

UNIVERSITÀ DEGLI STUDI DI CAGLIARI

FACOLTÀ DI INGEGNERIA
Corso di Laurea in Ingegneria Elettrica ed Elettronica

**RICOSTRUZIONE COMPRESSA DI UN
SEGNALE
MEDIANTE
COMPRESSIVE SENSING**

Tesi di Laurea Triennale

Relatore:
Prof. Giuseppe Rodriguez

Candidato:
Lorenzo Fanari

Anno Accademico 2014/2015

alla mia famiglia

Indice

1	Introduzione	5
2	Trasformata Wavelet	7
2.1	Cosa sono le Basi?	7
2.2	Trasformata Wavelet Continua	9
2.3	Trasformata Wavelet Discreta	10
3	Compressive Sensing	12
3.1	Il Problema del Sensing	12
3.2	Sparsità	14
3.3	Campionamento Incoerente	16
3.4	Sottocampionamento e Sparse Signal Recovery	17
3.5	Compressive Sampling Robusto	20
3.6	Isometrie Ristrette	21
3.7	Ricostruzione Generale del Segnale a Partire da Dati Sottocampionati	22
3.8	Ricostruzione Robusta di Segnali Affetti da Rumore	23
3.9	Acquisizione Casuale	24
3.10	Che Cos'è il Compressive Sampling?	26
4	Esempi e Applicazioni	27
4.1	Esempi	27
4.2	Single-Pixel Camera	32
4.3	Astronomia	33
4.4	Face Recognition	34
5	Conclusioni	35

Capitolo 1

Introduzione

Tipicamente il campionamento di segnali segue il teorema di Shannon: la frequenza di campionamento deve risultare, almeno, il doppio della frequenza massima del segnale (chiamata anche frequenza di Nyquist). Nei fatti, questo principio risulta alla base di tutti i protocolli utilizzati negli apparecchi elettronici audio-visivi, dispositivi medici per l'acquisizione di immagini, ricevitori radio e via dicendo. Per esempio, nel campo della conversione dei dati, la tecnologia alla base dei convertitori analogico-digitali (ADC) implementa la rappresentazione di Shannon solitamente quantizzata: il segnale è campionato uniformemente alla frequenza di Nyquist.

In questa tesi, si tratterà la teoria del Compressive Sensing, conosciuto come anche come Compressive Sampling o più semplicemente CS, un nuovo paradigma della campionatura che va contro le comuni conoscenze in campo dell'acquisizione dei dati. La teoria del CS afferma che si può ricostruire un segnale o un'immagine a partire da un numero ridotto di campioni o misure rispetto al metodo tradizionalmente utilizzato. Per rendere tutto ciò possibile, CS si basa su due principi: **Sparsità**, che riguarda il segnale d'interesse, l'**Incoerenza** che si basa sulle modalità di acquisizione.

1. La sparsità esprime l'idea che la “frequenza in cui è presente informazione” in un segnale a tempo continuo può risultare molto più piccola rispetto alla sua larghezza di banda, oppure, considerando segnali discreti, un numero di gradi di libertà che è comparabile con un valore inferiore rispetto alla sua lunghezza. Più precisamente, CS esprime il fatto che molti segnali naturali sono sparsi o comprimibili nel senso che hanno rappresentazioni concise quando vengono espressi con la corretta base Ψ .
2. L'incoerenza estende il dualismo tra tempo e frequenza e inoltre esprime l'idea che gli oggetti hanno una rappresentazione sparsa Ψ che deve

rispettare la condizione di sparsità nel dominio in cui questi sono stati acquisiti.

Il punto cruciale di tutto ciò è la possibilità di ideare protocolli di acquisizione che catturano le informazioni utili contenute all'interno di un segnale sparso, e condensarle in una piccola quantità di dati. Questi protocolli sono non adattativi e richiedono semplicemente la correlazione del segnale con un piccolo numero di forme d'onda. Il punto di forza di questa tipologia di protocolli è quello di consentire al sensore di catturare l'informazione presente in un segnale sparso utilizzando una piccola quantità di dati. A conclusione di questa breve introduzione si può dire che, CS è un protocollo molto semplice ed efficiente di acquisizione dei segnali che campiona a basse frequenze e che successivamente, attraverso la potenza computazionale, ricostruisce un segnale a partire da quello che risulterebbe un set incompleto di misure.

Capitolo 2

Trasformata Wavelet

Una delle parti fondamentali del CS, come verrà detto successivamente, si basa sull'utilizzo di basi con particolari caratteristiche la cui funzione sarà quella di rappresentare le funzioni campionate. Tra queste hanno notevole importanza le basi Wavelet. Queste vengono utilizzate, ad esempio, nelle codifica di immagini con lo standard JPEG-2000.

In poche parole, le Wavelets sono funzioni matematiche che suddividono il dato in esame in componenti di frequenze diverse, e quindi consente di studiare ogni componente con una risoluzione abbinata alla sua scala.

2.1 Cosa sono le Basi?

Per spiegare al meglio il concetto, si consideri il caso digitale (vettori), al posto del caso analogico (funzioni) .

Qualsiasi vettore bidimensionale (x, y) risulta essere combinazione del vettore $(1, 0)$ e di $(0, 1)$. Questi due vettori risultano essere la base per (x, y) . Perché? Nota quella x moltiplicata per $(1, 0)$, si ottiene il vettore $(x, 0)$, mentre moltiplicando y per il vettore $(0, 1)$, si ottiene $(0, y)$. La somma di questi due vettori sarà (x, y) .

Le migliori basi sono quelle che godono delle proprietà di perpendicolarità, o di ortogonalità tra vettori. Nel caso $(1, 0)$ e $(0, 1)$ questa proprietà viene soddisfatta.

Detto ciò, è possibile riprendere il discorso nel caso analogico. Al posto del vettore (x, y) , si consideri una funzione $f(x)$. Si ipotizzi, a titolo di esempio, che la $f(x)$ corrisponda ad una nota musicale, come il LA in un'ottava particolare. Questa è ricostruibile sommando tra loro seni e coseni attraverso una combinazione di ampiezze e frequenze. In questo caso le basi sono i seni e i coseni, e gli elementi della sintesi di Fourier. Per i seni e i coseni

scelti, è possibile scegliere quali di questi risultano ortogonali. Come? Semplicemente scegliendo una combinazione di funzioni di seno e coseno il cui prodotto interno dia zero. Il particolare set di funzioni che sono ortogonali e che costruiscono $f(x)$ sono definite basi.

2.2 Trasformata Wavelet Continua

La Trasformata Wavelet Continua di un segnale $x(t)$ è definita come una trasformazione lineare definita dall'integrale

$$\begin{aligned} W_{\psi x}(b, a) &= \int_{-\infty}^{\infty} x(t) \overline{\psi_{ba}(t)} dt \\ &= \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \overline{\psi\left(\frac{t-b}{a}\right)} dt \\ &= \langle x(t), \psi_{ba}(t) \rangle, \end{aligned}$$

dove a rappresenta il parametro di scala che determina l'espansione o la compressione della wavelet, e b centra l'intervallo d'interesse.

Detto ciò, la trasformata da informazioni sul segnale $f(t)$ con una finestra temporale

$$[b + at^* - a\Delta_{\psi}, b + at^* + a\Delta_{\psi}]$$

centrata in $b + at^*$ e con raggio $a\Delta_{\psi}$. Si può osservare che se a aumenta, la finestra temporale nel dominio spettrale tende a ridursi, mentre nel dominio temporale tende ad aumentare.

Oltre a ciò, la trasformata localizza anche nel dominio della frequenza lo spettro della funzione $f(t)$ con la finestra

$$\left[\frac{1}{a}(\omega^* - \Delta_{\Psi}), \frac{1}{a}(\omega^* + \Delta_{\Psi})\right].$$

Si può dire che con la CWT al variare della scala utilizzata, varia la banda di frequenza da analizzare. Più semplicemnte, per poter processare un segnale a basse frequenze, occorre utilizzare wavelet a scala larga.

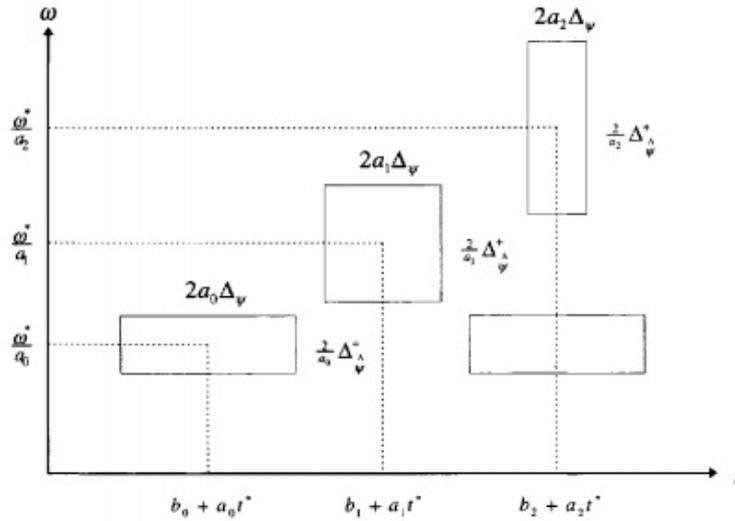


Figura 2.1: Finestre tempo-frequenza

2.3 Trasformata Wavelet Discreta

La CWT presenta 3 proprietà che rendono difficile il suo diretto utilizzo:

1. La ridondanza della CWT. In questo caso, la trasformata di wavelet è calcolata traslando continuamente una funzione continua scalabile sopra un segnale e infine si calcola la correlazione tra le due. In questo modo, le funzioni scalate non saranno mai vicine a basi ortogonali e perciò si otterranno coefficienti wavelet altamente ridondanti.
2. Infinito numero di coefficienti wavelet nella trasformata wavelet, si vorrebbe avere un numero limitato di questi coefficienti.
3. Per molte funzioni le trasformate wavelet non hanno soluzioni analitiche e queste possono essere calcolate esclusivamente numericamente o attraverso un computer.

Per queste ragioni, la CWT subisce un processo di discretizzazione che porterà alla sua discretizzazione, ovvero la DWT (Trasformata Discreta di Wavelet).

Ciò che consente di ottenere una trasformata meno ridondante è la scelta accurata dei parametri a e b , questi vengono scelti nel seguente modo:

$$a = 2^{-j}$$

$$b = k2^{-j},$$

in maniera tale da riscrivere la relazione per la CWT nel modo seguente:

$$\begin{aligned}(W_{\psi f})(k2^{-j}, 2^{-j}) &= 2^{\frac{j}{2}} \int_{-\infty}^{\infty} f(t) \overline{\psi\left(\frac{t - k2^j}{2^{-j}}\right)} dt \\ &= 2^{\frac{j}{2}} \int_{-\infty}^{\infty} f(t) \overline{\psi(2^j t - k)} dt\end{aligned}$$

dove con j e k si rappresentano punti del piano tempo-frequenza aventi coordinate $(\frac{k}{2^j}, \frac{1}{2^j})$.

Queste coordinate corrispondono ai coefficienti della DWT. In particolare, la DWT genera un insieme sparso di valori che rappresentano la correlazione tra il segnale analizzato e la wavelet.

Se in ingresso si considera una sequenza numerica, ottenuta ad esempio campionando il segnale $f(t)$, l'integrale della trasformata diventa una sommatoria

$$(W_{\psi f})(k2^{-j}, 2^{-j}) = 2^{\frac{j}{2}} \sum_n f(n) \overline{\psi(2^j t - k)}.$$

Capitolo 3

Compressive Sensing

3.1 Il Problema del Sensing

In questa tesi verrà discusso in che modo l'informazione di un segnale $f(t)$ è ottenuta da funzionali lineari che registrano i valori:

$$y_k = \langle f, \varphi_k \rangle, k = 1, \dots, m. \quad (3.1)$$

Semplicemente si correla l'oggetto che si vorrebbe acquisire con le forme d'onda φ .

Se ad esempio, le forme d'onda fossero dei delta di Dirac (impulsi) allora y sarebbe un vettore di valori campionati di f nel dominio del tempo o dello spazio.

Se le forme d'onda fossero funzioni che rappresentano i pixel di un'immagine da campionare, allora y sarebbero le immagini tipicamente acquisite da un sensore della fotocamera digitale.

Se le forme d'onda fossero sinusoidi, allora y sarebbe un vettore di coefficienti di Fourier; questa è la modalità di sensing utilizzata nell'imaging a risonanza magnetica (MRI).

Sebbene si potrebbe espandere la teoria del CS a segnali continui spazio/tempo, nel caso in esame si darà attenzione ai segnali discreti $f \in \mathbb{R}^n$. Le ragioni sono essenzialmente due: la prima è la semplicità di concetto, mentre la seconda motivazione è data dal fatto che la teoria del CS applicata a segnali discreti è molto più sviluppata.

Detto ciò, ci si concentrerà nelle situazioni di sottocampionamento dove il numero delle misurazioni m risulta essere più piccolo della dimensione n del segnale f .

Come tipologia di problema risulta essere molto comune per una varietà di ragioni. Per esempio, il numero di sensori può essere limitato oppure le



Figura 3.1: Esempio di registrazione di valori, utilizzando come forma d'onda φ una senoide

misure possono risultare troppo dispensive come nel processo di acquisizione immagini attraverso lo scattering, oppure potrebbe accadere che il processo di sensing risulti lento mentre l'oggetto potrebbe essere misurato in alcuni istanti di tempo come nel MRI e così via.

Queste circostanze fanno sorgere importanti domande:

Risulta possibile un'accurata ricostruzione con $\mathbf{m} \ll \mathbf{n}$ misurazioni?

È possibile rappresentare una forma d'onda che catturi la maggior parte delle informazioni di f con $\mathbf{m} \ll \mathbf{n}$?

Come si può approssimare f con queste informazioni?

La lettera A indica la matrice di sensing $m \times n$ con i vettori $\varphi_1^*, \dots, \varphi_m^*$ come righe (φ^* è il complesso coniugato di φ).

Il processo di ricostruzione di $f \in \mathbb{R}^n$ da $y = Af \in \mathbb{R}^m$, con $\mathbf{m} < \mathbf{n}$ è un tipico sistema **sottodeterminato** (ossia vi sono più incognite che equazioni: in questo caso il sistema potrebbe avere *infinite soluzioni*): ci sono infatti infiniti segnali \tilde{f} per cui $A\tilde{f} = y$. Questo risulta un problema mal posto e per poterlo risolvere si passa ad un **problema di minimi quadrati** (*least squares problem*), in cui è possibile trovare una soluzione al problema mal posto, ricavando la condizione di varianza minima in un generico sistema $Ax = b$:

$$\min_{x \in \mathbb{R}^n} \|Ax - b\|_2$$

La teoria di Shannon consente di ricostruire un segnale $f(t)$ a bassa banda con un piccolo numero (uniforme) di campioni.

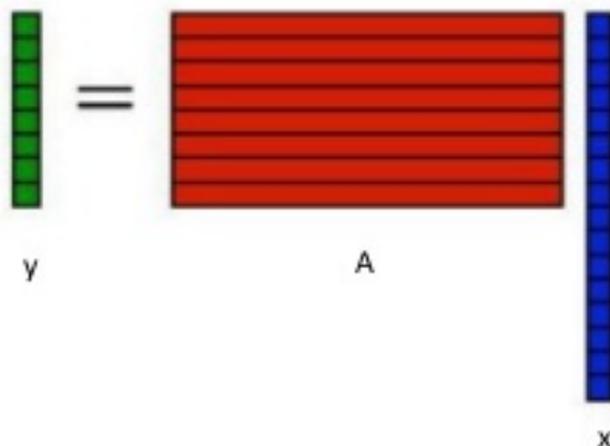


Figura 3.2: Rappresentazione di un sistema sottodeterminato

3.2 Sparsità

Molti segnali naturali hanno una rappresentazione compatta quando sono espressi con basi adeguate.

Con il simbolo f viene definito sia il segnale analogico, sia il segnale discretizzato (questa definizione sarà fondamentale nella comprensione dei calcoli).

Matematicamente parlando, un vettore $f \in \mathbb{R}^n$ (ad esempio un pixel n -esimo di un'immagine) si può espandere in una base ortonormale, (ad esempio una base wavelet) che conterrà le funzioni campionate, $\Psi = [\Psi_1 \Psi_2 \dots \Psi_n]$ in intervalli di tempo t_i come segue:

$$f(t_i) = \sum_{j=1}^n \Psi_j(t_i) x_j \quad (3.2)$$

dove x rappresenta la sequenza di coefficienti di f , $x_j = \langle f, \Psi_j \rangle$.

Risulterà conveniente esprimere f come Ψx (dove Ψ è la matrice $n \times n$) con Ψ_1, \dots, Ψ_n come colonne).

L'implicazione di sparsità ora è chiara: quando un segnale ha una rappresentazione sparsa, si possono scartare i piccoli coefficienti senza avere alte percentuali di perdita del segnale.

Una definizione di sparsità può essere la seguente: considerando un vettore di elementi x_j , esso viene definito *sparsa*, se molti suoi elementi risulteranno **nulli**.

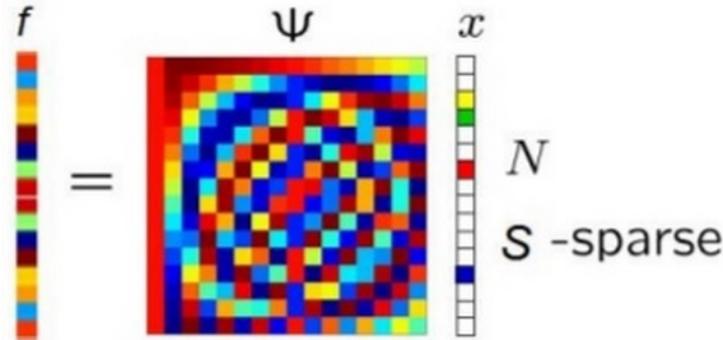


Figura 3.3: Rappresentazione grafica della equazione (3.2)

Formalmente, considerando l'espansione (3.2), f_S è ottenuta conservando solamente gli S coefficienti più grandi di x_j .

Per definizione, $f_s := \Psi x_S$, dove x_S è il vettore di coefficienti x_j in cui tutto, tranne S coefficienti, è posto a zero, ovvero fissato un valore τ , tutti gli x_j tali che: $|x_j| \leq \tau$ verranno inseriti in x_S .

Questo vettore è sparso in senso stretto dato che quasi tutti i suoi coefficienti sono nulli.

Si definisce $S - Sparso$ un oggetto con al massimo S valori non nulli.

L'utilizzo di basi ortonormali, cioè basi che rispettino la proprietà

$$\Psi^T \Psi = I,$$

consente di considerare

$$\| f_S \|_2 = \| \Psi x_S \|_2 = \| x_S \|_2 .$$

Quindi, dato che Ψ è una base ortonormale, allora $\| f - f_S \|_2 = \| x - x_S \|_2$ e se x risulta sparso o comprimibile, allora x risulta ben approssimato da x_S e l'errore dovuto a $\| f - f_S \|_2$ risulta piccolo. In parole povere, è possibile "buttar via" la maggior parte dei coefficienti senza avere grosse perdite.

Questo principio sta alla base delle più moderne tecnologie di compressione *lossy* (ad esempio in campo audio: mp3, dolby ac3, campo immagini: jpeg2000, campo video: xvid, dvix mpeg-4).

La sparsità, quindi, è una caratteristica fondamentale che permette una stima statistica accurata e un'efficiente compressione dei dati.

3.3 Campionamento Incoerente

Si supponga di avere una coppia (Φ, Ψ) di basi ortonormali \mathbb{R}^n .

La base Φ viene utilizzata per acquisire l'oggetto come visto in (3.1) e la seconda rappresenta f .

Definizione 3.1 *La coerenza tra la base di sensing Φ e la base delle rappresentazioni Ψ è:*

$$\mu(\Phi, \Psi) = \sqrt{n} \max_{1 \leq k, j \leq n} |\langle \Phi_k, \Psi_j \rangle| \quad (3.3)$$

La coerenza misura la massima correlazione tra gli elementi di Φ e Ψ . Se Φ e Ψ contengono elementi correlati, la coerenza è alta.

Quindi in base alla formula (3.3), la coerenza può assumere valori: $\mu(\Phi, \Psi) \in [1, \sqrt{n}]$.

Il concetto di coerenza può essere chiarito con il seguente *esempio*:

Considerando una base di Fourier $\Phi_k = n^{-1/2} e^{j \frac{2\pi}{n} kt}$ e la base di campionamento classica $\Psi = \delta(t - j)$ è facile osservare che

$$\max_{1 \leq k, j \leq n} \langle \Phi_k, \Psi_j \rangle = n^{-1/2}$$

quindi $\mu(\Phi, \Psi) = 1$.

Il compressive sampling utilizza coppie di basi con bassa coerenza. Le matrici casuali presentano alta incoerenza con le basi fissate Ψ .

Scegliendo una base ortonormale (base uniforme) Φ in maniera casuale, con alta probabilità, la coerenza tra Φ e Ψ sarà di $\sqrt{2 \log n}$.

Le forme d'onda casuali $(\varphi_k(t))$ con valori indipendentemente distribuiti avranno bassi valori di coerenza per qualsiasi rappresentazione fissata Ψ .

Quindi se il campionamento con sistemi incoerenti è buono, un meccanismo efficiente è quello che punta ad ottenere altre correlazioni con forme d'onda casuali come, ad esempio, il rumore bianco.

3.4 Sottocampionamento e Sparse Signal Recovery

Idealmente, si vorrebbero misurare tutti gli n coefficienti di f , ma nella realtà è possibile osservarne solamente un sottocampione da cui è possibile acquisire le misure

$$y_k = \langle f, \varphi_k \rangle, \forall k \in M \quad (3.4)$$

dove:

$$f(t) = \sum_{j=1}^m x_j \Psi_j(t),$$

mentre $M \subset \{1, \dots, n\}$ è un insieme di cardinalità $m < n$ (l'insieme degli elementi S di x_S). Da questa informazione è possibile ricostruire un segnale attraverso una minimizzazione con la *norma 1*; la ricostruzione proposta f^* è ottenuta da $f^* = \Psi x^*$, dove x^* è la soluzione del problema di ottimizzazione convessa ($\|x\|_1 : \sum_i |x_i|$)

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_1 \quad \text{sogetta a } y_k = \langle \varphi_k, \Psi \tilde{x} \rangle, \forall k \in M. \quad (3.5)$$

In cui

$$y_k = \sum_{j=1}^n \langle \Psi_j, \varphi_k \rangle \tilde{x}_j \quad \text{con } k=1 \dots m$$

dove $\langle \Psi_j, \varphi_k \rangle$ corrisponde all'elemento A_{jk} della matrice A di acquisizione.

Tra tutti gli oggetti $\tilde{f} = \Psi \tilde{x}$ consistenti con i dati, vengono scelti quelli che presentano sequenze della base sparsa con *norma 1 minima*.

L'utilizzo della norma 1 per promuovere la sparsità viene effettuato da diversi decenni. Una delle sue prime applicazioni è stata la *sismica a riflessione*, in cui veniva ricercata una funzione di riflessione sparsa attraverso una banda limitata di dati. Comunque questo non è l'unico metodo per ricavare una *sparse recovery*; esistono diversi algoritmi, tra cui gli algoritmi greedy.

Una prima conclusione è che nel caso in cui f sia sufficientemente sparsa, la ricostruzione attraverso la norma 1 è probabilmente esatta.

Teorema 3.1 *Fissato $f \in \mathbb{R}^n$ e supponendo che la sequenza di coefficienti x di f nella base Ψ sia sparsa. Si scelgano le m misure nel dominio Φ in maniera casuale. Se*

$$m \geq C \mu^2(\Phi, \Psi) S \log n \quad (3.6)$$

Per una costante positiva C , la soluzione di (3.6) è esatta con probabilità quasi assoluta.

Dove il numero minimo di misurazioni, che soddisfa il teorema sopraindicato, è dato da

$$m \geq CS \log n,$$

in questo caso la correlazione tra gli oggetti di Φ e di Ψ , risulta minima, infatti

$$\mu(\Phi, \Psi) = 1$$

Si possono effettuare 3 considerazioni:

1. Il ruolo della coerenza è completamente trasparente; più la coerenza risulta bassa, meno campioni risulteranno necessari.
2. Basta campionare un set di m campioni per evitare di ottenere perdita di informazione. Se $\mu(\Phi, \Psi) \cong 1$, allora l'ordine di $S \log n$ campionature risulta sufficiente.
3. Il segnale f può essere esattamente ricostruito a partire dai dati acquisiti senza conoscenze sul numero di campioni di f non nulli, sulla loro posizione, oppure sulla loro ampiezza. Se il segnale appare sufficientemente sparso, la ricostruzione sarà esatta.

È ora il momento di discutere riguardo al ruolo della probabilità in tutto questo. Il punto chiave è che per ottenere i migliori risultati possibili bisogna far ricorso a una descrizione di tipo statistico per poter comparare tutte le misure di dimensione m . Esistono segnali sparsi speciali che nel dominio Φ hanno la maggior parte dei campioni nulli. In altre parole si possono identificare segnali sparsi f e sottoinsiemi di dimensione n (più precisamente $n-S$) per cui $y_k = \langle f_k, \varphi_k \rangle = 0$ per tutti i $k \in M$.

Come si può intuire, non esiste un algoritmo che consente di ricostruire un segnale partendo da una percentuale troppo elevata di elementi posti a zero.

Dall'altro lato, il Teorema 1.1 garantisce che la frazione di dati per cui non è possibile un'esatta ricostruzione risulta trascurabile (una potenza di n con modulo grande e segno negativo). Quindi bisogna tollerare una probabilità di fallimenti estremamente piccola, nella pratica è approssimabile a zero, se la dimensione del campionamento è sufficientemente grande.

Studi riguardanti segnali speciali hanno mostrato che si ha bisogno di un numero minimo di campioni $\mu^2 S \log n$. Con minor numero di campioni la probabilità di perdere informazione è troppo elevata e qualsiasi metodo di ricostruzione risulta inutilizzabile.

In conclusione si consideri un'esempio in cui si hanno circa 25,000 coefficienti wavelet diversi da zero. Si acquisiscono le informazioni effettuando circa 96,000 misure incoerenti e si risolve la 3.6.

La minimizzazione con la norma 1 risulta perfetta, infatti $f^* = f$. Questo esempio mostra come un numero di campioni pari circa a quattro volte il livello di sparsità sembri bastare: molti altri studi hanno riportato risultati simili utilizzando questa indicazione puramente sperimentale.

3.5 Compressive Sampling Robusto

Si possono ricostruire segnali con poche misure, però per essere realmente efficace il CS deve riuscire a gestire situazioni in cui i segnali sparsi siano accompagnati dal rumore, essenzialmente per due motivi.

Primo, l'obiettivo d'interesse non sono esattamente segnali sparsi, bensì segnali approssimativamente sparsi. Il problema in questo caso è capire dove risulta possibile o non ottenere accurate ricostruzioni da misurazioni altamente sottocampionate. Secondo, in qualsiasi misurazione reale dei dati ci sarà un piccolo errore introdotto dal rumore a causa della non elevata precisione dei dispositivi.

Dunque è necessario che CS sia robusto di fronte a queste complicazioni, più precisamente: deve garantire che piccole perturbazioni nei dati rimangano piccole anche in fase di ricostruzione.

Considerando il problema di ricostruire il vettore $x \in \mathbb{R}^n$ dai dati:

$$y = Ax + z \quad (3.7)$$

Dove A è una matrice $m \times n$ chiamata "matrice di sensing", questa è la matrice di acquisizione, e z rappresenta l'errore stocastico o deterministico sconosciuto. La (3.7), grazie alle precedenti considerazioni, può essere scritta nel modo seguente:

$$y = R\Phi^T\Psi x + z = R\Phi^T f + z$$

Dato che $f = \Psi x$ e $y = R\Phi^T f$ (R è la matrice $m \times n$ che estrae le coordinate di M , cioè contiene solamente gli elementi **non nulli**), si può scrivere che $y = Ax$, dove $A = R\Phi^T\Psi$. Così facendo si può utilizzare il modello astratto (3.7) tenendo in mente che x è la sequenza dei coefficienti dell'oggetto d'interesse nella base appropriata.

Quindi, dalle conclusioni finora espresse, si può dire che il segnale campionato $\tilde{y} = R\Phi^T\Psi$ va considerato come $y \in \mathbb{R}^m$, in quanto R contiene esclusivamente elementi non nulli del segnale sparso f (in questo caso $m \ll n$).

3.6 Isometrie Ristrette

In questo paragrafo verrà introdotta una nozione molto utile per quanto riguarda lo studio generale della robustezza di CS: *La Proprietà di Isometria Ristretta* (RIP).

Definizione 3.2 Per qualsiasi intero $S = 1, 2, \dots$, si definisce **costante isometrica** δ_s di una matrice A il più piccolo numero tale che

$$(1 - \delta_s)\|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s)\|x\|_2^2 \quad (3.8)$$

verifichi la condizione per tutti i vettori x S -sparsi.

Si dirà che la matrice A rispetta la RIP di ordine S se δ_s non è troppo vicina a uno.

Quando la proprietà è rispettata, A conserva approssimativamente la norma euclidea del segnale S -sparso, questo implica che i vettori S -sparsi non possono essere nello spazio nullo di A (è come dire che questi vettori non possono essere ricostruiti). Tutto ciò sta a significare che tutti i sottoinsiemi di S colonne prese da A sono quasi ortogonali (le colonne di A non possono essere esattamente ortogonali perché ci sono più righe che colonne).

Per osservare la connessione tra la RIP e CS si immagini di voler acquisire segnali S -sparsi con A , si supponga che δ_{2S} sia sufficientemente minore di uno. Questo implica che tutte le distanze fra segnali S -sparsi devono mantenersi nello spazio di misura.

Quindi:

$$(1 - \delta_{2S})\|x_1 - x_2\|_2^2 \leq \|Ax_1 - Ax_2\|_2^2 \leq (1 + \delta_{2S})\|x_1 - x_2\|_2^2$$

va rispettata per i vettori S -sparsi x_1 e x_2 . Questo significa che:

$$Ax = A(x_1 - x_2) = Ax_1 - Ax_2 = 0 \longrightarrow Ax_1 = Ax_2$$

nel caso in cui le colonne di A non siano ortogonali e quindi risulterebbe impossibile distinguere i due vettori

Come dimostrato di seguito questa condizione garantisce l'esistenza di algoritmi robusti per discriminare segnali S -sparsi a seconda delle misurazioni compresse.

3.7 Ricostruzione Generale del Segnale a Partire da Dati Sottocampionati

Se la RIP risulta verificata, allora la minimizzazione in *norma 1* fornisce un'accurata ricostruzione:

$$\min_{\tilde{x} \in \mathbb{R}^n} \|\tilde{x}\|_1 \quad \text{soggetta a } A\tilde{x} = y = (Ax) \quad (3.9)$$

Teorema 3.2 *Assumendo che $\delta_{2S} < \sqrt{2} - 1$ la soluzione x^* a (3.9) rispetta*

$$\|x^* - x\|_2 \leq C_0 \|x - x_S\|_1 / \sqrt{S} \quad \text{e} \quad \|x^* - x\|_1 \leq C_0 \|x - x_S\|_1 \quad (3.10)$$

per una costante C_0 dove x_S è il vettore x con la maggior parte degli S componenti posti a zero.

Le conclusioni del Teorema 1.2 sono più forti rispetto a quelle del Teorema 1.1. Se x è S -sparso, allora $x = x_S$ e perciò, la ricostruzione è esatta. Questo teorema è valido per qualsiasi tipo di segnale. Se x non è S -sparso, allora (3.10) afferma che la qualità del segnale ricostruito è tanto buona quanto lo sarebbe stata se si fosse conosciuta, a priori, la posizione degli S valori maggiori di x (valori più importanti), come se questi fossero misurati direttamente.

In altre parole, la ricostruzione risulta buona se si ha una piena e perfetta conoscenza di x , estraendo le S informazioni più importanti.

Un'altra importante differenza tra i due teoremi è che quest'ultimo risulta di tipo deterministico, al contrario del primo, che risulta probabilistico.

Ciò significa che se si trova una matrice di acquisizione A che obbedisce alle ipotesi del teorema, questa si può utilizzare con la certezza di ricostruire esattamente tutti i vettori S -sparsi, o comunque gli S campioni più grandi di qualsiasi vettore senza possibilità di fallimento.

3.8 Ricostruzione Robusta di Segnali Affetti da Rumore

Considerando dati affetti da rumore come in 3.7 e utilizzando la minimizzazione con la norma 1 e facendo rilassare i vincoli lineari per la ricostruzione:

$$\min \|\tilde{x}\|_1 \quad \text{soggetto a} \quad \|A\tilde{x} - y\|_2 \leq \varepsilon. \quad (3.11)$$

Dove ε indica il valore limite di rumore presente nei dati. Infatti:

$$\|z\|_2 \leq \varepsilon$$

dove, appunto, z è il termine di rumore, avente energia limitata ε .

Il problema (3.11) è spesso chiamato *LASSO* (acronimo di *least absolute shrinkage and selection operator*); questo è un problema convesso e può essere risolto efficientemente.

Teorema 3.3 *Assumendo che $\delta_{2S} < \sqrt{2} - 1$, allora la soluzione x^* di (3.11) obbedisce a*

$$\|x^* - x\|_2 \leq C_0 \|x - x_S\|_1 / \sqrt{S} + C_1 \varepsilon \quad (3.12)$$

per certi valori di C_0 e C_1 .

Questo concetto è facile da spiegare: l'errore di ricostruzione dipende da due termini. Il primo è legato all'errore presente nei dati, non è legato alla presenza del rumore, mentre il secondo è proporzionale al livello di rumore. Inoltre, le costanti C_0 e C_1 sono tipicamente piccole. Da dati sperimentali è stato stabilito che CS è un meccanismo di acquisizione pratico e robusto, che lavora con tutti i tipi di segnali non necessariamente sparsi e gestisce al meglio i segnali rumorosi.

3.9 Acquisizione Casuale

Ciò che ora bisogna trovare è la matrice di acquisizione con la proprietà che i suoi vettori colonna presi da un sottoinsieme arbitrario siano quasi ortogonali; più grande risulta il sottoinsieme, meglio è.

Consideriamo le seguenti matrici di acquisizione A costruite attraverso le seguenti modalità:

1. Campionando n colonne uniformemente e casualmente dalla sfera unitaria di \mathbb{R}^n
2. Campionando a valori indipendenti e identicamente distribuiti dalla distribuzione normale con media nulla e varianza $1/m$
3. Campionando a valori indipendenti e identicamente distribuiti secondo una distribuzione simmetrica di Bernoulli o altre distribuzioni subgaussiane.

Con probabilità quasi certa, tutte queste matrici rispettano la RIP, dato che

$$m \geq CS \log(n/S), \quad (3.13)$$

dove C è una costante dipendente dalla matrice scelta. In tutti questi esempi, la probabilità di campionare una matrice che non rispetta RIP quando (3.13) è soddisfatta, è esponenzialmente piccola in m . Inoltre, non esistono matrici di misura o algoritmi di ricostruzione che diano le conclusioni del Teorema 1.2 con un numero inferiori di campioni rispetto a (3.13). Quindi l'utilizzo di matrici casuali con minimizzazione in norma 1 risulta una strategia quasi ottimale di acquisizione.

Per verificare la RIP per una coppia di basi ortogonali, con $\mathbf{A} = \mathbf{R}\Phi\Psi$ dove \mathbf{R} estrae uniformemente con casualità m coordinate, è sufficiente che:

$$m \geq CS(\log n)^4 \quad (3.14)$$

affinchè la proprietà regga con elevata probabilità. Se si vuole una probabilità di fallimento minore di $O(n^{-\beta})$ per $\beta < 0$, allora il miglior esponente in (3.14) è cinque al posto di quattro.

Questo prova che si possono ricostruire in modo stabile ed accurato tutti i segnali quasi sparsi da dati estremamente sottocampionati in un dominio incoerente. Inoltre, RIP è valida anche per matrici di acquisizione $A = \Phi\Psi$, dove Ψ è una base ortogonale arbitraria e Φ è una matrice di misure $m \times n$ ricavata casualmente da una distribuzione adatta. Se si fissa Ψ e si sceglie Φ in base alle modalità di acquisizione, precedentemente descritte, allora con

una probabilità quasi assoluta, la matrice $A = \Phi\Psi$ rispetta la RIP attraverso la (3.13).

Queste matrici di misurazioni casuali Φ sono in un certo senso universali; non è necessario conoscere la sparsità della base quando viene scelto il sistema di misura.

3.10 Che Cos'è il Compressive Sampling?

L'acquisizione di dati tipicamente si svolge nel modo seguente: una quantità massiccia di dati viene acquisita per poi essere, in gran parte, scartata durante la compressione per facilitare la trasmissione e conservazione dei dati stessi. Tecnicamente si acquisisce un array di pixel f ad alta risoluzione, poi si crea il set completo dei coefficienti della trasformata, quindi si codificano i coefficienti più grandi e gli altri vengono scartati, arrivando in fine, a f_S . Questo processo di acquisizione completa, seguito dalla compressione, è estremamente dispendioso.

CS opera in maniera differente, *“come se fosse possibile acquisire direttamente solo l'informazione importante sull'oggetto d'interesse”*. Utilizzando solamente $O(S \log(n/S))$ proiezioni casuali si hanno abbastanza informazioni per ricostituire il segnale con un'accuratezza almeno grande quanto quella fornita da f_S (la miglior rappresentazione compressa). In altre parole, i protocolli di misura CS traducono dati analogici in un formato compresso così da ottenere risoluzione massima utilizzando pochi sensori. Tutto ciò che occorre dopo la fase di acquisizione dati è la *“decompressione”* dei dati misurati (ovvero la ricostruzione del segnale misurato).

Capitolo 4

Esempi e Applicazioni

In questo capitolo verranno fatti esempi sul CS tramite `cvx`, un software utilizzabile in Matlab che consente di effettuare l'ottimizzazione convessa, in particolare verrà utilizzato per effettuare la minimizzazione con la norma 1 e la norma 2.

Oltre agli esempi in Matlab, verranno trattati casi reali in cui il CS viene utilizzato.

4.1 Esempi

In queste prime figure viene rappresentata la simulazione effettuata con norma 1 e successivamente con norma 2. Per velocizzare i calcoli è stata scelta una matrice di acquisizione $A = m \times n$, in cui n rappresenta la taglia del segnale ed è stata posta pari a 512, mentre il numero di misurazioni effettuate è pari a 128. Il vettore dei coefficienti sparsi presenterà 8 coefficienti diversi da zero.

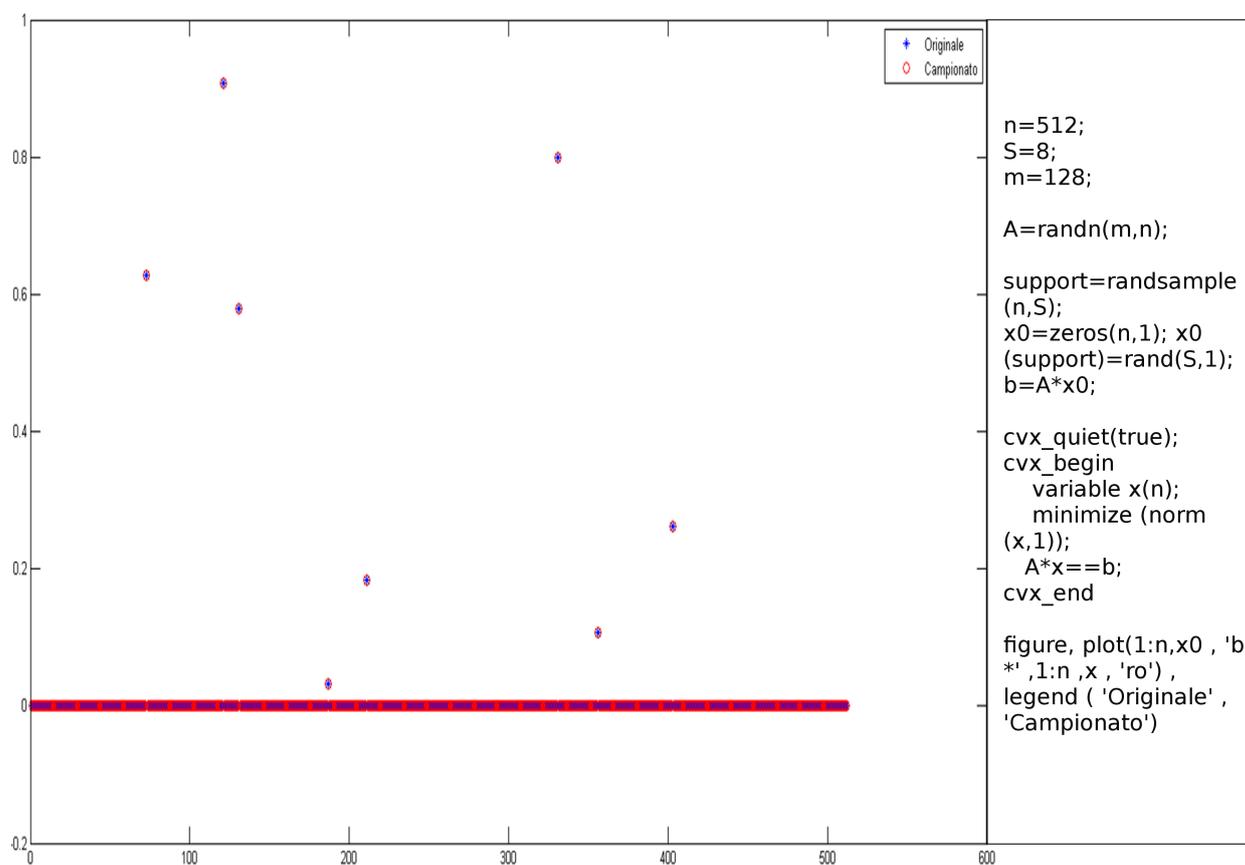


Figura 4.1: Esempio con norma 1

Come è possibile osservare dal grafico, la simulazione ha avuto buon esito: è stata rispettata la (3.6) e la norma 1, grazie alle sue caratteristiche, ha riprodotto perfettamente i dati da campionare.

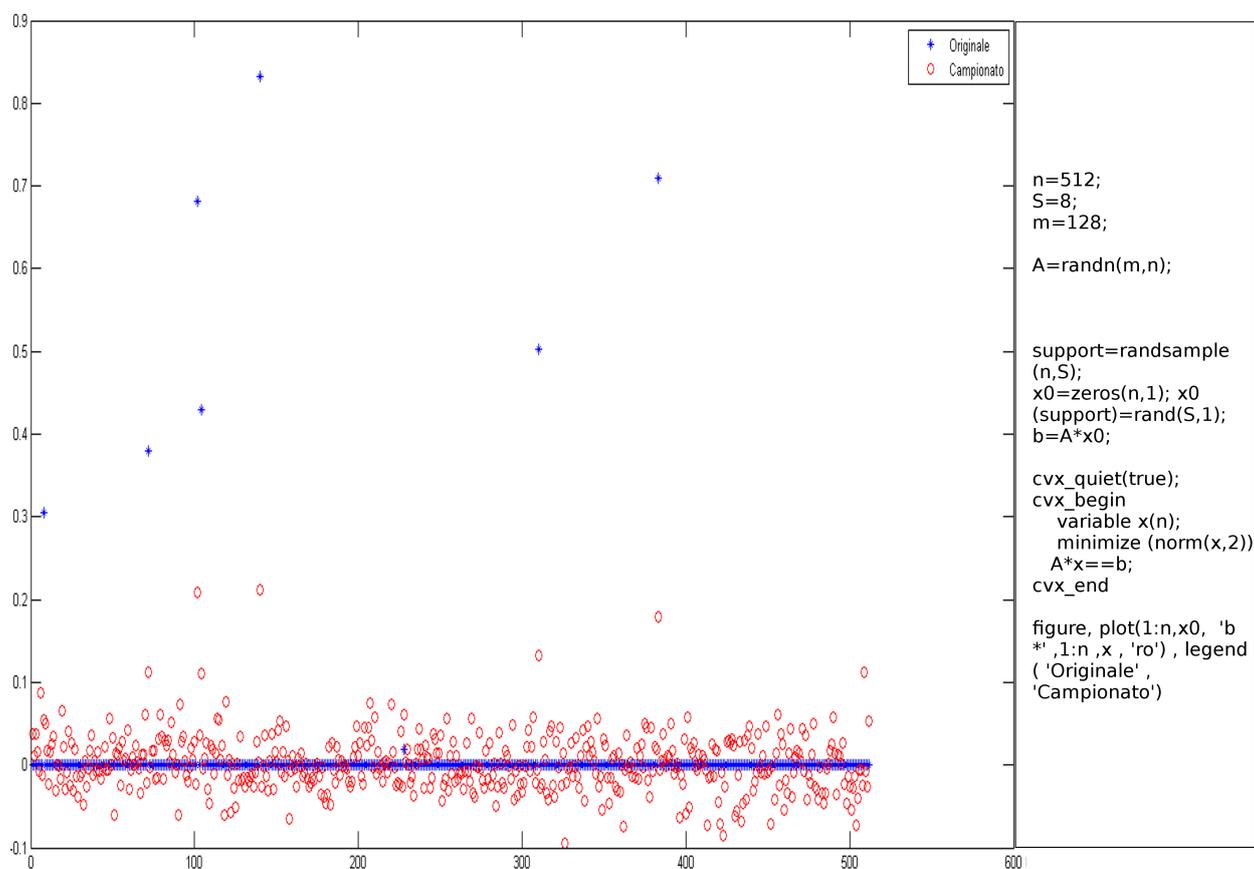


Figura 4.2: Esempio con norma 2

In questo caso, la simulazione non è andata a buon fine, a parità di sistema analizzato. Da ciò si può osservare che la norma 2, come descritto in precedenza non risulta efficace in questa tipologia di campionamento. Quindi è stato dimostrato che risulta corretta l'affermazione detta in precedenza, ovvero: *nel caso in cui f sia sufficientemente sparsa, la ricostruzione attraverso la norma 1 è probabilmente esatta.*

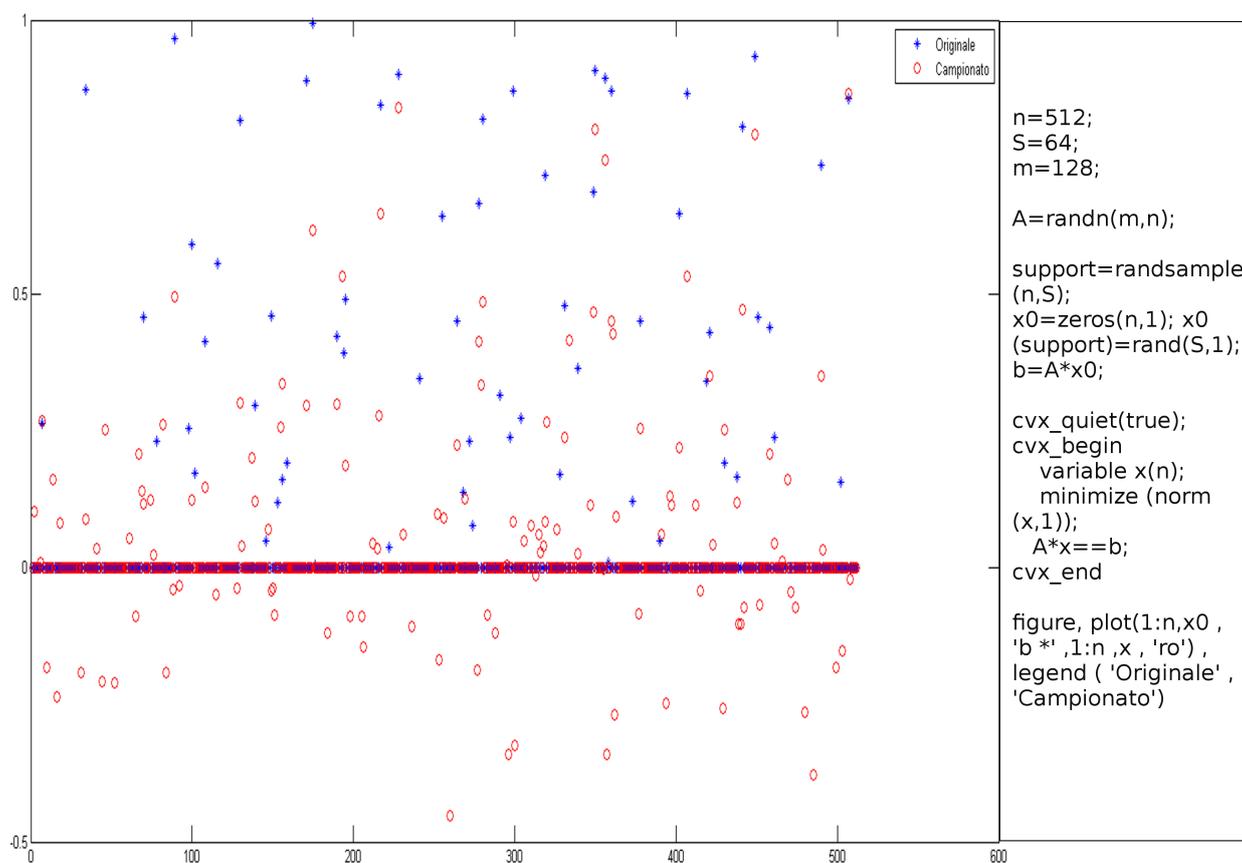
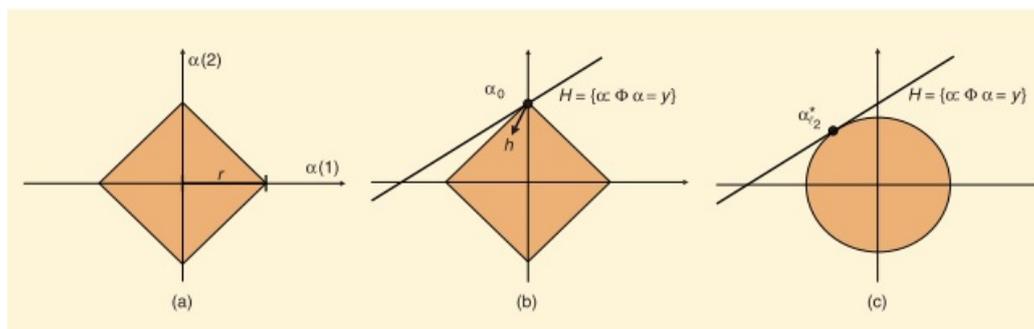


Figura 4.3: Minimizzazione con norma 1 nel caso di vettore poco sparso

Come si può osservare in questo caso, con f poco sparsa non si ottiene una fedele ricostruzione del segnale.

Da questi esempi, oltre agli effetti dovuti alla sparsità, si possono osservare differenze dovute al tipo di norma utilizzata per la minimizzazione. Per poter



spiegare questa differenza, si consideri la figura 3.4, dove (a) rappresenta la sfera l_1 avente un certo raggio r . Questa ha caratteristiche anisotrope.

L'immagine (b) rappresenta la ricostruzione mediante l_1 : il punto α_0 rappresenta un vettore sparso che effettua una sola misura. La riga denominata H è l'insieme di tutte le α che condividono lo stesso valore di misura. Lo scopo della minimizzazione con norma 1 è quello di individuare il punto appartenente a questa linea con norma 1 più piccola.

Considerando gli effetti della norma 2, si osservi l'immagine (c), dove la sfera anisotropica viene sostituita con una sfera avente caratteristiche isotropiche, in questo caso il punto di intersezione H con l'espansione della norma 2, non è affatto sparsa, e questo effetto si nota via via che aumentano le dimensioni.

Si può dire in conclusione che, nonostante la similitudine delle definizioni della norma 1 e della norma 2, si hanno 2 strumenti completamente diversi.

4.2 Single-Pixel Camera

L'architettura di questa fotocamera è basata principalmente su un computer ottico (può essere formato da un DMD, acronimo di dispositivo digitale a microspecchi, due lenti, un rivelatore di fotoni e un convertitore analogico digitale) che calcola le misure lineari casuali della scena osservata. L'immagine viene ricostruita o processata attraverso le misure di un computer. La struttura della fotocamera riduce le dimensioni richieste, la complessità e il costo del vettore rivelatore di fotoni crolla fino ad una singola unità, che consente l'uso di rivelatori inutilizzabili in una fotocamera digitale convenzionale. Le misure CS casuali permettono anche un compromesso tra spazio e tempo durante l'acquisizione dell'immagine.

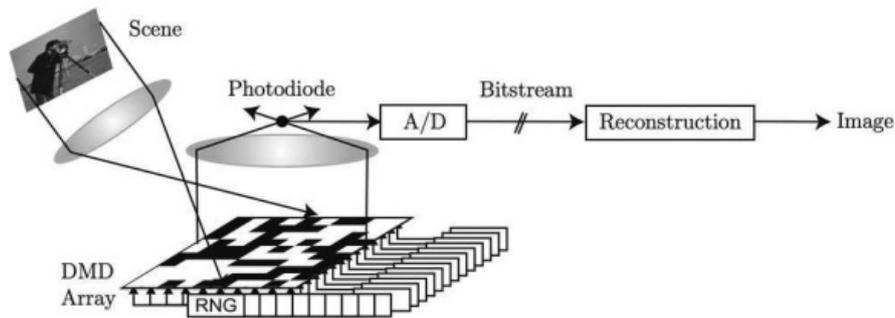


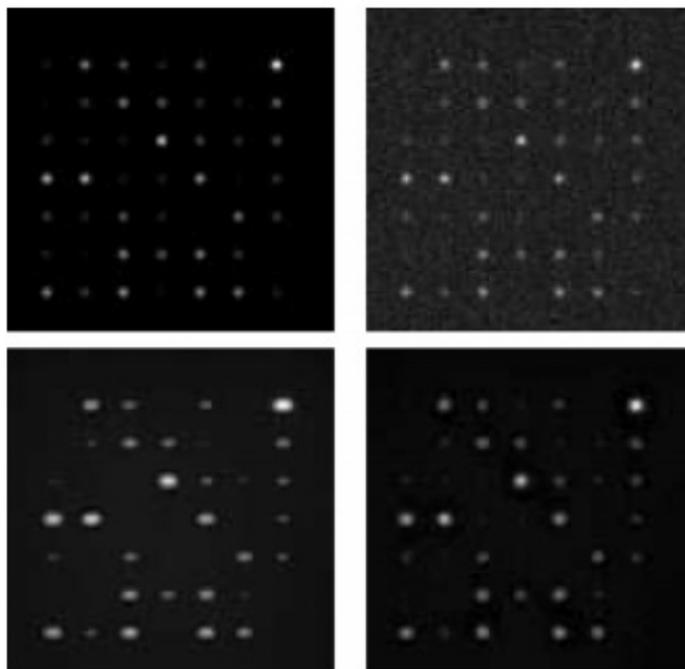
Figura 4.4: Struttura della fotocamera

Osservando la Figura 3.5 è possibile analizzare la struttura di questa fotocamera, ipotizzando di utilizzare le ipotesi valide per il CS è possibile comprendere il suo funzionamento: usando specchi riflettenti e non riflettenti (rappresentati in Figura 3.5 da zone chiare e zone scure) l'immagine viene ricevuta attraverso dei campionamenti casuali e con massima incoerenza rispetto alla base di campionamento. Le maschere vengono applicate per ogni fase di acquisizione per distinguere specchi riflettenti da specchi non riflettenti. Questo mezzo nella pratica per ovvi motivi non risulta conveniente, però ha un'ampia valenza teorica in quanto dimostra che è possibile ricostruire un'immagine velocemente e a basso costo.

4.3 Astronomia

Di anno in anno, la quantità dei dati astronomici aumenta a un ritmo sempre crescente. Le dimensioni degli array digitali continuano a crescere, a causa delle sempre più grandi quantità di dati dovute alla ricerca astronomica in periodi di tempo sempre più brevi. Come risultato, la comunità astronomica si è trovata di fronte alla necessità di dover adottare tecniche di compressione per immagazzinare i dati. Il caso del satellite Herschel, lanciato tra il 2008-2009 con la missione di studiare la formazione delle galassie attraverso il più grande telescopio spaziale mai costruito, è a stretto contatto con il problema di compressione in quanto i dati fotometrici dovevano essere compressi di un fattore 6 per essere inviati. Poiché il carico della CPU doveva essere piccolo, usare metodi convenzionali di compressione non era possibile.

Da questo punto di vista, il CS ha dato un enorme contributo in termini di compressione dati con grande efficienza, mantendo una bassa distorsione.



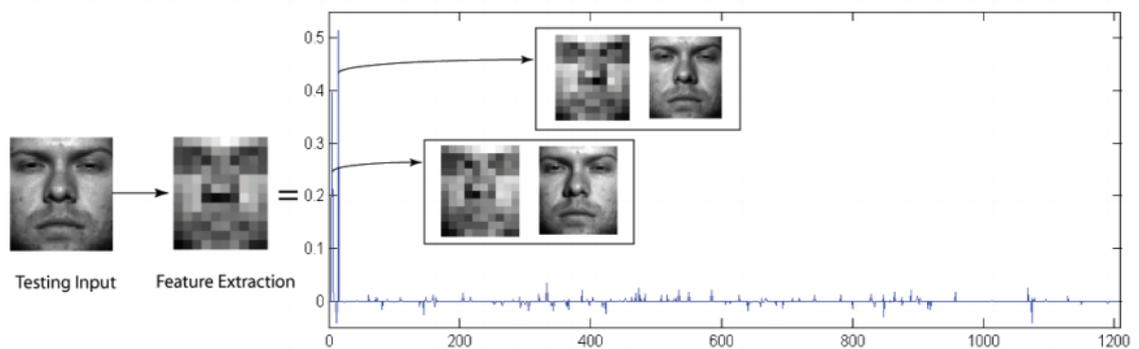
Nella figura è possibile osservare l'effetto del CS sull'immagine acquisita.

L'immagine in alto a sinistra è l'immagine originale, le immagini in alto a destra e in basso a sinistra sono, rispettivamente, la prima una delle sei immagini ricevute affette da rumore gaussiano bianco e l'altra è la media di queste sei immagini acquisite.

L'immagine in basso a destra, rappresenta la ricostruzione dell'immagine attraverso il Compressive Sensing, si può osservare come l'immagine originale è stata ben riprodotta.

4.4 Face Recognition

Il Compressed Sensing ha dimostrato un enorme potenziale per problemi come i pattern di rappresentazione e riconoscimento dei volti. L'approccio sfrutta la correlazione delle immagini dello stesso soggetto attraverso modelli sparsi, comuni nella letteratura degli algoritmi che consentono l'estrazione di caratteristiche e di riconoscimento dei volti.



Capitolo 5

Conclusioni

Il Compressive Sampling stabilisce, attualmente, una nuova teoria del campionamento, che stabilisce fondamentali e affascinanti connessioni con diversi campi della matematica, come probabilità, geometria degli spazi di Banach, analisi delle armoniche, teoria della computabilità e complessità basata sull'informazione. Il collegamento tra l'ottimizzazione convessa e lo sviluppo di metodi numerici estremamente efficienti e robusti rendono il Compressive Sensing un concetto utilizzabile per una vasta branca delle scienze naturali e delle applicazioni ingegneristiche, in particolare, nella acquisizione ed elaborazione dei segnali e delle immagini. Si può prevedere che il Compressive Sensing produrrà importanti effetti in diverse branche delle scienze e della tecnologia.

Recenti sviluppi riguardanti, in particolare, la ricostruzione di matrici con basso rango attraverso la minimizzazione con norma nucleare (o n -norma di Ky Fan), suggerisce nuove possibili estensioni del Compressive Sensing a strutture più complesse. Inoltre nuove sfide stanno emergendo ora nell'analisi numerica e nella simulazione, dove i problemi ad elevate dimensioni, come ad esempio le equazioni differenziali parziali stocastiche nella finanza e il calcolo della struttura degli elettroni nella chimica e nella biochimica, sono diventati la nuova frontiera. In questo contesto, come altre efficienti forme di approssimazione come space grid e metodi di tensor product, il compressive sensing è un promettente concetto, che rischia di far fronte alla *maledizione della dimensionalità*. In particolare, ulteriori sviluppi sistematici di adattività in presenza di differenti criteri, algoritmi randomizzati, un crescente ruolo di aspetti combinatori degli algoritmi sottostanti, sono un esempio di futuri sviluppi, che sono ispirati dai risultati del Compressive Sensing.

Bibliografia

- [1] An introduction to compressive sampling, Candès, Emmanuel J and Wakin, Michael B, *Signal Processing Magazine, IEEE*, 25, 02, 21–30, 2008, IEEE
- [2] Compressive sensing, Fornasier, Massimo and Rauhut, Holger, *Handbook of mathematical methods in imaging*, 187–228, 2011, Springer
- [3] An introduction to wavelets, Chui, Charles K, 1, 2014, Academic press
- [4] An introduction to wavelets, Graps, Amara, *Computational Science & Engineering, IEEE*, 2, 02, 50–61, 1995, IEEE
- [5] Imaging via compressive sampling [introduction to compressive sampling and recovery via convex programming], Romberg, Justin, *IEEE Signal Processing Magazine*, 25, 02, 14–20, 2008
- [6] Single-pixel imaging via compressive sampling, Duarte, Marco F and Davenport, Mark A and Takhar, Dharmpal and Laska, Jason N and Sun, Ting and Kelly, Kevin E and Baraniuk, Richard G and others, *IEEE Signal Processing Magazine*, 25, 02, 83, 2008
- [7] *Algoritmi numerici*, Rodriguez Giuseppe, 2008, Pitagora
- [8] A really friendly guide to wavelets, Valens, Clemens, accessible at: <http://perso.wanadoo.fr/polyvalens/clemens/wavelets/wavelets.html>, 1999
- [9] Compressed sensing in astronomy, Bobin, Jérôme and Starck, Jean-Luc and Ottensamer, Roland, *Selected Topics in Signal Processing, IEEE Journal of*, 2, 05, 718–726, 2008, IEEE
- [10] cvx user' guide for cvx version 1.1, Grant, Michael and Boyd, Stephen and Ye, Yinyu, accessible at: <http://www.stanford.edu/~boyd/index.html>, 2014

- [11] A compressive sensing approach for expression-invariant face recognition, Nagesh, Pradeep and Li, Baoxin, Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on, 1518–1525, 2009, IEEE