Bayesian Compressive Sensing with Belief Propagation

Based on the work of: Richard G. Baraniuk *et. al.* (Rice Univ.)

April 30, 2012

Compressed Sensing

Noiseless Measurements

Consider an *under-determined* measurement equation

$$\underline{Y}_{m\times 1} = \Psi_{m\times n} \cdot \underline{X}_{n\times 1}, \tag{1}$$

where $m < n$, and $||\underline{X}||_{\ell_0} = k$ (called k-sparse).

Compressed Sensing

Noiseless Measurements

Consider an *under-determined* measurement equation

$$\underline{Y}_{m\times 1} = \Psi_{m\times n} \cdot \underline{X}_{n\times 1}, \tag{1}$$

where $m < n$, and $||\underline{X}||_{\ell_0} = k$ (called k-sparse).

CS theory guarantees:

If Ψ is appropriately chosen (*e.g.* sub-Gaussian), $\underline{x}$ (realization of $\underline{X}$) can be *uniquely* recovered from $\underline{y}$, *without any distortion*:

$$\ell_0\text{-min decoding:} \quad \hat{\underline{x}} = \arg\min_{\underline{x}'} \left|\left|\underline{x}'\right|\right|_{\ell_0}, \quad s.t.\ \Psi \cdot \underline{x}' = \underline{y} \tag{2}$$

$$\ell_1\text{-min decoding:} \quad \hat{\underline{x}} = \arg\min_{\underline{x}'} \left|\left|\underline{x}'\right|\right|_{\ell_1}, \quad s.t.\ \Psi \cdot \underline{x}' = \underline{y} \tag{3}$$

Noisy Compressed Sensing

Noisy Measurements

Associates an additive noise in the measurements:

$$\underline{Y}_{m\times 1} = \Psi_{m\times n} \cdot \underline{X}_{n\times 1} + \underline{N}_{m\times 1}, \tag{4}$$

Noisy Compressed Sensing

Noisy Measurements

Associates an additive noise in the measurements:

$$\underline{Y}_{m\times 1} = \Psi_{m\times n} \cdot \underline{X}_{n\times 1} + \underline{N}_{m\times 1}, \tag{4}$$

Noisy CS Recovery:

- Convex Optimization: ℓ_1-regularization,
- Greedy Algorithms: MP, OMP, StOMP, *etc*,
- Combinational Algorithms,
- Bayesian.

Bayesian Compressed Sensing (BCS)

Consider when $\mathbf{p}_{\underline{X}}(.)$ is known:

$$\underline{Y}_{m\times 1} = \Psi_{m\times n} \cdot \underline{X}_{n\times 1} + \underline{N}_{m\times 1}$$

Optimal Decoding

$$\hat{\underline{x}}_i = \mathbf{E}\left[\underline{X}_i \middle| \underline{Y} = \underline{y}\right], \ 1 \leq i \leq n \tag{5}$$

Bayesian Compressed Sensing (BCS)

Consider when $\mathbf{p}_{\underline{X}}(.)$ is known:

$$\underline{Y}_{m\times 1} = \Psi_{m\times n} \cdot \underline{X}_{n\times 1} + \underline{N}_{m\times 1}$$

Optimal Decoding

$$\hat{\underline{x}}_i = \mathbf{E}\left[\underline{X}_i \middle| \underline{Y} = \underline{y}\right],\ 1 \leq i \leq n \tag{5}$$

This requires the decoder to calculate $\mathbf{p}_{\underline{X}}(\underline{x}|\underline{Y} = \underline{y})$.

Bayesian Compressed Sensing (BCS)

Consider when $\mathbf{p}_{\underline{X}}(.)$ is known:

$$\underline{Y}_{m\times 1} = \Psi_{m\times n} \cdot \underline{X}_{n\times 1} + \underline{N}_{m\times 1}$$

Optimal Decoding

$$\hat{\underline{x}}_i = \mathbf{E}\left[\underline{X}_i \middle| \underline{Y} = \underline{y}\right], \; 1 \le i \le n \tag{5}$$

This requires the decoder to calculate $\mathbf{p}_{\underline{X}}(\underline{x}|\underline{Y} = \underline{y})$.

Motivated by LDPC coding:

However, if Ψ **has a lot of zeros**, then computationally simple decoding may be possible.

Encoding by Sparse Matrix

$$\underline{Y}_{m\times 1} = \mathbf{\Psi}_{m\times n} \cdot \underline{X}_{n\times 1} + \underline{N}_{m\times 1}$$

Opposed to the conventional *dense* CS measurement matrices (*e.g.* sub-Gaussian), BCS measurement matrices have **low density**, which is the key to **fast** encoding and decoding.

Encoding by Sparse Matrix

$$\underline{Y}_{m\times 1} = \boldsymbol{\Psi}_{m\times n} \cdot \underline{X}_{n\times 1} + \underline{N}_{m\times 1}$$

Opposed to the conventional *dense* CS measurement matrices (*e.g.* sub-Gaussian), BCS measurement matrices have **low density**, which is the key to **fast** encoding and decoding.

Rice university CS group work:

- they adopt CS-LDPC Ψ with -1,1 non-zero entries,
- they *prove* possibility of robust decoding for tw-state Gaussian $\underline{X}$'s, under some conditions. $m = O\Big(k\log(n)\Big)$ measurements are sufficient.

Belief Propagation (BP) based Decoding

They consider a *two-state mixture Gaussian* prior:

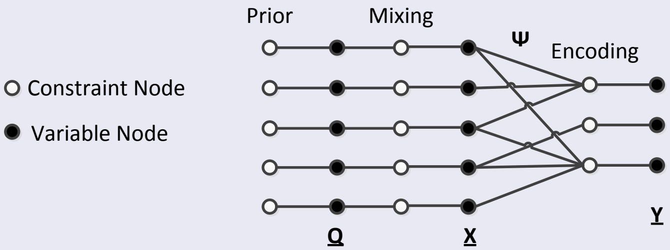
- For each $1 \leq i \leq n$, $Q_i \sim$ **Bernouli**$(\frac{k}{n})$
- if $Q_i = 1 : X_i \sim \mathcal{N}(0, \sigma_1^2)$,
- if $Q_i = 0 : X_i \sim \mathcal{N}(0, \sigma_0^2)$, where $\sigma_1^2 \gg \sigma_0^2$.

Belief Propagation (BP) based Decoding

They consider a *two-state mixture Gaussian* prior:

- For each $1 \leq i \leq n$, $Q_i \sim$ **Bernouli**$(\frac{k}{n})$
- if $Q_i = 1 : \; X_i \sim \mathcal{N}(0, \sigma_1^2)$,
- if $Q_i = 0 : \; X_i \sim \mathcal{N}(0, \sigma_0^2)$, where $\sigma_1^2 \gg \sigma_0^2$.

Factor Graph for CS-BP

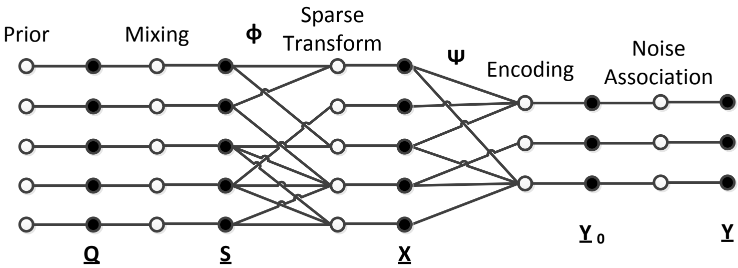


A Practical Case

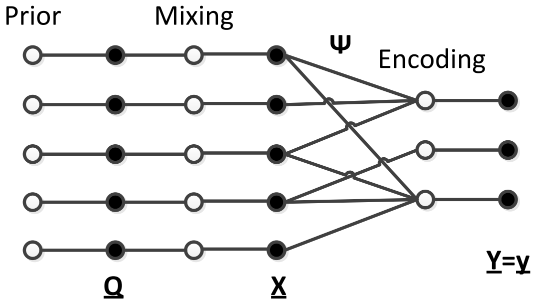
$\underline{X}$ is sparse in ϕ domain:

$$\underline{Y} \quad = \quad \underline{Y}_0 + \underline{N} = \Psi \cdot \underline{X} + \underline{N} = \Psi\ \phi \cdot \underline{S} + \underline{N}$$

$$\mathbf{E}[||\underline{S}||_{\ell_0}] \simeq k$$



Message Passing in CS-BP



Message Passing:

Types of messages:

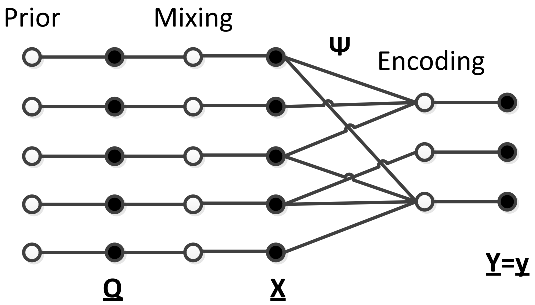
- $\mu_{v\to c}(v)$
- $\mu_{c\to v}(v)$

Updating Messages: Multiplication and Convolution of Beliefs

$$\mu_{v\to c}(v) = \prod_{u\in neib(v)-\{c\}} \mu_{u\to v}(v)$$

$$\mu_{c\to v}(v) = \sum_{\sim\{v\}} \left(cons(neib(c)) \prod_{w\in neib(c)-\{v\}} \mu_{w\to c}(w) \right)$$

Message Passing in CS-BP



$$f(v) = \prod_{u \in neib(v)} \mu_{u \to v}(v)$$

CS-BP is numerically shown to converge, if appropriate Ψ and messages are used.