Compressed Sensing & best k-term approximation

Albert Cohen, Wolfgang Dahmen and Ronald Devore

Contents

- Approximate representation of signals
  - Transform Coding
  - Non adaptive linear encoders
- Quick review of compressed sensing
- Can CS be used to approximate non sparse signals in R^N ?
  - What kind of performance can we expect ?
  - Proofs
- Conclusion

Approximate representation of signals

- Why approximate signals?
  - Aggressive signal acquisition
  - Acquired signal exists in very high dimensional space (R^N)
    - High memory required for storage
    - High computational costs in signal processing
  - Solution:
    - Represent acquired signal in a lower dimensional space (R^M) without losing much information about the signal.
    - Need (ϕ, Δ)
      - $\phi : X \rightarrow X_{coded}$ (encoder)
      - $\Delta: X_{coded} \rightarrow X_{reconstructed}$ (decoder)
      - Approximation/reconstruction error $= X - X_{reconstructed} = X - \Delta(\phi X)$

Approximate representation of signals

- Goal:
  - For given signal class $U \subset R^N$, design (φ,Δ) such that:
    - Encoder φ and decoder Δ should be easy to implement
      - Linear encoders (φ *is an m x N matrix*)
      - Decoders (iterative/recursive, linear filters, greedy algorithms)
    - Encoder φ should be non adaptive with respect to input signal.
    - Approximation/reconstruction error should be bounded
      - $e = x - x_{reconstructed}$
      - $\| e \| < B$ for $\| . \|$ of interest.

Approximate representation of signals

- How to analyze the performance of (φ, Δ) ?
  - Instantaneous approximation error:
    - E(x, U) = || x - Δ (φx) || for some || . || defined on R^N
  - Average approximation error:
    - E(U) = $\int_{x \in U} \|x - \Delta(\varphi x)\| P(x) dx$
  - Support recovery error:
    - E(x,U) = abs(|supp(x) ∩ supp(Δ(φx))| - |supp(x)|)
  - Best k-term approximation error:
    - $\sigma_K(X) = \inf_{z \in \Sigma_k} \|x - z\|$ where $\Sigma_k = \{ x \in R^N : \|x\|_0 \le k \}$

Transform Coding

- Encoder *φ :*
  - Project *X* onto basis *F(i)* to obtain Y : $Y = F X$
  - Retain top *k* coefficients of Y and corresponding index set T
  - Encoder output $\{Y_T, T\}$

Transform Coding

- Encoder *φ :*
  - Project *X* onto basis *F(i)* to obtain Y : $Y = F X$
  - Retain top *k* coefficients of Y and corresponding index set T
  - Encoder output $\{Y_T, T\}$

- Decoder Δ :
  - Generate $\underline{Y}$ such that: $\underline{Y}_i = Y_i$ for $i \in T$ and $\underline{Y}_i = 0$ for $i \in T^C$
  - $X_{reconstructed} = F^{-1}\underline{Y}$

Transform Coding

- Encoder *φ :*
  - Project *X* onto basis *F(i)* to obtain Y : $Y = F X$
  - Retain top *k* coefficients of Y and corresponding index set T
  - Encoder output $\{Y_T, T\}$
- Decoder Δ :
  - Generate $\underline{Y}$ such that: $\underline{Y}_i = Y_i$ for $i \in T$ and $\underline{Y}_i = 0$ for $i \in T^C$
  - $X_{reconstructed} = F^{-1}\underline{Y}$
- Approximation error :
  - $|| X - X_{reconstructed} ||_2 = ||Y - \underline{Y}||_2 = || Y_{T^c} ||_2 = \sigma_k(Y)$
- Main idea:
  - Change of basis such that signal energy is concentrated in a few coefficients in new basis.

Transform Coding

- Inefficiencies in transform coding
  - Encoding process is adaptive
    - It is not known beforehand which coefficients to retain.
    - Top k index set T varies with input signal X
  - High computational costs during encoding
    - Computations in encoder scale with input signal dimension N
    - Top k index set T varies with input signal X

Transform Coding

- Inefficiencies in transform coding
  - Encoding process is adaptive
    - It is not known beforehand which coefficients to retain.
    - Top k index set T varies with input signal X
  - High computational costs during encoding
    - Computations in encoder scale with input signal dimension N
    - Top k index set T varies with input signal X

Question:

*Since we retain only a few coefficients, is it possible to actually compute only a **few linear non adaptive measurements** and still retain necessary information about X?*

Non adaptive linear encoders

- Encoder φ :
  - Project *X* onto *n (<< N)* random basis *F(i)* to obtain Y

$$\begin{bmatrix} y_1 \\ y_2 \\ . \\ . \\ y_n \end{bmatrix}_{n\times 1} = \begin{bmatrix} -------F_1------- \\ -------F_2------- \\ . \\ . \\ -------F_n------- \end{bmatrix}_{n\times N} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ . \\ . \\ x_N \end{bmatrix}_{N\times 1}$$

  - Encoder output **y = φx**
  - Matrix φ is fixed (non adaptive)
- Decoder Δ :
  - ??
- Reconstruction error :
  - $|| x - x_{reconstructed} ||_2 = ||x - \Delta(\varphi x)||_2$
  - Does there exists a decoder such that reconstruction error is as good as transform coding ?
  - How small can *n* be ?

Non adaptive linear encoders

- Encoder φ *:*
  - Encoder output : $\mathbf{y}_{nx1} = \boldsymbol{\varphi}\,\mathbf{x}_{Nx1}$
  - $\boldsymbol{\varphi}$ is a n x N matrix (n << N)
- Decoder Δ :
  - ?? (practically feasible)
- Reconstruction error : (As good as transform coding !)
  - For all x ϵ R^N,

$$\left\| x - \Delta(\varphi x) \right\|_2 \leq C_o \sigma_K \left(x \right)$$

  - C_o is a constant independent of k and N
  - Want *instance optimality of (φ, Δ)*

Non adaptive linear encoders

- Problem Statement
  - Encoder φ : $\mathbf{y}_{n\times 1} = \boldsymbol{\varphi}\, \mathbf{x}_{N\times 1}$
  - Decoder Δ : ??
  - Reconstruction error :
    - For all x ϵ R^N, $\|x - \Delta(\varphi x)\|_2 \leq C_o \sigma_K(x)$
    - C_o is a constant independent of k and N

Question: Can we construct such (φ,Δ) ?

Non adaptive linear encoders

- <u>Problem Statement</u>

  - Encoder φ : $\mathbf{y}_{nx1} = \boldsymbol{\varphi}\, \mathbf{x}_{Nx1}$
  - Decoder Δ : ??
  - Reconstruction error :
    - For all x ϵ R^N, $\|x - \Delta(\varphi x)\|_2 \le C_o \sigma_K(x)$
    - C_o is a constant independent of k and N

Question: Can we construct such (φ,Δ) ?

Compressed Sensing says YES !

- Instance optimal (φ,Δ) exist for k-sparse signals.
- If φ is chosen properly, perfect recovery of x can be guaranteed using feasible decoding schemes.

Compressed Sensing Basics (1/6)

- Major work done by:
  - Candes, Romberg, Tao
  - Donoho
  - Baranuik, Rauhut and many more.

  - Problem Statement :
    - Given Y, reconstruct X
    - Simple measurement model:

$$Y = AX$$

Y_{nx1} - measurement vector
A_{nxN} - fat matrix (n << N)
X_{Nx1} - sparse vector with k non zero entries

Compressed Sensing Basics (2/6)

- Measurement model:

$$Y_{nx1} = A_{nxN} X_{Nx1} \qquad [\, n << N,\ ||X||_0 = k \,]$$

  - A_{nxN} has non trivial null space N(A)
    - Same Y can be caused due to infinitely many X
    - Which X to pick ?
  - Claim:

    For unique k-sparse solution, N(A) should not contain any 2k or less sparse non zero vector

$$N(A) \cap \Sigma_{2k} = \{0\}$$ *Null Space Property of order 2k*

  - Construction of matrix A which satisfies NSP of order 2k is difficult.

Compressed Sensing Basics (3/6)

- Measurement model: $Y_{nx1} = A_{nxN} X_{Nx1}$
  - $n << N$ and $||X||_0 = k$
  - Let $T \subset \{1,2\ldots.N\}$ is the support set of X, $|T| \leq k$
  - Compact measurement model:

$$Y = A_T X_T$$

  - For unique solution, all k columns of sub matrix A_T must be linearly independent for all possible index sets T.
  - For stable solution, Gram(A_T) must be well conditioned. (Why ?)

Compressed Sensing Basics (4/6)

- $Y = AX = A_T X_T$ (T is the support set of X, $|T| \le k$)

- For stable solution:
  - $(A_T)^t A_T$ must be well conditioned for all index sets T.
  - A must satisfy RIP of order K

- <u>Restricted Isometry Property (RIP)</u>

  - Matrix A satisfies RIP of order K if there is a δ_k in (0,1) such that

$$(1-\delta_K)\|X\|_2 \le \|AX\|_2 \le (1+\delta_K)\|X\|_2$$

holds for all $X \in \sum_k$

Compressed Sensing Basics (5/6)

$$Y = AX = A_T X_T$$

(T is the support set of X, $|T| \leq k$)

- For stable solution:
  - A must satisfy RIP of order K
- For unique k-sparse solution:
  - A must satisfy RIP of order 2K
  - RIP of order 2K ≡ NSP of order 2K
- For stable solution via L-1 minimization:
  - A must satisfy RIP of order 3K

Compressed Sensing Basics (6/6)

$$Y = AX = A_T X_T$$

(T is the support set of X, $|T| \leq k$)

- For stable, unique k-sparse solution:
  - A must satisfy RIP of order 2K

- Do random matrices satisfy RIP?
  - *Theorem:*
    Let φ be a n x N random matrix whose entries φ_{ij} are iid and drawn according to a Gaussian distribution with variance = 1/n.

    If $n \geq C\delta^{(-2)}\left(k\log\left(\frac{N}{k}\right) - \log(\varepsilon)\right)$ *for a constant C > 0. Then φ satisfies RIP of order K and RIP constant $\delta_k \leq \delta$ with probability atleast 1-ϵ.*

Non adaptive linear encoders

- <u>Coming back to original problem !</u>
  - Encoder φ : $\mathbf{y}_{nx1} = \boldsymbol{\varphi}\, \mathbf{x}_{Nx1}$
  - Decoder Δ : ??
  - Reconstruction error :
    - For all x ϵ R^N, $\|x - \Delta(\varphi x)\|_2 \leq C_o \sigma_K(x)$
    - C_o is a constant independent of k and N

Question: Can we construct such (φ,Δ) ?

YES ! If φ satisfies RIP of order 3K and *x* is k-sparse.

What can we say about reconstruction error if x is not sparse?

Main Result

- Encoder *:*
  - $y_{nx1} = \varphi \, x_{Nx1}$
  - φ is an n x N Gaussian random matrix
    - Variance of φ_{ij} = 1/n
    - $n \geq c_o \, k \log(N/n)$
- Decoder :
  - Δ = minimum squared residual decoder
- Reconstruction error :
  - There exists a high probability set $\Omega(\varphi) \subset \Omega$ such that for all $P(\Omega(\varphi)) \geq 1 - \epsilon$ such that for all $\omega \in \Omega(\varphi)$, we have

$$\left\| x(\omega) - \Delta(\varphi x(\omega)) \right\|_2 \leq C \sigma_K \left(x(\omega) \right)$$

    where C is a constant independent of k and N.
  - We have instance optimality w.h.p.!

High Level Proof

Gaussian Random Matrices

φ is an nxN Gaussian random matrix with variance = 1/n and $n \geq c_o k \log(N/k)$ **---- (1)**

High Level Proof

Gaussian Random Matrices

φ is an nxN Gaussian random matrix with variance = 1/n and $n \geq c_o k \log(N/k)$ **---- (1)**

RIP and Boundedness property

(a) φ satisfies RIP of order 2K w.h.p.

(b) For each $x \in R^N$, there exists a high probability set $\Omega(x) \subset \Omega$ such that for all $\omega \in \Omega(x)$, $\varphi(\omega)$ satisfies boundedness property** with respect to x and constant C

High Level Proof

Gaussian Random Matrices

φ is an nxN Gaussian random matrix with variance = 1/n and $n \geq c_o k \log(N/k)$ **---- (1)**

RIP and Boundedness property

(a) φ satisfies RIP of order 2K w.h.p.

(b) For each $x \in R^N$, there exists a high probability set $\Omega(x) \subset \Omega$ such that for all $\omega \in \Omega(x)$, $\varphi(\omega)$ satisfies boundedness property** with respect to x and constant C

Existence of good encoders w.h.p

For each $x \in R^N$, there exists a high probability set $\Omega(x) \subset \Omega$ such that for all $\omega \in \Omega(x)$, we have

$$\|x - \Delta(\varphi(\omega)x)\|_2 \leq C_o \sigma_K(x)_{l_2}$$

with $C_o = 1 + 2C/(1-\delta)$ and Δ as the proposed decoder

High Level Proof

Gaussian Random Matrices

φ is an nxN Gaussian random matrix with variance = 1/n and $n \geq c_o k \log(N/k)$ **---- (1)**

RIP and Boundedness property

(a) φ satisfies RIP of order 2K w.h.p.

(b) For each $x \in R^N$, there exists a high probability set $\Omega(x) \subset \Omega$ such that for all $\omega \in \Omega(x)$, $\varphi(\omega)$ satisfies <u>boundedness property**</u> with respect to x and constant C

Existence of good encoders w.h.p

For each $x \in R^N$, there exists a high probability set $\Omega(x) \subset \Omega$ such that for all $\omega \in \Omega(x)$, we have

$$\|x - \Delta(\varphi(\omega)x)\|_2 \leq C_o \sigma_K(x)_{l_2}$$

with $C_o = 1 + 2C/(1-\delta)$ and Δ as the proposed decoder

Instance optimality w.h.p

For every random matrix φ satisfying (1), there exists a high probability set $\Omega(\varphi) \subset \Omega$ such that for all $\omega \in \Omega(\varphi)$, we have:

$$\|x(\omega) - \Delta(\varphi x(\omega))\|_2 \leq C_o \sigma_k(x)_{l_2}$$

for some constant C_o is a constant independent of k and N.

Proof: RIP and BDDness implies existence of encoders w.h.p

- (a) φ satisfies RIP of order 2K w.h.p.
- (b) For each x ϵ R^N, there exists a high probability set Ω(x)⊂Ω such that for all ω ϵ Ω(x) , φ(ω) satisfies boundedness property** with respect to x and constant C

- Pick any arbitrary x ϵ R^N :
  - Let T be the index set of *k* largest coefficients of *x* => $\|x - x_T\|_2 = \sigma_K(x)_{l_2}$
  - Ω_0 : such that for *ω* ϵ $\Omega_{0,}$ *φ(ω) satisfies RIP or order 2k with constant δ*
    - $P(\Omega_0) \geq 1-\epsilon$
  - $\Omega_1(x-x_T)$: such that for *ω* ϵ $\Omega_{1,}$ *φ(ω) satisfies* boundedness probability for $x-x_T$ with constant *C*.
    - $P(\Omega_1(x-x_T)) \geq 1-\epsilon$
  - Let $\Omega' = \Omega_0 \cap \Omega_1(x-x_T)$, then $P(\Omega') \geq 1-2\epsilon$ (easy to show !)
  - Let φ be generated from Ω'
    - Encoder output $y = \varphi x$.
    - Decoder output $x^* = \Delta(y) = \underset{z \in R^N,\ \|z\|_0 = K}{\arg\min} \|y - \varphi z\|_2$
    - We work out the proof !

Proof: Existence of encoders w.h.p implies instance optimality w.h.p

For each $x \in R^N$, there exists a high probability set $\Omega(x) \subset \Omega$ such that for all $\omega \in \Omega(x)$, we have

$$\|x - \Delta(\varphi(\omega)x)\|_2 \leq C_o \sigma_K(x)_{l_2}$$

with $C_o = 1 + 2C/(1-\delta)$ and Δ as the proposed decoder

- We work out the proof !

Summary

- Encoder *:*
  - $y_{n\times1} = \varphi\, x_{N\times1}$
  - φ is an n x N Gaussian random matrix
    - Variance of φ_{ij} = 1/n
    - $n \geq c_o k \log(N/n)$
- Decoder :
  - Δ = minimum squared residual decoder
- Reconstruction error :
  - There exists a high probability set $\Omega(\varphi) \subset \Omega$ such that for all $P(\Omega(\varphi)) \geq 1-\epsilon$ such that for all $\omega \in \Omega(\varphi)$, we have

$$\left\| x(\omega) - \Delta(\varphi x(\omega)) \right\|_2 \leq C\sigma_K \left(x(\omega) \right)$$

    where C is a constant independent of k and N.
  - We have instance optimality w.h.p.!

- BACKUP

Proposed decoder Δ(y) for y = φx

- Decoder depends on φ.
- Decoder output is k-sparse.
- Decoder output minimizes the squared residual i.e.

$$\Delta(y) = \underset{z \in R^N,\ \|z\|_0 = K}{\arg\min} \|y - \varphi z\|_2$$

PROOF CONTINUED...

- Consider,

$$\|x - x^*\|_2 \leq \|x - x_T\|_2 + \|x_T - x^*\|_2$$

$$= \sigma_k(x)_{l_2} + \|x_T - x^*\|_2 \quad \text{----(i)}$$

- Consider second term,

$$\|x_T - x^*\|_2 \leq (1-\delta)^{-1}\|\varphi(x_T - x^*)\|_2$$

$$\leq (1-\delta)^{-1}(\|\varphi(x - x_T)\|_2 + \|\varphi(x - x^*)\|_2)$$

$$= (1-\delta)^{-1}(\|y - \varphi x_T)\|_2 + \|y - \varphi x^*\|_2)$$

$$\leq (1-\delta)^{-1}(\|y - \varphi x_T)\|_2 + \|y - \varphi x_T\|_2)$$

$$\leq 2(1-\delta)^{-1}\|\varphi(x - x_T)\|_2$$

$$\leq 2C(1-\delta)^{-1}\|(x - x_T)\|_2 \;=\; 2C(1-\delta)^{-1}\sigma_k(x)_{l_2} \quad \text{----(ii)}$$

- From (i) and (ii),

$$\|x - x^*\|_2 \leq \left(1 + \frac{2C}{1-\delta}\right)\sigma_k(x)_{l_2}$$