

# Compressed Sensing rumoroso

*Federico Carlesso*

Matricola: 597053

*Relatore: Professor Michele Pavon*

Università degli Studi di Padova

Facoltà di Ingegneria

Laurea Triennale in Ingegneria dell'Informazione

23 settembre 2011



# Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
<b>2</b>	<b>Segnali, serie di Fourier</b>	<b>7</b>
2.1	Segnale a tempo discreto . . . . .	7
2.2	Segnale a tempo discreto periodico . . . . .	8
2.2.1	Gli esponenziali periodici . . . . .	9
2.3	La formula di Eulero . . . . .	9
2.4	La serie di Fourier . . . . .	10
2.4.1	La serie di Fourier a tempo continuo . . . . .	10
2.4.2	Ortogonalità tra esponenziali . . . . .	10
2.4.3	La serie di Fourier a tempo discreto . . . . .	11
2.4.4	Alcune proprietà della serie discreta di Fourier . . . . .	12
<b>3</b>	<b>Richiami di algebra lineare</b>	<b>13</b>
3.1	Vincoli lineari . . . . .	13
<b>4</b>	<b>Variabili aleatorie, stima parametrica</b>	<b>17</b>
4.1	Teoria del rumore . . . . .	17
4.1.1	Processo aleatorio stazionario a tempo discreto . . . . .	17
4.1.2	Il rumore Gaussiano bianco . . . . .	19
4.2	Stima parametrica . . . . .	19
4.2.1	Il metodo dei minimi quadrati . . . . .	20
4.2.2	Il LASSO . . . . .	21
<b>5</b>	<b>Compressed Sensing senza rumore</b>	<b>23</b>
5.1	Numero di valori discreti indispensabili . . . . .	23
5.2	Unicità della soluzione . . . . .	25
5.2.1	Limitazioni del teorema: N numero primo . . . . .	26
5.3	Il parametro di incoerenza . . . . .	28
<b>6</b>	<b>Compressed Sensing rumoroso</b>	<b>29</b>
6.1	Coefficienti rumorosi: segnale sparso . . . . .	29
6.2	Ricostruzione di un segnale . . . . .	30

<b>7</b>	<b>Un'applicazione del CS rumoroso</b>	<b>33</b>
7.1	Il PAT: Photoacoustic computed tomography . . . . .	33
7.1.1	Misurazione sperimentale . . . . .	34
<b>8</b>	<b>Conclusioni</b>	<b>39</b>

# Capitolo 1

## Introduzione

Negli ultimi anni si è assistito a una crescita vertiginosa della quantità di dati da memorizzare. Ciò ha risolleavato una domanda: esiste un limite nel rapporto dati memorizzati/spazio dedicato? Esiste un limite di comprimibilità?

Shannon aveva già risposto in precedenza a questa domanda, sostenendo e dimostrando che tale limite esiste. Egli, assieme a Nyquist, sviluppò il *Teorema del Campionamento*, il quale stabilisce che per memorizzare un segnale reale e per poterlo recuperare completamente è necessario campionarlo con una frequenza sufficientemente elevata. Di conseguenza il minimo numero di campioni per recuperare un segnale è legato a questa frequenza. E' tuttavia possibile fare meglio?

Nell'ultimo decennio sono state sviluppate nuove teorie riguardo alla possibilità di poter recuperare integralmente un segnale partendo da un numero inferiore di campioni: è per questo proposito che nasce la teoria del *Compressed Sensing*. Essa prevede, imponendo opportune condizioni, di poter riottenere il segnale originario completamente.

Se dal punto di vista teorico il Compressed Sensing non rumoroso è rivoluzionario, ancor più stupefacente lo è quello riguardante campioni corrotti da rumore, in quanto tale modello ben si adatta al mondo reale, costantemente compromesso dal rumore.

Il presente elaborato si propone di presentare il problema che ha portato allo sviluppo del Compressed Sensing, analizzarne le caratteristiche e riassumere infine una sperimentazione pratica su uno strumento diagnostico. I primi capitoli della tesina riguardano richiami al concetto di segnale, di rumore e del problema del CS dal punto di vista dell'algebra lineare, per passare poi nel quarto capitolo alla presentazione delle caratteristiche del CS senza rumore e nel quinto al problema del CS rumoroso. L'ultimo capitolo è dedicato alla presentazione di una tra le varie applicazioni del CS.



## Capitolo 2

# Segnali, serie di Fourier

In questo capitolo verranno richiamati i concetti di segnale a tempo discreto, segnale a tempo discreto periodico e la rispettiva serie di Fourier.

### 2.1 Segnale a tempo discreto

Un segnale è una grandezza che può variare nel tempo. La sua rappresentazione matematica è una funzione complessa che, nel caso di un segnale a tempo discreto, ha come dominio un sottoinsieme dei numeri interi:

$$\begin{aligned}x &: [n_1, n_2] \subseteq \mathbb{Z} \rightarrow \mathbb{C}, \\n &\mapsto x(n), \\n_1 &\leq n_2.\end{aligned}$$

Si noti che il dominio di numeri interi può anche essere inteso come ottenuto per campionamento da un segnale a tempo continuo:

$$\begin{aligned}x &: [n_1T, n_2T] \rightarrow \mathbb{C}, \\T &\geq 0, n_1, n_2 \in \mathbb{Z}, n_1 \leq n_2,\end{aligned}$$

dove  $T$  è la spaziatura tra gli istanti di campionamento del segnale continuo

$$\begin{aligned}y &: [t_1, t_2] \subseteq \mathbb{R} \rightarrow \mathbb{C}, \\t &\mapsto y(t), \\t_1 &\leq t_2.\end{aligned}$$

e prende il nome di *periodo di campionamento*. Tali segnali hanno come codominio l'insieme dei numeri complessi; l'insieme può essere eventualmente ristretto ai segnali reali, aventi codominio reale, considerando la relazione

$$\mathbb{R} \subset \mathbb{C}.$$

Un segnale a tempo discreto può essere rappresentato mediante:

- una espressione analitica
- un grafico; esso può essere tridimensionale
  - Asse  $x$ : Asse dei tempi
  - Asse  $y$ : Asse della parte reale del segnale

$$\Re[x(n)]$$

- Asse  $z$ : Asse della parte immaginaria

$$\Im[x(n)]$$

oppure essere costituito di due grafici, uno rappresentante la parte reale e l'altro rappresentante la parte immaginaria

- una tabella

Si definisce *supporto* di un segnale a tempo discreto il più piccolo intervallo  $[n_1, n_2]$  tale che

$$x(n) = 0, \forall n \in ] - \infty, n_1] \cup [n_2, +\infty[$$

## 2.2 Segnale a tempo discreto periodico

Una distinzione importante fra segnali è quella fra segnali *periodici* e *aperiodici*. Un segnale a tempo discreto si dice *periodico* se esiste un numero intero  $N > 0$  tale che

$$x(n + N) = x(n), N \in \mathbb{Z}_+ \quad (2.1)$$

Tale valore  $N$  si dice inoltre *periodo fondamentale* se è il più piccolo per cui vale (2.1). Si noti che è possibile rendere un segnale periodico se questo presenta supporto finito: si dice ripetizione periodica di periodo  $T$  di  $x(n)$  il segnale

$$\text{rep}_T x(n) = \sum_{k=-\infty}^{k=+\infty} x(n - kT) \quad (2.2)$$

In tal modo si rende periodico un segnale con supporto limitato. La periodicità di segnali a tempo discreto è una proprietà che vale anche nel caso di somma e/o prodotto di segnali periodici. Vale infatti la seguente osservazione:

Somma e prodotto di segnali periodici a tempo discreto è sempre periodica.

### 2.2.1 Gli esponenziali periodici

Una importante famiglia di segnali periodici è quella degli esponenziali in *relazione armonica*.

Si dice segnale esponenziale a tempo discreto il segnale

$$\varphi(n) = e^{j\vartheta_0 n}$$

Si noti che esso è periodico solamente nel caso in cui

$$e^{j\vartheta_0 N} = e^{j\vartheta_0(n+N)}, \forall n$$

e quindi

$$e^{j\vartheta_0 N} = 1 \Rightarrow \vartheta_0 N = 2|k|\pi, k \in \mathbb{Z} \Rightarrow \frac{\vartheta_0}{2\pi} \in \mathbb{Q}$$

Di conseguenza si dirà *famiglia di esponenziali in relazione armonica* la famiglia di segnali esponenziali

$$\varphi_k(n) = e^{jk(\frac{2\pi}{N})n}, k \in \mathbb{Z}.$$

Tale famiglia presenta solamente N segnali distinti, infatti

$$e^{jN(\frac{2\pi}{N})n} = e^{j0(\frac{2\pi}{N})n} = 1$$

$$e^{j(N+1)(\frac{2\pi}{N})n} = e^{j(\frac{2\pi}{N})n}$$

$$e^{j(N+l)(\frac{2\pi}{N})n} = e^{jl(\frac{2\pi}{N})n}$$

con  $l \in \mathbb{Z}$ .

## 2.3 Richiami sulla rappresentazione dei numeri complessi: la formula di Eulero

I numeri complessi possono essere espressi in vari modi:

- forma algebrica

$$(a + jb), a, b \in \mathbb{R}$$

- forma trigonometrica

$$z = \rho(\cos \vartheta + j \sin \vartheta)$$

in cui  $\rho$  è il modulo del numero complesso e  $\vartheta$  è un argomento di z.

Dalla formula di Eulero:

$$e^{jn\vartheta} = \cos(n\vartheta) + j \sin(n\vartheta)$$

risulta evidente che gli esponenziali complessi rappresentano numeri complessi aventi modulo unitario e fase pari a  $n\vartheta$ . Essi perciò nel piano di Gauss appartengono alla circonferenza di centro l'origine e raggio unitario.

## 2.4 La serie di Fourier

La serie di Fourier è uno strumento importante che trova numerose applicazioni in ambito ingegneristico. Essa permette di rappresentare qualsiasi segnale periodico come serie di esponenziali complessi; inoltre, in caso di segnale reale, esso potrà essere espresso come serie di sinusoidi.

### 2.4.1 La serie di Fourier a tempo continuo

I segnali a tempo continuo periodici hanno una particolarità: essi possono essere rappresentati in maniera biunivoca da certi coefficienti  $a_k \in \mathbb{C}$ .

Si definisce *serie di Fourier a tempo continuo* del segnale  $x(t)$  l'espressione

$$x(t) = \sum_{k=-\infty}^{+\infty} a_k \varphi_k(t) \quad (2.3)$$

dove  $\varphi_k(t) = e^{jk\omega_0 t}$  è l'esponenziale complesso a tempo continuo. Si definisce coefficiente di Fourier a tempo continuo il valore

$$a_k = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-jk\omega_0 t} dt \quad (2.4)$$

Il teorema di Riesz-Fisher ci garantisce la convergenza fra il segnale e la sua rappresentazione in serie di Fourier:

**Teorema 1** Sia  $\{a_n\}_{n=-\infty}^{n=+\infty}$  una successione di numeri complessi tali che  $\sum_{n=-\infty}^{+\infty} |a_n|^2 < \infty$  allora esiste un segnale  $x(t)$  definito su  $[0, T_0]$  con  $\int_0^{T_0} |x(t)|^2 dt < \infty$  tale che vale  $a_k = \frac{1}{T_0} \int_0^{T_0} x(t) e^{-jk\omega_0 t} dt$  con  $\omega_0 = \frac{2\pi}{T_0}$

Dato  $x(t)$  segnale ad energia finita su  $[0, T_0]$  i suoi coefficienti di Fourier sono ben definiti da (2.4) e soddisfano la proprietà  $\sum_{n=-\infty}^{+\infty} |a_n|^2 \leq \infty$ ; inoltre definita  $s_n(t) = \sum_{k=-\infty}^{+\infty} a_k \varphi_k(t)$  vale  $\lim \int_0^T |s_n(t) - x(t)|^2 dt = 0$  ovvero  $s_n(t)$  converge in media quadratica a  $x(t)$ .

Il teorema quindi non ci garantisce che il passaggio da coefficienti di Fourier a segnale consenta di riottenere il segnale con convergenza puntuale, ma bensì la convergenza è nel senso della media quadratica.

### 2.4.2 Ortogonalità tra esponenziali

Rappresentare un segnale in serie di Fourier significa scrivere un segnale come combinazione lineare di un numero infinito di esponenziali complessi. È quindi possibile dare un'interpretazione geometrica della serie di Fourier, osservando come il calcolo di un coefficiente  $a_k$  corrisponde a calcolare la proiezione ortogonale del segnale  $x(n)$  su un elemento della base costituita

dagli esponenziali complessi. Tale base è di cardinalità infinita: gli esponenziali sono in numero infinito al variare di  $k$ . Si introduce perciò nello spazio vettoriale degli esponenziali complessi (generato dalla base  $[v_1, v_2, \dots]$ ) un prodotto scalare, così definito:

$$\langle \varphi_h(t), \varphi_k(t) \rangle = \int_0^{T_0} \varphi_h(t) \overline{\varphi_k(t)} dt$$

dove  $\overline{\varphi_k(t)}$  è il segnale complesso coniugato di  $\varphi_k(t)$ , da cui si osservi che

- se  $h = k \Rightarrow \int_0^{T_0} e^{j\omega(h-k)t} dt = T_0$
- se  $h \neq k \Rightarrow \int_0^{T_0} e^{j\omega(h-k)t} dt = \left[ \frac{1}{j\omega(h-k)} e^{j\omega(h-k)t} \right]_0^{T_0} = \varphi_{h-k}(t) - \varphi_{h-k}(t) = 0$

Ovvero la base di esponenziali si può considerare come base ortogonale. Da ciò segue la formula per il calcolo del coefficiente  $a_k$ , come dal teorema della proiezione ortogonale, ricordando che gli esponenziali devono essere normalizzati per ottenere una base ortonormale:

$$a_k = \langle x(t), \varphi_k(t) \rangle = \frac{1}{T_0} \int_0^{T_0} x(t) \overline{\varphi_k(t)} dt$$

### 2.4.3 La serie di Fourier a tempo discreto

Qualsiasi segnale a tempo discreto periodico può essere rappresentato come sommatoria di esponenziali complessi.

Dato un segnale  $x(n)$  periodico di periodo fondamentale  $N$ , si definisce *serie di Fourier a tempo discreto* l'espressione

$$x(n) = \sum_{k=\langle N \rangle} a_k \varphi_k(n) \quad (2.5)$$

dove  $\varphi_k(n) = e^{jk\frac{2\pi}{N}n}$  e dove  $k = \langle N \rangle$  indica  $N \in \mathbb{Z}$  numeri consecutivi.

In analogia con la serie di Fourier a tempo continuo, si può dimostrare che

$$a_k = \frac{1}{N} \sum_{n=\langle N \rangle} x(n) e^{-jK\vartheta_0 n} \quad (2.6)$$

dove  $\vartheta_0 = \frac{2\pi}{N}$ .

A tempo continuo la validità di tale espressione era garantita dall'ortogonalità tra i  $\varphi_k(t)$ . In questo caso la relazione di ortogonalità è la seguente:

$$\sum_{n=\langle N \rangle} e^{jk(\frac{2\pi}{N})n} = \begin{cases} N & \text{se } k = 0, N, -N, 2N, -2N \dots \\ 0 & \text{altrimenti} \end{cases}$$

dove tale espressione è vera perchè

$$\sum_{n=0}^{N-1} e^{jk(2\pi/N)n} = \frac{1 - e^{jk(\frac{2\pi}{N})N}}{1 - e^{jk(\frac{2\pi}{N})}} = 0$$

Si noti inoltre che se  $x(n)$  è periodico di periodo  $N$  allora  $x(n) - x(n - N) = 0$ . Tale espressione è sempre vera. Essa può essere anche vista come un'equazione alle differenze lineari con coefficienti costanti. In quanto tale essa ha come soluzione generale proprio la famiglia di esponenziali complessi in relazione armonica sopra citata.

#### 2.4.4 Alcune proprietà della serie discreta di Fourier

Considerando un segnale periodico di periodo  $N$  si può notare che la relazione fra valori  $x(n)$  e relativi  $a_k$  può anche essere vista come una funzione, una mappa

$$f : \mathbb{C}^N \rightarrow \mathbb{C}^N$$

$$\begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ \cdot \\ x(i) \\ \cdot \\ x(N-1) \end{pmatrix} \mapsto \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \cdot \\ a_i \\ \cdot \\ a_{N-1} \end{pmatrix}$$

che associa quindi ad un vettore di valori complessi un altro vettore di valori complessi. Tale mappa gode delle seguenti proprietà:

- è lineare;
- è isometrica.

Si noti che grazie all'isometria di cui gode la funzione, si può ancora ricorrere al teorema di Parseval per il calcolo dell'energia

$$\frac{1}{N} \sum_{n=\langle N \rangle} |x(n)|^2 = \sum_{k=\langle N \rangle} |a_k|^2 \quad (2.7)$$

Tale espressione ci consente di calcolare l'energia del segnale, operazione che può essere eseguita sia nel dominio del tempo che nel dominio della frequenza.

## Capitolo 3

# Richiami di algebra lineare

Di importanza chiave è l'algebra lineare nello studio del passaggio tra segnale periodico a tempo discreto e coefficienti della serie di Fourier a tempo discreto; le relazioni espresse tramite le sommatorie (2.5) e (2.6) sono infatti relazioni di tipo lineare.

Il Compressed Sensing prevede la possibilità di recupero del segnale partendo da un numero inferiore a  $N$  dei coefficienti della serie di Fourier discreta; si esporrà qui tale problema dal punto di vista dell'algebra lineare, ricorrendo ai due importanti teoremi di Rouchè -Capelli, osservando come sia necessario ricorrere ad altri strumenti per dimostrare questa opportunità.

### 3.1 Vincoli lineari

Un segnale discreto periodico può essere espresso come combinazione lineare di coefficienti  $a_k$ :

$$\begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(i) \\ \vdots \\ x(N-1) \end{pmatrix} = \begin{pmatrix} 1 & 1 & \dots & 1 & \dots & 1 \\ 1 & e^{j\frac{2\pi}{N}} & \dots & e^{ji\frac{2\pi}{N}} & \dots & e^{j(N-1)\frac{2\pi}{N}} \\ 1 & e^{j\frac{2\pi}{N}2} & \dots & e^{ji\frac{2\pi}{N}2} & \dots & e^{j(N-1)\frac{2\pi}{N}2} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ 1 & e^{j\frac{2\pi}{N}i} & \dots & e^{ji\frac{2\pi}{N}i} & \dots & e^{j(N-1)\frac{2\pi}{N}i} \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ \vdots & \vdots & \dots & \vdots & \dots & \vdots \\ 1 & e^{j\frac{2\pi}{N}(N-1)} & \dots & e^{ji\frac{2\pi}{N}(N-1)} & \dots & e^{j(N-1)\frac{2\pi}{N}(N-1)} \end{pmatrix} \times \begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_{N-1} \end{pmatrix}$$

Il sistema può anche essere espresso nella notazione simbolica

$$X = AY$$

dove  $X$  è il vettore dei segnale  $x(n)$ ,  $A$  è la matrice di esponenziali complessi (detta di *Fourier*) e  $Y$  è il vettore dei coefficienti  $a_k$ . Si osservi che questo è un

sistema di equazioni lineari, avente  $n$  equazioni e  $n$  incognite. Individuare un segnale a tempo discreto partendo dai relativi coefficienti di Fourier equivale quindi a risolvere un sistema lineare di equazioni.

Si richiamano di conseguenza i due importanti teoremi di Rouchè-Capelli sulla risolubilità dei sistemi lineari.

**Teorema di Rouchè-Capelli, 1 parte 1** *Il sistema  $AY=X$  ha soluzioni se e solo se il rango della matrice completa e quello della matrice incompleta coincidono.*

Si noti che in questo caso la matrice incompleta è la matrice  $A$ ; quella completa è costituita dalla matrice  $C$  le cui prime  $n$  colonne coincidono con le  $n$  colonne di  $A$  e la  $(n+1)$ -esima è la colonna  $Y$ .

**Teorema di Rouchè-Capelli, 1 parte 2** *Sia  $AY=X$  un sistema compatibile di  $m$  equazioni in  $n$  incognite.*

1. *Se il rango di  $A$  è uguale al rango di  $C$  ed è pari a  $n$  il sistema ammette una e una sola soluzione*
2. *Se il rango di  $A$  è uguale al rango di  $C$  ed è pari a  $r$ , con  $r < n$  allora le soluzioni dipendono da  $n-r$  parametri arbitrari*

Nel caso di ricostruzione del segnale partendo dai coefficienti di Furier si suppone che  $X$  sia il vettore incognito e che  $Y$  siano i coefficienti noti. Di conseguenza si considererà l'espressione

$$Y = BX$$

$$\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_i \\ \vdots \\ a_{N-1} \end{pmatrix} = B \times \begin{pmatrix} x(0) \\ x(1) \\ x(2) \\ \vdots \\ x(i) \\ \vdots \\ x(N-1) \end{pmatrix}$$

con

$$B = \begin{pmatrix} \frac{1}{N} & \frac{1}{N} & \cdots & \frac{1}{N} & \cdots & \frac{1}{N} \\ \frac{1}{N} & \frac{1}{N}e^{-j\frac{2\pi}{N}} & \cdots & \frac{1}{N}e^{-j\frac{2\pi}{N}i} & \cdots & \frac{1}{N}e^{-j\frac{2\pi}{N}(N-1)} \\ \frac{1}{N} & \frac{1}{N}e^{-j2\frac{2\pi}{N}} & \cdots & \frac{1}{N}e^{-j2\frac{2\pi}{N}i} & \cdots & \frac{1}{N}e^{-j2\frac{2\pi}{N}(N-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{1}{N} & \frac{1}{N}e^{-ji\frac{2\pi}{N}} & \cdots & \frac{1}{N}e^{-ji\frac{2\pi}{N}i} & \cdots & \frac{1}{N}e^{ji\frac{2\pi}{N}(N-1)} \\ \vdots & \vdots & \ddots & \vdots & \ddots & \vdots \\ \frac{1}{N} & \frac{1}{N}e^{-j(N-1)\frac{2\pi}{N}} & \cdots & \frac{1}{N}e^{-j(N-1)\frac{2\pi}{N}i} & \cdots & \frac{1}{N}e^{-j(N-1)\frac{2\pi}{N}(N-1)} \end{pmatrix}$$

e  $B = A^{-1}$ . In tal caso, si osserva che la matrice incompleta  $B$  avrà rango pari a quello della matrice completa, in quanto la matrice  $B$  è una matrice quadrata le cui righe sono vettori linearmente indipendenti (grazie alla proprietà della ortogonalità). Di conseguenza, se si hanno tutti i coefficienti  $a_k$  e quindi il rango della matrice è pari a  $n$ , si ha che il primo teorema di Rouchè-Capelli è rispettato e il sistema è risolubile. Il secondo teorema di Rouchè-Capelli assicura inoltre che il sistema ha una e una sola soluzione, grazie al rango  $n$  della matrice completa ed incompleta.

Il problema del Compressed Sensing sorge dalla necessità di ricostruire il segnale non avendo tutti i coefficienti  $a_k$ . In tal caso, supponendo di avere a disposizione  $l$  coefficienti  $a_k$ , si ottiene un sistema lineare di  $l$  equazioni in  $n$  incognite. Dal teorema di Rouchè-Capelli, dato che la matrice completa e incompleta hanno lo stesso rango  $l$ , segue che il sistema è risolubile. Tuttavia, sempre per il medesimo teorema, la soluzione non è unica, in quanto si ha che le soluzioni dipendono da  $n - l$  parametri arbitrari. In tal caso, si usa dire che il sistema ha come soluzione lo spazio somma, ovvero la somma di uno o più vettori e di un sottospazio vettoriale di dimensione  $n - l$  o che il sistema ha  $\infty^{n-l}$  soluzioni. Per ricavare  $x(n)$  bisognerà quindi aggiungere ulteriori condizioni a quelle lineari presenti.



## Capitolo 4

# Variabili aleatorie, stima parametrica

Importante nella trattazione di Compressed Sensing avente campioni rumorosi è la teoria probabilistica del rumore e alcune considerazioni di stima statistica parametrica. Verrà presentata la tecnica di stima parametrica con il metodo dei minimi quadrati; in seguito verranno dati dei cenni sulla stima con il metodo *LASSO*.

### 4.1 Teoria del rumore

Qualsiasi segnale fisico, una volta elaborato da dispositivi e/o trasmesso attraverso mezzo fisico, è soggetto a disturbi. Il risultato di una trasmissione, di una elaborazione è un segnale differente dall'originale, che presenta una componente di diversità che prende il nome di rumore. Il rumore viene modellizzato attraverso la teoria probabilistica e può essere diverso a seconda della tipologia che si va a considerare. È quindi opportuno scegliere un modello adatto al caso specifico che si va ad analizzare.

#### 4.1.1 Processo aleatorio stazionario a tempo discreto

Un processo aleatorio a tempo discreto è una funzione di due variabili tale che

$$\begin{aligned}x(n, \omega) &: \mathbb{Z} \times \Omega \rightarrow \mathbb{R} \\(n, \omega) &\mapsto x\end{aligned}$$

con  $\Omega$  spazio campionario. Di esso è caratteristico il valore medio, che è una funzione del tempo

$$\begin{aligned}m_x &: \mathbb{Z} \rightarrow \mathbb{C} \\m_x(n) &= E[x(n)]\end{aligned}$$

così come la potenza statistica

$$M_x : \mathbb{Z} \rightarrow \mathbb{R}$$

$$M_x(n) = E[|x(n)|^2]$$

così come la funzione di autocorrelazione:

essa è una funzione di due variabili,  $n$ , istante temporale, e  $\rho$ , tempo di ritardo.

$$r_x : \mathbb{Z} \times \mathbb{Z} \rightarrow \mathbb{C}$$

$$r_x(n, \rho) = E[x(n)\overline{x(n-\rho)}]$$

Un processo aleatorio  $x(n)$  si dice stazionario rispetto ad una descrizione statistica (es. media, autocorrelazione o PDF) se tale descrizione è invariante a qualsiasi traslazione del segnale nel tempo (ovvero  $x(n) = x(n - n_0)$ ).

In particolare, si può facilmente vedere che per un processo aleatorio stazionario la sua media  $m_x(n)$  è un valore costante, dato che:

$$m_x(n) = m_{x_1}(n) = E[x_1(n)] = E[x(n - n_0)] = m_x(n - n_0), n, n_0 \in \mathbb{Z}$$

cosicché si può scrivere semplicemente:  $m_x$  come media statistica,  $M_x$  come potenza statistica,  $\sigma_x^2$  come varianza.

Nello stesso modo, per un processo aleatorio stazionario la sua autocorrelazione  $r_x(n, \rho)$  è solo funzione della grandezza  $\rho$  e indipendente da  $n$ , quindi si può scrivere semplicemente solo  $r_x(\rho)$ . Un processo che è stazionario sia nella sua media che nella sua autocorrelazione si dice stazionario in senso lato (WSS).

Un processo aleatorio è descritto *al secondo ordine* da media e varianza; tuttavia un'ulteriore grandezza permette di descrivere la tipologia di un segnale: la densità spettrale di potenza, anche detta *Power spectral density (PSD)*. Tale grandezza rappresenta una descrizione del *secondo ordine* del processo aleatorio.

La densità spettrale di potenza (PSD) di un processo aleatorio WSS è la trasformata di Fourier della sua funzione di autocorrelazione

$$P_x(e^{j\vartheta}) = T \sum_{n=-\infty}^{+\infty} r_x(n) e^{-j\vartheta n}$$

dove la PSD gode delle seguenti proprietà:

- è a valori reali, anche per processi aleatori a valori complessi
- è non negativa
- il suo integrale porta alla potenza statistica
- è una funzione pari

Il potere di densità spettrale è un valore molto importante: esso ci dà informazioni sulla possibilità che valori precedenti del segnale influenzino i suoi valori futuri, esplicitandone la relazione in funzione della frequenza del segnale che si considera.

### 4.1.2 Il rumore Gaussiano bianco

Il rumore molto spesso viene modellizzato tramite un modello Gaussiano bianco, additivo al segnale (trascurando effetti di degrado e interazioni di Cross Talk) : con esso si intende un rumore che ha media nulla e il cui valore varia secondo una variabile aleatoria Gaussiana, un rumore in cui in ogni istante il suo valore è scorrelato dai suoi valori precedenti e futuri. Formalmente, un processo aleatorio WSS  $x(n)$  il cui potere di densità spettrale è costante attraverso l'intero asse delle frequenze si dice rumore bianco. Si osservi che per avere PSD costante un processo aleatorio bianco deve avere una autocorrelazione impulsiva

$$r_x(n) = P_0\delta(n)$$

dove la delta  $\delta$  è la delta di Kronecher nel caso discreto, delta distribuzione nel caso continuo.

In particolare un processo aleatorio Gaussiano bianco a media nulla si dice *rumore Gaussiano bianco (WGN)*.

## 4.2 Stima parametrica

Si supponga di avere un modello (può essere di vario tipo: economico, biologico...) che presenta dei parametri, dei valori ignoti. Avendo a disposizione dei dati (ottenuti da una misurazione per esempio) è possibile cercare di “indovinare tali parametri.

La stima parametrica è una tecnica di stima di valori ignoti che permette di ottenere un modello matematico (comprensivo dei parametri) che al meglio segue l'andamento dei dati ed effettua una previsione di essi. Tale tecnica non è unica; ne esistono molte versioni e varianti, fra cui la più nota è il metodo dei minimi quadrati. Si accennerà in seguito il metodo *LASSO* in quanto verrà utilizzato per il Compressed Sensing rumoroso, considerando come “parametro da determinare proprio il segnale  $x(n)$  (il suo vettore di valori).

Il procedimento di stima parametrica viene usato nel momento in cui si va a determinare il segnale  $x(n)$  sapendo che i coefficienti  $a_k$  sono corrotti da rumore. Si può infatti considerare la seguente espressione come “scomposizione dei coefficienti rumorosi:

$$Y = BX + Z$$

dove  $Y$  è il vettore dei coefficienti,  $B$  la matrice di esponenziali complessi,  $X$  il vettore dei *parametri* (in questo caso sono i valori  $x(n)$ ),  $Z$  è l'errore introdotto, supposto additivo al segnale. Utilizzando tale espressione, tramite le metodologie di stima, conoscendo tutti i valori  $a_k$  è possibile ottenere i valori  $x(n)$  che minimizzano l'errore del sistema.

Una commistione di tecniche di stima parametrica, vincoli sulla norma  $\ell_1$  e condizioni sul segnale come *segnale sparso* porteranno a risolvere il problema del Compressed Sensing con rumore.

### 4.2.1 Il metodo dei minimi quadrati

Si supponga di avere tutti i coefficienti  $a_k$ , considerati come costituenti del vettore  $Y$  e si supponga che tali coefficienti siano interessati da una componente rumorosa, intesa come additiva. Allora si può considerare l'espressione:

$$Y = BX + Z$$

dove  $Z$  è il vettore le cui componenti sono il valore del rumore introdotto. Tali componenti  $z_k$  sono assimilabili a una variabile aleatoria. Molto spesso tale variabile aleatoria (che rappresenta il rumore) è considerata come un processo casuale (spesso gaussiano) a media nulla e a campioni indipendenti, ovvero è un rumore *Gaussiano bianco*.

$$E[z_k] = 0$$

$$E[z_k, z_l] = 0, k \neq l$$

$$var[z_k] = \sigma_k^2$$

La varianza del processo può essere costante ( $\forall k$  cioè  $\sigma_k^2 = \sigma^2$ ) oppure non costante ( $\sigma_k^2$  dipendente da  $k$ ).

Si osservi che  $BX$  è il valore esatto del vettore degli  $a_k$ . Il problema di stima parametrica consiste nel determinare una stima ( $\hat{X}$ ) di  $X$  vettore del segnale noti i valori  $a_k$  contenuti in  $Y$  e la funzione  $BX$ . In questo caso, il sistema è lineare nei parametri; in quanto tale la procedura di stima ai minimi quadrati è semplificata.

Il vettore  $Z$ , vettore dei valori del rumore, presenta le seguenti caratteristiche:

$$E[Z] = 0$$

$$E[Z, Z^T] = \Sigma_Z$$

dove  $\Sigma_Z$  è la matrice di covarianza.

Si chiameranno *residui* il vettore di valori così definito:

$$R = Y - BX$$

dove tale vettore  $R$  è quindi una realizzazione della variabile aleatoria  $Z$ . Per risolvere il problema di stima parametrica bisognerà fare in modo che il vettore  $X$  minimizzi questi residui. A tale scopo si consideri una funzione costo  $j(X)$ :

$$j(X) = \sum_{i=1}^N r_i^2 = (Y - BX)^T(Y - BX)$$

dove  $r_i$  sono le componenti del vettore  $R$ . Per cercare il valore minimo di tale funzione *costo*, è sufficiente derivare rispetto al valore  $X$ .

$$\begin{aligned} j(X) &= (Y - BX)^T(Y - BX) \\ j(X) &= Y^T Y - 2Y^T B X + X^T B^T B X \end{aligned}$$

ha lo stesso minimo di

$$j(X) = \frac{1}{2} X^T B^T B X - Y^T B X$$

Derivando

$$\frac{\partial j(X)}{\partial X} = (B^T B)X - B^T Y = 0$$

si ottiene

$$\hat{X} = (B^T B)^{-1} B^T Y \quad (4.1)$$

che è il valore del vettore/segnale la cui trasformata al meglio si avvicina al vettore  $Y$ . Si osservi che la soluzione è unica, in quanto  $j(X)$  ha un solo minimo assoluto.

Di tale metodo esistono diverse varianti (per esempio considerando un peso diverso nei coefficienti  $a_k$  nel caso in cui il rumore sia maggiore a certe frequenze) utili a raffinare ulteriormente il risultato in casi specifici.

#### 4.2.2 IL LASSO

Il LASSO, acronimo di Least Absolute Shrinkage and Selection Operator, è una tecnica di stima parametrica che permette di evitare il fenomeno di *overfitting* (eccessivo adattamento), ovvero quel fenomeno per cui il modello statistico si adatta ai dati osservati usando un numero eccessivo di parametri. Tale fenomeno si verifica in generale quando si costruisce un modello statistico troppo complesso rispetto ai dati che si hanno a disposizione, non rispettando così la ricerca di un modello parsimonioso. Questo problema viene risolto dalla proprietà di regolarizzazione che utilizza il LASSO, algoritmo introdotto nel 1996 da Tibshirani.

Un esempio di uso di algoritmo LASSO consiste nella medesima operazione di stima parametrica dei minimi quadrati, introducendo però un vincolo sulla norma  $\ell_1$  dei parametri. Si ha la seguente formulazione:

$$\|Y - BX\|^2 + \theta \sum_{i=0}^{N-1} |x_i|$$

Dato  $\hat{X}$  un opportuno vettore di parametri stimati per dare il vettore di previsione  $\hat{Y}$ , con  $\hat{Y} = B\hat{X}$  e con l'errore quadratico totale pari a

$$j(X) = \|Y - \hat{Y}\|^2 \quad (4.2)$$

segue che LASSO sceglie  $\hat{X}$  in maniera tale da minimizzare la  $j(X)$ , soggetta ad un vincolo su

$$T(X) = \sum_{i=0}^{N-1} |x_i| \quad (4.3)$$

Il LASSO agisce tendendo a ridurre a 0 i coefficienti  $x_i$  uno alla volta; questa riduzione migliora la precisione della stima “selezionando le variabili più valide affinché esse vengano incluse nel risultato finale. LASSO ha una proprietà tale per cui si stipula che, per qualsiasi valore vincolato di  $T(X)$  solo un sottoinsieme di valori di  $x_i$  è diverso da 0.

La stima LASSO verrà utilizzata nel Compressed Sensing rumoroso per determinare la soluzione ad un problema di stima di segnale  $\bar{X}$ .

## Capitolo 5

# Compressed Sensing senza rumore

Un segnale può essere considerato come una serie di valori, di coefficienti relativi a una certa base prefissata (canonica, di esponenziali complessi...). Si supponga di voler comprimere un dato segnale: tale compressione può avvenire con o senza perdita di informazione. Se la compressione avviene senza perdita, il segnale, una volta compresso, può essere riportato alla sua condizione originaria. La compressione con perdita prevede invece l'impossibilità di riottenere il segnale originario, presupponendo però di poter ottenere un segnale simile a quello originario, eliminato delle sue componenti meno importanti. Se si suppone che il segnale sia *sparsso*, ovvero che il segnale discreto abbia almeno un certo numero di valori  $n$  nulli, allora è possibile eliminare le componenti più piccole del segnale senza perdere grandi quantità di informazione. Il Compressed Sensing consiste perciò nella possibilità di ricostruire un segnale sparso in maniera esatta partendo da un "sottocampionamento (una compressione con perdita) di valori di una rappresentazione particolare di tale segnale (nello specifico dalla sua rappresentazione in serie di Fourier). La chiave della trattazione del Compressed Sensing risiede nello studio delle norme, evidenziando come l'utilizzo di un vincolo sulla norma  $\ell_1$  del segnale consenta di ottenere il segnale più sparso possibile, segnale che è proprio quello cercato. In questo capitolo si chiamerà  $x(n)$  il segnale incognito, periodico di periodo  $N$  e  $a_k$  i coefficienti della serie a tempo discreto di Fourier che si suppongono privi di componente rumorosa. Si chiamerà  $X$  il vettore dei valori del segnale  $x(n)$  e si chiamerà  $Y$  il vettore dei coefficienti  $a_k$ .

### 5.1 Numero di valori discreti indispensabili

Sia data la serie discreta di Fourier, come definita da (2.6), espressa in forma matriciale da

$$Y = BX$$

Si supponga di avere a disposizione il vettore  $\hat{Y}$  parziale, ovvero il vettore sia relativo solo ad alcune frequenze  $k$  ( $k \in \Omega$  con  $\Omega \subset \mathbb{Z}_N$ , dove  $\mathbb{Z}_N$  è l'insieme di numeri  $n \in \mathbb{Z} : n \in [0, N - 1]$ ). La teoria dell'algebra lineare non ci consente di ricostruire esattamente il segnale, in quanto mancano dei dati, dei valori  $a_k$  relativi ad alcune frequenze. Si osservi infatti che il sistema ha  $N$  gradi di libertà, di cui solo  $|\Omega| < N$  sono vincolati, gli altri valori sono arbitrari.

Si supponga inoltre che il segnale  $x(n)$  abbia supporto su un sottoinsieme  $T$  di  $\mathbb{Z}_N$  (abbia dei valori non nulli solo su un  $T \subset \mathbb{Z}$ : si supponga dunque che il segnale sia *sparso*).

Nel caso in cui  $N$  sia primo, esiste un teorema che conferisce la possibilità di ricostruire esattamente il segnale  $x(n)$  se  $|T|$  è abbastanza piccolo.

**Teorema 1 ([2], pg. 5)** *Si supponga che la lunghezza  $N$  del vettore  $X$  sia un numero primo intero. Sia  $\Omega$  un sottoinsieme di  $0, \dots, N - 1$ , e sia  $X$  un vettore con supporto in  $T$  tale che*

$$|T| \leq \frac{1}{2}|\Omega|$$

*. Allora  $X$  può essere ricostruito unicamente da  $\Omega$  e  $\hat{Y}$ .*

*Viceversa, se  $\Omega$  non è l'insieme di tutte le  $N$  frequenze, allora esisteranno due vettori distinti  $X$  e  $V$  tali che  $|\text{supp}(X)|, |\text{supp}(V)| \leq \frac{1}{2}|\Omega| + 1$  e tali che abbiano lo stesso vettore di coefficienti  $\hat{Y}$ .*

In generale, nel caso in cui  $T$  e  $\Omega$  siano scelti entrambi a caso, il teorema (e in particolar modo il Lemma su cui esso si fonda, come spiegato nel paragrafo 5.2) vale ugualmente con alta probabilità. In sostanza, la condizione su  $N$  non è restrittiva.

Tale teorema asserisce che si può ricostruire  $X$  da  $2|T|$  campioni nel dominio della frequenza (coefficienti  $a_k$ ). La ricostruzione consiste nel considerare il vettore  $\bar{X}$  tale che

$$\min \|\bar{X}\|_{\ell_0} \tag{5.1}$$

dove la serie di Fourier di  $\bar{X}$  contiene i coefficienti di  $\hat{Y}$  e dove  $\|\bar{X}\|_{\ell_0}$  è il numero di termini non nulli del vettore  $\bar{X}$ . La computazione di tale valore è tuttavia irrealizzabile (richiede l'enumerazione di tutte le possibili posizioni di  $T$  elementi diversi da zero scelti tra  $N$  posizioni possibili); si ricostruisce perciò  $\bar{X}$  da  $\Omega$  e da  $\hat{Y}$  risolvendo il seguente problema:

$$\min_{\bar{X} \in \mathbb{C}^N} \|\bar{X}\|_{\ell_1} := \sum_{n \in \mathbb{Z}_N} |\bar{x}(n)| \tag{5.2}$$

dove la serie di Fourier di  $\bar{X}$  contiene i coefficienti di  $\hat{Y}$ . Tale condizione permette di ottenere il segnale più sparso possibile.

Le soluzioni di (5.1) e (5.2) sono equivalenti per la maggior parte di scelte di  $T$  e  $\Omega$  con  $|T| \leq \alpha \frac{|\Omega|}{\log N}$  ( $\alpha > 0$  valore costante): risolvere il problema (5.2) permette di ricostruire  $X$  completamente. Questo limite superiore è stabilito sotto l'assunzione che i coefficienti di Fourier siano campionati in maniera casuale (e che quindi il sottoinsieme  $\Omega$  sia scelto a caso tra tutti gli insiemi di quella numerosità). Sfruttando la casualità del campionamento dei valori nel dominio della frequenza (i coefficienti  $a_k$ ) è possibile dimostrare il seguente teorema:

**Teorema 2 ([2], pg. 7)** *Sia  $X \in \mathbb{C}^N$  un segnale discreto con supporto in un insieme sconosciuto  $T$ , e sia  $\Omega$  di grandezza  $|\Omega| = m$  scelto uniformemente a caso. Per un certo parametro di accuratezza  $M$  e con un certo valore  $C_M$  dipendente da  $M$ , se*

$$|T| \leq C_M (\log N)^{-1} |\Omega| \quad (5.3)$$

e quindi

$$|\Omega| \geq \frac{1}{C_M} (\log N) |T| \quad (5.4)$$

allora con probabilità pari a  $O(N^{-M})$  il minimizzatore  $\bar{X}$  del problema (5.2) è unico ed è uguale a  $X$ .

Tale teorema stabilisce quindi che è possibile ricostruire perfettamente un segnale sparso da circa  $|T| \log N$  campioni. Quindi il numero di campioni per la ricostruzione perfetta del segnale non deve andare al di sotto di tale valore, pena la perdita di una quantità considerevole di informazione.

Il risultato è sorprendente: avendo fatto la supposizione di segnale sparso si rende il segnale comprimibile senza perdita di informazione (con alta probabilità).

## 5.2 Unicità della soluzione: un'idea di dimostrazione

Il problema (5.2) pone molti interrogativi: è evidente che una sua soluzione esista, ma non è chiaro se questa soluzione sia unica e se essa sia o meno uguale a  $X$ . Per suggerire come è stata data una risposta a queste domande vengono qui presentati due lemmi, i quali danno un'idea di come Candes e Tao hanno costruito la loro soluzione.

**Lemma 1 ([2], pg.5)** *Sia  $N$  un numero primo intero e siano  $T, \Omega$  sottoinsiemi di  $\mathbb{Z}_N$ . Sia  $\ell_2(T)$  lo spazio dei segnali che sono nulli fuori di  $T$  (rispettivamente  $\Omega$ ). La serie di Fourier discreta  $\mathcal{F}_{T \rightarrow \Omega} : \ell_2(T) \rightarrow \ell_2(\Omega)$  è definita come*

$$\mathcal{F}_{T \rightarrow \Omega}\{X\} := \hat{X}$$

per tutti gli  $X \in \ell_2(T)$ . Se  $|T| = |\Omega|$ , allora  $\mathcal{F}_{T \rightarrow \Omega}$  è una funzione biettiva; se  $|T| \leq |\Omega|$  allora  $\mathcal{F}_{T \rightarrow \Omega}$  è una funzione iniettiva, se  $|T| \geq |\Omega|$  allora  $\mathcal{F}_{T \rightarrow \Omega}$  è suriettiva. La stessa proprietà vale se si sostituisce  $\mathcal{F}$  con la sua inversa  $\mathcal{F}^{-1}$ .

Tale teorema esplicita come la serie discreta di Fourier, intesa come funzione fra due vettori, è una mappa biunivoca, suriettiva e iniettiva sotto alcune condizioni. Si noti che la condizione di iniettività consente alla funzione di essere invertibile.

Il seguente lemma, utilizzando la proprietà di iniettività del sistema, permette di stabilire che la soluzione al problema (5.2) sia unica nel caso in cui esista un vettore  $P$  con alcune proprietà.

**Lemma 2 ([2], pg. 13)** *Sia  $\Omega \subset \mathbb{Z}_N$ . Per un vettore  $X \in \mathbb{C}^N$  con  $T := \text{supp}(X)$  si definisce vettore segno  $\text{sgn}(X)(n) := \frac{X(n)}{|X(n)|}$  quando  $n \in T$  e  $\text{sgn}(X) = 0$  altrove. Si supponga che esista un vettore  $P$  la cui serie discreta di Fourier  $\hat{P}$  abbia supporto in  $\Omega$  tale che*

$$P(n) = \text{sgn}(X)(n)$$

per tutti gli  $n \in T$  e

$$|P(n)| < 1$$

per ogni  $n$  non appartenente a  $T$ . Allora  $\mathcal{F}_{T \rightarrow \Omega}$  è iniettiva, il minimo  $\bar{X}$  del problema (5.2) è unico e uguale a  $X$ . Viceversa, se  $X$  è l'unico minimo di (5.2), allora esiste un vettore  $P$  con le suddette proprietà.

Tramite tale teorema si sposta l'attenzione del problema di unicità e uguaglianza della soluzione all'esistenza di un vettore  $P$  con determinate proprietà. Fissando  $X$  e il suo supporto  $T$ , nel caso in cui  $\Omega$  sia scelto a caso da tutti gli insiemi di grandezza  $m \geq C_M^{-1}|T| \log N$ , segue:

- $P$  può essere costruito con opportune tecniche [2] con probabilità  $1 - O(N^{-M})$
- $|P(n)| < 1$  per  $\forall n$  con probabilità  $1 - O(N^{-M})$

Avendo quindi costruito un vettore  $P$  si può utilizzarlo nel Lemma 2 ottenendo così la dimostrazione dell'unicità e dell'uguaglianza a  $X$  della soluzione di (5.2), sotto le condizioni del Teorema 1.

### 5.2.1 Limitazioni del teorema: $N$ numero primo

Il Teorema 1 e il Lemma 1 richiedono che  $N$ , periodo del segnale, sia un numero primo. Questa condizione non è in genere indispensabile perchè quando si sceglie  $T$  e  $\Omega$  casualmente difficilmente si incappa in una situazione di particolarità irrisolvibile. Tuttavia, in alcuni casi, se  $N$  non è un numero primo il

segnale non è ricostruibile. Si cita in seguito in questo paragrafo un esempio in cui, preso  $N$  non primo, non si è in grado di ottenere il segnale.

Si supponga che  $N$  sia un quadrato perfetto e si consideri il seguente treno di impulsi

$$x(n) = \sum_{l=0}^{N-1} \delta(n - l\sqrt{N})$$

La sua serie discreta di Fourier è la seguente:

$$x(n) = \sum_{k=0}^{N-1} a_k e^{-jk\frac{2\pi}{N}n}$$

con

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-jk\frac{2\pi}{N}n}$$

Si osservi che

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} \left[ \sum_{l=0}^{N-1} \delta(n - l\sqrt{N}) e^{-jk\frac{2\pi}{N}n} \right]$$

scambiando la sommatoria

$$a_k = \frac{1}{N} \sum_{l=0}^{N-1} \left[ \sum_{n=0}^{N-1} \delta(n - l\sqrt{N}) e^{-jk\frac{2\pi}{N}n} \right]$$

$$a_k = \frac{1}{N} \sum_{l=0}^{N-1} e^{-jk\frac{2\pi}{N}l\sqrt{N}}$$

$$a_k = \frac{1}{N} \sum_{l=0}^{N-1} e^{-jk\frac{2\pi}{\sqrt{N}}l}$$

Si osservi che per  $k$  valore multiplo di  $\sqrt{N}$  il coefficiente  $a_k$  è pari a 1; altrimenti è pari a 0, perchè:

$$a_k = \frac{1}{N} \frac{1 - e^{-jk\frac{2\pi}{\sqrt{N}}N}}{1 - e^{-jk\frac{2\pi}{\sqrt{N}}}} = 0$$

Se si prende la restrizione del segnale in maniera tale che i valori  $a_k$  appartengano a  $\Omega$  solo se  $k \neq p\sqrt{N}$  con  $p \in \mathbb{Z}$  allora  $\mathcal{F}\{X\} = 0$ . Il segnale ricostruito  $x(n)$  è dunque ovviamente pari a 0. Evidentemente, il segnale originale (costituito da impulsi) non è ricostruibile.

### 5.3 Il parametro di incoerenza

Nella trattazione del Compressed Sensing un fattore importante è il parametro di coerenza. Il numero di campioni necessari per la ricostruzione di un segnale dipende linearmente da questo parametro di coerenza, il quale a sua volta dipende dalla base di riferimento su cui vengono fatte le  $|\Omega|$  misurazioni. In questa trattazione si è sempre supposto che le misurazioni avvenissero nel dominio della frequenza (e quindi rispetto ad una base di esponenziali complessi); si ipotizzi per questo paragrafo che le misurazioni avvengano rispetto ad una base le cui componenti sono estratte da una popolazione  $F$ .

$F$  può godere di due proprietà:

- Proprietà di isotropia, si dice così se

$$E[bb^*] = I, b \sim F \quad (5.5)$$

- Proprietà di incoerenza, si dice parametro di coerenza  $\mu(F)$  il più piccolo numero tale che preso  $b = (b[1], \dots, b[n]) \sim F$

$$\max_{1 \leq t \leq n} |b[t]|^2 \leq \mu(F) \quad (5.6)$$

Nel caso in cui  $F$  sia la famiglia di esponenziali complessi, segue che  $\mu(F) = 1$ , in quanto

$$|e^{jk(\frac{2\pi}{N})n}| = 1, \forall k, n$$

Tale valore di parametro di coerenza viene utilizzato nel Teorema fondamentale del Compressed Sensing: il valore  $\frac{1}{C_M}$  del Teorema 2 può anche essere riscritto come

$$\frac{1}{C_M} = C_\beta \mu(F) \quad (5.7)$$

rendendo quindi evidente il legame fra parametro di coerenza e numero di campioni necessari per il recupero del segnale. Nel caso in cui  $b$  sia una famiglia di esponenziali complessi si ha un'ottima incoerenza, in quanto  $\mu(F)$  diventa molto basso, permettendo così di ricostruire il segnale partendo da un basso numero di campioni. Se invece si usasse un'altra base di rappresentazione del segnale, il numero dei campioni necessari alla ricostruzione del segnale (supponendo la medesima probabilità di ricostruzione) potrebbe essere anche superiore a quello ottenuto con la base di esponenziali complessi.

## Capitolo 6

# Compressed Sensing rumoroso

Sotto opportune condizioni si è osservato nel capitolo precedente che è possibile il recupero di un segnale discreto partendo da alcuni elementi della relativa serie discreta di Fourier. Questo capitolo presenta la possibilità di recupero di un segnale anche in caso di campioni di serie discreta di Fourier affetti da rumore. Tale caso è di utile applicazione: i segnali trasmessi tramite un mezzo fisico sono sempre interessati dalla presenza di rumore a causa di fenomeni fisici che intervengono in gran parte delle occasioni reali.

### 6.1 Coefficienti rumorosi: segnale sparso

Si consideri di avere un certo numero  $m$  di coefficienti  $a_k$  della serie di Fourier discreta affetti da rumore additivo. Si può immaginare che, a causa della presenza del rumore, il coefficiente  $a_k$  possa essere così costituito

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-jk \frac{2\pi}{N} n} + \sigma z_k \quad (6.1)$$

dove  $z_k$  è un campione del rumore bianco normalizzato. In relazione matriciale esso si può esprimere come

$$Y_1 = B_1 X + \sigma Z$$

dove  $Y_1$  è il vettore dei coefficienti di Fourier  $a_k$ ,  $B_1$  è la matrice di esponenziali complessi,  $X$  è il vettore del segnale  $x(n)$  e  $Z$  è il vettore del rumore. Si noti che in tale espressione si può dividere ambo i membri per  $\sqrt{m}$  in maniera tale da ottenere colonne di  $B_1$  aventi norma unitaria.

$$Y = BX + \sigma_m Z \quad (6.2)$$

Nel caso in cui si stia sottocampionando (ovvero il numero di coefficienti  $m$  sia minore del periodo del segnale  $N$  e sia dell'ordine di  $\mu(F)|T|\log N$ ) e il segnale  $x(n)$  sia sparso, per ricostruire  $X$  dai dati rumorosi  $Y$ , secondo la relazione (6.2) si può invocare la minimizzazione  $\ell_1$  per ottenere uno stimatore  $\hat{X}$  tale che

$$\|\hat{X} - X\|^2 \leq \text{polylog}(n) \frac{|T|}{m} \sigma^2 \quad (6.3)$$

Quindi il segnale ricostruito ha un errore quadratico che è molto basso se  $\sigma^2$  è piccolo, dato che il limite è stretto.

In sostanza, nel caso di un segnale sparso campionato  $m$  volte affetto da rumore additivo, è possibile ricostruire con buona approssimazione il segnale originario. Si osserverà nel paragrafo seguente che un segnale affetto da rumore può essere ricostruito e che l'errore di stima sarà determinato sia dall'errore in caso di assenza di rumore che dall'energia del rumore stesso.

## 6.2 Ricostruzione di un segnale

Sia dato il problema (6.2); si supponga che si stia sottocampionando e quindi che si abbiano  $m$  campioni della serie discreta di Fourier con  $m < N$ . Il seguente risultato si occupa di segnali arbitrari che non siano necessariamente sparsi (l'immagine non è esattamente sparsa) e di rumore.

Per ricostruire  $X$  partendo dai dati  $Y$  si considera l'algoritmo LASSO che risolve il seguente problema ai minimi quadrati

$$\min_{\bar{X} \in \mathbb{R}^N} \frac{1}{2} \|B\bar{X} - Y\|^2 + \lambda \sigma_m \|\bar{X}\|_{\ell_1} \quad (6.4)$$

dove tale richiesta consiste nell'identificare il vettore  $\bar{X}$  che minimizza la somma del rumore (il *residuo*, come si faceva nella stima con  $m = N$ ) e della norma  $\ell_1$  del vettore  $\bar{X}$  (ulteriore condizione determinata dal fatto che  $m < N$ ). Supponendo che il vettore del rumore  $Z$  sia Gaussiano ( $Z \sim N(0, I)$ ) e supponendo che il segnale  $X$  sia sparso e quindi con supporto in  $T$ , vale il seguente teorema:

**Teorema 1 ([1], pg. 7)** *Sia  $X$  un vettore arbitrario fissato in  $\mathbb{R}^N$ . Allora con probabilità minima  $1 - \frac{6}{N} - 6e^{-\beta}$  la soluzione al problema (6.4) con  $\lambda = 10\sqrt{\log N}$  obbedisce*

$$\|\hat{X} - X\| \leq \min_{1 \leq |T| \leq \bar{T}} C(1 + \alpha) \left( \frac{\|X - X_{|T|}\|_{\ell_1}}{\sqrt{|T|}} + \sigma \sqrt{\frac{|T| \log N}{m}} \right) \quad (6.5)$$

purchè  $m \geq C_\beta \mu(F) |\bar{T}| \log N$ . Se si misura l'errore nella norma  $\ell_1$ , allora

$$\|\hat{X} - X\|_{\ell_1} \leq \min_{1 \leq |T| \leq \bar{T}} C(1 + \alpha) \left( \|X - X_{|T|}\|_{\ell_1} + |T| \sigma \sqrt{\frac{\log N}{m}} \right) \quad (6.6)$$

dove  $C$  è una costante numerica  $X_{|T|}$  è la migliore  $|T|$ -sparsa approssimazione di  $X$  e  $\alpha = \sqrt{\frac{(1+\beta)|T|\log^5 N}{m}}$  che non è mai più grande di  $\log^2 N$  in questa impostazione.

Tale teorema specifica quindi che applicando la stima LASSO del segnale è possibile individuare una limitazione sulla norma euclidea e sulla norma  $\ell_1$  dell'errore di stima, permettendo perciò di ottenere il segnale sparso originale con buona probabilità.

Nell'interpretazione di tale limite si osservi che il primo termine corrisponde al limite conseguito in caso di mancanza dell'errore e che il secondo termine dipende dalle caratteristiche del rumore (la sua varianza).

Nel caso finora trattato si è osservato che con la tecnica LASSO è possibile ottenere come stima un segnale sparso, in quanto tale algoritmo azzeri i parametri meno significativi, quelli la cui perdita non implica l'impossibilità di ricostruzione del segnale.

Si osservi che l'algoritmo LASSO non è l'unico che permette di computare una regressione sparsa. Anche altri algoritmi, come per esempio l'algoritmo del Dantzig Selector è in grado di ricostruire il segnale originario, cercando

$$\min_{\bar{X} \in \mathbb{R}^N} \|\bar{X}\|_{\ell_1} \quad (6.7)$$

vincolato a

$$\|A^*(A\bar{X} - Y)\|_{\ell_\infty} \leq \lambda\sigma_m \quad (6.8)$$

Tale algoritmo permette di ottenere i seguenti limiti superiori all'errore di stima:

**Teorema 2** ([1], pg. 8) *Il Dantzig Selector, con  $\lambda = 10\sqrt{\log N}$  e tutte le altre relazione del Teorema 1 obbedisce*

$$\|\hat{X} - X\| \leq \min_{|T| \leq |\hat{T}|} C(1 + \alpha^2) \left( \frac{\|X - X_{|T|}\|_{\ell_1}}{\sqrt{|T|}} + \sigma\sqrt{\frac{|T|\log N}{m}} \right) \quad (6.9)$$

$$\|\hat{X} - X\|_{\ell_1} \leq \min_{|T| \leq |\hat{T}|} C(1 + \alpha^2) \left( \|X - X_{|T|}\|_{\ell_1} + |T|\sigma\sqrt{\frac{\log N}{m}} \right) \quad (6.10)$$

con le stesse probabilità del Teorema 1.

Quindi, seppur usando un algoritmo di stima diverso dal precedente, si ottengono dei limiti all'errore di stima simili a quelli individuati per il LASSO: la somma di due componenti, una dovuta all'errore in caso di mancanza di rumore e l'altra dovuta alle caratteristiche peculiari del rumore stesso.



## Capitolo 7

# Un' applicazione del Compressed Sensing rumoroso

La teoria del Compressed Sensing trova numerose applicazioni pratiche, inerenti vari ambiti scientifici. Nell'ambito delle telecomunicazioni, per esempio, la teoria del Compressed Sensing rumoroso può essere utile in termini di risparmio di dati inviati, permettendo la trasmissione di un segnale "compresso" senza la perdita sostanziale di informazioni. Altri esempi possono riguardare l'ambito biomedico; in seguito viene presentata un' applicazione particolare del Compressed Sensing al campo della diagnostica medica: il PAT.

### 7.1 Il PAT: Photoacoustic computed tomography

L'effetto fotoacustico venne osservato per la prima volta alla fine del XX secolo da Alexander Graham Bell. Esso consiste nella risposta acustica di un materiale sottoposto a radiazione luminosa, dovuta essenzialmente al riscaldamento del gas circostante la parte colpita dalla luce. Tale fenomeno viene sfruttato nel campo biomedico per la costruzione di dispositivi di diagnostica: si è verificato infatti che le sostanze biologiche reagiscono in modo diverso nel momento in cui vengono sottoposte a una radiazione luminosa appropriata. L'assorbimento ottico è infatti associato a proprietà fisiologiche, come la concentrazione dell'emoglobina e la saturazione dell'ossigeno; la grandezza delle onde acustiche è quindi associata a zone di particolare rilievo biologico.

Il PAT è uno strumento diagnostico che, tramite un rilevatore di onde acustiche, permette la ricostruzione di immagini di tessuti e apparati biologici. Esso sfrutta alcune equazioni, dette equazioni fotoacustiche, per calcolare e rappresentare, a partire dalla risposta acustica, l'immagine biologica.

Il problema nella ricostruzione di un'immagine biologica deriva dal fatto che non si è in grado di effettuare sperimentazioni su un angolo di  $4\pi$  (nel caso di immagini a 3D,  $2\pi$  nel caso di immagini 2D), ovvero non si è in grado di colpire a 360 gradi la parte interessata con le radiazioni luminose. In sostanza, ci si trova ad avere a che fare con un numero ridotto rispetto al necessario di misurazioni strumentali. Su base teorica (Agranovsky and Quinto, 1996 [6]) si è comunque in grado di ricostruire l'immagine completamente; tuttavia gli studi teorici non forniscono una procedura unica e inequivocabile per la ricostruzione e neppure garantiscono che tale operazione sia in pratica stabile. Numerosi algoritmi per la risoluzione sono stati presentati; anche il Compressed Sensing rientra come tecnica di soluzione del problema.

Si supponga che l'immagine che si va a catturare possa essere rappresentata come un segnale sparso relativo ad una certa base  $\psi$  (non necessariamente la base canonica). Si ipotizzi che la base su cui si vanno ad effettuare le misurazioni sia incoerente con la base su cui il segnale è sparso. Sotto queste condizioni si supponga di effettuare un certo numero di misurazioni (dell'ordine di  $|T|\log N$ , in cui  $|T|$  è la cardinalità del supporto del segnale nella sua rappresentazione sparsa e  $N$  la lunghezza del vettore di tale segnale), anche affette da rumore. Per la teoria del Compressed Sensing, è possibile, partendo da un numero limitato di acquisizioni (e quindi da un'angolazione limitata o da un numero basso di misurazioni riducendo così il tempo di esposizione), ricostruire completamente il segnale sparso con alta probabilità.

### 7.1.1 Misurazione sperimentale

La seguente misurazione sperimentale è stata effettuata da Dong Liang, Hao F. Zhang e Leslie Ying, presentata nel loro articolo *Compressed-sensing Photoacoustic Imaging based on random optical illumination*[6].

La loro misurazione sperimentale utilizza un dispositivo di rilevazione così costituito: si frappone fra zona da colpire e sorgente luminosa una maschera, la quale permetterà il passaggio di radiazioni luminose solo in determinati punti. Tali punti della maschera sono scelti casualmente (così come si campionava casualmente i coefficienti di Fourier nella trattazione precedente). Il dispositivo è dotato di due trasduttori, i quali rilevano il segnale acustico; essi sono posti ad un'angolazione ben precisa, in modo da avere due angoli di visuale. I valori rilevati dalle onde acustiche sono in relazione con il segnale dell'immagine tramite la seguente equazione nel dominio di Fourier:

$$\tilde{p}(\mathbf{r}_0, k, m) = -ik \int \int_{S'} dS' \tilde{G}_k^{(out)}(\mathbf{r}', \mathbf{r}_0) I_m(\mathbf{r}') p_0(\mathbf{r}') \quad (7.1)$$

dove  $\tilde{G}_k^{(out)}$  è una funzione esponenziale complessa che specifica la relazione fra posizione di rilevazione e posizione di emissione;  $\tilde{p}(\mathbf{r}_0, k, m)$  è la pressio-

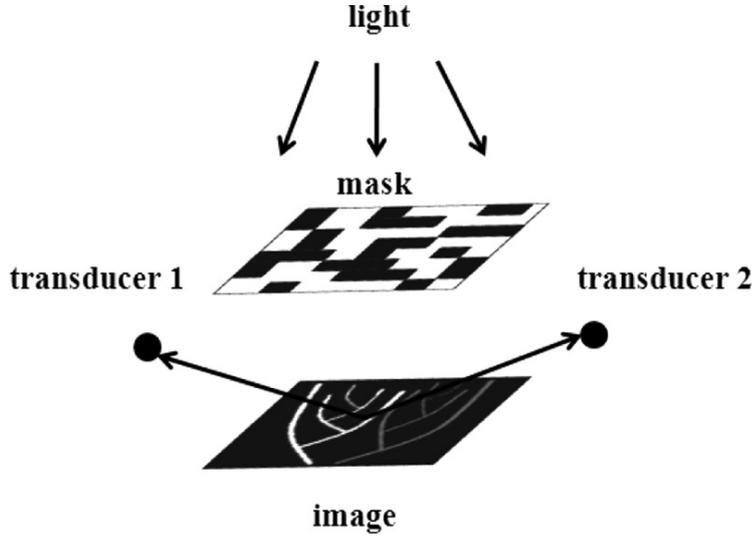


Figura 7.1: Riproduzione della modalità di cattura

ne rilevata nella posizione  $r_0$  sulla superficie da analizzare,  $k$  è l'ampiezza dell'onda acustica,  $m$  identifica la maschera usata;  $p_0(r')$  denota la sorgente acustica mentre  $I_m(r')$  denota l'illuminazione casuale dovuta alla maschera  $m$  nel punto  $r'$ . In seguito a operazione di discretizzazione, tale equazione diventa lineare nella forma:

$$Y = BX \quad (7.2)$$

dove il vettore  $Y$  rappresenta le misurazioni di pressione nel dominio di Fourier,  $B$  una matrice lineare che mette in relazione il segnale  $X$  con le rilevazioni effettuate dai due trasduttori espresse da  $Y$ . Si noti che il segnale  $X$  è riferito ad una certa base  $\phi$ . Il numero dei campioni totali è quindi determinato dal doppio del numero di maschere ( $2m$ ), in quanto due sono i trasduttori del segnale. Se esiste una certa base  $\psi$  per cui la rappresentazione del segnale è sparso, avendo un numero di campioni ridotti, è sufficiente fare in modo che:

$$X = \min_X \|\Phi X\|_{\ell_1} + TV(X) \quad (7.3)$$

dove  $\Phi$  è la matrice che consente il cambiamento di base fra la base  $\psi$  e quella  $\phi$  e  $TV(X)$  è un fattore detto *total variation* dell'immagine, incluso per migliorare l'immagine nel caso specifico.

Si consideri i seguenti risultati di simulazione (figura 7.2): la figura chiarisce come la ricostruzione con il metodo del Compressed Sensing permette di migliorare la qualità dell'immagine; inoltre tale proprietà tanto più migliora quanto più aumentano le maschere (e quindi le acquisizioni di dati).

Un altro esempio è dato dall'immagine *in vivo* del sistema cardiovascolare delle orecchie di un topo (figura 7.3):

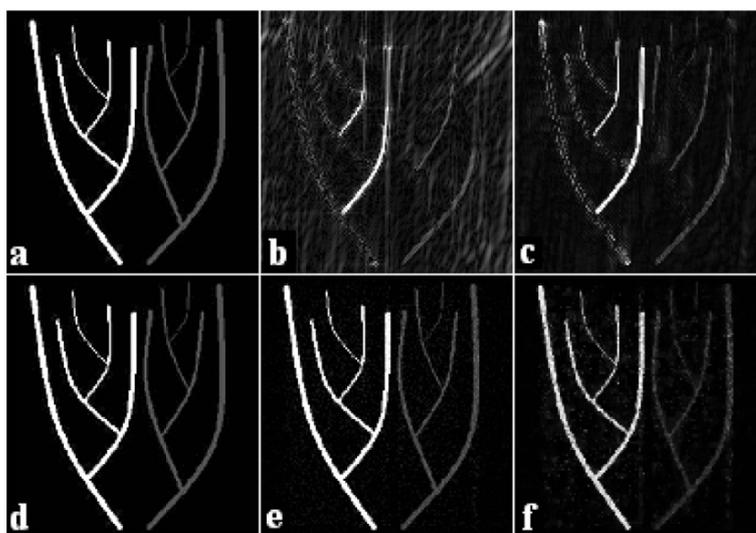


Figura 7.2: (a) Lo spettro originale; (b) Ricostruzione con il metodo usuale di proiezione inversa; (c) Ricostruzione con Compressed Sensing di una acquisizione usuale con PAT da 86 acquisizioni; (d) Ricostruzione con il metodo proposto usando 43 (e) 32 (f) 21 maschere usando 2 angoli di visuale

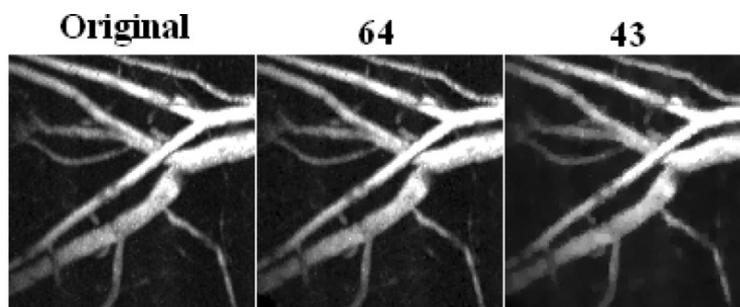


Figura 7.3: Un immagine PA di un orecchio *in vivo* di un topo nella colonna di sinistra; la ricostruzione con Compressed Sensing tramite 64 maschere nella colonna centrale e ricostruzione con 43 maschere nella colonna a destra.

Si noti anche in questo caso come il Compressed Sensing consente di diminuire il numero di acquisizioni per ottenere comunque un'immagine di buona qualità.

In conclusione, con questa analisi sperimentale si è verificato che tramite

la procedura del Compressed Sensing è possibile limitare il numero di acquisizioni rispetto a quelle necessarie usualmente per la ricostruzione di un segnale. Il CS permette inoltre di risolvere le difficoltà di acquisizione di dati quando non è possibile effettuare rilevazioni da varie angolazioni, adeguando lo studio all'esigenza di non invasività.



## Capitolo 8

# Conclusioni

Nella stesura di questo elaborato si è cercato di introdurre il concetto di Compressed Sensing, presentandolo con e senza rumore, evidenziandone le caratteristiche principali, le condizioni e le limitazioni nell'uso. Si è trascurata la parte dimostrativa, rimandando alla lettura diretta degli articoli scientifici le spiegazioni tecniche e le argomentazioni matematiche. L'ultima parte dell'elaborato tratta un aspetto un po' più pratico, magari più interessante dal punto di vista ingegneristico rispetto alla pura teoria. Proprio in questa parte, a mio parere, si evidenziano le molteplici applicazioni pratiche della teoria del Compressed Sensing e si rende chiaro l'importante ruolo che il Compressed Sensing potrebbe svolgere nel prossimo futuro.

Ringrazio sentitamente il Prof. Pavon per avermi presentato questo argomento attuale e interessante e per avermi seguito nella stesura di questa tesi.



# Bibliografia

- [1] Emmanuel J. Candès and Yaniv Plan. A probabilistic RIPless theory of compressed sensing. *IEEE Trans. Inform. Theory*, 2010.
- [2] Emmanuel J. Candès, J. Romberg, and T.Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 2006.
- [3] Emmanuel J. Candès and T.Tao. Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 2006
- [4] Cristina Ronconi. Appunti di geometria. Univer editrice, 2002
- [5] Robert Tibhsirani, Regression shrinkage and selection via the lasso *Statist. Soc. B.*, 1996
- [6] Dong Liang, Hao F. Zhang and Leslie Ying. Compressed-sensing Photoacoustic Imaging based on random optical illumination, *Int. J. Functional Informatics and Personalised Medicine*, Vol. 2, No. 4, pp.394,406, 2009
- [7] M.J. Roberts. Signals and Systems: Analysis of Signals Through Linear Systems. McGraw-Hill, 2004