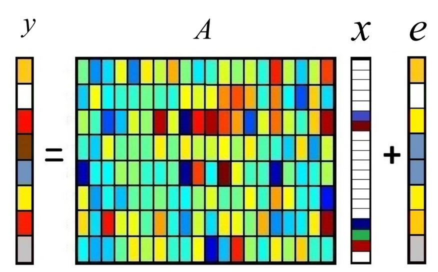
CAMP: An Algorithm to Recover Sparse Signals with Unknown Clustering Pattern Using Approximate Message Passing

Mohammad Shekaramiz, Todd K. Moon, and Jacob H. Gunther

Information Dynamics Laboratory, ECE Dept., Utah State University

Introduction

- Single measurement vector (SMV) model:
  - $\mathbf{y} = A\mathbf{x}_s + \mathbf{e}$: linear non-adaptive noisy measurements
  - $\mathbf{x}_s \in \mathbb{R}^{N\times 1}$ is the sparse signal of interest
  - $\mathbf{y} \in \mathbb{R}^{M\times 1}$ contains measurements ($M \ll N$)
  - solving the linear inverse problem to find $\mathbf{x}_s$

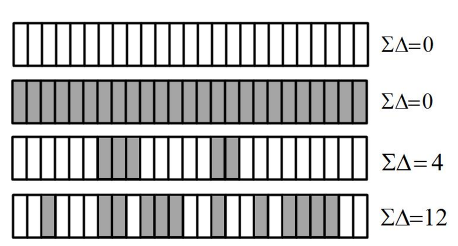


- **Assumption:**
  - *Sparse clustered pattern:* Non-zero elements of $\mathbf{x}$ may appear in clusters with an unknown structure
- **Proposed model:** $\mathbf{y} = A(\mathbf{s} \circ \mathbf{x}) + \mathbf{e}$

  $\mathbf{y} \in \mathbb{R}^{M\times 1}$, $A \in \mathbb{R}^{M\times N}$, $\mathbf{s} \in \{0,1\}^{N\times 1}$, $\mathbf{x} \in \mathbb{R}^{N\times 1}$, $\mathbf{e} \in \mathbb{R}^{M\times 1}$ ($M \ll N$)
  - $\mathbf{s}$ is the support learning vector and accounts for the non-zero locations of $\mathbf{x}$
  - In $(\mathbf{s} \circ \mathbf{x})$, $\circ$ denotes Hadamard product
- **Objective:**
  - Learning the sparsity pattern of $\mathbf{x}$
  - Recovering sparse signal $\mathbf{x}$ using the noisy SMV model
- **Proposed algorithm:**
  - *CAMP:* Algorithm to recover sparse signals with unknown clustering pattern using approximate message passing framework

Proposed Statistical Model and Defining Priors

- Measure of clumpiness [2]: $(\Sigma\Delta)_{\mathbf{s}} = \sum_{i=2}^{P} |s_i - s_{i-1}|$, where $\mathbf{s}$ is the support learning vector of the solution
  - There exist few transitions for the case where the supports of the solution have a clustered pattern
  - For example, a constant vector (all ones or all zeros) has a $\Sigma\Delta$ of 0
  - More examples:



- Change in the measure of clumpiness

$$(\Sigma\Delta)_{(\text{support of } \mathbf{x})} = \sum_{n=2}^{N} |b(x_n, T) - b(x_{n-1}, T)|,$$

  - T: a predetermined threshold
  - $b(.,.)$: returns a binary value

$$b(x_n, T) = \begin{cases} 1 & \text{if } |x_n| > T \\ 0 & \text{otherwise.} \end{cases}$$

- Prior on the solution vector $\mathbf{x}$:

$$\begin{aligned} &\forall n = 1, \ldots, N, \\ &x_n \sim \mathcal{N}(0, \alpha_n) \\ &\alpha_n \sim \mathcal{N}(e^{\left\{\frac{(\Sigma\Delta)|_{b(x_n,.)=0} - (\Sigma\Delta)|_{b(x_n,.)=1} - 1}{\theta_1}\right\}}, \theta_2), \end{aligned}$$

  - $(\Sigma\Delta)|_{b(x_n,.)=1}$: Sigma-Delta evaluation of the supports of the solution when s_n is set to be active
  - θ_1: A tuning parameter for the emphasis on the measure of clumpiness
  - θ_2: The prior variance on the variance of the variable x_n and is updated via the EM algorithm
- Joint probability distribution of the model:

$$P(\mathbf{y}, \mathbf{x}, \boldsymbol{\alpha}, \theta_1, \theta_2, \sigma^2) \propto$$
$$P(\mathbf{y}|\mathbf{x}, \sigma^2 I_N) \prod_{n=1}^{N} \Big(P(x_n; 0, \alpha_n) P(\alpha_n; e^{\left\{\frac{(\Sigma\Delta)|_{b(x_n,.)=0} - (\Sigma\Delta)|_{b(x_n,.)=1} - 1}{\theta_1},\right\}}, \theta_2) \Big)$$

- The measurement noise is assumed to be $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 I_N)$

Justification of the Priors

- CAMP algorithm adds an additional layer to AMP-SBL algorithm [1] using the described priors to encourage the clustered pattern
- To encourage sparsity, we assumed $x_n \sim \mathcal{N}(0, \alpha_n)$ as a prior
- The supports of the solution are then specified by the function $b(., T)$ where T is a predetermined threshold
- Based on the threshold, we discard the small-valued components of $\mathbf{x}$ from being considered as the support of the solution
- The smaller α_n is, the higher probability it provides to x_n becoming 0 i.e. (discarding s_n)

Behavior of α_n with respect to Sigma-Delta

- The reason behind assuming such prior on α_n:

$(\Sigma\Delta)\|_{b(x_n,.)=1}$	$(\Sigma\Delta)\|_{b(x_n,.)=0}$	α_n
cte	*cte*	↓
↑	*cte*	↓
cte	↓	↓
cte	↑	↑
↓	*cte*	↑

- For example consider the case where forcing either $s_n = 0$ or $s_n = 1$ does not make any change in the evaluation of Sigma-Delta
- In this case, though it promotes the clumpiness in the solution, it discourages the solution to be sparse
- Therefore, α_n needs to be decreased

CAMP Algorithm

CAMP:
Solver of linear inverse SMV problem for the clustered sparse signals:

- *Definitions*

$F_n(k_n, \alpha_n, c) = k_n \frac{\alpha_n}{c+\alpha_n}$
$G_n(\alpha_n, c) = \frac{c.\alpha_n}{c+\alpha_n}$
$F'_n(\alpha_n, c) = \frac{\alpha_n}{c+\alpha_n}$

- *Message updates using AMP*

For $n = 1, 2, \ldots, N$
  $k_n = \sum_{m=1}^{m} a^{\star}_{mn} z_m + \mu_n$
  $\mu_n = F_n(k_n, \alpha_n, c)$
  $\nu_n = G_n(\alpha_n, c)$
End
$c = \sigma^2 + \frac{1}{M}\sum_{n=1}^{N} \nu_n$
$z_m = y_m - \sum_{n=1}^{N} a_{mn}\mu_n + \frac{z_m}{M}\sum_{n=1}^{N} F'_n(\alpha_n, c), \forall m = 1, \ldots, M$

- *Parameter updates using EM algorithm*

% Updating $\boldsymbol{\alpha}$:
$\forall n = 1, 2, \ldots, N$, solve for α_n in
$\alpha_n^3 - e^{\left\{\frac{(\Sigma\Delta)|_{b(x_n,.)=0} - (\Sigma\Delta)|_{b(x_n,.)=1} - 1}{\theta_1}\right\}} \alpha_n^2 + \frac{\theta_2}{2}\alpha_n - \frac{\theta_2(\mu_n^2 + \nu_n)}{2} = 0$
which is the minimizer of

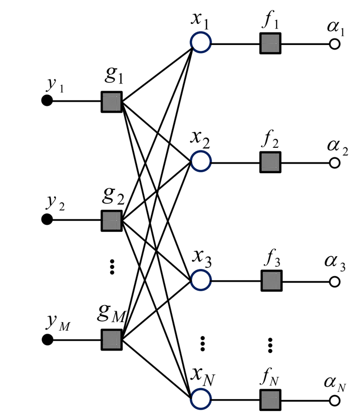
$$f(\alpha_n) = \ln(\alpha_n) + \frac{\mu_n^2 + \nu_n}{\alpha_n} + \frac{1}{\theta_2}(\alpha_n - e^{\left\{\frac{(\Sigma\Delta)|_{b(x_n,.)=0} - (\Sigma\Delta)|_{b(x_n,.)=1} - 1}{\theta_1}\right\}})^2$$

% Updating the noise variance σ^2:
$\sigma^{2[k+1]} = \frac{\|\mathbf{y} - A\boldsymbol{\mu}\|_2^2 + \sum_{n=1}^{N} \|\mathbf{a}_n\|_2^2 \nu_n}{M}$
% Updating the variance of $\boldsymbol{\alpha}$:
$\theta_2^{[k+1]} = \frac{1}{N}\sum_{n=1}^{N} \left(\alpha_n - e^{\left\{\frac{(\Sigma\Delta)|_{b(x_n,.)=0} - (\Sigma\Delta)|_{b(x_n,.)=1} - 1}{\theta_1}\right\}}\right)^2$

Factor Graph

- Under such modeling, all the distributions of the joint, conditional, and posterior densities become Gaussian

$$\begin{aligned} g_m &:= P(y_m|\mathbf{x}, \boldsymbol{\alpha}), \quad m = 1, 2, \ldots, M \\ f_n &:= P(x_n; 0, \alpha_n), \quad n = 1, 2, \ldots, N \end{aligned}$$



Message Passing

- Message from a function node to a variable node

$$M_{g_m \to x_n} \propto \mathcal{N}(a_{mn}x_n; z_{mn}, c_{mn}),$$

where

$$z_{mn} = y_m - \sum_{q \neq n} a_{mq}\mu_q, \text{ and } c_{mn} = \sigma^2 + \sum_{q \neq n} |a_{mq}|^2 \Sigma_q$$

- Message from a variable node to a function node

$$M_{x_n \to g_m} \propto \mathcal{N}\big(x_n; \sum_{l \neq m} a_{nl}z_{ln} \frac{\alpha_n}{c_n + \alpha_n \sum_{l \neq m} a_{ln}^2}, \frac{c_n\alpha_n}{c_n + \alpha_n \sum_{l \neq m} a_{ln}^2}\big)$$

- c_{nl} (under the large-system-limit) is approximated by

$$c_{nl} \simeq c_n := \frac{1}{M}\sum_{m=1}^{M} c_{mn}.$$

- Assumption: Matrix A is normalized with respect to its columns

$$\sum_{l \neq m} a_{ln}^2 \simeq \sum_{m=1}^{M} a_{mn}^2 = 1.$$

  Therefore,

$$M_{x_n \to g_m} \propto \mathcal{N}\big(x_n; \sum_{l \neq m} a_{nl}z_{ln} \frac{\alpha_n}{c_n + \alpha_n}, \frac{c_n\alpha_n}{c_n + \alpha_n}\big).$$

- Estimating the posterior on x_n

$$P(x_n|\mathbf{y}) \propto P(x_n; \alpha_n) \prod_{m=1}^{M} P(y_m|x_n)$$
$$\propto M_{f_n \to x_n} \prod_{m=1}^{M} M_{g_m \to x_n}) \propto \mathcal{N}(x_n; \mu_n, \nu_n)$$

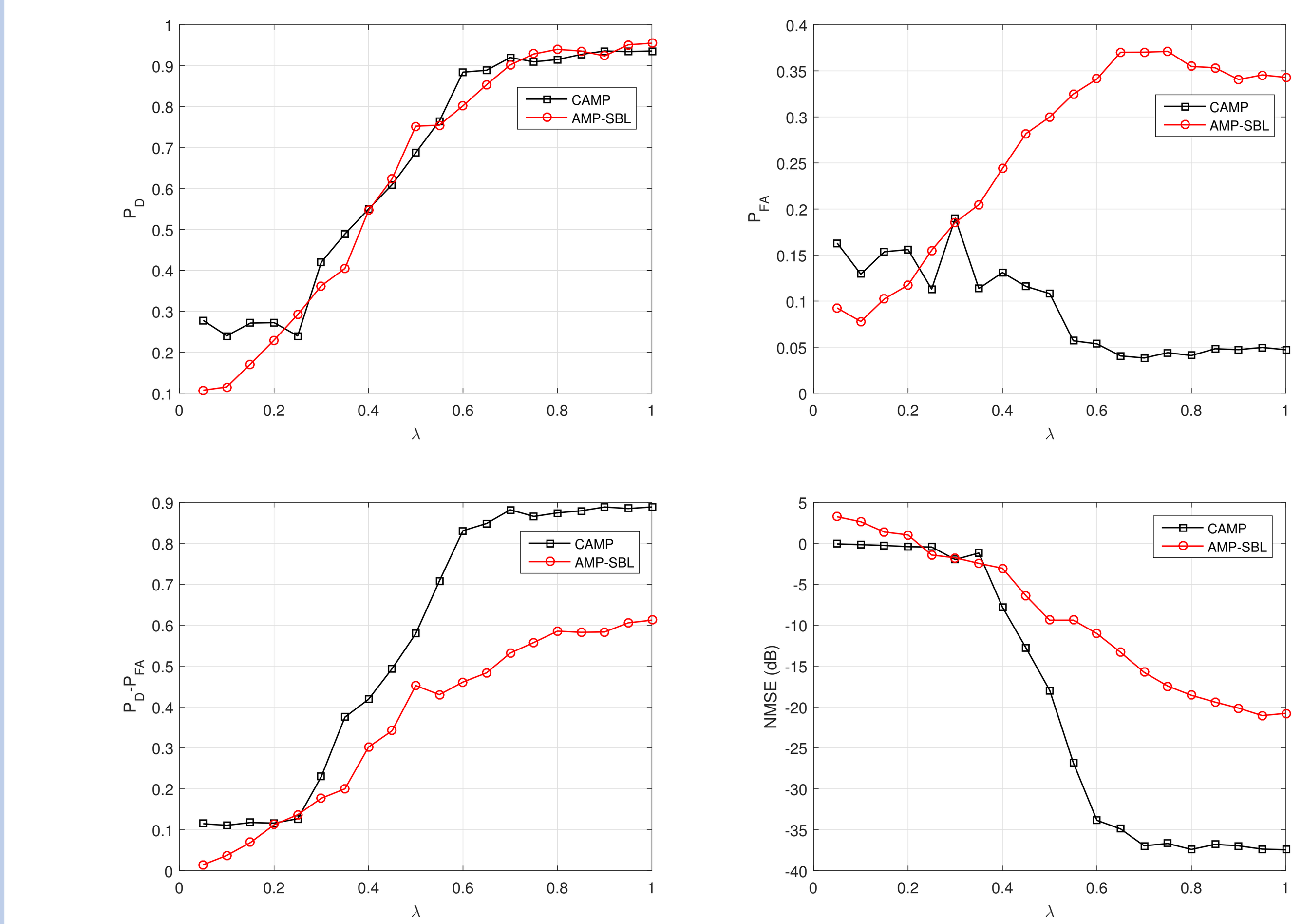
where

$$\mu_n = \sum_{m=1}^{M} a_{mn}z_{mn}(\frac{\alpha_n}{c_n + \alpha_n}), \text{ and } \nu_n = \frac{c_n\alpha_n}{c_n + \alpha_n}$$

Simulations Settings

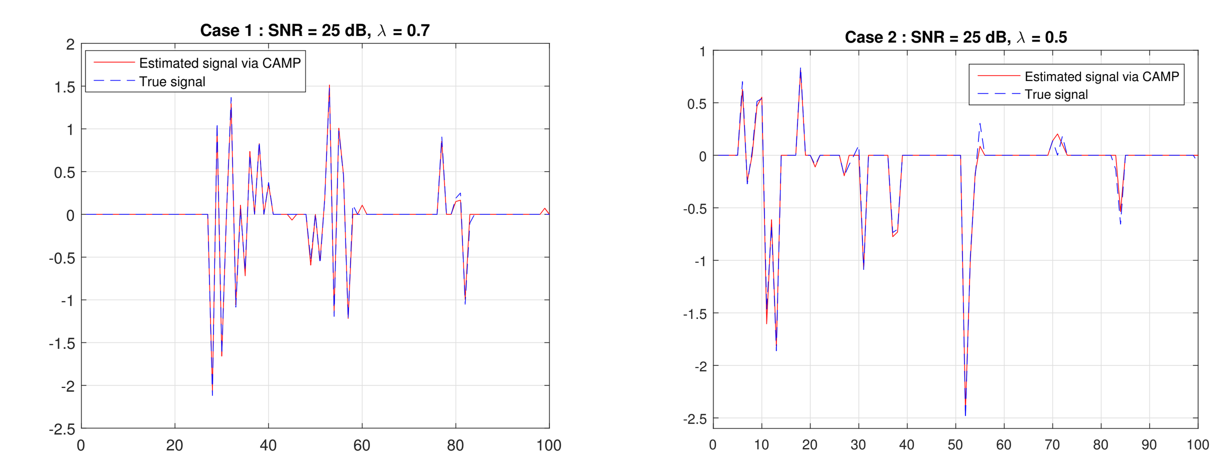
- The supports of the solution are binary and drawn from a Bernoulli distribution in such a way to have clustered sparsity structure
- The entries of $\mathbf{x}_{np}$ is drawn i.i.d. from Gaussian distribution with zero mean and variance $\sigma_x^2 = 1$
- The true solution is constructed from $\mathbf{x} = \mathbf{s} \circ \mathbf{x}_{np}$
- The sensing matrix $A \in \mathbb{R}^{M\times N}$, with $a_{mn} \sim \mathcal{N}(0, \frac{1}{\sqrt{M}})$, where M varies and $N = 100$
- The noise components are drawn i.i.d. from $\mathcal{N}(0, \sigma_n^2)$ with $SNR = 25dB$
- The measurement vector $\mathbf{y} \in \mathbb{R}^{M\times 1}$ is then computed from $\mathbf{y} = A\mathbf{x} + \mathbf{e}$
- The cardinality of $\mathbf{x}$ is set to $K_{sp} = 25$ for all the simulations
- In all the simulations, the total number of iterations is set to 1000
- To study the performance, we generate 100 random cases using the above settings and then averaging over all the obtained results
- In the figures, λ is the sampling rate and is defined as $\lambda = M/N$

Simulation Results (Performance Comparison)



Reconstruction Examples for Synthetic Data

- Parameters initialization for CAMP algorithm:
  - $\theta_1 = 10$, $c^{[0]} = 10$, $\theta_2^{[0]} = 0.5$, $T = 0.001$, and $ITER = 1000$
- *Case 1: SNR* $= 25dB$, $N = 100$, $\lambda = 0.7$, and $k_{sp} = 25$
- *Case 2: SNR* $= 25dB$, $N = 100$, $\lambda = 0.5$, and $k_{sp} = 25$



Conclusions

- A new algorithm for the recovery of sparse signals with unknown clustering pattern for the SMV problem was proposed
- **Performance evaluation:**
  - CAMP provides an overall lower false alarm rate (P_{FA}) compared to AMP-SBL
  - Based on the performance measure of ($P_D - P_{FA}$) we showed that CAMP performs better than AMP-SBL
  - The overall normalized mean-squared error (NMSE) between the true and the estimated solution for the CAMP is lower than the one for AMP-SBL

Main References

1. M. Al-Shoukairi and B. Rao, *Sparse Bayesian learning using approximate message passing*, 48th Asilomar Conf. on Sig., Syst. and Compt., pp. 1957–1961, Nov. 2014.
2. M. Shekaramiz, T. Moon, and J. Gunther, *On the block-sparse solution of single measurement vectors*, 49th Asilomar Conf. on Sig., Syst. and Compt., pp. 508–512, Nov. 2015.
3. D. L. Donoho, A. Maleki, and A. Montanari, *Message passing algorithms for compressed sensing: I. motivations and construction*, Proc. Inform. Theory Workshop, (Cairo, Egypt), pp. 1–5, Jan. 2010.