Compressive Imaging using Approximate Message Passing and a Markov-Tree Prior

Subhojit Som, Lee C. Potter, and Philip Schniter

Department of Electrical and Computer Engineering
The Ohio State University

Nov 8, 2010

The Compressive Imaging Problem

Linear observation of a sparse signal:

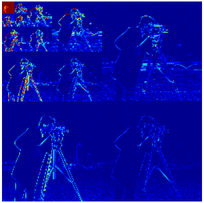
$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{w} = \boldsymbol{\Phi}\boldsymbol{\Psi}\boldsymbol{\theta} + \boldsymbol{w}.$$

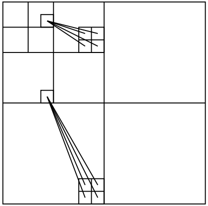
- $\boldsymbol{\Phi} \in \mathbb{R}^{M \times N}$, a known measurement matrix.
- $\boldsymbol{\Psi} \in \mathbb{R}^{N \times N}$, an orthonormal basis.
- Sparse: $\boldsymbol{\theta} = \boldsymbol{\Psi}^T \boldsymbol{x}$ has $K < M$ non-zero coefficients.
- Underdetermined when $M < N$.
- $\boldsymbol{w} \in \mathbb{R}^M$ is AWGN $\sim \mathcal{N}(0, \sigma_w^2 \mathbf{I}_M)$.

Our aim is to recover the image $\boldsymbol{x}$ from observation $\boldsymbol{y}$.

Quadtree Structure in Wavelet Transform of Natural Images

Persistence across scales: a large child coefficient usually has a large parent coefficient.





Hidden Markov Tree Model for Wavelet Coefficients

- Indicator vector $\boldsymbol{s} \in \{0,1\}^N$ denotes activity pattern.
- Wavelet coefficients θ_n are distributed independently given activity variable s_n:

$$p(\theta_n|s_n) = s_n \mathcal{N}(\theta_n; 0, \sigma_n^2) + (1 - s_n)\delta(\theta_n).$$

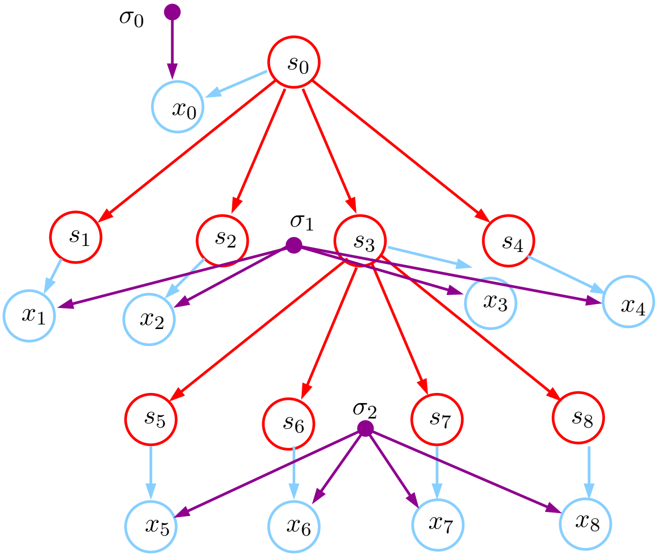
- The indicator variables are modeled as Markov tree $p(\boldsymbol{s})$.
  - Markov tree is specified by parent to child state transition matrices A_n and activity priors of the root $p(s_0 = 1)$.

$$A_n = \begin{bmatrix} p_n^{0\to 0} & 1 - p_n^{0\to 0} \\ 1 - p_n^{1\to 1} & p_n^{1\to 1} \end{bmatrix}$$

- Assume that the transition matrix A_j and variance σ_j^2 are constant at any scale but vary across scales.

[Crouse, Nowak, Baraniuk 1998; Romberg, Choi, Baraniuk 2001]

Hidden Markov Tree Model



Reconstruction w/ Probabilistically Structured Sparsity

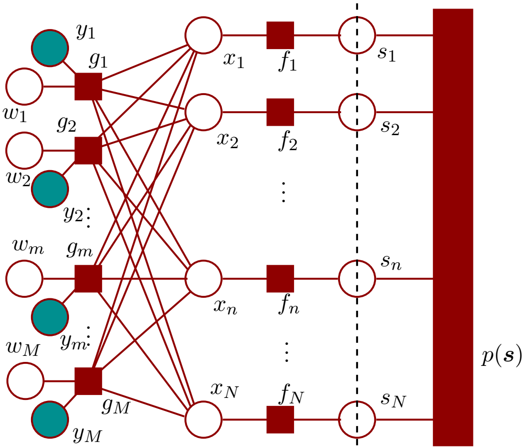
- Markov-chain Monte Carlo (MCMC):
  - Markov random field (MRF) [Wolfe, Godsill, Ng 2004]
  - Markov tree [He, Carin 2009]

  Drawbacks: slow convergence and difficulty in detecting convergence.
- Methods that iterate matching pursuit or ℓ_1-optimization with MAP sparsity-pattern detection:
  - Markov tree [Duarte, Wakin, Baraniuk 2008]
  - MRF [Cevher, Duarte, Hedge, Baraniuk 2008]

  Drawback: slow and ad hoc.
- Variational Bayes:
  - Markov tree [He, Chen, Carin 2010]

  Drawback: performance not always satisfactory
- Turbo reconstruction based on AMP:
  - Markov chain [Schniter 2010]

Factor Graph Representation



Turbo Reconstruction

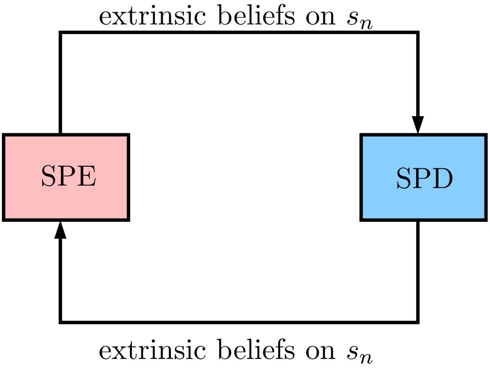
Inference problem can be tackled by *splitting it into two sub-problems and iterating between them*

- Reminiscent of noncoherent turbo equalization.
- The sparsity pattern equalization (SPE) block solves the inference problem using the observation structure (linear observation model).
- The sparsity pattern decoding (SPD) block solves the inference problem using the support structure (Markov model).

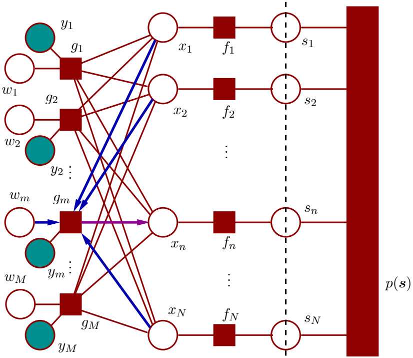
[Schniter 2010]

Message Passing between SPE and SPD

- Message passing within SPE is done via Approximate Message Passing (AMP). [Donoho, Maleki, Montanari 2009]
- SPD is done on HMT and forward-backward algorithm gives exact posterior.
- Beliefs on the indicator variables s_n are exchanged between these two blocks.



Gaussian Messages from g_m to x_n

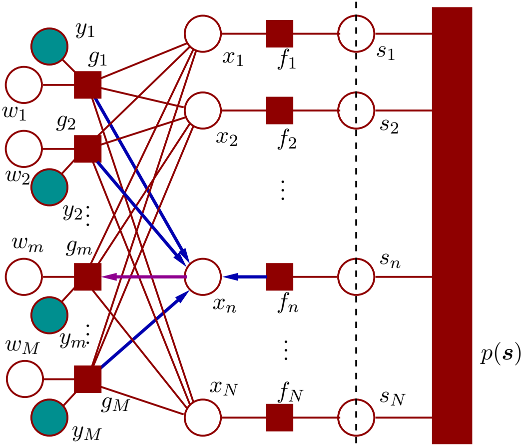


$$\nu_{g_m \to x_n}(x_n) \propto \int_{\{x_q\}_{q \neq n}} \mathcal{N}(y_m; A_{mn}x_n + \sum_{q \neq n} A_{mq}x_q, \sigma_w^2) \prod_{q \neq n} \nu_{x_q \to g_m}(x_q)$$

$$\nu_{g_m \to x_n}(x_n) = \mathcal{N}\left(x_n; \frac{z_{mn}}{A_{mn}}, \frac{c_{mn}}{|A_{mn}|^2}\right)$$

$$z_{mn} = y_m - \sum_{q \neq n} A_{mq}\mu_{mq} \qquad \text{and} \qquad c_{mn} = \sigma_w^2 + \sum_{q \neq n} A_{mq}^2 v_{mq}$$

Non-Gaussian Messages from x_n to g_m



- Outgoing message is product of incoming messages.
- Computation of means and variances suffice.

Message Update Complexity

- Message update complexity: MN updates of $\mathcal{O}(N)$ or $\mathcal{O}(M)$ corresponding to MN edges.
- Use two approximations:
  - Apply uniform variance approximations, e.g., $c_n \approx c_{mn}$.
  - Taylor series is used to approximate the deviations of messages across outgoing edges from the average message.
- These approximatations reduce the algorithm complexity to $\mathcal{O}(MN)$.
  - For subsampled DFT measurement matrix the complexity is $\mathcal{O}(N \log N)$.

Estimating Model Parameters

- Hyperpriors are assigned to model parameters σ_j, σ_w, $p_j^{0\to 0}$, $p_j^{1\to 1}$ at j^{th} scale and the probability p_0^1 that 0^{th} scale coefficients are active.
  - Gamma prior is assumed for the precision $\lambda_j = 1/\sigma_j^2$:

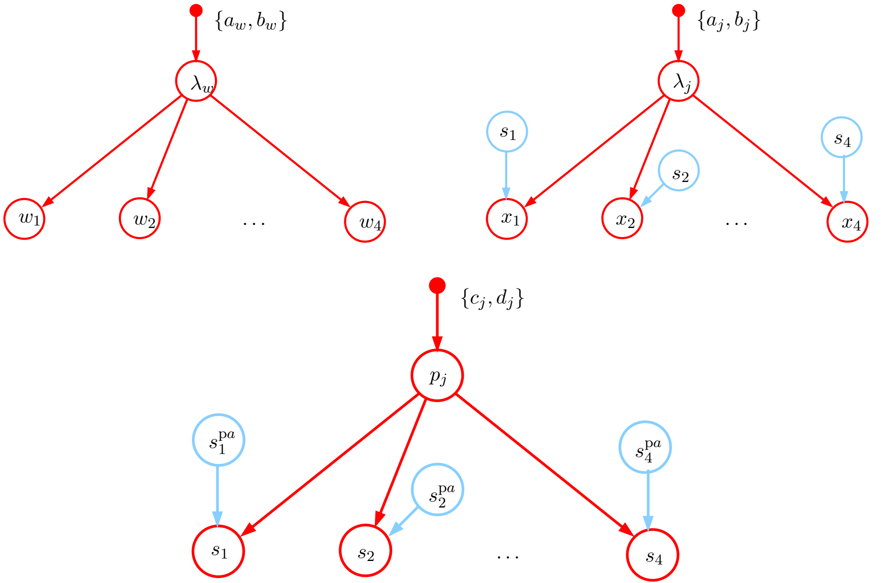
$$\mathrm{Gam}(\lambda_j) = \frac{1}{\Gamma(a_j)} b_j^{a_j} \lambda_j^{a_j - 1} \exp(-b_j \lambda).$$

  - Beta prior is assumed for activity and transition prior parameters p_j:

$$\mathrm{Beta}(p_j) = \frac{\Gamma(c_j + d_j)}{\Gamma(c_j)\Gamma(d_j)} p_j^{c_j - 1} (1 - p_j)^{d_j - 1}.$$

- At the end of every turbo iteration, the model is updated with the mean of the posteriors (MMSE estimates) of these parameters.
  - The MMSE estimates of x_n and s_n are used to obtain the posterior.

Hyperpriors on Model Parameters



Simulation Results: $M = 5000$, $N = 16384$

Iteration:1

Iteration:2

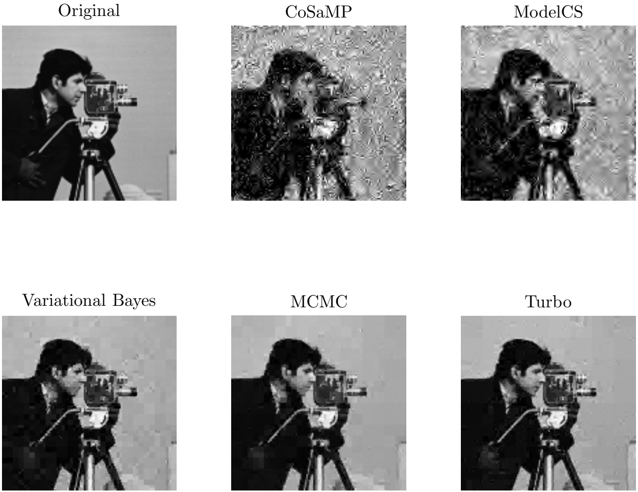
Iteration:3

Iteration:4

Iteration:5

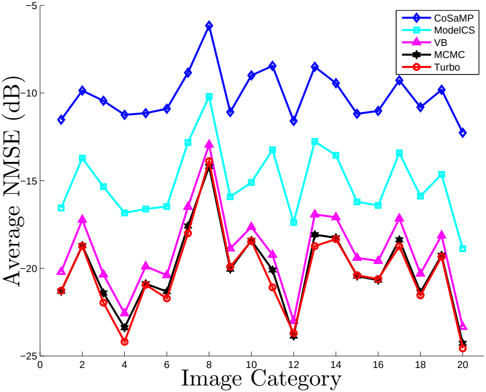
Iteration:6

Comparison with Existing Methods



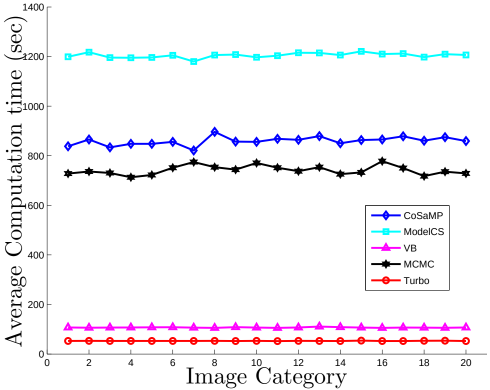
Reconstruction from $M = 5000$ observations. The images are of size 128×128 (i.e., $N = 16384$).

Comparison with Existing Methods



Average Normalized Mean Squared Error (NMSE) for 20 categories 128×128 images (i.e., $N = 16384$) from $M = 5000$ observations. In each category there are 30 images.
[http://research.microsoft.com/en-us/projects/objectclassrecognition]

Comparison with Existing Methods



Average computation time for 20 categories 128×128 images (i.e., $N = 16384$) from $M = 5000$ observations. In each category there are 30 images.

Summary on Comparison with Existing Methods

Algorithm	Computation Time (sec)	NMSE (dB)
CoSaMP	859	-10.13
ModelCS	1205	-15.10
Variational Bayes	107	-19.04
MCMC	742	-20.10
Turbo	53	-20.31

Table: Average computation time and Normalized Mean Squared Error (NMSE) from 584 natural images of size 128×128; $N = 16384$, $M = 5000$.

Conclusions

- We apply turbo reconstruction algorithm based on approximate message passing on hidden Markov tree modeled structured sparse signals.
- We propse to estimate the model parameters from the measured data.
- We apply the algorithm on natural images and demonstrate that it performs better than the existing algorithms in terms of both accuracy and computation time.