

# Prefazione

Come sappiamo, la matematica è uno strumento che ci permette di affrontare numerosi problemi del mondo reale. Ciò è vero anche nel campo della radioastronomia, ed in particolare del progetto SETI, il cui obiettivo è il riconoscimento di segnali di origine extraterrestre.

Ancor prima di tentare un'interpretazione del segnale, è infatti necessario isolarlo dalla miriade di interferenze terrestri ed emissioni di origine naturale. Abbiamo quindi bisogno di un metodo che ci permetta di ottenere la maggior quantità di informazioni dai dati ricevuti. L'utilizzo della trasformata di Fourier permette un'analisi molto accurata se il segnale che stiamo cercando è *stazionario*, ma se tale condizione viene a mancare è purtroppo indispensabile cercare di individuare uno strumento più efficiente.

Questa tesi, che si inserisce nell'ambito del progetto SETI ed è stata svolta presso il radiotelescopio "Croce del Nord" dell'Istituto Nazionale di Astrofisica (INAF), si pone l'obiettivo di studiare un nuovo metodo di analisi dei segnali nonstazionari, la trasformata di Hilbert-Huang. Punto base su cui poggia la tesi è il lavoro dello scienziato Norden E. Huang, ideatore della trasformata.

I primi due capitoli intendono descrivere l'ambiente in cui questa tesi si inserisce. In particolare, il primo capitolo presenta il radiotelescopio di Medicina e il progetto SETI, mentre il secondo illustra il procedimento di rilevazione dei segnali, attraverso l'uso della trasformata di Fourier.

Il terzo capitolo, cuore della tesi, espone in maniera approfondita la trasformata di Hilbert-Huang, evidenziandone pregi e difetti. Nel quarto capitolo ci si pone il problema dell'applicazione della trasformata nell'analisi di segnali, e si mostrano i risultati ottenuti. L'ultimo capitolo espone infine le conclusioni.



# Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
1.1	La radioastronomia . . . . .	5
1.2	La stazione radioastronomica di Medicina . . . . .	6
1.3	Il progetto SETI . . . . .	8
<b>2</b>	<b>Rilevazione del segnale</b>	<b>11</b>
2.1	La trasformata di Fourier . . . . .	11
2.2	Trasformata di Fourier discreta e FFT . . . . .	12
2.3	Rilevamento dei segnali e ipotesi di stazionarietà . . . . .	16
2.4	Alcuni metodi per analizzare segnali nonstazionari . . . . .	18
<b>3</b>	<b>La trasformata di Hilbert-Huang</b>	<b>21</b>
3.1	Frequenza istantanea . . . . .	21
3.1.1	Definizione di frequenza istantanea attraverso la trasformata di Hilbert . . . . .	22
3.1.2	Significato della definizione . . . . .	22
3.1.3	Discussione . . . . .	24
3.2	Intrinsic Mode Functions . . . . .	24
3.3	Empirical Mode Decomposition . . . . .	26
3.4	Spettro di Hilbert . . . . .	30
3.5	Discussione e problemi matematici aperti . . . . .	32
3.5.1	Trasformata di Hilbert normalizzata . . . . .	32
3.5.2	Problemi sulle spline . . . . .	34
3.5.3	Ottimizzazione sulla scelta delle IMF . . . . .	34
3.5.4	Altri problemi . . . . .	35
<b>4</b>	<b>Validazione</b>	<b>37</b>
4.1	Il problema del rumore . . . . .	37
4.2	Intermittency: un metodo per separare il segnale dal rumore . . . . .	38
4.3	Test . . . . .	40

<b>5 Conclusioni</b>	<b>49</b>
<b>A Generalized Zero Crossing</b>	<b>51</b>
<b>B Algoritmo EMD</b>	<b>53</b>
<b>Bibliografia</b>	<b>57</b>

# Capitolo 1

## Introduzione

Parte del materiale presentato in questo capitolo è stato liberamente tratto dalla tesi di laurea del dottor Giorgio Leone. [1]

### 1.1 La radioastronomia

Quasi tutto quello che conosciamo sul cosmo, riguardo le stelle, i sistemi stellari, le loro distribuzioni, la loro dinamica e cinematica, è stato ottenuto attraverso l'osservazione della radiazione elettromagnetica.

Per molti secoli, lo studio del cielo e delle stelle allo scopo di conoscere l'universo si è avvalso quasi esclusivamente di strumenti ottici come i telescopi; solo di recente agli strumenti che rilevano e analizzano la luce visibile (fotometri e spettrografi) si è affiancata l'osservazione della radiazione in altre lunghezze d'onda; ciò è stato reso possibile dallo sviluppo tecnologico che ha permesso la costruzione di strumenti in grado di captare onde elettromagnetiche a diverse frequenze.

I corpi celesti, infatti, in base al loro stato fisico, emettono una qualche forma di energia su tutta la gamma dello spettro elettromagnetico; questa radiazione, però, viene in larga misura assorbita dall'atmosfera terrestre, che lascia filtrare quasi solamente le lunghezze d'onda comprese nella banda ottica (fra i 400 e i 700  $nm$ ) e nella banda radio (tutte le lunghezze d'onda superiori a 1  $mm$ ). All'astronomia classica si sono dunque affiancate le osservazioni nella finestra radio, cioè la radioastronomia.

Il grosso vantaggio dell'osservazione radio rispetto a quella che sfrutta la radiazione visibile è che questa può essere effettuata anche di giorno ed è soggetta a meno disturbi dovuti alle condizioni atmosferiche e alla presenza di nubi opache di materia interstellare. Lo svantaggio è che a causa della minore frequenza delle onde radio si ha, a parità di superficie ricevente, una minore

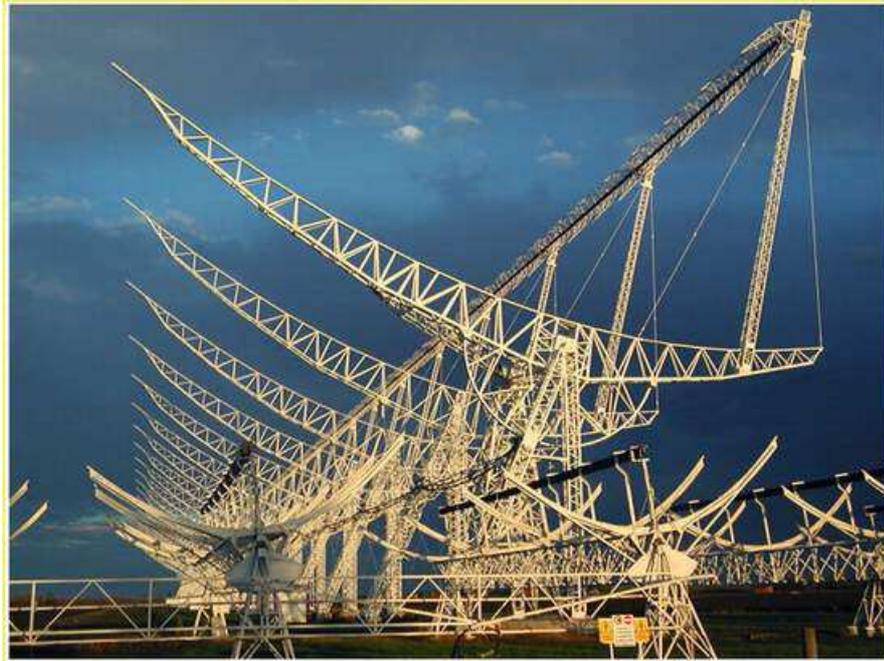


Figura 1.1: La Croce del Nord

risoluzione spaziale. Per rimediare a questo problema, è stato sviluppato il progetto VLBI (Very Long Baseline Interferometry).

Tale progetto consiste nell'osservazione della stessa sorgente da più radiotelescopi distribuiti sulla Terra. I dati acquisiti vengono poi correlati per ottenere l'immagine radio dell'oggetto osservato. In questo modo il potere risolutivo risulta equivalente a quello che si otterrebbe con un radiotelescopio di dimensioni pari alla distanza tra le antenne stesse (baseline).

## 1.2 La stazione radioastronomica di Medicina

Questa tesi di laurea è stata svolta presso l'osservatorio radioastronomico dell'Istituto Nazionale di Astrofisica (INAF) situato a Medicina (Bologna). Nell'Istituto esistono due grandi radiotelescopi: la "Croce del Nord" (da cui prende il nome l'osservatorio) e un'antenna parabolica da 32 metri.

Il radiotelescopio Croce del Nord (figura 1.1) è costituito da due bracci disposti a T. Il primo, orientato in direzione Nord-Sud, è composto da 64 antenne di forma cilindrico-parabolica lunghe 24 metri ciascuna, e copre una



Figura 1.2: L'antenna parabolica

superficie di circa  $10.000 \text{ m}^2$ . Il ramo Est-Ovest, invece, è costituito da un'unica antenna cilindrico-parabolica, lunga 564 metri e larga 35, per una superficie totale di raccolta di circa  $20.000 \text{ m}^2$ . La frequenza di funzionamento è centrata sui 408 MHz con una banda di 6 MHz.

Questo strumento, che è stato concepito per l'esplorazione sistematica del cielo alla ricerca di radiosorgenti (Sky Survey), è un radiotelescopio di *transito*, cioè è in grado di funzionare solo quando la sorgente, per effetto della rotazione terrestre, transita sul meridiano locale.

L'antenna parabolica (figura 1.2) invece, è completamente orientabile, e può quindi puntare gli oggetti celesti per tutto il tempo della loro "visibilità" in cielo. Quest'antenna è formata da uno specchio concavo di forma parabolica di 32 metri di diametro. Le frequenze di funzionamento vanno dai 327 MHz ai 43 GHz (cioè tra 90 e 0.7 cm di lunghezza d'onda).

L'antenna parabolica è coinvolta in diversi progetti, in rete o in *single dish* (ossia la modalità operativa in cui l'antenna compie le sue osservazioni singolarmente, senza cooperare con altri strumenti radioriceventi). Tra quelli in rete, quello maggiormente degno di nota è il progetto VLBI, a cui partecipa anche la sua antenna gemella installata a Noto (Sicilia).

Tra i progetti più significativi in *single dish* si annoverano quello di spettrometria, ricerca di comete ed il SETI.

## 1.3 Il progetto SETI

Molti studiosi ritengono che la vita, così come noi la conosciamo, non sia un evento unico. Ciò è suggerito da una serie di considerazioni, tra cui il fatto che il Sole non è che una delle trecento miliardi di stelle presenti nella nostra galassia, e che attorno a queste stelle è probabile l'esistenza di sistemi planetari, dato che i pianeti sono un possibile prodotto del processo di formazione delle stelle stesse. Non solo, ma le moderne indagini bioastronomiche hanno portato alla scoperta di numerose molecole organiche complesse, all'interno delle nubi presenti negli spazi interstellari; tali molecole sono alla base dello sviluppo della vita a noi nota.

Sulla base di tutto ciò, considerando la possibilità che la vita abbia potuto seguire un corso simile a quello avuto sulla Terra, evolvendosi fino alla nascita di una società tecnologicamente avanzata, è nato il progetto SETI (Search for ExtraTerrestrial Intelligence).

Il progetto fu avviato dalla NASA nel 1959, in seguito alla pubblicazione di un articolo scritto da due fisici, Cocconi e Morris, nel quale si teorizzava la possibilità che, qualora un'ipotetica civiltà extraterrestre in possesso dell'adeguata tecnologia avesse voluto farsi localizzare, avrebbe potuto farlo inviando nel cosmo un segnale radio artificiale a banda stretta in tutte le direzioni alla frequenza di 1420 MHz. [2]

Tale frequenza (relativa ad un a lunghezza d'onda di 21 *cm*) fu scelta in quanto corrisponde all'emissione spontanea di radiazione da parte degli atomi di idrogeno neutro interstellare, ed essendo l'idrogeno l'elemento più diffuso in natura, un'eventuale razza aliena tecnologicamente avanzata dovrebbe senz'altro conoscerla. Non solo, ma le frequenze collocate fra 1 GHz e 15 GHz sono quelle in cui l'effetto del rumore galattico e l'assorbimento dell'atmosfera è minore, e inoltre un segnale a banda stretta è nettamente distinguibile da quelli di origine naturale a banda larga.

Il primo moderno esperimento SETI (noto come *Project Ozma*) fu effettuato nel 1960, da parte dell'astronomo Frank Drake, il quale puntò un radiotelescopio di 25 metri di diametro in direzione delle stelle Tau Ceti ed Epsilon Eridani, ossia le due più vicine alla Terra, tra quelle di tipo solare e con probabili pianeti.

Anche se le ricerche non hanno portato fino ad ora alla ricezione di segnali extraterrestri, il progetto SETI è supportato dalla famosa equazione di Drake,  $N = R_* * f_p * n_e * f_l * f_i * f_t * L$ , che vuole enumerare il numero di civiltà presenti nella nostra galassia.  $R_*$  è il numero di stelle di tipo solare che nascono ogni anno,  $f_p$  frazione di queste stelle che possiede un sistema planetario,  $n_e$  numero medio di pianeti di tali sistemi che si trovano ad una distanza dalla stella tale da renderli abitabili,  $f_l$  frazione di questi pianeti

su cui la vita è presente se pur in modo primitivo,  $f_i$  quantità di pianeti su cui la vita si è evoluta fino a sviluppare l'intelligenza,  $f_t$  frazione di mondi intelligenti padroni di una tecnologia osservabile dalla Terra. Infine  $L$  indica la durata media in anni di tali civiltà. Nonostante le stime effettuate da Drake indichino la presenza di numerose civiltà extraterrestri, molti termini presenti nell'equazione di Drake si basano unicamente su congetture, dato che esiste un solo pianeta conosciuto in cui è presente la vita. Ciò ha portato alcuni studiosi a ritenere che l'equazione di Drake non abbia nulla a che fare con la scienza.

Il SETI Institute sta attualmente collaborando con il Laboratorio Radio Astronomico dell'Università di Berkeley per sviluppare un allineamento di radiotelescopi specializzati per gli studi SETI. Questo nuovo progetto viene detto *Allen Telescope Array* (ATA), e consisterà, una volta completato, nell'allineamento di 350 piatti radio di circa 6 metri di diametro. [3]

Dai primi mesi dl 1998 anche alla stazione di Medicina esiste un programma SETI, condotto tramite lo spettrometro digitale SERENDIP IV (Search for Extraterrestrial Radio Emissions from Nearby Developed Intelligent Populations). [4]

L'attività di ricerca, attualmente non limitata alla sola frequenza di 1420 MHz, viene condotta secondo la modalità operativa "Piggy back", ovvero i segnali che elabora il SERENDIP IV provengono dalle antenne della stazione durante la normale attività di osservazione, e quindi avviene in parallelo ad essa. Questa scelta è stata dettata dalla necessità di contenere i costi senza per questo diminuire le possibilità di successo. Infatti, anche ammettendo l'esistenza di una civiltà aliena, identificarla è una possibilità assai remota; basti pensare che non sappiamo dove, quando e a che frequenza cercarla.



# Capitolo 2

## Rilevazione del segnale

In questo capitolo intendiamo entrare nei dettagli dell'analisi comunemente effettuata sui dati rilevati dalle antenne. Per fare ciò sarà prima indispensabile introdurre la trasformata di Fourier, lo strumento matematico primario dell'analisi dei segnali, le cui applicazioni non sono limitate alla sola radioastronomia, ma si estendono ad innumerevoli altri campi (ottica, acustica, cristallografia...).

### 2.1 La trasformata di Fourier

Qualunque segnale può essere espresso in termini delle frequenze di cui è costituito attraverso l'applicazione della trasformata di Fourier; vediamo ora come questo è possibile.

Sia  $x(t)$  una funzione reale e  $C_0^\infty$ . La sua trasformata di Fourier è

$$X(\omega) = \int_{-\infty}^{\infty} x(t)e^{-i\omega t} dt \quad (2.1)$$

La funzione  $x(t)$  può essere ricavata a partire dalla sua trasformata come segue (formula di antitrasformazione)

$$x(t) = \frac{1}{2\pi} \int_{-\infty}^{\infty} X(\omega)e^{i\omega t} d\omega \quad (2.2)$$

Essendo  $x$  una funzione reale, vale la relazione

$$X(-\omega) = \overline{X(\omega)}$$

equivalente alle seguenti

$$|X(-\omega)| = |X(\omega)| \quad (2.3)$$

$$\arg X(-\omega) = -\arg X(\omega) \quad (2.4)$$

Dalle relazioni precedenti discende:

$$\begin{aligned} x(t) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| e^{i \arg X(\omega)} e^{i\omega t} d\omega = \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| \cos [\omega t + \arg X(\omega)] d\omega + \frac{i}{2\pi} \int_{-\infty}^{\infty} |X(\omega)| \sin [\omega t + \arg X(\omega)] d\omega \end{aligned}$$

La parte immaginaria di  $x$  è nulla, dunque otteniamo

$$x(t) = \frac{1}{\pi} \int_0^{\infty} |X(\omega)| \cos [\omega t + \arg X(\omega)] d\omega \quad (2.5)$$

La 2.5 esprime  $x(t)$  come somma integrale di infiniti termini sinusoidali, ognuno dei quali ha pulsazione  $\omega$  (e dunque frequenza  $\omega/2\pi$ ), ampiezza  $|X(\omega)|/\pi$  e fase  $\arg X(\omega)$ .

In altre parole, abbiamo mostrato che la 2.1, in modulo, altro non è che l'ampiezza della componente di  $x$  avente frequenza  $\omega$  (a meno di una costante moltiplicativa). L'insieme dei valori assunti da  $|X(\omega)|$  è detto *spettro di Fourier* di  $x$ . D'ora in avanti, ci riferiremo a  $x(t)$  come al segnale nel *dominio del tempo*, e a  $X(\omega)$  come al segnale nel *dominio della frequenza*.

Malgrado questo risultato, è evidente che la 2.1 non è applicabile, così com'è, ai dati sperimentali. Tali dati, infatti, non sono altro che una sequenza finita (una  $n$ -upla) di valori, che possiamo considerare un *campionamento* del segnale  $x(t)$ . In simboli

$$x_k = x(kT) \quad k = 1, 2, \dots, n$$

dove  $T$  è il periodo di campionamento del segnale.

Non solo, ma la 2.1 è un'equazione integrale in generale non risolvibile esplicitamente. Cerchiamo quindi una relazione che sia direttamente implementabile su un calcolatore.

## 2.2 Trasformata di Fourier discreta e FFT

Sia  $x = (x_1, x_2, \dots, x_n)$  una  $n$ -upla di valori reali, si definisce trasformata di Fourier discreta di  $x$  la  $n$ -upla  $X = (X_1, X_2, \dots, X_n)$  tale che:

$$X_q = \sum_{k=1}^n x_k e^{-i \frac{2\pi}{n} kq} \quad q = 1, 2, \dots, n \quad (2.6)$$

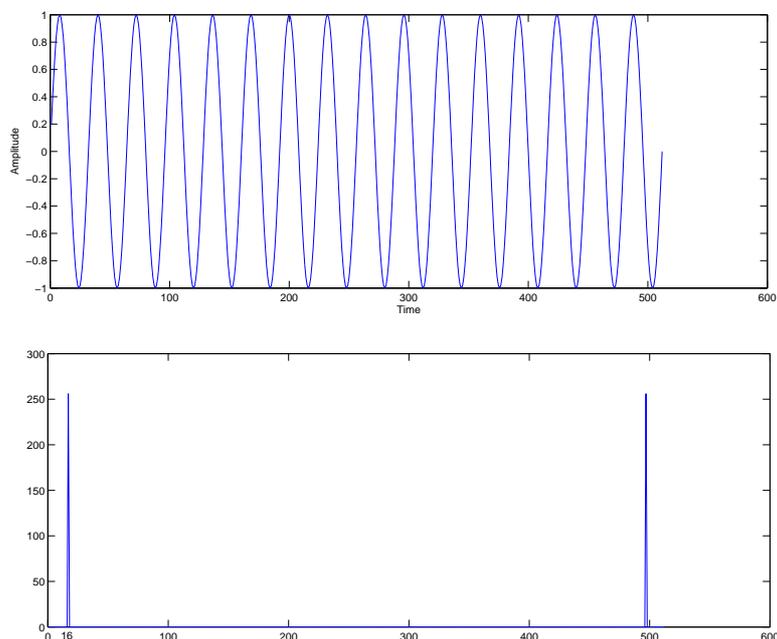


Figura 2.1: Il segnale  $x(t)$  (in alto) e il suo spettro  $X(\omega)$

Anche questa possiede una formula di antitrasformazione, espressa da:

$$x_k = \frac{1}{n} \sum_{q=1}^n X_q e^{i\frac{2\pi}{n}qk} \quad k = 1, 2, \dots, n \quad (2.7)$$

Confrontando la 2.1 con la 2.6 osserviamo immediatamente che l'integrale su tutto l'asse reale è stato sostituito dalla sommatoria di  $n$  addendi, corrispondenti agli  $n$  campionamenti effettuati; e il valore frequenziale  $\omega$ , che nella 2.1 poteva variare su tutti i numeri reali, è stato sostituito dalla sequenza  $2\pi q/n$ , con  $q = 1, 2, \dots, n$ , ossia tutti i multipli interi della partizione  $n$ -esima della circonferenza unitaria.

Sarebbe troppo lungo approfondire qui il legame e le differenze che intercorrono fra la trasformata di Fourier e la trasformata di Fourier discreta (il lettore interessato potrà comunque trovarne un'ampia trattazione in [5] e [6]). Limitiamoci per ora ad osservare che la 2.6 esprime una relazione biunivoca tra la  $n$ -upla dei dati e un'altra  $n$ -upla in cui il modulo delle componenti è, per definizione, il suo spettro. Cerchiamo ora di capire meglio questa relazione attraverso alcuni esempi.

In figura 2.1 è mostrato il grafico di un segnale sinusoidale di frequenza 16 Hz campionato a 512 Hz (cioè supponiamo che i campioni distino tra

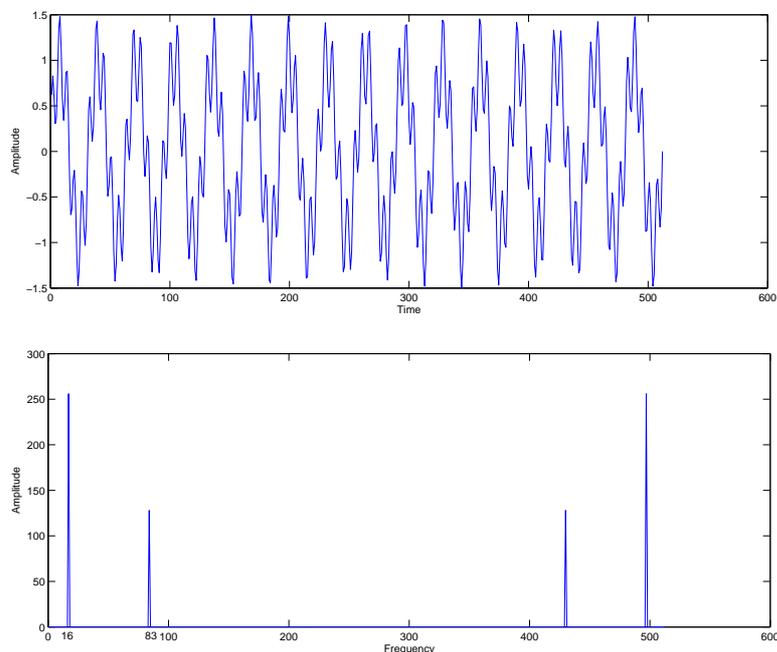


Figura 2.2: Il segnale  $x(t)$  (in alto) e il suo spettro  $X(\omega)$

loro  $1/512$  di secondo), e il suo spettro di Fourier, che mostra un picco in corrispondenza di 16 Hz. Si noti che in realtà il grafico presenta *due* picchi; ciò deriva dal fatto che lo spettro della trasformata di Fourier discreta è sempre simmetrico rispetto al punto medio della sequenza (per la precisione, rispetto al punto di posto  $n/2 + 1$ ), ma questo fenomeno non ha nessun significato fisico. Nella figura 2.2 abbiamo aggiunto un'ulteriore sinusoide con frequenza 83 Hz e ampiezza  $1/2$  di quella del segnale precedente. Come si può notare, lo spettro mostra ora due righe, corrispondenti alle due frequenze presenti nel set di dati analizzato.

Naturalmente, nell'analisi dei dati radioastronomici, i segnali non sono mai così "puliti". Al contrario, subiscono disturbi a causa di molteplici fattori, sia terrestri che extraterrestri. Queste interferenze, collettivamente chiamate "rumore" (*noise*), sono il problema principale da affrontare per poter sperare di riconoscere un segnale debole. Al fine di evidenziare l'influenza del rumore sulla trasformata di Fourier, abbiamo preso la sinusoide iniziale e l'abbiamo "immersa" in una sequenza di valori generati casualmente (che rappresentano il rumore). Dai grafici in figura 2.3 appare evidente che, malgrado nel dominio del tempo la sequenza di dati appaia del tutto casuale, la trasformata di Fourier è ancora in grado di individuarne la componente

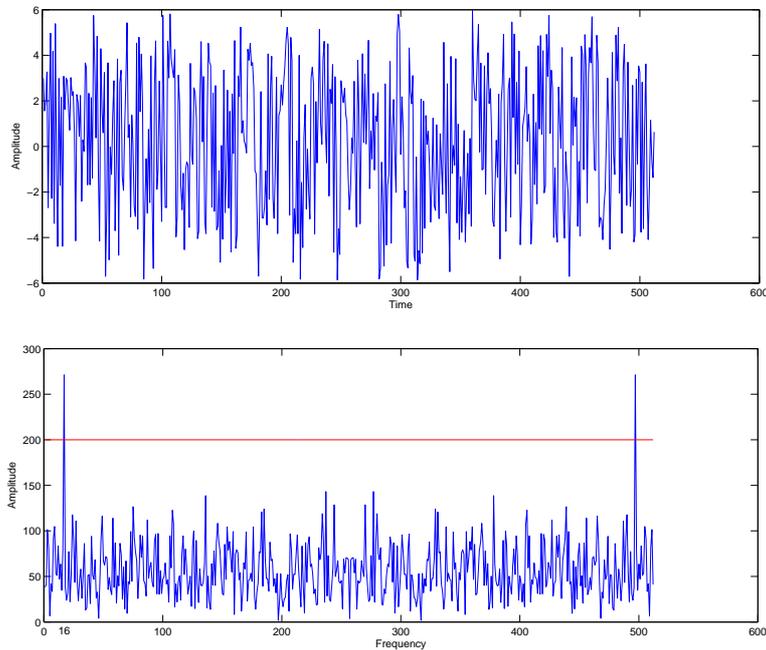


Figura 2.3: Il segnale  $x(t)$  (in alto) e il suo spettro  $X(\omega)$

alla frequenza 16 Hz. In ogni caso, questo esempio mostra anche che, ovviamente, all'aumentare del rumore, diventa sempre più difficile discernere una frequenza “importante”. A questo proposito, nella figura è stata posta una linea orizzontale che serve come livello di “soglia”; tutti i valori di  $|X_q|$  che oltrepassano questa soglia vengono considerati frequenze importanti dello spettro di Fourier.

Esistono algoritmi veloci di trasformazione che riducono il numero di calcoli richiesti per determinare i coefficienti  $X_q$ . Il modo più ovvio per ottenere tali coefficienti è attraverso la loro definizione espressa dalla 2.6. Tuttavia questo metodo non è molto efficiente in quanto richiede  $n^2$  moltiplicazioni e  $n^2$  addizioni. Utilizzando invece le trasformate veloci il numero di operazioni cala significativamente. Uno di questi algoritmi è la Fast Fourier Transform (FFT), che richiede un costo computazionale pari solo a  $O(n \log n)$  a condizione che il numero di punti  $n$  sia una potenza di 2.

## 2.3 Rilevamento dei segnali e ipotesi di stazionarietà

Il rilevamento dei segnali (*Signal Detection*) è il processo mediante cui un software in modo automatico analizza un set di dati e determina se in essi è presente un segnale rilevante. La più comune applicazione di questo procedimento si può riassumere come segue:

1. Si suppone che il segnale che ci si aspetta di rilevare non vari in frequenza e in ampiezza (*ipotesi di stazionarietà*).
2. Si usa la trasformata di Fourier sui dati per calcolarne lo spettro di frequenze.
3. Per ogni canale dello spettro, si fa un controllo sull'ampiezza; se tale valore oltrepassa una certa soglia, si ha una rilevazione del segnale e il processo restituisce la frequenza e l'ampiezza relative.

Anche se il procedimento sopra descritto è una semplificazione di quello effettivamente impiegato nell'analisi dei dati, esso ci dà comunque un'idea ben precisa dei suoi passaggi fondamentali. [7]

Soffermiamoci ora sul punto 1; perchè è così importante supporre che il segnale sia stazionario? Il lettore attento avrà certamente notato che negli esempi presentati nel capitolo precedente si sono sempre usati come segnali-campione delle semplici sinusoidi, con frequenze e ampiezze costanti. Ma se il segnale fosse nonstazionario (ossia la frequenza e l'ampiezza variano nel tempo), che risultato otterremmo applicando la trasformata di Fourier? Di certo non avremo uno spettro caratterizzato da picchi così evidenti come quelli mostrati sopra. Attraverso la 2.6 stiamo essenzialmente cercando di descrivere il segnale da cui abbiamo campionato la sequenza  $x_k$  utilizzando un numero finito di funzioni aventi frequenza e ampiezza sempre costanti; se nel segnale da analizzare tali valori non sono costanti, la rappresentazione che otteniamo non potrà essere veritiera. Cerchiamo di capire meglio quanto detto attraverso un esempio: il segnale mostrato in figura 2.4 presenta una decrescita esponenziale in ampiezza ed una crescita lineare della frequenza. Uno sguardo al suo spettro di Fourier ci basta a capire che esso non ci offre alcuna informazione significativa sul segnale.

Sfortunatamente, non c'è modo di sapere a priori se un segnale proveniente dal cosmo è stazionario o meno; il problema appena descritto rappresenta dunque la più grande limitazione dalla trasformata di Fourier. La capacità di analizzare segnali nonstazionari, come vedremo nel capitolo 3, è invece il maggior punto di forza della trasformata di Hilbert-Huang.

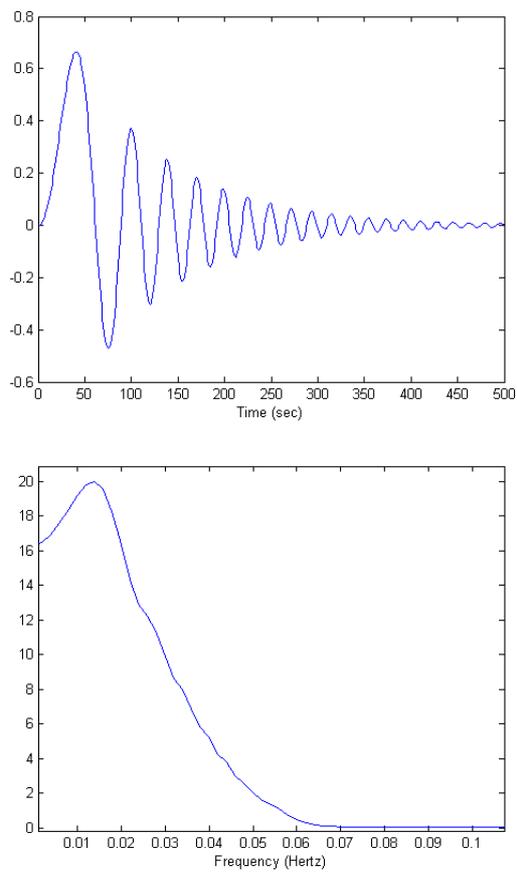


Figura 2.4: Il segnale  $x(t)$  (in alto) e il suo spettro  $X(\omega)$

## 2.4 Alcuni metodi per analizzare segnali non-stazionari

Prima di soffermarci sulla trasformata di Hilbert-Huang, che costituisce l'oggetto principale di questa tesi, presentiamo una breve panoramica dei principali metodi attualmente in uso nell'analisi di segnali nonstazionari. Di ciascuno di essi ci limiteremo a esporre come funziona e su quale principio è basato.

- *Lo spettrogramma*

Lo spettrogramma è il metodo più basilare, nulla di più di un'analisi di Fourier eseguita su *finestre*, cioè su piccole “porzioni” dell'intero set di dati. Spostando di volta in volta la finestra lungo l'asse temporale, è possibile ottenere una distribuzione tempo-frequenza. Perché questo metodo risulti efficace, è necessario comunque assumere che il segnale sia stazionario localmente (vale a dire all'interno di ogni finestra). Questo può creare problemi nella scelta delle dimensioni della finestra, soprattutto se, come nel nostro caso, non si conoscono le caratteristiche del segnale da rilevare.

- *L'analisi wavelet*

Questo metodo essenzialmente scompone il segnale non in una serie di funzioni sinusoidali, come la trasformata di Fourier, ma si serve di una diversa famiglia di funzioni; tali funzioni (tipicamente non nulle solo su un insieme compatto di valori, in modo da poter essere “localizzabili” sull'asse temporale) vengono ricavate da una “funzione-madre” attraverso una dilatazione e una traslazione. La trasformazione da applicare alla funzione  $x(t)$  è la seguente:

$$W(a, b, x, \Psi) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t) \Psi \left( \frac{t-b}{a} \right) dt \quad (2.8)$$

dove  $\Psi$  è la funzione-madre,  $a$  è il fattore di dilatazione e  $b$  quello di traslazione. Anche se tempo e frequenza non compaiono esplicitamente nella trasformata, la variabile  $1/a$  ci dà la scala di frequenza, mentre  $b$  la locazione temporale di un evento.[8]

- *La trasformata di Karhunen-Loève (KLT)*

Conosciuta anche come *analisi delle componenti principali*, la KLT impiega un metodo statistico per scomporre un qualunque segnale  $x(t)$  come segue:

$$x(t) = \sum_{k=1}^{\infty} Z_k f_k(t) \quad (2.9)$$

Le  $f_k$  sono le autofunzioni ortogonali che risolvono la seguente equazione integrale

$$\int_0^T R_x(t_1, t_2) f_k(t_2) dt_2 = \lambda_k f_k(t_1)$$

dove  $R_x(t_1, t_2)$  rappresenta la funzione di autocorrelazione di  $x(t)$ .

Una volta ricavata la 2.9, è sufficiente scartare tutte le  $f_k$  corrispondenti agli autovalori  $\lambda_k$  più bassi di una certa soglia<sup>1</sup> (esse costituiscono la componente “casuale” di  $x(t)$ , cioè il rumore), e mantenere quelle corrispondenti a valori “alti” di  $\lambda_k$  (che invece costituiscono la componente “deterministica” di  $x(t)$ , cioè il segnale che si vuole rilevare). [9]

Si noti che la KLT presenta una differenza radicale rispetto ai metodi precedenti; infatti, le funzioni in cui viene scomposto il segnale non sono determinate a priori, ma derivate direttamente dai dati. Ritroveremo questa importante caratteristica analizzando la trasformata di Hilbert-Huang.

---

<sup>1</sup>ciò è possibile in quanto si dimostra che

$$\sum_{k=1}^{\infty} \lambda_k < \infty$$



# Capitolo 3

## La trasformata di Hilbert-Huang

Recentemente sviluppata da Norden E. Huang, scienziato del NASA Goddard Space Flight Center, la trasformata di Hilbert-Huang (HHT) è un nuovo metodo creato per analizzare segnali nonstazionari. Si tratta di un metodo empirico, non ancora supportato da solide fondamenta teoriche. Questo capitolo si pone l'obiettivo di descriverne il funzionamento, insieme ad una breve trattazione dei problemi ad essa associati e rimasti ancora aperti.

Questo metodo si basa sull'*Empirical Mode Decomposition*, un algoritmo che scompone un qualunque set di dati in un numero finito di funzioni, definite *Intrinsic Mode Functions*. Per ciascuna di queste possiamo definire, attraverso la trasformata di Hilbert, la *frequenza istantanea*. Il risultato finale della trasformata è dunque una distribuzione tempo-energia-frequenza, chiamata *Spettro di Hilbert*. [10][11]

### 3.1 Frequenza istantanea

Quando si ha a che fare con un seno o un coseno, o con un qualunque altro segnale periodico, il concetto di frequenza è semplice e intuitivo. Naturalmente, tuttavia, se il segnale è nonstazionario e quindi aperiodico, il suo significato dev'essere approfondito. Si può, certamente, definire la frequenza in senso locale (vedi appendice A), ma il concetto di frequenza *istantanea*, cioè definita punto per punto, è quantomai lontano dalla nostra concezione. Questo fatto, unito ai numerosi problemi sollevati dalla sua definizione, ha reso molto difficile l'accettazione della nozione di frequenza istantanea, diversamente da quanto avvenuto, ad esempio, per quella di energia istantanea.

### 3.1.1 Definizione di frequenza istantanea attraverso la trasformata di Hilbert

Sia  $x(t)$  un'arbitraria funzione di classe  $L^p$ . Possiamo definire la sua trasformata di Hilbert come

$$y(t) = \frac{1}{\pi} PV \int_{-\infty}^{\infty} \frac{x(\tau)}{t - \tau} d\tau \quad (3.1)$$

dove  $PV$  indica il valore principale di Cauchy dell'integrale<sup>1</sup>. Essenzialmente abbiamo definito la trasformata di Hilbert come la convoluzione tra  $x(t)$  e la funzione  $1/t$ . E' possibile dimostrare che tale definizione è ben posta per ogni funzione di classe  $L^p$ . Una trattazione completa della trasformata di Hilbert e delle sue proprietà può essere trovata in [12].

A questo punto, definiamo il segnale analitico  $z(t)$  come

$$z(t) = x(t) + iy(t) = a(t)e^{i\vartheta(t)} \quad (3.2)$$

dove  $a(t)$  rappresenta il modulo e  $\vartheta(t)$  la fase di  $z(t)$ . La trasformata di Hilbert ci fornisce un sistema univoco di definire la parte immaginaria della funzione  $z(t)$ , in modo che il risultato sia una funzione analitica. E' possibile dimostrare che lo spettro di Fourier di  $z(t)$  è lo stesso di  $x(t)$ , dal quale sono però assenti tutte le frequenze negative. In altre parole, è possibile ottenere  $z(t)$  rimuovendo dallo spettro di  $x(t)$  tutte e sole le frequenze negative e poi antitrasformando (e infatti i metodi numerici per calcolare la trasformata di Hilbert utilizzano proprio questo procedimento).

Con questa definizione, la frequenza istantanea del segnale  $x(t)$  è semplicemente

$$\omega(t) = \frac{d\vartheta(t)}{dt} \quad (3.3)$$

### 3.1.2 Significato della definizione

Cerchiamo ora di capire le motivazioni alla base di una definizione così complessa e apparentemente lontana dal significato della parola "frequenza". Per farlo, domandiamoci se esiste una relazione fra  $\omega(t)$  e le frequenze presenti nello spettro di Fourier del segnale.

---

<sup>1</sup>Il valore principale di Cauchy di un integrale improprio in cui l'insieme d'integrazione è l'insieme dei numeri reali è per definizione:

$$PV \int_{-\infty}^{\infty} f(x) dx = \lim_{\varepsilon \rightarrow 0^+} \left[ \int_{b-\frac{1}{\varepsilon}}^{b-\varepsilon} f(x) dx + \int_{b+\varepsilon}^{b+\frac{1}{\varepsilon}} f(x) dx \right]$$

dove  $b$  è un punto di singolarità per la funzione  $f(x)$

Definiamo preliminarmente la frequenza media del segnale  $z(t)$ , che supponiamo essere una funzione  $C_0^\infty$ , nel modo seguente:

$$\langle \omega \rangle = \int \omega |S(\omega)|^2 d\omega \quad (3.4)$$

dove  $S(\omega)$  rappresenta la trasformata di Fourier di  $z(t)$ . Il fattore  $|S(\omega)|^2$  rappresenta il “peso” della componente frequenziale  $\omega$ . Ora, ricordiamo la nota relazione che sussiste fra la derivata di una funzione e la sua trasformata di Fourier, vale a dire:

$$\omega S(\omega) = \frac{1}{i} \int \frac{dz(t)}{dt} e^{-i\omega t} dt \quad (3.5)$$

Osserviamo inoltre che vale

$$\overline{S(\omega)} = \int \overline{z(t)} e^{i\omega t} dt \quad (3.6)$$

Andando a sostituire la 3.5 e la 3.6 nella 3.4, si ottiene

$$\langle \omega \rangle = \frac{1}{i} \int \int \int \overline{z(t')} \frac{dz(t)}{dt} e^{i\omega(t'-t)} dt dt' d\omega$$

Da cui troviamo

$$\langle \omega \rangle = \frac{1}{i} \int \overline{z(t)} \frac{dz(t)}{dt} dt \quad (3.7)$$

Ora sostituiamo  $z(t) = a(t)e^{i\vartheta(t)}$  nella 3.7 per ottenere

$$\begin{aligned} \langle \omega \rangle &= \frac{1}{i} \int a(t) e^{-i\vartheta(t)} \left( \frac{da(t)}{dt} e^{i\vartheta(t)} + a(t) i \frac{d\vartheta(t)}{dt} e^{i\vartheta(t)} \right) dt = \\ &= \int -ia(t) \frac{da(t)}{dt} + \frac{d\vartheta(t)}{dt} |z(t)|^2 dt \end{aligned}$$

Siccome avevamo supposto  $z(t) \in C_0^\infty$ , il primo termine, se integrato per parti, si annulla. Abbiamo ricavato dunque la relazione

$$\langle \omega \rangle = \int \frac{d\vartheta(t)}{dt} |z(t)|^2 dt \quad (3.8)$$

Se ora identifichiamo il fattore  $|z(t)|^2$  come il “peso” della quantità  $d\vartheta(t)/dt$ , possiamo interpretare la 3.8 come un'identificazione fra la frequenza media e il valor medio di  $d\vartheta(t)/dt$ :

$$\langle \omega \rangle = \left\langle \frac{d\vartheta}{dt} \right\rangle$$

Questo fatto ha suggerito che  $d\vartheta(t)/dt$  dovesse essere trattato come la frequenza istantanea.

### 3.1.3 Discussione

Malgrado quanto detto nel paragrafo precedente, si potrebbe obiettare che la relazione trovata tra la frequenza istantanea e le frequenze dello spettro di Fourier concerne la media nell'intervallo d'integrazione, è quindi di tipo globale, e non c'è alcun legame di uno-a-uno fra i due tipi di frequenze. In effetti, se la definizione di frequenza istantanea è applicata su un segnale qualsiasi, possiamo facilmente incorrere nel paradosso di ottenere frequenze negative o non appartenenti allo spettro di Fourier, e dunque prive di qualunque significato fisico.

Ragionando empiricamente, sono state individuate alcune restrizioni da imporre sulle funzioni affinché il calcolo della frequenza istantanea dia risultati significativi (nemmeno queste restrizioni ci danno tuttavia la certezza matematica che ciò avvenga, come vedremo più avanti). Al fine di illustrare quali siano queste limitazioni, osserviamo un semplice esempio.

Consideriamo la funzione

$$x(t) = \sin(t)$$

La sua trasformata di Hilbert è semplicemente  $y(t) = \cos(t)$ . Il grafico di  $x(t) + iy(t)$  è la circonferenza di raggio unitario, come mostrato in figura 3.1a. Il grafico della fase è una retta, illustrata in figura 3.1b, e la sua frequenza istantanea, come ci aspettiamo, è una costante (figura 3.1c). Spostiamo ora la media della funzione di una certa quantità  $\alpha$ , ottenendo quindi

$$x(t) = \alpha + \sin(t)$$

Il grafico di  $x(t) + iy(t)$  è ancora una circonferenza di raggio unitario, ma il suo centro risulterà spostato della quantità  $\alpha$ , come illustrato in figura 3.1a. Se  $\alpha < 1$  l'origine è ancora all'interno della circonferenza. La frequenza media di attraversamenti dello zero della funzione è la stessa del caso precedente, ma la fase e la frequenza istantanea sono piuttosto differenti, come mostrato nelle figure 3.1b e 3.1c. Se  $\alpha > 1$ , l'origine si trova all'esterno del cerchio. Dunque sia la fase che la frequenza istantanea assumeranno valori negativi, e quindi privi di significato.

In un segnale qualsiasi, il caso di onde sovrainposte sarà, localmente, equivalente al caso  $\alpha > 1$ ; mentre il caso di oscillazioni di forma asimmetrica (rispetto all'asse temporale) saranno equivalenti al caso  $\alpha < 1$ .

## 3.2 Intrinsic Mode Functions

Le restrizioni appena trovate, grazie all'aiuto dei semplici esempi mostrati sopra, ci suggeriscono il modo di definire una classe di funzioni per cui la

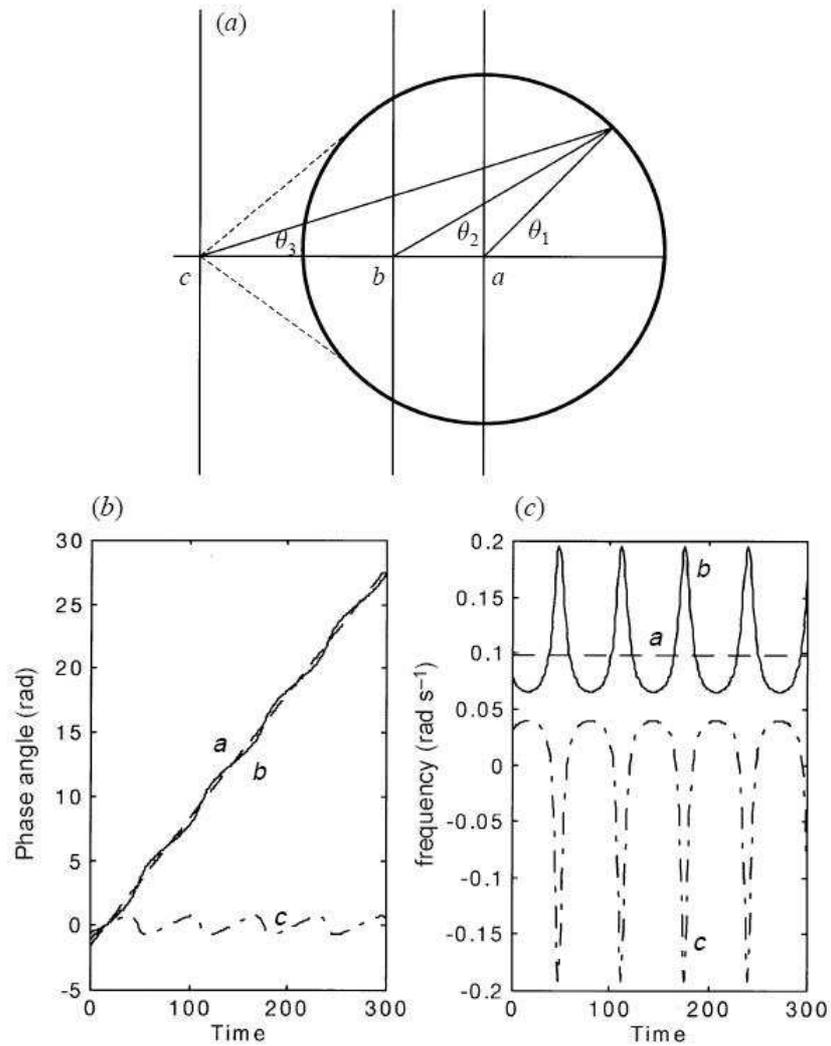


Figura 3.1: Interpretazione della frequenza istantanea. (a) Grafico del segnale analitico della funzione  $\alpha + \sin(t)$ . (a)  $\alpha = 0$ ; (b)  $\alpha < 1$ ; (c)  $\alpha > 1$ . (b) La fase della funzione  $x(t)$ . (c) La frequenza istantanea calcolata attraverso la 3.3

frequenza istantanea possa essere definita con successo in ogni punto. Definiamo dunque Intrinsic Mode Function (IMF) una funzione che soddisfi i seguenti requisiti:

1. Nell'intero set di dati, il numero di estremanti e di attraversamenti dello zero devono essere uguali o differire al più di uno;
2. in ogni punto, il valor medio dell'involuppo superiore e di quello inferiore<sup>2</sup> dev'essere zero.

La condizione (1) ci assicura che nel segnale non vi siano onde sovrainposte, o in altre parole che il segnale sia “monocomponente”; la (2) che la frequenza istantanea non abbia “fluttuazioni” dovute ad un profilo d'onda asimmetrico. Notiamo che nella pratica quest'ultima condizione diviene una semplice approssimazione: la media non sarà esattamente zero, ma si avvicinerà molto a tale valore.

Una IMF rappresenta una singola modalità di oscillazione; è la controparte di una semplice funzione armonica, ma la sua definizione è molto più generale; al posto di una frequenza e di un'ampiezza costanti, come nelle funzioni armoniche, una IMF può avere frequenza e ampiezza variabili nel tempo; si tratta dunque a tutti gli effetti di segnali nonstazionari.

Naturalmente, la maggior parte dei segnali non sono IMF; in ogni punto, può essere presente più di una modalità di oscillazione, e ciò impedisce alla trasformata di Hilbert di dare una descrizione accurata del segnale. Introduciamo dunque un metodo per scomporre un qualsiasi segnale in una somma di IMF.

### 3.3 Empirical Mode Decomposition

Sia  $x(t)$  un segnale arbitrario, che supponiamo possedere almeno un punto di massimo e uno di minimo. Per prima cosa, si costruisce la spline cubica passante per i tutti i punti di massimo e quella passante per i punti di minimo come mostrato in figura 3.2, per produrre l'involuppo superiore e inferiore di  $x(t)$ . Chiamata ora  $m_1$  la loro media, la differenza fra  $x(t)$  e  $m_1$  è la prima componente  $h_1$ , mostrata in figura 3.3,

$$h_1 = x(t) - m_1$$

---

<sup>2</sup>Dal testo originale di Norden E. Huang appare chiaro che i termini “involuppo inferiore” e “involuppo superiore” indichino le spline passanti per i punti di massimo o di minimo della funzione. Tuttavia, ho l'impressione che questi termini non vengano usati con rigore matematico, in quanto (nonostante l'algoritmo si basi su quelle cubiche) esistono numerosi tipi di spline, ognuno dei quali produrrebbe involuppi leggermente differenti

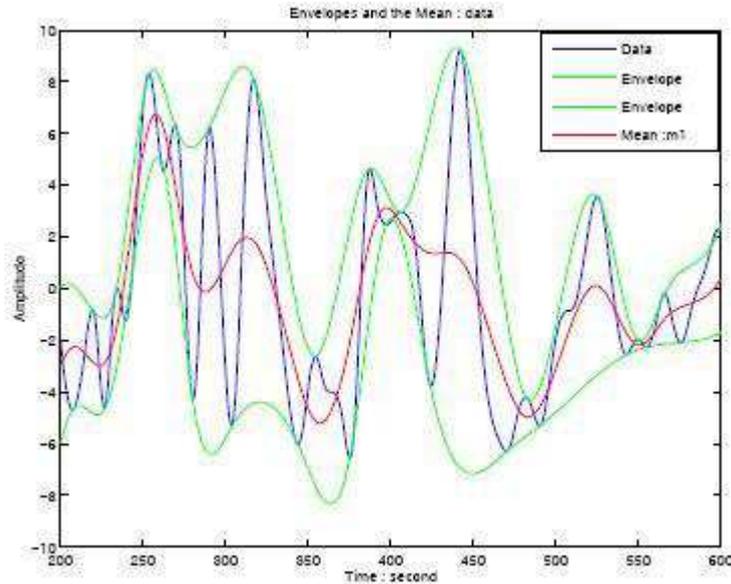


Figura 3.2: I dati (linea blu), l'involuppo superiore e inferiore (linea verde), individuati rispettivamente dai punti di massimo e di minimo, e il loro valor medio (linea rossa)

Idealmente,  $h_1$  dovrebbe soddisfare la definizione di IMF, poichè la sua costruzione descritta sopra dovrebbe averla resa simmetrica e avere tutti i massimi positivi e i minimi negativi. In realtà, tuttavia, non trattandosi di un metodo matematicamente rigoroso, questo procedimento dovrà essere reiterato più volte per ottenere una IMF. Nel secondo passaggio,  $h_1$  dovrà essere trattato come il nuovo segnale; si avrà dunque

$$h_1 - m_{11} = h_{11}$$

Dopo aver ripetuto questo processo un numero opportuno di volte, sia  $k$ , si ottiene una IMF, come mostrato in figura 3.4

$$h_{1k} = h_{1(k-1)} - m_{1k}$$

A questo punto poniamo

$$c_1 = h_{1k}$$

che rappresenta la prima IMF estratta dai dati.

Il procedimento appena descritto serve a due scopi: eliminare le onde sovrainposte e rendere il profilo d'onda più simmetrico. Sfortunatamente, il

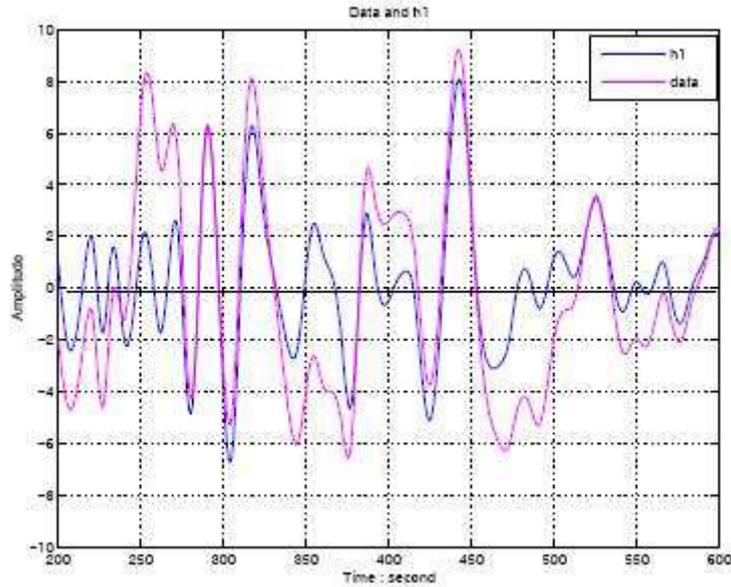


Figura 3.3: I dati (linea viola) e  $h_1$  (linea blu)

secondo effetto, portato all'estremo, potrebbe rendere la IMF risultante un segnale di ampiezza quasi costante, eliminando quindi anche le fluttuazioni in ampiezza fisicamente significative.

Per evitare questo, è necessario determinare un criterio che stabilisca quando terminare il processo. Questo può essere fatto limitando il valore della deviazione standard calcolata fra i risultati di due passaggi consecutivi. Più precisamente

$$SD_k = \sum_{t=0}^T \left[ \frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_{1(k-1)}^2(t)} \right] \quad (3.9)$$

dev'essere più piccolo di certo valore predeterminato (di solito posto fra 0.2 e 0.3), perchè il processo si arresti. Notiamo che, nonostante finora abbiamo sempre assunto che la variabile temporale  $t$  assumesse valori su un insieme continuo, nella pratica  $t$  può assumere solo un numero finito di valori discreti (in quanto si tratta del tempo di campionamento dei dati); a ciò è dovuta la presenza della sommatoria nella 3.9.

Assumiamo ora che il criterio sia stato soddisfatto e che la prima IMF,  $c_1$ , sia stata trovata. In base a come è stata costruita,  $c_1$  dovrebbe essere la modalità di oscillazione più veloce presente nei dati. A questo punto  $c_1$  può

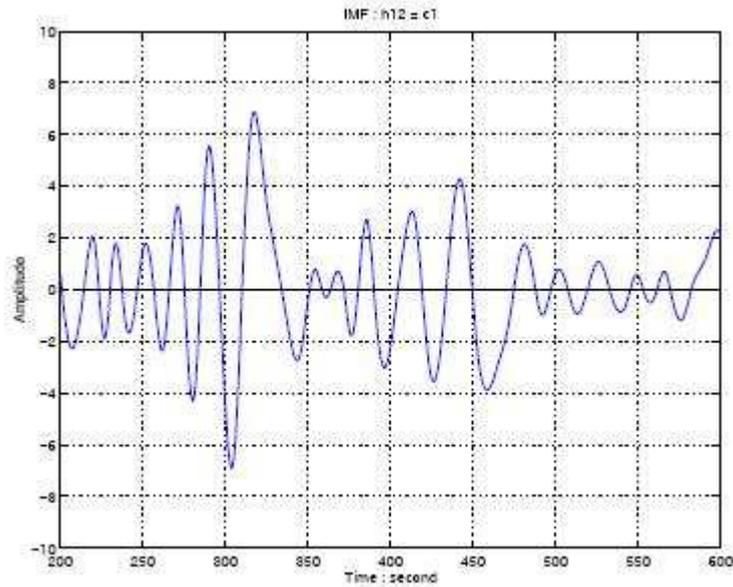


Figura 3.4: La prima IMF  $c_1$  dopo 12 passaggi

essere separata dal resto dei dati come segue

$$r_1 = x(t) - c_1 \quad (3.10)$$

Siccome il residuo  $r_1$  in generale contiene ancora oscillazioni di periodo più lungo, come mostrato in figura, esso è trattato come i nuovi dati e sottoposto allo stesso procedimento illustrato sopra. In questo modo, otterremo

$$\begin{aligned} r_2 &= r_1 - c_2 \\ &\vdots \end{aligned} \quad (3.11)$$

$$r_n = r_{n-1} - c_n$$

Questo procedimento può essere interrotto usando uno qualsiasi dei seguenti criteri: quando la IMF  $c_n$  (oppure  $r_n$ ) diventa più piccola di un valore predefinito, o quando il residuo  $r_n$  diviene una funzione priva di un punto di massimo e uno di minimo, dalla quale più nessuna IMF può essere estratta. Sommando la 3.10 e la 3.11, otteniamo finalmente

$$x(t) = \sum_{j=1}^n c_j + r_n \quad (3.12)$$

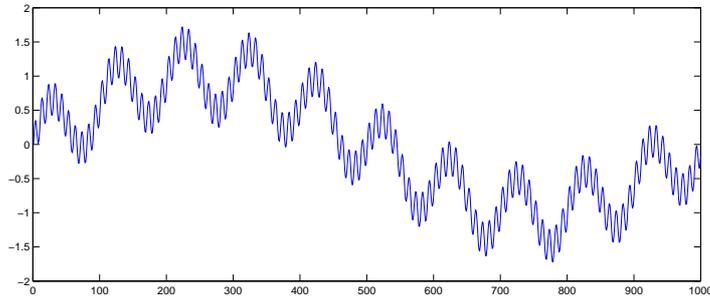


Figura 3.5: Sovrapposizioni di tre sinusoidi di diversa ampiezza e frequenza

Abbiamo dunque ottenuto una decomposizione di  $x(t)$  in  $n$  modalità di oscillazione e un residuo  $r_n$ .

A titolo esplicativo dell'intero procedimento, si consideri il semplice esempio di una sovrapposizione di sinusoidi di diversa ampiezza e frequenza, mostrato in figure 3.5 e 3.6 <sup>3</sup>.

### 3.4 Spettro di Hilbert

Avendo ottenuto le componenti IMF del segnale  $x(t)$ , non abbiamo ora nessun impedimento nell'applicare la trasformata di Hilbert ad ogni componente e calcolarne la frequenza istantanea, concordemente all'equazione 3.3. Dopo aver applicato la trasformata di Hilbert, possiamo esprimere i dati iniziali nel modo seguente:

$$x(t) = \sum_{j=1}^n a_j(t) \exp \left( i \int \omega_j(t) dt \right) \quad (3.13)$$

Qui abbiamo trascurato il residuo  $r_n$ , di proposito, poichè si tratta verosimilmente di una funzione monotona o di una costante.

L'equazione 3.13 ci permette inoltre di rappresentare l'ampiezza e la frequenza istantanea come funzioni del tempo in un grafico a tre dimensioni, in cui l'ampiezza può essere individuata sul piano tempo-frequenza. Essendo tuttavia il valore al quadrato dell'ampiezza la maniera più comunemente usata per esprimere la densità di energia, tale valore può essere sostituito all'ampiezza per ottenere quello che è definito lo spettro di Hilbert,  $H(\omega, t)$ .

---

<sup>3</sup>Questi grafici sono stati realizzati utilizzando un semplice algoritmo implementato su MATLAB che effettua l'EMD (vedi Appendice B)

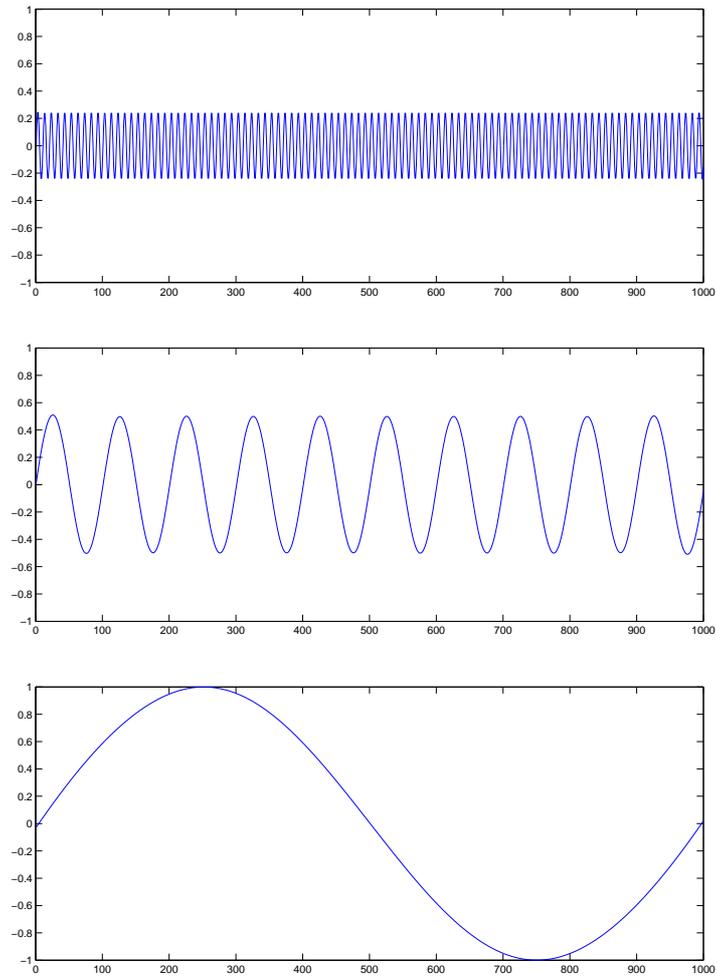


Figura 3.6: Le tre IMF ottenuto dopo aver applicato l'EMD, nell'ordine in cui vengono trovate. Il residuo è in questo caso la funzione identicamente nulla

E' anche possibile definire lo *spettro marginale*,  $h(\omega)$ , come

$$h(\omega) = \int_0^T H(\omega, t) dt$$

Lo spettro marginale ci offre una misura del contributo all'energia totale di ogni frequenza. Questo spettro rappresenta l'energia accumulata lungo tutto il set di dati in un senso probabilistico. Qui, la presenza di energia ad una certa frequenza significa che, all'interno del segnale, c'è un'alta probabilità che quella frequenza si sia manifestata localmente.

## 3.5 Discussione e problemi matematici aperti

Ora che abbiamo illustrato il funzionamento della trasformata di Hilbert-Huang, proviamo ad applicarla al segnale mostrato in figura 2.4, di cui ci eravamo serviti, nel capitolo 2, per mostrare l'inadeguatezza della trasformata di Fourier nell'analisi di segnali nonstazionari. In figura 3.7 è mostrato lo spettro di Hilbert del segnale; possiamo osservare che in questo caso sono chiaramente visibili sia l'aumentare della frequenza, sia il diminuire dell'energia.

Tuttavia, essendo la HHT un metodo empirico, privo di una consolidata base teorica, essa rimane aperta ad una serie di problemi di carattere matematico e numerico ancora solo in parte risolti. Questo paragrafo si pone l'obiettivo di esaminare questi problemi, e le eventuali soluzioni finora individuate.

### 3.5.1 Trasformata di Hilbert normalizzata

Come già discusso, malgrado la trasformata di Hilbert esista per ogni funzione di classe  $L^p$ , la frequenza istantanea può non possedere alcun significato fisico. Aver ridotto il campo d'azione alle sole IMF ha grandemente aumentato le probabilità di ottenere frequenze istantanee significative. Tuttavia la definizione di IMF soddisfa solo una condizione necessaria, e possono talvolta manifestarsi aberrazioni, come frequenze istantanee negative, in particolare se il segnale presenta grosse fluttuazioni in ampiezza. Devono dunque essere imposte condizioni aggiuntive.

Il teorema di Bedrosian (1963) afferma che la trasformata di Hilbert del prodotto di due funzioni,  $f(t)$  e  $h(t)$ , può essere scritta come

$$H[f(t)h(t)] = f(t)H[h(t)] \quad (3.14)$$

solo se gli spettri di Fourier di  $f$  e di  $h$  sono totalmente disgiunti nel dominio delle frequenze. Ora, ricordando le relazioni 3.1 e 3.2, possiamo esprimere il

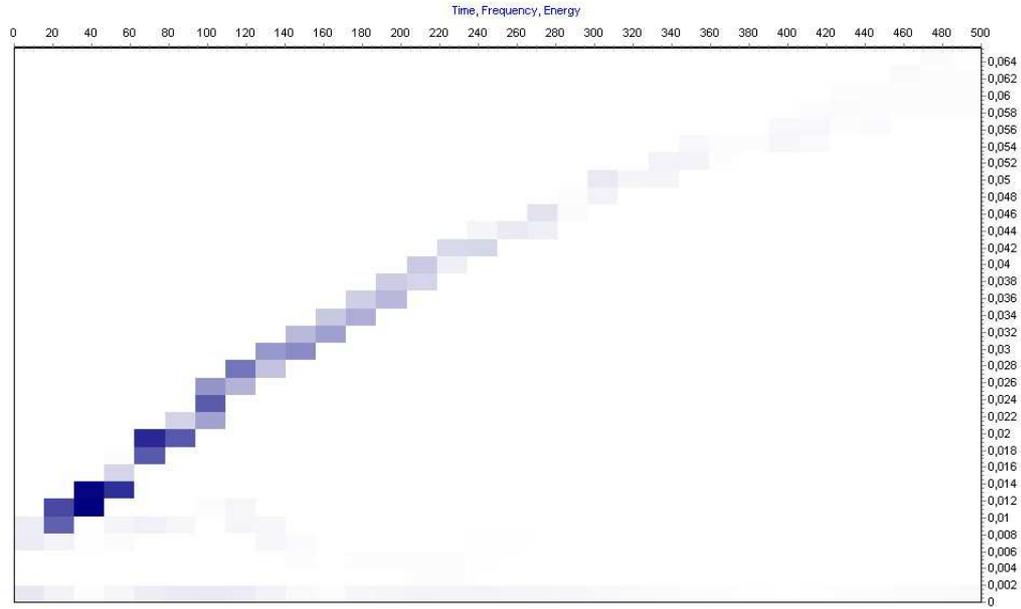


Figura 3.7: spettro di Hilbert del segnale in figura 2.4

segnale  $x(t)$  nella forma

$$x(t) = a(t)\cos[\vartheta(t)]$$

A questo punto, se applichiamo la 3.14, otteniamo

$$H\{a(t)\cos[\vartheta(t)]\} = a(t)H\{\cos[\vartheta(t)]\} \quad (3.15)$$

Perchè la relazione 3.15 è così importante? Al fine di ottenere un valore significativo per la frequenza istantanea, tale valore non può dipendere dall'ampiezza del segnale (rappresentata dalla funzione  $a(t)$ ), ma unicamente dalla sua modalità di oscillazione (rappresentata da  $\cos[\vartheta(t)]$ ).

In ogni caso, concordemente con il teorema di Bedrosian, la 3.15 può essere vera solo se l'ampiezza varia così lentamente che gli spettri di  $a(t)$  e  $\cos[\vartheta(t)]$  sono disgiunti. Per soddisfare questo requisito, si è proposto che, prima dell'applicazione della trasformata di Hilbert, le IMF vengano normalizzate come segue: anzitutto, si costruisce la spline passante per i massimi della IMF, sia  $E(t)$ ; poi, si normalizza la funzione dividendola per  $E(t)$ :

$$C(t) = \frac{x(t)}{E(t)} \quad (3.16)$$

Idealmente,  $C(t)$  dovrebbe avere ampiezza sempre uguale ad uno (e quindi soddisfare le condizioni per poter applicare la 3.15), ma chiaramente esistono casi anomali, in cui possono sorgere complicazioni nell'interpolazione nei punti in cui la fluttuazione è grande. Qui la spline potrebbe momentaneamente passare sotto ai dati, e far sì che in quei punti l'ampiezza della funzione normalizzata sia maggiore di uno. Malgrado questi casi siano rari, possono capitare.

Si noti che la normalizzazione espressa dalla 3.16 viene applicata unicamente per il calcolo della frequenza istantanea. Il contributo all'energia dello spettro di Hilbert rimane quello dato dalla funzione  $x(t)$ .

### 3.5.2 Problemi sulle spline

Innanzitutto, fra i tipi di spline utilizzabili, qual è il migliore? Come indicato nel processo dell'EMD, l'interpolazione è un passaggio fondamentale nella generazione delle IMF; e assume ancora maggiore importanza se consideriamo che non esiste alcun teorema che ci assicuri la convergenza della successione di funzioni  $h_{ik}$ ,  $i = 1, 2 \dots n$ , ad una IMF.

Nonostante il presente metodo, che utilizza spline cubiche, funzioni abbastanza bene nella maggior parte dei casi, talvolta possono manifestarsi dei problemi (vale a dire che l'algoritmo restituisce una funzione che non è una IMF). Vantaggi e svantaggi dell'utilizzo di spline di ordine superiore, sebbene finora siano stati giudicati marginali, non sono ancora stati definitivamente stabiliti e quantificati.

### 3.5.3 Ottimizzazione sulla scelta delle IMF

Dal punto di vista teorico, il set di IMF che possono essere estratte da un segnale attraverso l'EMD non è unico. Al contrario, esistono infiniti modi di scomporre un determinato set di dati. L'esperienza mostra che l'EMD può generare diversi set di IMF al variare dei parametri del processo. Un esempio è la scelta del criterio con cui terminare l'algoritmo di estrazione di una singola IMF, poichè troppi passaggi possono rovinare il significato fisico del segnale, e troppo pochi non permettono di ottenere una IMF pulita. Nonostante, come espresso nel paragrafo 3.3, sia già stato definito un criterio, basato sull'osservazione sperimentale, il problema dell'ottimizzazione delle IMF rimane aperto. Non solo, ma la domanda sull'unicità delle IMF rimanda ad un'altra, più generale: quale potrebbe essere un modo di definire le IMF più rigorosamente?

### 3.5.4 Altri problemi

Sia l'interpolazione spline che i metodi numerici utilizzati per il calcolo della trasformata di Hilbert possono manifestare indesiderati effetti agli estremi. Questi problemi sono stati finora risolti aggiungendo un'oscillazione caratteristica ad entrambi gli estremi.

Infine, l'algoritmo si basa sull'assunto che le differenti componenti di un segnale abbiano scale temporali distinte, e quindi scindibili attraverso l'EMD. Se tuttavia queste componenti hanno frequenze molto vicine fra loro, la decomposizione può non essere in grado di separarle, e dunque considerarle un'unica modalità di oscillazione. Si tratta, forse, della limitazione più grande della trasformata di Hilbert-Huang.



# Capitolo 4

## Validazione

Abbiamo finora illustrato il funzionamento della trasformata di Hilbert-Huang, insieme ad una panoramica dei problemi, sia teorici che pratici, ad essa legati. Poniamoci ora il problema di capire se e come la HHT può essere utile nello specifico caso dell'analisi di segnali radioastronomici. Per fare questo è indispensabile parlare del *rumore*.

### 4.1 Il problema del rumore

Come già accennato nel paragrafo 2.2, i dati sperimentali sono sempre contaminati da disturbi provenienti da radiosorgenti naturali (estremamente più potenti di qualunque onda elettromagnetica di origine artificiale), come le stelle, e da interferenze di origine terrestre (fulmini, canali televisivi, sonori, eccetera). Tali disturbi possono essere rimossi attraverso metodi opportuni, ma l'eliminazione non è mai netta come si vorrebbe. Ad esempio, la vicinanza del Sole può produrre un innalzamento dell'energia dello spettro in tutte le frequenze, alzando quindi il livello del rumore.

Il segnale che si vuole rilevare risulterà quindi impossibile da individuare senza l'utilizzo di appositi strumenti matematici. Non solo, ma non c'è modo di sapere a priori nemmeno se in un particolare set di dati, apparentemente casuali, il segnale è presente o meno. La capacità di una particolare trasformata di "estrarre" un segnale debole dal rumore di fondo è dunque di primaria importanza per i nostri fini.

Per quanto detto, possiamo scrivere

$$x(t) = s(t) + n(t) \tag{4.1}$$

dove  $x(t)$  rappresenta i dati,  $s(t)$  il segnale che ci interessa rilevare, e  $n(t)$  la componente casuale dovuta al rumore o *noise*.

Esiste un valore ben preciso che misura in che quantità il segnale è alterato dal rumore: tale valore, chiamato *rapporto segnale-rumore* o SNR (Signal to Noise Ratio), è definito nel modo seguente:

$$SNR = \frac{P_s}{P_n} \quad 0 \leq SNR \leq \infty \quad (4.2)$$

dove  $P_s$  e  $P_n$  sono rispettivamente la potenza (energia per unità di tempo) di  $s(t)$  e di  $n(t)$ .

E' evidente che non possiamo applicare la HHT così com'è su una funzione  $x(t)$  in cui il rapporto segnale-rumore non sia molto alto; quello che troveremmo sarebbe uno spettro di Hilbert incomprensibile, nel quale l'energia del segnale si confonde con quella del rumore. E' necessario trovare un metodo per scomporre il segnale in IMF in modo da poter distinguere quelle che contengono informazione utile da quelle che invece sono unicamente un prodotto del rumore.

## 4.2 Intermittency: un metodo per separare il segnale dal rumore

L'*Intermittency* è un metodo che permette di rimuovere certe frequenze da una IMF; esso entra in gioco durante il processo di estrazione della IMF dai dati, dopo che è stato soddisfatto il criterio di arresto dell'algoritmo (espresso dalla 3.9), ma prima che il risultato venga sottratto dai dati (equazione 3.10).

L'*Intermittency* funziona nel modo seguente: anzitutto vengono calcolate le distanze tra gli attraversamenti dello zero della IMF. Se una di queste distanze risulta maggiore di un valore predeterminato, tutta la sezione della IMF compresa tra i due attraversamenti dello zero viene azzerata. Altrimenti, quella sezione del segnale viene mantenuta. In altre parole, della IMF originale vengono prese unicamente le oscillazioni più veloci.

La figura 4.1 ci mostra come opera l'*Intermittency*; la linea tratteggiata blu rappresenta la IMF prima della sua applicazione, quella rossa la funzione risultante.

Per capire meglio come useremo questo procedimento, osserviamo ora il segnale in figura 4.2. Applicando l'*Intermittency*, riusciamo a scomporlo in base alle diverse velocità delle modalità di oscillazione, come mostrato in figura 4.3.

In che modo questo algoritmo può esserci utile nell'eliminazione del rumore? Siccome si suppone che la funzione  $n(t)$ , essendo composta di valori casuali, "oscilli" più velocemente di  $s(t)$ , applicando l'*Intermittency* (dopo

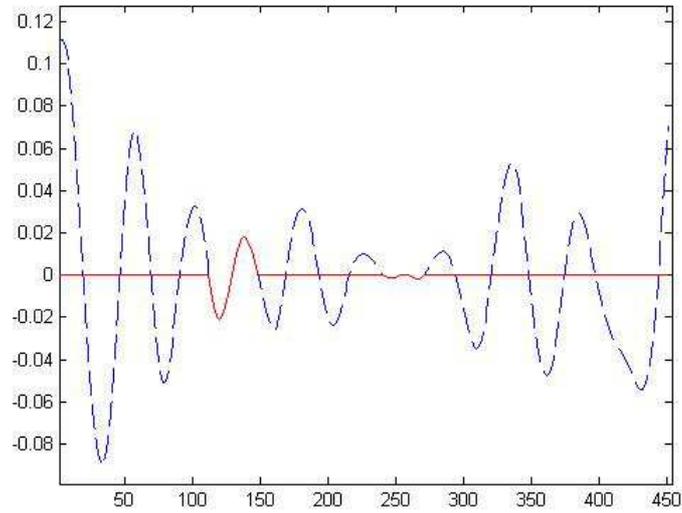


Figura 4.1: esempio di una IMF prima e dopo l'applicazione dell'Intermittency

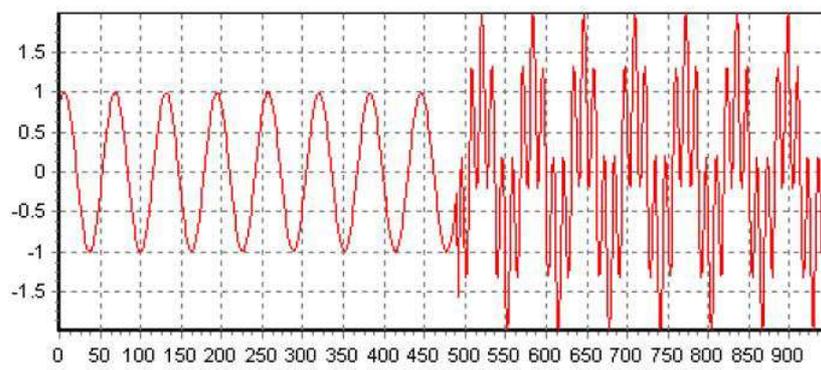


Figura 4.2: segnale

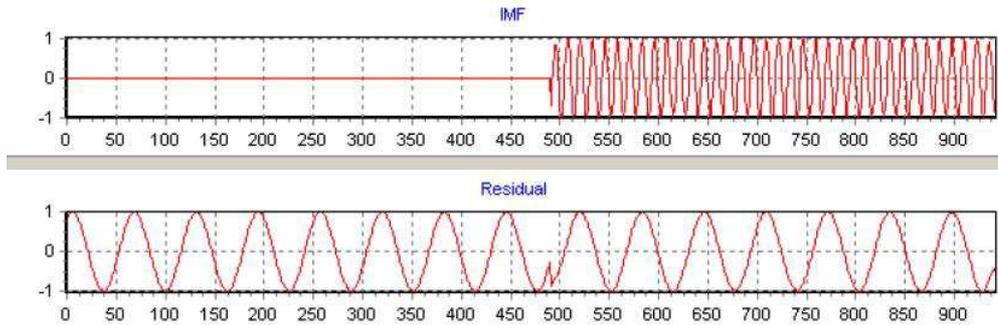


Figura 4.3: risultato ottenuto dopo l'applicazione dell'Intermittency; il segnale in alto rappresenta la IMF estratta dal segnale

aver selezionato un opportuno valore di soglia) dovremmo essere in grado di sottrarre ad  $x(t)$  tutta la componente dovuta al rumore.

Naturalmente, per poter applicare con successo questo procedimento, è necessario supporre le scale di tempo del segnale e quelle del rumore siano ben distinte. Non solo, ma se il rumore è alto, l'utilizzo dell'Intermittency descritto sopra rischia di compromettere anche il segnale; è quindi indispensabile effettuare alcune prove.

### 4.3 Test

Abbiamo generato un segnale sinusoidale rappresentato dalla funzione

$$s(t) = \sin\left(\frac{\pi}{100}t\right)$$

del quale prendiamo un campionamento di 1000 punti, il cui periodo supponiamo essere, per semplicità,  $\Delta t = 1$  s (la sua frequenza è dunque 0.005 Hz). Abbiamo quindi immerso il segnale in diversi livelli di rumore<sup>1</sup>.

1. SNR=5,9725 (figura 4.4)
2. SNR=1,4632 (figura 4.5)
3. SNR=0,6535 (figura 4.7)
4. SNR=0,3840 (figura 4.8)

---

<sup>1</sup>Il rumore è stato generato utilizzando una funzione `textitrandom` con distribuzione uniforme su un intervallo centrato sullo zero e ampiezza che varia caso per caso

## 5. SNR=0,0623 (figura 4.9)

Per ognuno di questi, le figure ci mostrano<sup>2</sup>, nell'ordine dall'alto verso il basso

- Un grafico in cui la linea verde rappresenta la ricostruzione del segnale ottenuto dall'applicazione dell'Intermittency, avendo fissato una soglia di 40 punti (si noti che, per come abbiamo campionato la funzione  $s(t)$ , la distanza fra due attraversamenti dello zero è sempre di 100 punti). La linea rossa rappresenta invece il risultato ottenuto applicando l'Intermittency sui dati in assenza del segnale, cioè sul solo rumore
- lo spettro di Hilbert del segnale ottenuto
- lo spettro ricavato applicando al segnale un metodo alternativo per il calcolo della frequenza, il *Generalized Zero Crossing* (si veda l'appendice A per una descrizione di questo algoritmo)

Nel primo caso (figura 4.4), possiamo notare che la ricostruzione del segnale è apparentemente molto buona. Tuttavia, lo spettro di Hilbert ci mostra che il segnale, pur rimanendo centrato sulla frequenza 0,005 Hz, assume frequenze istantanee che vanno da 0,003 Hz a 0,006 Hz. Tali fluttuazioni, pur mostrandoci la sensibilità della trasformata di Hilbert-Huang, sono indesiderate in quanto sono unicamente un effetto collaterale del processo di eliminazione del rumore.

Confrontando questo spettro con quello ottenuto con il GZC, notiamo che quest'ultimo presenta fluttuazioni molto meno marcate (si osservino i valori assunti sull'asse delle frequenze). Ciò è una diretta conseguenza del fatto che in questo algoritmo la frequenza istantanea non è calcolata “puntualmente”, ma “localmente”, e ciò rende i valori ottenuti maggiormente mediati. Siccome questa è la sola differenza sostanziale fra i due spettri, e che tale differenza rimane immutata nei casi seguenti, d'ora in poi ci limiteremo a commentare lo spettro ottenuto tramite il GZC.

Nel secondo caso (figura 4.5), cominciano a vedersi alcune distorsioni rispetto alla sinusoide originale; ne segue che lo spettro di Hilbert presenta fluttuazioni leggermente maggiori rispetto al caso precedente. Nel terzo e nel quarto caso (figure 4.7 e 4.8) le alterazioni del segnale sono ormai evidenti, soprattutto agli estremi, e ciò si traduce in grosse “impennate” della frequenza.

---

<sup>2</sup>Questi test sono stati effettuati usando il software HHT-DPS (*Data Processing System*), sviluppato da parte della NASA e concesso per un periodo di 60 giorni all'Istituto di Radioastronomia (IRA) dell'INAF

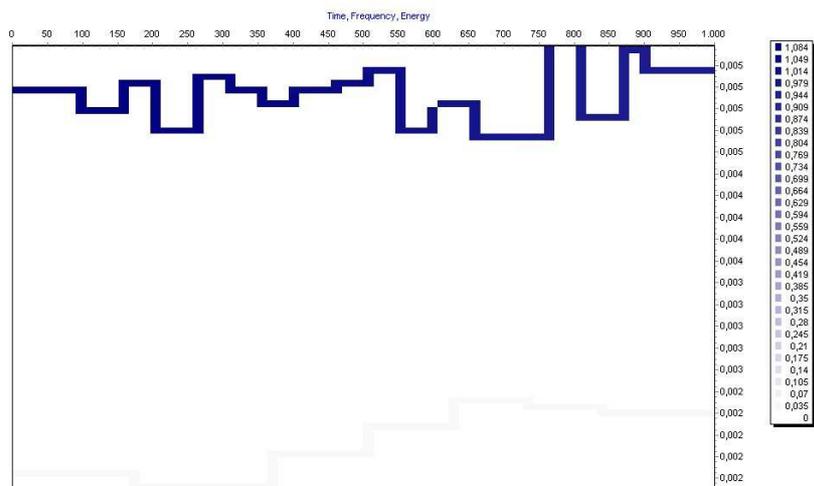
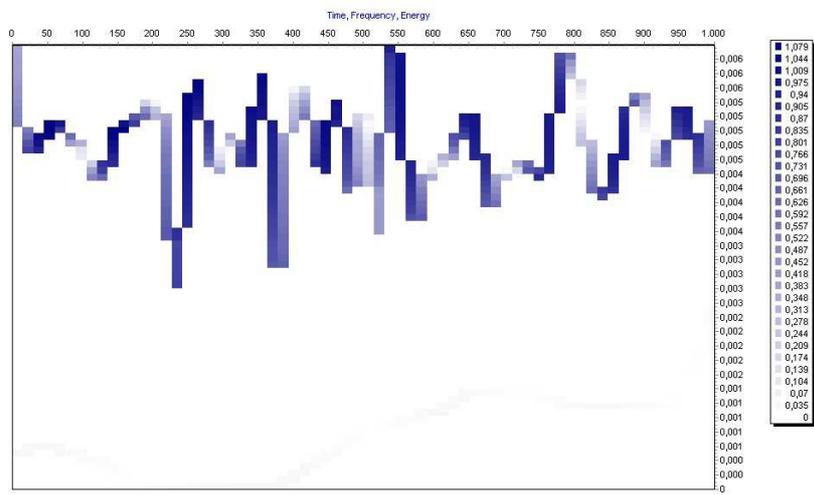
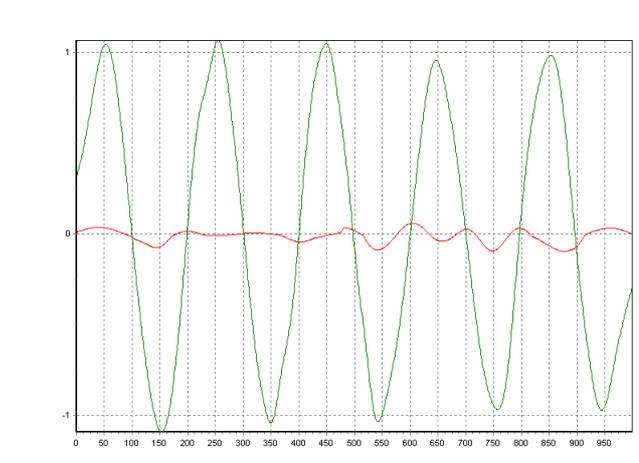


Figura 4.4: caso 1

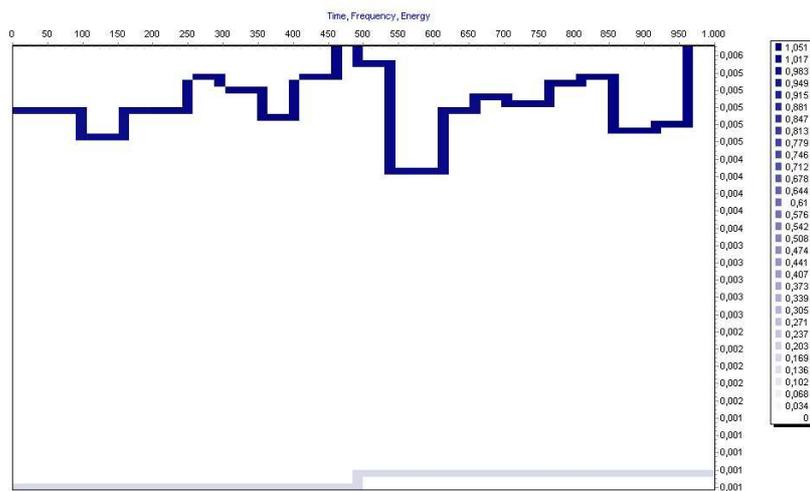
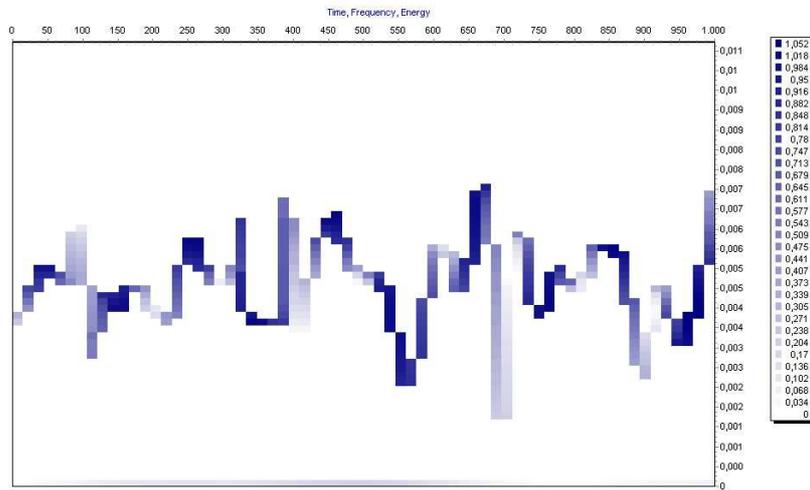
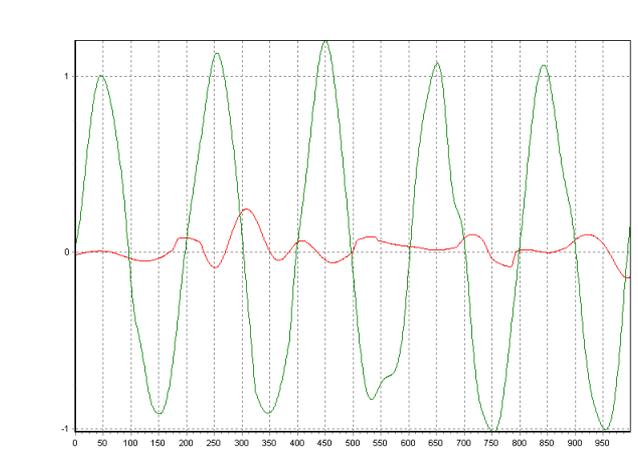


Figura 4.5: caso 2

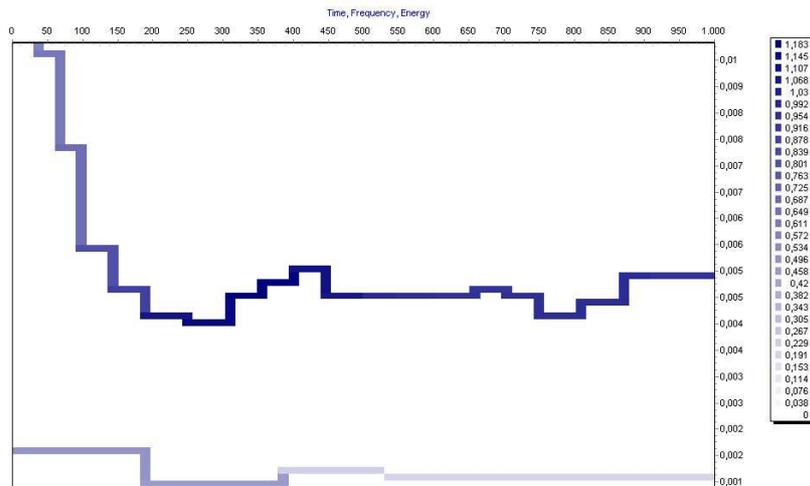
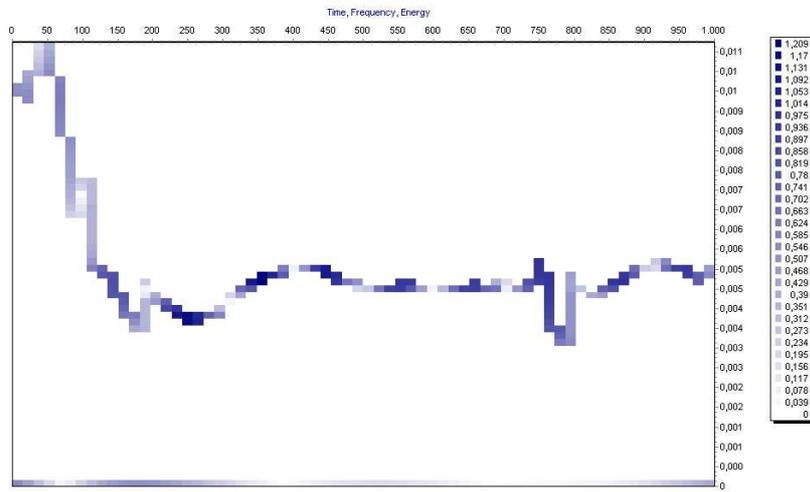
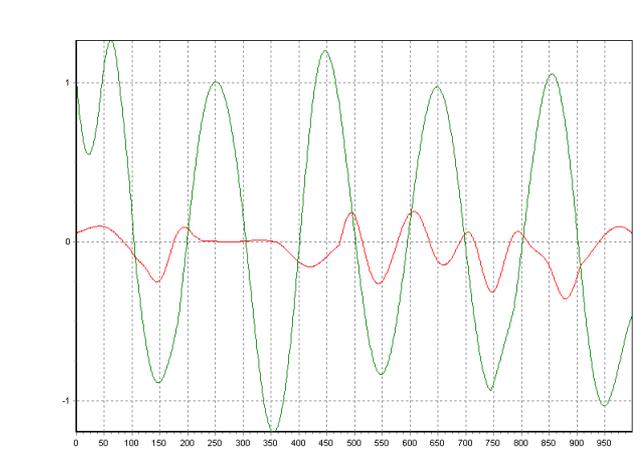


Figura 4.6: caso 3

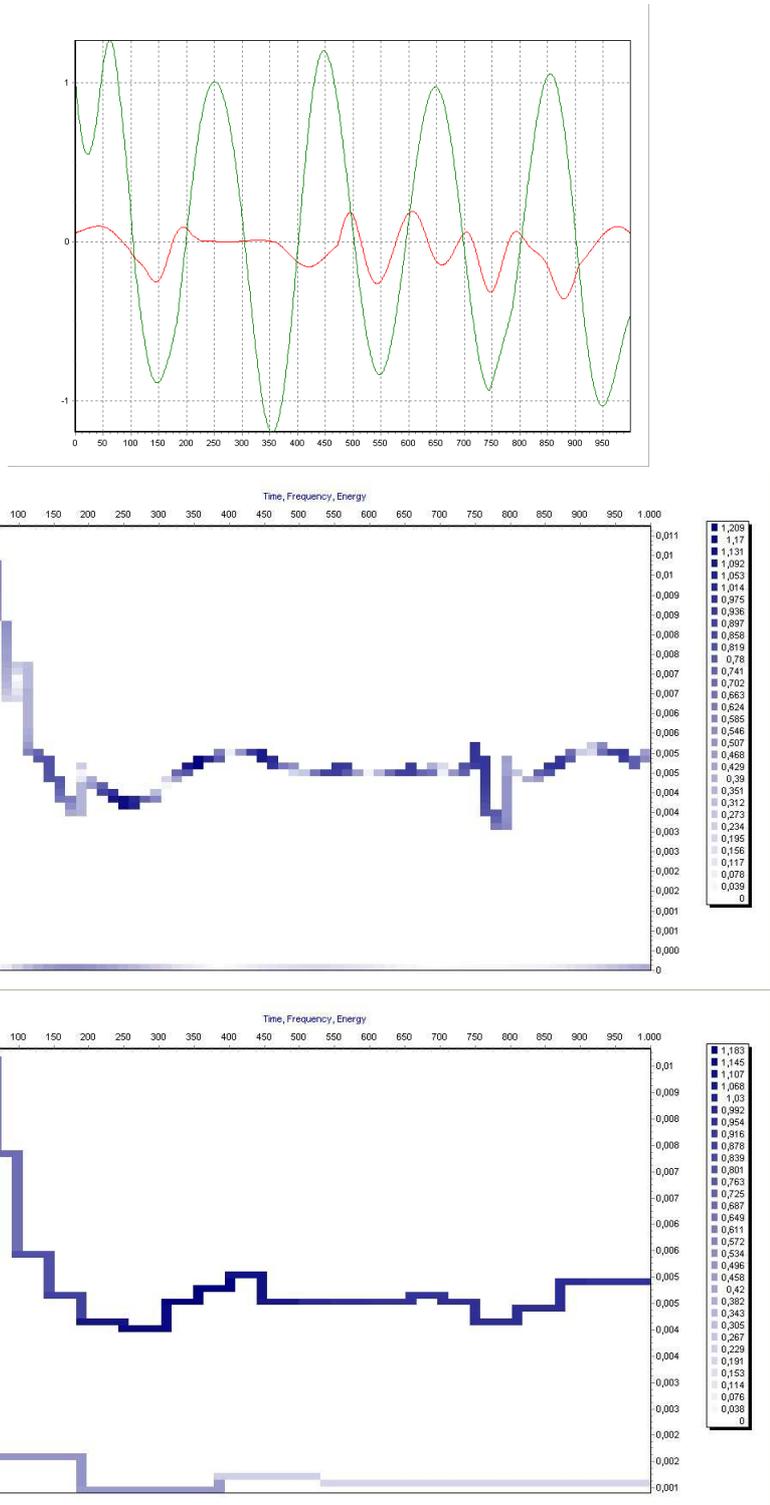


Figura 4.7: caso 3

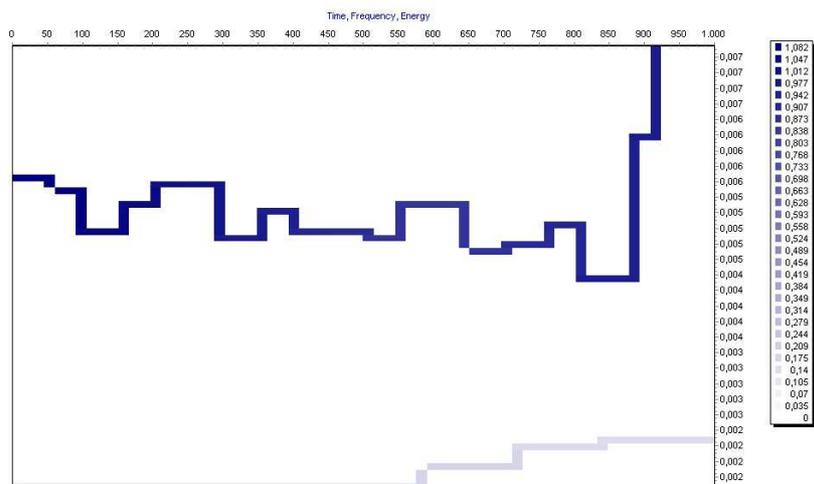
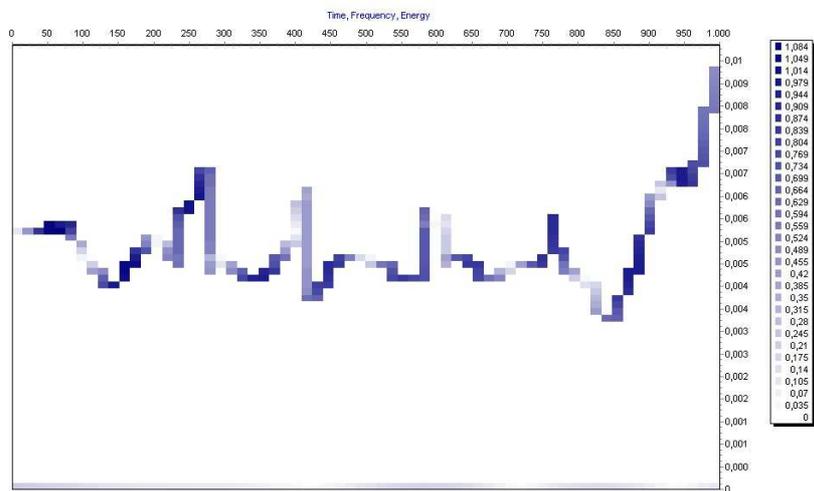
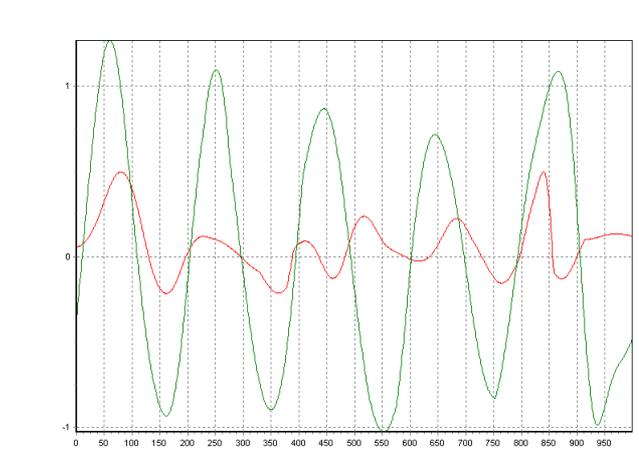


Figura 4.8: caso 4

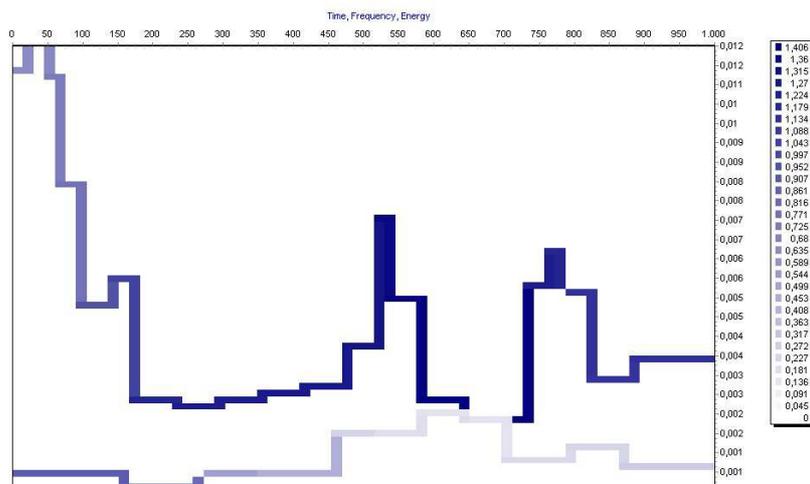
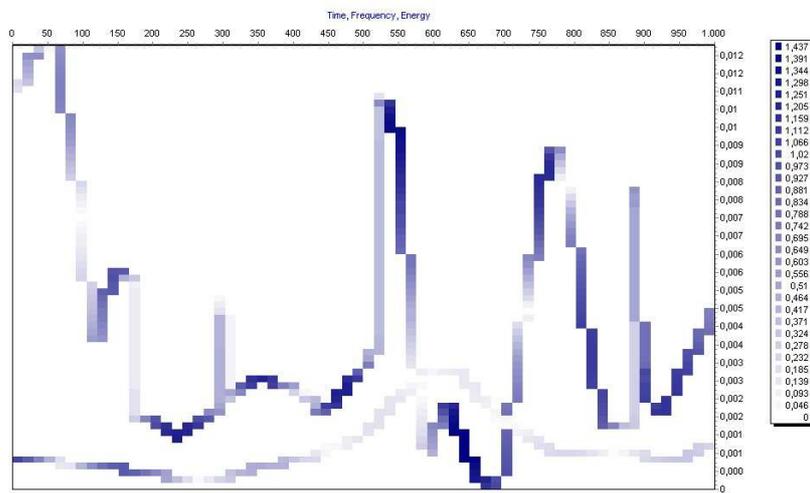
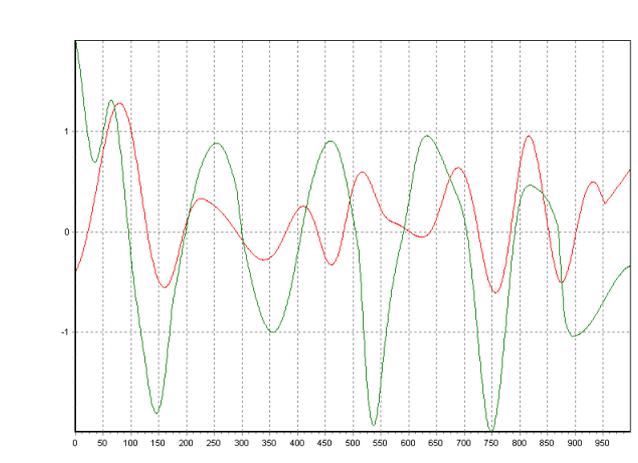


Figura 4.9: caso 5

Si noti che in tutti gli spettri eccetto il primo compare, in basso, una nuova componente, sempre più marcata a mano a mano che il rapporto segnale-rumore si abbassa. Ciò è una conseguenza del fatto che al segnale ottenuto dopo l'Intermittency viene applicata l'EMD, che inevitabilmente scompone il segnale in più componenti.

Nell'ultimo caso (figura 4.9), il segnale è completamente diverso dall'originale; anche lo spettro di Hilbert ci conferma questo risultato. Abbiamo ormai perso qualunque informazione utile sulla sinusoide originale.

Inoltre, osservando come varia il segnale che si ottiene applicando l'algoritmo sul solo rumore (linea rossa nei grafici), si è notato che la sua ampiezza cresce proporzionalmente all'aumentare dell'ampiezza di quest'ultimo. Da questa osservazione, possiamo facilmente concludere che, applicando l'Intermittency su un set di dati arbitrario, qualora ottenessimo come risultato un segnale in cui l'ampiezza massima è molto inferiore rispetto a quella dei dati iniziali (rapporto minore di  $1/4$ , ma si tratta di una stima molto grossolana), tale segnale è verosimilmente un semplice "artefatto" dell'algoritmo. L'ultimo caso analizzato è un esempio di questa situazione.

Sfortunatamente, i rapporti segnale-rumore dei dati radioastronomici sono in genere molto bassi. Non solo, ma i test effettuati rivelano che se il segnale è stazionario, la trasformata di Fourier rimane senza dubbio lo strumento di rilevazione più sensibile.

# Capitolo 5

## Conclusioni

Questa tesi, che si inserisce nell'ambito del progetto SETI, è stata sviluppata con l'obiettivo di studiare un approccio all'analisi dei segnali alternativo rispetto a quello della trasformata di Fourier. Abbiamo visto come la combinazione dell'Empirical Mode Decomposition (EMD) e della trasformata di Hilbert si sia dimostrata un valido strumento nella descrizione di segnali nonstazionari. Tuttavia, come discusso nel terzo capitolo, questa nuova trasformata presenta ancora diverse limitazioni e problemi da risolvere, tra i quali ricordiamo la sua incapacità di distinguere segnali di frequenze simili.

Nel tentativo di sfruttare la HHT nel caso di segnali corrotti da alti livelli di rumore, si è introdotto il metodo chiamato Intermittency. I test effettuati hanno mostrato tuttavia che tale metodo è valido solo in caso di rapporti segnale-rumore abbastanza alti.

Naturalmente, la presente tesi non ha la pretesa di chiudere l'argomento. Ulteriori ricerche e test sono necessari per meglio valutare le capacità della trasformata di Hilbert-Huang. In quest'ottica si inserisce il lavoro di tesi in Ingegneria Elettronica di Leonardo Amico, "Trasformata di Hilbert-Huang per la rilevazione di segnali non stazionari nel progetto SETI".

In ogni caso, questa nuova generazione di trasformate (non solo la HHT, ma anche la KLT, le Wavelet, ecc.) apre una nuova via, che finora era sempre stata ignorata per la mancanza di strumenti matematici adeguati: quella della ricerca di segnali nonstazionari.



# Appendice A

## Generalized Zero Crossing

L'algoritmo Generalized Zero Crossing (GZC) è forse il metodo più diretto per calcolare la frequenza locale di un segnale concordemente con la definizione classica di frequenza. I vantaggi di questo metodo includono il fatto che non si serve della trasformata di Hilbert, evitando quindi tutti i problemi legati al suo utilizzo; è inoltre computazionalmente più semplice. Lo svantaggio è che la risoluzione dei valori della frequenza è limitata ad un quarto della lunghezza d'onda.

Questo metodo può essere usato con successo solo con segnali con le proprietà di una IMF, dato che questi segnali non sono “sporcati” da onde sovrainposte o altre distorsioni. L'algoritmo utilizza alcune caratteristiche del segnale, quali gli estremanti, gli attraversamenti dello zero e i periodi delle singole oscillazioni, per ricavare la frequenza direttamente nel dominio del tempo. Ciascuna di queste caratteristiche da sola può essere usata per calcolare un diverso valore di frequenza; il GZC le utilizza tutte insieme in una media pesata che ci permette di cogliere la frequenza locale del segnale.

Come noto, la definizione classica di frequenza è l'inverso della lunghezza del periodo:

$$\omega = \frac{1}{\Delta t}$$

La frequenza locale definita dal Generalized Zero Crossing è:

$$\bar{\omega} = \frac{1}{12} \left( \frac{1}{q} + \sum_{j=1}^2 \frac{1}{h_j} + \sum_{j=1}^4 \frac{1}{p_k} \right)$$

Questo valore è definito in base ai “punti critici” (ossia estremanti e attraversamenti dello zero) della funzione. Come mostrato chiaramente in figura A.1,  $q$  rappresenta la distanza fra i due punti critici più vicini al punto considerato (“un quarto” del periodo),  $h_k$  la distanza fra tre punti critici successivi

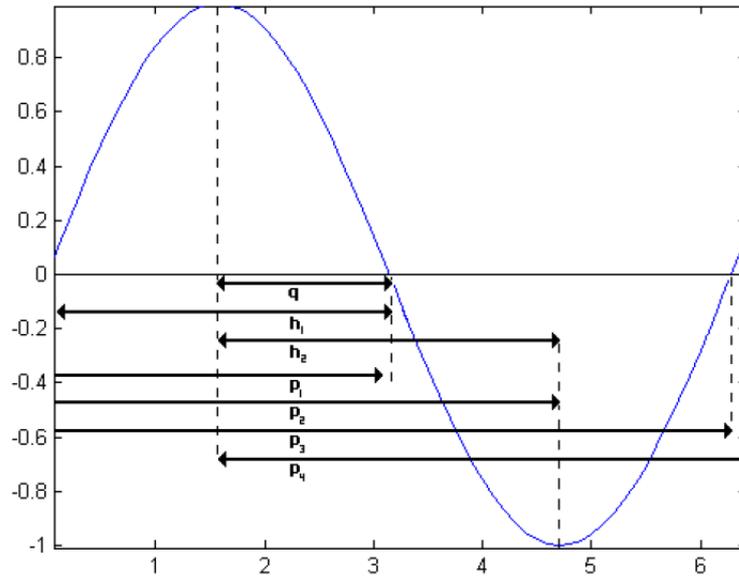


Figura A.1: Immagine di un sinusoide in cui è mostrata la relazione fra le variabili appena definite e il segnale

(“metà” del periodo), e  $p_k$  la distanza fra cinque punti critici successivi (un intero periodo).

I test da me effettuati hanno confermato che i risultati ottenuti su un segnale utilizzando l’algoritmo GZC sono sostanzialmente identici a quelli di una HHT in cui la derivata per ottenere la frequenza istantanea (equazione 3.3) è calcolata su un numero molto grande di punti (rendendolo quindi un valore ‘locale’ anzichè ‘puntuale’).

# Appendice B

## Algoritmo EMD

Come già detto, i test presentati in questa tesi sono stati effettuati utilizzando il software HHT-DPS, concesso dalla NASA per un periodo di valutazione di due mesi. Oltre a questo, è stato implementato attraverso MATLAB un algoritmo che esegue la prima parte della trasformata, vale a dire l'Empirical Mode Decomposition. Si tratta, comunque, di una versione abbastanza semplificata rispetto a quella esposta nel terzo capitolo (il criterio di arresto è diverso e non sono presenti i metodi numerici citati nel paragrafo 3.5). Di seguito è presentato il codice sviluppato.

### EMD.m

Questa è la funzione principale; essa prende in input il vettore  $x$  (il segnale) e lo scompone in una sequenza di IMF che vengono salvate per righe nella matrice  $A$ , e restituite come output. L'ultima riga della matrice conterrà il residuo. L'algoritmo è strutturato in modo da estrarre al massimo 50 IMF. Il processo si blocca anche se, dopo aver estratto una IMF, nel residuo non sono presenti almeno un punto di massimo e uno di minimo. Per trovare le IMF, viene utilizzata un'omonima funzione, descritta nel prossimo file.

```
function [A]= EMD(x)
    n=length(x);
    t=0;
    R=0;
    r=x;
    while t==0 && R<50
        nmax=0;
        nmin=0;
```

```

    for k=2:(n-1)
        if r(k-1)<r(k) && r(k)>r(k+1)
            nmax=nmax+1;
        elseif r(k-1)>r(k) \&\& r(k)<r(k+1)
            nmin=nmin+1;
        end
    end
    if nmax>0 && nmin>0
        A(R+1,:)=IMF(r);
        r=r-A(R+1,:);
    else
        t=1;
        A(R+1,1:n)=r;
    end
    R=R+1;
end
end
end

```

## IMF.m

Questa funzione, il cui output è il vettore `imf`, estrae una singola IMF dal segnale in input `x`. Per fare ciò, viene contato il numero di estremanti e di attraversamenti dello zero di `x`. Se questi due valori sono uguali o differiscono di 1, `x` è considerata già una IMF e viene restituita in output. In caso contrario, ad `x` viene applicata la funzione `ENVELOPES` (vedi sotto), e questo procedimento viene reiterato finchè la condizione sopra non viene soddisfatta. Si noti che il criterio di arresto è più semplice di quello espresso nel capitolo 3; tale criterio ci assicura che l'output soddisfa il primo punto della definizione di IMF, ma non il secondo.

```

function [imf]= IMF(x)
    n=length(x);
    k=0;
    imf=x;
    while(k==0)
        extrema=0;
        zero=0;
        for r=1:(n-1)
            if imf(r)<0 && imf(r+1)>=0
                zero=zero+1;
            end
        end
        k=zero-extrema;
    end
end

```

```

        elseif imf(r)>0 && imf(r+1)<=0
            zero=zero+1;
        end
    end
end
for r=2:(n-1)
    if imf(r-1)<imf(r) && imf(r)>imf(r+1)
        extrema=extrema+1;
    elseif imf(r-1)>imf(r) && imf(r)<imf(r+1)
        extrema=extrema+1;
    end
end
if extrema<zero-1 || extrema>zero+1
    imf=ENVELOPES(imf);
else
    k=1;
end
end
end
end

```

## ENVELOPES.m

Questa funzione costruisce le spline cubiche passanti per i punti di massimo e per i punti di minimo del vettore in input  $x$  (se i massimi o i minimi sono costituiti da un solo punto la funzione costruisce una retta piuttosto che una spline), ne fa la media e la sottrae ai dati. Il risultato (il vettore  $imf$ ) è l'output.

```

function [imf] = ENVELOPES(x)
    n=length(x);
    nmax=0;
    nmin=0;
    for k=2:(n-1)
        if x(k-1)<x(k) && x(k)>x(k+1)
            nmax=nmax+1;
            vmax(nmax)=x(k);
            pmax(nmax)=k;
        elseif x(k-1)>x(k) && x(k)<x(k+1)
            nmin=nmin+1;
            vmin(nmin)=x(k);
            pmin(nmin)=k;
        end
    end
end

```

```
        end
    end
    b=1:n
    if nmax==1
        splinemax(1:n)=vmax(1);
    else
        splinemax=spline(pmax,vmax,b);
    end
    if nmin==1
        splinemin(1:n)=vmin(1);
    else
        splinemin=spline(pmin,vmin,b);
    end
    media=(splinemax+splinemin)/2;
    imf=x-media;
end
```

# Bibliografia

- [1] Giorgio Leone: “Algoritmi per il riconoscimento di segnali per lo spettrometro SERENDIP IV nell’ambito del progetto SETI”, 2002
- [2] Giuseppe Cocconi, Philip Morrison: “Searching for interstellar communications”, Nature, vol 184, 1959
- [3] <http://www.seti-inst.edu/ata/>
- [4] <http://seti.ssl.berkeley.edu/serendip/>
- [5] L. Calandrino, G. Immovilli: “Schemi dalle lezioni di Comunicazioni elettriche”, 1991
- [6] L. R. Rabiner, B. Gold: “Theory and Application of Digital Signal Processing”, 1975
- [7] Pierpaolo Pari: “Rilevazione di segnali tramite trasformata KLT nel progetto SETI”, 2003
- [8] Charles K. Chui: “An Introduction to Wavelets”, 1992
- [9] C. Maccone: “Communications, KLT and relativity”, vol 1, 1995
- [10] <http://keck.ucsf.edu/~schenk/Huang-etal98.pdf>
- [11] Norden E. Huang, Samuel S. P. Shen: “Hilbert-Huang Transform and its applications”, 2005
- [12] S. L. Hahn: “Hilbert Transforms in signal processing”, 1996