Intro to Compressive Sensing

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 $\mathbf{x} \in \mathbb{R}^N$ is **compressible** in a given basis $\mathbf{W}$, where $\mathbf{W} = \mathbf{W}_1 \otimes \ldots \otimes \mathbf{W}_N$, if it can be represented in $\mathbf{W}$ with fewer than $N$ coefficients.

$$\mathbf{x} = \sum_{i=1}^{N} \mathbf{c}_i \mathbf{W}_i$$

• A signal $\mathbf{x} \in \mathbb{R}^N$ is **$S$-sparse** in $\mathbf{W}$ if it has at most $S$ nonzero coefficients in $\mathbf{W}$.

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 incoherent with the sparse basis.

Vocal

The coherence $\mu(\phi, \omega)$ between bases $\phi = \{\hat{\phi}_i\}_{i=1}^N$ and $\omega = \{\hat{\omega}_j\}_{j=1}^N$ is a measure of the maximum overlap between two sets of basis vectors,

$$\mu(\phi, \omega) = \sqrt{N} \max_{i,j} |\hat{\phi}_i \cdot \hat{\omega}_j| : \hat{\phi}_i = \frac{\phi_i}{\|\phi_i\|}$$

$\mu \in [1, N]$.

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

$$H(Q) + H(R) \geq \log(2\Omega)$$

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

- Two bases $\phi$ and $\omega$ are maximally incoherent with each other ($\mu = 1$), if any basis vector in $\phi$ is expressed as an even sum over all the basis vectors in $\omega$, and vice versa.

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

...are (3x3 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 structured basis.

- 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 vice versa.

  → 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.

\[
\tilde{y} = A \tilde{x}
\]

- \( \tilde{y} = M \times 1 \) vector of measurements
- \( A = M \times N \) "sensing matrix" (or measurement matrix)
- Each row of the sensing matrix is one of the incoherent basis vectors.

\[
\tilde{y}_{i} = \sum_{j=1}^{N} A_{ij} x_{j}
\]

- \( \tilde{x} = N \times 1 \) signal vector

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

**How Many incoherent measurements do you need?**

- If \( M = N \), then \( A \) is just an invertible linear transformation, and \( \tilde{x} = A^{-1} \tilde{y} \)

- For random measurements, one needs large enough \( M \) so that \( \text{rank}(A) = N \) (could take awhile!)
If $M<N$, then
\[ y - Ax = 0 \] has multiple solutions $(\tilde{x} : \tilde{y} = Ax)$ since this is an underdetermined system of linear equations.

However...

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

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

\[ \text{In fact, if } A \text{ is a restricted isometry for } 2S \text{-sparse vectors, and } \tilde{x} \text{ is } S \text{-sparse, then } \tilde{x} \text{ is the unique solution to} \]
\[ \min_{\tilde{x}} \| \tilde{x} \|_0 : y = A \tilde{x} \quad \text{ s.t. } \| \tilde{x} \|_0 = \lambda_0 \text{-norm} \]

(Comon, 2005)

Restricted Isometry Property (RIP)

A measurement matrix $A$ satisfies the RIP for $S$-sparse vectors with precision $\delta$ if it preserves the energy (i.e., $l_2$-norm) of any $S$-sparse vector $x$ to within a factor $(1+\delta)$. 

In short: If, for all $S$-sparse vectors $x$,
\[ (1-\delta) \| x \|_2^2 \leq \| Ax \|_2^2 \leq (1+\delta) \| x \|_2^2 \]

... then $A$ is a restricted isometry for $S$-sparse vectors with precision $\delta$.

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

$\Rightarrow$ A maps distinguishable signal vectors to distinguishable measurement vectors.

Why if $A$ satisfies the RIP for $2S$-sparse vectors, is $\tilde{x}$ uniquely determined?
Let there be two $S$-sparse vectors $\tilde{x}_1$ and $\tilde{x}_2$ such that $\tilde{y} = A\tilde{x}_1$.

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

\[(1-\delta)\|\tilde{x}_1 - \tilde{x}_2\|_2 \leq \|A(\tilde{x}_1 - \tilde{x}_2)\|_2 \leq (1+\delta)\|\tilde{x}_1 - \tilde{x}_2\|_2\]

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

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

\[0 \leq (1-\delta)\|\tilde{x}_1 - \tilde{x}_2\|_2 \leq \|\tilde{y} - A\tilde{x}_2\|_2 \leq (1+\delta)\|\tilde{x}_1 - \tilde{x}_2\|_2\]

If $\tilde{x}_1 \neq \tilde{x}_2$, then $\|\tilde{y} - A\tilde{x}_2\|_2 \geq \epsilon > 0$

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

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

Note:

$M = \#$ of rows of $A$
$M = \#$ of measurements

\[
\begin{pmatrix}
\text{(How many measurements we need to find } \tilde{x} \text{ by optimization)} \\
\text{(How many rows } A \text{ must have so that it satisfies the RIP for } 2S\text{-sparse vectors)}
\end{pmatrix}
\]

big problem!

Actually solving

\[
\min_{\tilde{x}} \|\tilde{x}\|_0 \quad : \quad \tilde{y} = A\tilde{x}
\]

is computationally intractable

$\Rightarrow$ one would have to check $\frac{N!}{(N-2S)!S!}$ possible decompositions of $\tilde{x}$
Easy Solution

If $A$ satisfies RIP for 25-sparse vectors, then $\hat{x}$ is the solution to

$$
\min \limits_{x} \|Ax - y\|_2^2 = \|Ax - \bar{y}\|_2^2 \\
\text{basis pursuit}
$$

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

What sort of measurement matrices satisfy the RIP for 25-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: $A_i \sim N(0, \frac{1}{M})$

   $$
P(\|x\|_2^2 < 16\|x\|_2^2) \geq \frac{1}{1 - 2e^{c\cdot\log(N/M) - M/8}}
   $$

   - Use Markov inequality to get this
   - $P(x > t) \leq \frac{E[e^{tx}]}{e^{ct}}$

   - Where $M = (\text{const}) \cdot \frac{\log(N/M)}{\log}$

   - $P(x > t) \leq e^{-ct^2}$

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

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

\[ \mathbf{A}_{13} = \text{ith entry of jth randomly drawn basis vector of incoherent basis.} \]

\[ P((1-\varepsilon)\mathbf{x}_i^2 \leq \mathbf{A}_x^2 \leq (1+\varepsilon)\mathbf{x}_i^2) \Rightarrow I \text{ quickly where} \]

\[ M \geq (\text{const}) \cdot S(\log(N)) \]

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

\[ \mathbf{\tilde{y}} = \mathbf{A} \mathbf{x}' = \mathbf{A} (\mathbf{Rx}) = (\mathbf{AR}) \mathbf{x} \]

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

Why random measurements are extremely useful:

- Since random measurements are incoherent with many structured measurement bases.

  \[ \Rightarrow \text{You don't need to know what the sparse basis is before you take the data.} \]

  \[ \Rightarrow \text{You can "test" different measurement bases, since signals with lower sparsity can be reconstructed more quickly!} \]
What if your signal is only approximately sparse?
What if there is noise in one's measurements?

Theorem from (Candes, Romberg, & Tao 2006)

If we observe \( \tilde{y} = A\tilde{x} + \tilde{\phi} \)
with \( \|\tilde{\phi}\|_2 < \epsilon \),
then the solution \( \hat{x} \) to
\[
\min_x \|x\|_1 : \|\tilde{y} - A\hat{x}\|_2 < \epsilon
\]
will satisfy
\[
\|\hat{x} - \bar{x}\|_2 \le (\text{const}) \left( \sqrt{\epsilon + \frac{1\|\bar{x} - \tilde{x}_{os}\|_1}{\sqrt{s}}} \right)
\]
\[
\text{where } \tilde{x}_{os} = s\text{-term approximation of } \bar{x}
\]
(Take s largest coefficients, and set rest to zero)

\[
\text{or } s = \text{largest value for which it is true that the sensing matrix satisfies the RIP.}
\]

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 \( \bar{x} \) \( \sim \tilde{x}_{os} \neq 0 \), and when our measurements are noisy \( \tilde{\phi} \neq 0 \).
Conclusion

What makes compressive sensing possible?

1. $\tilde{x}$ is sparse in a given basis (is compressible)
2. One can safely gather info about all the coefficients of $\tilde{x}$ 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 $\text{(three times)}$ the sparsity of $\tilde{x}$, then $\tilde{x}$ can be solved for exactly (with basis pursuit) even though $M < N$.
4. If $M \geq (\text{const}) S \log(\frac{N}{S})$, then sub-Gaussian random sensing matrices satisfy the RIP with exponentially high probability.

What can it be used for?

- Any high dimensional sparse signal describable as a vector in linear algebra:
  - Single pixel cameras (Doublepixel 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 5 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 $x$ uniquely determined when $A$ is a restricted isometry for 25-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?