobbiamo usare la distribuzione del χ^2 con $\mathbf{n} - \mathbf{m}$ gradi di libertà.

Un caso particolare di test del χ^2 è quello che si usa per testare la consistenza tra una distribuzione teorica e un istogramma di frequenza.

A partire dalla distribuzione teorica e dal numero di dati istogrammati, calcoliamo i valori teorici delle frequenze degli $\mathbf{m} < \mathbf{n}$ bin, a partire dalle probabilità $\{p_1, p_2, \dots, p_n\}$ e

moltiplicandole per n ; siano questi valori teorici $\{h_1^{(T)}, h_2^{(T)}, \dots, h_m^{(T)}\}$. Siano invece $\{h_1^{(S)}, h_2^{(S)}, \dots, h_m^{(S)}\}$ le frequenze trovate sperimentalmente per ciascun bin, istogrammando gli n dati. Si noti che il parametro n lo possiamo ricavare sommando tutte le frequenze ottenute per i vari bin, quindi il numero di gradi di libertà si riduce di 1. Ulteriori riduzioni possono aversi se ci occorrono altri parametri per calcolare le \mathbf{p}_i .

Notiamo inoltre che le distribuzioni della frequenza nell' i -esimo bin è distribuita secondo una distribuzione binomiale con parametri n e \mathbf{p}_i , e, se i bin sono tanti, \mathbf{p}_i è piccolo e quindi la binomiale si può approssimare, in tal caso, ad una poissoniana; quindi la varianza è proprio $h_i^{(T)}$. Costruiamo quindi la variabile

$$(2.215) \quad \chi^2 = \sum_{i=1}^n \frac{(h_i^{(S)} - h_i^{(T)})^2}{h_i^{(T)}}$$

(se i bin sono pochi, l'approssimazione poissoniana porta a sottovalutare il valore di χ^2).

Attenzione però! Abbiamo visto che le fluttuazioni casuali (le differenze tra $h_i^{(S)}$ e $h_i^{(T)}$) sono distribuiti secondo Poisson, mentre per poter usare il test del χ^2 ed avere la distribuzione del χ^2 sulla variabile costruita, queste poissoniane devono essere approssimabili da gaussiane, cosa che capita se il loro μ è abbastanza elevato (per esempio > 20). Ma ciò, se può essere verificato per i bin centrali, quasi sempre non lo è per quelli estremi che sono tipicamente più poveri. Possiamo allora seguire una delle seguenti due strade:

- trascuriamo i bin periferici (in tal caso il numero di gradi di libertà è dato semplicemente dal numero dei bin considerati)
- "accorpriamo" (cioè sommiamo i loro contenuti) più bin insieme (non necessariamente adiacenti); in questo caso il numero di gradi di libertà si calcola nel modo normale.

3 Sistemi discreti

3.1 Introduzione

Un sistema discreto (o a dati campionati) è un elaboratore di segnali discreti (o campionati). Cioè un sistema che accetta in ingresso una (o più) successioni di dati campionati e produce in uscita una (o più) successioni di dati campionati. Analogamente a quanto accade per i sistemi continui, che si possono definire in generale tramite l'equazione (1.8), essi possono definirsi tramite l'equazione (1.9), che riportiamo:

$$(3.1) \quad \begin{cases} \vec{y}_i = F(\vec{x}_i, \vec{s}_i) \\ \vec{s}_i = G(\vec{x}_i, \vec{s}_{i-1}) \end{cases}$$

dove \vec{x}_i , \vec{y}_i e \vec{s}_i sono rispettivamente i vettori d'ingresso, di uscita e di stato all'istante (campionato) i . Le dimensioni dei suddetti vettori sono in genere differenti; nel seguito ci occuperemo essenzialmente di sistemi con un ingresso ed una uscita, cioè con dimensioni di x e y uguali ad 1.

Se le funzioni F e G non dipendono dal tempo (cioè dall'indice i), il sistema è tempo-invariante. Nel seguito considereremo quasi esclusivamente sistemi tempo-invarianti (detti anche stazionari): trascureremo quindi di specificarlo.

Diciamo che un sistema è lineare se le funzioni F e G sono lineari, cioè sono equazioni matriciali. In tal caso

$$(3.2) \quad \begin{cases} \vec{y}_i = \mathbf{A} \cdot \vec{x}_i + \mathbf{B} \cdot \vec{s}_i \\ \vec{s}_i = \mathbf{C} \cdot \vec{x}_i + \mathbf{D} \cdot \vec{s}_{i-1} \end{cases}$$

con \mathbf{A} , \mathbf{B} , \mathbf{C} e \mathbf{D} delle matrici.

Ci sono altre rappresentazioni dei sistemi discreti lineari.

Definiamo la risposta impulsiva w_k del sistema come la risposta alla successione

$$(3.3) \quad \delta_i = \begin{cases} 0 & \text{per } i \neq 0 \\ 1 & \text{per } i = 0 \end{cases}$$

che è detta **funzione impulsiva discreta**, o **impulso unitario** o anche **delta discreta** ed ha nei sistemi discreti un ruolo analogo alla delta di Dirac per i sistemi continui.

Analogamente a quanto accade per i sistemi continui, per i sistemi lineari discreti (tempo-invarianti) la risposta a un ingresso generico x_i è data dalla **convoluzione discreta** della x_i per la risposta impulsiva w_k :

$$(3.4) \quad y_i = \sum_{k=-\infty}^{\infty} x_{i-k} \cdot w_k$$

Ovviamente nella (3.4), se la x_i è data dalla (3.3), allora $y_i = w_i$.

Se il sistema è causale, la (3.4) diventa

$$(3.5) \quad y_i = \sum_{k=0}^{\infty} x_{i-k} \cdot w_k$$

L'operazione di convoluzione discreta (3.4) viene spesso indicata in modo abbreviato, similmente alla convoluzione continua, con

$$(3.6) \quad y = x \otimes w$$

Se $w_k = 0$ per tutti i $k < 0$, il sistema è causale.

La (3.4) è un altro modo di rappresentare un sistema lineare, in genere più pratico di quello dell'equazione (3.2). In esso non compare lo stato, ma gli ingressi a tutti i tempi precedenti. Il vettore di stato può essere visto come la condizione in cui è stato pilotato il sistema dagli input precedenti.

3.2 Casi particolari

Ci sono poi casi particolari di sistemi lineari discreti, spesso molto utili:

- sistemi **moving average (MA)**, cioè “a media mobile”, detti anche FIR (Finite Impulse Response), rappresentati dall’equazione

$$(3.7) \quad y_i = \sum_{k=0}^m b_k \cdot x_{i-k}$$

cioè l’uscita y_i è ottenuta dalla convoluzione delle successioni x (infinita) e b (finita). Come si vede immediatamente, questi sistemi sono caratterizzati da una risposta impulsiva composta da soli $m+1$ campioni. Hanno quindi una memoria finita: l’uscita è indipendente completamente da ciò che gli è stato dato in ingresso prima degli ultimi $m+1$ campioni.

- sistemi **auto-regressivi (AR)**, detti anche IIR (Infinite Impulse Response), rappresentati dall’equazione

$$(3.8) \quad y_i = b_0 \cdot x_i - \sum_{k=1}^n a_k \cdot y_{i-k}$$

in cui l’uscita a un dato istante viene espressa in funzione dell’ingresso a quell’istante e da una combinazione delle uscite negli n istanti precedenti. Si dimostra che la risposta impulsiva di questo tipo di sistemi è di lunghezza infinita. Facciamo un esempio: consideriamo il caso in cui $n=1$ e $a_1 = -w$. La (3.8) diventa

$$(3.9) \quad y_i = b_0 \cdot x_i + w \cdot y_{i-1}$$

e, ponendo come ingresso la successione (3.3), abbiamo

$$(3.10) \quad \begin{aligned} y_i &= 0 \quad \text{per } i < 0 \\ y_0 &= b_0 \\ y_1 &= b_0 \cdot w \\ y_2 &= b_0 \cdot w^2 \\ y_3 &= b_0 \cdot w^3 \\ &\dots \end{aligned}$$

e quindi, in generale,

$$(3.11) \quad y_i = b_0 \cdot w^i$$

per qualsiasi $i > 0$.

- sistemi **ARMA** (auto-regressive moving average), rappresentati dall’equazione

$$(3.12) \quad y_i = \sum_{k=0}^m b_k \cdot x_{i-k} - \sum_{k=1}^n a_k \cdot y_{i-k}$$

che comprende una combinazione lineare degli $m+1$ ingressi precedenti e le n uscite precedenti. Generalizza i due casi precedenti.

3.2.1 Semplice applicazione di sistemi MA e AR

Supponiamo di avere una successione x_i di misure di una grandezza fisica s costante, “disturbate” da un “rumore” (cioè da errori casuali),

$$(3.13) \quad x_i = s + n_i$$

Supponiamo che gli errori n_i siano tra loro indipendenti e di voler stimare il valor medio di s . Come è noto, un buon metodo per fare questa stima è prendere la media di un certo numero di valori delle misure x_i . Supponiamo che gli errori (o il “rumore”) n_i abbiano valor medio nullo e deviazione standard σ ; possiamo definire σ “incertezza”¹⁷ su s da una singola misura. Se facciamo la media su N misure, diminuiamo l’incertezza su s di \sqrt{N} volte.

Supponiamo ora che la grandezza s vari lentamente nel tempo. Come possiamo stimarla? Un modo potrebbe essere prendere la media di un numero N di misure successive abbastanza piccolo che, durante le N , s sia variato in modo trascurabile¹⁸. Ciò significa che possiamo fare ad ogni istante i una stima $\hat{\mu}_i$ del valor medio di s . Abbiamo

$$(3.14) \quad \hat{\mu}_i = \frac{1}{N} \sum_{k=0}^{N-1} x_{i-k}$$

Ciò è realizzato tramite un sistema MA con coefficienti tutti eguali $b_k = \frac{1}{N}$, per $0 \leq k \leq N-1$.

Un problema di questo modo di procedere è che appare piuttosto innaturale considerare tutti e soli gli ultimi N campioni, tutti con lo stesso peso. Si possono fare scelte dei pesi che decrescono con k , in modo che i valori più lontani, e che quindi possono avere un s più diverso rispetto al valore all’istante i , pesino di meno.

Una scelta particolarmente “ragionevole” è “comoda” è usare un sistema AR del tipo di eq.

(3.9), con $w = e^{-\frac{1}{N}}$. Ciò equivale a fare una media mobile con infiniti termini (in linea di principio, in pratica si considerano tutti i termini precedenti disponibili), con pesi decrescenti

¹⁷ Potremo dare differenti definizioni dell’incertezza, per esempio 2σ , o 3σ .

¹⁸ Vedremo in seguito come le varie espressioni qualitative presenti in questo paragrafo possano essere rese quantitative.

in modo esponenziale con la distanza. Si noti che questa soluzione è anche molto “leggera” computazionalmente, perché ad ogni passo richiede solo una moltiplicazione e un prodotto¹⁹.

Abbiamo qui considerato solo sistemi “causali” che permettono di dare la stima $\hat{\mu}_i$ non appena è disponibile il campione x_i . In molti casi non c’è questa restrizione, e quindi per $\hat{\mu}_i$ possono utilizzarsi anche i campioni con indice maggiore di i .

In particolare il filtro MA visto precedentemente diventa, ovviamente con un diverso N ,

$$(3.15) \quad \hat{\mu}_i = \frac{1}{2N+1} \sum_{k=-N}^N x_{i-k}$$

Per quanto riguarda il filtro AR, possiamo applicarlo due volte, una volta (“in avanti”) normalmente e l’altra (“all’indietro”) invertendo l’ordine dei campioni; in tal modo il filtraggio avviene considerando per ciascun campione tutti i campioni, sia precedenti che seguenti.

¹⁹ Anche un filtro MA con tutti i pesi uguali può realizzarsi con un basso costo computazionale, ma con un maggiore “costo” di memoria.

3.3 Trasformata z

Definiamo trasformata z di un segnale discreto $\{x_i\}$

$$(3.16) \quad X(z) = \sum_{i=-\infty}^{\infty} x_i \cdot z^{-i}$$

dove $z = r \cdot e^{j\Omega}$ è una variabile complessa definita per tutti i valori per cui la (3.16) converge. Essa viene anche indicata con $X(z) = \mathcal{Z} \{x_i\}$ o, quando è chiaro che parliamo di trasformate z, con $x(t) \leftrightarrow X(z)$.

Con la trasformata z associamo a un segnale discreto un polinomiale.

Si noti che, se la serie $\{x_i\}$ è di lunghezza finita, la regione di convergenza copre tutto il piano della variabile z (piano di Gauss). Se è *infinita destra*, cioè nulla per i minore di un certo valore, la regione di convergenza è l'esterno di un cerchio di centro l'origine (il raggio del cerchio dipende dalla serie). Analogamente, se è infinita sinistra, cioè nulla per i maggiore di un certo valore, la regione di convergenza è l'interno di un cerchio di centro l'origine.

Facciamo un esempio di calcolo analitico della trasformata z. Supponiamo di avere una serie

$$(3.17) \quad \{x_i\} = \begin{cases} 0 & \text{per } i < 0 \\ w^i & \text{per } i \geq 0 \end{cases}$$

Allora, dalla (3.16),

$$(3.18) \quad X(z) = \sum_{i=0}^{\infty} w^i \cdot z^{-i} = \frac{1}{1 - w \cdot z^{-1}}$$

Se la trasformata X(z) converge nel dominio anulare definito da $R_1 < |z| < R_2$ allora si può calcolare la trasformata inversa con

$$(3.19) \quad x_k = \frac{1}{2\pi j} \oint_C X(z) \cdot z^{k-1} \cdot dz$$

dove C è un percorso chiuso che separa $|z| = R_1$ da $|z| = R_2$. Questo integrale può calcolarsi con l'ausilio del teorema dei residui di Cauchy:

$$(3.20) \quad \frac{1}{2\pi j} \oint_C X(z) \cdot z^{k-1} \cdot dz = \sum [\text{residui di } X(z) \cdot z^{k-1} \text{ nei poli interni a } C]$$

Se la X(z) è una funzione razionale, se cioè è data dal rapporto di due polinomi, possiamo ottenere il rapporto dei due polinomi come un polinomiale della variabile z^{-1} , l'antitrasformata si ottiene immediatamente.

In molti casi la trasformata inversa può calcolarsi usando le tavole e le proprietà della trasformata z.

3.3.1 Analogia con la trasformata di Laplace

Un segnale discreto $\{x_i\}$ può essere associato ad un segnale continuo, costituito da una successione di delte di Dirac,

$$(3.21) \quad x(t) = \sum_k x_k \cdot \delta(t - k \cdot \Delta t)$$

essendo Δt il tempo di campionamento. La trasformata di Laplace del segnale $x(t)$ è

$$(3.22) \quad X_L(s) = \sum_k x_k \cdot e^{-s \cdot k \cdot \Delta t}$$

e la trasformata di Fourier è

$$(3.23) \quad X_F(\omega) = \sum_k x_k \cdot e^{-j\omega \cdot k \cdot \Delta t}$$

Sostituendo $z = e^{s \cdot \Delta t}$, si ha

$$(3.24) \quad X_L(s) = X(z) \Big|_{z=e^{s\Delta t}}$$

e, se sul cerchio di raggio 1 c'è convergenza,

$$(3.25) \quad X_F(\omega) = X(z) \Big|_{z=e^{j\omega\Delta t}}$$

Si noti che $X_L(s)$ e $X_F(\omega)$ sono periodiche (la prima sulla direzione immaginaria), con periodo

$$(3.26) \quad \omega_0 = \frac{2\pi}{\Delta t}$$

Il mapping del piano s sul piano z (non univoco, il che spiega la suddetta periodicità) è tale che l'asse immaginario diventa il cerchio unitario, il semi-piano sinistro diventa l'interno di tale cerchio e il semi-piano destro l'esterno.

3.3.2 Proprietà della trasformata z

Sia $X(z)$, $X^{(1)}(z)$ e $X^{(2)}(z)$ le trasformate di $\{x_i\}$, $\{x_i^{(1)}\}$ e $\{x_i^{(2)}\}$ rispettivamente.

- **Linearità:**

$$(3.27) \quad \{a \cdot x_i^{(1)} + b \cdot x_i^{(2)}\} \leftrightarrow \{a \cdot X^{(1)}(z) + b \cdot X^{(2)}(z)\}$$

- **Shift** (scorrimento, ovvero ritardo o avanzamento):

$$(3.28) \quad \{x_{i-k}\} \leftrightarrow z^{-k} \cdot X(z)$$

Notiamo che la variabile z agisce come un operatore di avanzamento, ovvero uno scorrimento a sinistra; z^{-1} come un operatore di ritardo, ovvero uno scorrimento a destra.

- **Inversione temporale :**

$$(3.29) \quad \{x_{-i}\} \leftrightarrow X(z^{-1})$$

- **Moltiplicazione per una sequenza esponenziale :**

$$(3.30) \quad \{\alpha^i \cdot x_i\} \leftrightarrow X\left(\frac{1}{\alpha}z\right)$$

Se α è reale, ciò corrisponde ad uno scaling nel dominio z , se è $\alpha = e^{j\varphi}$, ciò corrisponde ad una rotazione nel piano z .

- **Convoluzione :**

La trasformata z del segnale ottenuto come convoluzione di due segnali discreti (vedi equazione (3.4)) è data dal prodotto delle due trasformate z dei due segnali. Essendo x e y due successioni con trasformata z rispettivamente X e Y , si ha

$$(3.31) \quad x \otimes y \leftrightarrow X \cdot Y$$

La dimostrazione di questa proprietà, nel caso di segnali discreti di lunghezza finita, è abbastanza semplice: è legata all'analogia tra l'operazione di convoluzione discreta e il prodotto di due polinomi. Per rendersene conto, sviluppare il prodotto

$$(3.32) \quad (x_0 + x_1 \cdot a + x_2 \cdot a^2 + \dots + x_n \cdot a^n) \cdot (y_0 + y_1 \cdot a + y_2 \cdot a^2 + \dots + y_m \cdot a^m) =$$

$$(x_0 \cdot y_0) + (x_0 \cdot y_1 + x_1 \cdot y_0) \cdot a + (x_0 \cdot y_2 + x_1 \cdot y_1 + x_2 \cdot y_0) \cdot a^2 + \dots + (x_0 \cdot y_0) \cdot a^{m+n}$$

- **Coniugazione :**

$$(3.33) \quad \{x_i^*\} \leftrightarrow X^*(z^*)$$

Da cui si deduce che, se x è reale,

$$(3.34) \quad X(z) = X^*(z^*)$$

3.3.3 Alcune trasformate z

Vediamo ora le trasformate z di alcune semplici successioni:

- **delta discreta** (vedi (3.3)),

$$(3.35) \quad \delta_i = \begin{cases} 0 & \text{per } i \neq 0 \\ 1 & \text{per } i = 0 \end{cases} \leftrightarrow 1$$

- **funzione gradino**

$$(3.36) \quad u_i = \begin{cases} 0 & \text{per } i < 0 \\ 1 & \text{per } i \geq 0 \end{cases} \leftrightarrow \sum_{i=0}^{\infty} z^{-i} = \frac{1}{1-z^{-1}} = \frac{z}{z-1}$$

- **esponenziale**

$$(3.37) \quad u_i w^i \leftrightarrow \frac{1}{1-w \cdot z^{-1}} = \frac{z}{z-w}$$

- **esponenziale “rampato”**

$$(3.38) \quad u_i \cdot i \cdot w^i \leftrightarrow \frac{w \cdot z^{-1}}{(1-w \cdot z^{-1})^2} = \frac{wz}{(z-w)^2}$$

- **coseno**

$$(3.39) \quad u_i \cdot \cos(\omega_0 \cdot i) \leftrightarrow \frac{1 - \cos(\omega_0) \cdot z^{-1}}{1 - 2 \cos(\omega_0) \cdot z^{-1} + z^{-2}} = \frac{z^2 - \cos(\omega_0) \cdot z}{z^2 - 2 \cos(\omega_0) \cdot z + 1}$$

- **seno**

$$(3.40) \quad u_i \cdot \sin(\omega_0 \cdot i) \leftrightarrow \frac{\sin(\omega_0) \cdot z^{-1}}{1 - 2 \cos(\omega_0) \cdot z^{-1} + z^{-2}} = \frac{\sin(\omega_0) \cdot z}{z^2 - 2 \cos(\omega_0) \cdot z + 1}$$

3.4 Trasformata di Fourier discreta ed FFT

3.4.1 La trasformata di Fourier per dati discreti (DTFT)

Come per i segnali continui abbiamo fatto derivare la trasformata di Fourier dalla trasformata di Laplace, così per i segnali discreti possiamo derivare dalla trasformata z la **trasformata di Fourier per dati discreti (DTFT, discrete time Fourier transform)**. Dalla (3.16), imponendo $|z|=1$, cioè $z = e^{j\Omega}$, dove

$$(3.41) \quad \Omega = \omega \cdot \Delta t$$

è la pulsazione normalizzata (essendo Δt è il tempo di campionamento), a cui corrisponde, dividendola per 2π , una frequenza normalizzata a-dimensionale che ha valore massimo 1, abbiamo

$$(3.42) \quad \mathcal{X}(\Omega) = \sum_{i=-\infty}^{\infty} x_i \cdot e^{-j \cdot i \cdot \Omega}$$

La trasformata inversa si ottiene come

$$(3.43) \quad x_i = \frac{1}{2\pi} \int_0^{2\pi} \mathcal{X}(\Omega) \cdot e^{j \cdot i \cdot \Omega} \cdot d\Omega$$

Ovviamente la (3.42) ha senso solo se la sommatoria converge e condizione sufficiente perché ciò avvenga è che $\sum |x_i| < \infty$.

Notiamo che la (3.42) è equivalente alla trasformata di Fourier per segnali continui per il segnale continuo

$$(3.44) \quad x(t) = \sum x_i \cdot \delta(t - i \cdot \Delta t)$$

Questa trasformata, oltre alle analoghe proprietà della trasformata di Fourier normale, gode delle seguenti:

- **periodicità:** la $\mathcal{X}(\Omega)$ è periodica di periodo 2π , come è ovvio
- **prima differenza**

$$(3.45) \quad x_i - x_{i-1} \leftrightarrow (1 - e^{-j\Omega}) \mathcal{X}(\Omega)$$

- **accumulazione** (analogo nel discreto dell'integrazione)

$$(3.46) \quad \sum_{k=-\infty}^i x_k \leftrightarrow \pi \cdot \mathcal{X}(0) \delta(\Omega) + \frac{1}{1 - e^{-j\Omega}} \mathcal{X}(\Omega)$$

La DTFT è analoga alla serie di Fourier (vedi (2.12)), solo che a domini invertiti: è discreta nel tempo e continua e periodica nelle frequenze, come la serie di Fourier è continua e periodica nel tempo e discreta nelle frequenze.

3.4.2 La DFT (e la FFT)

Introduciamo ora la **trasformata discreta di Fourier**, un argomento di grande interesse pratico, poiché è ciò che viene in effetti normalmente calcolato col computer. Per far ciò poniamo che la $\{x_i\}$ non sia di lunghezza infinita, ma lunga N (ovvero che sia nulla per $i < 0$ e per $i > N-1$).

Definiamo trasformata discreta di Fourier (**DFT**) la seguente successione:

$$(3.47) \quad X_k = \sum_{i=0}^{N-1} x_i \cdot W_N^{k \cdot i}$$

dove

$$(3.48) \quad W_N = e^{-j \frac{2\pi}{N}}$$

è la radice N -esima dell'unità. Possiamo calcolare la DFT inversa come

$$(3.49) \quad x_i = \frac{1}{N} \sum_0^{N-1} X_k \cdot W_N^{-k \cdot i}$$

Come si vede, abbiamo una trasformata tra due domini discreti e finiti: infatti sia il tempo che la frequenza (più precisamente la pulsazione) sono discreti e finiti. Quindi si capisce l'utilità di questa trasformata, ovviamente parente stretto della trasformata di Fourier per i dati discreti. Infatti

$$(3.50) \quad X_k = \mathcal{X}\left(\frac{k \cdot 2\pi}{N}\right)$$

Il calcolo della DFT, per grandi valori di N (che può essere per esempio dell'ordine di 10^6), richiede un grandissima potenza di calcolo. Infatti, trascurando il calcolo delle potenze di W_N , che possono essere calcolati a priori, dobbiamo eseguire circa N^2 moltiplicazioni ed altrettante addizioni (per $N=10^6$ abbiamo 10^{12} operazioni di ciascun tipo). È stato però introdotto (da Cooley e Tukey nel 1965) un algoritmo di calcolo, la **Fast Fourier Transform** o **FFT**, che riduce il numero di operazioni a un valore proporzionale a $N \cdot \log N$ (in una delle migliori implementazioni è $5 \cdot N \cdot \log_2 N$, per $N=10^6$ abbiamo circa 10^8 , cioè 10000 volte più veloce).

3.5 Equazioni alle differenze

L'equazione (3.12) può essere riscritta come

$$(3.51) \quad \sum_{k=0}^n a_k \cdot y_{i-k} = \sum_{k=0}^m b_k \cdot x_{i-k}$$

dove $a_0 = 1$. Sottolineiamo che questa equazione descrive un sistema causale²⁰.

Si noti la somiglianza con la (2.8): in quella l'operatore differenziale sostituisce l'operatore ritardo presente di questa. E infatti nei sistemi discreti l'operatore "avanzamento" ha una funzione analoga all'operatore differenziale dei sistemi continui: le equazioni alle differenze finite, come la (3.51), prendono il posto delle equazioni differenziali.

Si noti che i due termini della (3.51) possono essere visti come due convoluzioni, della sequenza finita $\{a_i\}$ e della sequenza infinita $\{y_i\}$ e della sequenza finita $\{b_i\}$ e della sequenza infinita $\{x_i\}$. Definendo le trasformate z

$$(3.52) \quad \begin{aligned} X(z) &= \sum_{i=-\infty}^{\infty} x_i \cdot z^{-i} \\ Y(z) &= \sum_{i=-\infty}^{\infty} y_i \cdot z^{-i} \\ A(z) &= \sum_{i=0}^n a_i \cdot z^{-i} \\ B(z) &= \sum_{i=0}^m b_i \cdot z^{-i} \end{aligned}$$

possiamo utilizzare la proprietà della trasformata z della convoluzione ed otteniamo l'equazione

$$(3.53) \quad A(z)Y(z) = B(z)X(z)$$

Un'equazione come la (3.51) è detta **equazione lineare alle differenze**.

²⁰ Questa trattazione è valida per sistemi causali, ma può facilmente adattarsi a sistemi non causali, con la limitazione che a) l'uscita dipenda da ingressi posteriori, ma non da uscite posteriori, b) ci sia un limite al massimo "futuro". Il caso di sistemi anti-causali è ovviamente banale.

3.6 Funzione di trasferimento discreta

Riscriviamo l'equazione (3.53) come

$$(3.54) \quad Y(z) = \frac{B(z)}{A(z)} \cdot X(z)$$

e quindi l'uscita y del sistema (o meglio la sua trasformata z) può ottenersi dall'ingresso x , z -trasformato e moltiplicato per la funzione

$$(3.55) \quad F(z) = \frac{B(z)}{A(z)}$$

o anche, esplicitando,

$$(3.56) \quad F(z) = \frac{\sum_{i=0}^m b_i \cdot z^{-i}}{1 + \sum_{i=1}^n a_i \cdot z^{-i}}$$

La funzione $F(z)$ è detta **funzione di trasferimento discreta** e opera analogamente alla funzione di trasferimento definita per i sistemi continui.

Possiamo sviluppare la (3.56) in serie di potenze di z , ottenendo

$$(3.57) \quad F(z) = \sum_{i=0}^{\infty} f_i \cdot z^{-i}$$

(se $n=0$, la serie è finita), dove le $\{f_i\}$ sono la risposta impulsiva.

È chiaro che la (3.54) risolve in modo formalmente semplice il problema del calcolo della risposta forzata dell'equazione (3.51) (ma in pratica potrebbe risultare piuttosto complesso, dato il problema del calcolo della trasformata dell'ingresso e soprattutto dell'anti-trasformata dell'uscita).

La funzione di trasferimento è il modo più usato di rappresentare un sistema lineare discreto tempo-invariante.

Una delle proprietà più importanti della funzione di trasferimento z (simile a quella delle funzioni di trasferimento con la trasformata di Laplace) è che la funzione di trasferimento di un sistema composto da due sistemi in cascata F_1 e F_2 , ha come funzione di trasferimento il prodotto delle due:

$$(3.58) \quad F = F_1 \cdot F_2$$

Possiamo anche definire il sistema inverso di uno dato, tale che se messo in cascata con esso il sistema risultante ha funzione di trasferimento 1, semplicemente come

$$(3.59) \quad F_I(z) = \frac{1}{F(z)}$$

tuttavia, perché sia “utilizzabile”, dobbiamo verificare che $F_I(z)$ sia stabile.

Poiché la trasformata z della delta discreta è 1, la $F(z)$ è anche la trasformata z della risposta impulsiva del sistema.

Ricordando che sul cerchio unitario $z = e^{j\Omega}$, possiamo facilmente calcolare la risposta in frequenza del sistema (la risposta a una sinusoide campionata) come

$$(3.60) \quad \mathcal{F}(\Omega) = F(e^{j\Omega}) = \frac{\sum_{k=0}^m b_k \cdot e^{-j\Omega \cdot k}}{1 + \sum_{k=1}^n a_k \cdot e^{-j\Omega \cdot k}}$$

La pulsazione Ω è definita per $0 \leq \Omega \leq 2\pi$ e, se i coefficienti a e b sono reali,

$$(3.61) \quad F(e^{j\Omega}) = F^*(e^{-j\Omega})$$

e quindi il valore massimo “effettivo” di Ω è π . I valori di Ω tra π e 2π possono vedersi come frequenze angolari negative (ricordiamo che l’“asse Ω ” è in effetti una circonferenza) e per sistemi a coefficienti reali la simmetria (3.61) è una simmetria tra frequenze positive e negative. Per i sistemi a coefficienti non reali, tale simmetria non sussiste.

Si noti che, nel caso dei coefficienti reali, per $\Omega = 0$ e $\Omega = \pi$ \mathcal{F} è reale, quindi lo sfasamento è nullo.

Usando la pulsazione (o frequenza angolare) “fisica” ω ,

$$(3.62) \quad \mathcal{F}(\omega \cdot \Delta t) = F(e^{j\omega \cdot \Delta t})$$

$\mathcal{F}(\Omega)$ è una funzione complessa della pulsazione normalizzata Ω : per ciascun valore di Ω abbiamo un numero complesso il cui modulo indica il guadagno del sistema e la fase lo sfasamento. Possiamo quindi rappresentare la $\mathcal{F}(\omega \cdot \Delta t)$ con i classici diagrammi di Bode.

Analogamente a come si è fatto per il continuo, la funzione di trasferimento è completamente descritta avendo dato

- gli zeri del polinomio al numeratore (chiamati semplicemente “zeri” z_k)

- gli zeri del polinomio al denominatore (chiamati "poli" p_k)
- il valore di b_0

3.6.1 Differenze e derivate

Data una successione $\{x_i\}$, chiamiamo **differenza prima** di $\{x_i\}$ la successione

$$(3.63) \quad y_i = x_i - x_{i-1}$$

La (3.63) rappresenta un sistema MA del primo ordine con funzione di trasferimento

$$(3.64) \quad F(z) = 1 - z^{-1}$$

Possiamo costruire la **differenza seconda** di $\{x_i\}$, facendo la differenza prima di $\{y_i\}$; il "sistema" che fa la differenza seconda equivale a due sistemi che fanno la differenza prima posti in serie, quindi la sua funzione di trasferimento è il quadrato della (3.64)

$$(3.65) \quad F(z) = (1 - z^{-1})^2 = 1 - 2z^{-1} + z^{-2}$$

Con le potenze successive della (3.64) si possono costruire le differenze di ordine superiore.

Si noti l'analogia della differenza prima (nel discreto) con la derivata prima (nel continuo).

Si può invertire l'operazione di differenza prima, con un sistema con funzione di trasferimento inversa

$$(3.66) \quad F(z) = \frac{1}{1 - z^{-1}}$$

che equivale al sistema AR del primo ordine

$$(3.67) \quad y_i = x_i + y_{i-1}$$

che esegue un'operazione analoga a quella dell'integrale nel continuo

$$(3.68) \quad y(t) = \int_{-\infty}^t x(\vartheta) \cdot d\vartheta$$

Ma quanto l'operazione di differenza è vicina a quella di derivazione? Questa questione potrà essere discussa quantitativamente alla luce di concetti che svilupperemo in seguito (lo spettro di potenza o di energia di un segnale), ma è chiaro che la differenza sarà una tanto migliore "stima" della derivata, quanto più lenta è la variazione del segnale (a parte un coefficiente

dovuto al tempo di campionamento, che noi normalmente, lavorando direttamente nel discreto, supponiamo pari all'unità di tempo).

3.7 Risposta di un sistema discreto

La soluzione generale di un'equazione lineare alle differenze è data dalla somma dell'*evoluzione libera*, cioè la soluzione ottenuta con ingresso nullo, e dalla *risposta forzata*, cioè dalla soluzione ottenuta con le condizioni iniziali (gli n valori precedenti il primo valore di ingresso non nullo) tutte poste a 0.

Il caso più semplice di risposta forzata è la risposta impulsiva, che è la risposta forzata in cui la "forzante", cioè l'ingresso, è un segnale a delta unitario. A partire dalla risposta impulsiva si può ricavare la risposta forzata e a quest'ultima può ricondursi l'evoluzione libera.

3.7.1 Risposta impulsiva

La risposta impulsiva di un sistema descritto dalla (3.51) può essere ottenuta semplicemente in modo ricorsivo, ponendo $x_0 = 1$ e tutti i valori successivi a 0.

Un altro modo per ricavarla è eseguire la divisione tra i polinomi (nella variabile z^{-1}) a numeratore e a denominatore, ottenendo (in genere) una equivalenza tra la funzione di trasferimento e un polinomiale infinito, in cui il coefficiente della potenza i -esima è l' i -esimo valore della risposta impulsiva.

Per ottenere una soluzione in forma analitica, osserviamo che il sistema associato alla (3.51), descritto dalla funzione di trasferimento (3.56), nel caso in cui tutti gli n poli siano differenti, può essere visto come il parallelo di n sistemi, ciascuno dei quali ha un solo polo.

Quindi il k -esimo polo $w_{(k)}$ darà una componente additiva al risultato proporzionale a

$$(3.69) \quad \eta_i^{(k)} = w_{(k)}^i$$

avremo quindi per la risposta impulsiva complessiva

$$(3.70) \quad \eta_i = \sum_{k=1}^n K_k \cdot \eta_i^{(k)}$$

dove le costanti K_k possono essere calcolate dal sistema formato dalle prime n

$$(3.71) \quad \sum_{k=1}^n K_k \cdot \eta_i^{(k)} = \tilde{\eta}_i$$

dove i $\tilde{\eta}_i$ sono i η_i calcolati in modo ricorsivo (e ovviamente $\tilde{\eta}_i = \eta_i$).

3.7.2 Risposta forzata

Se conosciamo la risposta impulsiva η_i , possiamo ottenere la risposta forzata, ad un ingresso $\{x_i\}$, come (vedi equazione (3.4))

$$(3.72) \quad y_i = \sum_{k=0}^{\infty} x_{i-k} \cdot \eta_k$$

3.7.3 Evoluzione libera

L'evoluzione libera si ha quando l'ingresso è nullo, ma non sono nulli tutti o alcuni dei valori di y_{i-k} per $1 \leq k \leq n$. Per l'evoluzione libera, possiamo considerare l'omogenea associata alla (3.51)

$$(3.73) \quad \sum_{k=0}^n a_k \cdot y_{i-k} = 0$$

o, ricordando $a_0 = 1$,

$$(3.74) \quad y_i = -\sum_{k=1}^n a_k \cdot y_{i-k}$$

A partire dalle condizioni iniziali

$$(3.75) \quad y_{-1}, y_{-2}, \dots, y_{-n}$$

possiamo calcolare ricorsivamente l'evoluzione libera come

$$(3.76) \quad \begin{aligned} y_0 &= -a_1 y_{-1} - a_2 y_{-2} - \dots - a_n y_{-n} \\ y_1 &= -a_1 y_0 - a_2 y_{-1} - \dots - a_n y_{-n+1} \\ y_2 &= -a_1 y_1 - a_2 y_0 - \dots - a_n y_{-n+2} \\ &\dots \end{aligned}$$

Possiamo tuttavia ricavare per essa un'espressione analitica. Per far ciò notiamo che l'evoluzione libera, nel caso delle condizioni (3.75), è equivalente alla risposta forzata della parte AR, in cui l'ingresso è dato da

$$(3.77) \quad \begin{aligned} x_0 &= -a_1 y_{-1} - a_2 y_{-2} - \dots - a_n y_{-n} \\ x_1 &= -a_2 y_{-1} - a_3 y_{-2} - \dots - a_n y_{-n+1} \\ x_2 &= -a_3 y_{-1} - a_2 y_0 - \dots - a_n y_{-n+2} \\ &\dots \\ x_{n-2} &= -a_{n-1} y_{-1} - a_n y_{-2} \\ x_{n-1} &= -a_n y_{-1} \end{aligned}$$

quindi basta calcolare questa risposta forzata.

Notiamo che l'evoluzione libera è indipendente dal numeratore della funzione di trasferimento (parte MA dell'equazione alle differenze).

3.8 Stabilità

Condizione necessaria e sufficiente per la stabilità è che il modulo dei poli della funzione di trasferimento sia minore di 1: i poli siano cioè interni al cerchio unitario.

Questa è la stabilità analitica. In pratica può accadere che un sistema analiticamente stabile non lo sia nella sua implementazione su calcolatore. Ciò può accadere, per sistemi complessi con poli molto vicini all'unità, a causa degli errori di arrotondamento.

Notiamo che i sistemi MA (o FIR), non avendo poli, sono intrinsecamente stabili.

3.9 Sistemi semplici

Analizziamo ora in dettaglio alcuni semplici sistemi MA e AR, cioè sistemi in cui nell'equazione (3.51), che per comodità riscriviamo,

$$(3.78) \quad \sum_{k=0}^n a_k \cdot y_{i-k} = \sum_{k=0}^m b_k \cdot x_{i-k}$$

n (detto **ordine AR**) e m (detto **ordine MA**) hanno valori piccoli.

Spesso sistemi più complessi possono essere ricondotti a più di questi sistemi in cascata..

3.9.1 Sistema di ordine 0

Il sistema più semplice è quello in cui nell'equazione (3.78), $n=m=0$ e abbiamo

$$(3.79) \quad y_i = b_0 \cdot x_i$$

Questo sistema non fa altro che amplificare di un fattore b_0 il segnale in ingresso. Questa amplificazione è la stessa per tutte le frequenze (la funzione di trasferimento è semplicemente b_0).

Attenzione ! Se b_0 è complesso e $b_0 = |b_0| \cdot e^{j\varphi}$, abbiamo un'amplificazione complessa che può essere vista come un'amplificazione reale $|b_0|$ più uno sfasamento (eguale per tutte le frequenze) di φ .

3.9.2 Sistema MA del primo ordine

Consideriamo il sistema in cui nell'equazione (3.78), $n=0$ e $m=1$; abbiamo

$$(3.80) \quad y_i = b_0 \cdot x_i + b_1 \cdot x_{i-1}$$

La funzione di trasferimento z è

$$(3.81) \quad F(z) = b_0 + b_1 \cdot z^{-1}$$

e la risposta in frequenza è

$$(3.82) \quad \mathcal{F}(\Omega) = F(e^{j\Omega}) = b_0 + b_1 \cdot e^{-j\Omega}$$

La risposta impulsiva, finita, è composta da due termini:

$$(3.83) \quad \begin{aligned} x_0 &= b_0 \\ x_1 &= b_1 \end{aligned}$$

Consideriamo dei casi particolari.

- $b_0 = 0, b_1 = 1$, ovvero un sistema che esegue semplicemente un ritardo di un campione. Analizziamo la risposta in frequenza:

$$(3.84) \quad \mathcal{F}(\Omega) = e^{-j\Omega}$$

Vediamo che abbiamo un'amplificazione unitaria per tutte le frequenze e uno sfasamento proporzionale a Ω (si dice anche **sfasamento lineare**). È questa una caratteristica di qualsiasi ritardo.

Un sistema come questo, in cui il guadagno non varia con la frequenza, ma in cui varia lo sfasamento è detto **sistema passa-tutto** (all-pass system). Questo tipo di sistemi possono essere molto utili. Dato un sistema $F(z)$ possiamo sempre costruire un sistema passa-tutto con la stessa caratteristica di fase di $F(z)$, come

$$(3.85) \quad G(z) = \frac{F(z)}{|F(z)|}$$

Tale sistema tuttavia in genere non sarà causale.

- $b_0 = -b_1 = 1$, ovvero un sistema che esegue la differenza prima, l'analogo discreto della derivata del continuo. Mostriamo i diagrammi di Bode²¹ per questo sistema:

²¹ Normalmente nei diagrammi di Bode il guadagno è espresso in decibel. Qui, per semplicità, lo lasciamo nelle sue unità naturali, in scala logaritmica. Ricordiamo che la misura in dB è $20 \cdot \log_{10} G$.

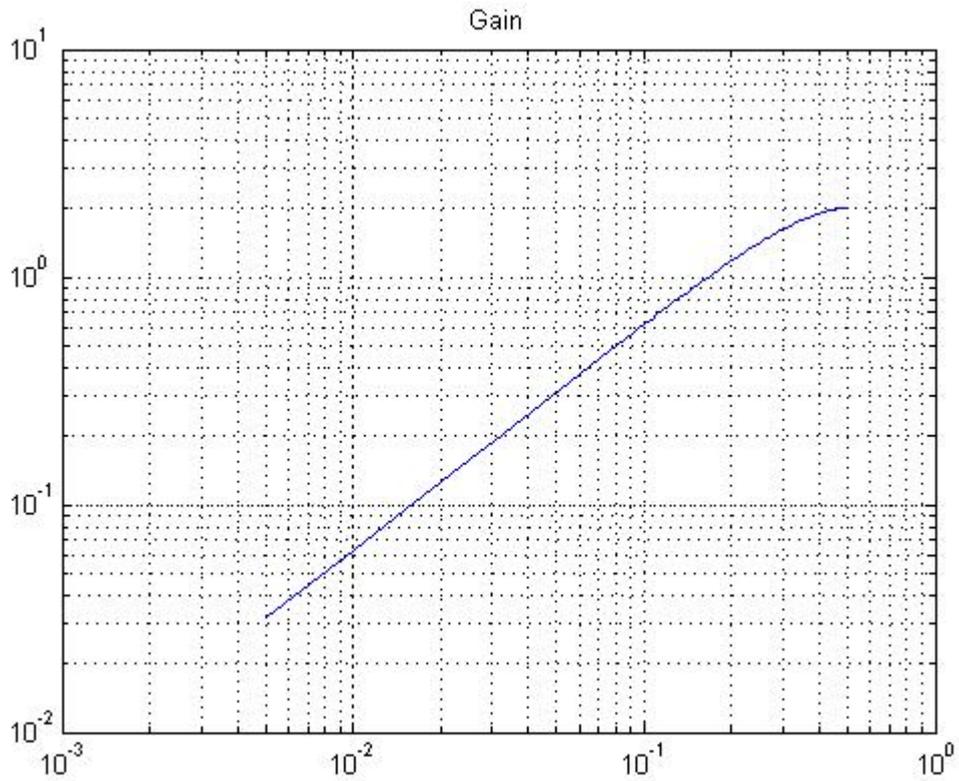


Figura 3-1

Si noti che la pendenza è di una decade di guadagno per una decade di frequenza.

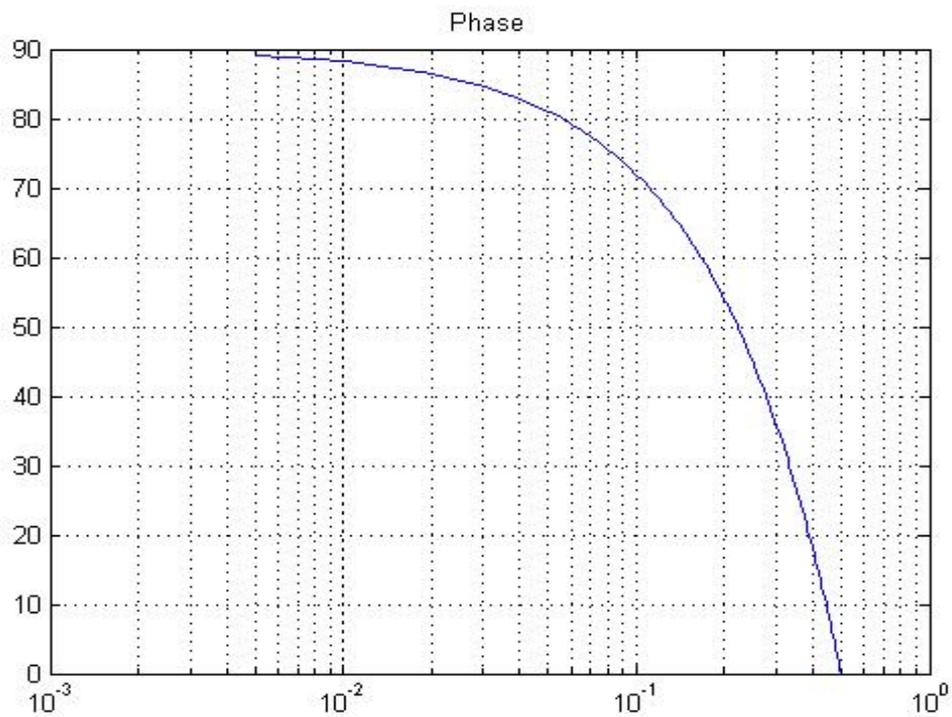
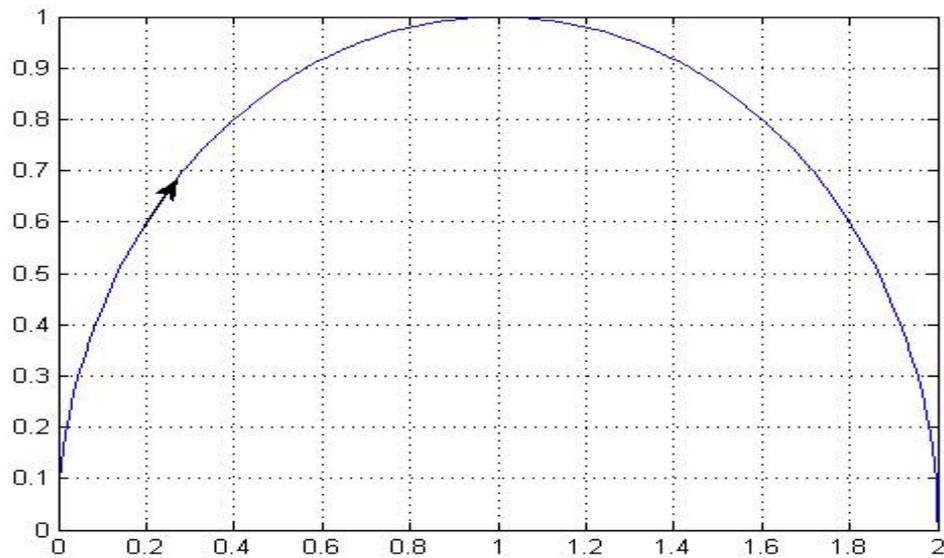


Figura 3-2

Si noti che a basse frequenze c'è uno sfasamento di 90 gradi positivo (in anticipo), mentre per alte frequenze lo sfasamento è nullo.

Possiamo anche graficare la funzione di trasferimento nel piano di Gauss (che ha per assi la parte reale e la parte immaginaria della variabile complessa), ottenendo il diagramma di Nyquist



(la freccia indica frequenza crescente).

Si noti che, come un derivatore, è un sistema passa alto.

- $b_0 = b_1 = \frac{1}{2}$, ovvero un sistema che fa la media degli ultimi due campioni. Mostriamo i grafici del guadagno e della fase per questo sistema:

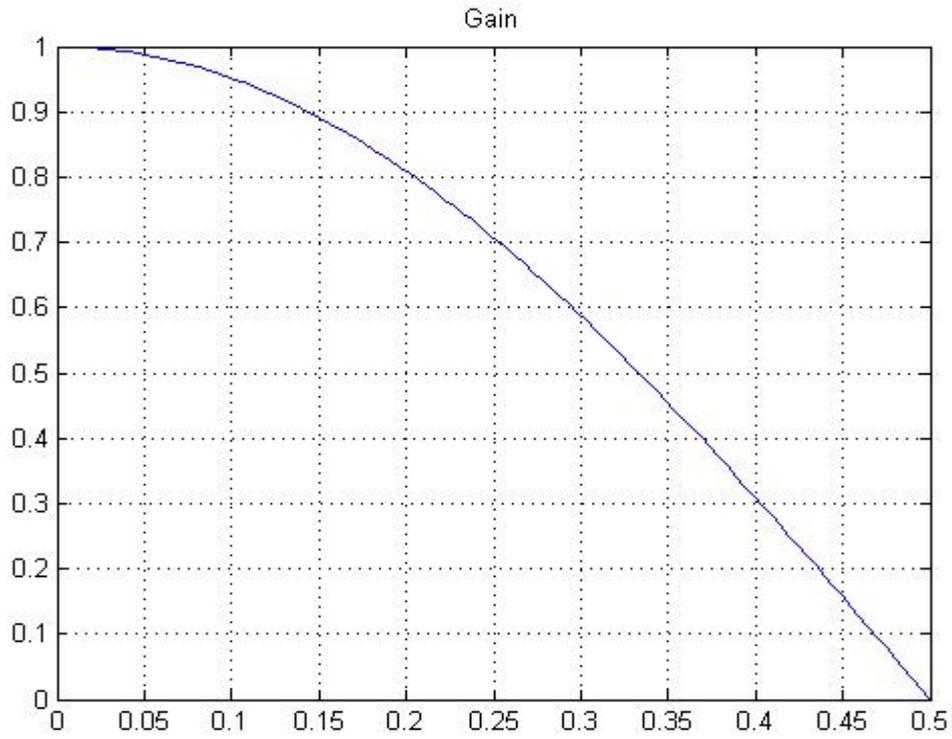


Figura 3-3

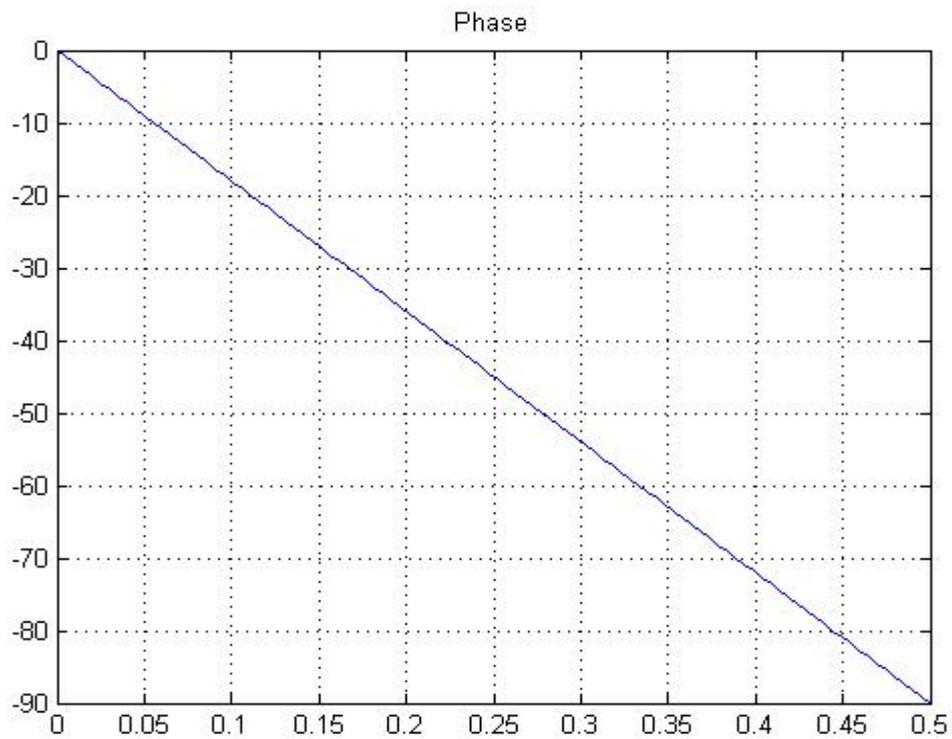


Figura 3-4

e il diagramma di Nyquist

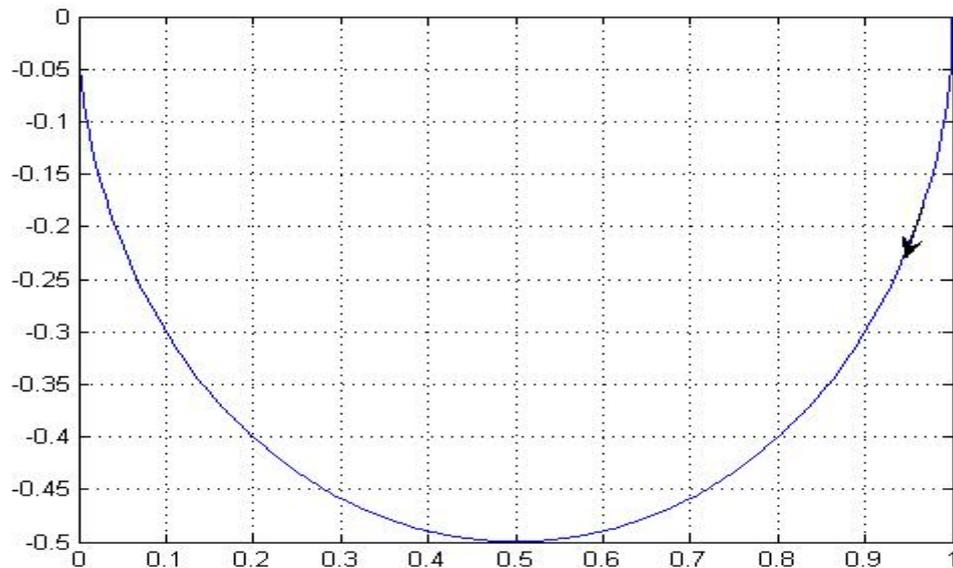


Figura 3-5

Si vede che si tratta di un sistema passa-basso.

Ci si può domandare cosa accade se la media la facciamo su un numero più elevato di campioni in ingresso, costruendo il sistema

$$(3.86) \quad y_i = \frac{1}{N} \cdot \sum_{k=0}^{N-1} x_{i-k}$$

Consideriamo per esempio un sistema MA di ordine 9, con tutti i 10 coefficienti eguali a 0.1. Ecco i relativi diagrammi:

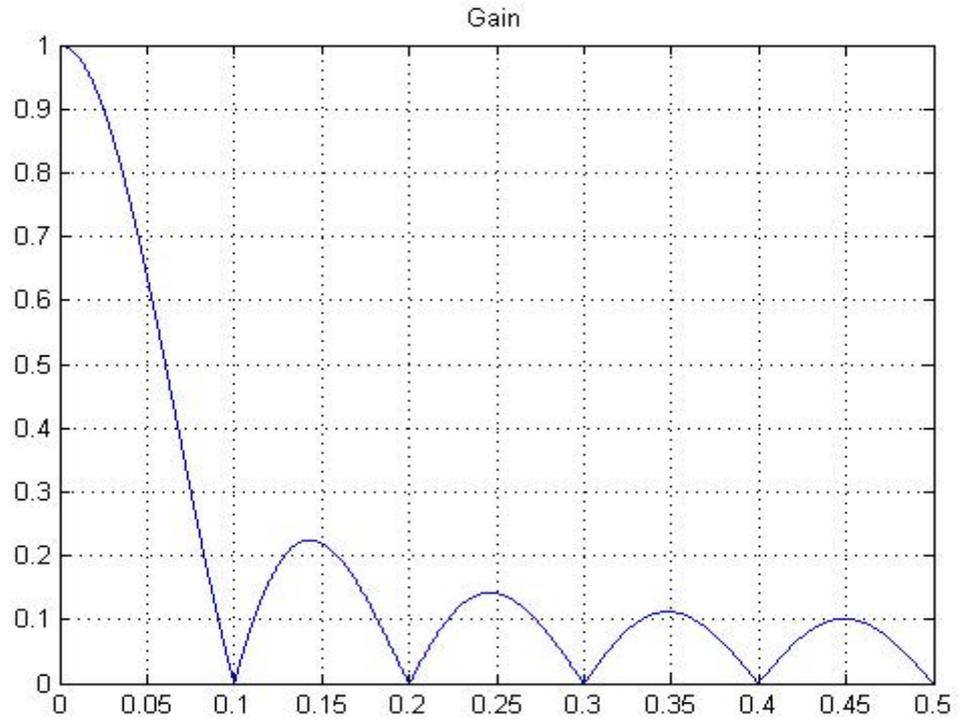


Figura 3-6

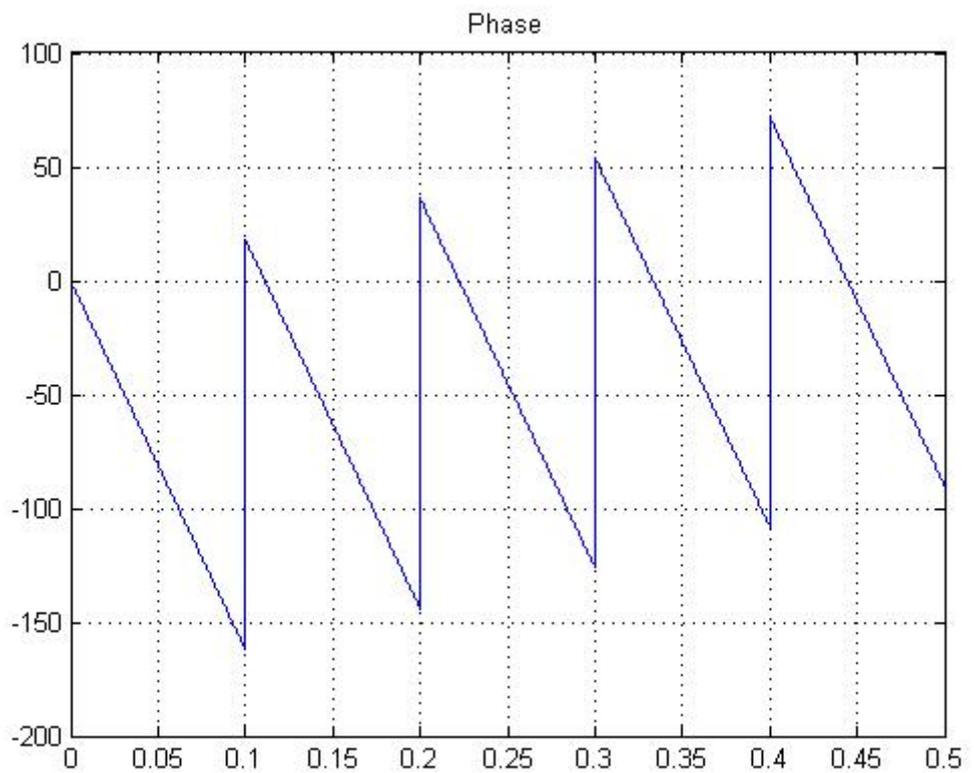


Figura 3-7

e Nyquist:

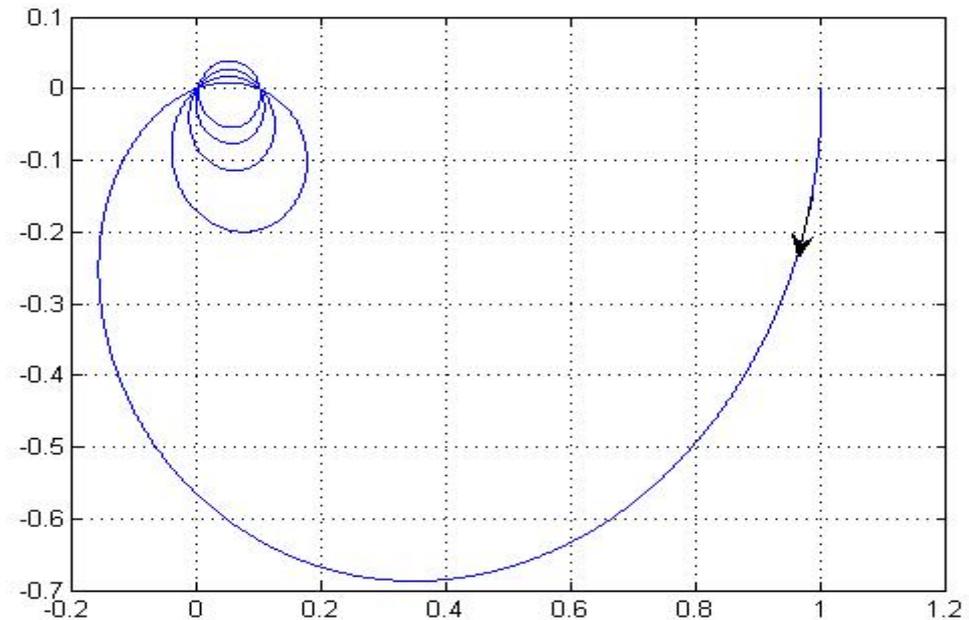


Figura 3-8

Si noti che è un più “pesante” passa basso, ma presenta delle “risonanze” che potrebbero essere fastidiose.

3.9.3 Due sistemi MA del primo ordine in cascata: MA del secondo ordine

Se poniamo due sistemi in cascata (o “in serie”), la funzione di trasferimento è data dal prodotto delle due funzioni di trasferimento. Se abbiamo due sistemi MA del primo ordine in cascata, con $B^{(1)}(z) = b_0^{(1)} + b_1^{(1)} \cdot z^{-1}$ e $B^{(2)}(z) = b_0^{(2)} + b_1^{(2)} \cdot z^{-1}$, abbiamo

$$(3.87) \quad B(z) = B^{(1)}(z)B^{(2)}(z) = b_0^{(1)}b_0^{(2)} + (b_0^{(1)}b_1^{(2)} + b_1^{(1)}b_0^{(2)})z^{-1} + b_1^{(1)}b_1^{(2)}z^{-2}$$

cioè un sistema MA del secondo ordine con coefficienti

$$(3.88) \quad \begin{aligned} b_0 &= b_0^{(1)}b_0^{(2)} \\ b_1 &= b_0^{(1)}b_1^{(2)} + b_1^{(1)}b_0^{(2)} \\ b_2 &= b_1^{(1)}b_1^{(2)} \end{aligned}$$

Se per esempio mettiamo in cascata due sistemi con $b_0 = -b_1 = 1$, cioè che eseguono la differenza prima, abbiamo un sistema con $b_0 = b_2 = 1$, $b_1 = -2$; questo è un sistema che esegue la differenza seconda (analoga alla derivata seconda del continuo). Abbiamo:

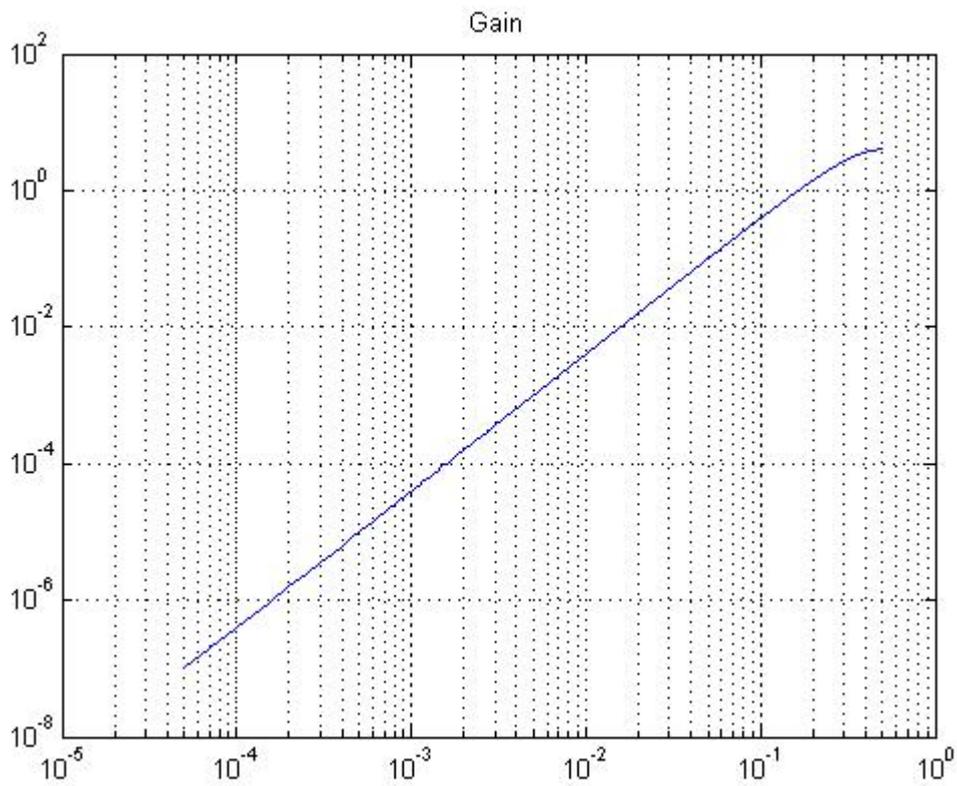


Figura 3-9

Si noti la pendenza di due decadi per ogni decade di frequenza, doppia che per la differenza prima.

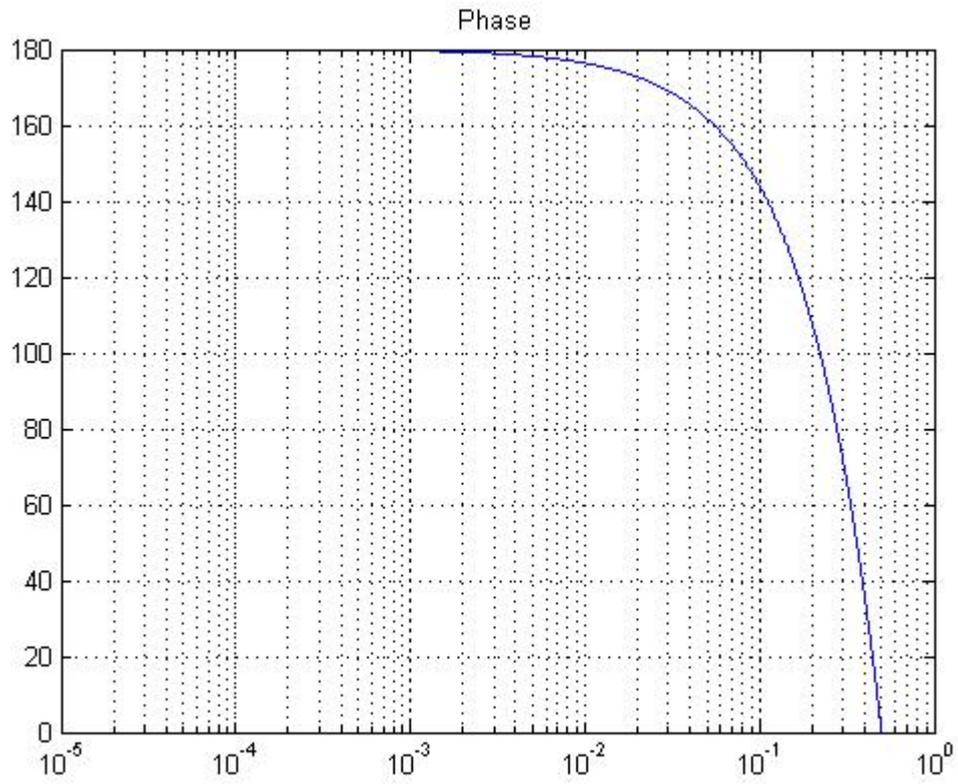


Figura 3-10

e Nyquist

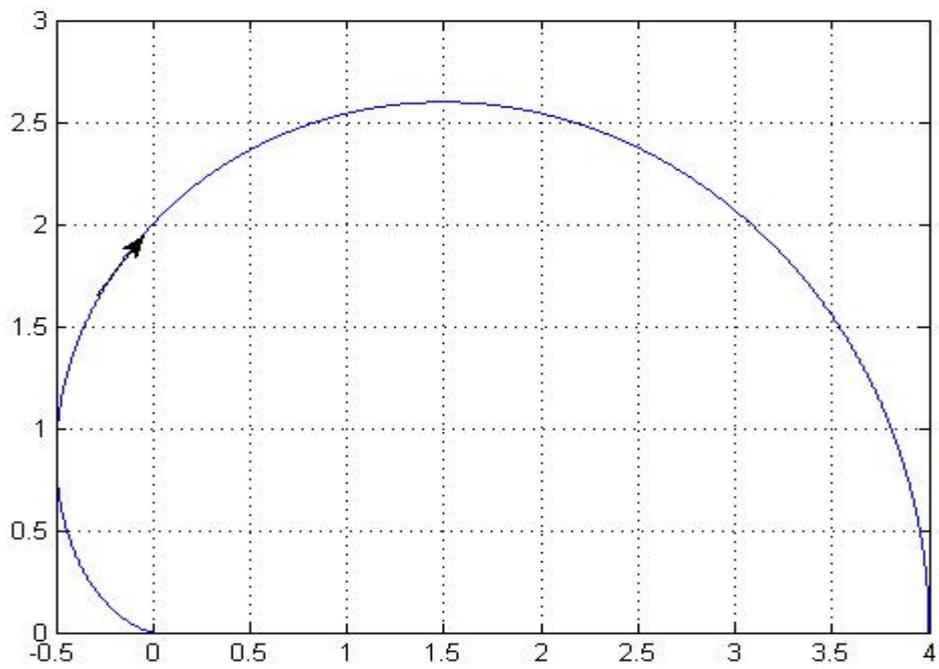


Figura 3-11

3.9.4 Sistema AR del primo ordine (reale)

Consideriamo il sistema in cui nell'equazione (3.78), $n=1$, $m=0$; abbiamo

$$(3.89) \quad y_i = b_0 \cdot x_i - a_1 \cdot y_{i-1}$$

In questo caso la risposta impulsiva è facilmente calcolabile come

$$(3.90) \quad \begin{aligned} x_0 &= b_0 \\ x_1 &= -b_0 \cdot a_1 \\ x_2 &= b_0 \cdot a_1^2 \\ x_3 &= -b_0 \cdot a_1^3 \\ &\dots \\ x_i &= b_0 \cdot (-a_1)^i \\ &\dots \end{aligned}$$

Si vede immediatamente che se $|a_1| > 1$ allora il sistema non è stabile (la risposta impulsiva ha un andamento esponenziale crescente).

Se $|a_1| \leq 1$ il sistema è stabile e, se $a_1 < 0$, la risposta impulsiva è esponenziale decrescente, altrimenti è in modulo esponenziale decrescente, ma con segni alternativamente positivi e negativi. L'andamento esponenziale ha decadimento τ dato da

$$(3.91) \quad \tau = -\frac{1}{\ln |a_1|}$$

che, per valori di $|a_1|$ molto vicini a 1, si può approssimare con

$$(3.92) \quad \tau \approx \frac{1}{1 - |a_1|}$$

Se $a_1 = -1$, la risposta impulsiva è un gradino di ampiezza b_0 .

Vediamo ora la risposta in frequenza. Si ha

$$(3.93) \quad \mathcal{F}(\Omega) = F(e^{j\Omega}) = \frac{b_0}{1 + a_1 \cdot e^{-j\Omega}}$$

Grafichiamo alcuni casi:

- $b_0 = 1$, $a_1 = -0.9$

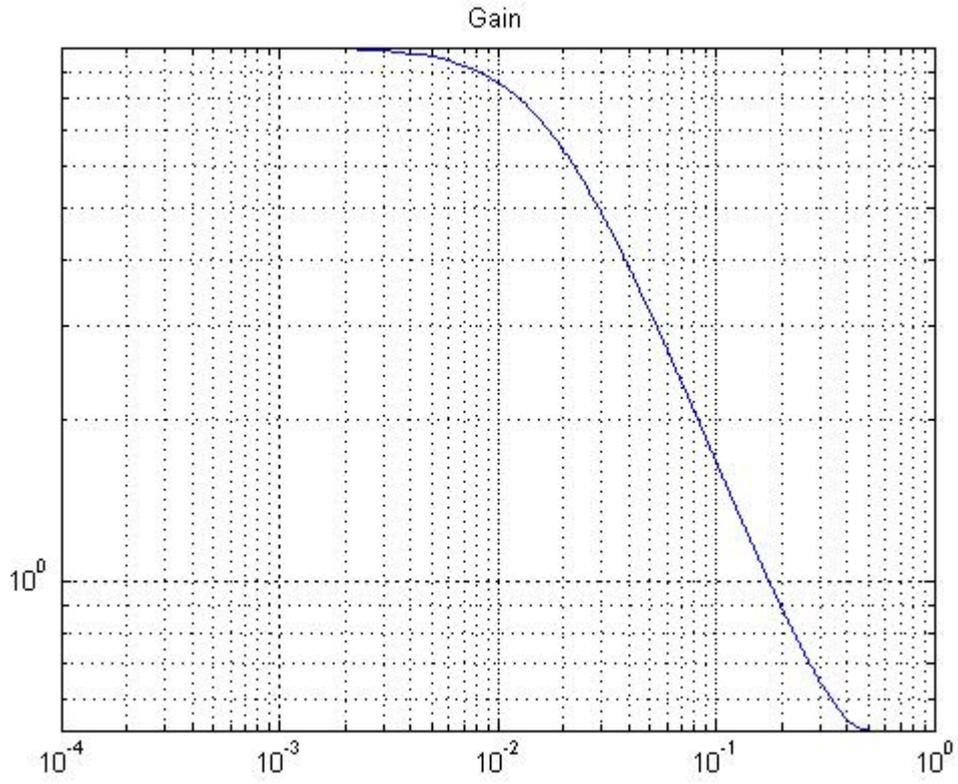


Figura 3-12

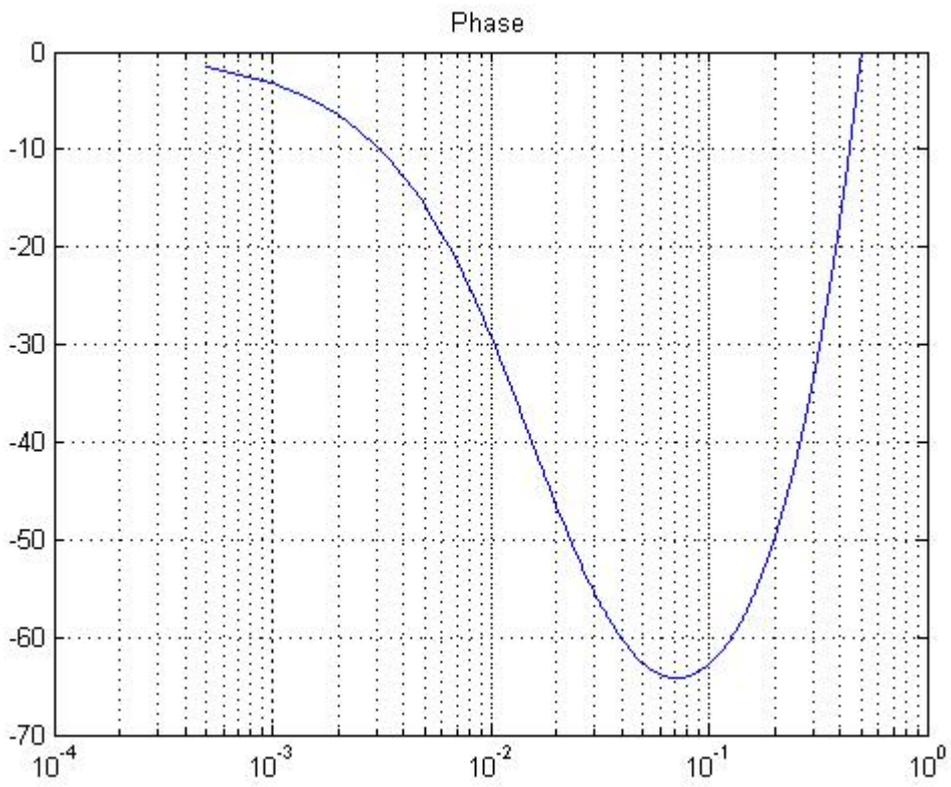


Figura 3-13

E il diagramma di Nyquist

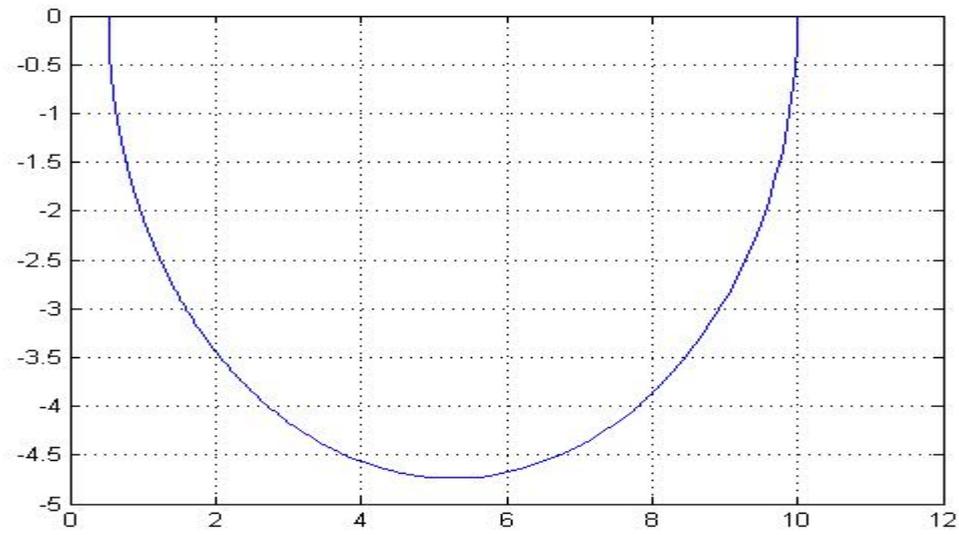


Figura 3-14

- $b_0 = 1, a_1 = -0.999$

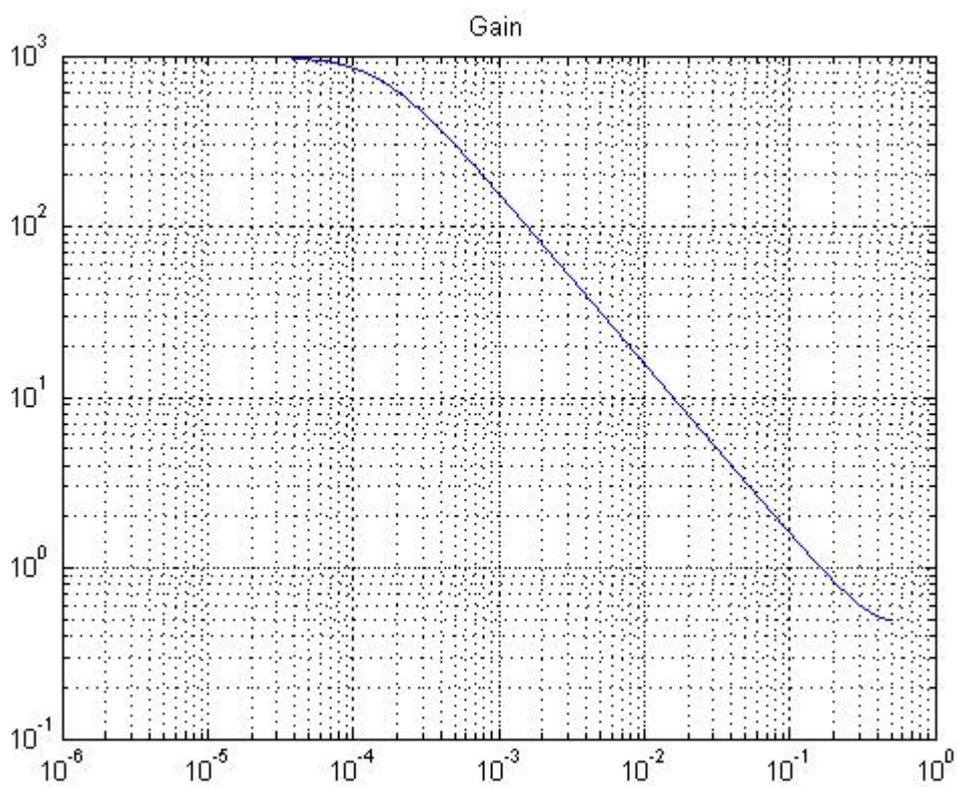


Figura 3-15

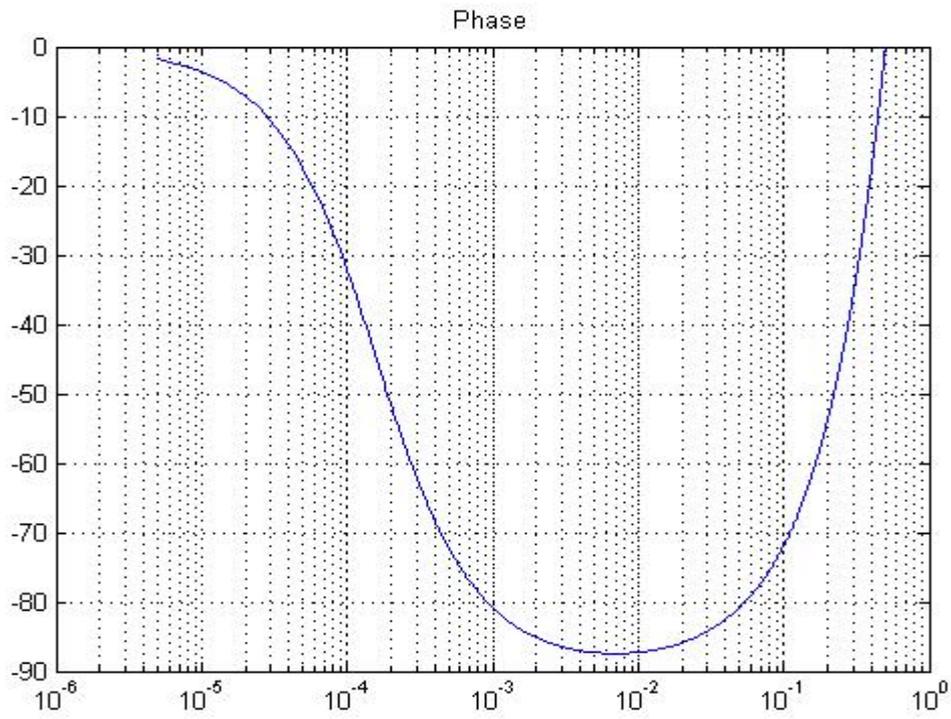


Figura 3-16

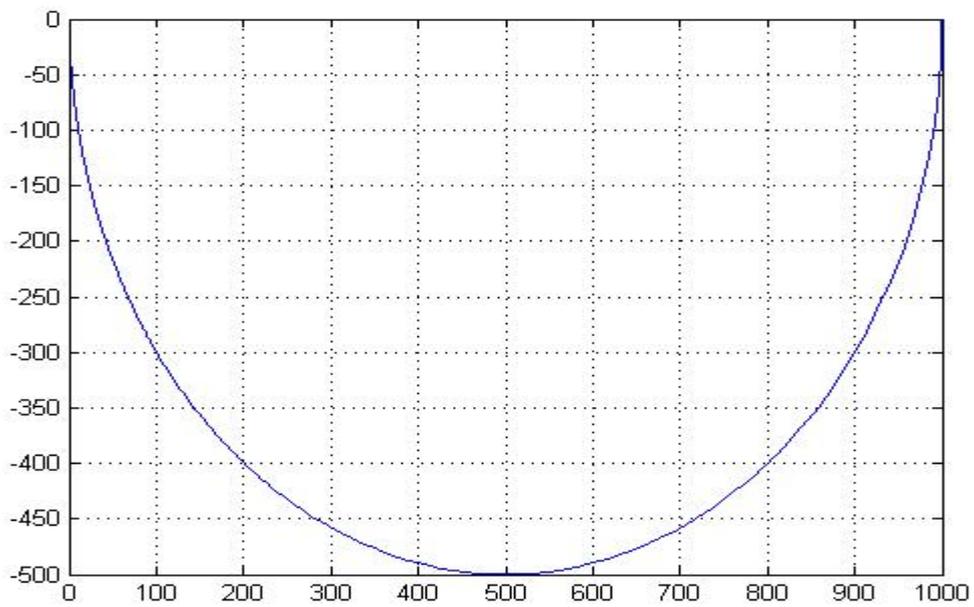


Figura 3-17

- $b_0 = 1$, $a_1 = 0.9$. In questo caso si ha

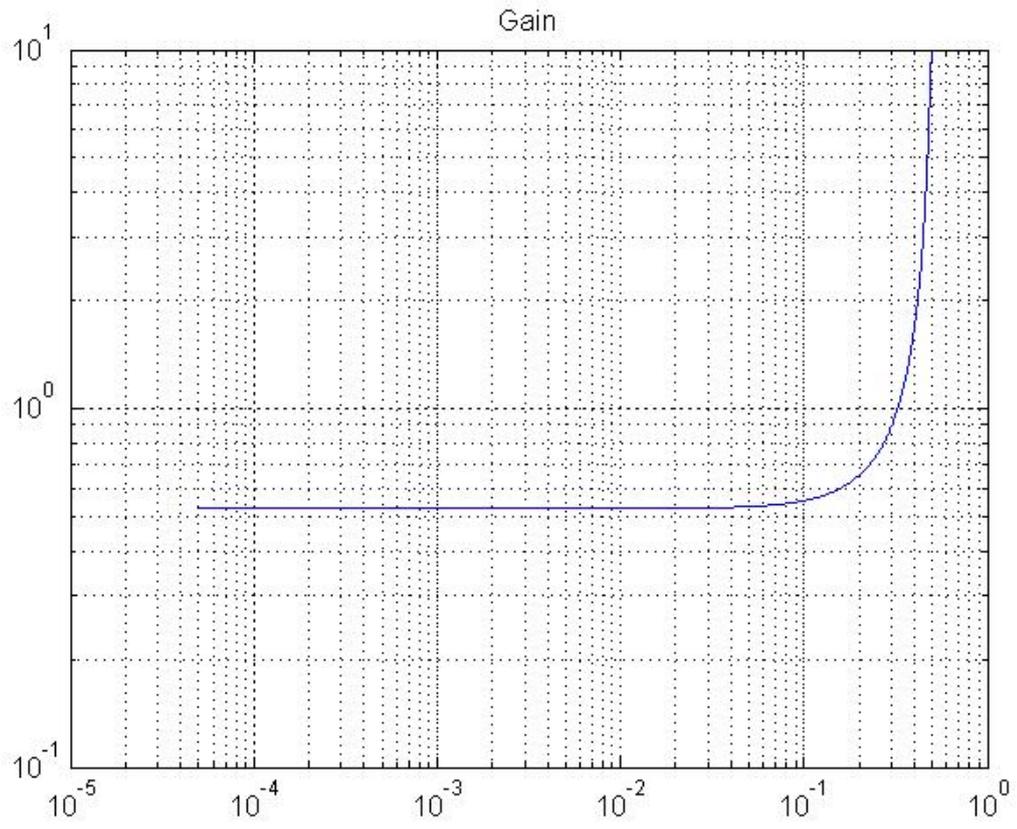


Figura 3-18

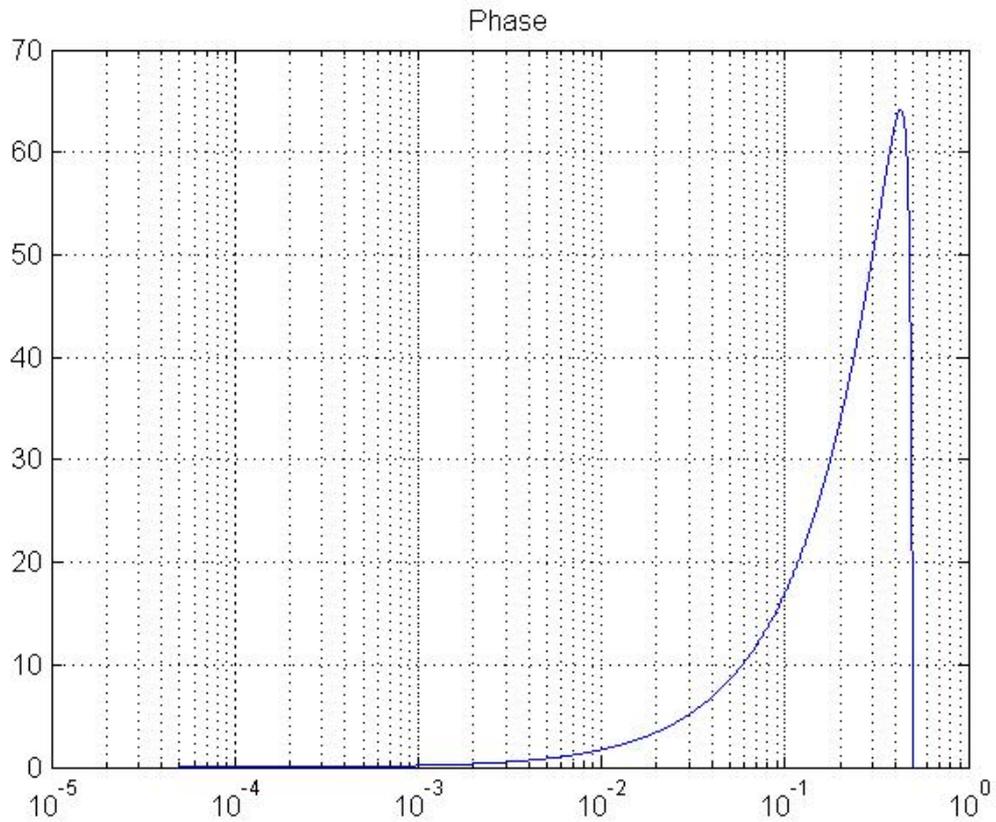


Figura 3-19

e Nyquist

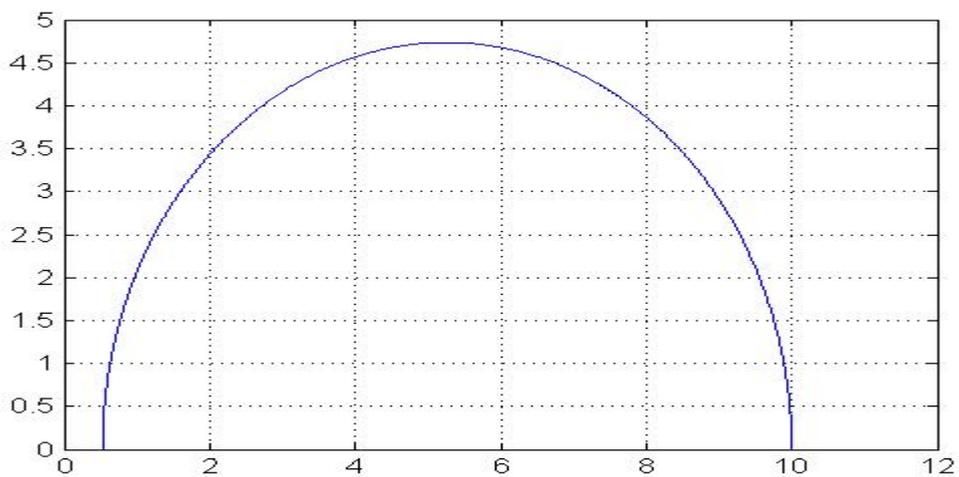


Figura 3-20

- Vediamo ora cosa accade se consideriamo l'inverso del sistema precedente, in cui semplicemente nella funzione di trasferimento z si è invertito il numeratore col denominatore. Il sistema era $b_0 = 1$, $a_1 = 0.9$; il suo inverso è un sistema MA che ha $b_0 = 1$, $b_1 = -0.9$. Ecco i grafici:

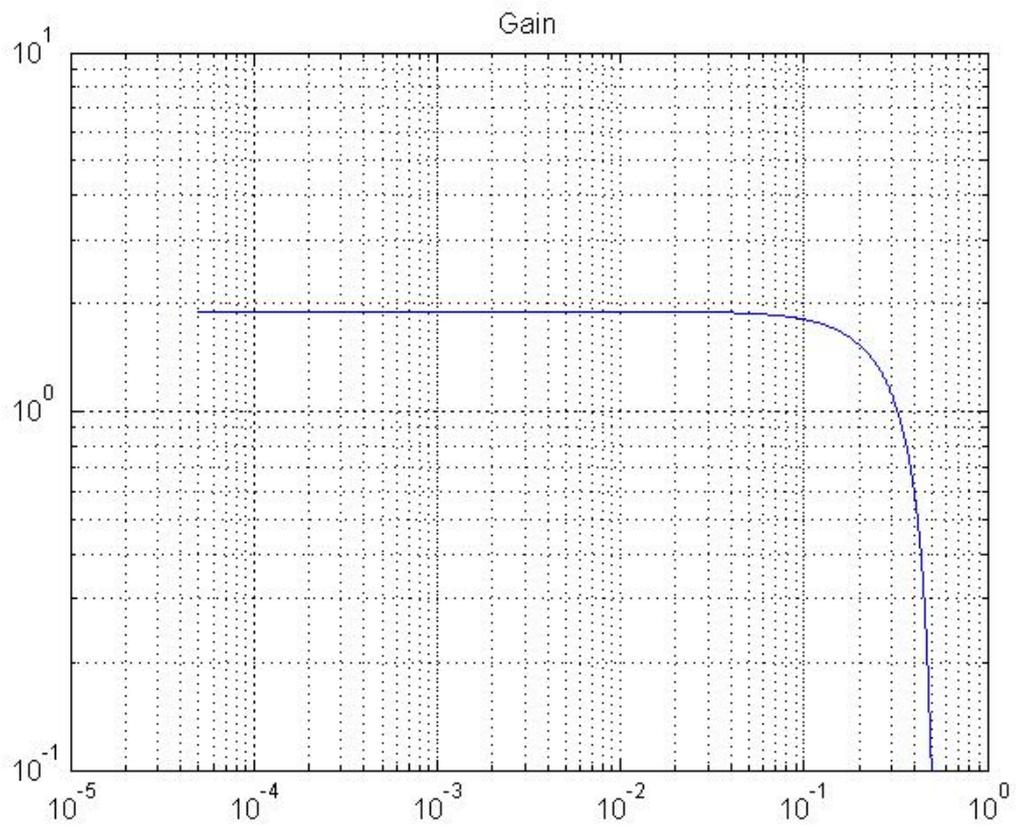


Figura 3-21

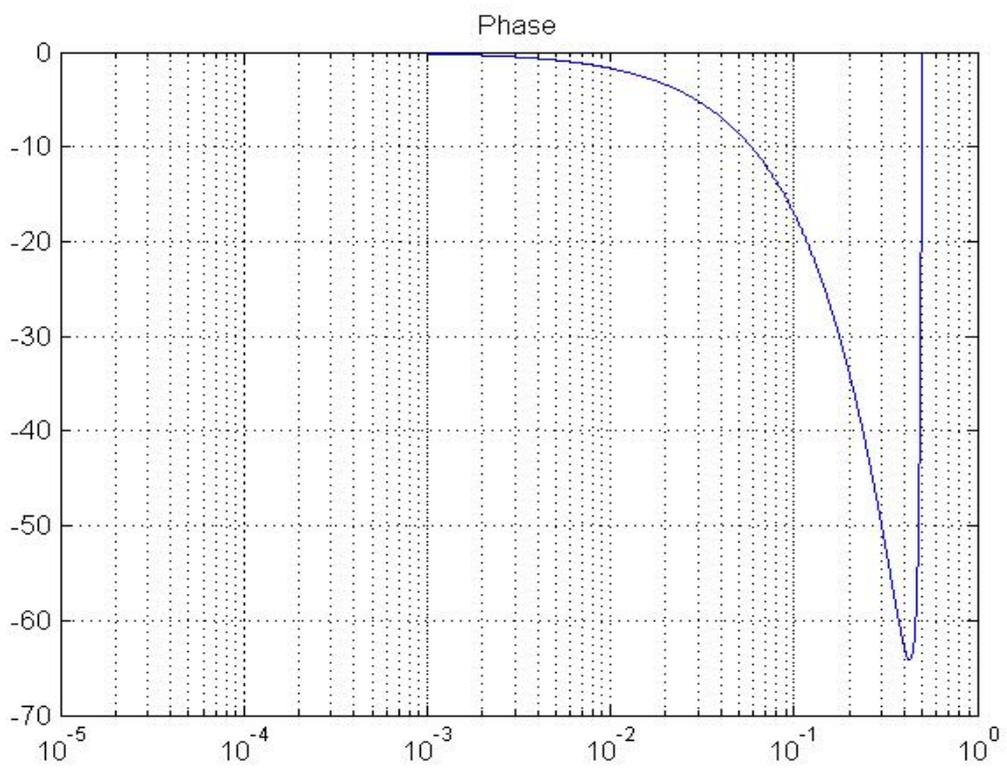


Figura 3-22

e Nyquist

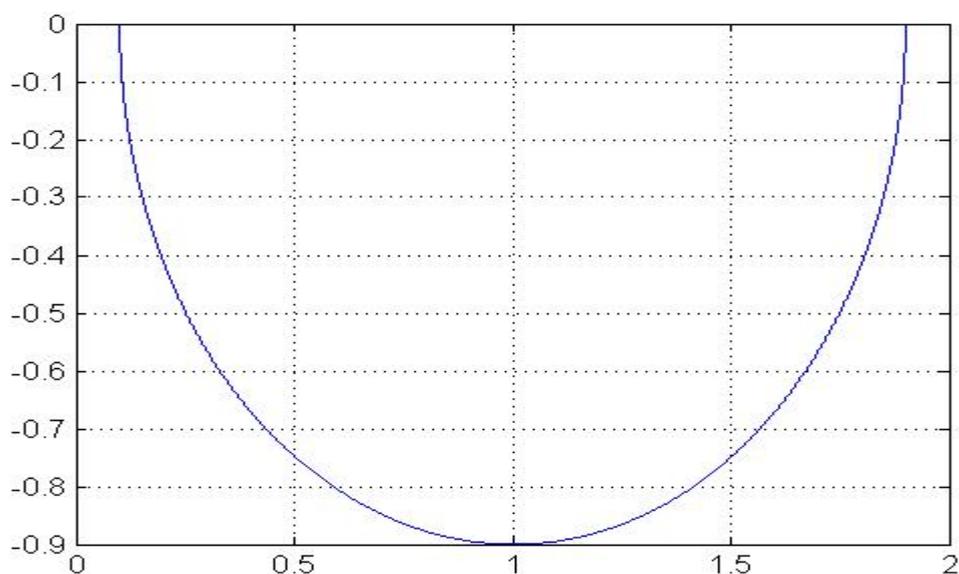


Figura 3-23

3.9.5 Sistema AR del primo ordine (complesso)

Se nell'equazione (3.89) il coefficiente a_1 è complesso, formalmente le cose non cambiano: la risposta impulsiva può essere espressa sempre come in equazione (3.90), ma ora, come vedremo, siamo in presenza di un'oscillazione smorzata.

Sia

$$(3.94) \quad w = -a_1 = r \cdot e^{j\varphi}$$

a parte il coefficiente di guadagno b_0 , la risposta impulsiva mostra tutte le potenze di w , da 0 a infinito; il k -esimo termine è dato da

$$(3.95) \quad x_k = b_0 \cdot w^k = b_0 \cdot (r^k \cdot e^{j \cdot k \cdot \varphi})$$

Notiamo che:

- il sistema è stabile se $r < 1$
- in tale ipotesi il valore assoluto decresce esponenzialmente
- la fase di w^k (e quindi di x_k) cresce proporzionalmente a k , quindi il vettore che rappresenta nel piano di Gauss il numero complesso x_k ruota intorno all'origine con velocità proporzionale a φ , riducendosi proporzionalmente di ampiezza. La frequenza di rotazione è data da

$$(3.96) \quad v_0 = \frac{\varphi}{2\pi}$$

e il tempo di decadimento τ è dato da (vedi equazione (3.91))

$$(3.97) \quad \tau = -\frac{1}{\ln |w|} = -\frac{1}{\ln r}$$

Consideriamo il caso di $b_0 = 1$, $r = 0.95$, $\varphi = 36^\circ$. Ecco l'evoluzione nel piano di Gauss

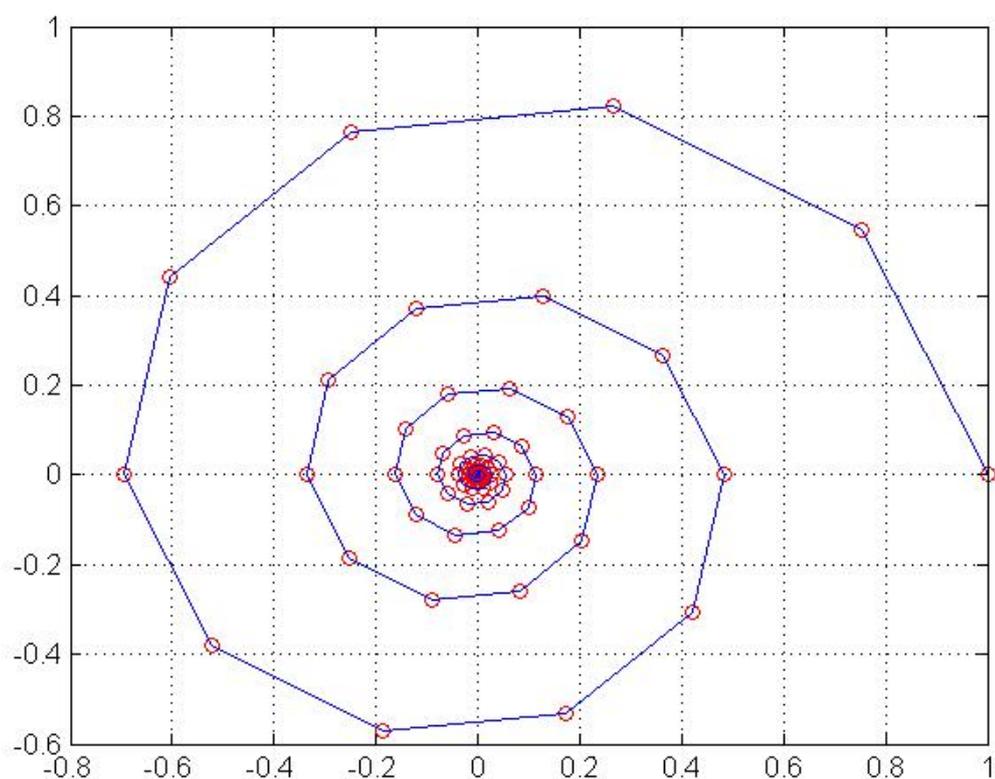


Figura 3-24

La risposta in frequenza è data da

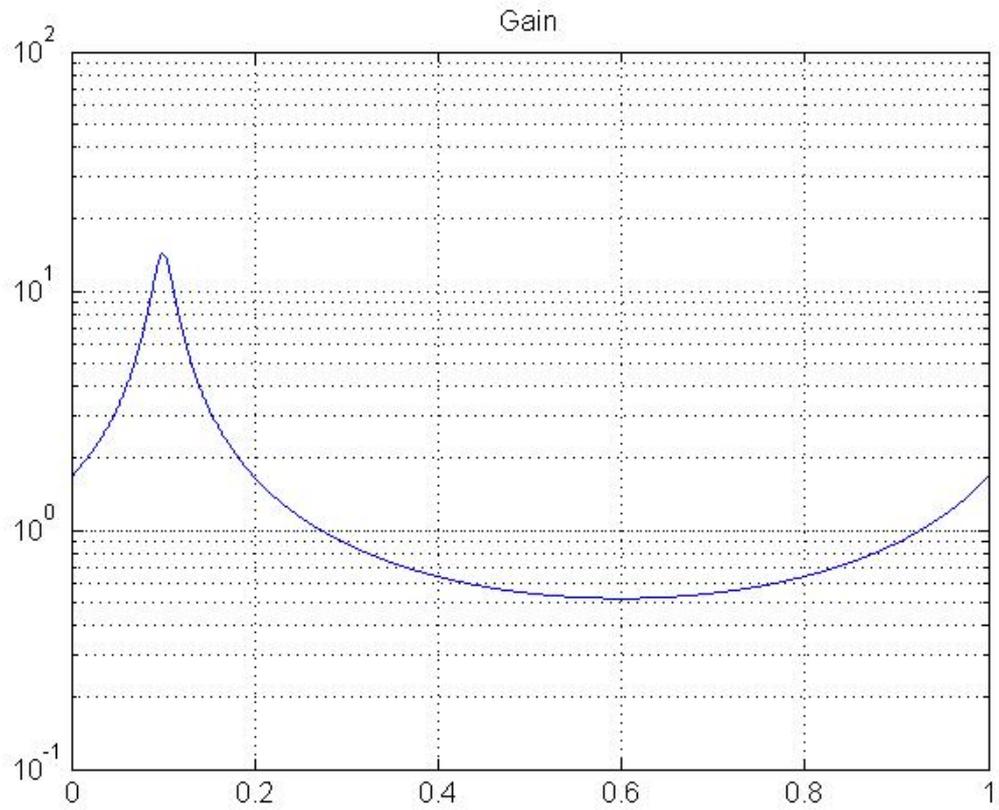


Figura 3-25

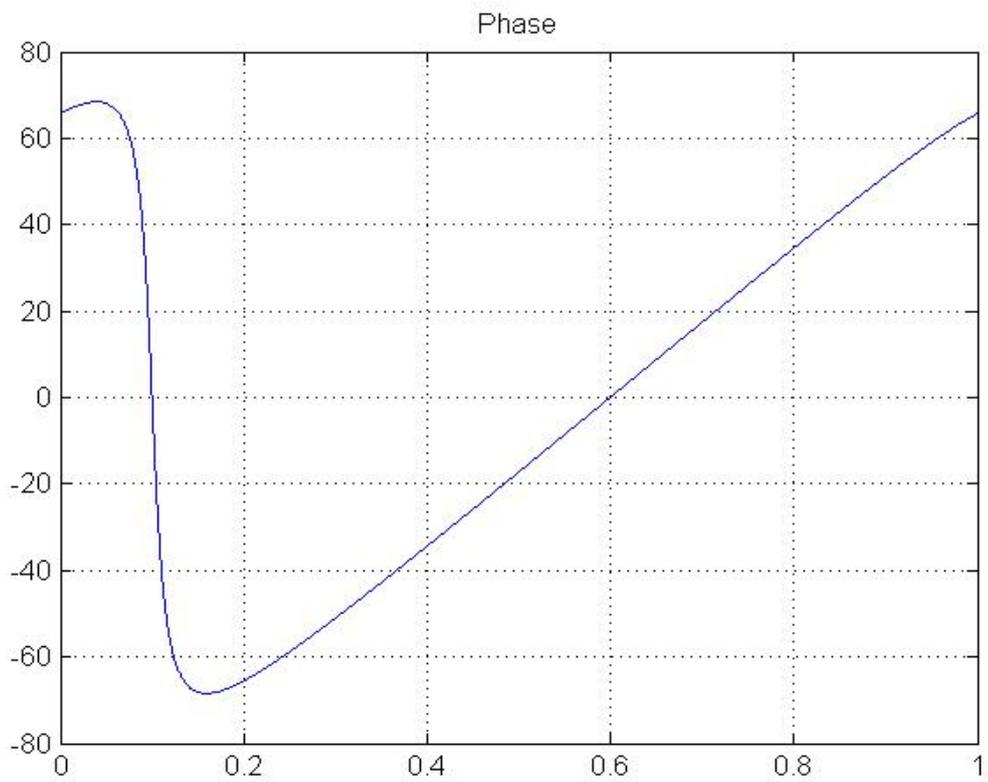


Figura 3-26

Si noti la presenza di una risonanza a frequenza 0.1 $\left(\frac{36}{360} = 0.1\right)$.

Il diagramma di Nyquist è

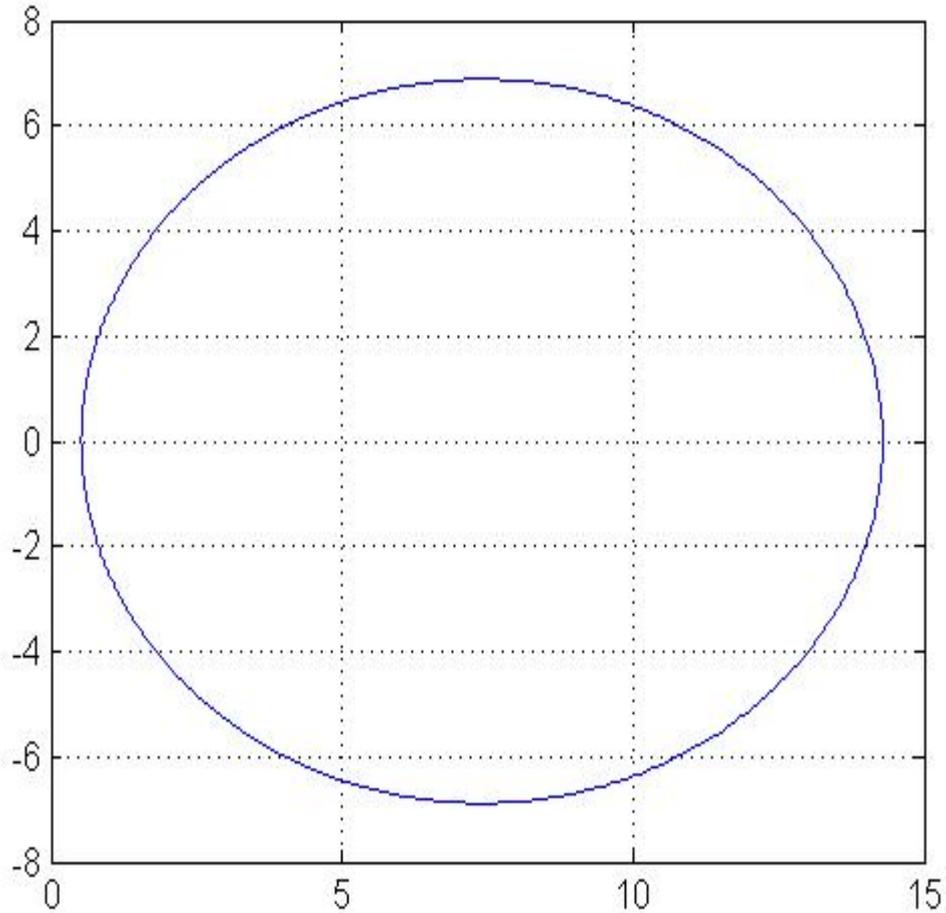


Figura 3-27

3.9.6 Sistema AR del secondo ordine

Consideriamo il sistema in cui nell'equazione (3.78), $n=2$, $m=0$; abbiamo

$$(3.98) \quad y_i = b_0 \cdot x_i - a_1 \cdot y_{i-1} - a_2 \cdot y_{i-2}$$

Il denominatore della funzione di trasferimento è

$$(3.99) \quad A(z) = 1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2} = (1 - p_1 \cdot z^{-1}) \cdot (1 - p_2 \cdot z^{-1})$$

dove p_1 e p_2 sono i due poli (soluzioni della $A(z) = 0$).

Se i due poli sono complessi coniugati, li possiamo porre come

$$(3.100) \quad p_{1,2} = r \cdot e^{\pm j\theta}$$

e quindi

$$(3.101) \quad \begin{aligned} A(z) &= 1 + a_1 \cdot z^{-1} + a_2 \cdot z^{-2} = 1 - (p_1 + p_2) \cdot z^{-1} + (p_1 \cdot p_2) \cdot z^{-2} = \\ &= 1 - 2 \cdot r \cdot \cos \theta \cdot z^{-1} + r^2 \cdot z^{-2} \end{aligned}$$

Se $r < 1$ il sistema è stabile e, se $a_1 < 0$, la risposta impulsiva è una sinusoide smorzata esponenzialmente

$$(3.102) \quad \eta_i = \frac{b_0}{2} \cdot (p_1^i + p_2^i) = b^0 \cdot r^i \cdot \cos(i \cdot \theta)$$

L'andamento esponenziale ha decadimento τ dato da

$$(3.103) \quad \tau = -\frac{1}{\ln r}$$

Grafichiamo il caso in cui $b_0 = 1$, $r = 0.99$, $\theta = \frac{\pi}{6}$. Abbiamo

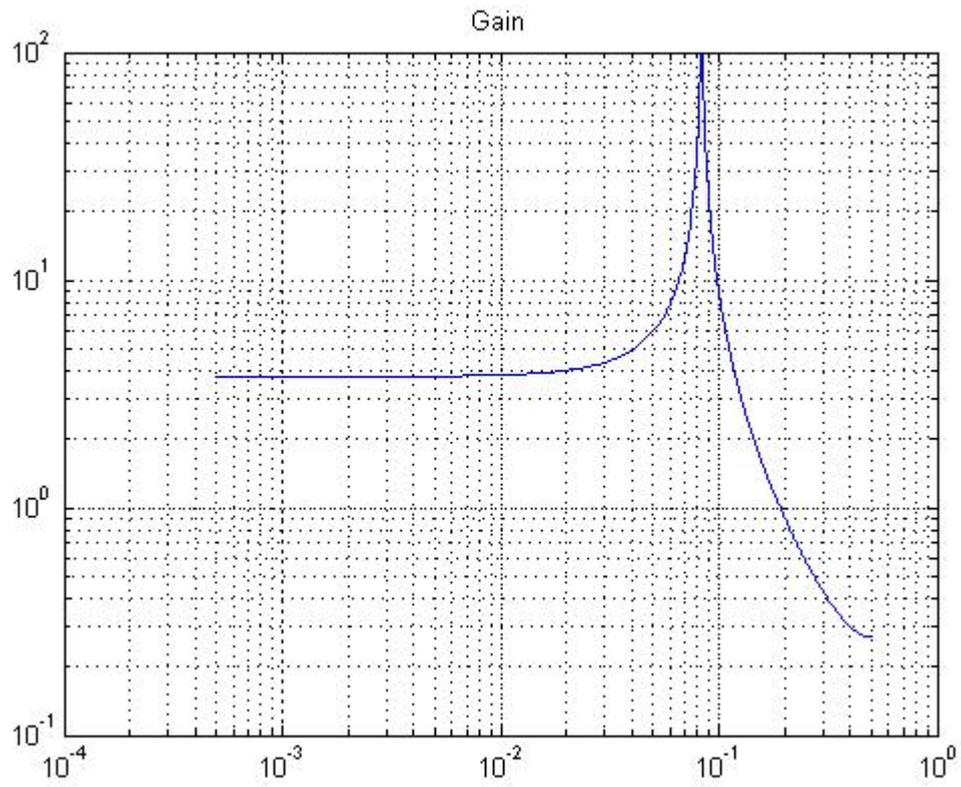


Figura 3-28

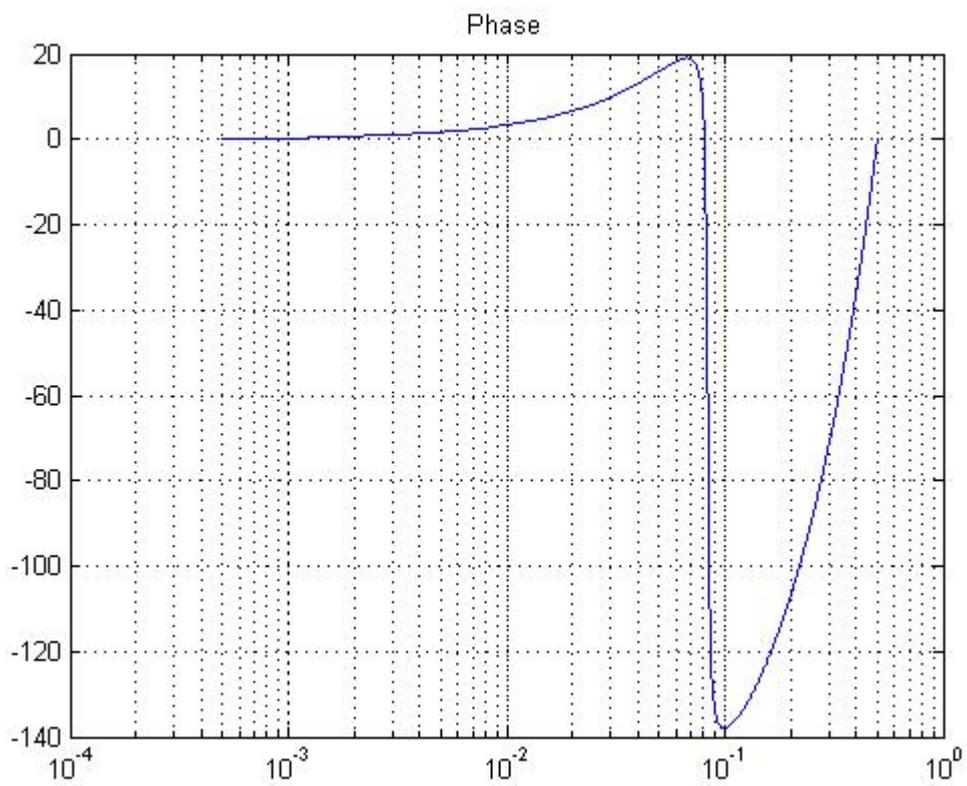


Figura 3-29-

e il diagramma di Nyquist

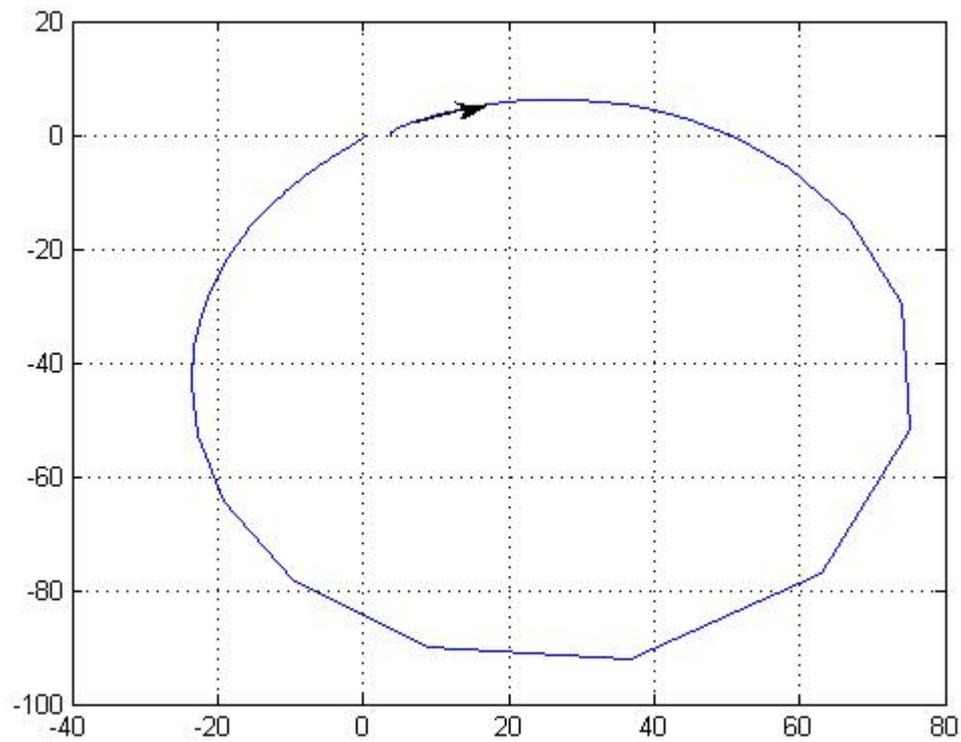


Figura 3-30

3.9.7 Semplici sistemi ARMA

Come esempi di sistemi di questo tipo prendiamo i seguenti:

- **sistema passa-alto** (ARMA[1,1]), con funzione di trasferimento

$$(3.104) \quad F(z) = \frac{1 - z^{-1}}{1 - w \cdot z^{-1}}$$

derivato dal passa-basso.
Per $w=0.9$ si ha

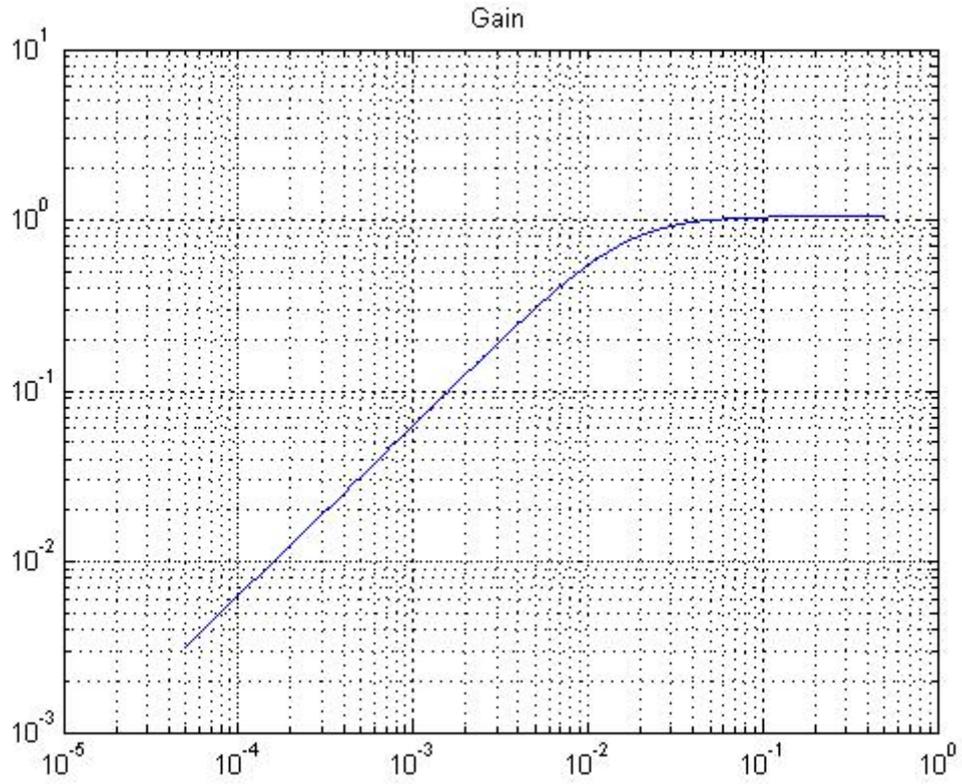


Figura 3-31

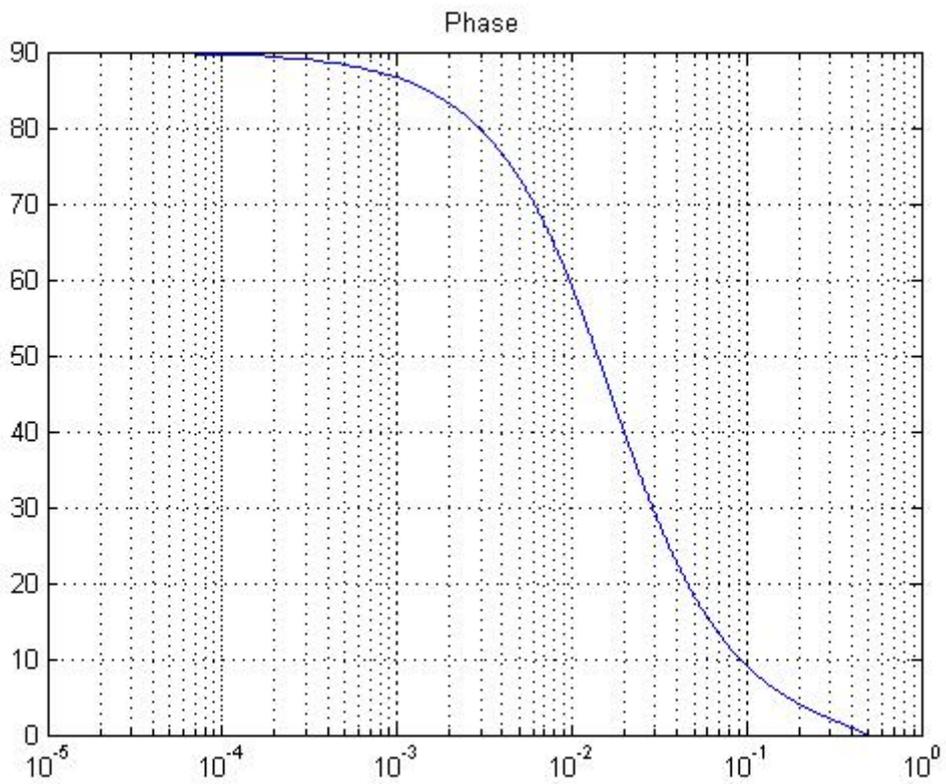


Figura 3-32

Diagramma di Nyquist

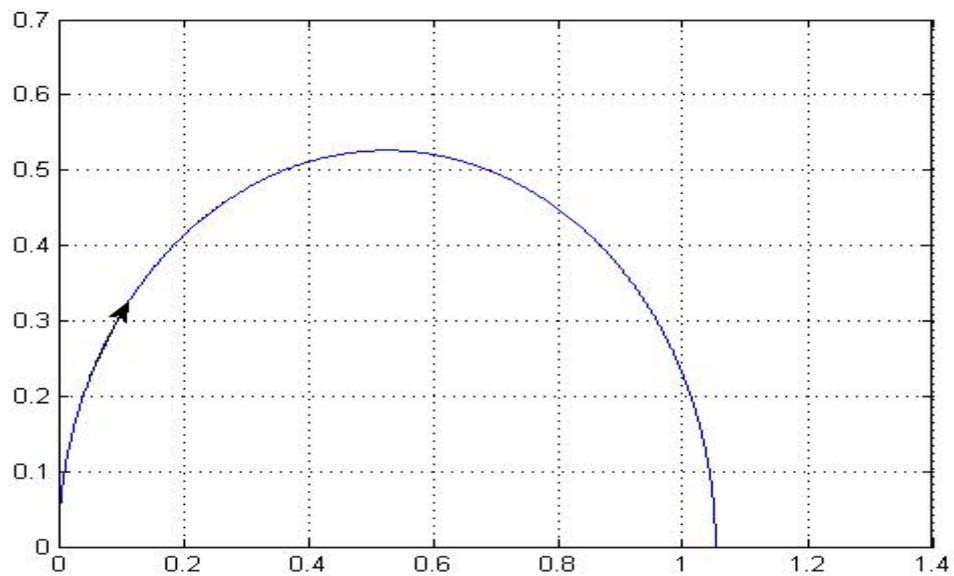


Figura 3-33

- **risonanza passa-banda** (ARMA[2,1]), con funzione di trasferimento

$$(3.105) \quad F(z) = \frac{1 - z^{-1}}{1 - 2r \cos \theta \cdot z^{-1} + r^2 \cdot z^{-2}}$$

derivato dall'AR del secondo ordine.

Per $r = 0.99$, $\theta = \frac{\pi}{60}$ si ha

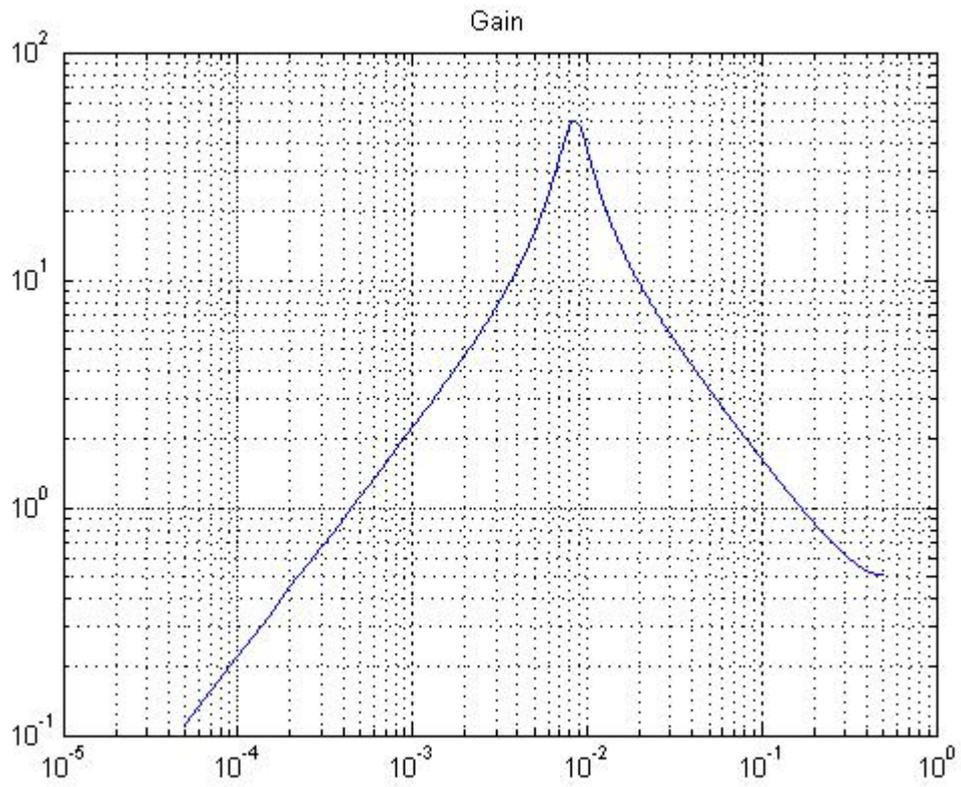


Figura 3-34

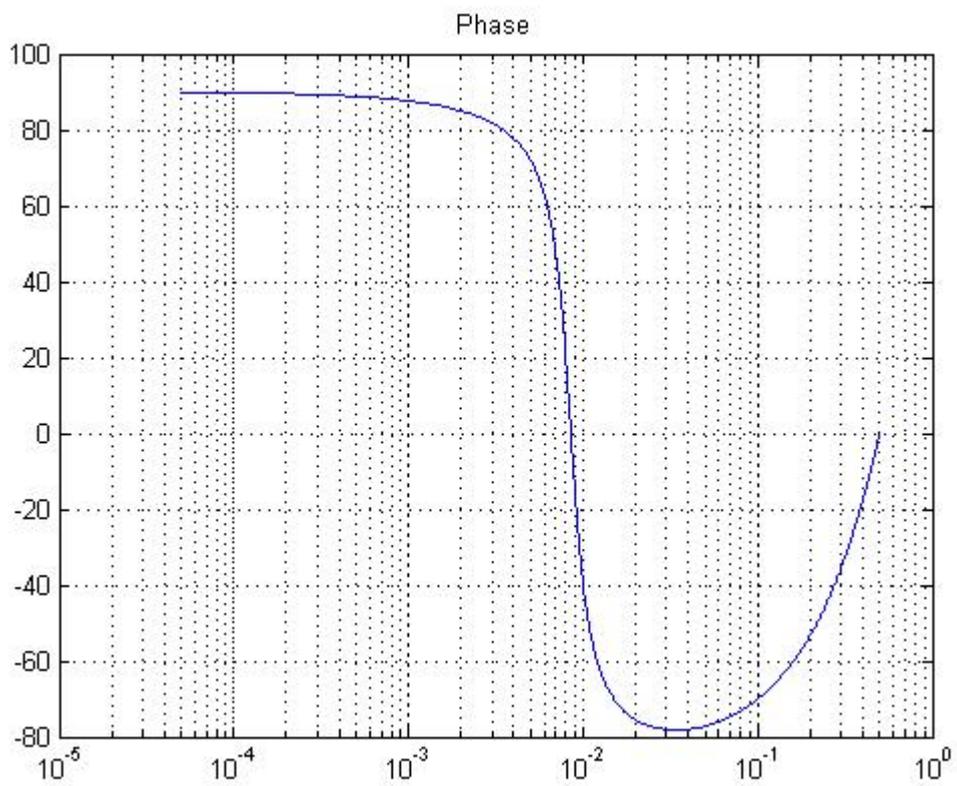


Figura 3-35

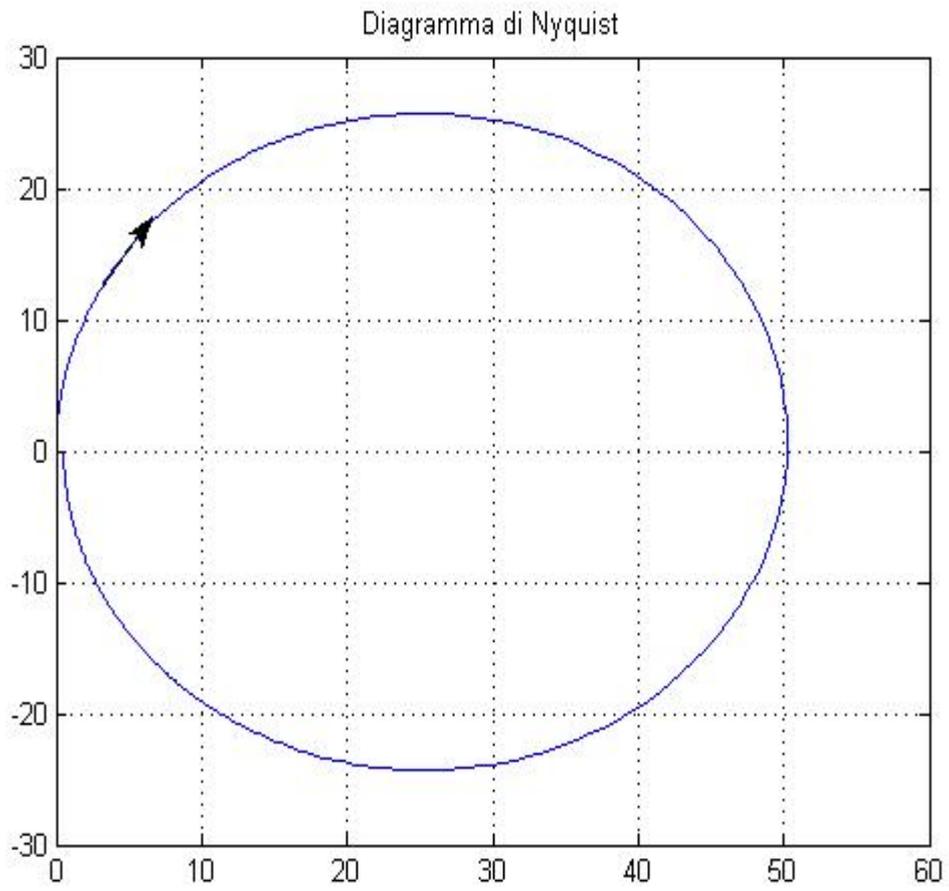


Figura 3-36

3.10 Sistemi non-lineari – i sistemi di Volterra

Facciamo un breve cenno ai sistemi non-lineari, una categoria così vasta che non esistono metodi di analisi generali (a parte cose piuttosto generiche).

Ci limiteremo a quelli che sono considerati il primo passo della non-linearità: **i sistemi di Volterra del secondo ordine** (i sistemi di Volterra del primo ordine sono semplicemente i sistemi lineari discreti). Essi sono descritti dall'equazione

$$(3.106) \quad y_i = \sum_{k=0}^{\infty} h_k \cdot x_{i-k} + \sum_{k=0}^{\infty} \sum_{l=0}^{\infty} q_{kl} \cdot x_{i-k} \cdot x_{i-l}$$

come si vede, oltre al membro lineare presente già nella (3.5), ce ne è un altro quadratico.

Si trova che nel dominio trasformato si ha

$$(3.107) \quad Y(\omega) = H(\omega) \cdot X(\omega) + \frac{1}{2} \cdot \int_{-\infty}^{\infty} Q(\omega_1, \omega - \omega_1) \cdot X(\omega_1) \cdot X(\omega - \omega_1) \cdot d\omega_1$$

4 Segnali transitori

4.1 Energia, posizione e lunghezza di un segnale transitorio

Un segnale **transitorio** (detto anche **transiente** o **impulsivo**) è un segnale che va a zero per il tempo tendente all'infinito in entrambi i sensi. Chiamiamo **energia del segnale**, la grandezza E_w

- Caso continuo, segnale $w(t)$

$$(4.1) \quad E_w = \int_{-\infty}^{\infty} |w(t)|^2 \cdot dt$$

- Caso discreto, segnale w_i

$$(4.2) \quad E_w = \sum_{i=-\infty}^{\infty} |w_i|^2$$

e supporremo inoltre che sia E finito.

È spesso comodo definire la “**posizione**” e la “**lunghezza**” (o “**durata**”) di un segnale parametri analoghi al valor medio e alla deviazione standard delle distribuzioni di probabilità :

- Caso continuo, segnale $w(t)$

$$(4.3) \quad \text{posizione} = t_w = \int_{-\infty}^{\infty} t \cdot \frac{|w(t)|^2}{E_w} \cdot dt$$

$$(4.4) \quad \text{lunghezza} = l_w = \sqrt{\int_{-\infty}^{\infty} (t - t_w)^2 \cdot \frac{|w(t)|^2}{E_w} \cdot dt}$$

- Caso discreto, segnale w_i

$$(4.5) \quad \text{posizione} = t_w = \sum_{i=-\infty}^{\infty} i \cdot \frac{|w_i|^2}{E_w}$$

$$(4.6) \quad \text{lunghezza} = l_w = \sqrt{\sum_{i=-\infty}^{\infty} (i-t_w)^2 \cdot \frac{|w_i|^2}{E_w}}$$

Se il segnale w è reale positivo, sono spesso più comode le definizioni di posizione e durata analoghe alle precedenti, ma che utilizzano $w(t)$ o, nel caso discreto, w_i (normalizzate con il loro integrale o somma) invece che i loro moduli quadri. Facciamo due esempi.

Prendiamo il segnale “gaussiano”

$$(4.7) \quad w(t) = A \cdot \exp\left(-\frac{t^2}{2 \cdot \sigma^2}\right)$$

Se prendiamo quest’ultima definizione di larghezza, abbiamo che la larghezza è σ . Con la definizione precedente abbiamo

$$|w(t)|^2 = \left| A \cdot \exp\left(-\frac{t^2}{2 \cdot \sigma^2}\right) \right|^2 = A^2 \cdot \exp\left(-\frac{t^2}{\sigma^2}\right)$$

e quindi

$$(4.8) \quad l_w = \frac{\sigma}{\sqrt{2}}$$

cioè un valore più piccolo.

Si noti che col nostro concetto di lunghezza o durata, funzionerebbe meglio $2 \cdot \sigma$.

Prendiamo ora un segnale pari a 1 durante il periodo da 0 a T e nullo altrove. In questo caso le due definizioni danno lo stesso valore, ma non quello che ci piacerebbe, cioè T , ma $\frac{T}{\sqrt{12}}$.

4.2 Convoluzione, cross-correlazione e distanza di due impulsi

Dati due impulsi x_i e y_i , possiamo calcolare alcune utili parametri e funzioni²²:

- il **prodotto scalare**, detto anche **cross-energia**

$$(4.9) \quad r_{xy} = \sum_{i=-\infty}^{\infty} x_i \cdot y_i^*$$

- la **distanza**

$$(4.10) \quad d_{xy} = \sqrt{\sum_{i=-\infty}^{\infty} |x_i - y_i|^2}$$

- la **convoluzione**,

$$(4.11) \quad y_k = \sum_{i=-\infty}^{\infty} x_{k-i} \cdot y_i$$

che già conosciamo.

- la **correlazione**

$$(4.12) \quad R_{xy}(k) = \sum_{i=-\infty}^{\infty} x_{k+i} \cdot y_i^*$$

²² Qui vengono date le espressioni per dati discreti. Analoghi concetti sono sviluppati nel caso continuo.

4.3 Trasformata di Fourier di un segnale ad energia finita

Data la condizione (4.1), un segnale transitorio, o impulso, può essere Fourier-trasformato

$$(4.13) \quad W(\omega) = \int_{-\infty}^{\infty} w(t) \cdot e^{-j\omega t} \cdot dt$$

Vale ovviamente il teorema di Parseval per cui

$$(4.14) \quad \int_{-\infty}^{\infty} |w(t)|^2 \cdot dt = \frac{1}{2\pi} \int_{-\infty}^{\infty} |W(\omega)|^2 \cdot d\omega = \int_{-\infty}^{\infty} |W(2\pi \cdot \nu)|^2 \cdot d\nu$$

dove $\nu = \frac{\omega}{2\pi}$ è la frequenza.

Nel discreto, usando la DTFT (la trasformata discreta di Fourier), abbiamo

$$(4.15) \quad W(\Omega) = \sum_{i=-\infty}^{\infty} w_i \cdot e^{-j\Omega i}$$

(ricordiamo che $\Omega = \omega \cdot \Delta t$ è la pulsazione normalizzata).

Il teorema di Parseval nel discreto diventa

$$(4.16) \quad \sum_{i=-\infty}^{\infty} |w_i|^2 = \frac{1}{2\pi} \int_0^{2\pi} |W(\Omega)|^2 \cdot d\Omega$$

Analogamente a come abbiamo fatto per $w(t)$, possiamo definire posizione e larghezza per $W(\Omega)$. Sia L_w la larghezza di $W(\Omega)$; si ha

$$(4.17) \quad l_w \cdot L_w \geq \frac{1}{2}$$

L'eguaglianza vale se

$$(4.18) \quad w(t) = A \cdot \exp\left(-\frac{t^2}{4 \cdot l_w^2}\right)$$

In tal caso la trasformata di Fourier è

$$(4.19) \quad W(\omega) = A \cdot l_w \cdot 2\sqrt{\pi} \cdot \exp(-l_w^2 \cdot \omega^2)$$

e quindi

$$(4.20) \quad l_w \cdot L_w = \frac{1}{2}$$

q.e.d..

Attenzione ! Se il segnale è reale, la trasformata di Fourier è una funzione hermitiana e quindi $|W(\omega)|^2$, che è usata per calcolare la larghezza, è simmetrica rispetto all'origine. Questo fa sì che, nel caso di $w(t)$ "a banda stretta", la larghezza di $W(\omega)$ sia molto più grande di quella che sarebbe se si considerassero solo le frequenze positive. Se, per i segnali reali, si considerano le sole frequenze positive, si ricava una nuova relazione di "indeterminazione"

$$(4.21) \quad l_w \cdot L_w \geq \frac{1}{2}$$

dove l_w è lo stesso, ma L_w (anche molto) più piccolo, che è molto più "pregnante".

Facciamo come esempio

$$(4.22) \quad w(t) = A \cdot \exp\left(-\frac{t^2}{4 \cdot l_w^2}\right) \cdot \cos(\omega_0 \cdot t)$$

la cui trasformata (vedi (2.22)) è proporzionale a

$$(4.23) \quad \exp\left(-\frac{(\omega + \omega_0)^2}{4 \cdot \Lambda_w^2}\right) + \exp\left(-\frac{(\omega - \omega_0)^2}{4 \cdot \Lambda_w^2}\right)$$

con $\Lambda_w = \frac{1}{2 \cdot l_w}$. Se $\Lambda_w \ll \omega_0$, il calcolo su tutte le frequenze porta a $L_w \approx \omega_0$ e quindi

$l_w \cdot L_w \gg \frac{1}{2}$; se ci si limita alle frequenze positive invece $L_w \approx \Lambda_w$ e quindi $l_w \cdot L_w \approx \frac{1}{2}$.

4.4 Autocorrelazione e spettro di energia

Se facciamo la correlazione incrociata di un segnale con se stesso, abbiamo l'autocorrelazione

$$(4.24) \quad R_x(k) = \sum_{i=-\infty}^{\infty} x_{k+i} \cdot x_i^*$$

Questa definizione di autocorrelazione per i segnali impulsivi è ovviamente diversa, ma analoga, dall'autocorrelazione nel caso dei processi stocastici. Come l'autocorrelazione nel caso dei processi stocastici, gode di varie proprietà:

- è la convoluzione del segnale con se stesso invertito temporalmente (e coniugato, se il segnale è complesso)
- è simmetrica rispetto a 0
- ha ivi il suo massimo assoluto (e in questo caso in 0 vale 1)
- la sua trasformata di Fourier è sempre positiva (ed è lo spettro di energia)
- la "larghezza" dell'autocorrelazione (definita come in (4.6) e (4.4)) è $\sqrt{2}$ volte maggiore della w .

Possiamo definire **spettro di energia** la trasformata di Fourier dell'autocorrelazione

$$(4.25) \quad S_x(\omega) = \sum_k R_x(k) \cdot e^{-j\omega k}$$

essendo $R_x(k)$ ricavato dalla (4.24).

Vale il seguente risultato (teorema di Wiener-Kinchin):

$$(4.26) \quad S_x(\omega) = \left| \sum_k x_k \cdot e^{-j\omega k} \right|^2$$

4.5 Segnale analitico

La trasformata di Hilbert di un segnale $x(t)$ è definita come

$$(4.27) \quad x'(t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x(\tau)}{t-\tau} \cdot d\tau$$

e l'antitrasformata è

$$(4.28) \quad x(t) = -\frac{1}{\pi} \int_{-\infty}^{\infty} \frac{x'(\tau)}{t-\tau} \cdot d\tau$$

Questa trasformata è in effetti un filtro lineare che ha la proprietà fondamentale che

$$(4.29) \quad \begin{aligned} \sin(\omega \cdot t) &\leftrightarrow \cos(\omega \cdot t) \\ \cos(\omega \cdot t) &\leftrightarrow -\sin(\omega \cdot t) \end{aligned}$$

per questo motivo è detta anche **filtro di quadratura**, che “sfasa” ogni componente sinusoidale di 90 gradi. Tale filtro ha la funzione di trasferimento

$$(4.30) \quad H(j\omega) = -j \cdot \text{sign}(\omega) = \begin{cases} -j & \text{per } \omega > 0 \\ j & \text{per } \omega < 0 \end{cases}$$

Se $x'(t)$ è la trasformata di Hilbert del segnale reale $x(t)$, allora definiamo segnale analitico il segnale

$$(4.31) \quad x_A(t) = x(t) + j \cdot x'(t)$$

Se $X(\omega)$ è la trasformata di Fourier di $x(t)$, allora la trasformata di Fourier di $x_A(t)$ è

$$(4.32) \quad X_A(\omega) = \begin{cases} 0 & \text{per } \omega < 0 \\ 2 \cdot X(\omega) & \text{per } \omega > 0 \end{cases}$$

La trasformata di Hilbert si può definire anche nel discreto come un filtro di risposta impulsiva

$$(4.33) \quad h_i = \begin{cases} 0 & \text{per } i \text{ pari} \\ \frac{2}{\pi \cdot i} & \text{per } i \text{ dispari} \end{cases}$$

Vediamo un esempio.

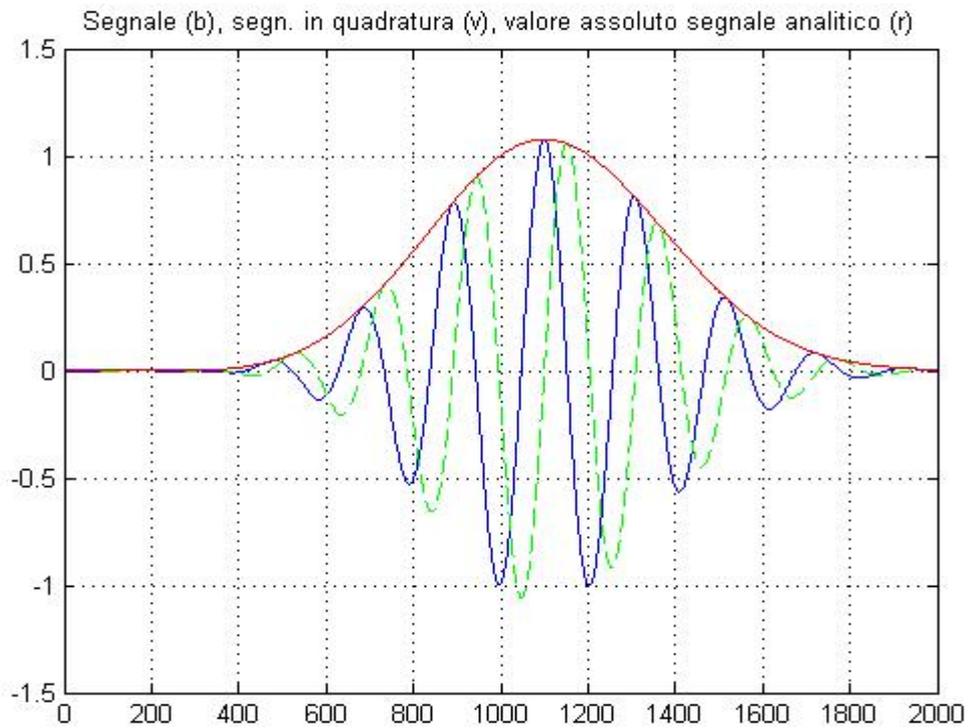


Figura 4-1

Come si vede, il valore assoluto del segnale analitico è l'involuppo del segnale originario.

Ed ecco questo segnale analitico nel piano di Gauss

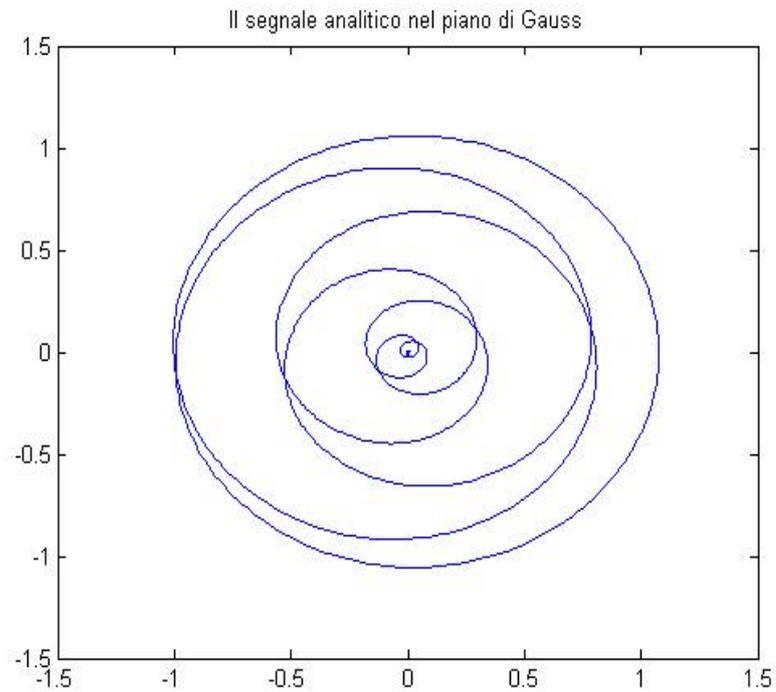


Figura 4-2

4.6 Segnali permanenti

Chiamiamo segnale permanente un segnale che non è “localizzato” come un impulso. Per tali segnali non possiamo definire l’energia (che sarebbe infinita), né la posizione, né la larghezza.

Per trattare questo tipo di segnali useremo la teoria dei processi stocastici. Come vedremo, in taluni casi (processi stazionari) potremo parlare di un concetto analogo a quello di energia, la potenza

- Caso continuo, segnale $w(t)$

$$(4.34) \quad P_w = \lim_{T \rightarrow \infty} \frac{1}{2 \cdot T} \int_{-T}^T |w(t)|^2 \cdot dt$$

- Caso discreto, segnale w_i

$$(4.35) \quad P_w = \lim_{K \rightarrow \infty} \frac{1}{2 \cdot K} \sum_{i=-K}^K |w_i|^2$$

Per un segnale impulsivo, la P_w (che non ha senso calcolare) sarebbe 0.

Analogamente all’autocorrelazione e allo spettro di energia che abbiamo definito per i segnali transitori, definiremo le analoghe funzioni per i segnali permanenti, autocorrelazione e spettro di potenza, che possono essere viste come valori medi per unità di tempo.

5 Processi stocastici

5.1 Introduzione

5.1.1 Definizione

Ricordiamo la nostra definizione di variabile casuale:

“Una variabile casuale è una variabile (reale o complessa) al cui valore è associata una distribuzione di probabilità.”

In analogia, possiamo dare la seguente definizione di processo stocastico:

Un processo stocastico è un insieme di funzioni del tempo alle quali è associata una distribuzione di probabilità.

Ciascuna delle funzioni del tempo viene chiamata **realizzazione** del processo stocastico. Mentre il processo stocastico è il modello generale, quello che si può osservare sperimentalmente è soltanto una realizzazione.

Se le funzioni del tempo sono continue, parliamo di **processo stocastico continuo**, se sono discrete di **processo stocastico discreto**. Nel seguito svilupperemo il formalismo per lo più per il caso continuo, ma analoghi sviluppi si hanno nel caso discreto (dove, ricordiamo, la variabile t è discreta).

Per un processo stocastico $X(t)$, possiamo definire la funzione di distribuzione del primo ordine

$$(5.1) \quad F(x;t) = P\{X(t) \leq x\}$$

da cui la densità

$$(5.2) \quad f(x;t) = \frac{\partial F(x;t)}{\partial x}$$

Abbiamo quindi il valor medio (o valore aspettato)

$$(5.3) \quad \mu(t) = E[X(t)] = \int_{-\infty}^{\infty} x \cdot f(x;t) \cdot dx$$

e la varianza

$$(5.4) \quad \sigma^2(t) = E\left[(X(t) - \mu(t))^2\right] = \int_{-\infty}^{\infty} (x - \mu(t))^2 \cdot f(x;t) \cdot dx$$

Per un dato valore del tempo t , un processo stocastico si riduce a una variabile casuale.

Si definiscono inoltre, per qualsiasi $n > 0$, le funzioni di distribuzione di ordine n come

$$(5.5) \quad F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2, \dots, X(t_n) \leq x_n\}$$

e le relative densità

$$(5.6) \quad f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{\partial^n F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$$

Un processo stocastico è statisticamente definito se sono definite tutte le funzioni di distribuzione (o di densità) per tutti gli ordini.

Dato un processo stocastico, possiamo crearne uno nuovo facendolo passare attraverso un opportuno sistema. Il nuovo processo ha in genere caratteristiche differenti da quello di partenza.

5.1.2 Funzioni del secondo ordine

Consideriamo le funzioni del secondo ordine,

$$(5.7) \quad F(x_1, x_2; t_1, t_2) = P\{X(t_1) \leq x_1, X(t_2) \leq x_2\}$$

e

$$(5.8) \quad f(x_1, x_2; t_1, t_2) = \frac{\partial^2 F(x_1, x_2; t_1, t_2)}{\partial x_1 \partial x_2}$$

Si noti che

$$(5.9) \quad F(x_1, \infty; t_1, t_2) = F(x_1; t_1)$$

e

$$(5.10) \quad f(x_1; t_1) = \int_{-\infty}^{\infty} f(x_1, x_2; t_1, t_2) \cdot dx_2$$

Possiamo calcolare la densità condizionale come

$$(5.11) \quad f(x_1; t_1 | X(t_2) = x_2) = \frac{f(x_1, x_2; t_1, t_2)}{f(x_2; t_2)}$$

Definiamo **autocorrelazione** la funzione

$$(5.12) \quad R(t_1, t_2) = E[X(t_1) \cdot X(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2 \cdot f(x_1, x_2; t_1, t_2) \cdot dx_1 \cdot dx_2$$

e **autocovarianza** la funzione

$$\begin{aligned}
 C(t_1, t_2) &= E\left[\left(X(t_1) - \mu(t_1)\right) \cdot \left(X(t_2) - \mu(t_2)\right)\right] = \\
 (5.13) \quad &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x_1 - \mu(t_1)) \cdot (x_2 - \mu(t_2)) \cdot f(x_1, x_2; t_1, t_2) \cdot dx_1 \cdot dx_2 = \\
 &= R(t_1, t_2) - \mu(t_1) \cdot \mu(t_2)
 \end{aligned}$$

la varianza si può ricavare come

$$(5.14) \quad \sigma^2(t) = C(t, t) = R(t, t) - \mu^2(t)$$

Se il processo è complesso (cioè se sono complesse le funzioni del tempo), allora

$$(5.15) \quad R(t_1, t_2) = E\left[X(t_1) \cdot X^*(t_2)\right]$$

e

$$(5.16) \quad C(t_1, t_2) = E\left[\left(X(t_1) - \mu(t_1)\right) \cdot \left(X^*(t_2) - \mu^*(t_2)\right)\right]$$

In questo caso C ed R sono funzioni complesse. Si noti che è una generalizzazione del caso reale.

Dall'autocorrelazione $R(t_1, t_2) = R(t_1, t_1 + \tau) = R(t, t + \tau)$ possiamo ricavare lo **spettro di potenza** (spesso chiamato semplicemente spettro) come la trasformata di Fourier

$$(5.17) \quad S(t, \omega) = \int_{-\infty}^{\infty} R(t, \tau) \cdot e^{-j\omega\tau} d\tau$$

Sebbene lo spettro di potenza contenga esattamente la stessa informazione dell'autocorrelazione, spesso quest'ultimo è più facilmente interpretabile.

Lo spettro di potenza, più correttamente la **densità spettrale di potenza**, indica il contenuto "in potenza" (cioè quadratico) di un segnale. Ne parleremo estesamente in un prossimo capitolo. Lì discuteremo anche il senso delle frequenze negative.

5.1.3 Il caso di due processi

Supponiamo di avere due processi X(t) e Y(t), che per generalità consideriamo complessi.

Possiamo definire la **correlazione incrociata**²³ (**cross-correlation**)

$$(5.18) \quad R_{xy}(t_1, t_2) = E[X(t_1) \cdot Y^*(t_2)] = R_{yx}^*(t_1, t_2)$$

e la **covarianza incrociata** (**cross-covariance**)

$$(5.19) \quad C_{xy}(t_1, t_2) = R_{xy}(t_1, t_2) - \mu_x(t_1) \cdot \mu_y(t_2)$$

Diciamo che i due processi sono **scorrelati** se

$$(5.20) \quad R_{xy}(t_1, t_2) = \mu_x(t_1) \cdot \mu_y(t_2)$$

e diciamo che sono **ortogonali** se

$$(5.21) \quad R_{xy}(t_1, t_2) = 0$$

A partire dalla correlazione incrociata $R_{xy}(t_1, t_2) = R_{xy}(t_1, t_1 + \tau) = R_{xy}(t, t + \tau)$, possiamo definire lo **spettro incrociato** (**cross-spectrum**) come

$$(5.22) \quad S_{xy}(t, \omega) = \int_{-\infty}^{\infty} R_{xy}(t, \tau) \cdot e^{-j\omega\tau} d\tau = S_{yx}^*(t, \omega)$$

5.1.4 Stazionarietà, ergodicità

Diciamo che un processo $X(t)$ è stazionario²⁴ se le funzioni che lo definiscono sono indipendenti da uno spostamento nel tempo di qualsiasi ampiezza. Cioè se per qualsiasi T $X(t)$ è statisticamente indistinguibile da $X(t+T)$. Si ha cioè

$$(5.23) \quad F(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = F(x_1, x_2, \dots, x_n; t_1 + T, t_2 + T, \dots, t_n + T)$$

e

$$(5.24) \quad f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = f(x_1, x_2, \dots, x_n; t_1 + T, t_2 + T, \dots, t_n + T)$$

Ne deduciamo che per un processo stocastico stazionario si ha

²³ Detta anche correlazione mutua, intercorrelazione o cross-correlazione.

²⁴ Questa è la definizione di un processo stocastico stazionario “in senso stretto”. Si definisce anche la stazionarietà “in senso lato”.

$$(5.25) \quad f(x;t) = f(x)$$

$$\mu(t) = \mu = \text{costante}$$

$\mu = E[x(t)]$ è anche detto **cumulante del primo ordine**.

ponendo $\tau = t_2 - t_1$, abbiamo

$$f(x_1, x_2; t_1, t_2) = f(x_1, x_2; \tau)$$

$$(5.26) \quad R(t_1, t_2) = R(\tau) = R^*(-\tau)$$

$$C(t_1, t_2) = C(\tau) = C^*(-\tau)$$

$R_{xx}(\tau) = R(\tau) = E[x^*(t) \cdot x(t+\tau)]$ è, nell'ipotesi che sia $E[x(t)] = 0$, anche il **cumulante del secondo ordine** $C_{xx}(\tau)$.

Notiamo che

$$(5.27) \quad R_{xx}(0) - \mu^2 = \sigma^2$$

è la varianza del processo.

Lo spettro di potenza in questo caso perde la dipendenza dal tempo e viene definito come

$$(5.28) \quad S(\omega) = \int_{-\infty}^{\infty} R(\tau) \cdot e^{-j\omega\tau} d\tau$$

Possiamo definire la potenza totale del processo l'integrale totale su tutta la banda di frequenza (non pulsazione), ottenendo

$$(5.29) \quad R(0) = \frac{1}{2\pi} \int_{-\infty}^{\infty} S(\omega) d\omega = \int_{-\infty}^{\infty} S(2\pi \cdot \nu) d\nu$$

Possono definirsi cumulanti di altri ordini, che, nell'ipotesi che sia $E[x(t)] = 0$, sono nel caso del terzo ordine

$$(5.30) \quad C_{xxx}(\tau_1, \tau_2) = E[x^*(t) \cdot x(t+\tau_1) \cdot x(t+\tau_2)]$$

e del quarto ordine

$$(5.31) \quad C_{xxxx}(\tau_1, \tau_2, \tau_3) = E[x^*(t) \cdot x(t+\tau_1) \cdot x(t+\tau_2) \cdot x^*(t+\tau_3)] - \\ - C_{xx}(\tau_1) \cdot C_{xx}(\tau_2 - \tau_3) - C_{xx}(\tau_2) \cdot C_{xx}(\tau_1 - \tau_3) - M_{xx}^*(\tau_3) \cdot M_{xx}(\tau_1 - \tau_2)$$

dove

$$(5.32) \quad M_{xx}(\tau) = E[x(t) \cdot x(t+\tau)]$$

è il momento del secondo ordine (eguale al cumulante del secondo ordine nel caso di processi reali).

e notiamo che da $C_{xxx}(0,0)$ si ricava il momento del terzo ordine e da $C_{xxxx}(0,0,0)$ quello del quart'ordine.

Tutti i cumulanti di processi stazionari reali sono simmetrici nei loro argomenti, cioè

$$C_{xx}(\tau) = C_{xx}(-\tau)$$

$$(5.33) \quad C_{xxx}(\tau_1, \tau_2) = C_{xxx}(\tau_2, \tau_1) = C_{xxx}(-\tau_1, \tau_2 - \tau_1)$$

$$C_{xxxx}(\tau_1, \tau_2, \tau_3) = C_{xxxx}(\tau_2, \tau_1, \tau_3) = C_{xxxx}(\tau_1, \tau_3, \tau_2) = C_{xxxx}(-\tau_1, \tau_2 - \tau_1, \tau_3 - \tau_1)$$

La trasformata di Fourier di $C_{xxx}(\tau_1, \tau_2)$ è detto **bispettro** e si indica con $S_{xxx}(\omega_1, \omega_2)$,

$$(5.34) \quad S_{xxx}(\omega_1, \omega_2) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{xxx}(\tau_1, \tau_2) \cdot e^{-j\omega_1\tau_1} \cdot e^{-j\omega_2\tau_2} \cdot d\tau_1 \cdot d\tau_2$$

la trasformata di Fourier di $C_{xxxx}(\tau_1, \tau_2, \tau_3)$ è detto **trispettro** e si indica con $S_{xxxx}(\omega_1, \omega_2, \omega_3)$

$$(5.35) \quad S_{xxxx}(\omega_1, \omega_2, \omega_3) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{xxxx}(\tau_1, \tau_2, \tau_3) \cdot e^{-j\omega_1\tau_1} \cdot e^{-j\omega_2\tau_2} \cdot e^{-j\omega_3\tau_3} \cdot d\tau_1 \cdot d\tau_2 \cdot d\tau_3$$

Per due processi **congiuntamente stazionari (jointly stationary)** abbiamo

$$(5.36) \quad R_{xy}(t_1, t_2) = R_{xy}(\tau) = E[X(t+\tau) \cdot Y^*(t)]$$

Diciamo **ergodico** un processo in cui le statistiche di insieme sono eguali alle statistiche temporali. In altre parole possiamo conoscere tutte le proprietà del processo stocastico osservando la realizzazione effettiva. È intuitivo che un processo non può essere ergodico se non è stazionario.

La stazionarietà e l'ergodicità possono essere ristrette ad alcune caratteristiche del processo (per esempio la media, la varianza o l'autocorrelazione).

5.1.5 Esempi

Facciamo tre esempi (di processi continui):

1. Consideriamo un processo che contiene 3 funzioni del tempo

$$x_1(t) = 0 \quad \rightarrow \text{Prob. } \frac{1}{2}$$

$$x_2(t) = \sin(2\pi \cdot 50 \cdot t) \quad \rightarrow \text{Prob. } \frac{1}{4}$$

$$x_3(t) = \sin(2\pi \cdot 100 \cdot t) \quad \rightarrow \text{Prob. } \frac{1}{4}$$

Questo processo stocastico potrebbe essere il modello di un dispositivo elettrico che con probabilità 0.5 è acceso (e ovviamente con la stessa probabilità è spento) e quando è acceso può generare con eguale probabilità una sinusoide a 100 Hz o una a 50.

2. Consideriamo un processo che è composto di infinite funzioni del tempo definite da

$$x_A(t) = A \cdot \sin(10 \cdot t)$$

dove A è distribuita uniformemente tra 0 e 2.

Questo processo stocastico potrebbe essere il modello di un oscillatore elettrico che genera una tensione sinusoidale, con una manopola che ne stabilisce l'ampiezza, e la manopola può stare in una qualsiasi posizione con la stessa probabilità (ma non può essere cambiata di posizione).

3. Consideriamo tutte le possibili funzioni del tempo che possono aversi come uscita di un contatore di particelle che rivela per esempio raggi cosmici. Anche in questo caso ci troviamo di fronte a un processo stocastico, anche se è più difficile conoscere la distribuzione di probabilità delle varie realizzazioni (e in parecchi casi non serve, essendo sufficienti altre funzioni, più semplici, legate alla distribuzione di probabilità).

5.2 Trasformazioni di processi stocastici

Dato un processo stocastico X, possiamo assegnare secondo certe regole a ciascuna funzione del tempo $x(t)$ del processo, una funzione $y(t)$, ottenendo un nuovo processo stocastico Y. Y può essere considerato l'uscita di un opportuno sistema.

Far passare un processo stocastico X attraverso un sistema F significa costruire un nuovo processo Y in cui tutte le funzioni del tempo $x(t)$ sono sostituite dalle uscite $y(t)$ di F alle $x(t)$.

Consideriamo alcuni casi particolari.

5.2.1 Sistema statico (senza memoria)

Sia in tal caso

$$(5.37) \quad Y(t) = g(X(t))$$

Le funzioni di densità del processo Y sono ricavabili con le tecniche di trasformazione delle variabili aleatorie. Si ha

$$(5.38) \quad f_y(y_1, y_2, \dots, y_n; t_1, t_2, \dots, t_n) = \frac{f_x(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n)}{|g'(x_1)| \cdot |g'(x_2)| \cdot \dots \cdot |g'(x_n)|}$$

con $y_i = g(x_i)$.

Per quanto riguarda il valor medio e l'autocorrelazione abbiamo

$$(5.39) \quad E[Y(t)] = \int_{-\infty}^{\infty} g(x) \cdot f_x(x; t) \cdot dx$$

e

$$(5.40) \quad R_{yy}(t_1, t_2) = E[y(t_1) \cdot y(t_2)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1) \cdot g(x_2) \cdot f_x(x_1, x_2; t_1, t_2) \cdot dx_1 \cdot dx_2$$

Inoltre, se X è stazionario, anche Y lo è.

5.2.2 Sistema lineare (tempo invariante)

Indichiamo con L l'operatore lineare che descrive il sistema. (vedi (1.10)). Si hanno i seguenti risultati:

$$(5.41) \quad E[Y(t)] = L[E[X(t)]]$$

$$(5.42) \quad R_{xy}(t_1, t_2) = L_{t_2}[R_{xx}(t_1, t_2)]$$

$$(5.43) \quad R_{yy}(t_1, t_2) = L_{t_1}[R_{xy}(t_1, t_2)]$$

Con L_{t_2} indichiamo “rispetto alla variabile t_2 ”, considerandola variabile corrente e t_1 parametro; analogamente per L_{t_1} .

Se il processo in ingresso a un sistema lineare tempo invariante è stazionario, lo è anche quello in uscita.

Nell'ipotesi di stazionarietà, supponiamo di avere un sistema (continuo) definito da una funzione di trasferimento $F(s)$, ovvero dalla risposta impulsiva $f(t)$. Per il valor medio si ha

$$(5.44) \quad E[Y(t)] = \mu_Y = E\left[\int_{-\infty}^{\infty} X(t-\theta) f(\theta) \cdot d\theta\right] = \mu_X \cdot \int_{-\infty}^{\infty} f(\theta) \cdot d\theta = \mu_X \cdot F(0)$$

Per calcolare l'autocorrelazione dell'uscita è opportuno calcolare prima la correlazione incrociata tra l'uscita e l'ingresso:

$$(5.45) \quad R_{YX}(\tau) = E[Y(t) \cdot X^*(t-\tau)] = E\left[\int_{-\infty}^{\infty} x(t-\theta) \cdot x^*(t-\tau) \cdot f(\theta) d\theta\right] = \\ = \int_{-\infty}^{\infty} R_{XX}(\tau-\theta) \cdot f(\theta) d\theta$$

e

$$(5.46) \quad R_{YX}(\tau) = \int_{-\infty}^{\infty} R_{XX}(\tau-\theta) \cdot f^*(-\theta) d\theta$$

e quindi

$$(5.47) \quad R_{YY}(\tau) = E[Y(t+\tau) \cdot Y^*(t)] = E\left[\int_{-\infty}^{\infty} y(t+\tau) \cdot x^*(t-\theta) \cdot f^*(\theta) d\theta\right] = \\ = \int_{-\infty}^{\infty} R_{YX}(\tau+\theta) \cdot f^*(\theta) d\theta$$

Ora, ricordando la proprietà della convoluzione e la definizione di spettro di potenza, troviamo

$$(5.48) \quad S_{XY}(\omega) = S_{XX}(\omega) \cdot F^*(j\omega)$$

e

$$(5.49) \quad S_{YY}(\omega) = S_{XX}(\omega) \cdot |F(j\omega)|^2$$

Lo stesso risultato per gli spettri possiamo ottenerlo ragionando sul significato di spettro di potenza e di funzione di trasferimento.

Infatti, in una banda molto piccola intorno a ω_0 il segnale è approssimabile con una sinusoide con una certa ampiezza A e una certa fase; lo spettro di potenza è kA^2 , dove k è una

opportuna costante, e indipendente dalla fase. Il passaggio attraverso il sistema lineare modifica ovviamente tale valore di un fattore $|F(j\omega_0)|^2$, come descritto dalla (5.49). Analogamente si può ragionare per la (5.48).

5.2.3 Un caso particolare: il derivatore

Il **derivatore**, che esegue la derivata di un processo stocastico, spesso detta **derivata stocastica** (ovviamente le funzioni che compongono il processo devono essere derivabili). Abbiamo in questo caso

$$(5.50) \quad E[x'(t)] = \frac{dE[x(t)]}{dt}$$

$$(5.51) \quad R_{xx'}(t_1, t_2) = \frac{\partial R_{xx}(t_1, t_2)}{\partial t_2}$$

$$(5.52) \quad R_{x'x'}(t_1, t_2) = \frac{\partial R_{xx'}(t_1, t_2)}{\partial t_1} = \frac{\partial^2 R_{xx}(t_1, t_2)}{\partial t_1 \partial t_2}$$

e se il processo è stazionario, con $\tau = t_1 - t_2$,

$$(5.53) \quad R_{xx'}(\tau) = -\frac{dR_{xx}(\tau)}{d\tau}$$

$$(5.54) \quad R_{x'x'}(\tau) = \frac{dR_{xx'}(\tau)}{d\tau^2} = -\frac{d^2 R_{xx}(\tau)}{d\tau^2}$$

5.3 Processi stocastici normali

5.3.1 Proprietà fondamentali

Chiamiamo **normale** (o **gaussiano**) un processo stocastico. in cui sono normali le funzioni di densità di qualsiasi ordine.

Ci sono due motivi per cui i processi stocastici normali sono di grande importanza:

1. Molto spesso fenomeni naturali (e non) sono ben rappresentati da processi stocastici normali
2. Grazie alle proprietà matematiche che appresso enunceremo, si può costruire una semplice ed elegante teoria dei processi stocastici normali, per cui, anche se i fenomeni di interesse non sono esattamente rappresentabili con processi normali, si preferisce approssimarli, quando possibile, con essi.

Ecco le due proprietà fondamentali:

PROPRIETÀ I - Per un processo stocastico normale la funzione densità di ordine n (per qualsiasi n) è determinata se sono note la media $\mu(t)$ e l'autocorrelazione $R(t_1, t_2)$ (o l'autocovarianza $C(t_1, t_2) = R(t_1, t_2) - \mu(t_1) \cdot \mu(t_2)$).

Infatti, definita la matrice \mathbf{M} di covarianza tra le n $x(t_i)$ con elementi

$$(5.55) \quad m_{ij} = C(t_i, t_j)$$

abbiamo

$$(5.56) \quad f(x_1, x_2, \dots, x_n; t_1, t_2, \dots, t_n) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{M}|}} \cdot \exp \left[-\frac{(\mathbf{x} - \boldsymbol{\mu}) \cdot \mathbf{M}^{-1} \cdot (\mathbf{x}^T - \boldsymbol{\mu}^T)}{2} \right]$$

dove $\mathbf{x} = \{x_1, x_2, \dots, x_n\}$ e $\boldsymbol{\mu} = \{\mu(t_1), \mu(t_2), \dots, \mu(t_n)\}$.

Come sappiamo, una trasformazione lineare di una variabile casuale normale produce una variabile casuale normale. Analogamente abbiamo che:

PROPRIETÀ II - Una trasformazione lineare di un processo stocastico normale produce un processo stocastico normale.

Cioè se facciamo passare un processo normale attraverso un sistema lineare, otteniamo in uscita un processo normale.

Particolarmente semplice e illuminante è lo sviluppo nell'ipotesi stazionaria, che d'ora in poi sottintenderemo verificata.

5.3.2 Il rumore bianco

Il rumore bianco. è un processo stocastico normale stazionario (ed ergodico) definito da

$$\mu(t) = 0$$

(5.57)

$$R(\tau) = \delta(\tau)$$

Cioè due valori del processo a due istanti qualsiasi, comunque vicini, sono scorrelati (e indipendenti: ricordiamo che due variabili casuali normali con correlazione nulla sono indipendenti).

È immediato vedere che la densità spettrale di potenza è

$$(5.58) \quad S(\omega) = \int_{-\infty}^{\infty} R(\tau) \cdot e^{-j\omega\tau} d\tau = \int_{-\infty}^{\infty} \delta(\tau) \cdot e^{-j\omega\tau} d\tau = 1$$

Lo spettro quindi contiene tutte le frequenze, con la stessa ampiezza. Comprendiamo allora il motivo del nome “bianco”: è in analogia con la luce bianca, che è composta da tutti i colori, cioè tutte le frequenze (in realtà solo del visibile e non tutte con la stessa potenza)²⁵.

Notiamo che il rumore bianco, a parte altre caratteristiche “ideali”, è caratterizzato da potenza infinita. Ciò per la presenza di potenza a tutte le frequenze. Tutti i sistemi fisici sono limitati in frequenza, e quindi tutti i segnali fisici, che sono da essi generati, lo sono.

5.3.3 Processi stocastici normali e sistemi lineari

Ricordiamo che un processo stocastico normale è statisticamente definito se è dato il valor medio e l'autocorrelazione. O, in altre parole, se è dato lo spettro di potenza. Ciò significa che dal punto di vista statistico due processi stocastici normali che hanno lo stesso spettro sono statisticamente indistinguibili.

Ciò ci induce a modellare qualsiasi processo stocastico normale come l'uscita di un opportuno sistema lineare al cui ingresso sia stato posto rumore bianco. Ricordando la (5.49) e la (5.58), si vede che, se vogliamo modellizzare (o anche simulare) un processo stocastico normale con spettro $S(\omega)$, il sistema deve avere una funzione di trasferimento tale che

$$(5.59) \quad |F(j\omega)|^2 = S(\omega)$$

Ci sono infiniti sistemi che godono della proprietà (5.59): infatti questa relazione impone vincoli sul modulo della funzione di trasferimento, non sulla fase. Se osserviamo il processo in uscita, non abbiamo modo di capire quale sia la vera funzione di trasferimento usata. Non così

²⁵ Talvolta si parla di rumore “rosa” per indicare un rumore con spettro abbastanza piatto alle basse frequenze e praticamente nullo alle alte. Questo, a differenza del rumore bianco, è un rumore “possibile”.

se possiamo osservare anche il rumore bianco in ingresso, come possiamo renderci conto considerando la (5.48).

Abbiamo quindi trovato un modo di rappresentare e possibilmente simulare un processo normale con un sistema lineare.

Ma le proprietà dei processi stocastici normali e dei sistemi lineari ci permetteranno di sviluppare semplici soluzioni a una varietà di problemi, come per esempio quello della rivelazione dei segnali immersi nel rumore.

5.4 Processi discreti: processi ARMA

5.4.1 Résumé dei risultati

La teoria dei processi stocastici è stata fin qui sviluppata p.er lo più per i processi continui. Analoghi sviluppi possono farsi nel caso dei processi discreti. In particolare per quanto riguarda la teoria dei processi normali e i sistemi lineari, per i processi discreti si ottengono analoghi risultati considerando sistemi discreti. Sintetizziamo qui i risultati più imporanti.

Analogamente al caso continuo, possiamo definire lo spettro di potenza come la trasformata di Fourier per dati discreti (vedi (3.42) dell'autocorrelazione $R_{xx}(i)$)

$$(5.60) \quad S_x(\Omega) = \sum_{i=-\infty}^{\infty} R_{xx}(i) \cdot e^{-j \cdot i \cdot \Omega}$$

e lo spettro incrociato dalla correlazione incrociata $R_{xy}(i)$

$$(5.61) \quad S_{xy}(\Omega) = \sum_{i=-\infty}^{\infty} R_{xy}(i) \cdot e^{-j \cdot i \cdot \Omega}$$

Ovviamente vale la formula di inversione (3.43).

Se invece della pulsazione normalizzata Ω vogliamo usare quella fisica $\omega = \frac{\Omega}{\Delta t}$, la densità spettrale va opportunamente normalizzata

$$(5.62) \quad S'(\omega) = S(\Omega) \cdot \Delta t$$

Ciò garantisce la proprietà fondamentale che l'integrale sull'intera banda della pulsazione ($\frac{2\pi}{\Delta t}$ nel caso di ω , 2π nel caso di Ω) sia pari alla varianza del segnale moltiplicato 2π . Se l'integrale viene fatto nella frequenza invece che nella pulsazione, si ha semplicemente la varianza del segnale.

Consideriamo lo schema seguente:

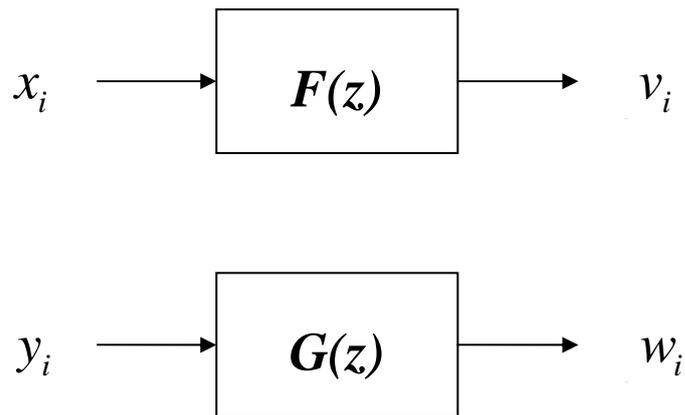


Figura 5-1

In esso x_i, y_i, v_i, w_i sono processi stocastici normali discreti.

Dalla (3.57) abbiamo

$$(5.63) \quad F(z) = \sum_{i=0}^{\infty} f_i \cdot z^{-i}$$

$$G(z) = \sum_{i=0}^{\infty} g_i \cdot z^{-i}$$

e quindi

$$(5.64) \quad v_i = \sum_{k=0}^{\infty} f_k \cdot x_{i-k}$$

$$w_i = \sum_{k=0}^{\infty} g_k \cdot y_{i-k}$$

Allora

$$(5.65) \quad R_{vx}(j) = E[v_{i+j} \cdot x_i^*] = E\left[\sum_{k=0}^{\infty} f_k \cdot x_{i-k+j} \cdot x_i^*\right] = \sum_{k=0}^{\infty} f_k \cdot R_{xx}(j)$$

$$(5.66) \quad R_{vv}(j) = E[v_{i+j} \cdot v_i^*] = E\left[\sum_{k=0}^{\infty} f_k \cdot x_{i-k+j} \cdot v_i^*\right] = \sum_{k=0}^{\infty} f_k \cdot R_{xv}(j)$$

e poiché $R_{xv}(j) = R_{vx}^*(-j)$,

.....

e inoltre

$$(5.67) \quad R_{vy}(j) = E[v_{i+j} \cdot y_i^*] = E\left[\sum_{k=0}^{\infty} f_k \cdot x_{i-k+j} \cdot y_i^*\right] = \sum_{k=0}^{\infty} f_k \cdot R_{xy}(j)$$

e

$$(5.68) \quad R_{vw}(j) =$$

Per passare agli spettri si usa la proprietà della convoluzione e si ha

$$(5.69) \quad S_{vy}(z) = F(z) \cdot S_{xy}(z)$$

$$(5.70) \quad S_{vw}(z) = G(z^{-1}) \cdot S_{vy}(z) = F(z) \cdot G(z^{-1}) \cdot S_{xy}(z)$$

che comprende come caso particolare (per $x_i = y_i$ e $F=G$)

$$(5.71) \quad S_{vx}(z) = F(z) \cdot S_{xx}(z)$$

$$(5.72) \quad S_{vw}(z) = F(z^{-1}) \cdot S_{vx}(z) = F(z) \cdot F(z^{-1}) \cdot S_{xx}(z) = |F(z)|^2 \cdot S_{xx}(z)$$

5.4.2 Rumore bianco discreto

Anche nel caso discreto possiamo definire il rumore gaussiano bianco.. Esso consiste in una successione di campioni distribuiti normalmente con valor medio nullo (e spesso si prende con varianza unitaria).

L'autocorrelazione è

$$(5.73) \quad R(j) = \sigma^2 \cdot \delta_j$$

e lo spettro, usando la (3.42), è

$$(5.74) \quad S(\Omega) = \sigma^2$$

Se vogliamo “lavorare” con le frequenze fisiche, allora

$$(5.75) \quad S'(\omega) = S(\Omega) \cdot \Delta t = \sigma^2 \cdot \Delta t$$

dove $0 \leq \Omega < 2\pi$, ovvero $-\pi \leq \omega < \pi$. Come si vede, lo spettro è piatto come nel caso del rumore bianco continuo, ma la banda del segnale è limitata, quindi, a differenza del rumore

bianco continuo, questo non crea paradossi, è fisicamente realizzabile: infatti non è altro che una successione di valori distribuiti gaussianamente con varianza σ^2 , indipendenti.

Si noti come la densità spettrale “fisica” si abbassi proporzionalmente al tempo di campionamento.

Ecco una realizzazione di rumore bianco discreto (1000 campioni)

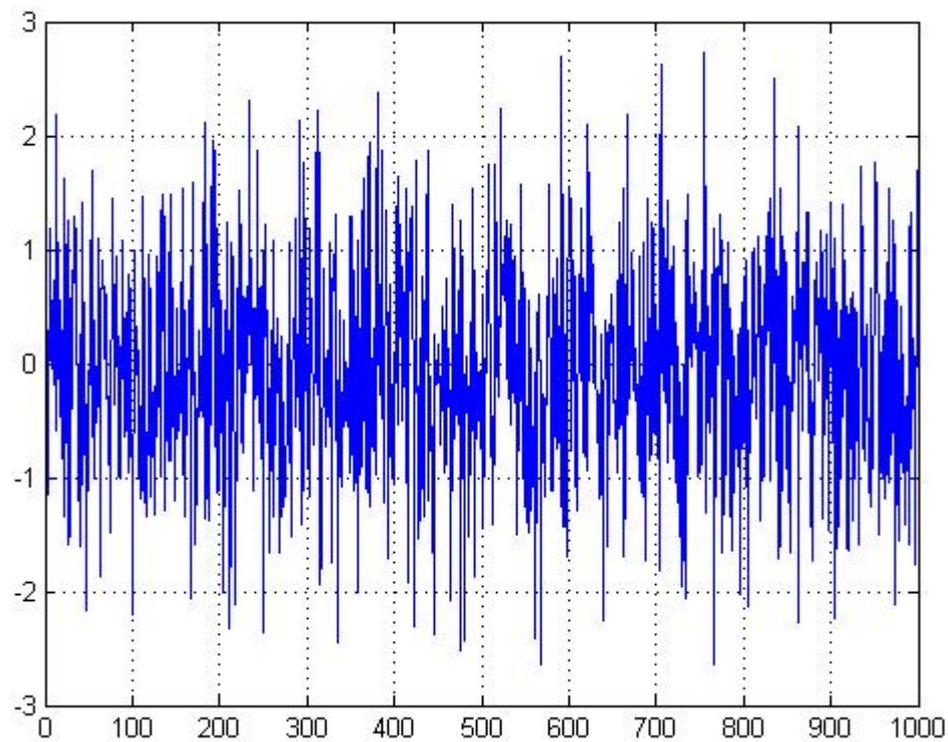


Figura 5-2

ed eccone l'istogramma

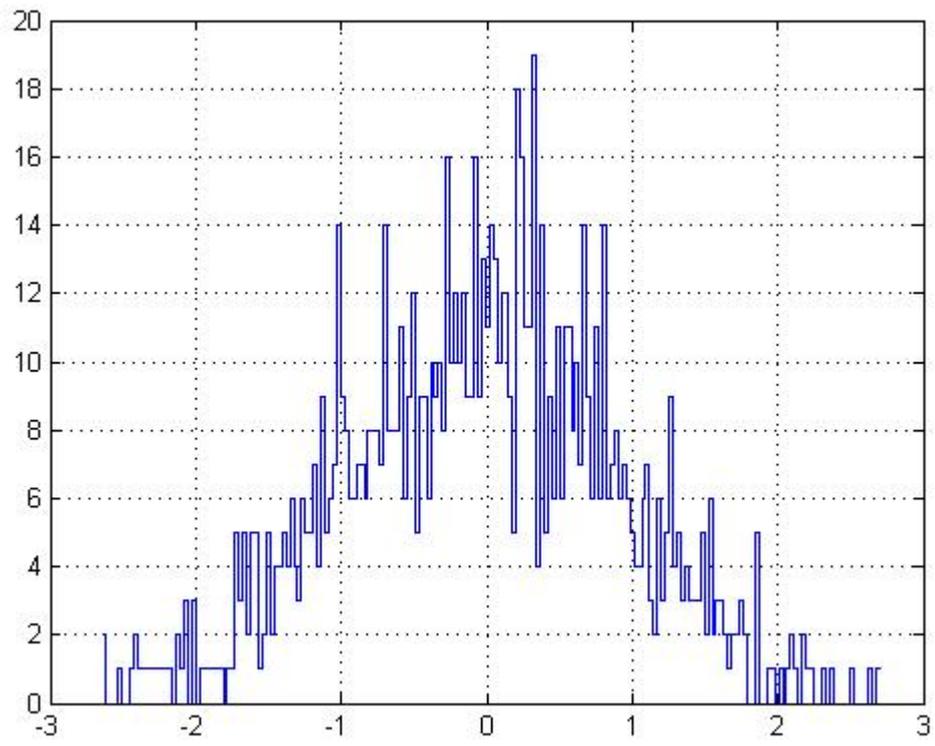


Figura 5-3

e la (stima della) autocorrelazione

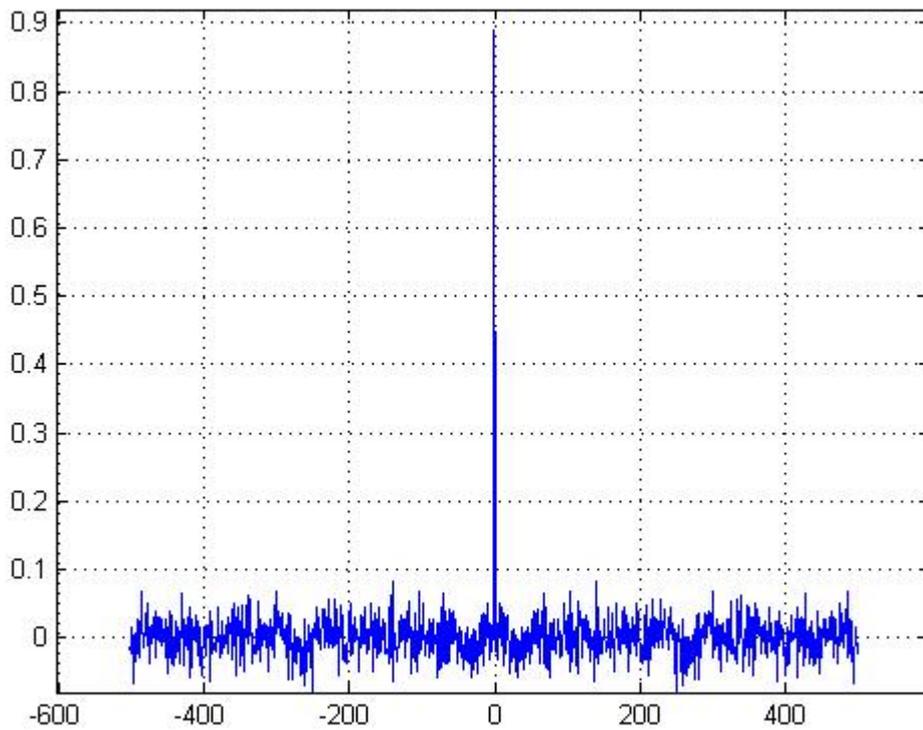


Figura 5-4

5.4.3 Processi AR, MA e ARMA

Se mettiamo in ingresso a un sistema lineare discreto, di funzione di trasferimento $F(z)$, del rumore bianco di varianza σ^2 , produciamo un processo stocastico normale con media nulla e spettro di potenza

$$(5.76) \quad S(\Omega) = \sigma^2 \cdot |F(e^{j\Omega})|^2$$

Se il sistema è un sistema AR o MA o ARMA, diremo che il processo è rispettivamente AR, MA o ARMA. Mostriamo ora alcuni casi.

5.4.4 Processo AR del primo ordine (reale)

Consideriamo il caso in cui $b_0=1$, $a_0=1$, $a_1=-0.99$, essendo 1 il tempo di campionamento. Ecco i grafici dell'autocorrelazione, dello spettro e 1000 campioni di una realizzazione

Autocorrelazione

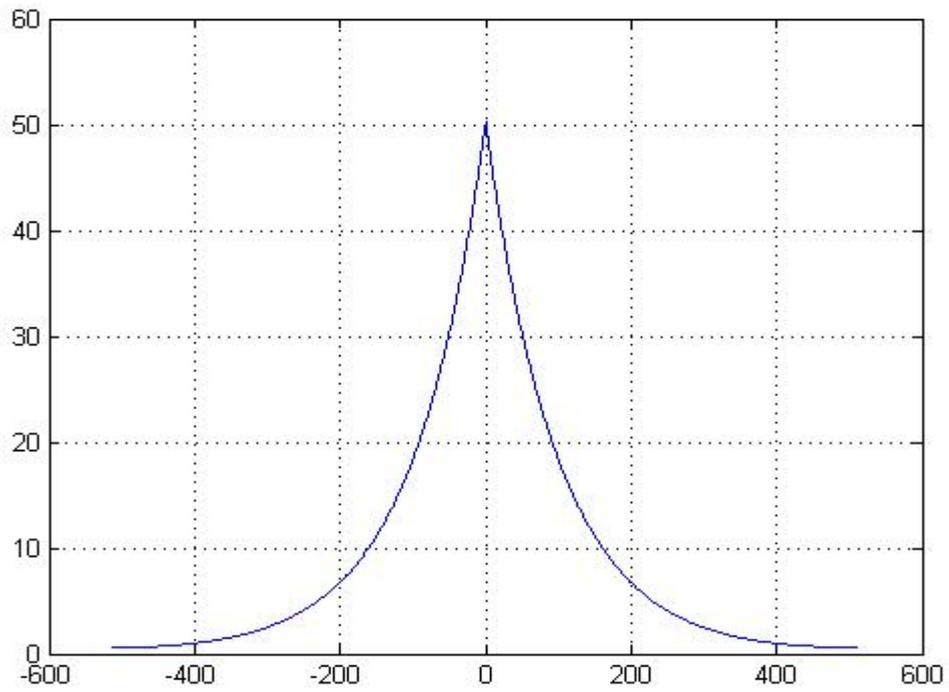


Figura 5-5

Spettro (in ascissa la frequenza)

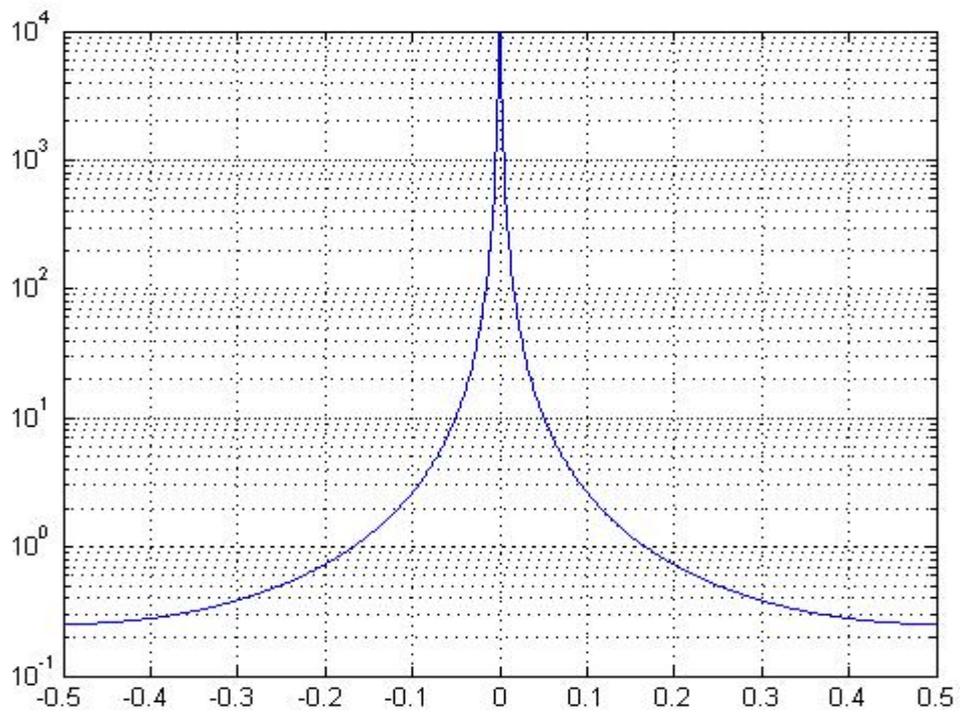


Figura 5-6

Un pezzo di una realizzazione

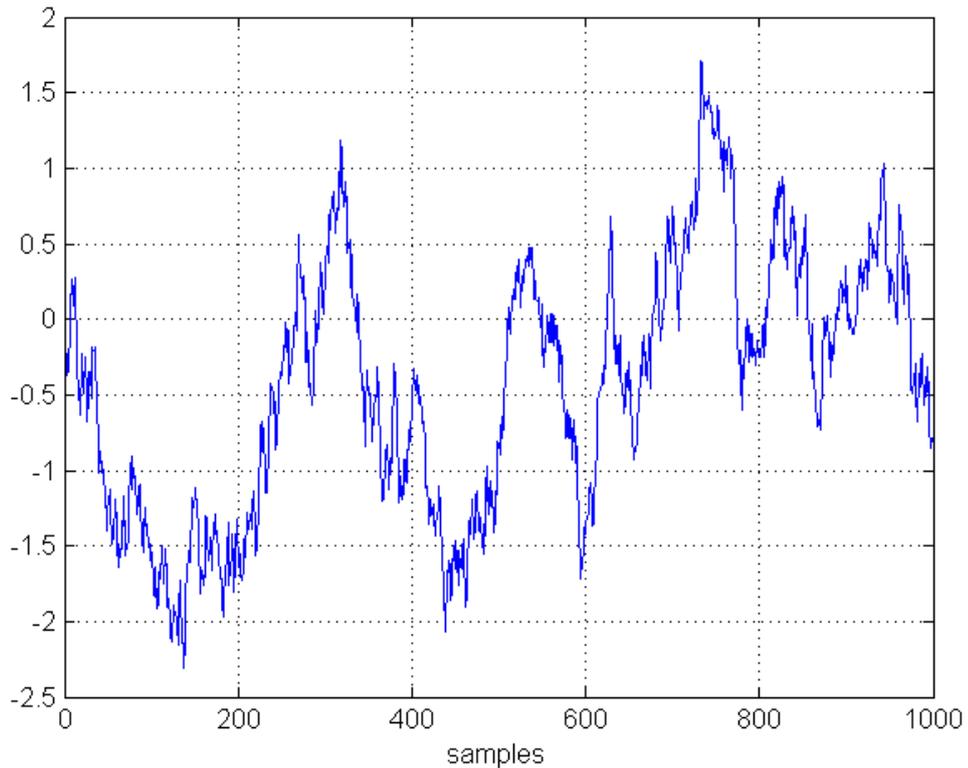


Figura 5-7

5.4.5 Processo AR del primo ordine (complesso)

Consideriamo il caso in cui

$$b_0 = 1, a_0 = 1, a_1 = 0.99 \cdot \exp(-j \cdot \pi / 20) = (-0.9778 + j \cdot 0.1549)$$

essendo 1 il tempo di campionamento. Ecco i grafici dell'autocorrelazione, dello spettro e 1000 campioni di una realizzazione.

Si noti che in questo esempio il tempo di decadimento è lo stesso dell'esempio precedente, ma

abbiamo una risonanza a frequenza negativa, a $\nu = -\frac{1}{40} = -0.025$.

Possiamo dire che questo processo è una rotazione oraria di $360/40=9$ gradi nel piano di Gauss del processo reale precedente.

Autocorrelazione (la parte immaginaria è punteggiata)

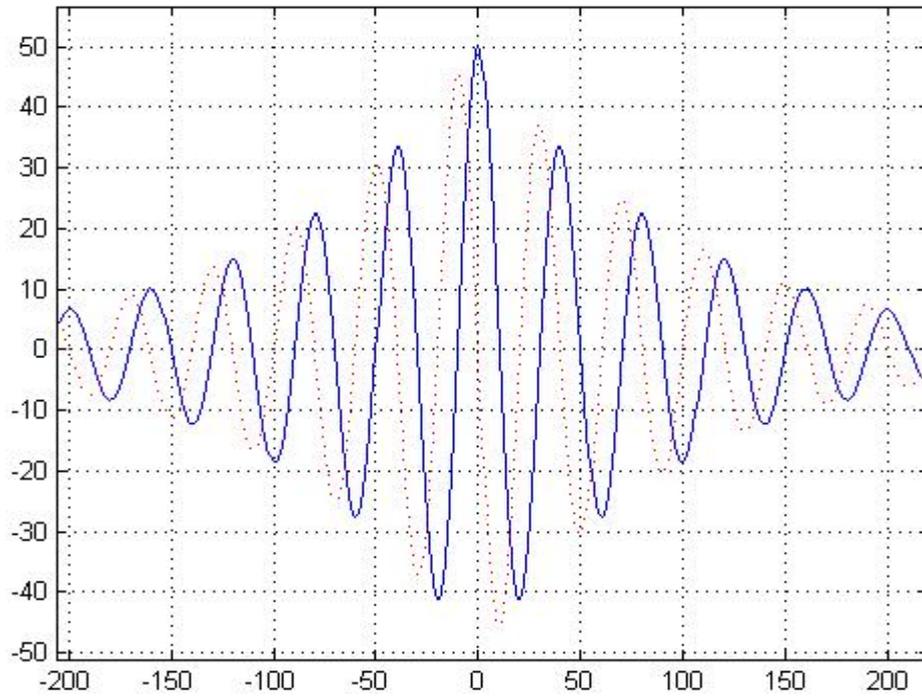


Figura 5-8

Spettro

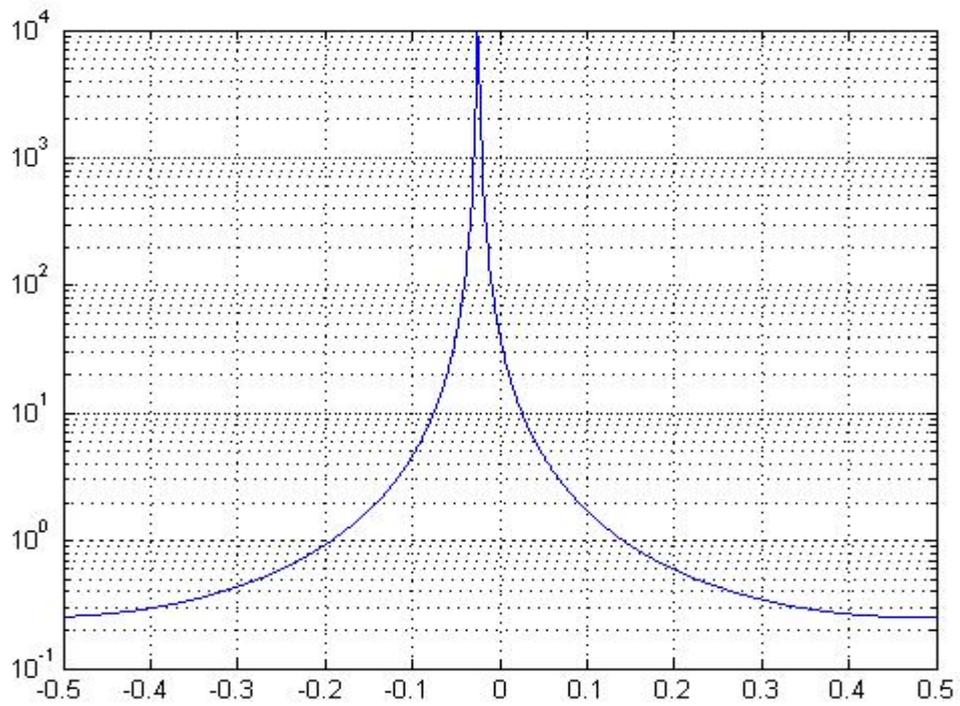


Figura 5-9

Un pezzo di una realizzazione (la parte immaginaria è punteggiata)

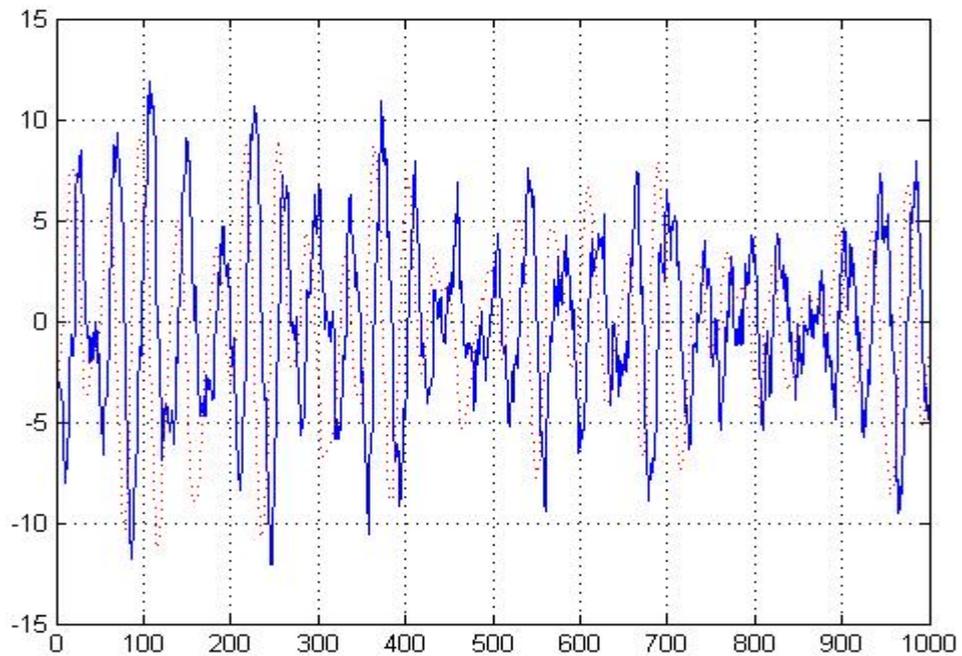


Figura 5-10

Parte immaginaria vs reale

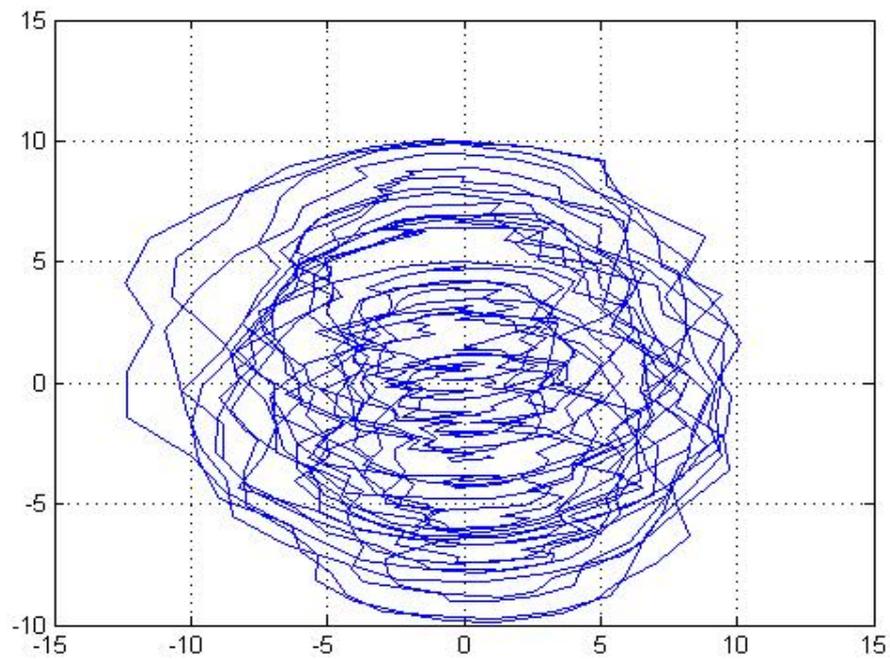


Figura 5-11

5.4.6 Processo AR del secondo ordine

Consideriamo il caso in cui

$$b_0 = 1, a_0 = 1, a_1 = -2 \cdot 0.99 \cdot \cos\left(\frac{\pi}{6}\right) = -1.7147, a_2 = r^2 = 0.9801$$

Abbiamo

Autocorrelazione

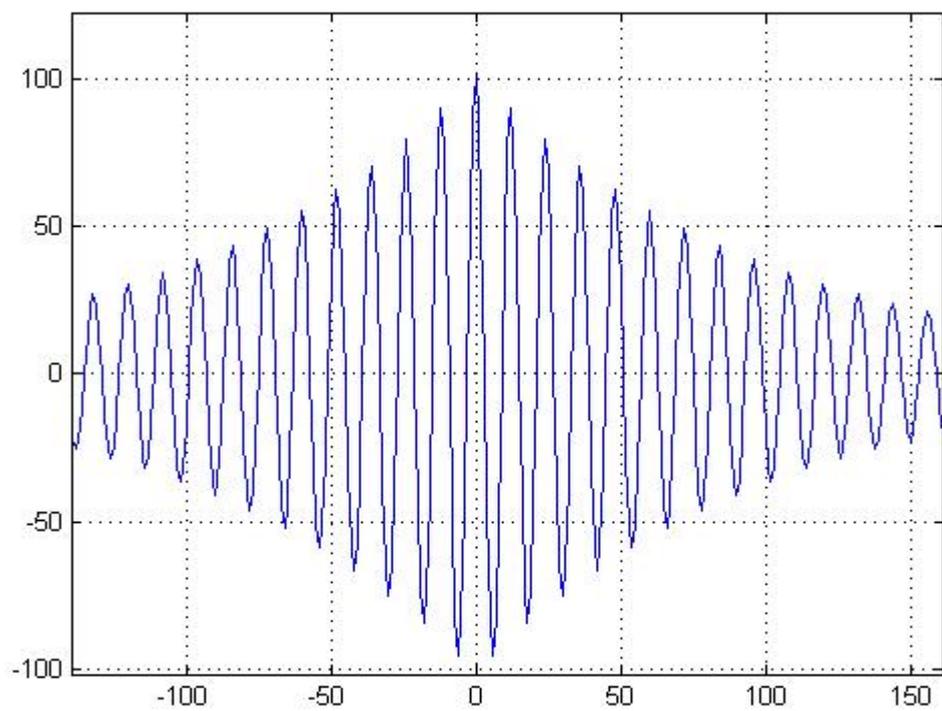


Figura 5-12

Spettro

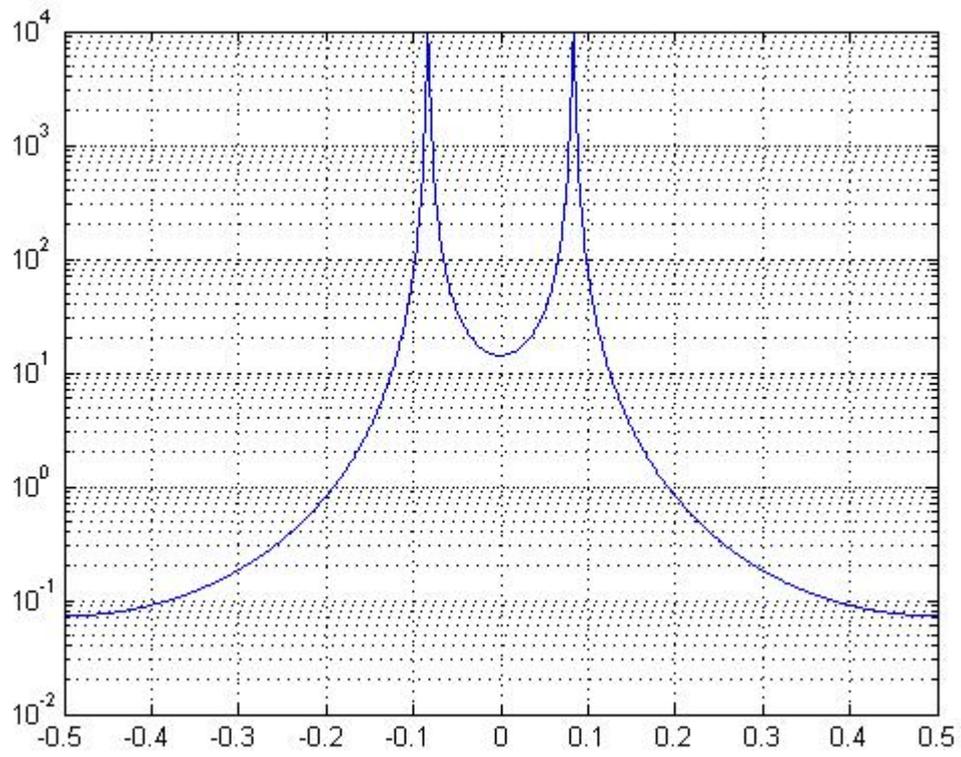


Figura 5-13

Pezzo di dati

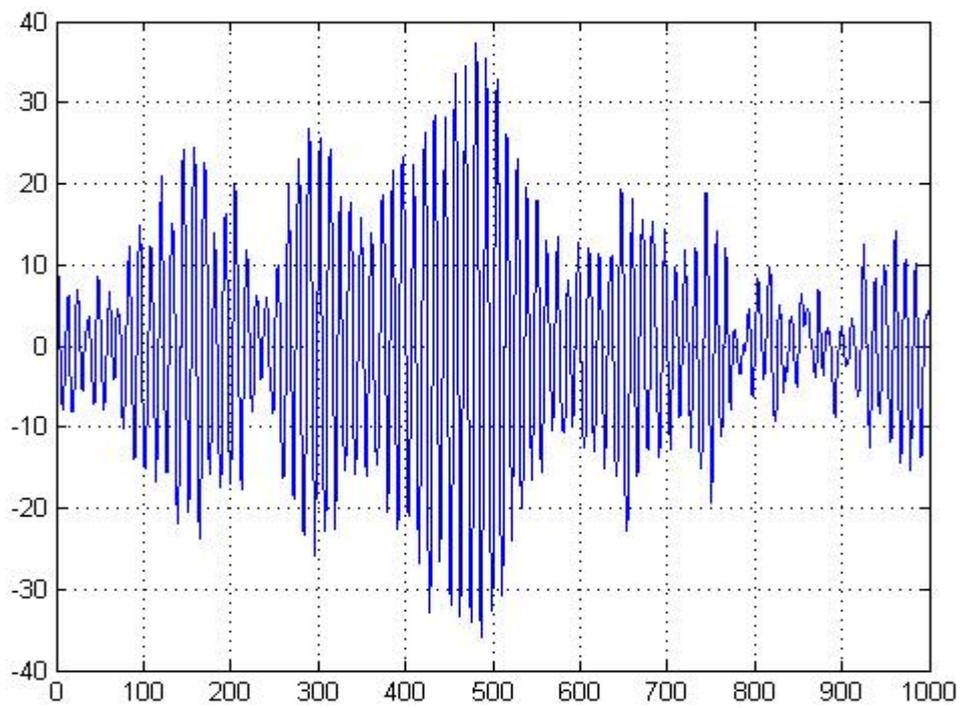


Figura 5-14

5.5 Processo di Poisson

Analizziamo un altro processo stocastico, di grande interesse non solo in Fisica, il processo di Poisson, che, come suggerisce il nome, è collegato alla distribuzione di Poisson.

Il problema è quello di modellare un processo fisico che consiste nella generazione di eventi, tutti eguali, distribuiti in modo uniforme nel tempo.

Un esempio è il decadimento radioattivo di un campione di materiale; gli eventi sono i singoli eventi rivelati da un contatore. In questo caso il tempo di osservazione deve essere molto più breve del tempo di decadimento dei nuclei del materiale, altrimenti gli eventi non sono distribuiti uniformemente nel tempo.

Un altro esempio di interesse fisico è il passaggio di una giunzione da parte di elettroni o lacune. Questo processo, in cui il numero di eventi è molto grande nell'unità di tempo²⁶, è la causa del rumore shot.

Sebbene si possa definire opportunamente anche nel discreto, il processo di Poisson è un processo stocastico continuo.

Le funzioni del tempo del processo possono essere la successione di delte di Dirac

$$(5.77) \quad x(t) = \sum_{i=-\infty}^{\infty} \delta(t-t_i)$$

dove con t_i si indicano gli istanti di occorrenza degli eventi. Spesso come funzioni del processo vengono prese le funzioni non decrescenti a valore intero che contano il numero di arrivi prima del tempo t : in pratica l'integrale della $x(t)$ di (5.77), da 0 a t .

Il processo ha un unico parametro, indicato con λ , che indica la densità degli eventi, cioè il numero atteso di eventi nell'unità di tempo. Nel periodo di tempo T si aspetta il numero di eventi

$$(5.78) \quad \mu = \lambda \cdot T$$

La distribuzione degli eventi in tale periodo è data dalla distribuzione di Poisson

$$(5.79) \quad P(k) = \frac{(\lambda \cdot T)^k}{k!} e^{-\lambda \cdot T}$$

e il numero di eventi in due intervalli disgiunti sono due variabili casuali indipendenti.

Si può calcolare che il tempo τ tra due eventi successivi ha densità di probabilità esponenziale

$$(5.80) \quad f(\tau) = \lambda \cdot u(\tau) \cdot e^{-\lambda \cdot \tau}$$

con valor medio

²⁶ Il passaggio della giunzione non è istantaneo, quindi, se il numero di eventi è elevato, dalla corrente risultante possono non essere "risolti" i singoli eventi.

$$(5.81) \quad E[\tau] = \frac{1}{\lambda}$$

Non sempre il processo di Poisson con λ costante è un buon modello per processi fisici a eventi (in cui la cosa importante è il tempo di occorrenza degli eventi, non altre loro caratteristiche). Ciò perché la densità non è uniforme nel tempo, ma può variare; il processo cioè è descritto non più dal parametro λ costante, ma dalla funzione del tempo $\lambda(t)$. In tal caso possiamo calcolare il numero aspettato di eventi nel periodo $t_1 < t < t_2$ come

$$(5.82) \quad \mu = \int_{t_1}^{t_2} \lambda(t) \cdot dt$$

e la distribuzione è

$$(5.83) \quad P(k) = \frac{\left(\int_{t_1}^{t_2} \lambda(t) \cdot dt \right)^k}{k!} e^{-\int_{t_1}^{t_2} \lambda(t) \cdot dt}$$

6 *Analisi statistica dei segnali*

Finora abbiamo sviluppato strumenti teorici per descrivere e manipolare segnali.

Passiamo ora alla parte applicativa, cioè, dato un segnale, come ottenere le informazioni di nostro interesse.

Supponiamo che i nostri dati siano schematizzabili come un **processo stocastico stazionario ed ergodico**.

Fare l'analisi significa ricavare informazioni su questo processo, cioè stimare i suoi parametri. Tali stime vengono eseguite a partire dalla singola realizzazione del processo, l'unica disponibile; l'ipotesi di ergodicità ci garantisce che ciò sia possibile.

Si ricavano essenzialmente informazioni sulla distribuzione del primo ordine e su quella del secondo (o meglio, sulle funzioni a questa collegate di autocorrelazione e spettro di potenza). Nel caso gaussiano ciò esaurisce la descrizione del processo; non nel caso generale.

Prima di far ciò, tuttavia, discuteremo dei problemi connessi alla discretizzazione dei segnali continui, procedura fondamentale per far uso dei calcolatori digitali.

Bisogna dire che alcune procedure di analisi e di elaborazione dei segnali possono essere eseguite con l'uso di dispositivi analogici, ma questa alternativa è diventata sempre meno appetibile con lo sviluppo dei sistemi di elaborazione digitale.

Compiti preliminari dell'analisi del segnale sono:

- determinazione della quantità dei dati da analizzare (ovviamente è ben diverso analizzare 100 numeri o 10^{10} , sia per le tecniche, sia per gli obiettivi)
- identificare la “dinamica” del segnale in gioco, cioè quale è il presumibile intervallo dei valori delle ampiezze che potrà prendere
- indagare preliminarmente sulla banda del segnale osservato
- indagare sulla presenza di trend (cioè di tendenze di fondo, per esempio a salire)
- indagare sulla stazionarietà del fenomeno che si sta osservando
- indagare sulla presenza di periodicità

Una volta che siano chiariti, anche se in modo non minuzioso, gli interrogativi di questa pre-analisi, si può procedere ad una corretta acquisizione dei dati e ad analisi più approfondite.

6.1 Il campionamento

I segnali prodotti dai processi fisici sono in genere continui. Per analizzarli ed elaborarli con i calcolatori digitali devono essere campionati. Come possiamo campionarli? Quali problemi comporta il campionamento?

Campionare significa prendere i valori di un segnale continuo a certi istanti temporali che costituiscono un insieme discreto. Come scegliamo questi istanti? Una scelta semplice è quella di prendere gli istanti di tempo equispaziati (**campionamento uniforme**). Anche così dobbiamo scegliere la spaziatura, cioè il tempo di campionamento. Per far ciò possiamo utilizzare il teorema del campionamento.

Ci sono però casi in cui è meglio usare un campionamento non uniforme. Per esempio

- per adattarlo alla variabilità del segnale
- perché non siamo nelle ipotesi del teorema del campionamento, né possiamo metterci
- perché una certa analisi è semplificata da un campionamento non uniforme.

6.1.1 Teorema del campionamento

Il problema è questo: se campioniamo un segnale discreto, ne cogliamo tutte le caratteristiche? In altre parole, possiamo, a partire dai campioni, ricostruire il segnale originale?

Il teorema del campionamento, formulato nel 1928 da Nyquist e provato nel 1949 da Shannon, è il seguente:

Se un segnale reale ha spettro di potenza nullo per frequenze superiori a una data ν_0 , esso può essere completamente ricostruito se è campionato ad una frequenza

$$(6.1) \quad \nu_s \geq 2 \cdot \nu_0$$

Data una frequenza di campionamento $\nu_s = \frac{1}{\Delta t_s}$, dove Δt_s è il tempo di campionamento, la

frequenza $\nu_N = \frac{\nu_s}{2}$ è detta **frequenza di Nyquist**.

Dati i campioni $\{x_i\}$, nell'ipotesi del teorema, possiamo ricavare il segnale originale come

$$(6.2) \quad x(t) = \sum_{i=-\infty}^{\infty} x_i \cdot \frac{\sin\left(\pi\left(\frac{2t}{\Delta t_s} - i\right)\right)}{\pi\left(\frac{2t}{\Delta t_s} - i\right)}$$

la funzione $\mathbf{sinc}(\alpha) = \frac{\sin(\pi \cdot \alpha)}{\pi \cdot \alpha}$ che compare nella (6.2), detta funzione interpolante, ha il seguente andamento

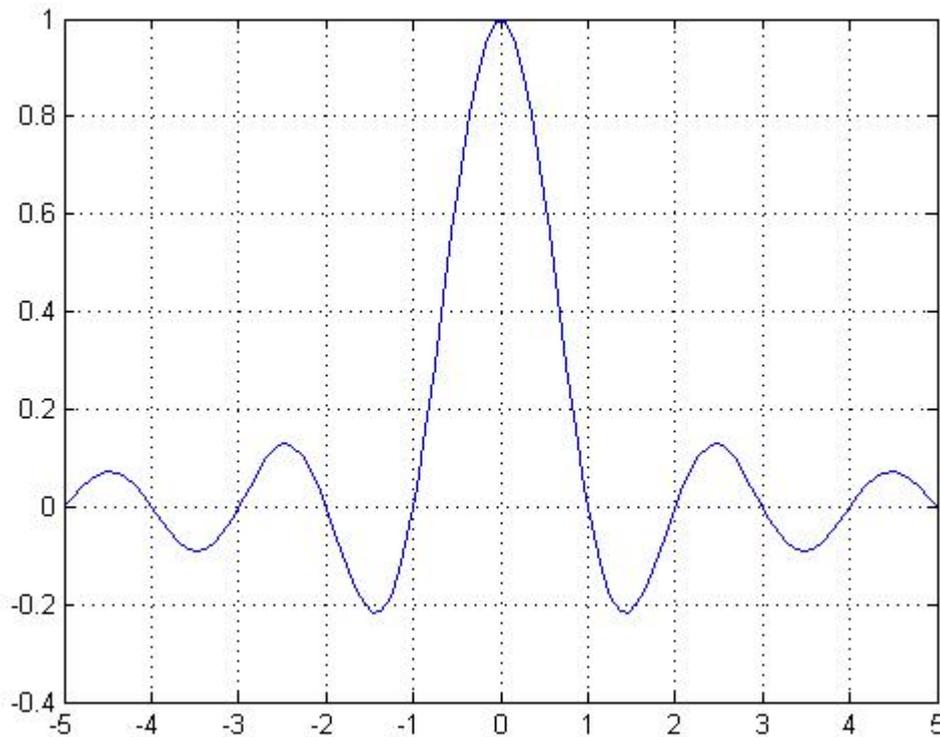


Figura 6-1

ed è la trasformata di Fourier della funzione

$$(6.3) \quad f(t) = \begin{cases} \frac{1}{2\pi} & \text{per } -\pi \leq t < \pi \\ 0 & \text{altrove} \end{cases}$$

Osservando la funzione d'interpolazione, notiamo che correttamente essa si annulla nei punti in cui ci sono altri campioni (infatti a quegli istanti la $x(t)$ è completamente definita dal campione), ma può stupire che in certi punti sia addirittura negativa: come mai? Sono delle correlazioni negative introdotte dal taglio in frequenza.

Notiamo che la (6.2) non è altro che l'uscita di un filtro passa-basso "perfetto", con funzione di trasferimento

$$(6.4) \quad F(j\omega) = \begin{cases} 1 & \text{per } -2\pi \cdot \nu_N \leq \omega < 2\pi \cdot \nu_N \\ 0 & \text{altrove} \end{cases}$$

al cui ingresso sia stato posto il segnale

$$(6.5) \quad \tilde{x}(t) = \sum_{i=-\infty}^{\infty} x_i \cdot \delta(t - i \cdot \Delta t_s)$$

6.1.2 Aliasing e filtri anti-aliasing

Che succede se si campiona con un tempo di campionamento che non rispetta la condizione posta dall'equazione (6.1) del teorema del campionamento ?

Ciò che succede è che le frequenze più elevate di $\nu_N = \frac{\nu_s}{2}$, frequenza di Nyquist, si “mascherano” da frequenze inferiori a ν_N . Questo fenomeno viene detto con termine inglese **aliasing**.

In figura sono mostrate due sinusoidi, una a frequenza $\nu = \frac{1}{5} \text{ Hz}$ (la frequenza di campionamento è 1) e una a $\nu' = \nu_s - \nu = \frac{4}{5} \text{ Hz}$.

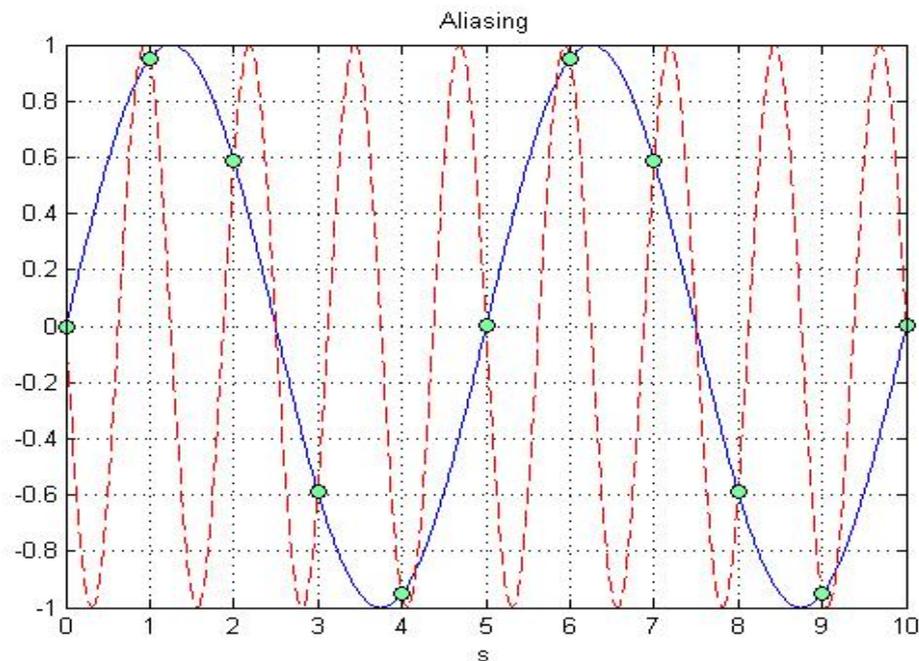


Figura 6-2

Si vede che le due sinusoidi condividono esattamente gli stessi campioni, anche se la sinusoida a più bassa frequenza ha 5 campioni a periodo, quella a più alta 1.25 campioni (meno dei 2 prescritti dal teorema del campionamento).

In genere una frequenza

$$(6.6) \quad \nu = N \cdot \nu_s + \nu_1 \quad \text{essendo} \quad \nu_1 < \nu_s$$

si “maschera” da ν_1 se $\nu_1 < \frac{\nu_s}{2}$, altrimenti si maschera da $\nu_s - \nu_1$, o, più correttamente, da $-\nu_1$.

Il fenomeno dell’aliasing è ben noto nella vita pratica. Per esempio è dovuto all’aliasing il fatto che a volte nei film le ruote delle carrozze sembrano girare al contrario (un film è infatti una successione di immagini campionate, in genere a 24, 25 o 30 Hz, se i raggi delle carrozze

hanno tra di loro un angolo α e tra un fotogramma e l’altro la ruota gira di un po’ più di $\frac{\alpha}{2}$,

questa appare girare al contrario. Accelerando, apparirà girare nel verso giusto, ma più lentamente, quindi di nuovo nell’altro verso e così via.

Ci sono usi pratici importanti dell’aliasing, per esempio lo stroboscopio.

Il fenomeno dell’aliasing è in genere dannoso nell’analisi dei segnali perché aumenta il livello del rumore. Va quindi eliminato o per lo meno limitato, e ciò si fa con opportuni filtri passa basso, **analogici**, prima del campionamento, in modo da porsi nelle condizioni del teorema del campionamento (nessuna componente a frequenze superiori alla frequenza di Nyquist).

Talvolta, data la molto maggiore facilità di realizzazione dei filtri digitali, se si vuole un campionamento a una frequenza ν_s , si usa un “rozzo” filtro anti-aliasing analogico, una frequenza di campionamento molto più elevata di quella voluta, quindi un filtro anti-aliasing digitale “perfetto” che limita il segnale nella banda di Nyquist voluta. Si procede quindi a un sottocampionamento alla frequenza voluta.

Infine vogliamo ricordare che talvolta l’aliasing nell’analisi dei segnali può essere usato utilmente, per esempio per abbassare la frequenza di campionamento (vedi il prossimo paragrafo).

6.1.3 Generalizzazione

Se il segnale ha lo spettro nullo al di fuori di una certa banda B (che non necessariamente inizia da 0), possiamo teoricamente generalizzare il teorema del campionamento, limitando la frequenza di campionamento a $2 \cdot B$. Ben più difficile è però la procedura di ricostruzione.

Dato una certa frequenza di campionamento, abbiamo visto che l’asse delle frequenze si suddivide in bande “aliasate”: se la banda del segnale entra tutta in una di queste bande aliasate, allora la ricostruzione è semplificata. Ciò però può significare dover scegliere una frequenza di campionamento abbastanza maggiore di $2 \cdot B$.

6.1.4 Il caso dei segnali complessi

Se il segnale è complesso, c'è differenza tra le frequenze positive e negative, quindi la banda del segnale da considerare è doppia. Tuttavia il numero di campioni necessari è la metà, perché ciascun campione è “doppio” (parte reale e parte immaginaria). Detto in altre parole, abbiamo il doppio di campioni reali, ma descriviamo una banda doppia: la simmetria rispetto alla frequenza di Nyquist non c'è.

6.1.5 Rumore di quantizzazione

Quando si digitalizza un segnale continuo, oltre ai problemi relativi al campionamento, si devono considerare i problemi relativi alla quantizzazione. Il segnale campionato infatti viene trasformato in un numero intero (e memorizzato come tale) tramite l'uso di un convertitore analogico-digitale (ADC). Questo numero è il numero che, moltiplicato per una quantità Δx chiamata **quanto di conversione**, meglio approssima il valore del campione (che è un numero reale). Se il campione vale x , ricaviamo un valore n tale che

$$(6.7) \quad -\frac{\Delta x}{2} < x - n \cdot \Delta x \leq \frac{\Delta x}{2}$$

Al valore vero x del campione assegneremo il valore “quantizzato” $\tilde{x} = n \cdot \Delta x = x + \varepsilon$.

Possiamo schematizzare questo processo come l'aggiunta di un rumore ε , detto **rumore di quantizzazione**.

Se il segnale $\{x_i\}$ è schematizzabile con un rumore bianco (cioè i campioni sono indipendenti) i campioni $\{\varepsilon_i\}$ del rumore di quantizzazione sono indipendenti e distribuiti uniformemente tra $-\frac{\Delta x}{2}$ e $\frac{\Delta x}{2}$. La varianza è (vedi eq. (2.90))

$$(6.8) \quad \sigma_\varepsilon^2 = \frac{(\Delta x)^2}{12}$$

e lo spettro di potenza è (vedi eq. (5.74) e (5.75))

$$(6.9) \quad S(\omega) = \frac{(\Delta x)^2}{12} \cdot \Delta t$$

costante, indipendente da ω .

Si noti che la densità spettrale decresce con l'aumentare della frequenza di campionamento. Ciò può essere utilizzato per ridurre l'errore di quantizzazione: vediamo come.

La “classe” di un ADC è data dal numero di bit N usati per la conversione, in genere 8 o 12 o 16. Quanto maggiore è il numero di bit, tanto migliore è il convertitore: essendo M il massimo range coperto dal convertitore, il quanto di conversione è

$$(6.10) \quad \Delta x = \frac{M}{2^N}$$

L'idea è quella di campionare i dati analogici a una frequenza ν_s' molto più elevata di quella richiesta ν_s , applicare un buon filtro anti-aliasing (in software) e sotto-campionare a ν_s : i campioni ottenuti ora sono numeri "reali" (non più "interi"), perché l'uscita del filtro è composta da numeri floating point (a 32 bit, che equivale a circa 24 bit di mantissa, o più se si lavora in doppia precisione) e avranno un errore di quantizzazione equivalente con spettro $\frac{\nu_s'}{\nu_s}$ volte più basso e quindi è come se avessimo

$$(6.11) \quad \frac{1}{2} \log_2 \left(\frac{\nu_s'}{\nu_s} \right)$$

bit di conversione in più.

6.2 Caratteristiche statiche – Istogramma, momenti campionari

Una stima della densità del primo ordine (vedi (5.2), che nel caso stazionario si riduce a $f(x)$) può ottenersi facendo un istogramma dei dati, normalizzato per il numero totale dei dati utilizzati.

Per fare questo istogramma occorre scegliere bene i bin, che non necessariamente saranno tutti uguali. Se si usano pochi bin, si hanno informazioni un po' "rozze" sulla $f(x)$, se se ne usano troppi, pochi dati andranno nei singoli bin e quindi si avranno grosse fluttuazioni percentuali e quindi errori.

Notare che solo raramente i dati istogrammati saranno indipendenti: in questo caso il numero di dati che entrano in un dato bin è distribuito approssimativamente secondo una distribuzione di Poisson con parametro $\mu = N \cdot p$, essendo p la probabilità che un valore vada nel bin i oggetto ed N il numero totale dei dati. Ricordiamo che la deviazione standard è $\sigma = \sqrt{\mu}$. In genere i dati sono correlati e la situazione è peggiore.

Un altro modo per avere informazione sulla $f(x)$ è stimarne il valor medio (per esempio con la media aritmetica), la varianza ed eventualmente i momenti superiori. Anche in questo caso la convergenza statistica verso il valore vero è rallentata dalla presenza di correlazioni. Per esempio, per il caso del valor medio, supponiamo che il segnale abbia una autocovarianza $C(k)$ e vogliamo stimarne il valor medio facendo la media di n campioni successivi

$$(6.12) \quad \hat{\mu} = \frac{1}{N} \sum_{i=1}^N x_i$$

e la varianza su questa stima è

$$(6.13) \quad \sigma_{\hat{\mu}}^2 = \frac{1}{N} \sum_{k=-N}^N C(k) \cdot \left(1 - \frac{|k|}{N}\right)$$

Nel caso in cui

$$(6.14) \quad C(k) = \sigma^2 \cdot e^{-\frac{|k|}{\tau}}$$

abbiamo approssimativamente (per $\tau > 1$)

$$(6.15) \quad \sigma_{\hat{\mu}}^2 \approx \frac{2 \cdot \tau \cdot \sigma^2}{N} \cdot \left(1 - \frac{1 - e^{-\frac{N}{\tau}}}{\frac{N}{\tau}}\right) = \frac{2 \cdot \sigma^2}{N'} \cdot \left(1 - \frac{1 - e^{-N'}}{N'}\right)$$

dove $N' = \frac{N}{\tau}$ è la "lunghezza equivalente" del pezzo di dati. Se N' è grande, si ha circa

$$(6.16) \quad \sigma_{\hat{\mu}}^2 \approx \frac{2 \cdot \tau \cdot \sigma^2}{N} \cdot \left(1 - \frac{1 - e^{-\frac{N}{\tau}}}{\frac{N}{\tau}} \right) = \frac{2 \cdot \sigma^2}{N'} \cdot \left(1 - \frac{1 - e^{-N'}}{N'} \right)$$

6.3 Autocorrelazione

La definizione di autocorrelazione di un processo stocastico è data dalla (5.12), che nell'ipotesi stazionaria diventa

$$(6.17) \quad R(\tau) = E[X(t+\tau) \cdot X^*(t)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 x_2^* \cdot f(x_1, x_2; \tau) \cdot dx_1 \cdot dx_2$$

Data la commutatività della moltiplicazione, la parte reale è una funzione pari, la parte immaginaria (se presente) è dispari.

Inoltre gode delle seguenti proprietà:

- $R(0)$ è reale ed è il massimo assoluto; possono esserci altri valori di τ in cui $R(\tau) = R(0)$, ma solo se la X è periodica e τ è un multiplo del periodo
- $R(\tau) = R^*(-\tau)$, quindi la sua trasformata di Fourier è reale
- La sua trasformata di Fourier è non-negativa. Questa proprietà deriva dalla proprietà di essere **definita positiva**, cioè che, comunque presi n numeri arbitrari a_1, a_2, \dots, a_n ed n numeri reali $\tau_1, \tau_2, \dots, \tau_n$, si deve avere

$$(6.18) \quad \sum_{i=1}^n \sum_{l=1}^n R(\tau_i - \tau_l) \cdot a_i \cdot a_l \geq 0$$

Nel caso discreto la situazione è analoga, ma sostituiamo la variabile continua τ con la discreta k e la densità di probabilità con la distribuzione:

$$(6.19) \quad R(k) = E[X(i+k) \cdot X^*(i)] = \sum_{i_1=-\infty}^{\infty} \sum_{i_2=-\infty}^{\infty} x_{i_1} x_{i_2}^* \cdot f(x_{i_1}, x_{i_2}; k)$$

Data l'ipotesi di ergodicità, possiamo stimare la $R(k)$ sostituendo il valore atteso della (6.17) con la media su un pezzo di dati $\{x_1, x_2, \dots, x_N\}$:

$$(6.20) \quad \hat{R}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} x_{i+k} \cdot x_i^*$$

Una diversa stima è

$$(6.21) \quad \hat{R}(k) = \frac{1}{N} \sum_{i=1}^{N-k} x_{i+k} \cdot x_i^*$$

Come si vede, la differenza è nel coefficiente di normalizzazione. Il primo stimatore è non-distorto (unbiased), ma sui valori di k vicini a N ci sono forti incertezze (fluttuazioni).

La seconda stima è chiaramente distorta (il valore aspettato di $\hat{R}(k)$ è inferiore al valor vero di un fattore $\frac{N}{N-k}$, ma così sono ridotte anche le fluttuazioni), ma la (6.21) gode della proprietà b).

Spesso nell'analisi dei segnali si usa la tecnica di moltiplicare i dati per una "finestra", cioè dei pesi opportuni, in genere minori alle estremità. Questa operazione viene detta "finestratura" (windowing).

La (6.21) si ricava dalla (6.20) tramite una "finestra" a triangolo isoscele detta "finestra di Bartlett".

Le procedure (6.20) e (6.21) sono usate raramente, a parte il caso in cui siamo interessati solo a un piccolo sottoinsieme dei ritardi k . Ciò perché il calcolo è molto pesante (abbiamo $N^2/2$ moltiplicazioni ed addizioni). Esse (o meglio la (6.21)) vengono sostituite dalla seguente procedura che si avvale della FFT (Fast Fourier Transform, vedi paragrafo 3.4.2).

Consideriamo la successione infinita $\{x_i\}$ composta da tutti 0, a parte i valori per i da 1 a N , pari ai nostri campioni. La stima (6.21) è ottenuta tramite la

$$(6.22) \quad \hat{R}(k) = \frac{1}{N} \sum_{i=-\infty}^{\infty} x_{i+k} \cdot x_i^*$$

Ricordiamo che la convoluzione tra le successioni a e b (eq. (3.4)) è

$$(6.23) \quad y_k = \sum_{i=-\infty}^{\infty} a_{k-i} \cdot b_i$$

e si vede quindi che la (6.22) è la convoluzione tra la x invertita temporalmente e la coniugata della x . Ne segue che, essendo $X(\Omega)$ la trasformata di Fourier della successione x ,

$$(6.24) \quad \hat{R}(k) = \frac{1}{N} \cdot \mathcal{F}^{-1}(|X(\omega)|^2)$$

dove con \mathcal{F}^{-1} indichiamo la trasformata inversa di Fourier. Per eseguire questa procedura ci si avvale della trasformata di Fourier (per i dettagli, si veda al capitolo sui filtri in frequenza).

6.4 Spettro di potenza

Lo spettro di potenza è definito come la trasformata di Fourier dell'autocorrelazione (vedi eq. (5.17)). Nel caso stazionario si ha

$$(6.25) \quad S(\omega) = \int_{-\infty}^{\infty} R(\tau) \cdot e^{-j\omega\tau} d\tau$$

e, date le proprietà dell'autocorrelazione, è reale e non-negativo. Se inoltre il processo è reale, lo spettro è una funzione pari.

Si noti che le proprietà dello spettro e dell'autocorrelazione sono analoghe a quelle della densità di probabilità e della funzione caratteristica.

Mentre l'autocorrelazione deve avere le proprietà ("complicate") indicate in 6.3, lo spettro ha l'unica limitazione di essere positivo (o meglio non negativo). In genere comunque data una funzione non negativa di ω , a singolo valore, possiamo costruire un processo stocastico che la abbia come spettro di potenza.

Per i processi discreti, abbiamo (eq. (5.60))

$$(6.26) \quad S(\Omega) = \sum_{i=-\infty}^{\infty} R(i) \cdot e^{-j\cdot i \cdot \Omega}$$

essendo Ω la pulsazione normalizzata; ricordiamo che per avere lo spettro nelle unità "fisiche", $\Omega = \omega \cdot \Delta t$, e

$$(6.27) \quad S'(\omega) = S(\Omega) \cdot \Delta t$$

Se il segnale è reale, la potenza del segnale nella banda di frequenza $\{\nu_1, \nu_2\}$ è data da

$$(6.28) \quad 2 \cdot \int_{\nu_1}^{\nu_2} S'(2\pi \cdot \nu) d\nu$$

Il fattore 2 deriva dal fatto che per segnali reali non c'è differenza tra frequenze positive e negative.

Se il segnale è complesso, $S'(\omega)$ non è pari e bisogna distinguere tra frequenze positive e

negative e quindi $\int_{\nu_1}^{\nu_2} S'(2\pi \cdot \nu) d\nu$ è diverso da $\int_{-\nu_2}^{-\nu_1} S'(2\pi \cdot \nu) d\nu$. Perciò la potenza del segnale nella banda di frequenza $\{\nu_1, \nu_2\}$ è data da

$$(6.29) \quad \int_{\nu_1}^{\nu_2} S'(2\pi \cdot \nu) d\nu$$

Vogliamo notare che spesso nella fisica sperimentale e nella tecnica, quando i segnali sono sempre reali e non interessano molto le questioni teoriche, si usa una diversa definizione di spettro, con le sole frequenze positive e con un valore doppio di quello ottenuto da (5.60). Questo viene chiamato **spettro unilatero**. Scompare così lo “scomodo” fattore 2 della (6.28).

Vediamo ora come stimare lo spettro di potenza.

6.4.1 Stimatori spettrali non parametrici

Un modo per stimare lo spettro è quello di stimare l'autocorrelazione con lo stimatore (6.21), cioè lo stimatore non distorto con la finestra di Bartlett applicata e farne la trasformata di Fourier.

(Se si usasse la stima (6.20), avremmo una stima spettrale più “scomoda”, con errori più grandi e, per certe frequenze, con valori negativi. Ciò è dovuto al grosso peso che hanno le fluttuazioni sui valori estremali).

Questa procedura tuttavia non è usata quasi mai. essa infatti è computazionalmente molto pesante. Lo stesso risultato lo si ottiene usando la procedura usata per il calcolo veloce dell'autocorrelazione (escludendo la trasformata inversa). La ripetiamo:

Siano dati N campioni successivi $\{x_1, x_2, \dots, x_N\}$. Se ne faccia la trasformata di Fourier (per dati discreti)

$$(6.30) \quad X(\Omega) = \sum_{i=1}^N x_i \cdot e^{-j(i-1)\Omega}$$

La stima spettrale è data da

$$(6.31) \quad \hat{S}(\Omega) = \frac{|X(\Omega)|^2}{N}$$

È questa una conseguenza del teorema di Wiener-Kinchin²⁷. La trasformata (6.30) viene eseguita in modo veloce con la FFT.

Se si vogliono usare le frequenze “fisiche”, con tempo di campionamento Δt , ricordiamo che $\Omega = \omega \cdot \Delta t$ e

$$(6.32) \quad \hat{S}(\omega) = \frac{|X(\omega \cdot \Delta t)|^2}{N} \cdot \Delta t$$

Lo spettro (e la sua stima) è una grandezza statistica: da esso ovviamente non possono ricavarsi i dati originari e ciò perché si è persa l'informazione di fase delle componenti armoniche.

Vediamo tre problemi dello stimatore (6.31):

²⁷ In effetti in passato la (6.31) definiva lo spettro di potenza e il teorema di Wiener-Kinchin (noto già, precedentemente, ad Einstein in una forma un po' diversa) stabiliva che la stessa informazione si aveva facendo la trasformata di Fourier dell'autocorrelazione.

- **La risoluzione spettrale**

Per come è costruita la stima spettrale, notiamo che la risoluzione in frequenza della stima spettrale, dipende dalla lunghezza del pezzo di dati. Se abbiamo N campioni, essa è, convenzionalmente, in unità della frequenza fisica,

$$(6.33) \quad \Delta \nu = \frac{1}{N \cdot \Delta t}$$

Discuteremo meglio questo argomento quando avremo introdotto il concetto di finestra.

- **Problema del finestramento**

La stima è fatta su un pezzo di dati non infinito (come dovrebbe essere in teoria), ma di lunghezza finita. Questo può essere visto come il prodotto della successione infinita per una finestra “rettangolare” (boxcar in Inglese)²⁸

$$(6.34) \quad w_i = \begin{cases} 1 & \text{per } 1 \leq i \leq N \\ 0 & \text{altrove} \end{cases}$$

Ora sappiamo che la trasformata di Fourier del prodotto di due funzioni è pari alla convoluzione delle trasformate delle due funzioni, quindi la stima spettrale data dalla (6.31) “smussa” lo spettro vero: cioè è come se ci si passasse sopra un passa basso simile a quello di equazione (3.86) (in effetti la risposta in fase è diversa).

Questo smussamento può essere particolarmente fastidioso (se si vuole stimare spettri con rapide variazioni). Si riduce il problema usando una finestra non rettangolare.

Lo stimatore è quindi così costruito:

$$(6.35) \quad \hat{S}(\Omega) = \frac{\left| \sum_{i=1}^N w_i \cdot x_i \cdot e^{-j(i-1)\Omega} \right|^2}{\sum_{i=1}^N |w_i|^2}$$

La successione $\{w_i\}$ è la “finestra”. In genere vale 1 (il valore massimo) per i valori centrali, è simmetrica e si riduce agli estremi.

Una finestra molto usata è la finestra che prende il nome da Von Hann²⁹ (detta anche *banning window*) è la seguente

²⁸ La presenza di questa finestra sui dati temporali implica la finestra di Bartlett sull'autocorrelazione: l'autocorrelazione di una funzione rettangolare è una funzione triangolare isoscele.

²⁹ Julius Von Hann (1839-1921), meteorologo austriaco. Blackmann e Tukey hanno dato il suo nome a questa finestra.

(6.36)
$$w_i = \frac{1 + \cos\left(\frac{2\pi \cdot (i - N/2)}{N}\right)}{2}$$

Ecco la finestra di Von Hann normalizzata per $\sum_{i=1}^N |w_i|^2$ (100000 punti):

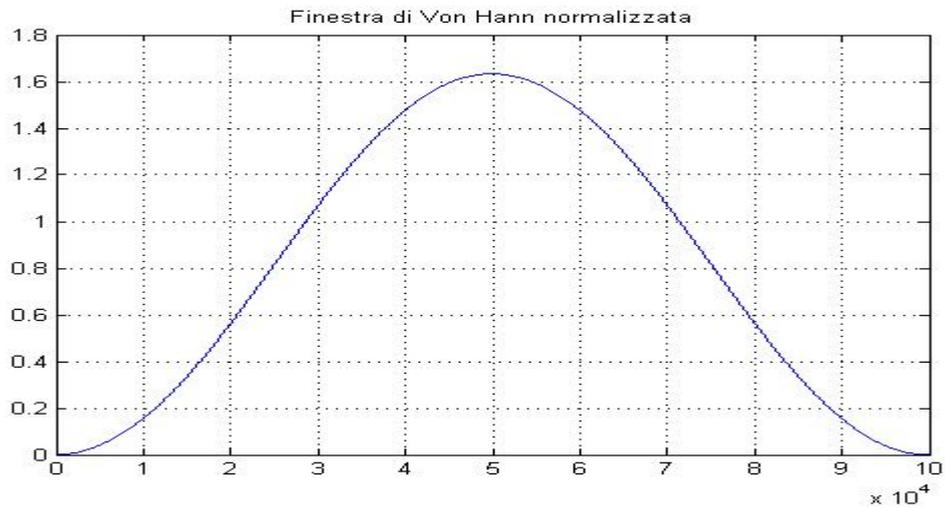


Figura 6-3

Vediamo ora l'effetto della finestra. Nel grafico sono riportate le stime spettrali per il caso di una sinusoidale di frequenza 0.1, eseguite con la finestra di hanning e con la finestra rettangolare (o "boxcar", cioè usando semplicemente la (6.31)):

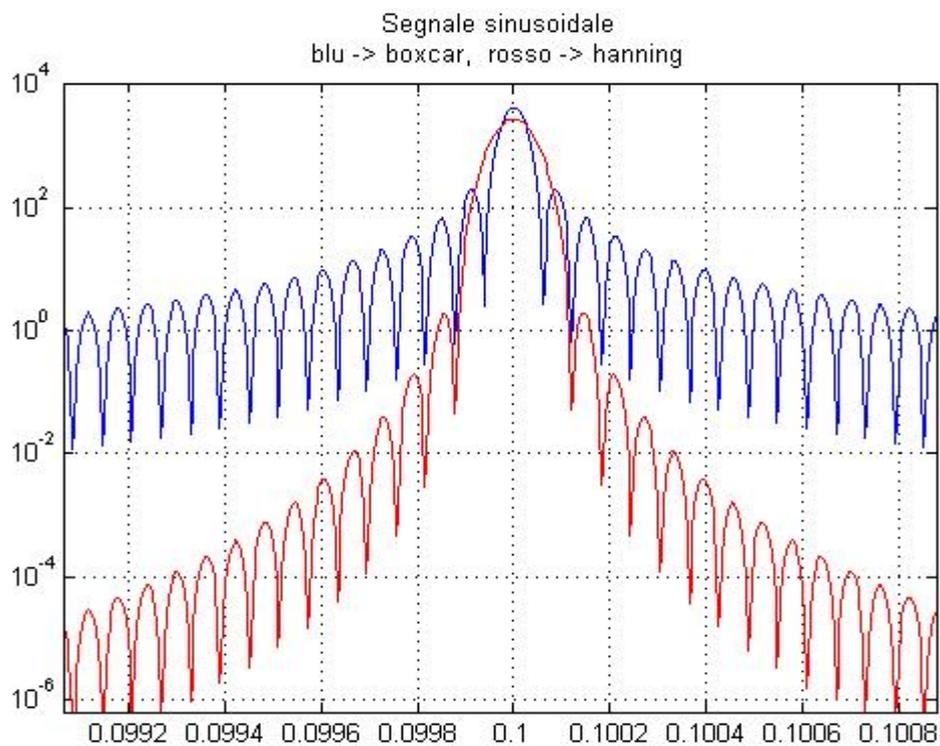


Figura 6-4

Lo spettro “ideale” del segnale sarebbe una delta a frequenza 0.1. In effetti invece vediamo un picco con dei lobi laterali (artefatti causati dalla finestrazione dei dati).

Notiamo che l’aver introdotto la finestra di Von Hann ha ridotto la risoluzione (di circa un fattore 2; ma questo in genere non è un problema, infatti se occorre una migliore risoluzione basta prendere un pezzo più lungo di dati), ma ha ridotto di parecchio l’ampiezza dei lobi laterali.

Nella figura seguente rappresentiamo le stime spettrali con le due finestre di un segnale composto di due sinusoidi (una 100 volte più piccola dell’altra)

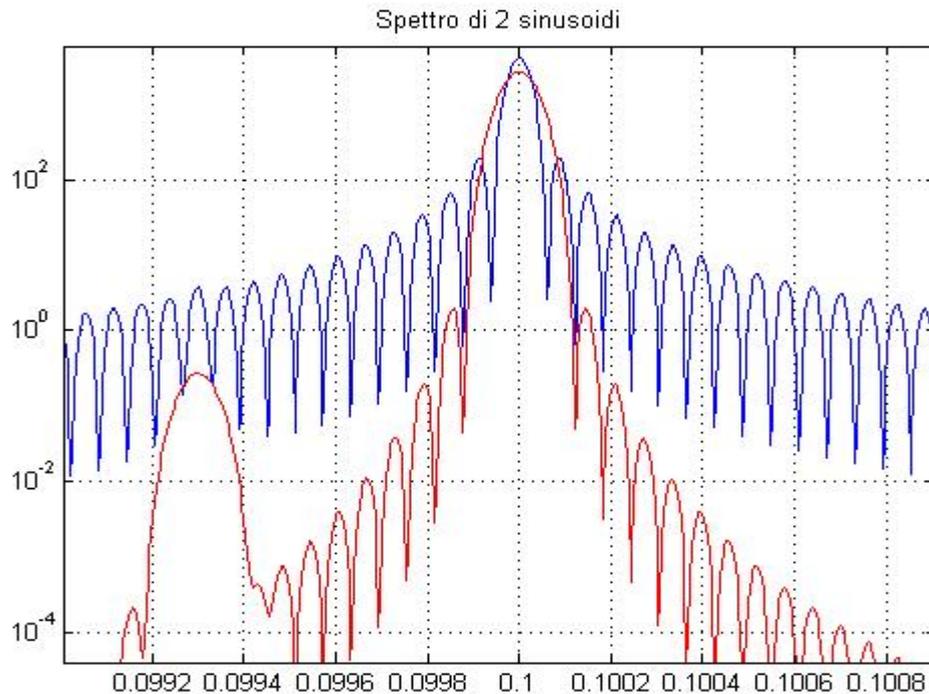


Figura 6-5

Si nota che i lobi laterali della finestra rettangolare nascondono completamente il picco più piccolo, perfettamente visibile invece con la finestra di Von Hann.

Sono state studiate numerose finestre, con proprietà diverse. Quale utilizzare dipende dal tipo di dati che si hanno.

Un ultimo esempio: la finestra triangolare (confrontata con quella di hanning)

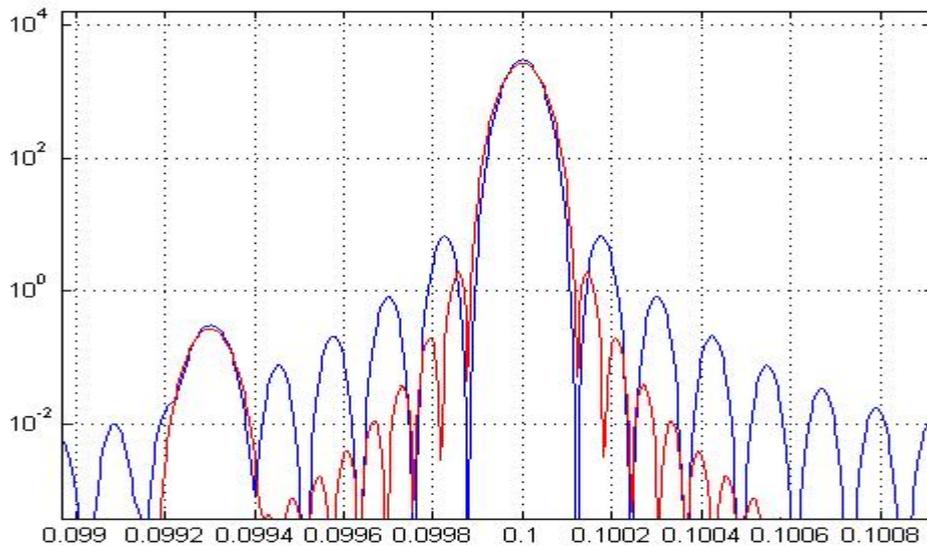


Figura 6-6

Si noti che i lobi laterali sono più ampi, ma c'è una risoluzione (ed un'ampiezza di picco) leggermente migliore. I lobi poi hanno una larghezza doppia.

Infine torniamo al concetto di risoluzione: introdurre una finestra (anche la rettangolare, che spesso è una “finestra involontaria”) significa “smussare” lo spettro col modulo quadro della trasformata di Fourier della finestra. Se lo spettro è $S(\omega)$ e se la finestra $w(t)$ ha trasformata di Fourier $W(\omega)$, il valore aspettato dello spettro “finestrato” è

$$(6.37) \quad S_w(\omega) = k \cdot S(\omega) \otimes |W(\omega)|^2$$

dove k è una costante di normalizzazione.

Se prendiamo il segnale di più alta risoluzione possibile nello spettro, una delta spettrale, cioè una sinusoidale nel tempo, possiamo definire la risoluzione come la larghezza con cui è visto questo segnale (per esempio la larghezza a mezza altezza; definizioni migliori per la risoluzione possono farsi, per esempio, conoscendo il rapporto segnale-rumore). Questo non è approssimativamente che il modulo quadro della trasformata di Fourier della finestra (a parte una costante moltiplicativa e una traslazione).

Ecco il grafico, in scala logaritmica, per le finestre rettangolare e hanning

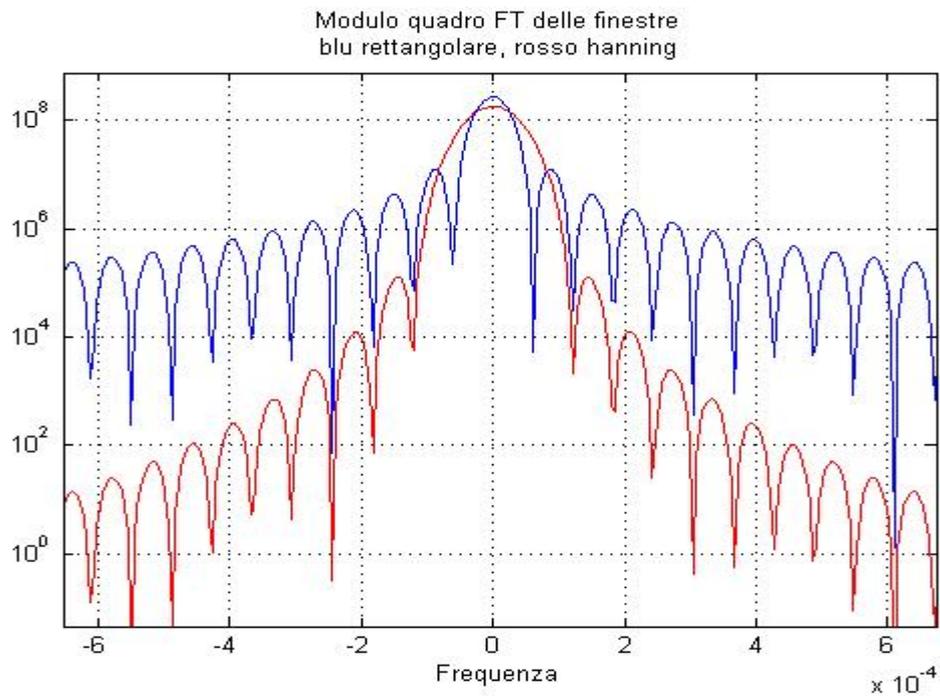


Figura 6-7

ed ecco lo stesso grafico in scala lineare

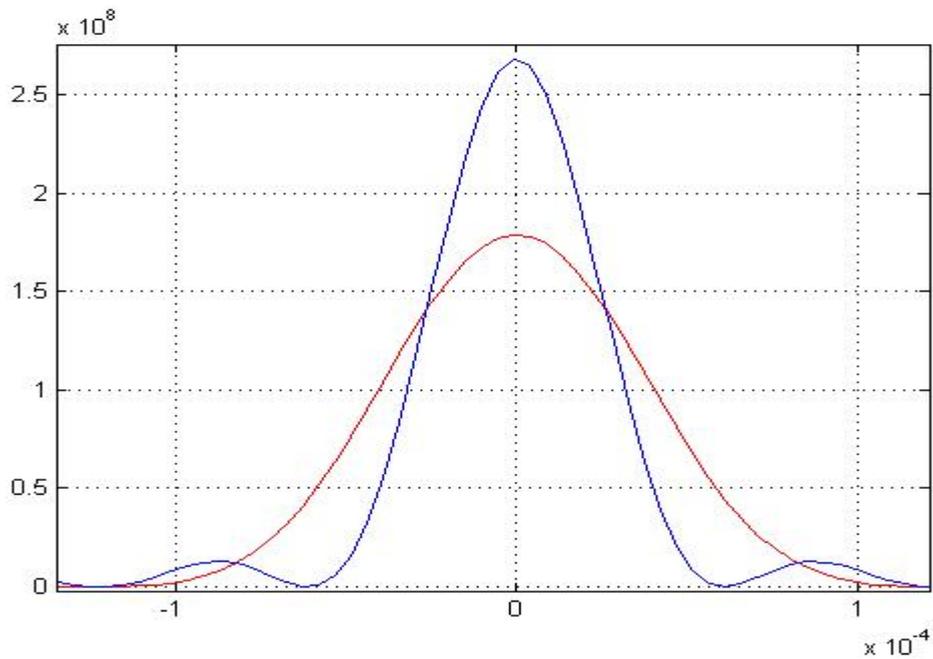


Figura 6-8

Da questo vediamo che la risoluzione (col metodo della larghezza a mezza altezza) è circa 1.5 volte peggiore nell'hanning.

- **Incertezza della stima**

Un problema dello stimatore (6.31) o di quello (6.35) è che l'incertezza della stima è molto elevata, e, soprattutto, non si riduce con l'aumentare della lunghezza del pezzo di dati.

Per evidenziare la cosa, supponiamo che il nostro segnale sia una successione correlata di campioni gaussiani normalizzati, cioè un rumore bianco con densità spettrale $S(\Omega) = 1$.

Per semplicità usiamo lo stimatore (6.31). La trasformata di Fourier è data da

$$(6.38) \quad X(\Omega) = \sum_{i=1}^N x_i \cdot e^{-j(i-1)\Omega} = \sum_{i=1}^N x_i \cdot \cos((i-1)\cdot\Omega) + j \cdot \sum_{i=1}^N x_i \cdot \sin((i-1)\cdot\Omega)$$

Consideriamo il pezzo $\sum_{i=1}^N x_i \cdot \cos((i-1)\cdot\Omega)$. Per ogni valore di Ω abbiamo la somma di N

numeri casuali $x_i \cdot \cos((i-1)\cdot\Omega)$ che hanno media 0 e varianza $\frac{1}{2}$. Per il teorema del limite centrale quindi il valore della sommatoria è una variabile normale con media nulla e varianza $\frac{N}{2}$. Lo stesso può dirsi dell'altra sommatoria con i seni. $|X(\Omega)|^2$ è una variabile casuale

ottenuta dalla somma dei quadrati di due variabili gaussiane: ha quindi distribuzione esponenziale³⁰, con valore atteso N. Poi, poiché per la stima dello spettro, il modulo quadro della trasformata di Fourier va divisa per N, abbiamo che **la distribuzione dei valori dello spettro è esponenziale con media 1**. (Ricordiamo che ci aspetteremmo semplicemente 1).

Ecco un pezzo dello spettro di questo genere

³⁰ La distribuzione esponenziale, a parte un coefficiente di normalizzazione, è la distribuzione del χ^2 con 2 gradi di libertà (vedi (2.93). In essa la deviazione standard è uguale al valore atteso. La probabilità di avere valori superiori di x è $P = e^{-\frac{x}{\mu}}$. (Vedi pagina 43)

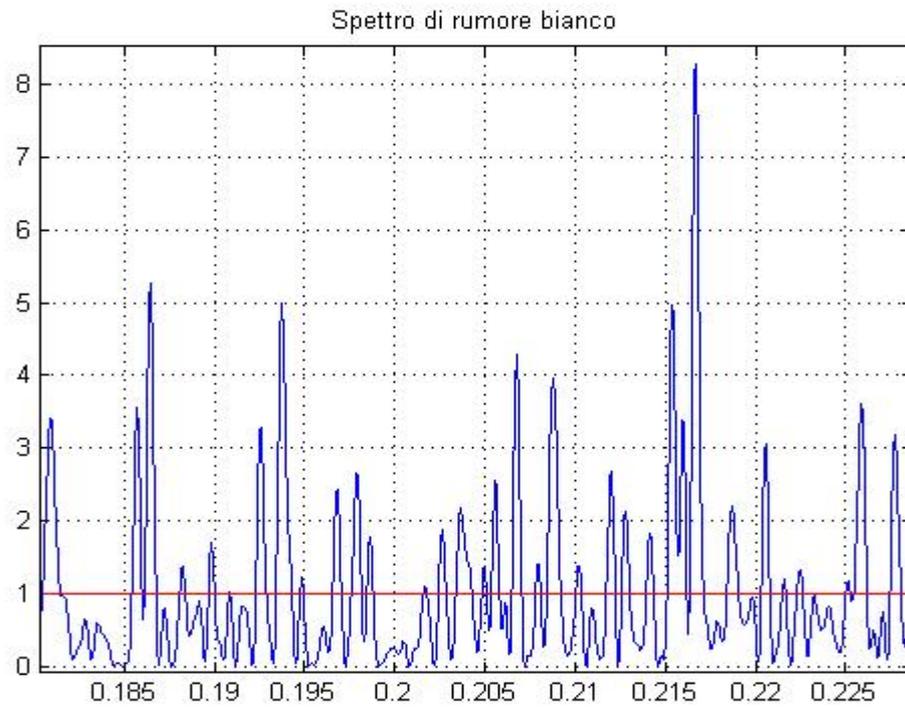


Figura 6-9

Si noti che ci aspetteremmo una linea costante a 1 (linea rossa). L'istogramma delle ampiezze spettrali è

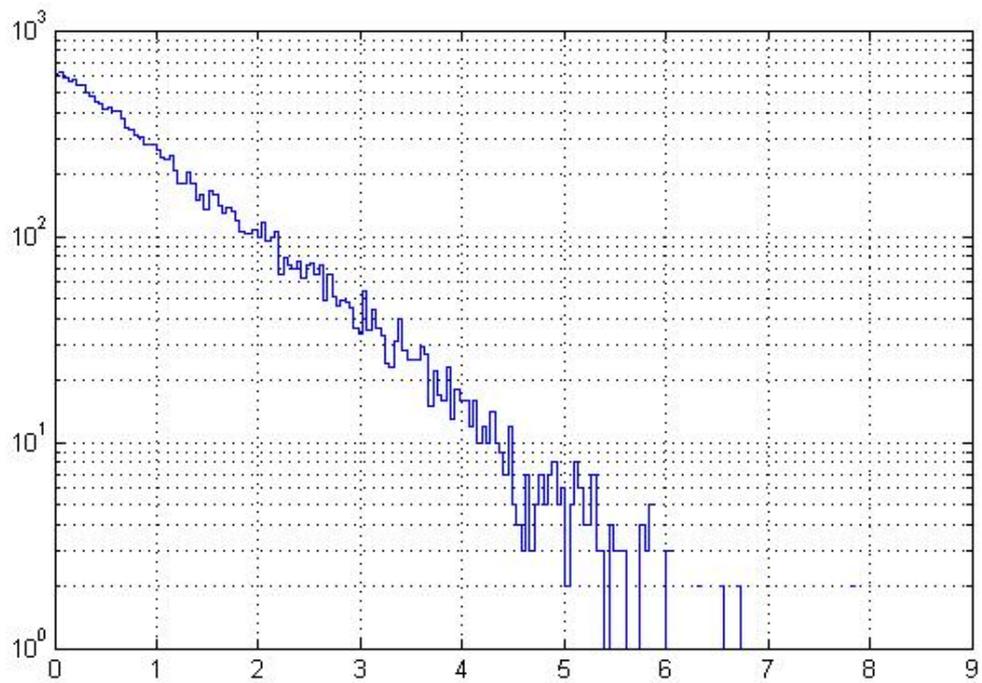


Figura 6-10

Si noti che la distribuzione è esponenziale (una retta in plot semilog).

Attenzione ! la situazione non cambia minimamente con l'aumentare del numero di campioni.

Col numero dei campioni invece aumenta il numero di punti indipendenti dello spettro (la risoluzione).

Notiamo inoltre che la distribuzione esponenziale ha code abbastanza pesanti, cioè valori di probabilità elevate anche per alti rapporti critici³¹. Per esempio, per CR=3 abbiamo $P(x > 4) = e^{-4} = 0.0183$.

Per ridurre quindi le fluttuazioni, possiamo operare in due modi:

- smussare la stima spettrale con un opportuno filtro passa-basso
- ricavare varie stime spettrali del tipo (6.31) o (6.35) da pezzi di dati indipendenti e farne la media.

Quest'ultima è la procedura più usata (in genere è più comodo fare FFT più corte).

Avendo a disposizione un pezzo di $M \cdot N$ campioni, lo stimatore più generale è quindi

$$(6.39) \quad \hat{S}(\Omega) = \frac{1}{M} \cdot \sum_{k=1}^M \frac{\left| \sum_{i=1}^N w_i \cdot x_{i+(k-1) \cdot N} \cdot e^{-j \cdot (i-1) \cdot \Omega} \right|^2}{\sum_{i=1}^N |w_i|^2}$$

La scelta di M (il numero di pezzi) si fa mediando tra la necessità di una buona risoluzione e quella di una bassa incertezza.

Nell'ipotesi che abbiamo fatto in precedenza, di rumore bianco e finestra rettangolare, la distribuzione attesa è una distribuzione del χ^2 normalizzata (a valore atteso 1) con 2M gradi di libertà.

Ecco il grafico di un pezzo di spettro stimato con finestra rettangolare ed M=10.

³¹ Ricordiamo che il rapporto critico (critical ratio) è definito come $CR = \frac{|x - \mu|}{\sigma}$.

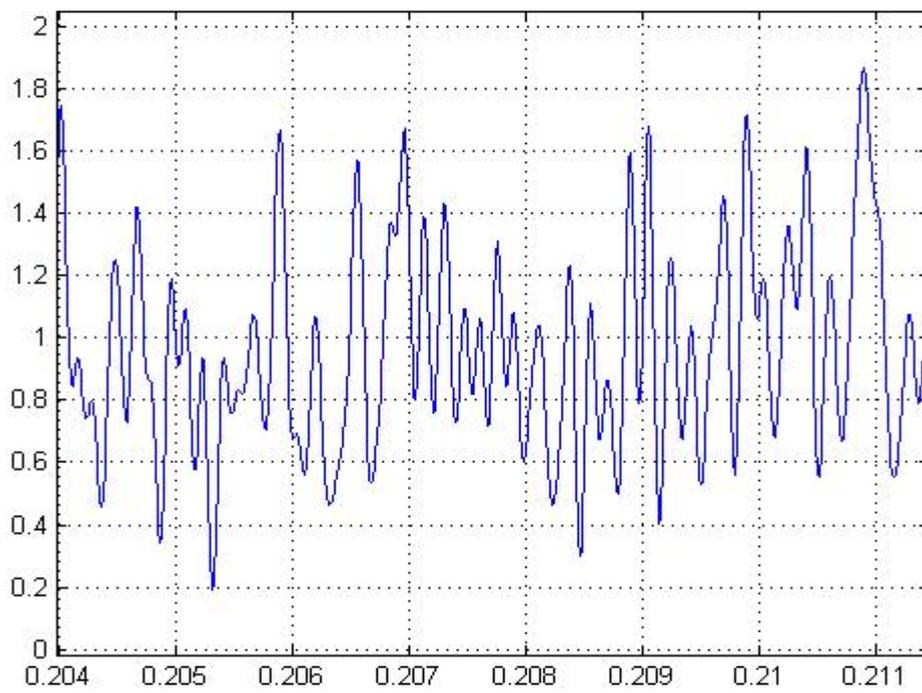


Figura 6-11

ed ecco l'istogramma

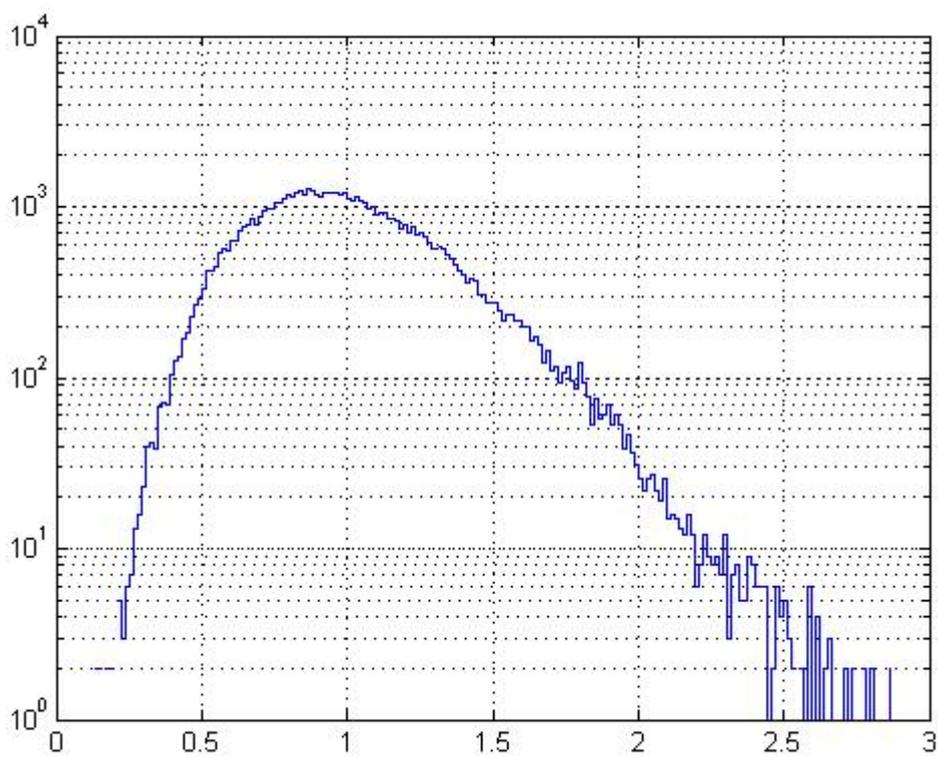


Figura 6-12

Ed ecco con $M=100$:



Figura 6-13

e la distribuzione è

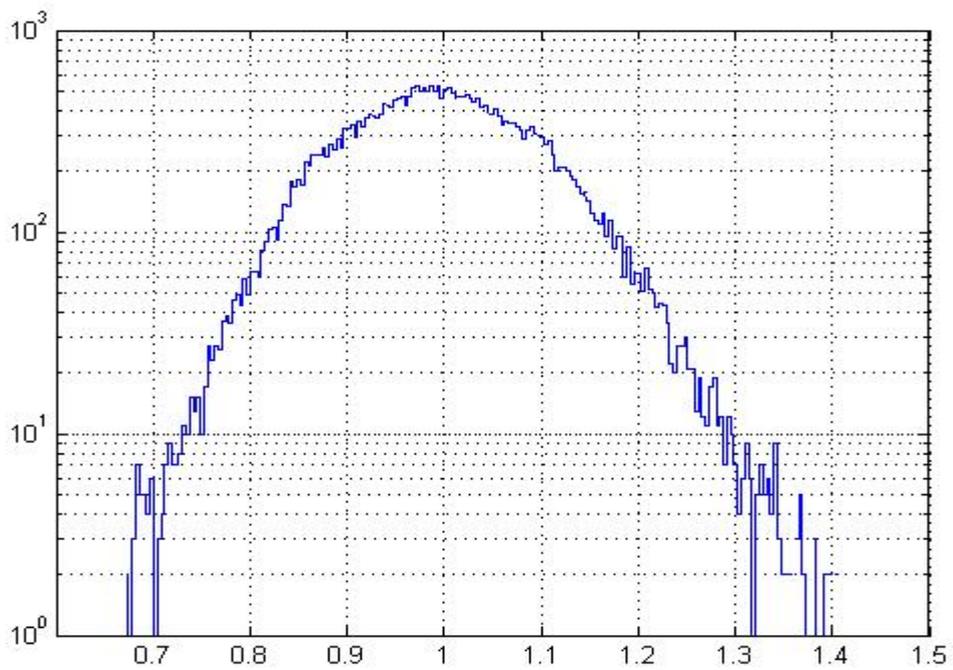


Figura 6-14

Nel caso di finestre non rettangolari, è ragionevole che i pezzi siano interallacciati. Infatti le parti iniziali e finali di ciascun pezzo contribuiscono molto poco al valore spettrale: l'interallacciamento (per esempio a metà) fa sì che tutti i dati siano considerati.

6.4.2 Stimatori spettrali parametrici

Un modo alternativo di stimare lo spettro di potenza di un segnale è scegliere un modello che riteniamo adeguato e stimarne i parametri, quindi calcolare lo spettro di potenza del modello con quei parametri.

Uno dei modelli più usati è il modello AR, detto anche “a massima entropia”, perché presuppone un andamento asintotico esponenziale dell'autocorrelazione, che corrisponde alla massima entropia.

Facciamo un esempio:

- riteniamo che un certo segnale che stiamo misurando sia adeguatamente modellizzato da un processo AR di un certo ordine:

$$(6.40) \quad y_i = b_0 \cdot x_i + w_1 \cdot y_{i-1} + w_2 \cdot y_{i-2} \dots + w_n \cdot y_{i-n}$$

Ora moltiplichiamo il membro a destra e a sinistra per y_{i-k} , con $k > 0$, e prendiamone il valore aspettato. Poiché $R_{xy}(k) = 0$ per $k > 0$, si ha

$$(6.41) \quad R_{yy}(k) = w_1 \cdot R_{yy}(k-1) + w_2 \cdot R_{yy}(k-2) \dots + w_n \cdot R_{yy}(k-n)$$

e, prendendo tutti i $1 \leq k \leq n$ e indicando $r_k = R_{yy}(k)$, abbiamo

$$(6.42) \quad \begin{pmatrix} r_0 & r_1 & \dots & r_{n-1} \\ r_1 & r_0 & \dots & r_{n-2} \\ \vdots & \vdots & & \vdots \\ r_{n-1} & r_{n-2} & \dots & r_0 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

Queste sono le equazioni di Yule e Walker. Con esse si possono ricavare i coefficienti del modello AR dai valori dell'autocorrelazione.

La matrice simmetrica che compare viene detta matrice di Toeplitz e il sistema viene risolto con l'efficiente algoritmo di Levinson e Durbin.

Se nelle (6.42) mettiamo, al posto dei valori dell'autocorrelazione, delle loro stime, possiamo stimare i valori dei parametri w_i del modello AR(n). Da questi lo spettro viene stimato (vedi (3.60)) come

$$(6.43) \quad \hat{S}_y(\Omega) = \sigma_x^2 \cdot |F(e^{j\Omega})|^2 = \frac{b_0^2 \cdot \sigma_x^2}{\left| 1 - \sum_{k=1}^n \tilde{w}_k \cdot e^{-j\Omega k} \right|^2}$$

Questo tipo di stima spettrale viene implementata in vari modi, a seconda di come viene eseguita la stima dell'autocorrelazione o risolto il sistema (6.42).

Gli stimatori spettrali parametrici sono particolarmente buoni ed efficienti se il modello usato è a) aderente ai dati, b) non ha un numero troppo elevato di parametri.

6.5 Cross-correlazione e cross-spettro

La **correlazione incrociata** (o **cross-correlazione** o anche **intercorrelazione**) tra due processi stocastici stazionari X e Y è data da

$$(6.44) \quad R_{XY}(\tau) = E[X(t+\tau) \cdot Y^*(t)]$$

La trasformata di Fourier di $R_{XY}(\tau)$, indicata in genere con $S_{xy}(\omega)$, è detta **cross-spettro** o **spettro incrociato di potenza**.

Una funzione di autocorrelazione non ha le simmetrie e limitazioni di forma dell'autocorrelazione. Così il cross-spettro può assumere valori negativi o anche complessi.

Nel caso di processi ergodici, si può eseguire una stima di $R_{XY}(\tau)$ con

Data l'ipotesi di ergodicità, possiamo stimare la $R_{XY}(\tau)$ sostituendo il valore aspettato della (6.44) con la media su due pezzi di dati $\{x_1, x_2, \dots, x_N\}$ e $\{y_1, y_2, \dots, y_N\}$

$$(6.45) \quad \hat{R}_{xy}(k) = \frac{1}{N-k} \sum_{i=1}^{N-k} x_{i+k} \cdot y_i^*$$

Una diversa stima è

$$(6.46) \quad \hat{R}_{xy}(k) = \frac{1}{N} \sum_{i=1}^{N-k} x_{i+k} \cdot y_i^*$$

Per quanto riguarda la stima del cross-spettro, questa può' ottenersi come trasformata di Fourier di una stima della cross-correlazione, o tramite l'uso del teorema di Wiener-Khinchin, per esempio nel modo seguente.

Si dividono i dati disponibili per ciascun processo in M pezzi di N dati. Si calcola quindi

$$(6.47) \quad \hat{S}_{xy}(\Omega) = \frac{1}{M} \cdot \sum_{k=1}^M \frac{\left(\sum_{i=1}^N w_i \cdot x_{i+(k-1) \cdot N} \cdot e^{-j(i-1)\Omega} \right) \cdot \left(\sum_{i=1}^N w_i \cdot y_{i+(k-1) \cdot N}^* \cdot e^{-j(i-1)\Omega} \right)}{\sum_{i=1}^N |w_i|^2}$$

La scelta di M (il numero di pezzi) si fa mediando tra la necessità di una buona risoluzione e quella di una bassa incertezza.

6.6 Coerenza

Lo spettro di potenza incrociato, come la cross-correlazione, è una misura della similarità di due processi. La **coerenza** (coherence) è una forma normalizzata del cross-spettro. La possiamo definire³² come

$$(6.48) \quad C_{xy}(\Omega) = \frac{S_{xy}(\Omega)}{\sqrt{S_x(\Omega) \cdot S_y(\Omega)}}$$

La stima della coerenza si fa con

$$(6.49) \quad \hat{C}_{xy}(\Omega) = \frac{\hat{S}_{xy}(\Omega)}{\sqrt{\hat{S}_x(\Omega) \cdot \hat{S}_y(\Omega)}}$$

Attenzione però. Le stime vanno fatte con equazioni come la (6.47) e la (6.39), con un valore di M abbastanza elevato. Per M=1 la (6.49) dà 1.

³² Talvolta la coerenza viene definita come il quadrato della presente definizione.

7 Filtraggio e trasformazione dei dati

7.1 Segnali e rumori, rapporto segnale/rumore

Il tipico problema di misura è il seguente:

Siamo interessati a stimare un segnale $x(t)$ (o la sua trasformata $X(\omega)$), ma possiamo osservare $y(t)$ che

- contiene una trasformazione x' di x
- contiene addizionato un altro segnale non desiderato

Supporremo che la trasformazione del segnale x sia lineare e descritta da un sistema F , e che il segnale non desiderato aggiuntivo (che chiamiamo **rumore**) sia un processo gaussiano descritto da uno spettro

$$(7.1) \quad S_n(\omega) = S_0 \cdot |N(\omega)|^2$$

dove S_0 è una costante e N è uno degli infiniti sistemi che risolvono l'equazione (7.1).

La situazione è descritta, nel dominio trasformato, dall'equazione

$$(7.2) \quad Y(\omega) = X(\omega) \cdot F(\omega) + S_0 \cdot N(\omega)$$

Mostriamo lo schema di questo processo

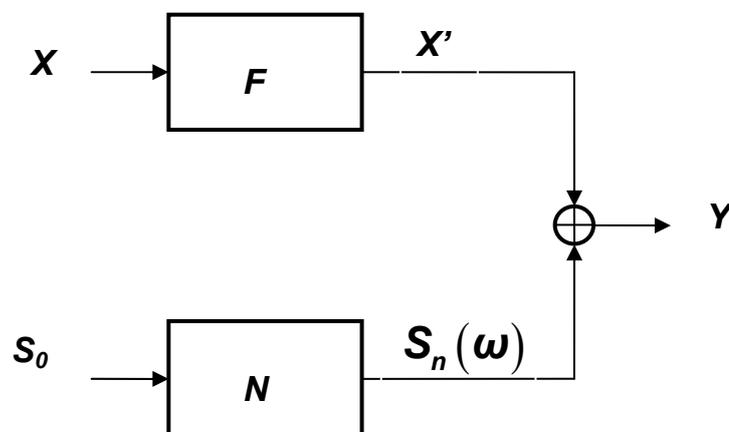


Figura 7-1

Nell'ipotesi che il rumore non esista (o sia trascurabile) il problema si riduce a trovare il sistema inverso di F (se questo è invertibile) o una sua buona approssimazione.

Più complessa (e praticamente interessante) è il caso in cui il rumore non sia trascurabile.

Distinguiamo tre casi particolarmente interessanti:

1. il segnale $x(t)$ ha una forma nota,

$$(7.3) \quad x(t - t_0) = A \cdot w(t)$$

e va individuata la sua presenza, e stimati il tempo di occorrenza t_0 e l'ampiezza A .

Questo è il problema della **rivelazione (detection)**. In tal caso si definisce **rapporto segnale/rumore (SNR – signal-to-noise ratio)** il rapporto tra il massimo del segnale x e la deviazione standard del rumore

$$(7.4) \quad SNR = \frac{\max(x(t))}{\sigma_n}$$

2. il segnale $x(t)$ è un processo stocastico che vogliamo stimare. In questo caso possiamo definire SNR il rapporto tra la deviazione standard del segnale e quella del rumore³³

$$(7.5) \quad SNR = \frac{\sigma_x}{\sigma_n}$$

3. il segnale $x(t)$ è un processo stocastico di cui vogliamo stimare uno o più parametri (per esempio il valor medio e/o la varianza). In questo caso ci interessa l'incertezza sulla stima del parametro. L'SNR può essere in questo caso l'incertezza relativa (quando questa ha senso).

In tutti questi casi ci interessa di ridurre l'effetto del rumore. Un altro modo di dire ciò è “aumentare il rapporto segnale/rumore”³⁴. Ciò però ha senso se la distribuzione del rumore rimane la stessa dopo la stima (per esempio è e rimane gaussiana). Altrimenti il SNR non è un buon indice, ciò che conta sono le probabilità, per esempio di avere un falso segnale o di avere un certo errore di stima o di farsi “sfuggire” segnali non abbastanza ampi.

Problemi come questi si risolvono con la costruzione di particolari sistemi detti **filtri**, con particolari caratteristiche. Vedremo in questo capitolo come calcolare i migliori filtri per un dato scopo e come realizzarli praticamente.

³³ Talora si usa come SNR il rapporto delle varianze, cioè il quadrato di questa quantità. Definiremo tale grandezza **rapporto segnale/rumore di potenza (power SNR)**

³⁴ Ovviamente non basta ridurre il rumore: ciò si avrebbe, per esempio, semplicemente moltiplicando i dati per una costante piccola, ma ridurre il rumore non sacrificando l'ampiezza del segnale.

7.2 Il filtro adattato

Tratteremo qui il primo caso elencato nel paragrafo 7.1 e cioè il caso in cui il segnale $x(t)$ ha una forma nota $w(t)$,

$$(7.6) \quad x(t+t_0) = A \cdot w(t)$$

e va individuata la sua presenza, e stimati il tempo di occorrenza t_0 e l'ampiezza A . La soluzione di questo problema viene chiamata **filtro adattato (matched filter)**. Si noti che ottimizzare la stima di A significa anche ottimizzare il rapporto segnale/rumore, ma non viceversa (l'ottimizzazione dell'SNR lascia libero un parametro di ampiezza).

Supponiamo inoltre che $F=1$, cioè $x'=x$, per cui

$$(7.7) \quad y(t) = A \cdot w(t-t_0) + \varepsilon(t)$$

dove $\varepsilon(t)$ è il rumore.

Possiamo scrivere ciò in forma discretizzata

$$(7.8) \quad y_i = A \cdot w_{i-i_0} + \varepsilon_i$$

e discutere direttamente questo caso.

Ovviamente per definire la forma w abbiamo infinite scelte (ognuna con una diversa ampiezza); scegliamo quella che “normalizza” l'energia

$$(7.9) \quad \sum_{k=-\infty}^{\infty} w_k^2 = 1$$

7.2.1 Caso del rumore bianco

Supponiamo ora che il rumore sia bianco, quindi gli ε_i sono campioni gaussiani con spettro

$$S_n(\omega) = \sigma_\varepsilon^2.$$

Cerchiamo il filtro migliore con la tecnica della massima verosimiglianza.

Possiamo scrivere la verosimiglianza ad un dato istante (per semplicità di notazione abbiamo scelto l'istante 0) come

$$(7.10) \quad L(A; \{y_i\}) = \prod_i \frac{1}{\sqrt{2\pi \cdot \sigma_\varepsilon}} \cdot e^{-\frac{(y_i - A \cdot w_i)^2}{2\sigma_\varepsilon^2}}$$

e prendendone il logaritmo ed ignorandone la parte costante abbiamo

$$(7.11) \quad l(A; \{y_i\}) = \sum_i -\frac{(y_i - A \cdot w_i)^2}{2\sigma_\varepsilon^2}$$

Eguagliando la derivata a 0 e sostituendo \hat{A} per A, si ha

$$(7.12) \quad \sum_i A \cdot w_i^2 = \sum_i y_i \cdot w_i$$

$$\hat{A} = \frac{\sum_i y_i \cdot w_i}{\sum_i w_i^2} = \sum_i y_i \cdot w_i$$

quindi, per ogni istante i,

$$(7.13) \quad \hat{A}_i = y_i \otimes w_{-k} = \sum_{k=-\infty}^{\infty} y_{i+k} \cdot w_k = \sum_{k=-\infty}^{\infty} (A \cdot w_{i+k-i_0} + \varepsilon_{i+k}) \cdot w_k$$

Cioè la stima di A è la cross-correlazione dei dati y per la forma del segnale w.

Poiché il problema è gaussiano (le statistiche coinvolte sono per ipotesi tutte normali), ci aspettavamo che la soluzione sarebbe stata un filtro lineare, come in effetti è.

Se non è presente alcun segnale, avremo in uscita un nuovo rumore

$$(7.14) \quad \eta_i = \sum_{k=-\infty}^{\infty} w_k \varepsilon_{i+k}$$

stazionario con varianza

$$(7.15) \quad \sigma_\eta^2 = \sigma_\varepsilon^2 \cdot \sum_{k=-\infty}^{\infty} w_k^2 = \sigma_\varepsilon^2$$

Se ci fosse solo il segnale, avremmo

$$(7.16) \quad \hat{A}_i = \sum_{k=-\infty}^{\infty} y_{i+k} \cdot w_k = A \cdot \sum_{k=-\infty}^{\infty} w_{i+k-i_0} \cdot w_k = A \cdot R_w(i - i_0)$$

$R_w(l)$ è l'autocorrelazione del segnale w_i , definita come

$$(7.17) \quad R_w(l) = \sum_{k=-\infty}^{\infty} w_{l+k} \cdot w_k = w_k \otimes w_{-k}$$

Se vogliamo un filtro che massimizzi il SNR, allora possiamo fare la correlazione tra i dati e $B \cdot w_k$, dove B è una qualsiasi costante, e avremo lo stesso risultato. Ma porre B=1 semplifica la valutazione dell'ampiezza del segnale.

Vediamo la soluzione nel dominio della frequenza.

Indichiamo l'uscita del filtro adattato con $\{v_i\}$, con trasformata V. Abbiamo

$$(7.18) \quad V(\omega) = Y(\omega) \cdot W^*(\omega)$$

Possiamo vederne lo schema:

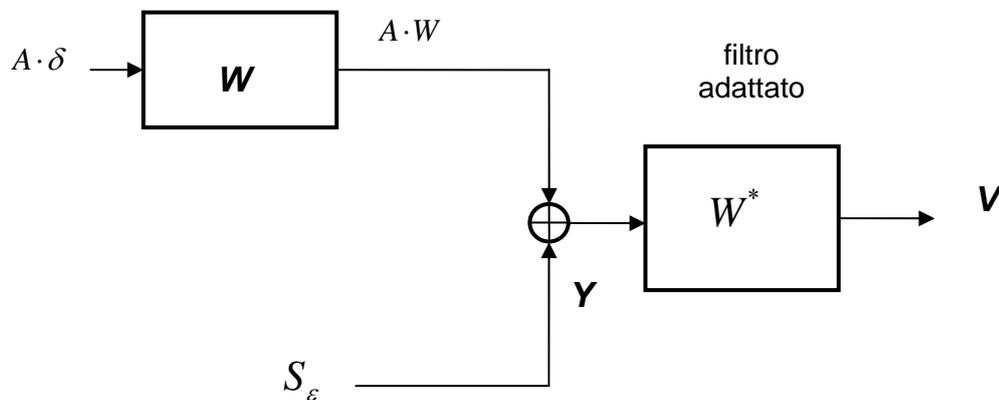


Figura 7-2

Nel nodo somma vediamo la somma del segnale w (quando e se presente la delta all'ingresso di W) e del rumore bianco di densità spettrale S_ϵ .

Se i segnali w sono rari (o assenti), $\{v_i\}$ è la risposta del filtro adattato al rumore bianco; avrà quindi spettro

$$(7.19) \quad S_w(\omega) = S_\epsilon \cdot |W|^2$$

e autocorrelazione pari all'autocorrelazione del segnale w moltiplicato per σ_ϵ^2 . Questa è un'interessante proprietà del filtro adattato.

C'è stato miglioramento di SNR? Dato che la varianza del rumore è la stessa prima e dopo, vediamo cosa succede sul segnale. Prima il massimo valore era

$$(7.20) \quad Max_{pre} = A \cdot \max(|w_i|)$$

dopo diventa

$$(7.21) \quad Max_{post} = A$$

Ora è evidente che, per la (7.9), $\max(|w_i|) \leq 1$ (l'eguaglianza si ha se w_i è una delta). E il guadagno del filtro è

$$(7.22) \quad G = \frac{1}{\max(|w_i|)}$$

7.2.2 Altre dimostrazioni

Dato il grande interesse pratico e teorico del filtro adattato, vediamo altre metodologie di calcolo.

Si voglia massimizzare il SNR con un filtro lineare. Sia questo filtro B definito da una risposta impulsiva (non necessariamente causale) $\{a_i\}$; cioè

$$(7.23) \quad \frac{A \cdot \sum_i w_i \cdot b_i}{\sigma_\varepsilon^2 \cdot \sum_i |b_i|^2}$$

Per la disuguaglianza di Schwartz (vedi (2.140)), il massimo si ha se $b_i \propto w_i$ (di cui $b_i = w_i$ è un caso particolare).

Un altro modo di calcolare il filtro adattato, che giunge allo stesso risultato (7.12), è il seguente:

- si voglia stimare, col metodo dei minimi quadrati, l'ampiezza A di un impulso $\{A \cdot w_i\}$.

Il rumore gaussiano bianco aggiunto può essere visto come un errore gaussiano additivo. La stima sarà lineare per la gaussianità (e potremo indicare lo stimatore con il filtro F di risposta impulsiva $\{f_i\}$). L'errore da minimizzare sarà

$$(7.24) \quad P = E \left[\left(A - \sum_i f_i \cdot y_i \right)^2 \right]$$

Poniamo nella (7.24) $f_i = w_i + a_i$. Abbiamo, ricordando la (7.9),

$$\begin{aligned}
(7.25) \quad P &= E \left[\left(A - \sum_i (w_i + a_i) \cdot (A \cdot w_i + \varepsilon_i) \right)^2 \right] = \\
&E \left[\left(A - A \cdot \sum_i w_i^2 - A \cdot \sum_i w_i a_i - \sum_i w_i \varepsilon_i - \sum_i \varepsilon_i a_i \right)^2 \right] = \\
&E \left[\left(A \cdot \sum_i w_i a_i + \sum_i w_i \varepsilon_i + \sum_i \varepsilon_i a_i \right)^2 \right]
\end{aligned}$$

Ora, se vogliamo minimizzare P, il primo termine in parentesi, che dipende da A che può essere grande a piacere, si deve annullare, quindi deve essere

$$(7.26) \quad \sum_i w_i a_i$$

(ortogonalità di \mathbf{w} e \mathbf{a}). Quindi, ricordando che le ε_i sono scorrelate e $E[\varepsilon_i^2] = \sigma_\varepsilon^2$,

$$\begin{aligned}
(7.27) \quad E \left[\left(\sum_i (w_i + a_i) \varepsilon_i \right)^2 \right] &= \sigma_\varepsilon^2 \cdot \left(\sum_i (w_i^2 + a_i^2) + 2 \cdot \sum_i w_i \cdot a_i \right) = \\
&= \sigma_\varepsilon^2 \cdot \sum_i (w_i^2 + a_i^2)
\end{aligned}$$

da cui deve essere $a_i = 0$ per tutti gli i e quindi $f_i = w_i$, come avevamo trovato con le altre ipotesi.

7.2.3 Caso generale

Vediamo ora cosa possiamo fare nel caso in cui il rumore non è bianco.

Poiché la nostra soluzione ottima dà un filtro lineare, possiamo “sbiancare” il rumore (e quindi modificare di conseguenza il segnale) e applicare a ciò il metodo appena applicato³⁵; attenzione però perché il nuovo filtro andrà adattato al segnale modificato dallo sbiancante.

Per chiarire la situazione, mostriamo lo schema delle operazioni:

³⁵ Si deve supporre che il filtro sbiancante sia realizzabile (e stabile). Nella nostra ipotesi di rumore gaussiano stazionario, ciò è sempre possibile.

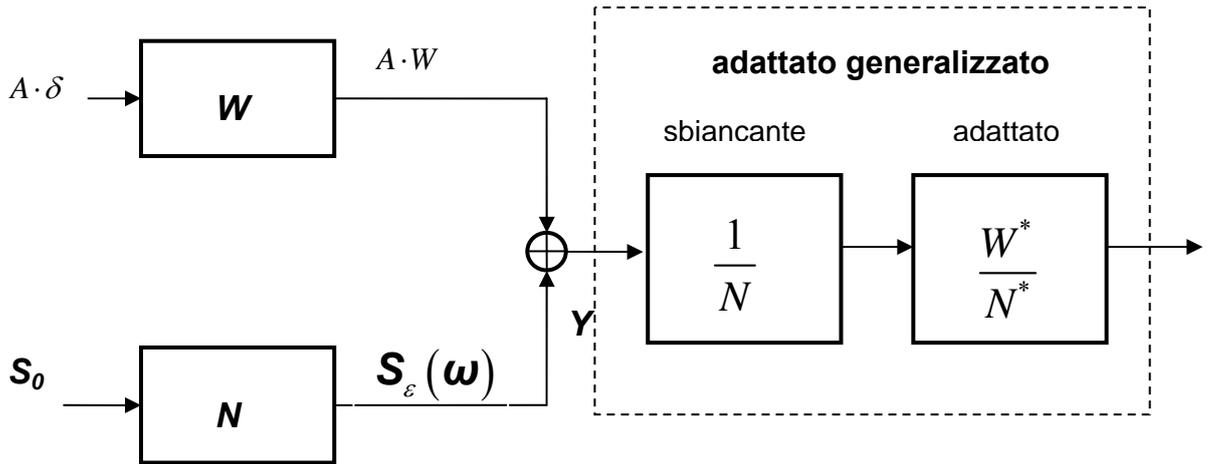


Figura 7-3

Se lavoriamo con le frequenze fisiche, per semplicità supporremo $S_0 = 1 \text{ Hz}^{-1}$ (l'ampiezza viene così conglobata in N); se usiamo le frequenze normalizzate (cioè quando usiamo la variabile pulsazione normalizzata Ω definita in (3.41)), possiamo porre $S_0 = 1$.

In formule, nel dominio della frequenza, abbiamo

$$(7.28) \quad V(\omega) = Y(\omega) \cdot \frac{W^*(\omega)}{|N(\omega)|^2} = Y(\omega) \cdot \frac{W^*(\omega)}{S_\varepsilon(\omega)}$$

Si noti che il filtro “sbiancante” (whitening filter) N è applicato due volte, una volta in modo “causale” (dal passato al futuro) e una volta in modo anti-causale (dal futuro al passato, quando il filtro applicato nel dominio della frequenza è N^*): ciò rende nullo qualsiasi sfasamento che N da solo possa introdurre.

La risposta al segnale, nel dominio della frequenza, è data da

$$(7.29) \quad V(\omega) = \frac{|W(\omega)|^2}{S_\varepsilon(\omega)}$$

Questa è sempre reale positiva e quindi la sua anti-trasformata è una funzione hermitiana. Se il segnale e il rumore sono reali, è anche una funzione pari.

Se invece c'è solo rumore, all'uscita abbiamo un rumore che ha spettro

$$(7.30) \quad S_v(\omega) = S_0 \cdot \left| \frac{W^*(\omega)}{N^*(\omega)} \right|^2 = S_0 \cdot \frac{|W(\omega)|^2}{S_\varepsilon(\omega)}$$

Si noti che, S_0 è essenziale per motivi di dimensioni fisiche, se usiamo le misure fisiche.

Per calcolare la risposta impulsiva nel tempo e l'autocorrelazione del rumore occorre calcolare

$$(7.31) \quad R_k = \mathcal{F}^{-1} \left(\frac{|W(\omega)|^2}{S_\varepsilon(\omega)} \right)$$

In particolare è importante il valore R_0 , dove ha il massimo assoluto.

Un importante risultato è che la risposta impulsiva in frequenza (7.29) è proporzionale allo spettro di potenza del rumore (7.30), e la risposta impulsiva (7.31) nel tempo è proporzionale all'autocorrelazione del rumore.

L'SNR dopo il filtro è

$$(7.32) \quad SNR = \frac{R_0}{\sqrt{S_0 \cdot R_0}} = \sqrt{\frac{R_0}{S_0}}$$

7.3 Teoria della rivelazione (Detection Theory)

Abbiamo costruito il filtro adattato, ma questo dà un valore ad ogni istante di tempo, quindi “mente” quasi sempre: infatti la sua stima è ragionevole solo quando il segnale è effettivamente presente.

Per poter “**rivelare**” (**to detect** in Inglese) effettivamente segnali, dobbiamo **decidere** quando il segnale è plausibilmente presente e quando invece no.

Nel caso del problema che abbiamo posto nel paragrafo 7.2, la decisione viene presa sulla base dell’ampiezza dell’uscita del filtro adattato:

- **decidiamo che ci sia un segnale del tipo cercato se la $\{v_i\}$ ha un massimo relativo e supera una data soglia.**

La situazione è analoga a quella descritta nel paragrafo 2.3.10 sui test statistici.

Supponiamo che l’uscita del filtro abbia una densità di probabilità $f(x; s)$, dove s è l’ampiezza del segnale eventualmente presente. Se poniamo una soglia θ , abbiamo i due casi

A. Il segnale non è presente. Con probabilità

$$(7.33) \quad P_0 = \int_{-\infty}^{\theta} f(x; 0) \cdot dx$$

il segnale non supererà la soglia (e quindi abbiamo indovinato) e con probabilità $1 - P_0$ la soglia viene superata anche in assenza di segnale e quindi commettiamo un **errore di primo tipo**, ovvero abbiamo un “**falso allarme**” (**false alarm**).

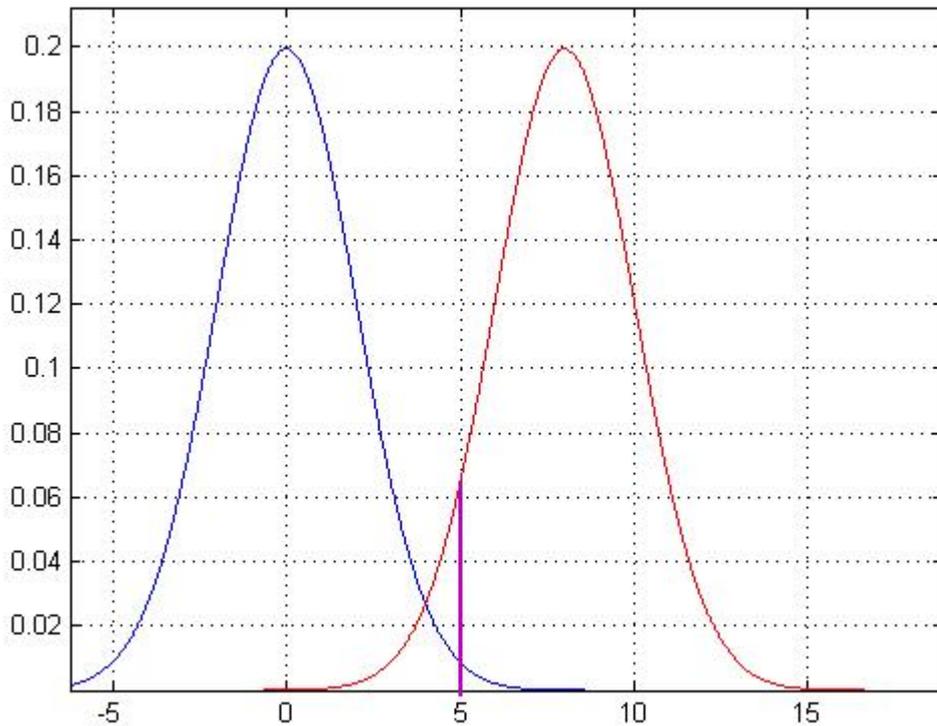
B. Il segnale è presente. Con probabilità

$$(7.34) \quad P_s = \int_{\theta}^{\infty} f(x; s) \cdot dx$$

il segnale supererà la soglia (e quindi abbiamo indovinato) e con probabilità $1 - P_s$ la soglia non viene superata anche in presenza di segnale che quindi non viene rivelato; in tal caso commettiamo un **errore di secondo tipo**, ovvero abbiamo un “**falso rigetto**” (**false dismissal**).

Se “alziamo” la soglia, cioè aumentiamo θ , possiamo ridurre la probabilità di errori del primo tipo, ma aumentando la probabilità di errori del secondo tipo.

Nella figura seguente è presentato un caso



In questo grafico sono rappresentate le distribuzioni dell'uscita di un rivelatore (che potrebbe essere un filtro adattato), nell'ipotesi di assenza di segnale e nell'ipotesi in cui sia presente un segnale di ampiezza 8.

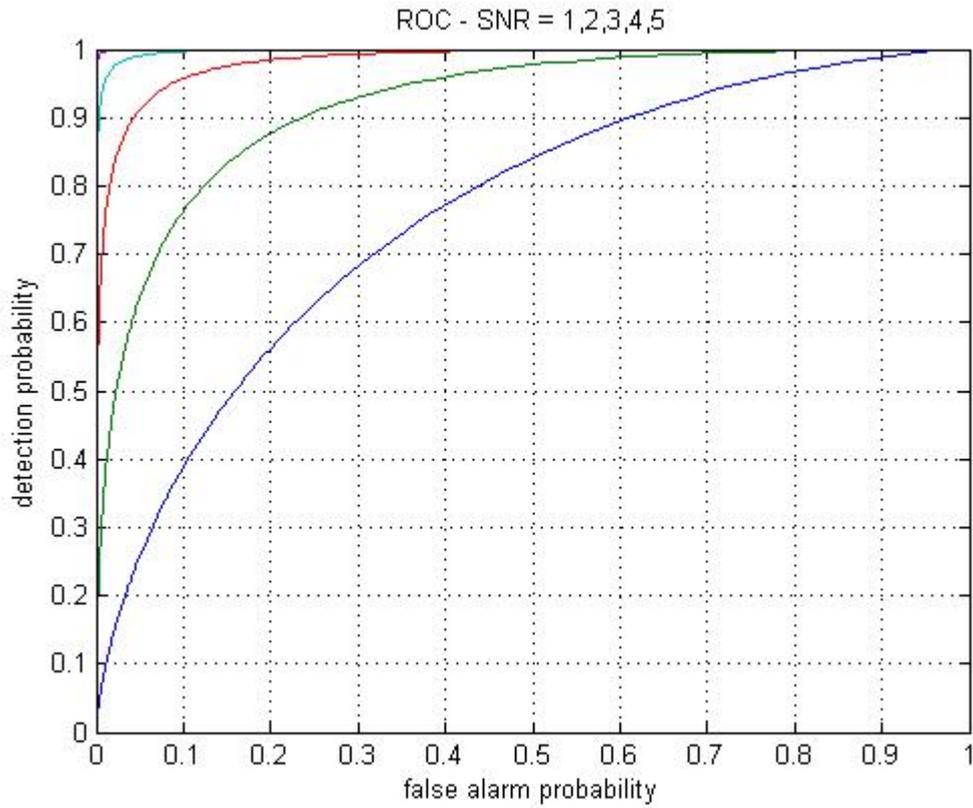
In questo caso il rumore è additivo (vedi in seguito l'equazione (7.35)) e gaussiano, di deviazione standard 2. Poniamo la soglia a 5 ed abbiamo

- **probabilità di corretto "rigetto":** $P_0 = 0.99379$
- **probabilità di falso allarme:** $1 - P_0 = 0.00621$
- **probabilità di corretta rivelazione:** $P_1 = 0.933193$
- **probabilità di falso rigetto:** $1 - P_1 = 0.066807$

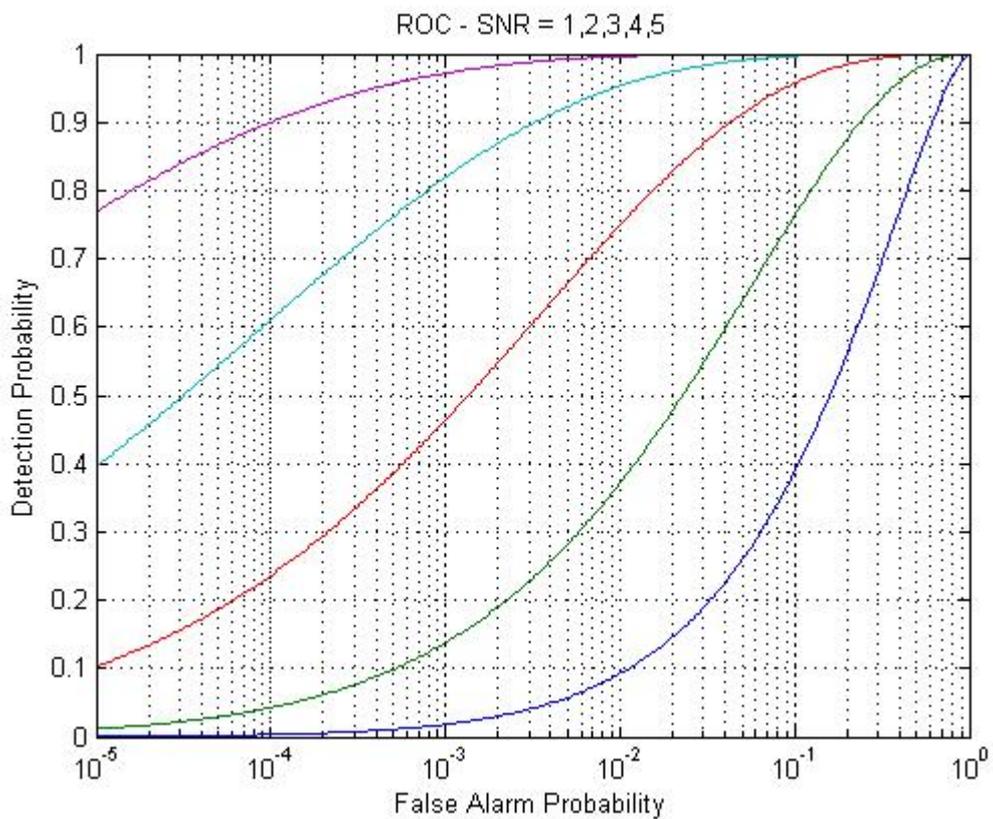
Attenzione: spesso le ipotesi non sono due, ma se il segnale s può assumere molti valori, cioè se s è una variabile casuale (quindi, per esempio, può prendere tutti i valori di un dato intervallo, secondo una data distribuzione di probabilità, che potrebbe anche non essere nota), l'ipotesi A è in effetti molteplice.

Per rappresentare la situazione si usano spesso le **Receiver Operating Characteristics (ROC)**, delle curve che danno, per ogni valore dell'ampiezza del segnale, la probabilità di rivelazione P_s in funzione della probabilità di falso allarme $1 - P_0$, al variare della soglia.

Eccone un esempio, per rumore gaussiano additivo



e, in scala semi-logaritmica,



Se, come accade di frequente, il rumore è additivo, allora

$$(7.35) \quad f(x; s) = f(x - s)$$

e quindi tutto il processo è descritto soltanto da $f(x)$. O, ancora meglio, dalla funzione

$$(7.36) \quad P(x) = \int_x^{\infty} f(\xi) \cdot d\xi = 1 - F(x)$$

che indica la probabilità di falso allarme.

Se vogliamo confrontare due rivelatori (detector) di segnali o due filtri applicati allo stesso rivelatore, basta confrontare le relative due $f(x)$. Se le due $f(x)$ sono dello stesso tipo, basta confrontarne i parametri. Per esempio, se le $f(x)$ sono gaussiane, basta dare le due deviazioni standard (ed eventualmente i valori aspettati, se ci sono degli offset).

Se lo scopo è il confronto, in tali casi la pletora delle ROC non è una buona scelta.

7.4 Filtro di Wiener

Consideriamo ora il problema di stimare un segnale che è modellizzato da un processo stocastico stazionario gaussiano, da dei dati in cui a tale segnale è aggiunto un rumore anch'esso gaussiano e stazionario. Cioè, nel discreto,

$$(7.37) \quad y_i = x_i + n_i$$

Lo schema è il seguente

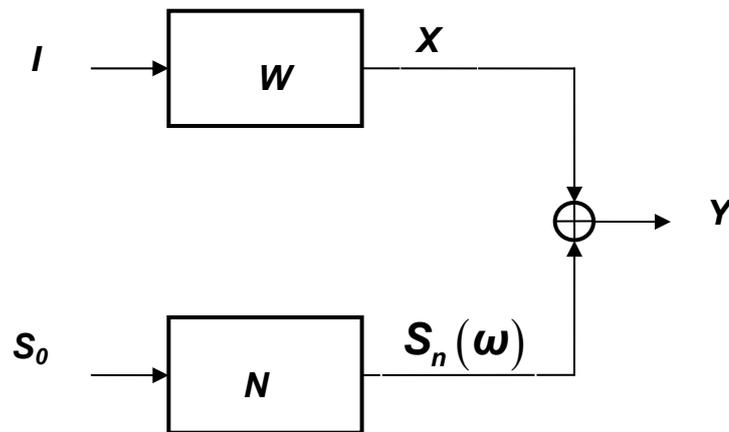


Figura 7-4

Il segnale è modellizzato come l'uscita di un filtro lineare W con all'ingresso un segnale con le caratteristiche statistiche di un rumore bianco e che chiamiamo "processo di innovazione".

Dobbiamo costruire la stima ottima nel senso dei minimi quadrati, cioè

$$(7.38) \quad \hat{x}_i = \sum_{k=-\infty}^{\infty} y_{i-k} \cdot f_k$$

tale che

$$(7.39) \quad E \left[(x_i - \hat{x}_i)^2 \right]$$

sia minimo.

Per il principio di ortogonalità, la condizione di minimo sulla (7.39) porta a che sia

$$(7.40) \quad E \left[(x_i - \hat{x}_i) \cdot y_k \right] = 0$$

ovvero

$$E \left[\left(x_i - \sum_{l=-\infty}^{\infty} (x_{i-l} + n_{i-l}) \cdot f_l \right) \cdot (x_k + n_k) \right] = 0$$

e, passando ai valori aspettati, ricordando la stazionarietà di x e n e che x e n sono indipendenti e quindi scorrelati,

$$(7.41) \quad R_{xx}(k) - \sum_l f_l \cdot (R_{xx}(l+k)) - \sum_l f_l \cdot (R_{nn}(l+k)) = 0$$

Imponendo il minimo errore secondo i minimi quadrati, troviamo, nel dominio della frequenza,

$$(7.42) \quad W(\omega) = \frac{S_x(\omega)}{S_y(\omega)} = \frac{S_x(\omega)}{S_n(\omega) + S_x(\omega)}$$

dove $S_y(\omega)$ è lo spettro di potenza dei dati, $S_x(\omega)$ è lo spettro di potenza del segnale e $S_n(\omega)$ lo spettro di potenza del rumore. Il filtro $W(\omega)$ è detto filtro di Wiener, dal nome del matematico americano Norbert Wiener.

Nel caso più generale, il filtro si scrive

$$(7.43) \quad W(\omega) = \frac{S_{xy}(\omega)}{S_y(\omega)}$$

dove $S_{xy}(\omega)$ è lo spettro incrociato tra il processo x e i dati y .

7.5 Realizzazione di filtri

Un filtro, in pratica, non è altro che un sistema discreto. Dei sistemi discreti abbiamo parlato ampiamente. Qui faremo alcune osservazioni sulla loro realizzazione.

Spesso infatti possono esservi varie possibili implementazioni di uno stesso filtro, soprattutto tenendo presente che raramente si possono realizzare filtri “ottimi”, ma ci si deve accontentare approssimazioni, spesso buone. Ciò anche perché l'informazione sui segnali e sui rumori non sempre è completa.

7.5.1 Filtri FIR

I filtri FIR (Finite Impulsive Response) non sono che un altro nome dei sistemi MA (moving average) visti nel capitolo 3. La caratteristica fondamentale è che questi filtri sono intrinsecamente stabili.

Spesso tuttavia possono essere molto pesanti dal punto di vista computazionale.

7.5.2 Filtri IIR

I filtri IIR (Infinite Impulse Response, AKA filtri AR) sono invece, in genere, meno pesanti computazionalmente (poiché spesso hanno pochi coefficienti) e sono più “naturali” per via della dipendenza a tutte le distanze. In certi casi, piccoli errori nel calcolo dei coefficienti (anche per arrotondamenti) possono dar luogo ad instabilità.

7.5.3 Filtri a sfasamento nullo

I filtri di tipo FIR o IIR sono causali e introducono inevitabilmente (a parte casi banali) uno sfasamento dipendente dalla frequenza.

Talvolta è utile non introdurre sfasamenti, ovvero applicare filtri con funzione di trasferimento reale.

Lo sfasamento introdotto dai filtri FIR può essere lineare con la frequenza, come è lo sfasamento di un semplice ritardo. Infatti la funzione di trasferimento di un ritardo Δt è

$$(7.44) \quad F_D(j\omega; \Delta t) = e^{-j\omega\Delta t}$$

e quindi la fase è

$$(7.45) \quad \varphi = -\omega \cdot \Delta t$$

Ciò accade se i coefficienti del filtro FIR sono simmetrici rispetto a un valore, per esempio i seguenti due:

$$y_i = x_i - 2 \cdot x_{i-1} + x_{i-2}$$

(7.46)

$$y_i = x_i + 2 \cdot x_{i-1} + 2 \cdot x_{i-2} + x_{i-3}$$

Questo ritardo è quasi sempre inessenziale e quindi si parla spesso di filtri a fase lineare come filtri a sfasamento nullo.

Un modo per ottenere filtri a fase nulla, è far passare lo stesso filtro prima in modo causale, cioè dal passato al futuro, e quindi in modo anti-causale, cioè dal futuro al passato. Ovviamente per far ciò dobbiamo conoscere i “dati futuri” (cioè dobbiamo lavorare offline, quindi con un effettivo ritardo). Ciò che accade è che a ciascuna frequenza, se il filtro “causale” la sfasa di φ , quello anti-causale la sfasa di $-\varphi$ e quindi alla fine lo sfasamento è nullo. Ma attenzione! ciascuno dei due filtri applica a quella frequenza un guadagno G , quindi alla fine il guadagno è G^2 .

Vista nel dominio trasformato la situazione è la seguente: se il filtro applicato ha una funzione di trasferimento complessa $F(j\omega)$, applicato in modo anticausale la sua funzione di trasferimento diventa $F^*(j\omega)$ e quindi il filtro effettivamente applicato dopo le due operazioni diventa $|F(j\omega)|^2$.

7.5.4 Filtri in frequenza

Applicare un filtro FIR significa eseguire una convoluzione. Per eseguire una convoluzione di due “vettori”, uno, il pezzo di dati, lungo N e uno, il filtro FIR con M coefficienti, occorre eseguire $M \cdot N$ moltiplicazioni ed altrettante somme. Se M ed N sono numeri abbastanza grandi (per esempio³⁶ $N = 10^9$ ed $M = 10^5$), il carico computazionale diventa elevatissimo.

Si può allora far uso del fatto che nel dominio trasformato di Fourier la convoluzione è una semplice moltiplicazione (quindi, in linea di principio, M volte più veloce).

Dobbiamo però:

- dividere gli N dati in gruppi di lunghezza M (preferibilmente M deve essere una potenza di 2)
- trasformare i dati di ingresso al filtro e quindi anti-trasformare i dati di uscita. E ciò può essere fatto con la FFT (vedi paragrafo 3.4.2), con un numero di operazioni proporzionale a $M \cdot \log M$ (invece che M^2 , come sarebbe con la convoluzione).

³⁶ Sono valori tipici per un giorno di dati di un’antenna gravitazionale, con un filtro ben adattato al rumore reale. Spesso poi si devono applicare contemporaneamente migliaia di filtri.

Con M dell'ordine di 10^5 , si guadagna un fattore circa 1000 nel numero delle operazioni elementari svolte.

Il fatto di lavorare “a pezzi” sui dati, permette di realizzare facilmente filtri non-causali e/o a sfasamento nullo.

Ma attenzione ! La FFT, che è un modo veloce di calcolare la DFT, considera i dati periodici di periodo M , e questo creerebbe gravissimi artefatti nel filtraggio. L'escamotage che si usa per questo problema è, in genere, il seguente:

- si divide il pezzo di dati da analizzare in M -pezzi di lunghezza M interallacciati per la metà
- ogni M -pezzo viene utilizzato per analizzarne solo la metà centrale. Infatti con una FFT di lunghezza M , si implementa un filtro FIR di lunghezza effettiva $M/2$ (tipicamente $M/4$ con indice negativo e $M/4$ con indice positivo), e quindi con metà dei coefficienti pari a zero (per filtri non-causali, la parte nulla è la centrale)

7.6 I disturbi impulsivi - Filtro a mediana

Fin'ora abbiamo visto solo filtri lineari. I filtri lineari operano prendendo tutti i campioni del segnale, indipendentemente dall'ampiezza, e il valore d'uscita dipende linearmente dai valori dei singoli campioni, quindi se un campione è molto grande, ha un peso grande sull'uscita.

Spesso il rumore non è solo un processo gaussiano, o comunque continuo, ma contiene dei disturbi localizzati ("burst" o impulsi) rari, ma di grande ampiezza.

Come la caratteristica del rumore che abbiamo considerato nei paragrafi 7.2 e 7.4 era lo spettro e utilizzando questa conoscenza possiamo ridurre l'effetto, così nel caso di questi burst la caratteristica che li contraddistingue è l'ampiezza e quindi grazie a questo in parecchi casi possiamo eliminarli tramite un'opportuna procedura non lineare, che dipende dall'ampiezza (purtroppo però spesso cancellando la parte del segnale che è presente nello stesso periodo).

Vediamo ora un'altra procedura non lineare per stimare il valor medio del segnale in presenza di rumore impulsivo.

Nel caso di rumore stazionario continuo, possiamo fare una media mobile (o operare con un filtro AR, come in 3.2.1). Se ci sono grossi impulsi, possiamo dividere i dati in intervalli e per ciascuno di questi calcolare la mediana. La mediana è molto meno sensibile alla presenza di grossi impulsi rispetto alla media (che è un'operazione lineare).

Questa procedura viene indicata come **filtro a mediana**.

8 Cenno alla trasformata wavelet

9 Cenni al caso di dati non stazionari

9.1 *Analisi tempo-frequenza*

9.2 *Filtri adattivi*

10 Cenni ai modelli non lineari

10.1 Filtri non lineari

10.2 Il bispettro

Quando abbiamo a che fare con modelli non-lineari, in genere l'ipotesi di gaussianità cade e quindi diventano interessanti le statistiche di ordine superiore al secondo.

Il bispettro, come abbiamo visto in (5.34), è la trasformata di Fourier del cumulante del terzo ordine di un processo stocastico stazionario

$$(9.1) \quad C_{xxx}(\tau_1, \tau_2) = E[x^*(t) \cdot x(t + \tau_1) \cdot x(t + \tau_2)]$$

Notiamo che in questo cumulante ci sono varie simmetrie, e cioè, per segnali reali,

$$(9.2) \quad \begin{aligned} ABC &\rightarrow \tau_1, \tau_2 \\ ACB &\rightarrow \tau_2, \tau_1 \\ BAC &\rightarrow -\tau_1, \tau_2 - \tau_1 \\ BCA &\rightarrow \tau_2 - \tau_1, -\tau_1 \\ CAB &\rightarrow -\tau_2, \tau_1 - \tau_2 \\ CBA &\rightarrow \tau_1 - \tau_2, -\tau_2 \end{aligned}$$

Per segnali complessi invece c'è solo

$$(9.3) \quad \begin{aligned} ABC &\rightarrow \tau_1, \tau_2 \\ ACB &\rightarrow \tau_2, \tau_1 \end{aligned}$$

Queste simmetrie si riflettono sul bispettro.

Vediamo come si presenta il bispettro di una sinusoide:

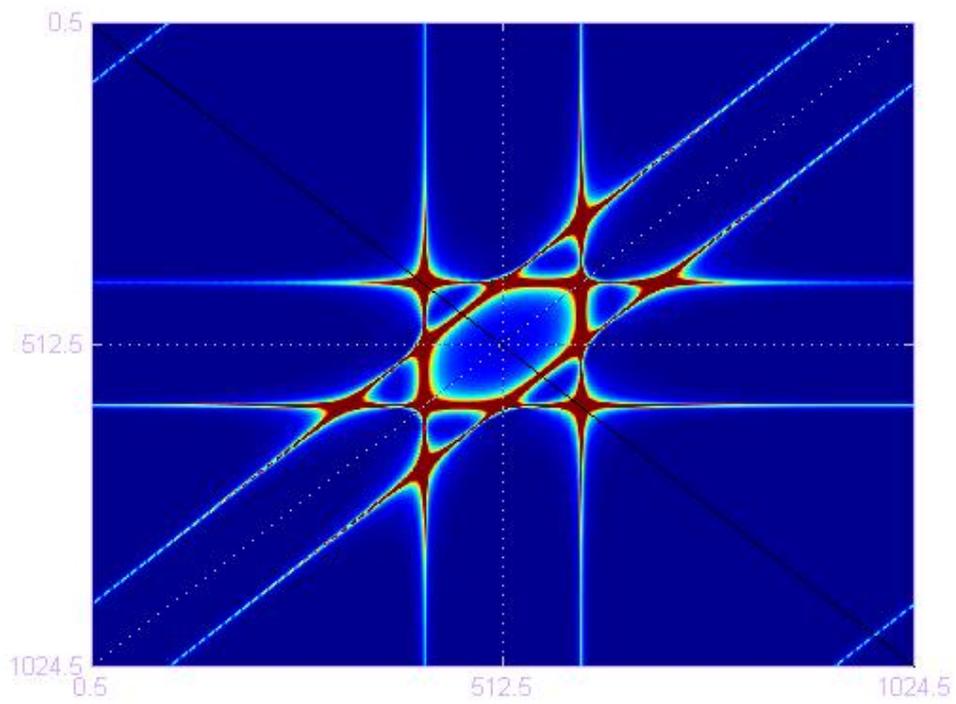


Figura 10-1

Nei sistemi non lineari raramente sono presenti frequenze “pure”, ecco come appare il bispettro se si introduce una non-linearità che genera la terza armonica

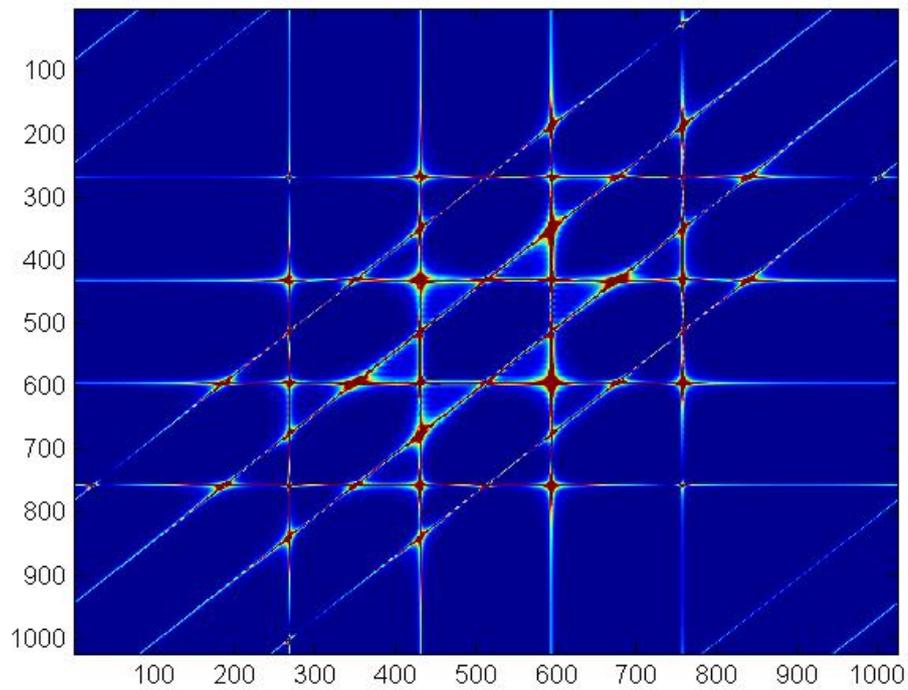


Figura 10-2

e nel caso che generi la seconda e la terza:

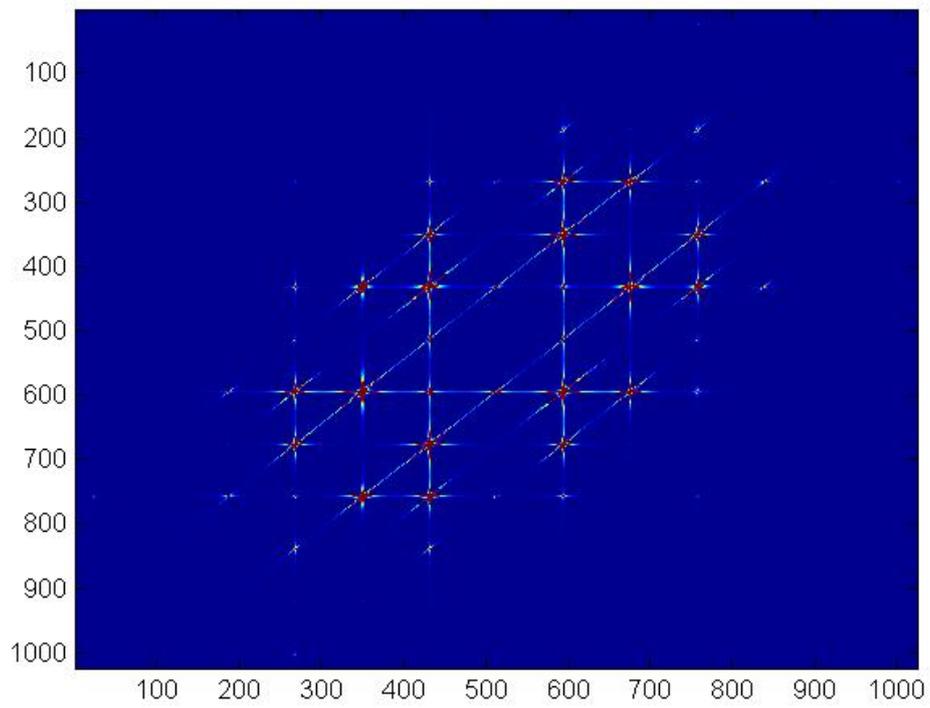


Figura 10-3

Vediamo cosa accade per il bispettro di un esponenziale immaginario:

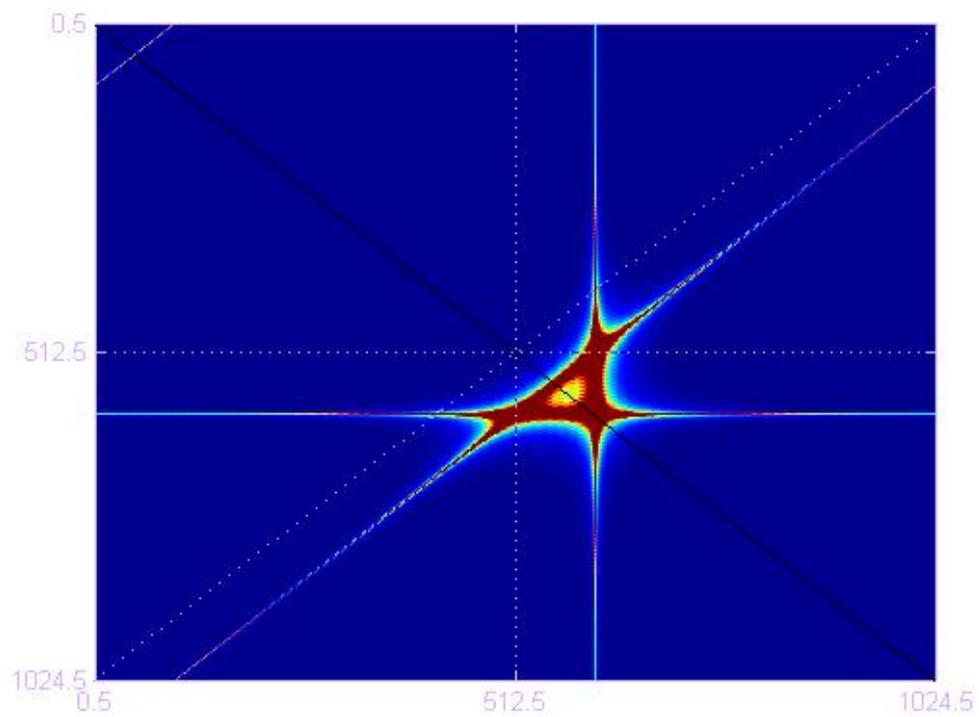


Figura 10-4

Particolarmente interessante è ciò che accade con segnali casuali. Di seguito sono presentati i bispettri nel caso di rumore bianco, ma con distribuzione rispettivamente

- gaussiano
- esponenziale
- chi quadro con 10 gradi di libertà
- uniforme

Lo spettro di questi quattro rumori è indistinguibile, mentre ecco come si presentano i 4 bispettri:

Gaussiano

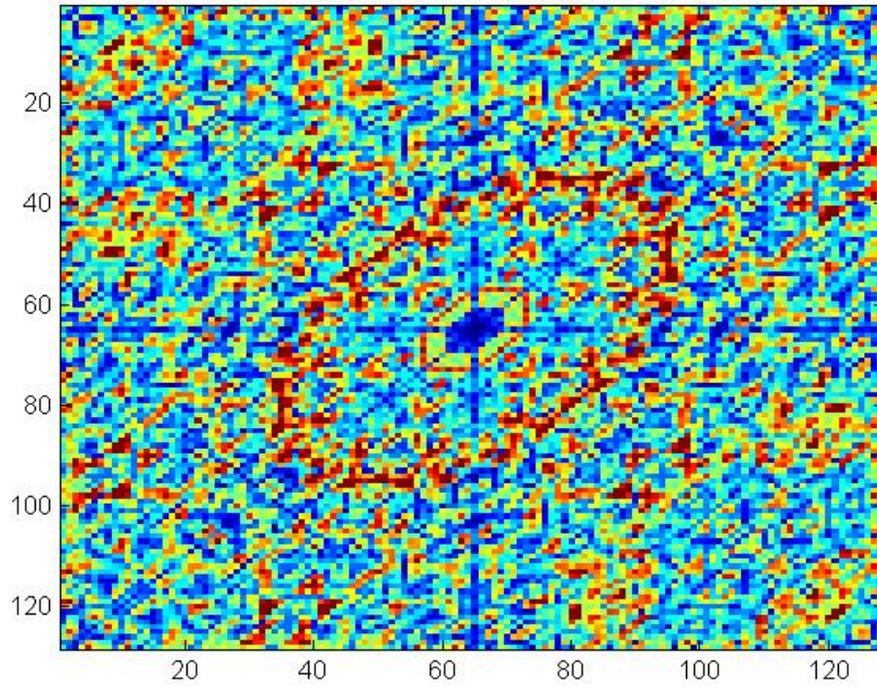


Figura 10-5

Esponenziale

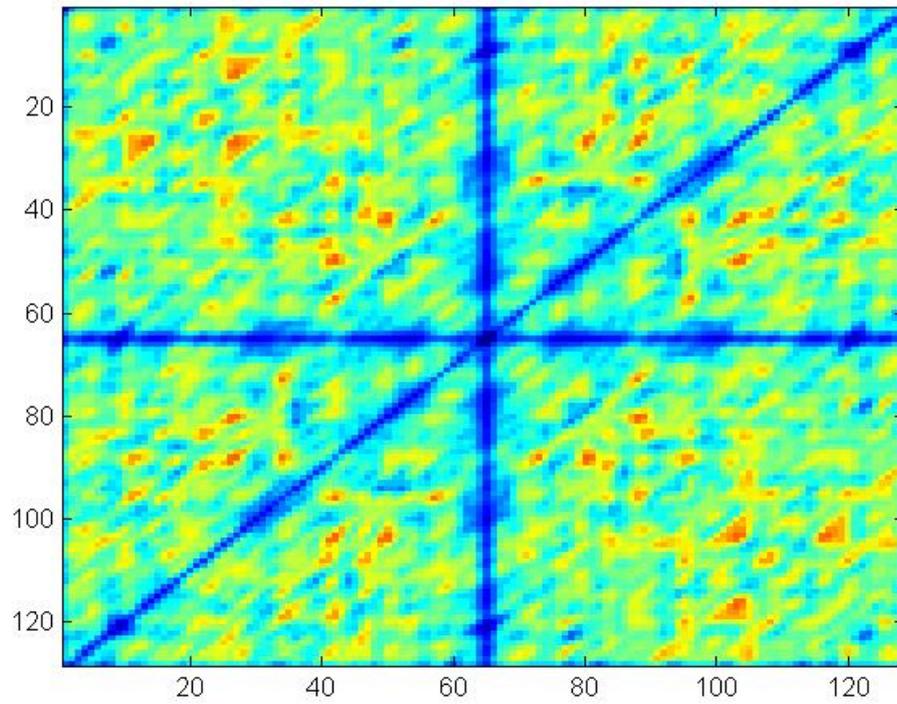


Figura 10-6

Chi quadro con 10 gradi di libertà

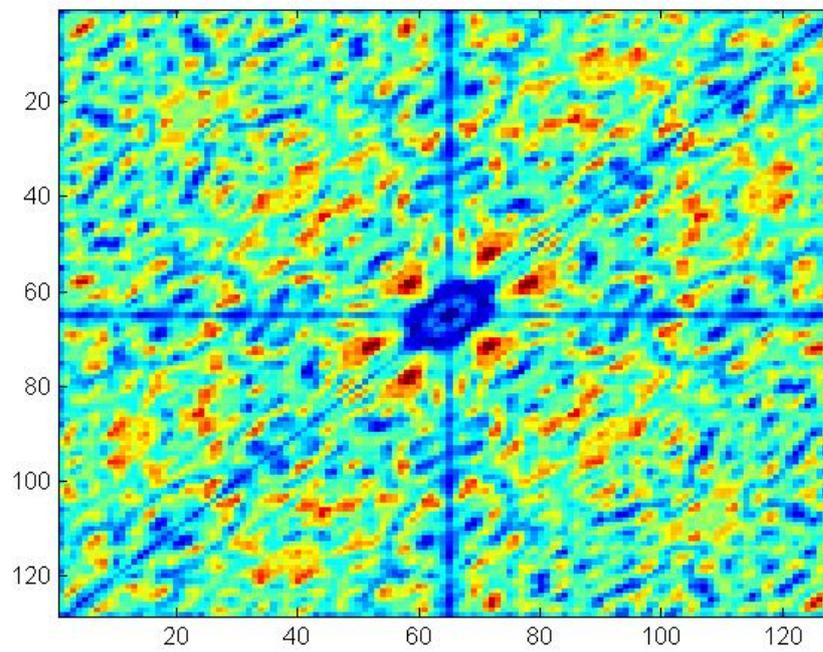


Figura 10-7

Uniforme

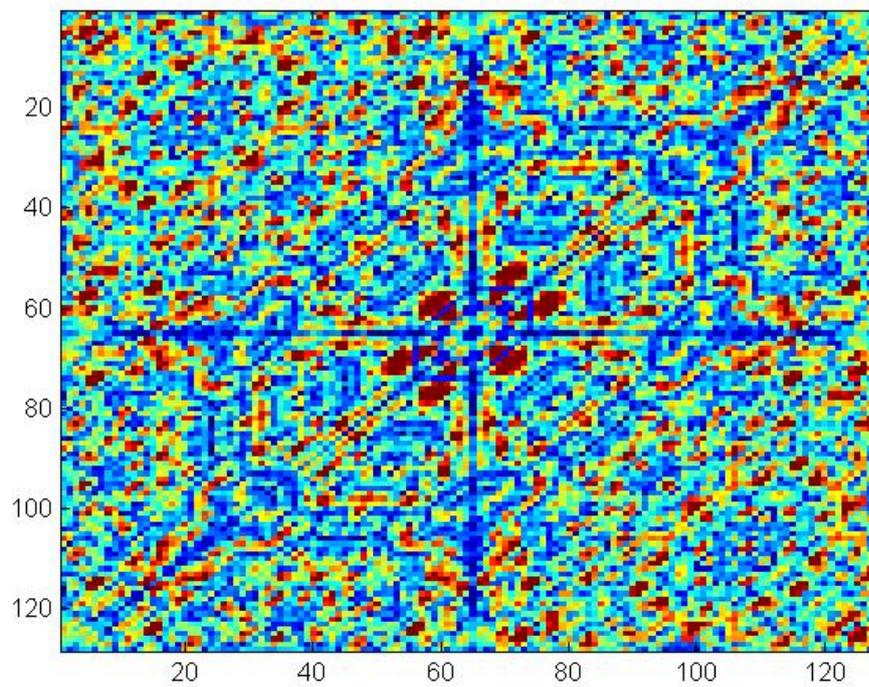


Figura 10-8

Per quanto riguarda il valore, avendo posto segnali con la stessa potenza, abbiamo il seguente risultato:

Tipo di rumore bianco	Media valore assoluto
Gaussiano	0.038
Esponenziale	0.210
Chi quadro 10gdl	0.098
Uniforme	0.002

Notiamo poi che mentre nel caso gaussiano e in quello uniforme abbiamo un pattern con variazioni rapide, nel caso dell'esponenziale e del $\chi^2(10)$ le variazioni sono più morbide. Questo si evidenzia facendo lo spettro di potenza dell'immagine (vedi paragrafo 11.2). Ecco i risultati:

Gaussiano

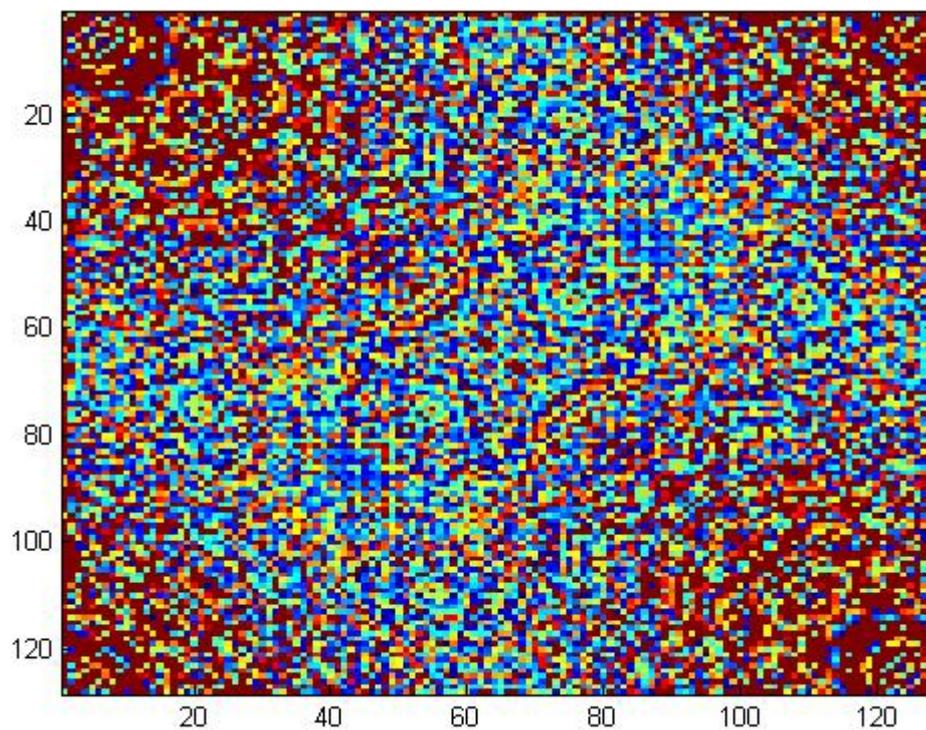


Figura 10-9

Esponenziale

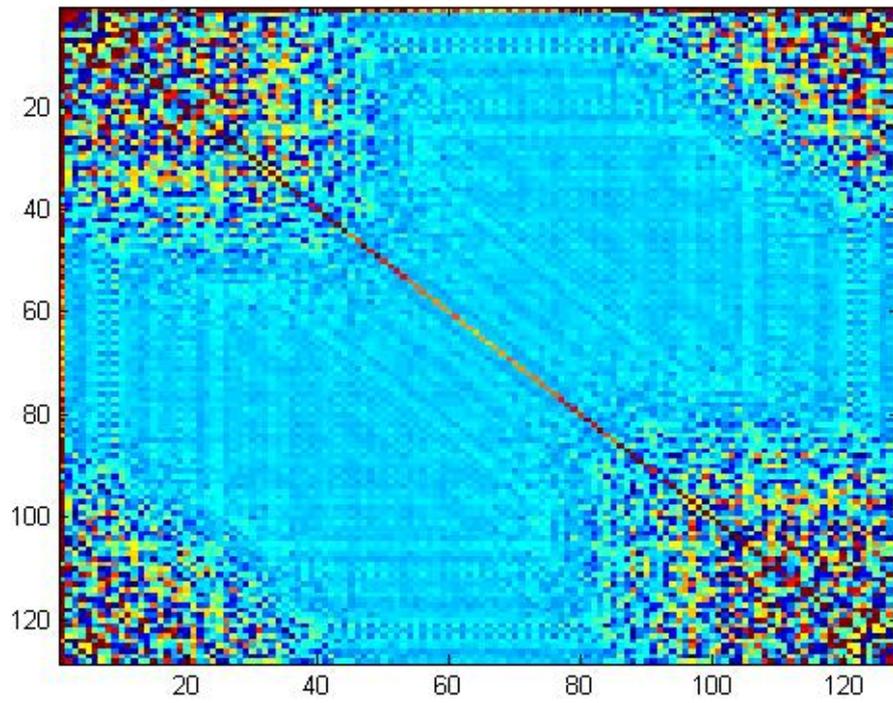


Figura 10-10

Chi Quadro (10)

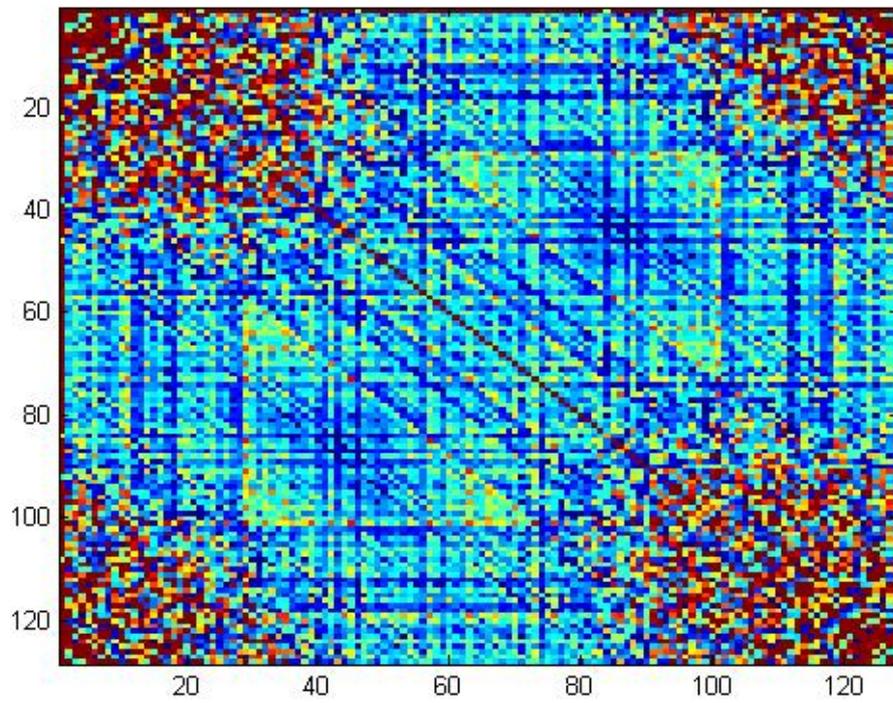


Figura 10-11

Uniforme

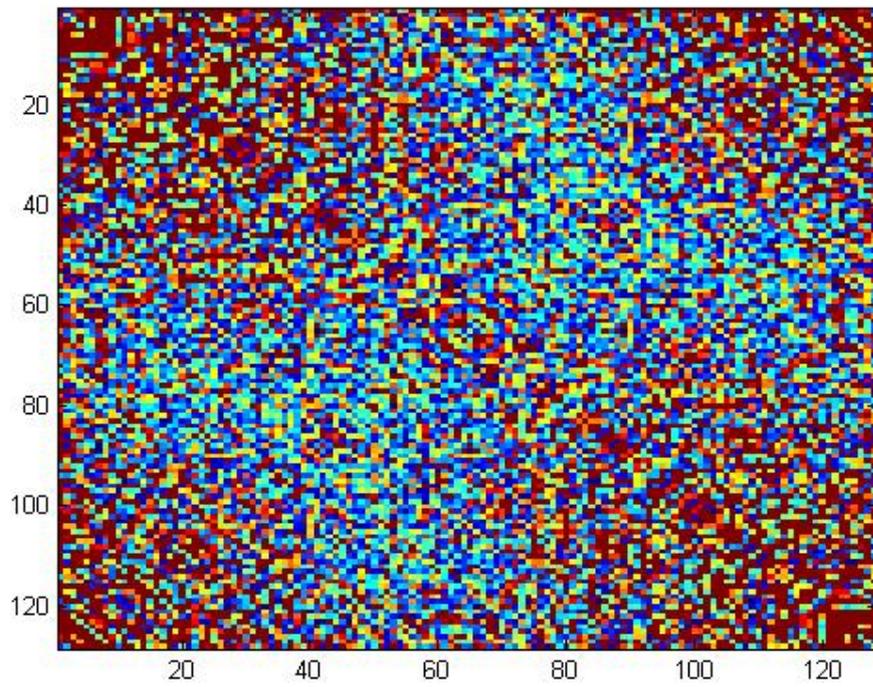


Figura 10-12

Ed infine vediamo il bispettro di rumore esponenziale complesso

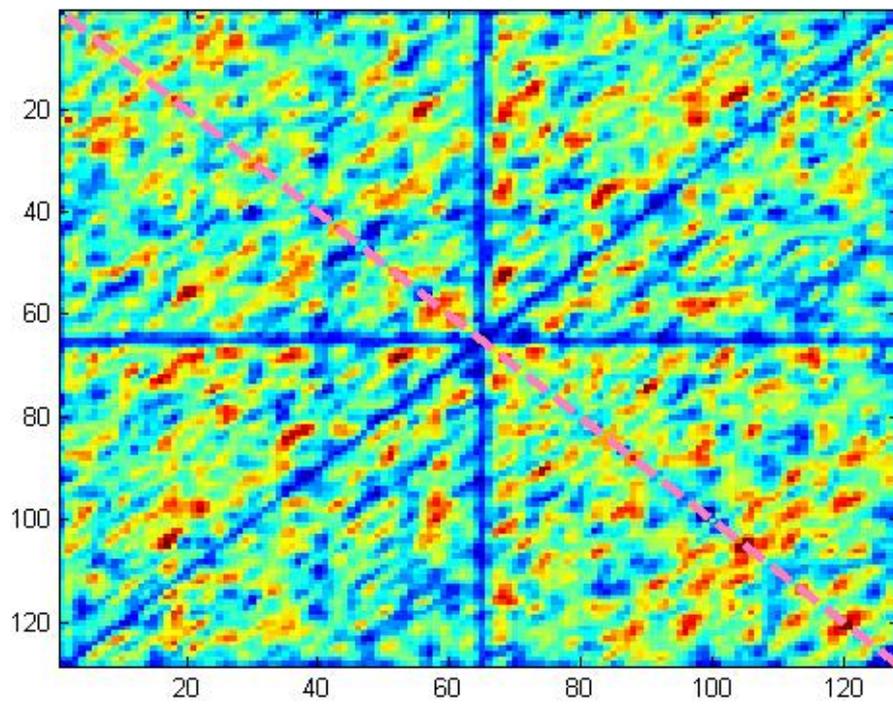


Figura 10-13

da cui si nota la riduzione della simmetri a bilatera.

11 Cenno all'immagine processing

11.1 Immagini ed elaborazione delle immagini

Un'immagine è descritta da una funzione di due variabili

$$(10.1) \quad g(x, y)$$

Distinguiamo tra immagini in bianco e nero e a colori. Nel primo caso in ogni punto la $g(x,y)$ è una funzione scalare, nel secondo a ciascun punto è associato un vettore a 3 dimensioni. Qui ci occuperemo solo di immagini in bianco e nero (anche se esistono modi di visualizzazione a colori di immagini in bianco e nero).

Come avviene per i segnali temporali, così anche le immagini vengono discretizzate per essere "lette" e elaborate dai computer. la funzione $g(x,y)$ diventa quindi una matrice

$$(10.2) \quad g_{ik}$$

Ogni elemento di questa matrice viene detto **pixel**.

La visualizzazione di una immagine può essere eseguita per esempio per scala di grigi. Nella rappresentazione a scala di grigi, rappresentiamo lo 0 (o il valore minimo) col nero e il valore massimo di nostro interesse col bianco.

Possiamo però anche usare scale "colorate" differenti che talvolta possono risultare più comprensibili o più suggestive.

Eccone un esempio. Qui l'immagine è una foto³⁷; ogni pixel ha un valore tra 0 e 63.

³⁷ Non sempre è così: per esempio un'immagine può essere una rappresentazione tempo-frequenza di un segnale.

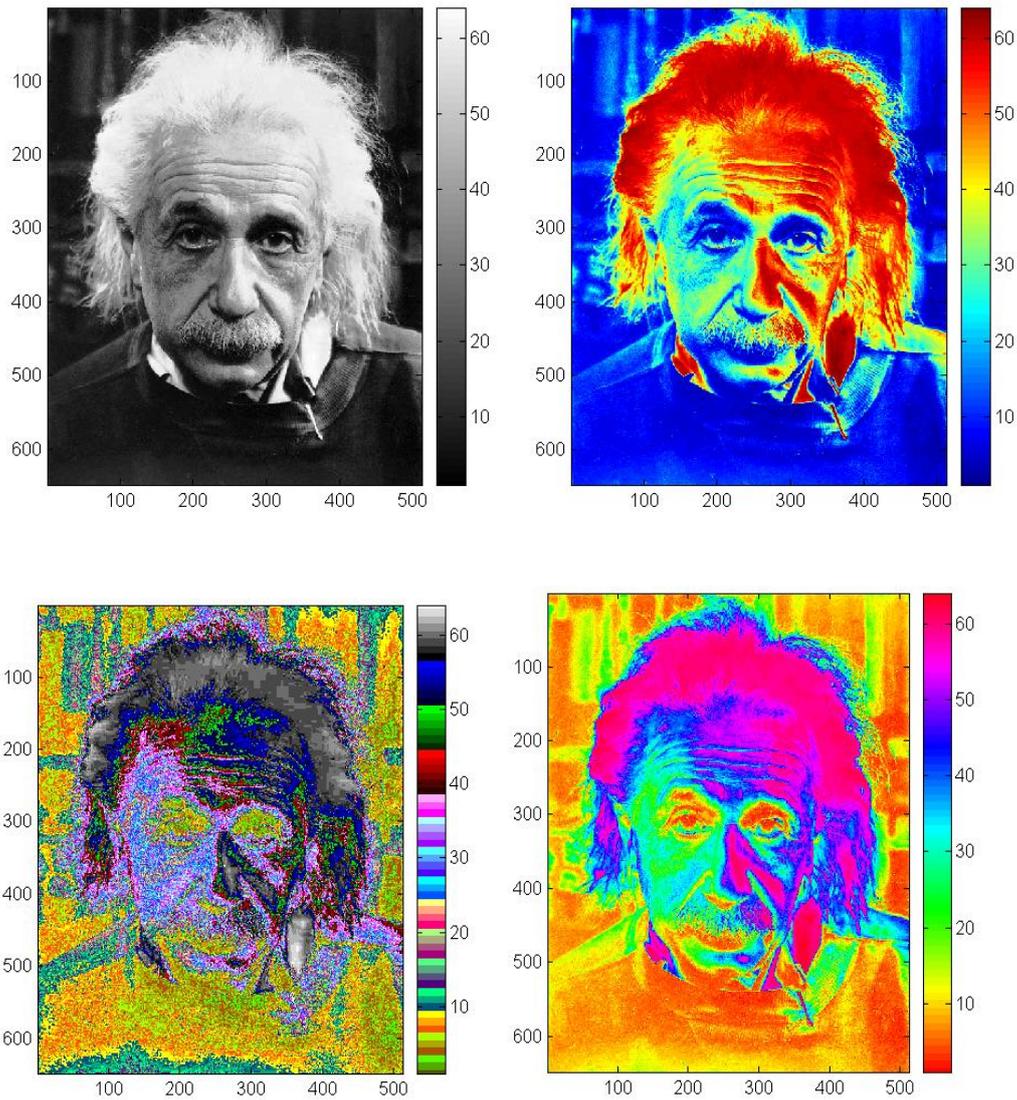


Figura 11-1

L'elaborazione delle immagini è un campo molto esteso, in cui si trasforma opportunamente un'immagine, al fine, per esempio, di evidenziare (o nascondere) certe caratteristiche³⁸.

Distinguiamo tre tipi di elaborazione:

- **puntuale** : ogni pixel viene modificato dipendentemente dal valore che ha
- **locale** : ogni pixel viene modificato dipendentemente dal valore che hanno quelli vicini (anche non solo gli adiacenti)
- **globale** : tutti i pixel vengono modificati nello stesso modo (per esempio sommando una costante)

Facciamo un paio di esempi di elaborazione puntuale.

³⁸ Il fine puo' anche essere artistico.

- il negativo : $g'_{ik} = \max(g_{ik}) - g_{ik}$

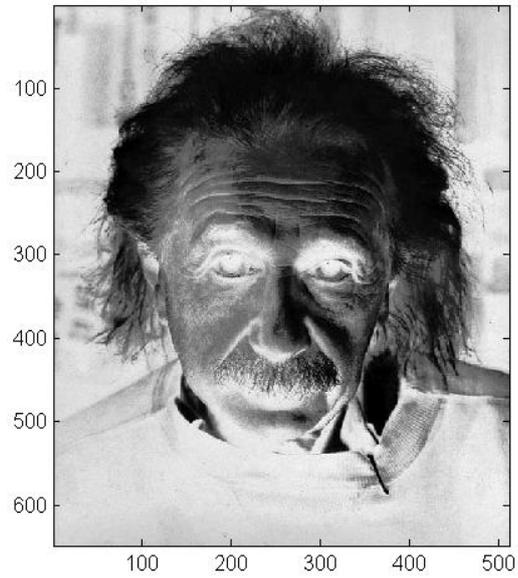


Figura 11-2

- alto contrasto : $g'_{ik} = \text{sign}(g_{ik} - \theta)$ (dove θ è una soglia)

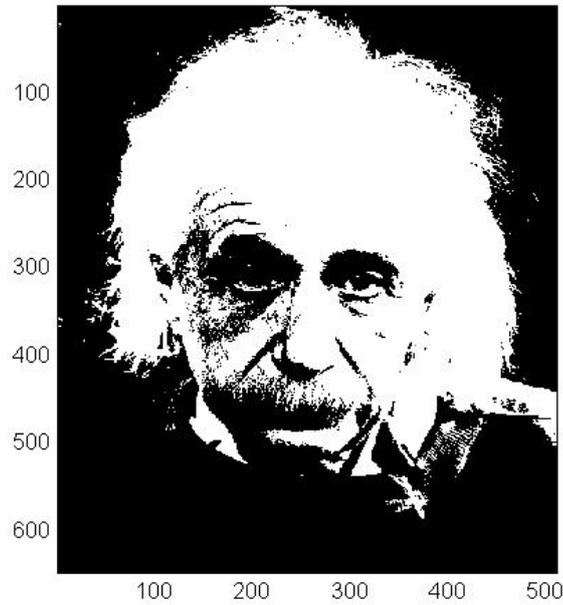


Figura 11-3

Ve diamo ora una tipica elaborazione “locale”:

- “**derivata**”: a ogni pixel è sottratto quello che lo precede (per esempio verticalmente)

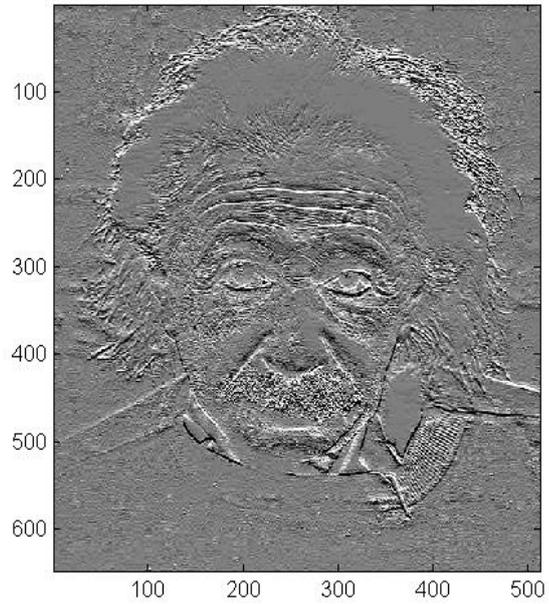


Figura 11-4

11.2 Elaborazione lineare delle immagini

Le immagini possono essere elaborate in modo analogo ad i segnali. La differenza è che al posto dell'ascissa temporale (o dell'indice "temporale" nel caso dei segnali discreti), abbiamo due ascisse spaziali ortogonali (ovvero due indici discreti) (vedi (10.1) e (10.2)).

Analogamente ai segnali temporali, possiamo introdurre l'operazione di convoluzione che nel continuo si definisce

$$(10.3) \quad a(x, y) \otimes b(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} a(\xi, \zeta) \cdot b(x - \xi, y - \zeta) \cdot d\xi \cdot d\zeta$$

Nel caso discreto si ha

$$(10.4) \quad a_{ik} \otimes b_{ik} = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} a_{mn} \cdot b_{(i-m)(k-n)}$$

Elaborazioni di una immagine del tipo della convoluzione vengono dette elaborazioni lineari.

Analogamente al caso dei segnali temporali, possiamo definire la trasformata di Fourier di una immagine $g(x, y)$ come

$$(10.5) \quad G(u, v) = \mathcal{F}(g(x, y)) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) \cdot e^{-j(u \cdot x + v \cdot y)} \cdot dx \cdot dy$$

(si suppone $g(x, y) = 0$ dove non esiste l'immagine). La sua inversa è

$$(10.6) \quad g(x, y) = \mathcal{F}^{-1}(G(u, v)) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u, v) \cdot e^{j(u \cdot x + v \cdot y)} \cdot dx \cdot dy$$

e, nel caso discreto, equivalentemente alla DFT,

$$(10.7) \quad G(\Phi, \Psi) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g_{ik} \cdot e^{-j(i \cdot \Phi + k \cdot \Psi)}$$

e la sua inversa

$$(10.8) \quad g_{ik} = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(\Phi, \Psi) \cdot e^{j(i \cdot \Phi + k \cdot \Psi)} \cdot d\Phi \cdot d\Psi$$

Possiamo definire l'analogo dello spettro di potenza come

$$(10.9) \quad S(u, v) = k \cdot |G(u, v)|^2$$

dove k è una costante; chiamiamolo semplicemente **spettro**, perché ovviamente non è uno spettro di potenza (né di energia). Per l'autocorrelazione abbiamo

$$(10.10) \quad R(\xi, \zeta) = \mathcal{F}^{-1}(S(u, v))$$

Facciamo un esempio. Ecco l'immagine di un cerchio

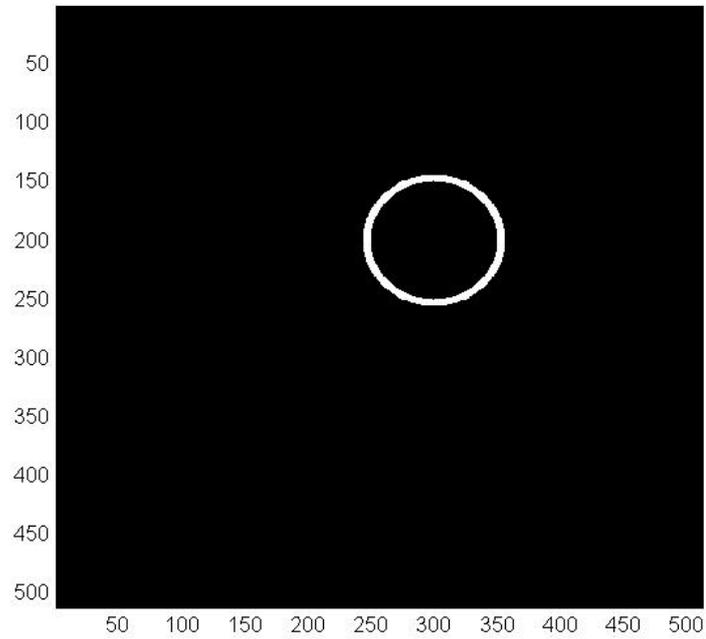


Figura 11-5

Ed ecco il logaritmo dello spettro

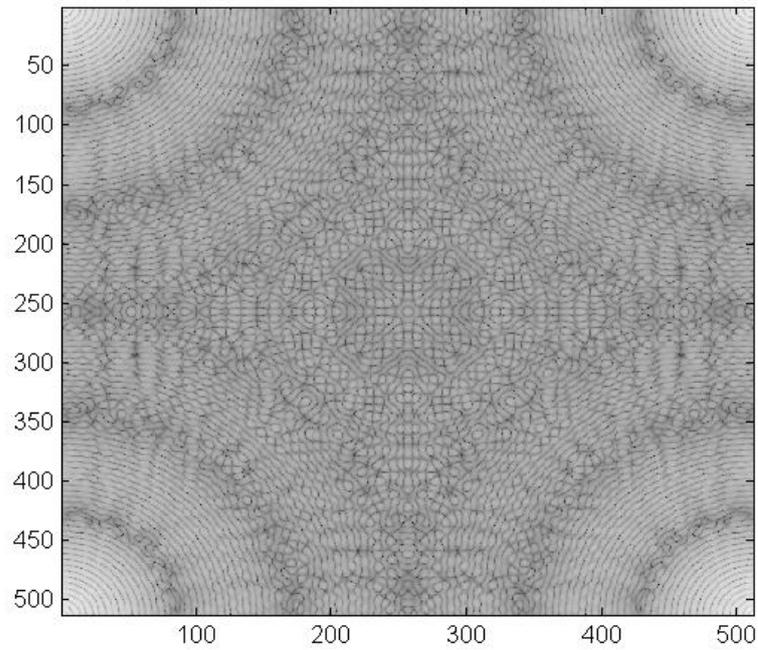


Figura 11-6

Si noti che questo “spettro” è indipendente dalla posizione del cerchio nella figura (anzi, sul “toro” costruito unendo i due lati verticali e quindi i due orizzontali).

Ed ecco l'autocorrelazione:

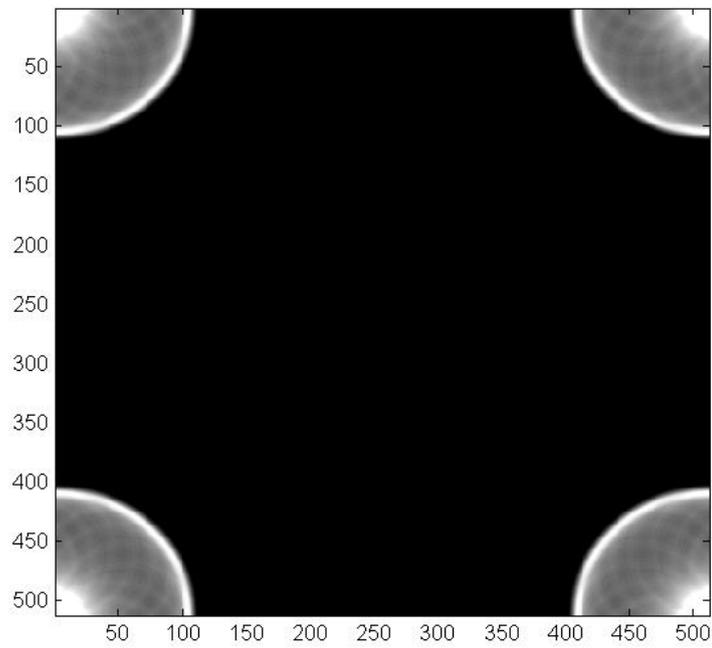


Figura 11-7

Vediamo ora come migliorare il rapporto segnale rumore in una immagine.

Costruiamo la seguente immagine

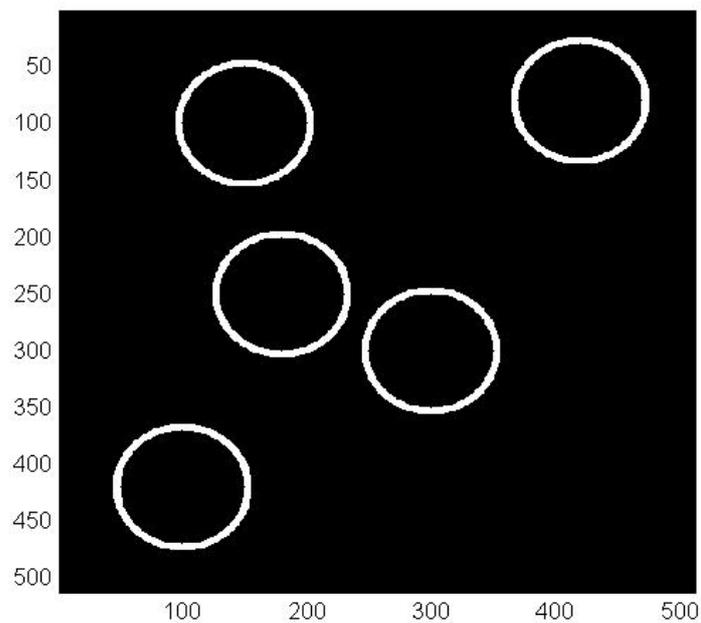


Figura 11-8

Sono 5 cerchi uguali a quello precedente di cui abbiamo fatto lo spettro.

Ora sommiamo a questa immagine del “rumore”, cioè a ciascun pixel una variabile normale. Se il rumore aggiunto è abbastanza “grande”, i cerchi scompariranno. Ecco il risultato avendo aggiunto rumore gaussiano con deviazione standard 4 volte maggiore:

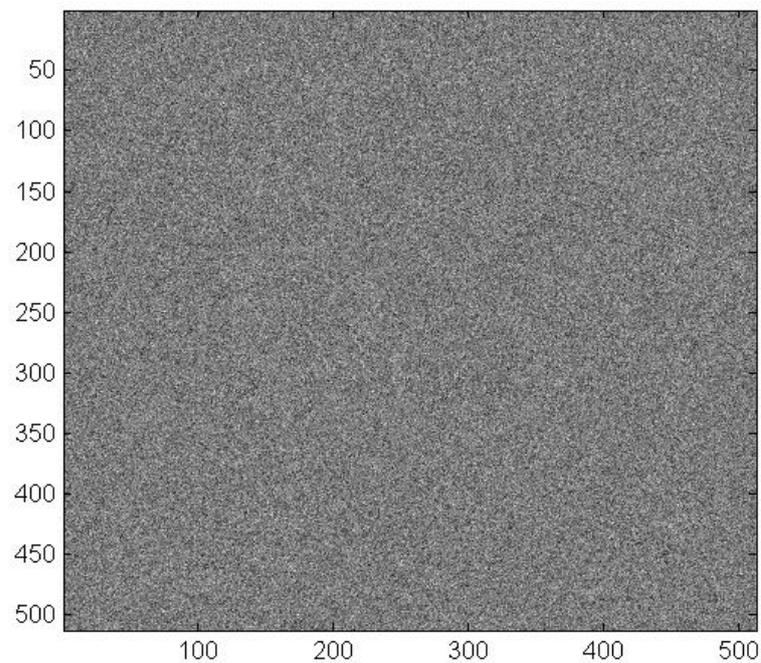


Figura 11-9

Ora costruiamo un semplice filtro così fatto:

- si prende la trasformata di Fourier dell'immagine disturbata
- si moltiplica per lo spettro del singolo cerchio
- se ne fa la trasformata inversa

Ed ecco il risultato:

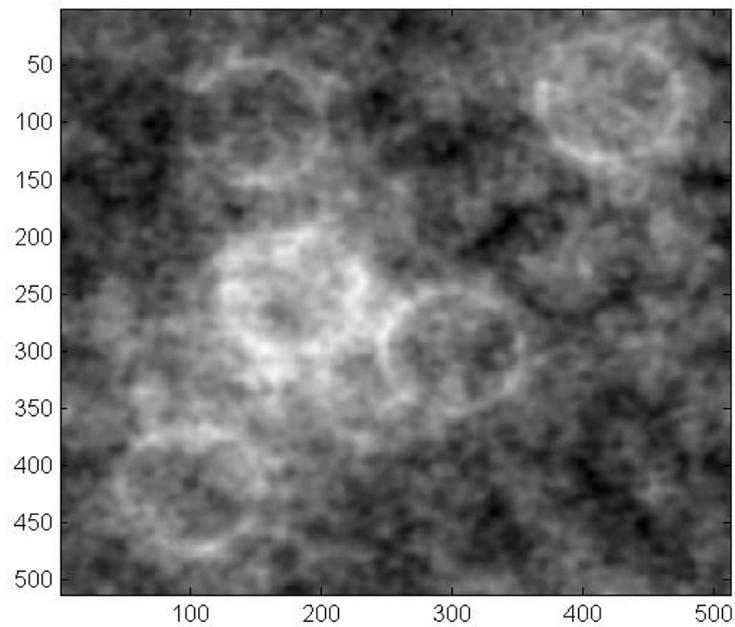


Figura 11-10

Come si vede, anche se disturbati, i cerchi sono ben visibili, nella posizione giusta.

Questo è un filtro piuttosto rozzo. Possiamo fare il filtro adattato (del tutto analogo a quello nel caso dei segnali), prendendo un cerchio nell'origine

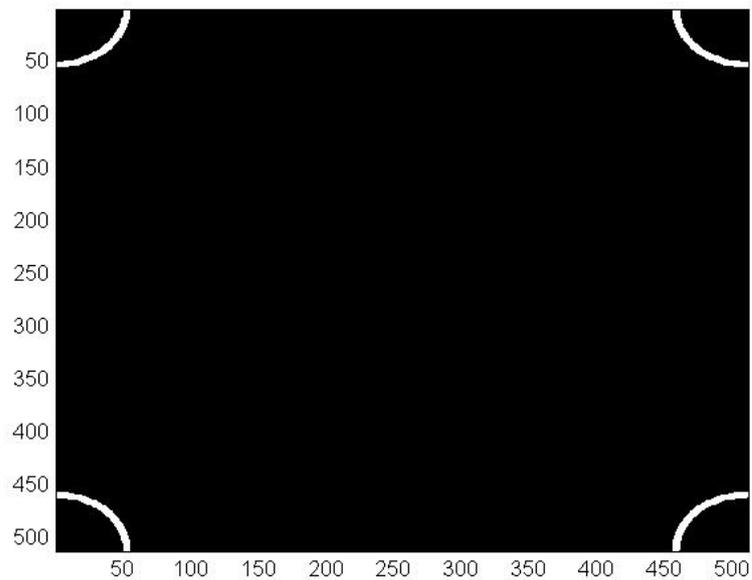


Figura 11-11

e quindi usarlo come “filtro adattato” ed applicarlo all'immagine, eventualmente lavorando con le trasformate di Fourier, cioè essendo M questa immagine e X l'immagine originale, otteniamo

$$(10.11) \quad Y = \mathcal{F}^{-1} \left[\mathcal{F}(X) \cdot \mathcal{F}(M)^* \right]$$

Otteniamo

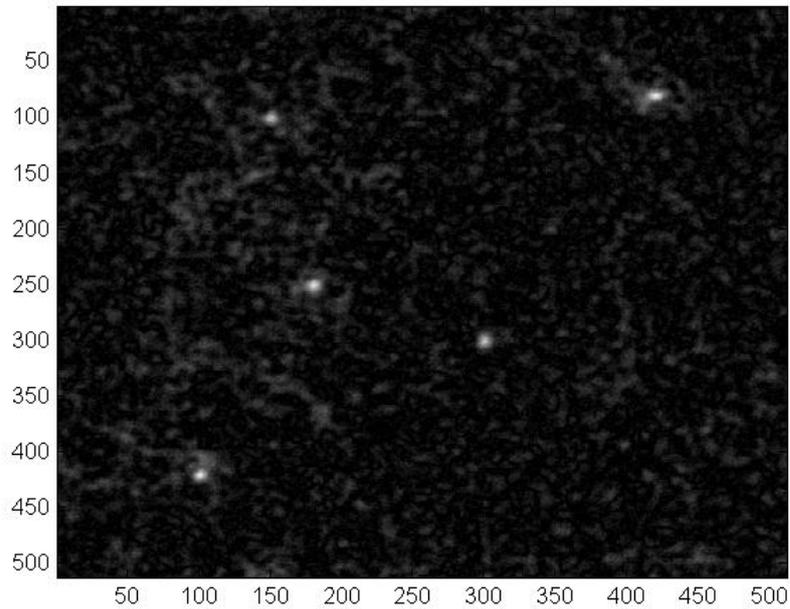


Figura 11-12

dove, come si vede, appaiono dei punti luminosi nei centri dei cerchi. Da questa, se il rapporto segnale-rumore è abbastanza buoni, possiamo riottenere l'immagine originale con

$$(10.12) \quad X' = \mathcal{F}^{-1} \left[\mathcal{F}(Y \cdot |Y|^3) \cdot \mathcal{F}(M) \right]$$

ottenendo

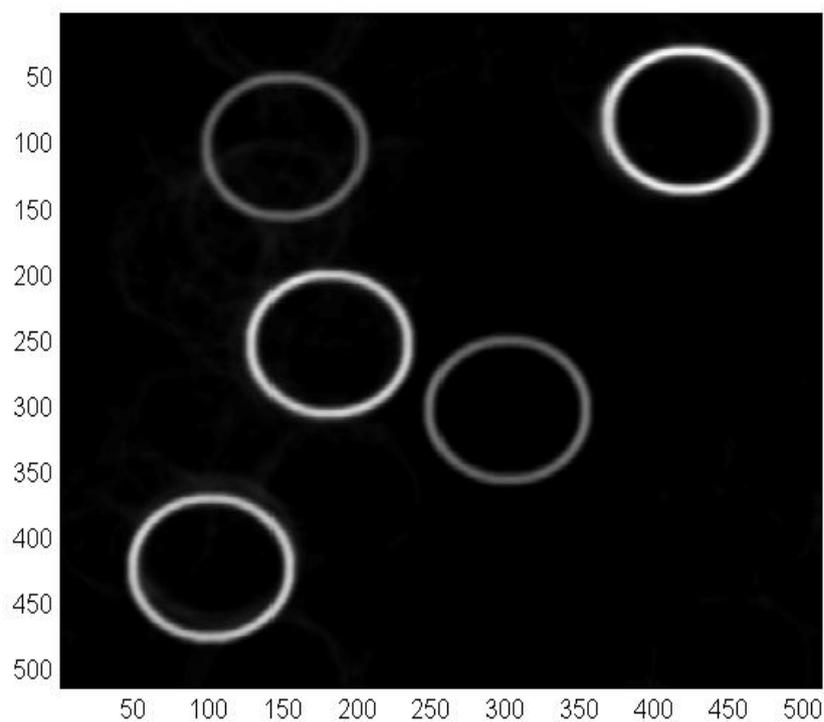


Figura 11-13

che è decisamente un buon risultato. Notare che quest'ultima operazione è non-lineare, e, come spesso accade con i metodi non lineari, dà buoni risultati solo se il rapporto segnale-rumore è abbastanza elevato. Se si operasse in modo lineare, se cioè si usasse

$$(10.13) \quad X' = \mathcal{F}^{-1}[\mathcal{F}(Y) \cdot \mathcal{F}(M)]$$

avremmo il risultato di Figura 11-10.

11.3 La compressione JPEG

12 Un'applicazione: la ricerca di periodicità

12.1 Introduzione al problema

La ricerca e l'analisi di periodicità in un segnale sono un problema che si presenta di frequente. Per affrontarlo sono stati messi a punto decine di metodi di analisi, alcuni dei quali analizzeremo in questo capitolo. Qui vogliamo mostrare alcuni dei modi in cui si presenta il problema.

Innanzitutto individuiamo tre tipi di periodicità:

- **periodicità coerente**, per cui

$$(11.1) \quad x(t) = x(t+T)$$

dove T è il "periodo". Si ha, di conseguenza

$$(11.2) \quad x(t) = x(t+k \cdot T)$$

dove k è qualsiasi intero.

Si parla di periodicità coerente anche nel caso in cui vari l'ampiezza ma non la fase. Per esempio nel caso di

$$(11.3) \quad x(t) = A(t) \cdot \sin\left(\frac{2\pi \cdot t}{T}\right)$$

dove $A(t)$ è una funzione positiva, in genere "lenta" (cioè con frequenze basse rispetto a $1/T$).

- **periodicità non coerente**, per cui

$$(11.4) \quad x(t) \approx x(t+T)$$

Ogni periodo, in questo caso, è quasi uguale al precedente, ma ci possono essere (in genere lievi) differenze di fase, che tuttavia si possono accumulare nel tempo. Un esempio di periodicità non coerente è un processo AR del secondo ordine.

- **periodicità variabile**, in genere "lentamente" (altrimenti viene meno l'idea di periodicità). In tal caso

$$(11.5) \quad x(t) = A(t) \cdot x(t+T(t))$$

dove $A(t)$ e $T(t)$ variano lentamente. $A(t)$ e $T(t)$ potrebbero essere note oppure potrebbe essere l'oggetto della ricerca

I problemi di analisi poi variano, a seconda di

- cosa è noto (per esempio il periodo, il periodo e la fase, la forma d'onda,...). Se fosse noto tutto, a parte l'ampiezza, teoricamente si potrebbe applicare il filtro adattato, ma raramente è ragionevole usare un filtro adattato per la ricerca di segnali periodici (almeno nella forma che abbiamo visto in 7.2)
- cosa cerchiamo (per esempio l'ampiezza, il periodo, il periodo e la fase, la forma d'onda,...)
- che tipo di dati abbiamo a disposizione
- con che precisione ci interessa il risultato
- quale è il rapporto segnale rumore

12.2 Il periodogramma

Lo strumento principale per analizzare le periodicità è la stima dello spettro di potenza tramite l'uso dei periodogrammi, di cui abbiamo parlato nel paragrafo 6.4.1. È questo sicuramente lo strumento migliore per un'analisi esplorativa.

Se il segnale è descritto da

$$(11.6) \quad x(t) = A \cdot \sin(\omega_0 t + \varphi)$$

e il rumore, nei dintorni della frequenza del segnale, ha spettro abbastanza piatto che vale S_n (bilatero), se sono trascurabili artefatti dell'analisi, abbiamo nello spettro stimato con un singolo periodogramma

$$(11.7) \quad SNR = \frac{A^2 \cdot T_{obs}}{4 \cdot S_n}$$

dove T_{obs} è il tempo di osservazione. Il rumore in tal caso, come sappiamo, è distribuito esponenzialmente.

Se facciamo la media di M periodogrammi (ciascuno lungo T_{obs} / M), abbiamo

$$(11.8) \quad SNR = \frac{A^2 \cdot T_{obs}}{4 \cdot \sqrt{M} \cdot S_n}$$

I vantaggi di questo strumento sono

- è molto veloce e si possono osservare molte periodicità contemporaneamente
- è affidabile, cioè solo in casi particolari sono generati artefatti e gli errori di stima sono sotto controllo

I problemi di questo strumento sono

- è difficile applicarlo a tantissimi dati (una FFT oggi può essere fatta senza difficoltà fino a qualche milione di dati)
- non si osservano variazioni nel tempo
- la risoluzione in frequenza è limitata (nella definizione più banale è M / T_{obs} , in effetti è in genere migliore di questo valore (se si usa FFT con soprarisoluzione), dipendentemente dal SNR.
- ignora la fase della periodicità

- è uno strumento non molto adeguato per la ricerca di periodicità non sinusoidali
- non è molto adeguato per periodicità non coerenti (a meno che non si riduca la risoluzione)

12.3 Metodi parametrici

Abbiamo accennato a questi metodi in 6.4.2. Per la ricerca di periodicità, questi metodi sono adeguati in casi non molto complessi. Permettono in genere migliori risoluzioni rispetto al periodogramma.

12.4 Il lock-in

Supponiamo di conoscere la frequenza di un segnale, e dobbiamo stimarne l'ampiezza e la fase. Uno strumento per questo scopo è il lock-in³⁹.

Dato un segnale $x(t)$, il lock-in produce il seguente segnale complesso⁴⁰

$$(11.9) \quad y(t) = \int_{-\infty}^t x(\theta) \cdot e^{-\frac{t-\theta}{\tau}} \cdot e^{j\frac{2\pi\theta}{T}} \cdot d\theta$$

qui T è il periodo che si suppone noto e τ è il “tempo d'integrazione”, cioè il tempo scala delle eventuali variazioni.

Il rapporto segnale/rumore è, nell'ipotesi di eq. (11.6),

$$(11.10) \quad SNR = \frac{A^2 \cdot \tau}{4 \cdot S_n}$$

e la distribuzione del modulo quadro del rumore è esponenziale.

Dal numero complesso $y(t)$ ricaviamo la stima dell'ampiezza del segnale

$$(11.11) \quad \hat{A} = |y(t)|$$

(attenzione ! è biasata per la presenza del rumore; il bias va tolto quadraticamente) e la sua fase

$$(11.12) \quad \hat{\varphi} = \angle y(t)$$

Si noti che entrambe le stime sono variabili nel tempo, e τ è il tempo scala della più rapida variazione osservabile.

Una forma diversa di lock-in è la seguente

$$(11.13) \quad y(t) = \int_0^t x(\theta) \cdot e^{j\frac{2\pi\theta}{T}} \cdot d\theta$$

Come si vede, la costante di tempo è scomparsa, e l'integrale si estende dall'inizio delle misure fino al tempo t . Si può graficare la $y(t)$ nel piano di Gauss. Chiamiamo questa figura il “verme”, per l'aspetto che assume nel caso di periodicità di periodo T .

³⁹ Con questo nome viene indicato un particolare amplificatore elettronico. Qui noi indichiamo l'algoritmo utilizzato da questo amplificatore.

⁴⁰ L'amplificatore elettronico produce due segnali separate, detti “segnale in fase” e “segnale in quadratura”. Noi li vedremo come un singolo segnale complesso.

Ciò può essere molto utile per vedere variazioni di frequenza. Se la frequenza del segnale è mediamente diversa da $\nu = \frac{1}{T}$, allora il “verme” gira intorno all’origine: dal numero di giri che compie (e dal verso) si deduce la differenza tra la frequenza “vera” e $\nu = \frac{1}{T}$.

Vediamo ora il “verme” per una realizzazione di un processo AR(2) con frequenza 100 Hz e $Q=10000$, osservato a 100 Hz:

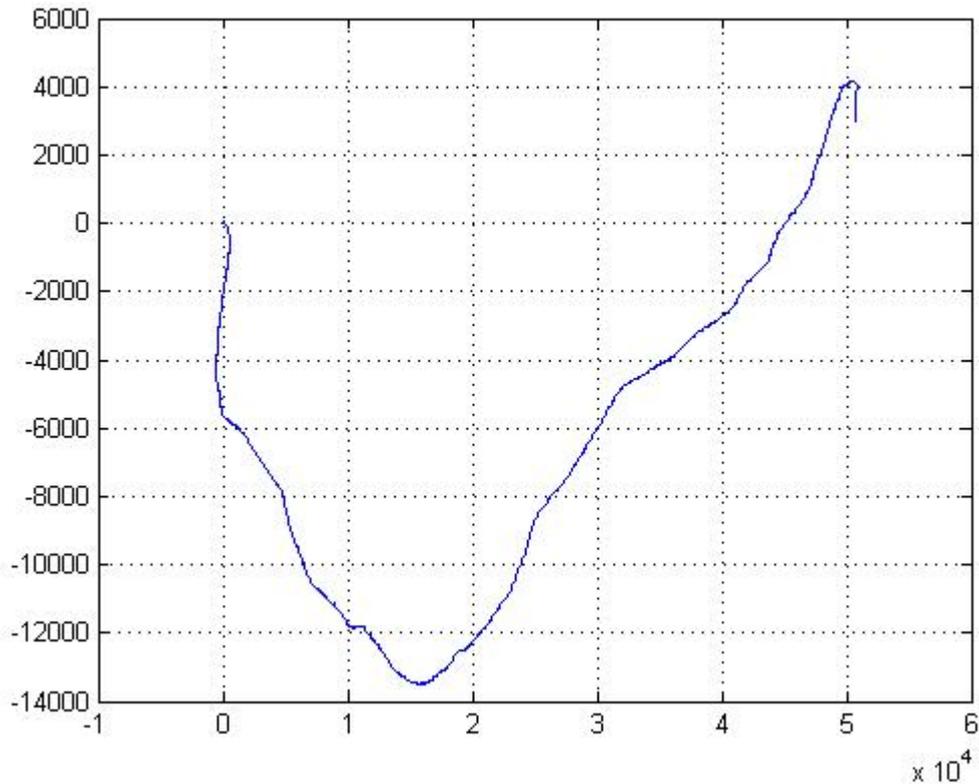


Figura 12-1

e ecco cosa avviene se si “sintonizza” il verme a 100.1 Hz (sugli stessi dati)

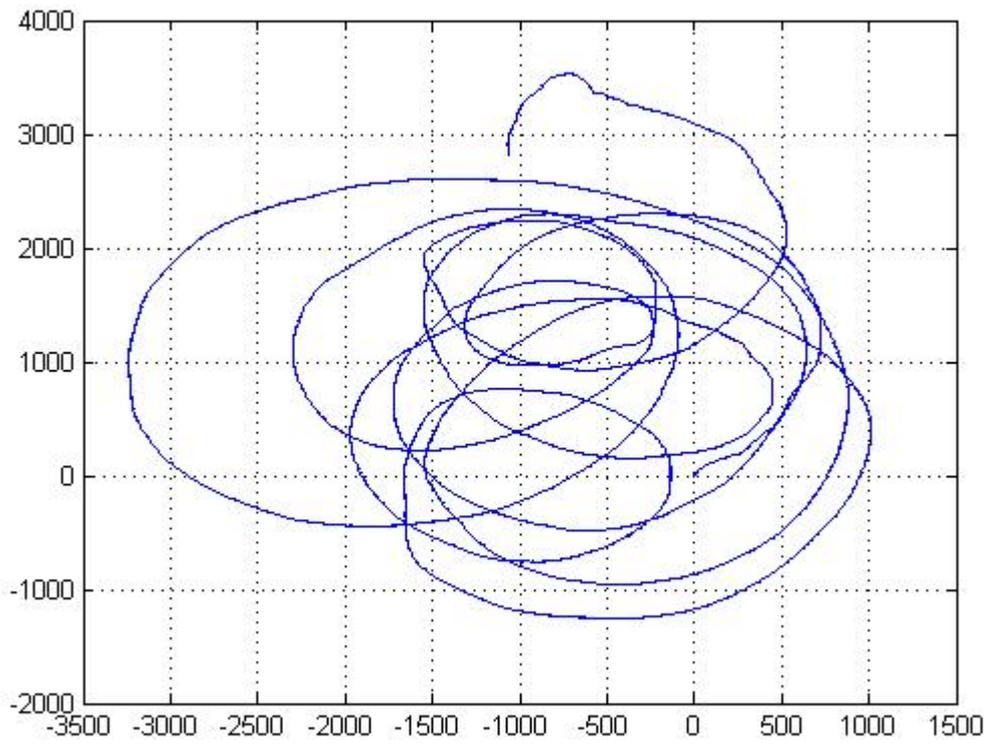


Figura 12-2

12.5L'analisi di fase

Talora si conosce il periodo di un fenomeno, ma siamo interessati a come ne va l'ampiezza all'interno del periodo.

Per esempio è questo il caso in cui siamo interessati al rumore di fondo in un laboratorio, sappiamo che questo è correlato con l'attività umana e quindi ha una periodicità giornaliera, ma vogliamo sapere come varia nelle varie ore del giorno.

Un altro esempio è il segnale delle pulsar: è questo un segnale periodico ed è di grande importanza conoscerne il profilo. Ecco alcuni diagrammi di fase per il segnale della pulsar della nebulosa del Granchio (Crab pulsar):

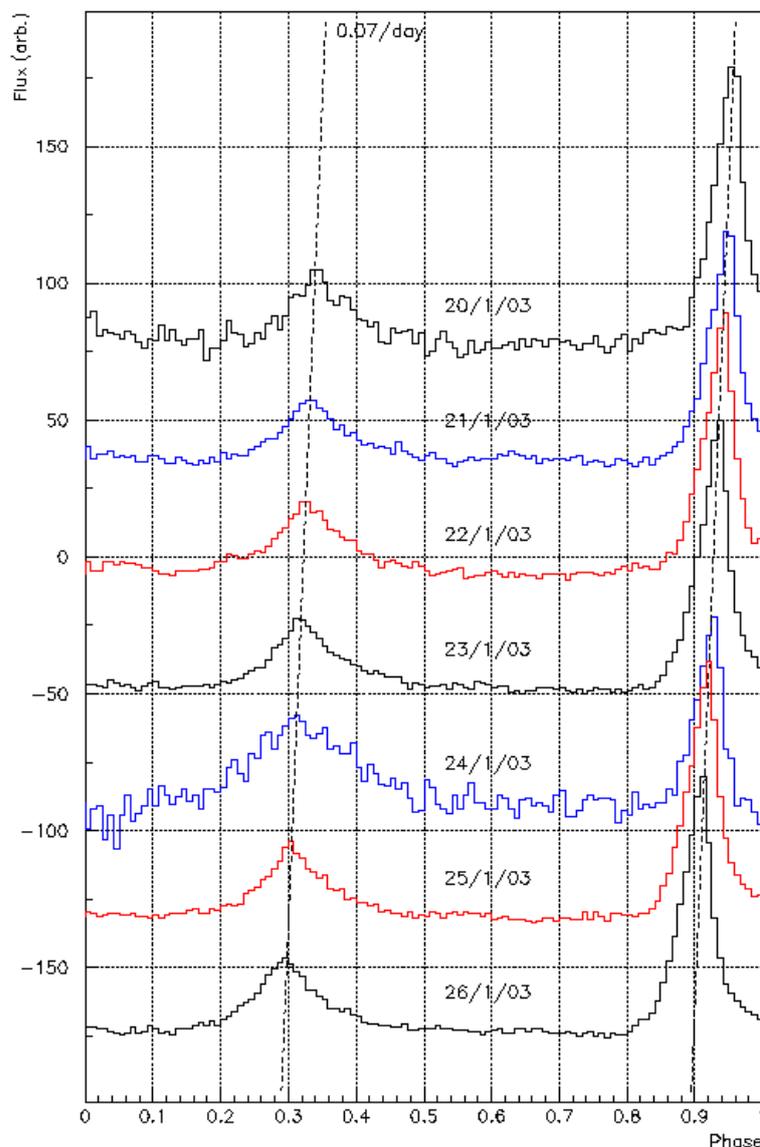


Figura 12-3

Questo tipo di grafici si eseguono mediando in fase i vari periodi del segnale.

12.6 Il fit armonico

Un altro metodo di analisi di periodicità non sinusoidali è il fit del segnale armonico con

$$(11.14) \quad \sum_{k=1}^n \left(a_k \cdot \cos\left(\frac{2\pi \cdot k \cdot t}{T}\right) + b_k \cdot \sin\left(\frac{2\pi \cdot k \cdot t}{T}\right) \right)$$

cioè trovare i coefficienti a_k e b_k che meglio descrivono il segnale (per esempio secondo il criterio dei minimi quadrati). Il segnale è quindi descritto da una $2n$ parametri. È buona norma calcolare anche l'incertezza su questi parametri.

Se T non è noto, si può fare un'indagine con vari valori di T , in un range ragionevole, e graficando l'inverso dell'errore quadratico medio, facendo così una specie di spettro. Questo metodo, anche per $n=1$, ha una migliore risoluzione rispetto al periodogramma (anche se è molto più costoso computazionalmente).

Se T è noto, si misurano in pratica le varie armoniche del diagramma di fase.

12.7 Il caso del campionamento non uniforme

Talvolta i dati non sono campionati in modo uniforme: le osservazioni possono essere sporadiche (come per esempio può capitare in astronomia).

In altri casi può essere utile un campionamento non uniforme per superare i limiti della frequenza di Nyquist.

In questo caso, se le osservazioni sono $\{x_1, x_2, \dots, x_N\}$, una stima dello spettro di potenza è

$$(11.15) \quad \hat{S}(\omega) = \frac{1}{N} \cdot \left| \sum_{k=1}^N x_k \cdot \exp(j\omega \cdot t_k) \right|^2$$

Facciamo un esempio. Campioniamo un segnale di forma

$$(11.16) \quad x(t) = \sin(2\pi \cdot 900 \cdot t) + \cos(2\pi \cdot 211 \cdot t)$$

consideriamo prima il caso di 100 campioni uniformi in un secondo. In tal caso è evidente che le condizioni del teorema del campionamento non sono soddisfatte per le due frequenze di 211 e 900 Hz (la frequenza di Nyquist è 50 Hz) e infatti ecco lo spettro:

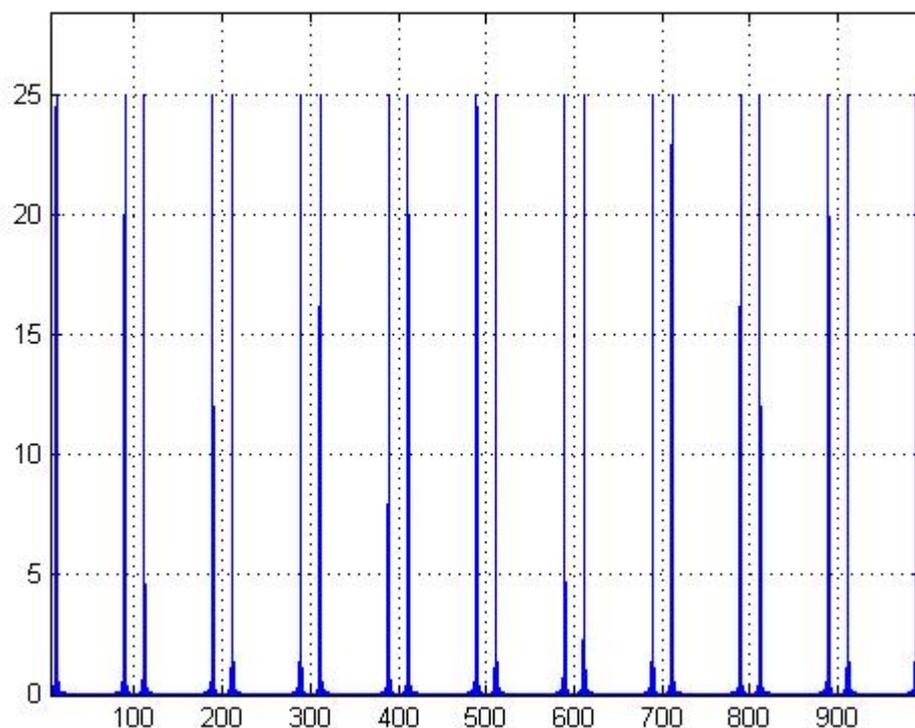


Figura 12-4

Si vedono varie frequenze aliasate più volte. Tra queste compare 211 Hz, ma non 900 (che è aliasata a 0 perché è un multiplo pari della frequenza di Nyquist).

Ed ecco invece lo spettro ottenuto con la (11.15), avendo preso 100 istanti di campionamento a caso (distribuzione uniforme) in un secondo

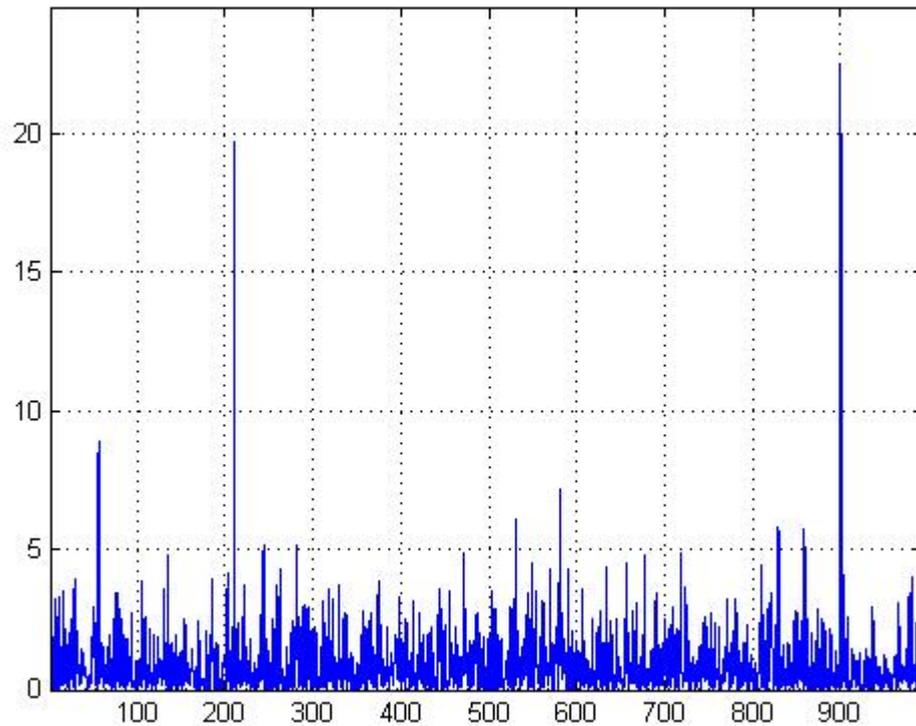


Figura 12-5

Come si vede, le due frequenze compaiono al valore giusto (anche se c'è un fondo casuale che è facile dimostrare aver distribuzione esponenziale con valore atteso uguale alla varianza del segnale).

Se i tempi di campionamento hanno delle periodicità, queste si riflettono su $\hat{S}(\omega)$.

12.8 Periodicità non coerenti

Se si ha una periodicità non coerente, il modello più semplice è quello di un processo gaussiano del secondo ordine (nel discreto, un AR(2)).

In tal caso ci sono 4 parametri da stimare:

- la **frequenza**
- il **tempo di decadimento** τ
- la **varianza**
- il **rumore di fondo** su cui è “seduto” il picco spettrale

Un modo per far ciò è:

- estrarre il pezzo di spettro che contiene il picco
- aggiungere una seconda parte di eguale lunghezza e porla eguale a 0
- antitrasformare questo spettro sintetico, ottenendo un'autocorrelazione complessa “analitica”
- dedurre la sua frequenza (ci sono vari modi, per esempio contare il numero di giri nel piano di Gauss) e da questa sommata alla frequenza iniziale del pezzo ricavare la **frequenza** cercata
- dall'involuppo (semplicemente il valore assoluto) dell'autocorrelazione vicino a 0 ricavare il **rumore di fondo** (dal picchetto ripido vicino a 0), il **tau** (dalla pendenza della discesa “lenta”) e la **varianza** (l'estrapolazione a 0 della discesa “lenta”).

12.9 Periodicità di eventi

Talvolta il problema è quello di trovare le periodicità di impulsi. Di questi non è di interesse la forma o l'ampiezza, ma solo il tempo di occorrenza.

Il modello è quello di un processo di Poisson con λ variabile in modo periodico nel tempo. È questo uno dei casi a cui fa riferimento l'equazione (5.83).

Essendo t_i gli istanti di tempo in cui sono capitati gli impulsi, costruiamo in questo caso lo **spettro impulsivo**

$$(11.17) \quad S(\omega) = \frac{1}{N} \cdot \left| \sum_{k=1}^N \exp(\omega \cdot t_i) \right|^2$$

Onde evitare errori di arrotondamento è bene sostituire nella (11.17) $(t_i - t_1) \rightarrow t_i$.

A parte che a bassissime frequenze, se gli impulsi hanno densità uniforme $S(\omega)$ ha distribuzione esponenziale con valore aspettato 1.

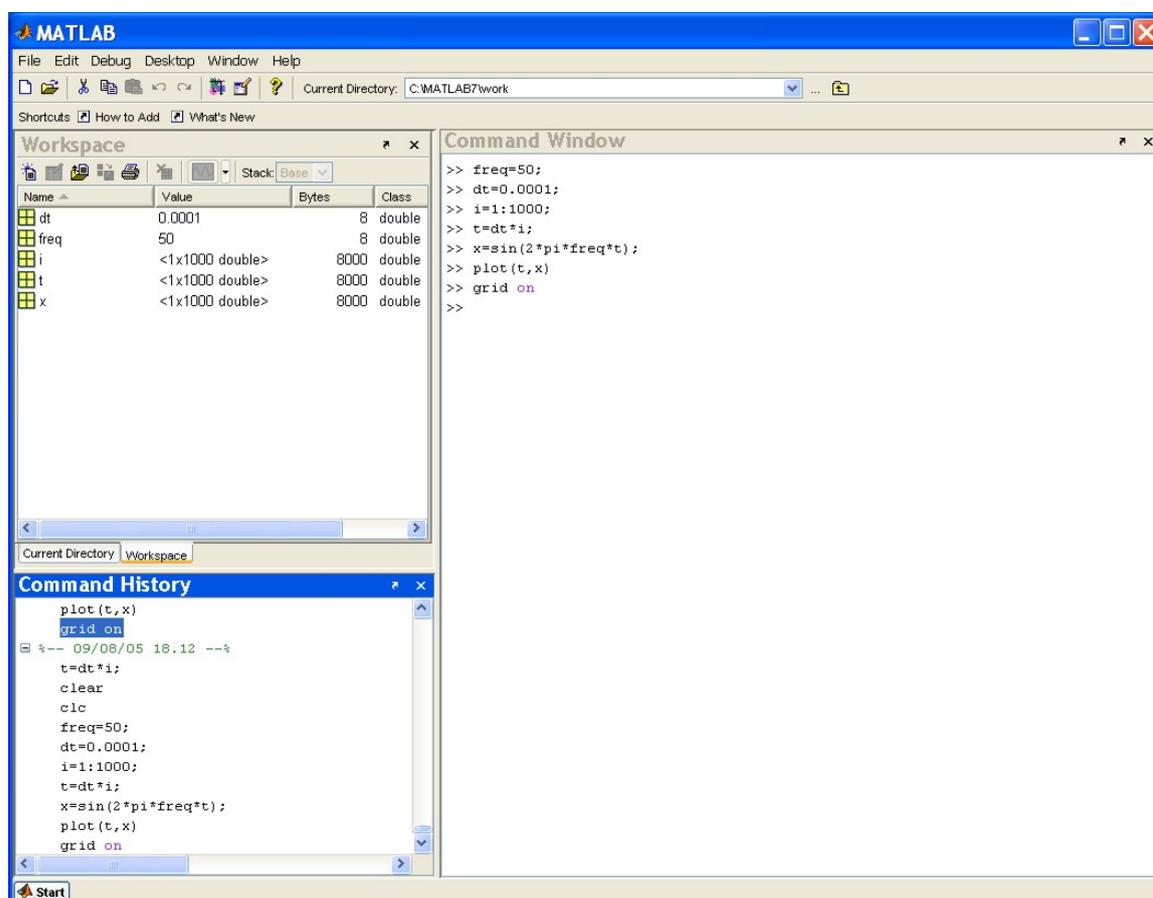
Anche nel caso degli eventi possono farsi diagrammi di fase, fit armonici, eccetera.

13 Appendici

Esercitazioni con Matlab

Introduzione a Matlab

Matlab è un software che comprende un linguaggio di programmazione di alto livello per problemi tecnici e scientifici e un ambiente interattivo per l'analisi e lo sviluppo di programmi (vedi figura).



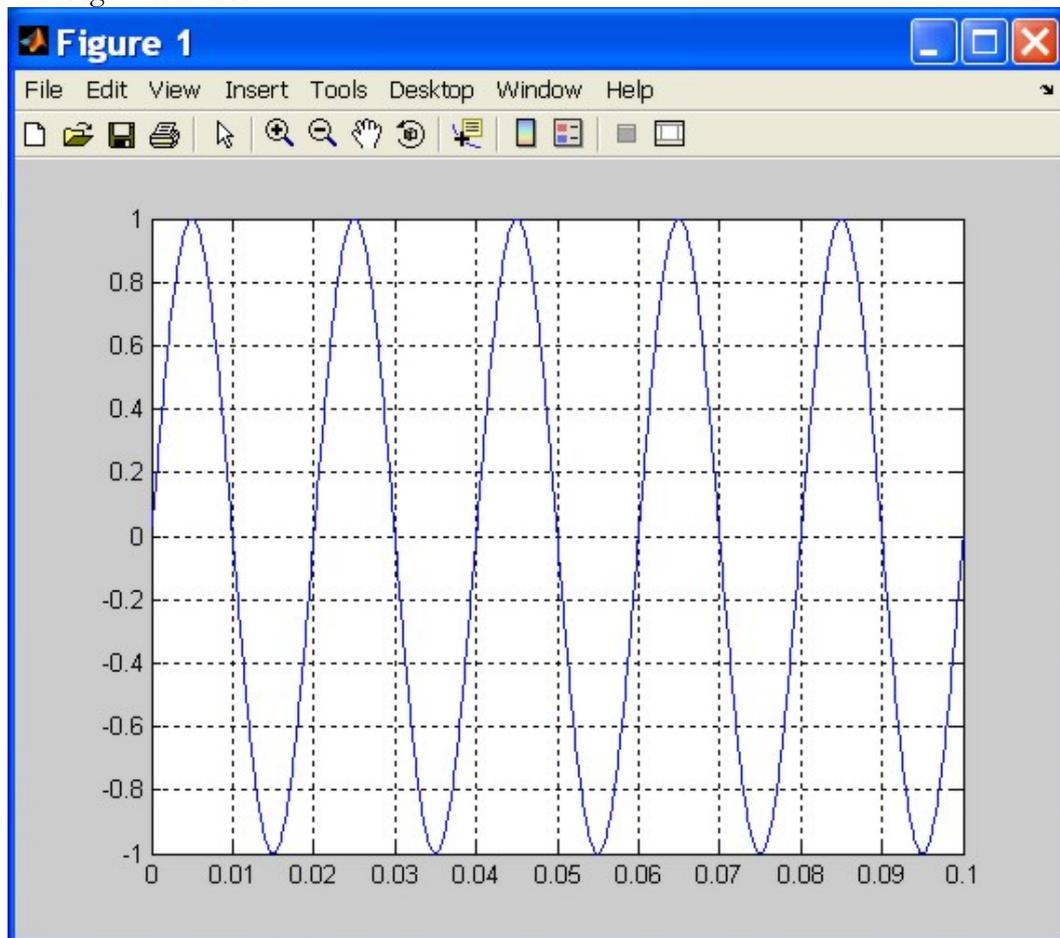
Per esempio, se nella finestra interattiva si digita (i “>>” sono il prompt)

```
>> freq=50;
>> dt=0.0001;
>> i=1:1000;
>> t=dt*i;
>> x=sin(2*pi*freq*t);
>> plot(t,x)
```

assegna 50 alla variabile freq
assegna 0.0001 alla variabile dt
crea un vettore i con i numeri da 1 a 1000
moltiplica i valori di i per dt
crea un vettore x con i valori del seno
crea un grafico con t ascissa e x ordinata

```
>> grid on          aggiunge una griglia al grafico
>>
```

appare la seguente finestra:



Si noti che se un'istruzione non viene terminata da un ";" appare immediatamente il risultato:

```
>> a=2*8
a =
    16
>> i=1:0.5:2
i =
    1.0000    1.5000    2.0000
>>
```

I comandi possono essere messi in file (con estensione .m) che possono essere eseguiti semplicemente digitando il nome del file (senza l'estensione). Possono mettersi in file anche "functions" (analoghe, con qualche differenza, alle subroutine del Fortran e alle funzioni del C).

Nell'ambiente interattivo è presente un esteso Help, con la descrizione delle migliaia di funzioni interne e tutorial.

Sono disponibili (con licenza a parte) numerosi toolbox per specifici ambiti applicativi (per esempio la statistica, l'analisi dei segnali, le equazioni differenziali alle derivate parziali, l'immagine processing e così via).

Matlab è prodotto dalla MathWorks (<http://www.mathworks.com/>) e nel sito sono disponibili ottimi manuali d'uso.

Esistono “cloni” gratuiti di Matlab. I due migliori sono

- **Scilab** (<http://scilabsoft.inria.fr/>) sviluppato a partire dal 1990 dall'INRIA, un istituto nazionale di ricerca francese
- **Octave** (<http://www.octave.org/>) che però gira solo sotto Linux (si può usare sotto Windows usando Cygwin, l'ambiente Linux-like di Windows sviluppato dalla Cygnus (<http://www.cygwin.com/>)).

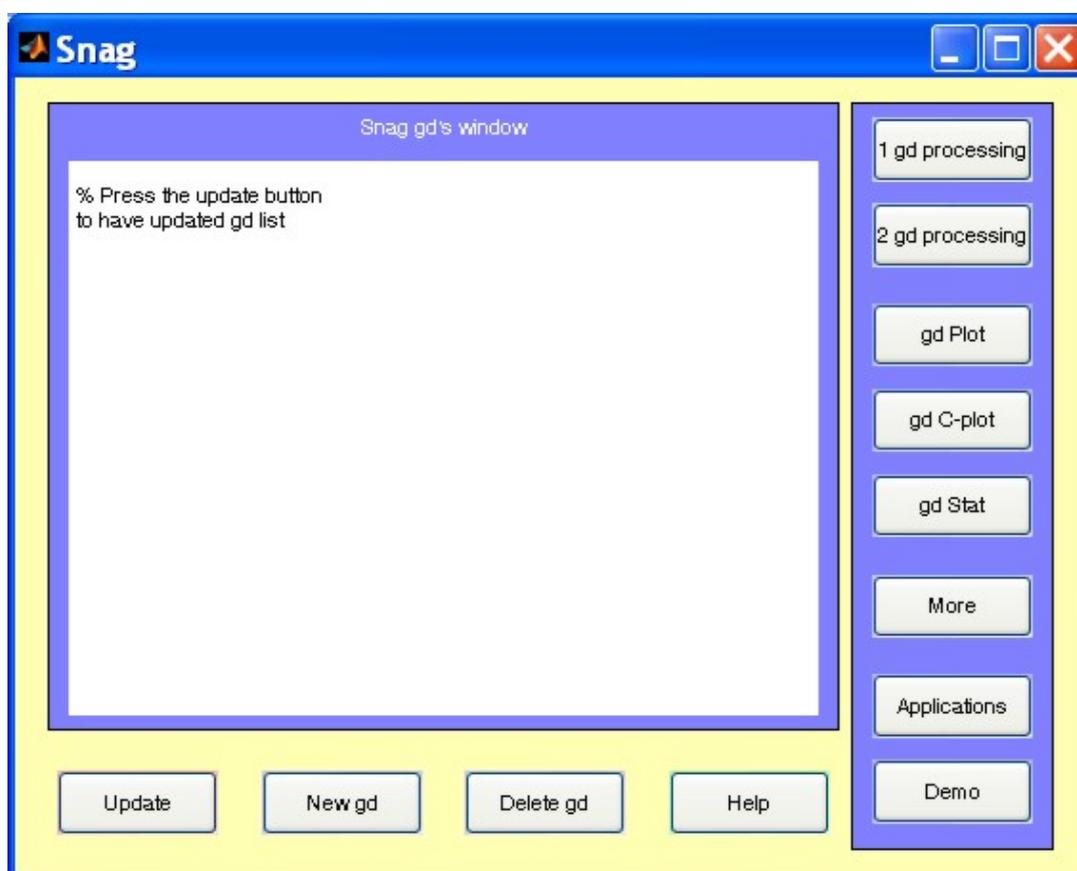
Snag

Snag è un toolbox di Matlab sviluppato presso il gruppo di ricerca sulle onde gravitazionali del dipartimento, per l'analisi dei dati e la simulazione dei dati delle antenne gravitazionali. Comprende oltre 800 funzioni Matlab e alcuni strumenti grafici interattivi.

Può essere scaricato dal sito <http://grwavs.f.roma1.infn.it/snag/default.htm> dove sono disponibili guide di uso e di programmazione.

È stato sviluppato tramite programmazione a oggetti e la classe più importante è il **gd** (gruppo dati), che descrive un insieme di dati, per ciascuno dei quali c'è un'ascissa (per esempio temporale) e un'ordinata.

Lo strumento interattivo basilare si invoca digitando al prompt `snag` e appare la finestra



Per ulteriori dettagli si rimanda alle guide disponibili nel sito.

Uso di SnagLab (versione VB)

Introduzione

SnagLab è un programma orientato all'analisi dei dati sperimentali.

I dati possono essere letti da file o immessi in modo interattivo.

La versione attuale è scritta in Visual Basic 6, in ambiente Windows (Windows 98 e successivi); è in sviluppo una versione in C#.

Alla partenza del programma viene creato nel folder temporaneo dell'utente un file di log SnagLab.log con varie informazioni e parametri calcolati.

Istallazione

L'ultima versione del programma è scaricabile da grwavs.f.roma1.infn.it/snaglab. Per installarlo, creare una cartella (per es. SnagLab) e metterci l'eseguibile e tutti gli eventuali file di supporto (dll e ocx) che dovessero mancare al sistema.

Uso

I GD

I dati sono immagazzinati in strutture chiamate GD (gruppi dati); questi sono numerati, a partire dal GD(0) che è il GD dove normalmente “entrano” i dati. I dati del GD(0) possono essere spostati in altri GD.

Esistono tre “gusti” di GD, a seconda del tipo di dati:

- **standard data**, in cui ogni dato è caratterizzato da ascissa, ordinata e loro incertezze
- **equal uncertainties**, in cui le incertezze sono le stesse per tutti i dati
- **sampled data**, in cui l'ascissa è virtuale, poiché si suppone che siano presi a tutti i multipli del tempo di campionamento. Questo è utile per i segnali.

La finestra “GD manager” mostra lo stato dei GD.

In ogni momento esiste un GD “a fuoco”, su cui si opera automaticamente. Il GD a fuoco è in genere l'ultimo considerato. Si può modificare il GD a fuoco 256ompatta su una casella del GD manager o modificando le apposite caselle in alcune finestre di dialogo.

Altre strutture dati

MD: (Multiple GD) strutture dati contenenti più GD con le stesse ascisse.

DM: (Data Map: GD a due dimensioni) strutture dati con ascissa a due dimensioni (x,y). La dimensione x può essere “reale” (campionamento non uniforme) o “virtuale” (campionamento uniforme); la dimensione y è sempre virtuale. I DM sono trattati in modo analogo ai GD: c'è un DM manager, a cui si accede dal GD manager, da cui si possono visualizzare, come matrici o come mappe.

Alcune delle procedure di SnagLab sono specifiche dei DM, alcune possono applicarsi ai DM e ai GD.

Interfaccia iniziale

La finestra iniziale è composta da tre parti:

- l'albero delle procedure disponibili; le procedure si attivano 257ompattarl sopra. Non tutte le procedure sono già attive; alcune procedure, come i fit, sono più comodamente attivabili dall'apposito menu dei grafici.
- la finestra delle informazioni; quando ci sono GD in uso, contiene un preview del GD "a fuoco"; se è "a fuoco" un GD non usato, viene visualizzato questo file.
- i tasti per accedere al GD manager, alla finestra di input standard dei dati e al Setup (anche il setup contiene alcune caratteristiche non ancora o non ancora bene implementate).

Input dei dati

I dati possono essere immessi e modificati tramite l'interfaccia interattiva: occorre cliccare sulla griglia, alla posizione desiderata e modificare le caselle della pop-up window che appare. (È in sviluppo un modo di input direttamente da griglia, selezionabile dal set-up, ma non è ancora consigliabile). Notare che i dati immessi possono essere attivi o no: ciò è utile per esempio per fare grafici o fit parziali. I dati così prodotti sono salvabili su file.

I dati possono essere scritti in file leggibili da SnagLab. Attualmente esistono tre formati diversi:

- **UD**, per dati standard con incertezze individuali, così composto:

```
#UD
<stringa di didascalia>
<numero di dati>
<x1> <dx1> <y1> <dy1> <act1>
<x2> <dx2> <y2> <dy2> <act2>
.....
```

dove x_n , dx_n , y_n e dy_n sono ascissa, sua incertezza (assoluta), ordinata e sua incertezza, per ogni punto; act è una flag di attivazione che vale 1 o 0.

- **MD**, per dati multipli, su più colonne, una per un'ordinata diversa, così composto:

```
#MD
<numero ordinate (n)> <numero di dati>
<incertezza ascissa> <incertezza ordinata>
<stringa di didascalia ordinata (colonna) 1>
<stringa di didascalia ordinata 2>
....
<stringa di didascalia ordinata n>
<x1> <y11> <y12> ... <y1n>
<x2> <y21> <y22> ... <y2n>
.....
```

- **SD**, così composto:

```
#SD
<stringa di didascalia>
<numero di dati> <tempo iniziale> <tempo di campionamento>
```

<y1> <y2> ...

I dati possono continuare su più righe, come si preferisce.

I dati nei file possono essere letti con la Open del menu file della finestra “Standard Data Input” o 258ompatta sull'icona open file della task bar. Se si apre un file con formato MD, viene chiesto se si vuole leggerlo lo stesso anche se non è UD, quale colonna si vuole leggere; se si risponde 0, le legge tutte e le mette in GD successivi, a partire da quello richiesto dall'utente (ATTENZIONE ! la prima colonna va anche nel GD 0).

Altri formati, in lettura e scrittura, saranno supportati.

Grafico

Si può graficare il GD “a fuoco”. Quindi ogni grafico ha un GD “padrone”. Possono tuttavia essere visualizzati sullo stesso grafico altri GD, tramite il menu “New input”.

Dell'ultimo GD immesso nel grafico può farsi il fit (in vari modi).

Può essere creato un nuovo grafico (con l'ultimo GD immesso), con nuove scelte di range e di tipo per le coordinate; il tipo delle coordinate è normale, logaritmico, quadrato, radice quadrata, inverso, per entrambi gli assi. Il fit è visualizzabile correttamente solo per coordinate normali o logaritmiche.

Fit

Sono attivabili vari tipi di fit per i dati presenti. In particolare:

- fit lineari, generici o con un punto fisso o con fissata pendenza
- fit polinomiali (fino al sesto grado, ma si può aumentare il limite col set-up)
- fit di funzioni di trasferimento (passa basso, passa alto e passa banda), fino al sesto ordine..

Dai fit lineari e polinomiali si può fare, attivando le flag di conversione logaritmica, fit di potenza, esponenziali e logaritmici. Nel caso del fit esponenziale si calcola anche il tau, nel caso del fit di potenza si calcola anche l'esponente.

È visualizzabile la zona di piano che comprende il range di variazione dei parametri (senza l'imposizione della dovuta correlazione tra i coefficienti stimati).

Per quanto riguarda i fit di funzioni di trasferimento, si noti che il passa alto e il passa basso, a partire dal secondo ordine, anche le risonanze. Per il primo e il secondo ordine vengono calcolate frequenze di taglio, frequenze di risonanza, tau e Q.

Analisi statistiche

Sono attivabili semplici analisi statistiche del singolo GD o il coefficiente di correlazione (ed eventualmente lo scatter-plot, cioè un GD con ascissa l'ordinata del primo GD e come ordinata quella del secondo) di due GD.

È possibile fare il test del chi-quadro tra due GD o tra un GD e il fit calcolato.

Analisi dei segnali

Le operazioni di analisi dei segnali sono orientate ai GD di tipo *sampled data*. Sono previsti:

- il calcolo dello spettro di potenza
- il calcolo dell'autocorrelazione (attivabile o dalla finestra dello spettro di potenza o come caso particolare della correlazione incrociata)
- il calcolo della correlazione incrociata (cross-correlation) tra due GD
- applicazione di vari filtri digitali, in modo causale, anti-causale o bifronte

Calcoli (menu “Computations”)

Può calcolarsi l'integrale definito (in un intervallo di indici o di ascissa) di un GD.

Per i GD che rappresentano istogrammi o densità di probabilità, si possono calcolare media, deviazione standard, integrale totale, asimmetria e curtosi.

Varie operazioni sui GD

Ai GD possono essere applicate varie procedure. A parte la derivata e l'integrale (eseguito con semplice sommatoria), ci sono:

- operazioni sugli assi x e/o y (quadrato, radice quadrata, logaritmo decimale, logaritmo naturale, inverso,...)
- operazioni lineari su due GD ($a*GD1+b*GD2+c$) [se si vuole, si può ignorare il secondo, con $b=0$, ma non il primo]
- operazioni non-lineari su due GD (prodotto, divisione, minimo, massimo)
- operazioni su GD complessi (il supporto ai GD complessi è, per ora, limitato)
- integrali, derivate, convoluzioni, FFT (che può produrre un GD complesso)
- selezioni, basate sull'ampiezza, sul valore dell'ascissa e/o sul valore dell'indice dello stesso GD o di un altro
- operazioni di insieme (cancellamento, rotazione e “inversione” per un singolo GD, concatenazione di due GD.

Segnali teorici, simulati e distribuzioni

Possono crearsi GD contenenti decine di tipi di segnali teorici, simulati o distribuzioni di probabilità.

Esercizi

Gli esercizi sono raccolti in vari capitoli, in file rtf. Possono essere letti con un normale word processor, come Word, WordPad e simili, o anche visualizzati nella finestra di testo della finestra principale.

Programmi esterni

E' possibile aggiungere programmi esterni che creano o modificano GD, oppure fanno calcoli su GD. Il modello è client-server, per cui SnagLab agisce da client rispetto a dll esterne che fungono da server di procedure.

Il nome delle dll esterne ammissibili è

csnagdll.dll	per il C/C++
fsnagdll.dll	per il Fortran
bsnagdll.dll	per il Visual Basic

Tali file devono essere nel folder del programma SnagLab.

Ciascuna dll può fornire fino a 50 programmi.

Viene fornita un progetto "template" in Visual C++ per la costruzione della csnagdll.dll .

Il colloquio client-server funziona così:

- il client chiede quali sono le procedure disponibili
- il server gli manda la lista che il client visualizza
- il client sceglie quale procedura vuole e su quali GD deve operare
- il server manda le liste dei parametri reali e interi necessari, fornendo dei valori default
- il client visualizza la lista dei parametri ("di input") e l'utente (il client) digita i valori (se sono diversi dai default)
- il server fornisce l'eventuale GD di uscita e gli eventuali parametri di output (visualizzati dal client).

Per ora, solo GD di tipo 3 sono gestibili.

Le routine basilari (che devono essere modificate dall'utente) sono

```
getcepl  
getextprneeds  
cextpr3
```

oltre a crea_prog_list, tutte contenute in csnagdll.cpp .

Il progetto è contenuto in uno zip file. Occorre scompattarlo, girarlo in Visual C++ e fare le modifiche.

Altre dll saranno fornite per accedere a dati e procedure esterne (per esempio snagframedll.dll e snagr87.dll per accedere ai dati di antenne gravitazionali)

Tabelle

Distribuzione cumulativa normale standardizzata

z	centesimi z									
	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.500000	0.503989	0.507978	0.511967	0.515953	0.519939	0.523922	0.527903	0.531881	0.535856
0.1	0.539828	0.543795	0.547758	0.551717	0.555670	0.559618	0.563559	0.567495	0.571424	0.575345
0.2	0.579260	0.583166	0.587064	0.590954	0.594835	0.598706	0.602568	0.606420	0.610261	0.614092
0.3	0.617911	0.621719	0.625516	0.629300	0.633072	0.636831	0.640576	0.644309	0.648027	0.651732
0.4	0.655422	0.659097	0.662757	0.666402	0.670031	0.673645	0.677242	0.680822	0.684386	0.687933
0.5	0.691462	0.694974	0.698468	0.701944	0.705402	0.708840	0.712260	0.715661	0.719043	0.722405
0.6	0.725747	0.729069	0.732371	0.735653	0.738914	0.742154	0.745373	0.748571	0.751748	0.754903
0.7	0.758036	0.761148	0.764238	0.767305	0.770350	0.773373	0.776373	0.779350	0.782305	0.785236
0.8	0.788145	0.791030	0.793892	0.796731	0.799546	0.802338	0.805106	0.807850	0.810570	0.813267
0.9	0.815940	0.818589	0.821214	0.823814	0.826391	0.828944	0.831472	0.833977	0.836457	0.838913
1	0.841345	0.843752	0.846136	0.848495	0.850830	0.853141	0.855428	0.857690	0.859929	0.862143
1.1	0.864334	0.866500	0.868643	0.870762	0.872857	0.874928	0.876976	0.878999	0.881000	0.882977
1.2	0.884930	0.886860	0.888767	0.890651	0.892512	0.894350	0.896165	0.897958	0.899727	0.901475
1.3	0.903199	0.904902	0.906582	0.908241	0.909877	0.911492	0.913085	0.914656	0.916207	0.917736
1.4	0.919243	0.920730	0.922196	0.923641	0.925066	0.926471	0.927855	0.929219	0.930563	0.931888
1.5	0.933193	0.934478	0.935744	0.936992	0.938220	0.939429	0.940620	0.941792	0.942947	0.944083
1.6	0.945201	0.946301	0.947384	0.948449	0.949497	0.950529	0.951543	0.952540	0.953521	0.954486
1.7	0.955435	0.956367	0.957284	0.958185	0.959071	0.959941	0.960796	0.961636	0.962462	0.963273
1.8	0.964070	0.964852	0.965621	0.966375	0.967116	0.967843	0.968557	0.969258	0.969946	0.970621
1.9	0.971284	0.971933	0.972571	0.973197	0.973810	0.974412	0.975002	0.975581	0.976148	0.976705
2	0.977250	0.977784	0.978308	0.978822	0.979325	0.979818	0.980301	0.980774	0.981237	0.981691
2.1	0.982136	0.982571	0.982997	0.983414	0.983823	0.984222	0.984614	0.984997	0.985371	0.985738
2.2	0.986097	0.986447	0.986791	0.987126	0.987455	0.987776	0.988089	0.988396	0.988696	0.988989
2.3	0.989276	0.989556	0.989830	0.990097	0.990358	0.990613	0.990863	0.991106	0.991344	0.991576
2.4	0.991802	0.992024	0.992240	0.992451	0.992656	0.992857	0.993053	0.993244	0.993431	0.993613
2.5	0.993790	0.993963	0.994132	0.994297	0.994457	0.994614	0.994766	0.994915	0.995060	0.995201
2.6	0.995339	0.995473	0.995603	0.995731	0.995855	0.995975	0.996093	0.996207	0.996319	0.996427
2.7	0.996533	0.996636	0.996736	0.996833	0.996928	0.997020	0.997110	0.997197	0.997282	0.997365
2.8	0.997445	0.997523	0.997599	0.997673	0.997744	0.997814	0.997882	0.997948	0.998012	0.998074
2.9	0.998134	0.998193	0.998250	0.998305	0.998359	0.998411	0.998462	0.998511	0.998559	0.998605
3	0.998650	0.998694	0.998736	0.998777	0.998817	0.998856	0.998893	0.998930	0.998965	0.998999
3.1	0.999032	0.999064	0.999096	0.999126	0.999155	0.999184	0.999211	0.999238	0.999264	0.999289
3.2	0.999313	0.999336	0.999359	0.999381	0.999402	0.999423	0.999443	0.999462	0.999481	0.999499
3.3	0.999517	0.999533	0.999550	0.999566	0.999581	0.999596	0.999610	0.999624	0.999638	0.999650
3.4	0.999663	0.999675	0.999687	0.999698	0.999709	0.999720	0.999730	0.999740	0.999749	0.999758
3.5	0.999767	0.999776	0.999784	0.999792	0.999800	0.999807	0.999815	0.999821	0.999828	0.999835
3.6	0.999841	0.999847	0.999853	0.999858	0.999864	0.999869	0.999874	0.999879	0.999883	0.999888
3.7	0.999892	0.999896	0.999900	0.999904	0.999908	0.999912	0.999915	0.999918	0.999922	0.999925
3.8	0.999928	0.999930	0.999933	0.999936	0.999938	0.999941	0.999943	0.999946	0.999948	0.999950
3.9	0.999952	0.999954	0.999956	0.999958	0.999959	0.999961	0.999963	0.999964	0.999966	0.999967
4	0.999968	0.999970	0.999971	0.999972	0.999973	0.999974	0.999975	0.999976	0.999977	0.999978
4.1	0.999979	0.999980	0.999981	0.999982	0.999983	0.999983	0.999984	0.999985	0.999985	0.999986

Valori del χ^2 per un dato livello di fiducia

Livello di fiducia	0.9	0.95	0.99	0.995	0.999	0.9995	0.9999
Gradi di libertà							
1	2.706	3.841	6.635	7.879	10.827	12.115	15.134
2	4.605	5.991	9.210	10.597	13.815	15.201	18.425
3	6.251	7.815	11.345	12.838	16.266	17.731	21.104
4	7.779	9.488	13.277	14.860	18.466	19.998	23.506
5	9.236	11.070	15.086	16.750	20.515	22.106	25.751
6	10.645	12.592	16.812	18.548	22.457	24.102	27.853
7	12.017	14.067	18.475	20.278	24.321	26.018	29.881
8	13.362	15.507	20.090	21.955	26.124	27.867	31.827
9	14.684	16.919	21.666	23.589	27.877	29.667	33.725
10	15.987	18.307	23.209	25.188	29.588	31.419	35.557
11	17.275	19.675	24.725	26.757	31.264	33.138	37.365
12	18.549	21.026	26.217	28.300	32.909	34.821	39.131
13	19.812	22.362	27.688	29.819	34.527	36.477	40.873
14	21.064	23.685	29.141	31.319	36.124	38.109	42.575
15	22.307	24.996	30.578	32.801	37.698	39.717	44.260
16	23.542	26.296	32.000	34.267	39.252	41.308	45.926
17	24.769	27.587	33.409	35.718	40.791	42.881	47.559
18	25.989	28.869	34.805	37.156	42.312	44.434	49.185
19	27.204	30.144	36.191	38.582	43.819	45.974	50.787
20	28.412	31.410	37.566	39.997	45.314	47.498	52.383
21	29.615	32.671	38.932	41.401	46.796	49.010	53.960
22	30.813	33.924	40.289	42.796	48.268	50.510	55.524
23	32.007	35.172	41.638	44.181	49.728	51.999	57.067
24	33.196	36.415	42.980	45.558	51.179	53.478	58.607
25	34.382	37.652	44.314	46.928	52.619	54.948	60.136
26	35.563	38.885	45.642	48.290	54.051	56.407	61.667
27	36.741	40.113	46.963	49.645	55.475	57.856	63.166
28	37.916	41.337	48.278	50.994	56.892	59.299	64.656
29	39.087	42.557	49.588	52.335	58.301	60.734	66.152
30	40.256	43.773	50.892	53.672	59.702	62.160	67.623
31	41.422	44.985	52.191	55.002	61.098	63.581	69.097
32	42.585	46.194	53.486	56.328	62.487	64.993	70.564
33	43.745	47.400	54.775	57.648	63.869	66.401	72.029
34	44.903	48.602	56.061	58.964	65.247	67.804	73.475
35	46.059	49.802	57.342	60.275	66.619	69.197	74.925
36	47.212	50.998	58.619	61.581	67.985	70.588	76.372
37	48.363	52.192	59.893	62.883	69.348	71.971	77.800
38	49.513	53.384	61.162	64.181	70.704	73.350	79.218
39	50.660	54.572	62.428	65.475	72.055	74.724	80.637
40	51.805	55.758	63.691	66.766	73.403	76.096	82.055
41	52.949	56.942	64.950	68.053	74.744	77.458	83.475
42	54.090	58.124	66.206	69.336	76.084	78.818	84.874
43	55.230	59.304	67.459	70.616	77.418	80.174	86.272
44	56.369	60.481	68.710	71.892	78.749	81.527	87.672
45	57.505	61.656	69.957	73.166	80.078	82.873	89.070

Bibliografia

In genere sono indicati solo gli argomenti più caratterizzanti.

Generale

Autore/Editore	Titolo	Argomenti/Commenti
vari / CRC – Chapman&Hall	Digital Signal Processing Handbook	
Poularikas (ed.by)	Handbook of Formulas and Tables for Signal Processing	
Bendat e Piersol / Wiley	Random Data: Analysis and Measurement Procedures	
MatLab	Statistics Toolbox	
MatLab	Signal Processing Toolbox	

Segnali deterministici – Filtraggio

Autore/Editore	Titolo	Argomenti/Commenti
B.P. Lathy / Berkeley-Cambridge	Signal Processing and Linear Systems	State space analysis
Oppenheim-Willsky / Prentice-Hall	Signals and Systems	Modulation, Partial fraction expansion
A.V.Oppenheim, R.W.Schafer, J.R.Buck / Prentice Hall 1999	Discrete-Time Signal Processing	
Haykin – Van Veen	Signals and Systems	Applications to communication, filters and equalizers and feed-back systems
Hsu / Shaum	Signals and Systems	State Space Analysis, Mathematical Reviews

Hayes / Shaum	Digital Signal Processing	Implementation, Filter design
Karu / Zizi Press	Signals and Systems Made Ridiculously Simple	Bode plots, Deconvolution, Causality and Stability, Feedback, Modulation
Lyons / Prentice Hall	Understanding Digital Signal Processing	Advanced sampling techniques, Signal averaging, Digital data formats, DSP tricks
White	Digital Signal Processing	Coefficient precision
Britton Rorabaugh / TAB Books	Digital Filter Designer's Handbook	Bilinear transformations, Practical considerations
S.W. Smith / California Technical Publishing	The Scientist and Engineer's Guide to Digital Signal Processing	Applications to audio processing, image processing, neural networks, DSPs
S.T. Karris / Orchard	Signals and Systems	Inverse Laplace transform, Circuit analysis, State variables
R.L.Allen, D.W.Mills	SIGNAL ANALYSIS TIME, FREQUENCY, SCALE, AND STRUCTURE	Generalized Fourier Transform, Frequency domain analysis, Time-Frequency Signal Transform, Time-Scale signal transform, Mixed domain signal analysis
A. Mertins / Wiley	Signal Analysis Wavelets, Filter Banks, Time-Frequency Transforms and Applications	Hartley, Hilbert, Band-pass signals, Discrete signals representations, whitening transforms, power spectrum and autocorrelation estimation, uncertainty principle, wavelet transform
MatLab	Filter Design	Fixed point filters, adaptive filters, multirate filters, frequency transformations

Segnali stocastici

Autore/Editore	Titolo	Argomenti/Commenti
R.M. Howard / Wiley	Principles of Random Signal Analysis and Low Noise Design	Power spectral density analysis, memoryless transformations of random processes, low noise design

P.Z. Peebles	Probability, random variables, and random signal principles	Probability, Multiple random variables, random processes, Linear systems with random inputs, Optimum linear systems, Noise in practical systems
D.S.G. Pollock / Academic Press	A Handbook of Time-Series Analysis, Signal Processing and Dynamics	Polynomial methods, State space equations, Least square methods, Linear filters, ARMA processes, Statistical distributions, Estimation theory
J.D. Hamilton / Princeton University Press	Time Series Analysis	ARMA processes, forecasting, spectral analysis, Asymptotic distributions, Bayesian analysis, Kalman filter, Non-stationary time series
Hsu / Shaum	Theory and Problems of Probability, Random Variables, and Random Processes	Probability, Random processes processing and analysis, Estimation theory, Decision theory, Queueing theory
S.V. Vaseghi / Wiley	Advanced Digital Signal Processing and Noise Reduction	Probability models, Bayesian estimation, Markov models, Wiener filters, Adaptive filters, Prediction, Power spectrum and correlation, Interpolation, Impulsive noise, Equalization
I.Glover, P.Grant / Prentice Hall	Digital Communications	Digital communication, Decision theory, Optimum filtering, Information theory, TDMA, CDMA, PAL, JPEG, MPEG, Networks
A. Papoulis, S. Unnikrishna Pillai, McGraw-Hill 2002	Probability, Random Variables and Stochastic Processes	
Hwei P. Hsu / McGraw-Hill, 1997	Theory and Problems of Probability, Random Variables, and Random Processes	
MatLab	Financial Time Series Toolbox	
MatLab	GARCHToolbox	

Aspetti matematici

Autore/Editore	Titolo	Argomenti/Commenti
----------------	--------	--------------------

S.M.Kay / Prentice Hall	Fundamentals of Statistical Signal Processing: Estimation Theory	Estimation, Maximum likelihood, Least squares, Linear Models, Kalman filter, Bayesian philosophy
R. Baraniuk	Signals and Systems	Hilbert spaces, convergence, transforms (?)

Immagini

Autore/Editore	Titolo	Argomenti/Commenti
R.C. Gonzales, R.E. Woods / Prentice Hall	Digital Image Processing	Visual perception, Image enhancement in space and frequency domains, Image restoration, Wavelet, Image compression, Representation and description, Object recognition
W.K.Pratt / Wiley	Digital Image Processing	Continuous and digital image characterization, Linear processing, Image improvement, Image analysis, PIKS software
Hany Farid / www.cs.dartmouth.edu/ ~farid/tutorials/	Fundamentals of Image Processing	Linear filtering, Wiener filter, Non- linear filtering, Median filter, Dithering, Motion estimation, Principal components analysis
A. Erhardt-Ferron	Theory and Applications of Digital Image Processing	Basic, practical
Adhemar Bultheel	Wavelets with applications in signal and image processing	Heisenberg uncertainty, Time- frequency plane, Filter banks, Multiresolution, Multidimensional wavelets, Signal processing, Image processing
G.X.Ritter, J.N.Wilson / CRC Press	Handbook of Computer Vision	Image algebra, Image enhancement, Edge detection, Thinning and skeletonizing, Morphological transforms, Linear image transform, Pattern matching, Feature description, Neural network and cellular automata
MatLab	Image processing toolbox	

Altro

Autore/Editore	Titolo	Argomenti/Commenti
H.L. Van Trees / Wiley	Detection, Estimation, and Modulation Theory, Part I	Classical detection and Estimation Theory, Random processes, Estimation of signal parameters and continuous waveforms
G. Wornell / Prentice Hall	Signal Processing with Fractals: a Wavelet-Based Approach	Statistically self-similar processes, Fractal modulation
F. Hlawatsch, G. Matz / http://www.nt.tuwien.ac.at/dspgroup/tfgroup/time.html	Time-Frequency Signal Processing: A Statistical Perspective	Molti altri articoli nel sito
C. Torrente, C.P.Compo / Bulletin of the American Meteorological Society, Vol. 79, No. 1, January 1998	A Practical Guide to Wavelet Analysis	Semplice introduzione
H.Tong / Oxford Science Publications	Non-linear Time Series	
MatLab (5)	Higher-Order Spectral Analysis Toolbox	
MatLab	Wavelet Toolbox	
MatLab	Fuzzy Logic Toolbox	
MatLab	Neural Network Toolbox	
MatLab	System Identification Toolbox	

Indice Analitico

A

aliasing.....	165
analisi preliminare.....	162
asimmetria.....	41
autocorrelazione.....	171

B

bispettro.....	215
----------------	-----

C

campionamento.....	8
campionamento non uniforme.....	247
coefficiente di correlazione.....	52
coerenza.....	189
consistenza.....	56
convoluzione.....	16; 19; 22
convoluzione discreta.....	75
correttezza.....	56
covarianza.....	51
curtosi.....	41

D

densità marginali.....	49; 50
deviazione standard.....	37
diagramma di fase.....	245
distribuzione	
t di Student.....	45
distorsione.....	56
distribuzione	
binomiale.....	38
cumulativa.....	41
del chi quadro.....	42
di Cauchy.....	44
di Gauss o normale.....	42
di Poisson.....	38
di Raleigh.....	44
esponenziale.....	43
uniforme.....	41
disturbi impulsivi.....	208
disuguaglianza di Chebyshev.....	40
disuguaglianza di Schwartz.....	53

E

efficienza.....	56
equazioni alle differenze.....	86
equazioni di Yule e Walker.....	186
esperimenti controllati.....	55

esperimenti osservativi.....	55
evoluzione libera.....	92

F

falso allarme.....	199
FFT.....	85
filtro	
a mediana.....	208
a sfasamento nullo.....	205
adattato.....	192
di Wiener.....	203
FIR.....	205
IIR205	
in frequenza.....	206
lineare per immagini.....	229
sbiancante.....	196
filtro di quadratura.....	130
fit armonico.....	246
fit lineare.....	65
frequenza di Nyquist.....	163
funzione caratteristica.....	47
funzione di trasferimento.....	26
funzione di trasferimento discreta.....	87
funzione di verosimiglianza.....	59

G

GD.....	255
---------	-----

I

immagini.....	225
indipendenza stocastica.....	50
inferenza statistica.....	55
intervallo di confidenza.....	63
ipotesi alternativa.....	71
ipotesi nulla.....	71
istogramma.....	169

K

kurtosis.....	41
---------------	----

L

likelihood.....	59
livello di confidenza.....	63
livello di fiducia.....	63; 72
livello di significatività.....	72
lock-in.....	242

M

massima verosimiglianza	58
Matlab	251
matrice di covarianza	52
mediana	39
minimi quadrati	56
moda	39

P

periodicità coerente	237
periodicità di eventi	250
periodicità non coerente	237; 249
principio dei minimi quadrati	66
Principio di Ortogonalità	57
processi stocastici	134
autocorrelazione	135; 138
autocovarianza	136
bispettro	139
correlazione incrociata	137
covarianza incrociata	137
cumulante del primo ordine	138
cumulante del quarto ordine	138
cumulante del secondo ordine	138
cumulante del terzo ordine	138
e sistemi lineari	141
ergodicità	139
normali	144
ortogonali	137
processi ARMA	147
processo AR(1) complesso	154
processo AR(1) reale	152
processo AR(2)	157
processo di Poisson	160
rumore bianco	145
rumore bianco discreto	149
scorrelati	137
spettro di potenza	136; 138
spettro incrociato	137
stazionari	137
trasformazioni	140
trispettro	139
pulsazione normalizzata	84

R

rapporto segnale/rumore	191
risposta forzata	92
risposta impulsiva	75; 91
rivelazione	191
robustezza	56
ROC (Receiver Operating Characteristics)	200
rumore	190
rumore di quantizzazione	167

S

segnale analitico	130
segnali	
analogici e digitali	9

autocorrelazione	129
continui e discreti	8
convoluzione	126
correlazione	126
cross-energia	126
delta	10; 25
deterministici o casuali	9
distanza	126
durata	124
energia	124
esponenziale complesso	11
esponenziale decrescente	25
esponenziale simmetrico	25
gaussiana	23
gradino	9; 25
impulsivi	124
larghezza	124
lorentziana	25
pacchetto gaussiano	23
pacchetto rettangolare	24
periodici	9
permanenti	133
posizione	124
prodotto vettoriale	126
reali e complessi	9
rettangolo	24
sinusoidale	10
transitori	124
sistemi	
AR del primo ordine (complesso)	112
AR del primo ordine (reale)	105
AR del secondo ordine	115
AR o auto-regressivi	77
ARMA	77
ARMA (1,1)	118
causali	14
di ordine 0	95
di Volterra	123
in parallelo	13
in serie	12
lineari	14
<i>lineari continui</i>	16
MA del primo ordine	95
MA del secondo ordine	102
MA o a media mobile	77
non lineari	123
passa-alto continuo	31
passa-basso continuo	29
passa-tutto	96
risonanza continua	33
stabili	15
statici o dinamici	13
tempo-invarianti	15
skewness	41
Snag	254
SnagLab	255
spettro di energia	129
spettro di potenza	173
spettro impulsivo	250
stabilità	94
stima	55
correlazione incrociata	188
dei parametri di una retta sperimentale	65
del valor medio	62
della varianza	64
dell'autocorrelazione	171
dello spettro di potenza	173

spettro incrociato	188
stima bayesiana	61
stima spettrale	
finestramento	175
incertezza	180
non parametrica	174
parametrica	186
risoluzione	175
stimatore	56

T

teorema	
Bayes	61
del campionamento	163
teoria della rivelazione	199
test statistici	71
trasformata	
di Fourier	19

di Fourier discreta	85
di Fourier per dati discreti	84
di Hilbert	130
di Laplace	18
wavelet	209
z 80	

V

valor medio	37; 39
valor medio di $g(x)$	39
valore aspettato	37; 39
variabili casuali	
continue	39
discrete	37
multiple	49
varianza	37