

Compressed sensing

Alla ricerca del pixel perduto

Di Giovanni Naldi

Una conseguenza evidente dell'evoluzione dei mezzi di comunicazione è la presenza pervasiva nella nostra vita quotidiana di immagini e filmati e dei corrispondenti strumenti di acquisizione come telecamere e fotocamere, tutte rigorosamente digitali.

Anche i produttori di telefoni cellulari si sfidano a colpi di *megapixel* o di capacità di memoria per la conservazione di immagini o sequenze. Forse non è la presenza ossessiva tipica delle atmosfere dei film tratti dai libri di Philip K. Dick come *Minority Report* o *Blade Runner*, anche se a volte percepiamo un certo abuso.

Forse meno evidente e meno noto è il fatto che la matematica sia diventata uno strumento indispensabile per il trattamento, l'elaborazione delle immagini: dal filtraggio ai metodi efficienti per la memorizzazione, dall'estrazione di particolari caratteristiche alla correzione di artefatti e ad altro ancora.

Un mondo digitale

La tecnologia dei sistemi di acquisizione evolve velocemente (camere digitali ad alta risoluzione, convertitori analogico-digitale, DNA *microarray*, ...) e singoli sensori vengono messi in rete (sistemi distribuiti *wireless* di sensori, archivi di immagini, ...) . Anche le modalità di acquisizione e i segnali fisici corrispondenti sono tra i più vari: segnali acustici, segnali radio, segnali video, raggi X, raggi gamma, segnali infrarossi e tanti altri ancora. Come acquisire, gestire, memorizzare, elaborare in modo efficiente questo diluvio di dati? Ci occuperemo qui di un paradigma recente noto come *compressive sensing* (o *compressive sampling*) che consente, in generale, di ricostruire un segnale (immagine) attraverso un numero decisamente minore di misure (campionamenti) rispetto ai metodi tradizionali. Questa possibilità apre la strada alle applicazioni più svariate oltre che essere l'origine di problemi matematici interessanti.

Per fare un esempio, tra i tanti possibili, facciamo un salto nello spazio e nel tempo. Spostiamoci nella primavera del 2009 con un gruppo di medici in servizio presso il Lucile Packard Children's Hospital della Stanford University. Occorreva provvedere, attraverso immagini a risonanza magnetica (MRI) ad alta risoluzione, all'analisi delle condizioni di un piccolo paziente di due anni. Il bambino aveva ricevuto, dieci mesi prima, una porzione di fegato da un donatore. Qualcosa purtroppo non andava, c'era la possibilità che uno o entrambi i dotti biliari del fegato si fossero bloccati. Il radiologo pediatrico Shreyas Vasanawala sperava che la MRI potesse fornire la risposta, ma per avere qualche speranza di successo, era necessaria, anzi indispensabile, un'immagine di scansione ad alta risoluzione. Ma come si poteva tenere perfettamente immobile un bimbo di due anni? Anche un solo respiro e l'immagine sarebbe risultata offuscata. Ci sarebbe voluta un'anestesia sufficiente per fermare la respirazione (almeno due minuti) che per il piccolo avrebbe potuto essere estremamente pericolosa. Allora, Vasanawala e uno dei suoi colleghi, l'ingegnere elettronico Michael Lustig, sperimentarono un nuovo

algoritmo di rilevamento basato sulla tecnica del *compressed sensing* e... furono necessari solo 40 secondi per raccogliere dati sufficienti a produrre immagini cristalline del fegato del bambino.

Sinteticamente, potremmo dire che il *compressed sensing* è una nuova teoria di campionamento (o di strategia di misurazione) che cerca di catturare gli aspetti “essenziali” di un oggetto attraverso il minor numero possibile di misure.

Supponiamo che l’oggetto in questione sia un segnale, per esempio un suono, e che sia rappresentabile con una funzione reale $f(t)$, dove t potrebbe rappresentare il tempo. Per acquisire informazioni sul segnale stesso facciamo delle misure.

In Figura 1 sono schematizzati due differenti tipi di misura (campionamento): nel primo caso si valuta semplicemente il segnale f in un punto t_k e il valore della misurazione è $y_k = f(t_k)$. Nel secondo caso la situazione è apparentemente più complessa ma l’idea è piuttosto semplice: decomponiamo il segnale rispetto a un insieme di “funzioni elementari”, ovvero rispetto a funzioni sinusoidali elementari (funzioni *seno* e *coseno* a diversa frequenza, $\sin(\omega t)$, $\cos(\omega t)$). In altri termini, per chi ha più confidenza con problemi di questo tipo, stiamo considerando l’analisi di Fourier (discreta) per il segnale in esame. Il campionamento convenzionale di un segnale, utilizzato nella maggior parte degli attuali dispositivi di conversione analogico/digitale, consiste nella memorizzazione dei valori, per esempio dei valori medi, $y_k = f(t_k)$, come nel primo esempio di Figura 1, per una griglia di campionamento uniforme, cioè per una griglia in cui il cosiddetto *intervallo di campionamento* $\Delta t = (t_{k+1} - t_k)$ è costante per ogni k .

Esiste un teorema importante che mette in relazione il contenuto di un segnale campionato con la *frequenza di campionamento* (cioè il rapporto $F_s = 1/\Delta t$) e con le componenti di frequenza minima e massima presenti nel segnale originale, definendo così la minima frequenza necessaria per campionare un segnale senza perdere informazioni. Si tratta del “Teorema di campionamento di Nyquist-Shannon”, secondo il quale il segnale $f(t)$ se è a banda limitata, cioè se non ha nessuna componente con una frequenza maggiore di una frequenza massima F_M , allora può essere ricostruito esattamente partendo dai suoi campioni $f(t_k)$ se si ha $F_s \geq 2 F_M$.

Volendo quindi utilizzare nella pratica campionamenti con griglie uniformi per segnali con un contenuto in frequenza a “banda larga” siamo costretti ad un passo di campionamento molto piccolo e quindi siamo costretti alla valutazione e alla memorizzazione di molte misurazioni.

Nel caso discreto abbiamo, solitamente, che il segnale originale è rappresentato da una sequenza di N numeri reali, cioè da un vettore reale \mathbf{x} N -dimensionale, $\mathbf{x} = (x_1, x_2, \dots, x_N)^T \in \mathbb{R}^N$ (nelle applicazioni N è “grande”, diciamo dell’ordine di 10^6 o 10^7) dove con T si intende il vettore “trasposto” ovvero, in questo caso, un vettore verticale invece che orizzontale.

Per acquisire il segnale (o per avere informazioni su di esso), faremo M misure: per interpretarle con i nostri strumenti di lavoro (matematico) consueti, ci immaginiamo che ognuna di esse (y_j) sia un numero reale ottenuto come il prodotto scalare euclideo in \mathbb{R}^N tra il vettore segnale \mathbf{x} e un opportuno vettore φ_j . Se costruiamo il vettore $\mathbf{y} = (y_1, y_2, \dots, y_M)$ delle misure, si osserva che \mathbf{y} si può scrivere come $\mathbf{y} = \Phi \mathbf{x}$ dove Φ indica la matrice di M righe e N colonne avente come j -esima riga il vettore φ_j . A questo punto la domanda è: possiamo ricostruire il segnale \mathbf{x} dalle misure \mathbf{y} fatte quando $M < N$ o, di maggior significato applicativo, quando $M \ll N$, cioè M è molto più piccolo di N ?

In generale il problema è mal posto perché equivale a risolvere un sistema lineare sotto determinato (cioè con troppo poche equazioni rispetto al numero delle incognite) e quindi, generalmente, ci sono infiniti candidati ad essere il segnale originale \mathbf{x} . I metodi del *compressed sensing* cercano di chiarire quando, invece, il problema può essere risolto producendo un solo segnale \mathbf{x} .

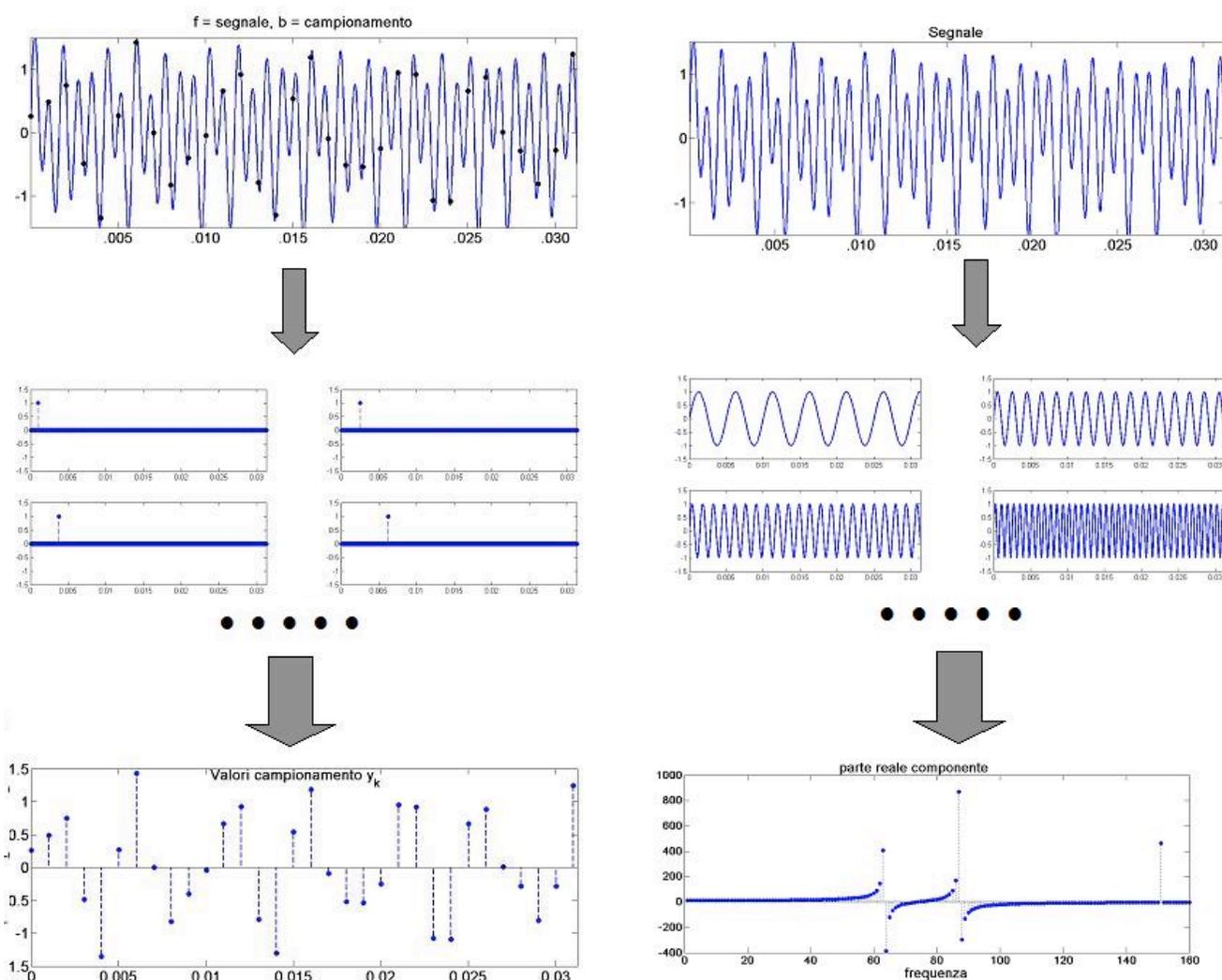


Figura 1. Due differenti metodi di misurazione, il primo (colonna di sinistra) considera la semplice valutazione puntuale $y_k=f(t_k)$; il secondo (colonna di destra) prevede invece di valutare le componenti rispetto alle funzioni seno e coseno per differenti valori di frequenza (analisi di Fourier).

Prima di descrivere il modo di procedere ricordiamo che un'immagine è pensata semplicemente come una funzione f in due variabili. Il valore $f(x,y)$ rappresenta una misura di qualche caratteristica (luminanza, colori, ...) in una "posizione" (x,y) di una scena (ovviamente si pensa di aver definito un sistema di riferimento per poter parlare senza ambiguità di posizione). Per esempio, pensiamo a una antica e suggestiva fotografia in bianco e nero (B/N): in ogni punto, la funzione f rappresenta la tonalità di grigio presente nel punto stesso. Il processo di formazione di un'immagine da analizzare attraverso un computer può essere molto diverso da quello essenzialmente di natura ottica delle varie tecniche di fotografia (per esempio, nelle immagini mediche spesso viene misurato il coefficiente di assorbimento dei tessuti rispetto a particolari radiazioni elettromagnetiche).

Torniamo all'esempio della fotografia B/N: il "piano immagine" I , ovvero il dominio della funzione f , è un sottoinsieme limitato del piano \mathbb{R}^2 (un rettangolo tipicamente), un codominio della f è un intervallo $[0, I_{\max}]$ dove I_{\max} è il valore massimo della variabile fisica legata all'immagine (per esempio, la

luminanza). Una funzione associata a un'immagine raramente può essere espressa mediante una semplice funzione analitica per cui, per effettuare su di essa elaborazioni mediante il calcolatore, è necessario che la funzione f venga campionata e quantizzata. Campionare un'immagine significa prendere sul piano immagine un numero finito di punti (il Teorema del campionamento che abbiamo citato è un tipico risultato per stabilire le condizioni sufficienti per non avere perdita di informazione o per non perderne "troppa"). Quantizzare significa prendere un numero finito di valori interi nell'intervallo $[0, I_{\max}]$.

Naturalmente, un campionamento e una quantizzazione non adatti o inadeguati comportano una grave perdita di informazione con relativo deterioramento dell'immagine stessa. In particolare il numero di livelli di un'eventuale quantizzazione deve essere tale da non far percepire "salti di livello di grigio" all'osservatore. Nella nostra immagine ciò può accadere se i livelli di grigio (numero di suddivisioni dell'intervallo $[0, I_{\max}]$), sono meno di 5. In genere si usano 256 livelli di grigio, ma per alcune applicazioni questo valore può cambiare: per esempio, in ambito medico le immagini NMR (Nuclear Magnetic Resonance) utilizzano 4096 livelli di grigio o più.

In Figura 2 si mostra un esempio di immagine digitale e la rappresentazione di una sua piccola parte come matrice di valori interi. Ogni valore intero rappresenta un certo livello di valore di grigio. Il fatto che l'immagine si possa rappresentare attraverso una matrice è conseguenza del fatto che la fotografia è stata suddivisa in un certo numero di celle elementari attraverso un numero opportuno numero di righe e di colonne.

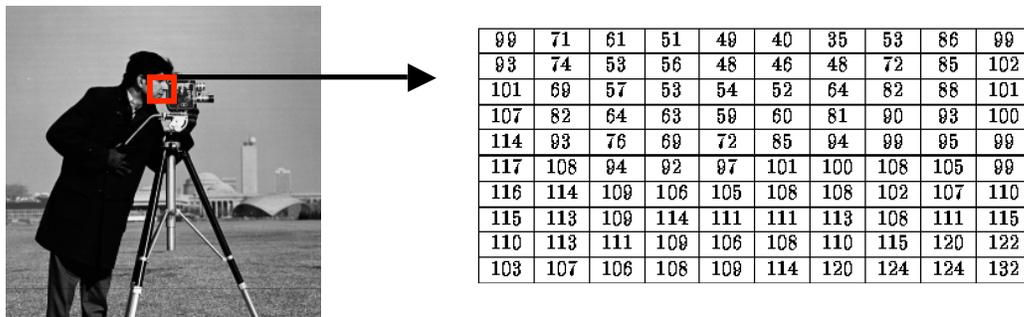


Figura 2. Un'immagine digitale a livelli di grigio: la tabella a destra rappresenta il valore dei singoli pixel dell'area incorniciata.

Cambiare numero di righe e di colonne implica cambiare l'area della cella elementare e quindi la risoluzione con cui si decompone l'immagine. I valori di grigio, cioè i valori nella tabella che rappresenta l'immagine, si possono pensare come medie locali dei valori di grigio nella cella (rappresentazione vicina a come operano, per esempio, le fotocamere digitali). Un'immagine digitale ottenuta come ora descritto è un'immagine *raster*. Le celle elementari dette si chiamano anche *pixel* (*picture elements*). In certe applicazioni l'immagine potrebbe non essere descritta *pixel per pixel* ma attraverso un insieme di primitive geometriche (linee, archi, rettangoli, ...), abbiamo allora un'immagine vettoriale.

In Figura 3 si mostra l'effetto del cambiamento di risoluzione spaziale e di quantizzazione per i valori di livello di grigio (l'immagine è "Lena", un'immagine famosa spesso utilizzata come immagine *test* in esperimenti di analisi delle immagini e *Computer Vision*).



Figura 3. Esempio di immagine digitale (A); con una diversa risoluzione, più grossolana, e con la conseguente formazione di quadretti più grandi (B); con diversa quantizzazione dei livelli di grigio con meno livelli (C).

Per le immagini a colori ricordiamo che possiamo pensarle come sovrapposizione di tre immagini monocromatiche (verde, rosso e blu): la funzione f diventa una funzione a valori vettoriali. Tale scelta dei colori fondamentali non è certo l'unica possibile ma la più utilizzata (codifica RGB). Ogni singola immagine monocromatica può essere pensata come un'immagine a livelli di valori di verde, di rosso e di blu e quindi formalmente può essere pensata come tre immagini del tipo livello di grigio (un codominio è un intervallo limitato). Possiamo quindi ricondurre anche le immagini al caso di un segnale rappresentato da uno (o più) vettori \mathbf{z} a $(N_1 \times N_2)$ componenti, dove N_1 ed N_2 rappresentano, rispettivamente, il numero di righe e di colonne di cui è composta l'immagine (o una sua componente monocromatica).

2. Il Problema.

I segnali (immagini, suoni, ...) possono essere rappresentati in vario modo come abbiamo detto poche righe sopra: per esempio attraverso valori puntuali (o medi) oppure attraverso la decomposizione in componenti elementari con diversa frequenza (analisi di Fourier) o altro ancora. Molte di queste codifiche sono a disposizione per una successiva elaborazione del segnale di partenza. Un caso tipico è la "compressione" dei segnali che tutti, sia pure inconsapevolmente, utilizziamo quando siamo alle prese con un file mp3, una immagine jpeg o un dvd. Esistono due tipi di compressione: senza perdita di informazione e con perdita di informazione. Nel primo caso, essenzialmente adatto per i testi, si vogliono utilizzare tecniche che permettano un risparmio in termini di occupazione di memoria ma che non facciano perdere alcuna informazione contenuta nel file in esame: si sfruttano ridondanze, ripetizioni, dizionari di codifica efficienti. Nel caso di compressione con perdita di informazione si manipola una codifica opportuna del segnale per individuare le componenti "importanti" tralasciando quei dettagli inessenziali per la fruizione del segnale stesso: si codifica quindi una sorta di sintesi efficace del segnale memorizzando solo le informazioni essenziali con un risparmio nell'uso delle risorse.

Dal punto di vista matematico, abbiamo un vettore, $\mathbf{f} \in \mathbf{R}^N$ (per esempio un'immagine con N pixel), che può essere rappresentato attraverso una base (ortonormale) $\Psi = \{\psi_1, \psi_2, \dots, \psi_N\}$,

$\mathbf{f}(t) = \sum_i c_i \psi_i(t)$, dove i c_i sono i valori delle componenti di \mathbf{f} lungo la "direzione"

$$\psi_i, c_i = (\mathbf{f}, \psi_i), i = 1, 2, \dots, N.$$

Quando il segnale è comprimibile, o consente una rappresentazione sparsa, è possibile scartare le componenti più piccole senza perdere una parte troppo grande del contenuto di informazione del segnale dato.

Formalmente, possiamo considerare il segnale compresso $f_s(t)$ ottenuto scegliendo K componenti più grandi nella rappresentazione di f . Come misurare l'errore commesso? Ricordiamo che possiamo stimare la lunghezza di un vettore attraverso la scelta di un'opportuna norma, per esempio la norma euclidea $\|x\|_2 = (\mathbf{x}, \mathbf{x})^{1/2}$, abbiamo che, dato che stiamo considerando uno sviluppo rispetto ad una base ortonormale, $\|f - f_s\|_2 = \|C - C_s\|_2$, dove C e C_s sono i vettori delle componenti (ordinate) del segnale iniziale e del segnale sparso (compresso). Il segnale è quindi comprimibile quando i valori c_i decadono rapidamente e C è ben approssimato da C_s , che contiene valori uguali a zero dalla componente $K+1$ -esima in avanti, per cui l'errore $\|f - f_s\|_2$ è piccolo. In Figura 4 si mostra un esempio in cui la base utilizzata è una base di *wavelet* che produce un vettore di coefficienti con un significato multiscala (componente in una certa posizione per una determinata scala).

Sempre in Figura 4 si mostra un esempio di compressione in cui l'immagine ricostruita è ottenuta scartando il 75% dei coefficienti dello sviluppo mantenendo solo quelli in modulo maggiori: apparentemente l'immagine ricostruita non ne ha risentito molto. Per una introduzione alla trasformata di Haar, caso più semplice di Wavelet, si veda il riquadro corrispondente.

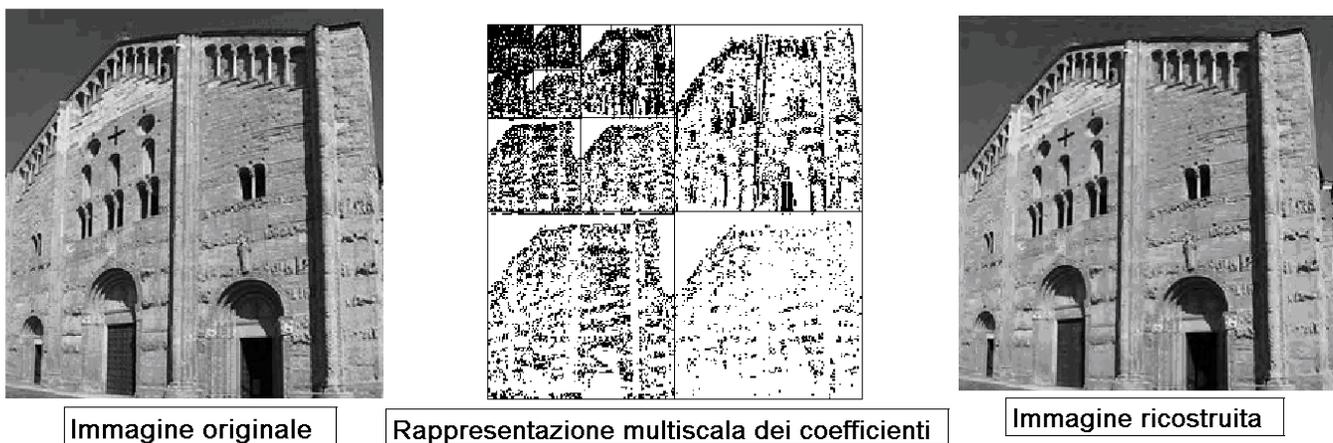


Figura 4. Un esempio di “compressione” di un’immagine: a sinistra abbiamo l’immagine originale, al centro i coefficienti della trasformata Wavelet (gli spazi bianchi rappresentano coefficienti “piccoli”), a sinistra la ricostruzione dell’immagine in base ai soli coefficienti “grandi” della trasformata. In questo esempio il 75% dei coefficienti è stato scartato con notevole guadagno di spazio.

Il fatto che un segnale sia “sparsificabile” ha molte applicazioni tra cui stime statistiche accurate, classificazione, compressione, ricerca di caratteristiche importanti nel segnale. Notiamo infine che il segnale potrebbe anche essere sparso all’inizio (per esempio pochi valori diversi da zero). In ogni caso per la rappresentazione sparsa del segnale stesso, quando è possibile farla, si procede solitamente acquisendo tutti gli N valori, si calcola la rappresentazione rispetto a una opportuna base e si selezionano i coefficienti. Veniamo al nostro problema:

sapendo che il nostro segnale è comprimibile possiamo adottare una strategia di acquisizione del segnale con $M \ll N$ campionamenti (misure)?

In altri termini: possiamo catturare direttamente in fase di acquisizione le caratteristiche peculiari del segnale? In questo modo avrebbe un notevole risparmio sia nella fase di acquisizione che in quella di memorizzazione e in quella di trasmissione del campionamento. Resterebbe il problema di ricostruire il segnale stesso dalle M misure fatte. Questo il problema che il *compressive sensing* risolve: come fare le domande giuste al segnale.

3. L'insostenibile leggerezza del caso

Riassumiamo la situazione. Abbiamo un segnale \mathbf{f} comprimibile, K -sparso rispetto ad una data base Ψ , ed una base Φ con cui effettuare il campionamento (per semplicità le basi siano ortonormali). A priori non sappiamo quali sono le K componenti importanti del segnale, ma sappiamo che ci sono: vogliamo trovare Φ in modo tale che l'acquisizione comporti già la cattura del segnale compresso e che non distrugga quindi le componenti salienti del segnale. Per realizzare questo scopo, le basi Φ e Ψ devono essere *incoerenti* ovvero che i vettori di Ψ non abbiano una rappresentazione sparsa in Φ e viceversa. Un indice della coerenza tra le due basi è la seguente quantità

$$\mu(\Phi, \Psi) = N^{1/2} \cdot \max_{i,j} |(\varphi_i, \psi_j)|$$

che misura la correlazione più grande tra gli elementi delle due basi: due elementi molto correlati portano ad un valore $\mu(\Phi, \Psi)$ grande. In generale $\mu(\Phi, \Psi) \in [1, N^{1/2}]$ e nel *compressive sensing* si mira a valori bassi. Per esempio, il campionamento con la base canonica, quindi con valori puntuali, risulta incoerente con la base di Fourier scelta come base di rappresentazione.

Cerchiamo una strategia generale, indipendente dal tipo di segnale e quindi possibilmente dal tipo di base Ψ con cui lo si rappresenta. Una soluzione viene fornita dal caso o meglio dalle matrici casuali, che sono incoerenti, con alta probabilità, con ogni base Ψ . Risulta complicato qui definire cosa intendiamo per Φ casuale: ci affidiamo all'intuizione del lettore. Per esempio, possiamo intendere i valori degli elementi φ della base Φ come variabili casuali indipendenti ed identicamente distribuite da una densità di probabilità Gaussiana con media zero e varianza $1/N$ (l'usuale distribuzione "a campana").

A solo titolo illustrativo del tipo di risultati teorici citiamo il seguente. Sia $\mathbf{f} \in \mathbf{R}^N$ un segnale K -sparso rispetto alla base Ψ , selezioniamo M misure attraverso una base casuale Φ , se

$$M \geq C \cdot \mu^2(\Phi, \Psi) \cdot K \cdot \log(N)$$

Per una costante positiva C , allora posso ricostruire il segnale \mathbf{f} con alta probabilità. Risultati di questo tipo possono lasciare insoddisfatti perché non abbiamo quantificato con che probabilità riusciamo a ricostruire esattamente, come stimare le quantità nella disuguaglianza scritta sopra, come esemplificare.

Qui ci limitiamo a ribadire che attraverso una selezione di vettori casuali riusciamo a determinare una strategia generale che ci porta ad una acquisizione economica, $M < N$, del segnale e ad una sua ricostruzione con una confidenza ragionevole. Per la ricostruzione del segnale partiamo dal vettore delle misure $\mathbf{y} = \Phi \mathbf{f}$, dalle basi Φ e Ψ e vogliamo ricostruire \mathbf{f} . Il problema, dato che $M < N$, non ha soluzione unica e gli algoritmi di ricostruzione mirano a ricostruire il segnale K -sparso a meno di traslazioni. Un approccio classico a questo problema inverso di ricostruzione consiste nel ricercare il vettore \mathbf{f}^* di "energia" minima cioè di norma euclidea minima risolvendo il problema di minimo vincolato

$\min \| \mathbf{g} \|_2$ tale che $y_k = (\varphi_k, \Psi \mathbf{g})$, $k = 1, \dots, M$.

Questo problema di minimizzazione ha una soluzione ottenibile in forma chiusa ma sfortunatamente non consente di ottenere soluzioni K-sparse: in generale la soluzione ha molti elementi non zero. Una norma naturale potrebbe essere la norma indicata con $\| \mathbf{x} \|_0$ che conta il numero di elementi diversi da zero del vettore \mathbf{x} . Il problema di minimo diventa,

$\min \| \mathbf{g} \|_0$ tale che $y_k = (\varphi_k, \Psi \mathbf{g})$, $k = 1, \dots, M$

che può certamente ritrovare le soluzione K-sparse ma che è numericamente instabile e richiede l'enumerazione esaustiva di tutte le possibili posizioni di K elementi diversi da zero scelti tra N posizioni possibili: combinazioni di N oggetti presi K a K. Quest'ultimo problema risulta essere uno di quei problemi allo stato attuale computazionalmente "intrattabili": è un problema NP.

Fortunatamente una norma che conduce ad un problema di ottimizzazione trattabile (problema di ottimizzazione convessa) e che consente la ricostruzione di segnali K-sparsi è la norma $\| \mathbf{x} \|_1 = \sum |x_i|$.

Il risultato citato sopra si riferisce a questo tipo di ricostruzione, ovvero alla ricostruzione ottenuta attraverso la soluzione del seguente problema di minimo,

$\min \| \mathbf{g} \|_1$ tale che $y_k = (\varphi_k, \Psi \mathbf{g})$, $k = 1, \dots, M$.

Per un confronto tra le varie norme abbiamo riportato in Figura 5 la rappresentazione geometrica di varie "sfere" rispetto a norme differenti. La porzione di piano rappresenta i vincoli del problema, trovare il minimo significa trovare la sfera "tangente", ovvero la più vicina. Si può osservare che cambiando la norma si possono ottenere soluzioni più o meno sparse.

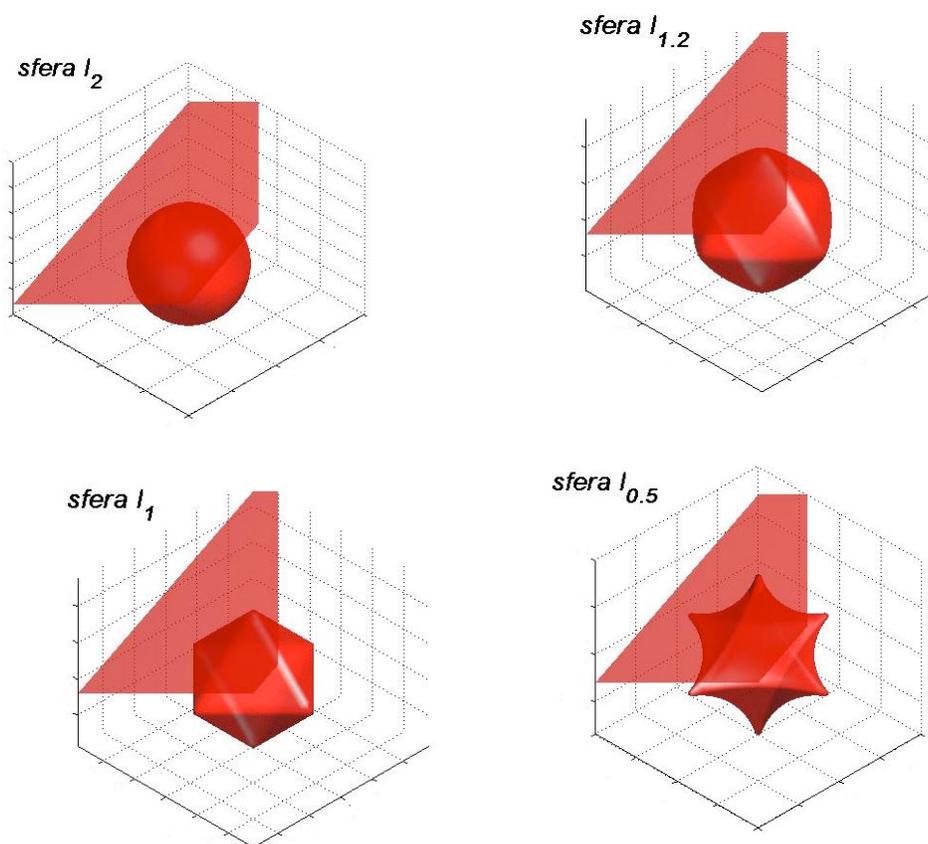
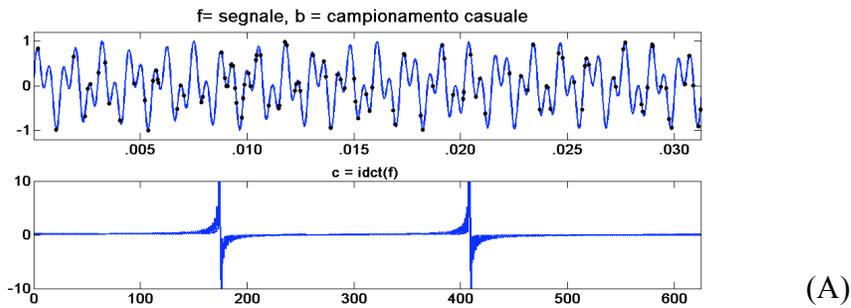
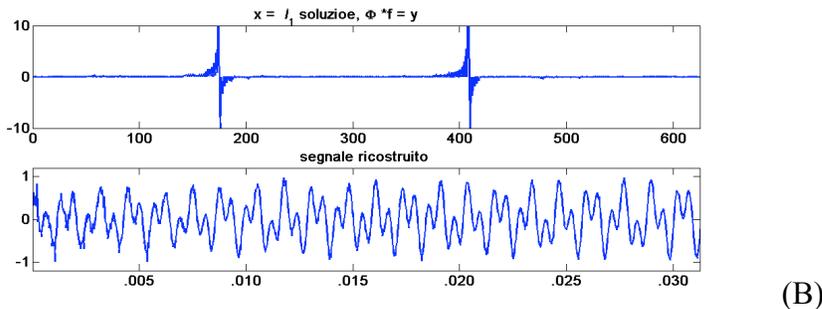


Figura 5. Soluzione del problema di ricostruzione come minimo vincolato, qui i vincoli sono rappresentati da una porzione di piano, rispetto a norme differenti. La norma L_p significa $(\|x\|_p)^p = \sum |x_i|^p$.

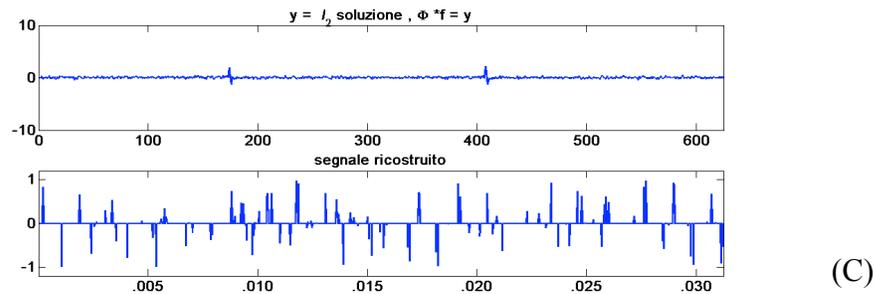
In Figura 6 mostriamo un esempio di ricostruzione di un segnale che risulta sparso rispetto alla base di Fourier, qui considerata solo come componenti rispetto a funzioni coseno di differente frequenza. Si confronta la soluzione del problema di ottimizzazione con differenti norme e partendo da un campionamento casuale del segnale. La scelta della norma comporta la possibilità di una ricostruzione corretta del segnale comprimibile.



(A)



(B)



(C)

Figura 6. Soluzione del problema di ricostruzione come minimo vincolato partendo da un campionamento casuale di un segnale che risulta sparso rispetto alla rappresentazione rispetto alla base di Fourier ristretta alle funzioni coseno di varia frequenza, (A) segnale originale e componenti in frequenza; (B) ricostruzione delle componenti e del segnale utilizzando la norma L_1 ; (C) ricostruzione delle componenti in frequenza e del segnale utilizzando la norma L_2 .

4. Conclusioni sparse

Il *compressed sensing* risponde in fondo ad una domanda naturale e semplice, almeno nella formulazione: se ho un segnale che è ben rappresentabile da 1000 componenti rispetto ad una rappresentazione perché devo fare un milione di campionamenti per acquisire il segnale stesso?

Il campionamento casuale associato ad una possibile rappresentazione sparsa e a un opportuno problema di ottimizzazione risponde alla domanda e fornisce uno strumento computazionale adeguato. Un esempio pratico recente consiste in *single-pixel camera* in cui il campionamento casuale è fisicamente realizzato attraverso un sistema di microspecchi che permettono di fare misurazioni di una scena secondo le indicazioni del *compressed sensing*.

Altre applicazioni riguardano la classificazione di immagini, per esempio per il riconoscimento automatico di volti, e la ricostruzione di immagini in presenza di rumore (cioè di un elemento di disturbo) o di informazioni mancanti.

Dal punto di vista matematico, i collegamenti tra teoria dell'approssimazione, analisi armonica, calcolo delle probabilità, analisi matematica sono fertili e di notevole interesse. Problemi aperti rimangono e una sbirciata al blog di Terence Tao possono aggiornare sulle ultime novità (<http://terrytao.wordpress.com/>)
Certamente il *compressed sensing* rappresenta un esempio notevole di matematica in azione.

Per saperne di più

Riferimenti Bibliografici

[EP] E. Prestini, *Applicazioni dell'Analisi Armonica*, Hoepli, Milano, 1996

[Ma] S. Mallat, *A Wavelet tour of signal processing*. Academic Press, 1998.

[GW] R. C. Gonzalez, R. E. Woods, *Digital Image Processing*, Addison-Wesley, Reading MA, 1992.

[Ca] K.R. Castleman, *Digital Image Processing*, Prentice Hall, 1996.

Webgrafia

<http://terrytao.wordpress.com/> (blog di Terence Tao)

<http://dsp.rice.edu/cs> (pagina ricca di informazioni, lavori di ricerca, tutorial,...)

<http://nuit-blanche.blogspot.com/search/label/CS> (blog su compressive sensing)

<http://www.ams.org/samplings/math-history/hap7-pixel.pdf> (articolo divulgativo AMS)

<http://www-stat.stanford.edu/~donoho/> (pagina di D. Donoho)

<http://www-stat.stanford.edu/~candes/> (pagina di E. Candes)

Una introduzione computazionale alla trasformata di Haar

Facciamo un solo esempio per capire il problema della “compressione” delle informazioni contenute in un’immagine al fine di selezionare solo quelle significative e al fine di ottimizzare l’occupazione di memoria e la velocità di trasmissione. Consideriamo per semplicità un vettore s in una dimensione e non un’immagine (matrice di pixel). Supponiamo che i valori di s siano,

$$s = [10 \ 20 \ -5 \ 15 \ 12 \ 12 \ 30 \ 25]$$

invece del segnale originale consideriamo una sua rappresentazione equivalente: ad ogni coppia di valori a, b sostituiamo i valori della media e della loro differenza,

$$a, b \Rightarrow m = (a + b)/2 ; d = (a - b)/2.$$

Operando, come detto per ogni coppia di valori contigui si ottiene,

$$s = [10 \ 20 \ -5 \ 15 \ 12 \ 12 \ 30 \ 25] \Rightarrow s^1 = [15 \ -5 \ 5 \ -10 \ 12 \ 0 \ 27.5 \ 2.5]$$

per esempio per la prima coppia $a=10, b=20$, quindi $m = (20+10)/2=15$, $d=(10-20)/2=-5$ e analogamente per le altre coppie. A questo punto possiamo raggruppare nelle prime posizioni tutte le medie e nelle restanti posizioni del vettore s^1 le differenze:

$$s^1 = [15 \ -5 \ 5 \ -10 \ 12 \ 0 \ 27.5 \ 2.5] \Rightarrow \underline{s}^1 = [15 \ 5 \ 12 \ 27.5 \ | \ -5 \ -10 \ 0 \ 2.5].$$

Ripetiamo l’operazione media/differenze sulle prime componenti (le quattro medie delle coppie del vettore originale s), e sostituiamo ancora le coppie di valori con le coppie medie/differenze:

$$\underline{s}^1 = [15 \ 5 \ 12 \ 27.5 \ | \ -5 \ -10 \ 0 \ 2.5] \Rightarrow s^2 = [10 \ 5 \ 19.75 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5].$$

Ancora raggruppiamo le medie nella prima parte del vettore e le differenze nella seconda metà,

$$s^2 = [10 \ 5 \ 19.75 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5] \Rightarrow \underline{s}^2 = [10 \ 19.75 \ | \ 5 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5].$$

Infine possiamo sostituire alla prima coppia del vettore \underline{s}^2 la coppia media/differenza (questa volta la permutazione finale non è necessaria),

$$\underline{s}^2 = [10 \ 19.75 \ | \ 5 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5] \Rightarrow s^3 = [14.875 \ | \ -4.875 \ | \ 5 \ -7.75 \ | \ -5 \ -10 \ 0 \ 2.5].$$

Il contenuto d’informazione del vettore s^3 è il medesimo del vettore di partenza s . Infatti le operazioni fatte sono invertibili, dalla coppia media/differenza m, d possiamo risalire univocamente alla coppia originale a, b :

$$m, d \Rightarrow a = (m + d); \quad b = (m - d).$$

Ogni singolo passo è quindi invertibile e dalle componenti dell’ultimo vettore possiamo ricostruire il vettore originale (ovviamente occorre tenere conto delle permutazioni fatte). In Figura 7 si mostra il vettore originale, abbiamo collegato i punti con una spezzata, ed il vettore trasformato finale (in ascissa ci sono gli indici delle componenti).

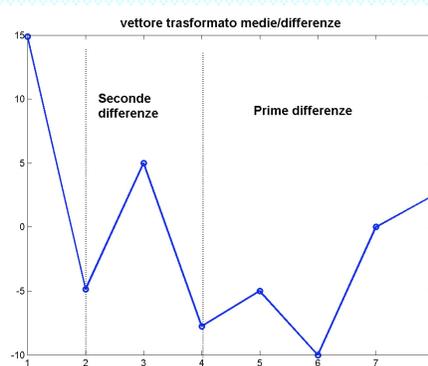
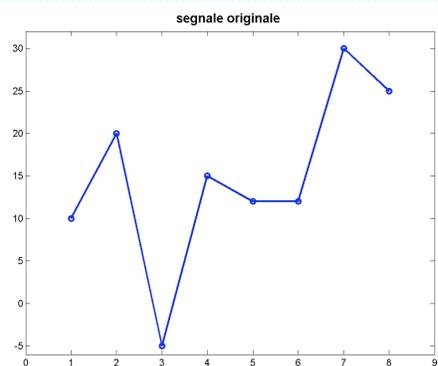


Figura 7. Rappresentazione grafica di un vettore e del vettore trasformato con successive operazioni di media/differenza.

Interpretiamo meglio i due grafici in Figura 7. Le medie successive indicano l'andamento locale medio dove, però, la località cambia nel senso che si mediano valori prima adiacenti poi coppie di medie e via dicendo. In sostanza intervengono valori appartenenti a sottoinsiemi sempre più ampi, in un certo senso stiamo guardando il segnale a scale differenti e ci spostiamo da una scala piccola (valori adiacenti) ad una scala grande (ultime medie). Allo stesso modo le differenze indicano la variazione locale e su scale differenti. A questo punto: come selezionare le informazioni? Un primo metodo consiste nello stabilire una soglia e nel porre uguali a zero le differenze in modulo sotto la soglia prescelta. In altri termini: se la differenza è considerata piccola possiamo rappresentare i valori di una coppia attraverso il loro valore medio. Potremmo commettere un errore, anzi quasi certamente lo commetteremo, ma stiamo accettando di fare un'approssimazione (controllata). In Figura 8 si mostra un esempio di tale procedimento con soglia $T_1 = 5$ e, rispettivamente $T_2 = 8$, quindi ponendo uguale a zero le componenti di tipo differenza in modulo minore od uguale alla soglia e ricostruendo il segnale dal vettore trasformato e con le differenze "alterate". Nei casi reali molte componenti differenza saranno azzerate e si mettono in atto tecniche opportune di memorizzazione di vettori con molti zeri (vettori o matrici sparse) con un guadagno in termini di occupazione di memoria (attenzione, senza un trattamento particolare del vettore modificato non ci sarebbe nessun vantaggio). Le domande che si pongono sono varie, tra cui: possiamo stimare l'errore tra il vettore originale e quello modificato? Vi è una alterazione ragionevole del vettore originale in modo da mantenere proprietà qualitative essenziali? È possibile generalizzare il procedimento? È possibile utilizzare operazioni più sofisticate di media e differenza con un maggiore guadagno?

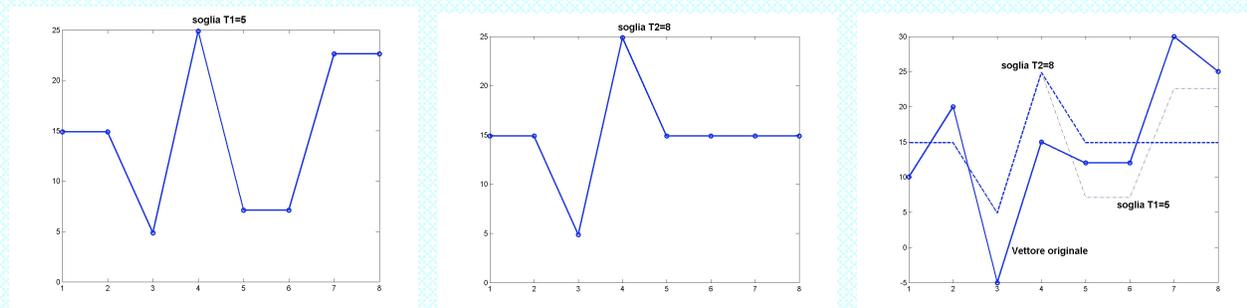


Figura 13. Segnali ricostruiti dopo operazione di azzeramento in base a soglie differenti delle componenti di tipo differenza.

Il procedimento visto può anche essere interpretato utilizzando l'algebra lineare, le diverse rappresentazioni corrispondono a diverse basi. Nell'esempio fatto si passa dalla base canonica di vettori $e_i = [0 \ 0 \ \dots \ 1 \ \dots \ 0]$, con un solo 1 nella posizione i -esima, ad una base detta base di Haar. Cambiando punto di vista e passando da vettori a funzioni e spazi funzionali si costruiscono nuove basi per vari spazi: le basi di Wavelet. A questo punto il racconto diventa di estremo interesse sia teorico che applicativo ma, come si suol dire, è un'altra storia. Citiamo solo un testo ormai considerato un classico di S. Mallat sulle Wavelet e oltre [MA] e rimandiamo alle risorse disponibili in rete per una ricerca personale che lasciamo al lettore interessato (per esempio partendo dal link www.wavelet.org).