Intro to Compressive Sensing

James Schneeloch
9-9-2014

Notes from Justin Romberg's course on compressive sensing in Tsinghua University on 10-14-2013 in Beijing, China

What is it?

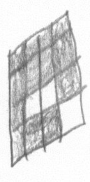
- It is a way of measuring a high-dimensional signal with a relatively low number of measurements by exploiting its compressibility in a given basis.

Vocab

- A signal $\vec{x} \in \mathbb{R}^N$ is compressible in a given basis W, where $W = \{\vec{w}_i\}_{i=1}^N$, if it can be represented in W with fewer than N coefficients.

$$\underbrace{\vec{x}}_{\text{signal}} = \sum_{i=1}^{N} \underbrace{c_i}_{\text{coefficients}} \underbrace{\vec{w}_i}_{\text{basis vectors}}$$

- A signal $\vec{x} \in \mathbb{R}^N$ is S-sparse in W if it has AT MOST S nonzero coefficients in W.

ex  A digital image with three nonzero pixels is 3-sparse in the pixel basis (and also 4-sparse, 5-sparse ...)

How do we usually measure high-dimensional sparse signals?

- We have to measure all N of its coefficients
  - Then, we store the S nonzero coefficients in memory, discarding the rest

If you knew where the S nonzero coefficients were, then you could just measure those ($S < N$) and be done with it.

→ But mostly, you don't know where they are beforehand!

How can one find the S nonzero coefficients without knowing where they are?

(Besides measuring all N of them, that is)

☆ To find these nonzero coefficients with fewer than N measurements, you can measure in a basis <u>incoherent</u> with the sparse basis.

<u>Vocab</u>

- The <u>coherence</u> $\mu(\Phi, \omega)$ between bases $\Phi = \{\vec{\Phi}_i\}_{i=1}^{N}$ and $\omega = \{\vec{\omega}_i\}_{i=1}^{N}$ is a measure of the maximum overlap between two sets of basis vectors (or inner product)

$$\mu(\Phi, \omega) \equiv \sqrt{N} \max_{ij} |\hat{\Phi}_i \cdot \hat{\omega}_j| : \quad \hat{\Phi}_i = \frac{\vec{\Phi}_i}{|\vec{\Phi}_i|}$$

$$\mu \in [1, \sqrt{N}]$$

Note: In the entropic uncertainty principle for observables Q and R:

$$H(Q) + H(R) \geq \log(\Omega)$$

$$\Omega = \frac{N}{\mu(Q,R)^2}$$

- Two bases Φ and ω are <u>maximally incoherent</u> with each other, $(\mu \to 1)$, if any basis vector in Φ is expressed as an even sum over all the basis vectors in ω, and vise versa

<u>ex</u>

Pixel basis and DCT (Discrete Cosine Transform) basis
Pixel basis and DFT (Discrete Fourier Transform) basis

...are both pairs of maximally incoherent (or mutually unbiased) bases.

Note: In practice, we often do measurements from a basis of random vectors.

- These random bases are approximately highly incoherent with any <u>structured basis</u>
  - Structured bases include almost all standard measurement bases
    - Pixel basis
    - DCT, DFT bases
    - wavelet bases

Why measure in a basis incoherent with the sparse basis?

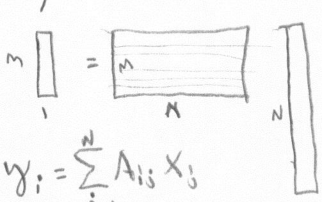
- Each incoherent basis vector is expressed as a sum over almost all the sparse basis vectors, and vise versa.

  $\longrightarrow$ Each incoherent measurement is likely to sample some relevant information about all the important coefficients in the sparse basis.

Note: Each measurement can be thought of as the overlap (or inner product) of your signal with a given basis vector.

ex/

$$\vec{y} = A\vec{x}$$



$$y_i = \sum_{j=1}^{N} A_{ij} x_j$$

$\vec{y} = M \times 1$ vector of measurements

$A = M \times N$ "sensing matrix" (or measurement matrix)

$\rightarrow$ Each row of the sensing matrix is one of the incoherent basis vectors.

$\vec{x} = N \times 1$ signal vector

With M measurements in the incoherent basis, you narrow down $\vec{x}$ to an $(N-M)$ dimensional subspace.

How Many incoherent measurements do you need?

- If $M = N$, then A is just an invertable linear transformation, and $\vec{x} = A^{-1}\vec{y}$
- For random measurements, one needs large enough M so that $\text{rank}(A) = N$ (could take awhile!)

If $M < N$, then

$\vec{y} - A\vec{x} = 0$ has multiple solutions $(\vec{x} : \vec{y} = A\vec{x})$

since this is an underdetermined system of linear equations.

HOWEVER...

If we know that $\vec{x}$ is S-sparse...

Then we can use that information to narrow down our solution set.

☆ In fact, if A is a <u>restricted isometry</u> for 2S-sparse vectors, and $\vec{x}$ is S-sparse, then <u>$\vec{x}$ is the unique solution</u> to

$$\min_{x} |\vec{x}|_0^0 : \vec{y} = A\vec{x}$$

$|x|_0^0$ = "ℓ_0-norm"
= number of nonzero coefficients

(Candes & Tao
IEEE Trans. Inf. Th.
2005)

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

A measurement matrix A satisfies the RIP for S-sparse vectors with precision δ if it preserves the <u>energy</u> (i.e. ℓ_2-norm) of any S-sparse vector $\vec{x}$ to within a factor $(1 \pm \delta)$.

In short: If, for all S-sparse vectors $\vec{x}$,

$$\boxed{(1-\delta)|\vec{x}|_2^2 \leq |A\vec{x}|_2^2 \leq (1+\delta)|\vec{x}|_2^2},$$

... then A is a restricted isometry for S-sparse vectors with precision δ.

Note: If A satisfies RIP for 2S-sparse vectors, then the distances (and inner products) between any two S-sparse vectors is approximately preserved.

→ A maps distinguishable signal vectors to distinguishable measurement vectors.

<u>Why if A satisfies the RIP for 2S-sparse vectors, is an S-sparse $\vec{x}$ uniquely determined?</u>

- Let there be two S-sparse vectors $\vec{x}_1$ and $\vec{x}_2$

  such that $\vec{y} = A\vec{x}_1$.

- We are given that A satisfies the RIP for $2S$-sparse vectors with precision δ

$$(1-\delta)|\vec{x}_1-\vec{x}_2|_2^2 \leq |A(\vec{x}_1-\vec{x}_2)|_2^2 \leq (1+\delta)|\vec{x}_1-\vec{x}_2|_2^2$$

Note: $(\vec{x}_1-\vec{x}_2)$ is a $2S$-sparse vector.

Since $\vec{y} = A\vec{x}_1$, then,

$$0 \leq \underbrace{(1-\delta)|\vec{x}_1-\vec{x}_2|_2^2}_{\varepsilon} \leq |\vec{y}-A\vec{x}_2|_2^2 \leq (1+\delta)|\vec{x}_1-\vec{x}_2|_2^2$$

If $\vec{x}_1 \neq \vec{x}_2$, then $|\vec{y}-A\vec{x}_2|_2^2 \geq \varepsilon > 0$

If $\vec{x}_1 \neq \vec{x}_2$, then $\vec{y} \neq A\vec{x}_2$

$\therefore$ If A satisfies RIP for $2S$-sparse vectors, and $\vec{x}$ is an S-sparse signal consistent with the measurement vector, $\vec{y} = A\vec{x}$, there is no other S-sparse signal consistent with the data.

Note:

M = # of rows of A

M = # of measurements

$$\left(\begin{array}{l}\text{How many measurements}\\ \text{we need to find } \vec{x} \text{ by}\\ \text{optimization}\end{array}\right) \cong \left(\begin{array}{l}\text{How many rows } A \text{ must have}\\ \text{so that it satisfies the RIP for}\\ 2S\text{-sparse vectors}\end{array}\right)$$

big problem!

Actually solving

$\min_{\vec{x}} |\vec{x}|_0^0 : \vec{y} = A\vec{x}$ is computationally intractable

$\rightarrow$ One would have to check $\frac{N!}{(N-S)!\,S!}$ possible decompositions of $\vec{x}$

easy Solution

If A satisfies RIP for 3S-sparse vectors, then $\vec{x}$ is the solution to

$$\min_{\vec{x}} |\vec{x}|_1^1 : \vec{y} = A\vec{x}$$ [this method is called basis pursuit]

(Candes, Romberg, & Tao 2004)

ℓ_1 norm of x

$$|x|_1^1 = \sum_{i=1}^{N} |x_i|$$

ℓ_p norm of x

$$|x|_p^p = \left(\sum_{i=1}^{N} |x_i|^p\right)^{\frac{1}{p}}$$

- Basis pursuit is a convex optimization problem where (relatively) easy algorithms exist for solving them!

What sort of measurement matrices satisfy the RIP for 2S-sparse vectors?

- It is hard to design such sensing matrices, but certain classes of random matrices can be shown to satisfy the RIP with exponentially high probabilities.

① Sub-Gaussian (i.i.d.) random matrices

Gaussian: $A_{ij} \sim N(0, \frac{1}{M})$

$$\boxed{P\left((1-\delta)|\vec{x}|_2^2 \le |A\vec{x}|_2^2 \le (1+\delta)|\vec{x}|_2^2\right) \ge 1 - 2e^{\left(c \cdot S\log\left(\frac{N}{S}\right) - \frac{M}{8}\delta^2\right)}}$$

→ Use Markov inequality to get this

$$P(x > t) \le \frac{E[e^{\lambda x}]}{e^{\lambda t}}$$

→ Where $M > (\text{const})\, S \log\left(\frac{N}{S}\right)$ → (nearly optimal!)

$p \to 1$ quickly!

- A random variable X is sub-Gaussian if there is a $b > 0$:

$$E[e^{tX}] \le e^{\frac{b^2 t^2}{2}}$$

$\Rightarrow E[X] = 0$ and $\mathrm{Var}[X] \le b^2$

- Any centered bounded random variable is sub-Gaussian
- If X is sub-Gaussian, then there is some $(c, \lambda) > 0$:

$$P(|X| \ge \lambda) \le 2e^{-c\lambda^2}$$

→ decays faster than Gaussian

IMPORTANT: For different classes of random sub-Gaussian matrices, the constant in front is different.

(2) Structured random matrices

i.e., Random rows from an incoherent orthonormal basis.

A_{ij} = j^{th} entry of i^{th} randomly drawn basis vector of incoherent basis.

$$P\left((1-\delta)|\vec{x}|_2^2 \leq |A\vec{x}|_2^2 \leq (1+\delta)|\vec{x}|_2^2\right) \to 1 \text{ quickly where}$$

$$M > (\text{const}) \cdot S(\log(N))^4$$

What if the sparse basis is different than the measurement basis?

$$\vec{y} = A\vec{x}' = A(\underbrace{\overset{\text{basis transformation}}{R}\vec{x}}_{\text{signal in measurement basis}}) = (AR)\underset{\text{signal in sparse basis}}{\vec{x}}$$

- If the modified sensing matrix (AR) also satisfies the RIP, then one can still get $\vec{x}' = R\vec{x}$ from basis pursuit using (AR) as one's measurement matrix

Why random measurements are extremely useful:

- Since random measurements are incoherent with many structured measurement bases...
  - ⇒ You don't need to know what the sparse basis is before you take the data.
  - ⇒ You can "test" different measurement bases, since signals with lower sparsity can be reconstructed more quickly!

What if your signal is only approximately sparse?

What if there is noise in one's measurements?

· Theorem from (Candès, Romberg, & Tao; 2006)

If we observe $\vec{y} = A\vec{x} + \vec{\Phi}$

with $|\vec{\Phi}|_2 < \varepsilon$,

then the solution $\hat{\vec{x}}$ to

$$\min_{\vec{x}} |\vec{x}|_1 \; : \; |\vec{y} - A\vec{x}|_2 < \varepsilon$$

will satisfy

$$|\hat{\vec{x}} - \vec{x}|_2 \leq (\text{const}) \left(\varepsilon + \frac{|\vec{x} - \vec{x}_{0S}|_1}{\sqrt{S}} \right)$$

: $\vec{x}_{0S}$ = S-term approximation of $\vec{x}$

(Take S largest coefficients, and set rest to zero)

: S = largest value for which it is true that the sensing matrix satisfies the RIP.

- In short, we can bound the mean-squared-error (MSE) between our solution and the true signal, both for when our signal is only approximately sparse ($\vec{x} - \vec{x}_{0S} \neq 0$), and when our measurements are noisy ($\vec{\Phi} \neq 0$)

Conclusion

- What makes compressive sensing possible?

  1. $\vec{x}$ is sparse in a given basis (is compressible)
  2. One can safely gather info about all the coefficients of $\vec{x}$ by measuring in a basis incoherent with the sparse basis.
  3. If one's set of measurements forms a restricted isometry for vectors with at most (three times) the sparsity of $\vec{x}$, then $\vec{x}$ can be solved for exactly (with basis pursuit) even though $M < N$.
  4. If $M > (\text{const})\, S \log\left(\frac{N}{S}\right)$, then sub-Gaussian random sensing matrices satisfy the RIP with exponentially high probability.

- What can it be used for?

  - Any high dimensional sparse signal describeable as a vector in linear algebra.
    - Single pixel cameras (Double pixel cameras)
    - Matrix completion by random projections
      - Quantum state tomography.

- When is it especially useful?

  - Whenever measurements are expensive (time, energy, money)
    - Cheap IR cameras
    - High-res low light imaging
    - Medical imaging

Recap: Intro to Compressive Sensing

- What is it?
- How do we usually measure high-dimensional sparse signals?
- How can one find the S nonzero coefficients without knowing where they are?
- Why measure in a basis incoherent with the sparse basis?
- How many incoherent measurements do you need?
- Why is $\vec{x}$ uniquely determined when A is a restricted isometry for $2S$-sparse vectors?
- What sort of measurement matrices are restricted isometries?
- What if the sparse basis is different than the measurement basis?
- What if your signal is only approximately sparse?
- What if there is noise in one's measurements?

Conclusion

- What makes compressive sensing possible?
- What can it be used for?
- When is it especially useful?