Sparse solution of linear systems:

- Given a matrix \( \Phi \) of size \( m \times n \) with \( m \leq n \), define:

\[ R_k = \{ \Phi x \mid x \in \mathbb{R}^n \text{ s.t. } \| x \|_0 \leq k \} \]

where \( \| x \|_0 = \text{number of nonzero entries of } x \)

- Problem of interest: \( \min \| x \|_0 \text{ s.t. } \Phi x = y \) where \( y \in R_k \)

- We aim to find a sparse solution of \( y = \Phi x \).

- Naive strategy if \( \Phi \) is full rank:
  - Choose \( m \) columns of \( \Phi \) and form a matrix \( A = [\phi_1, \phi_2, \ldots, \phi_m] \).
  - Solve \( y = Az \) if \( A \) is non-singular.
  - Among all possible \( A \)'s, find the one leading to the sparsest \( z \).
  - Design \( y \) by adding 0 to \( z \).

- This is not a good strategy because if \( n = 2m \), we need to solve about \( 2^n \) equations.

- If \( m = 512 \) and \( n = 1024 \) \( \Rightarrow \) solve at least \( 2^{512} \) systems of size \( 512 \times 512 \).

- Question: How to find a good algorithm?

- Motivations:

  1. Compressed sensing: economically recording information
     about a vector \( x \).
     - \( \text{data} = \text{signal, image, ...} \)
     - \( x \to \Phi \to A \to \hat{x} \) where \( \hat{x} = x \)
     - \( x \): compressible data, \( \text{goal: design } (\Phi, A) \)
2. Error Correcting Codes:

\[ x \rightarrow \begin{array}{c} R \rightarrow Z \\ A \end{array} \]

\[ Z = Ax \] where \( A \in \mathbb{R}^{m \times n} \) and \( m > n \)

- The channel corrupts some random entries of \( Z \).

- The received signal is \( W = Z + V \) \( \text{sparse signal; need to correct the errors.} \)

- Define \( B = \begin{bmatrix} A & A^T \end{bmatrix} \) such that \( BB^T = I \)

\[ \Rightarrow B^TW = B^TZ + B^TV = \begin{bmatrix} X + A^TV \\ (A^T)^TV \end{bmatrix} \]

- Define \( y = (A^T)^TV \) \( \Rightarrow \) Given \( y \) and \( A^T \), find the sparsest vector \( V \).

- Other motivations: Cryptography, recovery of lost data, ...

- Convex relaxation:

\[ \min_x \|x\|_0 \quad \text{s.t.} \quad \Phi x = y \]

\[ \Rightarrow \min_x \|x\|_1 \quad \text{s.t.} \quad \Phi x = y \]

- Two techniques to study the relaxation: 1. Mutual coherence, 2. Restricted Isometric Property (RIP)

- Mutual coherence:

- Define spark of \( A \), \( \text{spark}(A) \), as the minimum number of linearly dependent columns of \( A \).

\[ \Rightarrow \text{spark}(A) \leq \text{rank}(A) + 1 \]

- Thm: If \( y = \Phi x^* \) and \( \|x\|_0 < \text{spark}(\Phi)/2 \), then \( y \) \( \text{is the sparsest solution.} \)
Proof: Let \( \tilde{x} \) be another solution such that 
\[ \| \tilde{x} \|_0 \leq \| x^* \|_0. \] 
Then,
\[ y = \Phi x^* = \Phi \tilde{x} \Rightarrow \Phi (x^* - \tilde{x}) = 0 \]
and 
\[ \| x^* - \tilde{x} \|_0 \leq \| x^* \|_0 + \| \tilde{x} \|_0 < \text{spark } (\Phi) \]
but \( \| x^* - \tilde{x} \|_0 \) columns of \( \Phi \) must be linearly dependent. \( \Rightarrow \) Contradiction.

By-product: The sparsest solution is unique.

---

Assume that each column of \( \Phi \) is normalized (\( \Phi \) is called a dictionary)

\[ M(\Phi) = \max_{\{i,j \leq n \}} |(\Phi^T \Phi)_{ij}| \]

Mutual coherence

- Property: \( M(\Phi) \leq 1 \)
- Goal: Make \( M(\Phi) \) small.
- Fact: \( M(\Phi) \geq (2^m)^{-\frac{1}{2}} \) if \( n \geq 2m \).

Thm: \( \text{spark } (\Phi) > \frac{1}{M(\Phi)} \)

Main Theorem: Consider the optimization

\[ \min \| x \|_1 \text{ s.t. } \| \Phi x - y \|_2 \leq \varepsilon \]

Where \( y = \Phi x_0 + \varepsilon \) with \( \| \varepsilon \|_2 \leq \varepsilon < \varepsilon \) and \( \| x_0 \|_0 = k \).

(\( \varepsilon \): measurement error, \( \varepsilon \): accuracy of solution)

Denote the solution as \( \hat{x}_0 \). If

\[ k < \frac{1}{M(4k-1)} \]

\[ \Rightarrow \| \hat{x}_0 - x_0 \|_2^2 \leq \frac{(\varepsilon + \delta)^2}{1 - M(4k-1)} \]
- Restricted Isometry Property:

- Given an index set $T \subset \{1, 2, \ldots, n\}$, let $A_T$ be a matrix consisting of those columns of $A$ with indices in $T$.

- Definition of $k$ restricted isometry constant $\delta_k$:
  
  Smallest number $\delta_k > 0$ (if exists) such that
  
  $$
  (1 - \delta_k) \|x\|_2^2 \leq \|A_T x\|_2^2 \leq (1 + \delta_k) \|x\|_2^2 \quad \forall x
  $$

  for all $T$'s satisfying $\#(T) \leq k$.

- Thm: Consider $y = \Phi x_0$ with $\|x_0\|_0 \leq k$. If $\delta_{2k} < 1$, then $x_0$ is the sparsest solution of $y = \Phi x$.

- Proof: If $x_0$ and $\tilde{x}_0$ are both sparse, then:
  
  $\Phi(x_0 - \tilde{x}_0) = 0$

  $$
  \implies (1 - \delta_{2k}) \|x_0 - \tilde{x}_0\|_2^2 \leq \|\Phi_T (x_0 - \tilde{x}_0)\|_2^2 = 0
  $$

  where $y_0$ and $\tilde{y}_0$ are the non-zero parts of $x_0$ and $\tilde{x}_0$, and $T$ contains the indices of non-zero entries of $x_0 - \tilde{x}_0$.

- Main Thm: If $\delta_{3k} + 3 \delta_{4k} < 2$, then the convex relaxation is exact.

- How to find an RIP matrix $\Phi$?

- Thm: Let $A$ be a matrix with entries being iid Gaussian random variables with zero mean and variance $\frac{1}{\sqrt{m}}$.
  Then for any $\epsilon > 0$, we have:
\[ P \left( |\|\mathbf{A}\|_2^2 - \|\mathbf{x}\|_2^2| < \varepsilon \|\|\mathbf{x}\|_2^2 \right) \geq 1 - 2\varepsilon e^{-\frac{3m}{2}} \quad \forall \mathbf{x} \]

where \( \varepsilon \) is a positive number independent of \( \varepsilon \) and \( \|\mathbf{x}\|_2 \).

\[ \Rightarrow \text{ when } m \text{ is large enough, we can find an RIP matrix.} \]

- **Summary:** \( \mathbf{y} = \Phi \mathbf{x}_0 \) and assume \( \Phi = \text{full rank} \)
  
  if \( \|\mathbf{x}_0\| \leq \frac{m}{\varepsilon} \) \( \Rightarrow \) \( \mathbf{y} = \Phi \mathbf{x} \) has a unique solution \( \text{sparse} \)

\[ \text{spark}(\Phi) = m + 1 \]

- **Convex relaxation** \( \min_{\|\mathbf{x}\|_1} \text{ s.t. } \mathbf{y} = \Phi \mathbf{x} \)

  **Thm:** The relaxation is exact with a high probability in a random case if \( m \geq \text{constant} \cdot k \log \frac{n}{k} \)