1. Sparsity in Signal/Image Processing
- Parsimony in signal representation has value in many applications.
- A large class of signals can be represented in a compact manner with respect to some basis (dictionary).
- Recent advances using sparse representation for a variety of applications:
  - A measurement (dictionary) matrix $A \in \mathbb{R}^{m \times n}$
  - Compressive Sensing framework ($k \in \mathbb{R}^n$):
    $$x = \arg \min_{x} |y - Ax|_2 \leq \epsilon$$
- The $\ell_0$ relaxation of regularized problem:
  $$x = \arg \min_{x} |y - Ax|_2 + \lambda|x|_1$$

2. Motivation and Idea
- Model-based Compressive Sensing:
  $$\max f(x) \text{ subject to } |y - Ax|_2 < \epsilon$$

  Laplacian MAP estimation
  $$\arg \max_{x} \exp \left(-\lambda \sum_{i=1}^{p} |x_i| \right) + \arg \min_{x} |x|_1$$

- Introduce a prior $f(x)$ which can capture the sparse structure
- Spike-and-Slab Priors: The gold standard to induce sparsity [1]
- Goal: Find the sparse signal $x \in \mathbb{R}^p$ from a set of fewer measurements $y \in \mathbb{R}^q$ ($q < p$)

3. Bayesian Interpreation

Bayesian formulation
$$y|A, x, \gamma, \sigma^2 \sim \mathcal{N}(Ax, \sigma^2 I),$$
$$x|\gamma, \lambda, \sigma^2 \sim \prod_{i=1}^{p} \mathcal{N}(0, \sigma^2 \lambda^{-1}) + (1 - \gamma_i) \mathcal{N}(x_i = 0),$$
$$\gamma \sim \prod_{i=1}^{p} \text{Bernoulli}(\nu_{i}),$$

- The optimal $x^*, \gamma^*$ are obtained by MAP estimation on joint posterior density:
  $$(x^*, \gamma^*) = \arg \max_{x, \gamma} \{p(x, \gamma|A, y, K, \lambda, \sigma^2)\}.$$  

Optimization problem [2]
$$\left(x^*, \gamma^*\right) = \arg \min_{x, \gamma} |y - Ax|_2^2 + \lambda |x|_1 + \sum_{i=1}^{p} \rho_i \gamma_i,$$  

$$\text{subject to } \gamma_i \in \{0, 1\}$$  

4. Existing Solutions/Special Cases
- Markov Chain Monte Carlo (MCMC algorithm) [3]
- Majorization minimization (MM algorithm) [4] by assuming only one $\kappa$
  $$\arg \min_{x} |y - Ax|_2^2 + \lambda |x|_2^2 + |x|_1$$
- Adaptive Elastic Net [5] by relaxing $\ell_0$ norm to $\ell_1$ norm
  $$\arg \min_{x} |y - Ax|_2^2 + \lambda |x|_2^2 + |x|_1$$
- Sparse Reconstruction by Separable Approximation (SpaRSA) [6]

5. AMP: Adaptive Matching Pursuit - First observations
- First, let $D = \frac{A}{\sqrt{\lambda T}}$ and $z = \frac{y}{\epsilon}$ with $I \in \mathbb{R}^{p \times p}$ and $p \in \mathbb{R}^{n \times 1}$ being the identity matrix and zero vector, we can rewrite (2) as:
  $$(x^*, \gamma^*) = \arg \min_{x, \gamma} |y - Dx|_2^2 + \sum_{i=1}^{p} \rho_i \gamma_i.$$  

where $|d_i|_2^2 = |a_i|_2^2 + \lambda + 1, \forall i = 1, \ldots, p$

- If we know the true support of the signal, i.e. $S = \{i : \gamma_i \neq 0\}$, we can infer the solution of (3) by:
  $$x_S = \arg \min_{x_S} |x - D_S x_S|_2^2 = \frac{D_S^T D_S x_S}{\mathbf{1}_S{D_S^T D_S} \mathbf{1}_S} = w_S$$
  $$r_S = z - D_S x_S$$

6. AMP: The central idea

Define: $g(S) = \min_{S} |x - D_S x_S|_2^2 + \sum_{i\in S} \rho_i$, best improvement if we insert: $U_S = \min_{S \subseteq \theta} g(S \cup \{i\}) - g(S)$, best improvement if we remove: $V_S = \min_{S \subseteq \theta} g(S \setminus \{i\}) - g(S)$.

- If $U_2, V_2 > 0$ stop. If $U_2 < V_2 \Rightarrow$ insert, else remove.

7. AMP: Supporting lemmas and Cholesky decomposition

- **Lemma 1**: Initialization of the support set $S$
  If $\kappa_i < \epsilon$ ($\kappa_i > \epsilon$, then $i \in \hat{S}$ – the optimal active set $\Rightarrow$ initialize $S_0 = \{i : \kappa_i < \epsilon\}$

- **Lemma 2**: Approximation of $U_S$
  $$U_S \leq \min_{i \in \hat{S}} [\rho_i - |r_i|_2^2] \frac{1}{3} + \frac{1}{4}$$

- **Lemma 3**: Approximation of $V_S$
  $$V_S \leq \min_{i \in \hat{S}} [\{1 + \lambda (|r_i|_2^2 + 2|r_i|_2^2 - \rho_i\}] \frac{1}{4}$$

- **Challenge**: Once $S$ is found, we need to solve: $D_S^2 D_S x_S = w_S$ which may be expensive.

- **Solution**: Use the Cholesky decomposition: $D_S^2 D_S = L_L L_L^T$, where $L_L$ is a low triangular matrix.

8. AMP: Supporting lemmas and Cholesky decomposition

- **Forward Cholesky decomposition**
  If $y$ is the solution of $L_L^T y = L_L x_S$, via forward substitution, then:
  $$L_L^T y = \begin{bmatrix} \lambda^2 + \lambda + 1 & 0 \\ \lambda^2 + \lambda + 1 & \lambda^2 + \lambda + 1 \end{bmatrix} L_L x_S$$

- **Backward Cholesky decomposition**
  If $L L^T = \begin{bmatrix} L_0 & 0 \\ 0 & L_0 \end{bmatrix}$, then:
  $$L L^T y = \begin{bmatrix} 0 & 0 \\ 0 & \lambda^2 + \lambda + 1 \end{bmatrix} L_0 y$$


9. Experiments with simulated data
- 1000 realizations of $A$, sparse $x_0$ and Gaussian noise $\varepsilon$: $y = Ax_0 + \varepsilon$
- Coefficient vector $x_0 \in \mathbb{R}^{1024}$. Test vector $y \in \mathbb{R}^{256}$. Sparsity level = 100.

10. Experiments with real data
- Reconstruction of handwritten digit images from the MNIST dataset.
- MNIST contains 60000 digits images (0 to 9) of size 28x28 pixels.
- Sparse signal $x_0$ (vectorized image) is to be reconstructed from a smaller set of random measurements $y = Ax_0 + \varepsilon$.
- Coefficient vector $x_0 \in \mathbb{R}^{784}$. Measured vector $y \in \mathbb{R}^{256}$.