An Introduction to Compressive Sensing

Olga V. Holtz
UC Berkeley & TU Berlin

Stanford
January 2009
Compressed Sensing: History

Compressed Sensing (CS)

People involved, (Right to left: J. Claerbout, B. Logan, D. Donoho, E. Candés, T. Tao and R. DeVore)
Compressed Sensing: Introduction

Old-fashioned Thinking

- Collect data at grid points
- For $n$ pixels, take $n$ observations

Compressed Sensing (CS)

- Takes only $O(n^{1/4} \log^5(n))$ random measurements instead of $n$

(CS camera at Rice)
Model signals as band-limited functions $x(t)$

Support of $\hat{x}$ is contained in $[-\Omega\pi, \Omega\pi]$ 

Shannon-Nyquist

Uniform time sampling with spacing $h \leq 1/\Omega$ gives exact reconstruction

A/D converters: sample and quantize

Problem: if $\Omega$ is very large, one cannot build circuits to sample at the desired rate
Compressive sensing seeks a way out of this dilemma

Two new components:

- New model classes for signals: signals are **sparse** in some representation system (basis/frame)
- New meaning of samples: sample is a **linear functional** applied to the signal
Compressive sensing seeks a way out of this dilemma

Two new components:

- New model classes for signals: signals are sparse in some representation system (basis/frame)
- New meaning of samples: sample is a linear functional applied to the signal

Given $x \in \mathbb{R}^n$ with $n$ large, ask $m$ non-adaptive questions about $x$

- Question means inner product $v \cdot x$ with $v \in \mathbb{R}^n$ means sample
- Such sampling is described by an $m \times n$ linear system $\Phi x = y$
With no additional information on $x$ cannot say anything
But we are interested in those $x$ that have structure
Typically $x$ can be represented by sparse linear combinations of certain building blocks (e.g., a basis)
Issue: in many problems, we do not know the basis
Here we assume the basis is known (for now)
Ansatz: look for $k$-sparse solutions:

$$x \in \Sigma_k \text{ that is } \# \text{supp}(x) \leq k.$$
Sparsest Solutions of Linear equations

Find a **sparsest** solution of linear system

\[(P_0) \quad \min \{ \|x\|_0 : \Phi x = b, \ x \in \mathbb{R}^n \}\]

where \( \|x\|_0 = \) number of nonzeros of \(x\) and \(\Phi \in \mathbb{R}^{m \times n}\) with \(m < n\).

- The solution is in general **not** unique.
- Moreover, this problem is **NP-Hard**
Basis Pursuit

Main idea:
Use the convex relaxation

\[(P_1) \quad \min \{ \|x\|_1 : \Phi x = b, \ x \in \mathbb{R}^n \}\]

Basis Pursuit [Chen, Donoho, and Saunders (1999)]

Solving \((P_1)\) in polynomial time
Can be solved by linear programming:

\[
\begin{align*}
\min & \quad \mathbb{1}^T y \\
\text{s.t.} & \quad \Phi x = b \\
& \quad -y \leq x \leq y
\end{align*}
\]
Mutual incoherence:

\[ M(\Phi) = \max_{i \neq j} |\phi_i^T \phi_j| \]

where \( \Phi = [\phi_1 \ldots \phi_n] \in \mathbb{R}^{m \times n} \) and \( \|\phi_i\|_2 = 1 \).

**Theorem (Elad and Bruckstein (2002))**

Suppose that for the sparsest solution \( x^* \) we have

\[ \| x^* \|_0 < \frac{(\sqrt{2} - \frac{1}{2})}{M(\Phi)}. \]

Then the solution of \((P_1)\) is equal to the solution of \((P_0)\).
Restricted Isometry Property of Order $k$ [Candès, Romberg, Tao (2006)]: Let $\delta_k$ be the smallest number such that

$$(1 - \delta_k)\|x\|_2^2 \leq \|\Phi x\|_2^2 \leq (1 + \delta_k)\|x\|_2^2$$

for all $k$-sparse vectors $x \in \mathbb{R}^n$ where $\Phi = [\phi_1 \ldots \phi_n] \in \mathbb{R}^{m \times n}$.

**Theorem (E. J. Candès (2008))**

If $\delta_{2k} < \sqrt{2} - 1$, then for all $k$-sparse vectors $x$ such that $\Phi x = b$, the solution of $(P_1)$ is equal to the solution of $(P_0)$. 
Basis Pursuit De-Noising (BPDN):

\[
(P_1^\epsilon) \quad \min \{ \| x \|_1 : \| \Phi x - b \|_2 \leq \epsilon \}
\]

[Chen, Donoho, and Saunders (1999)]

Theorem (E. J. Candès (2008))

Suppose that the matrix \( \Phi \) is given and \( b = \Phi \hat{x} + e \) where \( \| e \|_2 \leq \epsilon \). If \( \delta_{2k} < \sqrt{2} - 1 \), then

\[
\| x^* - \hat{x} \|_2 \leq C_0 k^{-1/2} \sigma_k(\hat{x})_1 + C_1 \epsilon,
\]

where \( x^* \) is the solution of \( (P_1^\epsilon) \) and

\[
\sigma_k(\hat{x})_1 = \min_{z \in \Sigma_k} \| \hat{x} - z \|_1.
\]

Other Heuristics: Orthogonal Matching Pursuit, Mangasarian’s approach, Bilinear formulation, etc.
End of Part I.
Good compressive sensing (CS) matrices:

**Known Result for Random matrices**

- Known reconstruction bounds for matrices with entries drawn at random from various probability distributions:

  \[ k \leq Cm/\log(n/m). \]

- Specific recipes include Gaussian, Bernoulli and other classical matrix ensembles.

- Particular case: there is a probabilistic construction of matrices \( \Phi \) of size \( m \times n \) with entries \( \{\pm \frac{1}{\sqrt{m}}\} \) satisfying RIP of order \( k \) with the above bound.
Introduce the **Concentration of Measure Inequality (CMI)** property on a probability space \((\Omega, \varrho)\)

Suppose \(\Phi = \Phi(\omega)\) is a collection of random \(m \times n\) matrices

**Property \(PO(\delta)\):** the collection is said to have CMI if, for each \(x \in \mathbb{R}^n\), there is a set \(\Omega_0(x, \delta) \subset \Omega\) s.t.

\[
(1 - \delta) \|x\|_2 \leq \|\Phi x\|_2 \leq (1 + \delta) \|x\|_2, \quad \omega \in \Omega(x, \delta)
\]

and \(\varrho(\Omega(x, \delta)^c) \leq C_0 e^{-c_0 m \delta^2}\)

Gaussian, Bernoulli and many other families have this property
Johnson-Lindenstrauss Lemma.

Given $\epsilon \in (0, 1)$, a set $X$ of points in $\mathbb{R}^n$ such that $\#X =: \sigma > \sigma_0 = O(\ln m / \epsilon^2)$, there is a Lipschitz function $\Phi : \mathbb{R}^n \to \mathbb{R}^m$ s.t.

$$(1 - \epsilon)\|u - v\|_2 \leq \|\Phi(u) - \Phi(v)\|_2 \leq (1 + \epsilon)\|u - v\|_2.$$ 

If $X$ is a set of points and $m > c \ln(\#X)\epsilon^{-2}$ with $c$ sufficiently large, then the set $\Omega_0 := \cap_{x, x' \in X} \Omega(x - x', \epsilon)$ satisfies

$$\varrho(\Omega_0^c) \leq C_0(\#X)^2 e^{-c_0 n \epsilon^2} = e^{2 \ln(\#X) - c_0 m \epsilon^2 + \ln C_0}$$

If $m \geq (2 \ln(\#X) + \ln C_0) / c_0 \epsilon^2$, then the measure is $< 1$ hence we get the JL lemma.
If \( k \leq cm/\ln(n/m) \) and \( \Phi \) satisfies JL, then we have RIP of order \( k \).

For \( c \) sufficiently small, the probability that a Gaussian ensemble satisfies RIP is \( 1 - Ce^{-cm} \).

If we have a collection of \( O(e^{cm}) \) bases, then a random draw of Gaussian will satisfy RIP with respect to all of these bases simultaneously.

**Basic reason why this works**

\[
\Pr \left[ \left| \left\| \Phi(\omega)x \right\|_{\ell^2}^2 - \left\| x \right\|_{\ell^2}^2 \right| \geq \varepsilon \left\| x \right\|_{\ell^2}^2 \right] \leq 2e^{-mc_0(\varepsilon)}, \quad 0 < \varepsilon < 1.
\]
Difficulties in Deterministic Case

[DeVore]: Proposes a deterministic constructions of order

\[ k \leq C\sqrt{m} \log n / \log(n/m) \]

which is still far from probabilistic results.

Very recent ideas

Find deterministic constructions with better bounds

- using bipartite **expander graphs** [Indyk, Hassibi, Xu];
- using **structured matrices** such as Toeplitz, cyclic, generalized Vandermonde matrices.
- Subgoal: deterministic **polynomial time** algorithm for constructing good CS matrices.
- Holy grail: \( k \leq Cm/\log(n/m) \) – achievable for random matrices but not yet in deterministic constructions.
Cohen, Dahmen, DeVore [2009]: Compressed sensing and best $k$-term approximation, JAMS.

**Best $k$-term approximation error:**

$$
\sigma_k(x) := \inf_{z \in \Sigma_k} \| x - z \|_X.
$$

**Encoder-decoder viewpoint**

The matrix $\Phi$ serves as an encoder producing $y = \Phi x$. To extract $x$ / approximation to $x$, use a decoder $\Delta$ (not necessarily linear). Thus

$$\Delta(y) = \Delta(\Phi x) \text{ approximates } x.$$
Ask for the largest value of $k$ s.t.

$$x \in \Sigma_k \implies \Delta(\Phi x) = x.$$  

Ask for the largest value of $k$ s.t., for a given class $K$,

$$E_n(K)_X \leq C\sigma_k(K)_X, \text{ where }$$

$$E_n(K)_X := \inf_{(\Phi, \Delta)} \sup_{x \in K} \|x - \Delta(\Phi x)\|,$$

$$\sigma_k(K)_X := \sup_{x \in K} \sigma_k(X).$$

A pair $(\Phi, \Delta)$ is called instance-optimal of order $k$ with constant $C$ for the space $X$ is

$$\|x - \Delta(\Phi x)\|_X \leq C\sigma_k(X)_X$$

for all $x \in X$ with a constant $C$ independent of $k$ and $n$. 
Connection with Gelfand widths

**Gelfand widths**

For $K$ a compact set in $X$ and $m \in \mathbb{N}$, the **Gelfand width** of $K$ of order $m$ is

$$d^m(K)_X := \inf_{\text{codim } Y \leq m} \sup \{ \|x\|_X : x \in K \cap Y \}.$$ 

**Basic result**

**Lemma.** Let $K \subseteq \mathbb{R}^n$ be symmetric, i.e., $K = -K$, and satisfy $K + K \subseteq C_0K$ for some $C_0$. If $X \subseteq \mathbb{R}^n$ is any normed space, then

$$d^m(K)_X \leq E_m(K)_X \leq C_0 d^m(K)_X, \quad 1 \leq m \leq n.$$
Orders of Gelfand widths

Orders of Gelfand widths of $\ell_q$ balls

Theorem [Gluskin, Garnaev, Kashin (1977,1984)]

$$C_1 \psi(m, n, q, p) \leq d^m(U(\ell^n_q)) \leq C_2 \psi(m, n, q, p)$$

where

$$\psi(m, n, q, p) := \left( \min\{1, n^{1-1/q} m^{-1/2} \} \right)^{\frac{1/q-1/p}{1/q-1/2}},$$

$$\psi(m, n, 1, 2) := \min\{1, \sqrt{\frac{\log(n/m)}{m}} \}.$$ 

Corollary. The necessary number of measurements $k$ satisfies

$$k \leq c_0 m/ \log(n/m).$$
Denote $\mathcal{N} := \mathcal{N}(\Phi) := \{x : \Phi x = 0\}$.

**Uniqueness of recovery**

**Lemma.** For an $m \times n$ matrix $\Phi$ and for $2k \leq m$, the following are equivalent:

- There is a decoder $\Delta$ s.t. $\Delta(\Phi x) = x$ for all $x \in \Sigma_k$.
- $\Sigma_{2k} \cap \mathcal{N} = \{0\}$.
- For any set $T$ with $\#T = 2k$, the matrix $\Phi_T$ has rank $2k$.
- For any $T$ as above, the matrix $\Phi^*_T \Phi_T$ is positive definite.
Approximate recovery

Approximation to accuracy $\sigma_k$

**Theorem [Cohen, Dahmen, DeVore (2009)].** Given an $m \times n$ matrix $\Phi$, a norm $\| \cdot \|_X$ and a value of $k$, a sufficient condition that there exists a decoder $\Delta$ s.t.

$$\| x - \Delta(\Phi x) \|_X \leq C\sigma_k(x)_X$$

is that

$$\| \eta \|_X \leq C/2 \cdot \sigma_{2k}(\eta)_X, \quad \eta \in \mathcal{N}.$$  

A necessary condition is that

$$\| \eta \|_X \leq C \cdot \sigma_{2k}(\eta)_X, \quad \eta \in \mathcal{N}.$$  

This gives rise to the **null space property** (in $X$ of order $2k$):

$$\| \eta \|_X \leq C \cdot \sigma_{2k}(\eta)_X, \quad \eta \in \mathcal{N}.$$
The null space property

Approximation and the null space property

Corollary [Cohen, Dahmen, DeVore (2009)]. Suppose that $X$ is an $l_p^n$ space, $k \in \mathbb{N}$ and $\Phi$ is an encoding matrix. If $\Phi$ has the null space property in $X$ of order $2k$ with constant $C/2$, then there exists a decoder $\Delta$ so that

$$\|x - \Delta(\Phi x)\|_X \leq C\sigma_k(x)_X.$$ 

Conversely, the validity of the above condition for some decoder $\Delta$ implies that $\Phi$ has the null space property in $X$ of order $2k$ with constant $C$. 
Theorem [Candès-Romberg-Tao (2006)]. Let $\Phi$ be any matrix with satisfies the RIP of order $3k$ with $\delta_{3k} \leq \delta < (\sqrt{2} - 1)^2 / 3$. Define the decoder $\Delta$ by

$$\Delta(y) := \arg\min_{\Phi z = y} \|z\|_{\ell_1}.$$  

Then $(\Phi, \Delta)$ satisfies

$$\|x - \Delta(\Phi x)\|_X \leq C\sigma_k(x)_X$$

in $X = \ell_1$ with $C = \frac{2\sqrt{2}+2-(2\sqrt{2}-2)\delta}{\sqrt{2}-1-(\sqrt{2}+1)\delta}$. 
The End.