Efficient Message Passing-Based Inference in the Multiple Measurement Vector Problem

Justin Ziniel   Philip Schniter

Department of Electrical and Computer Engineering
The Ohio State University

Asilomar Conference on Signals, Systems, and Computers, 2011

Work supported in part by NSF grant CCF-1018368 and DARPA/ONR grant N66001-10-1-4090

Outline

Background
- The Multiple Measurement Vector (MMV) Problem
- Existing Approaches
- Signal Model

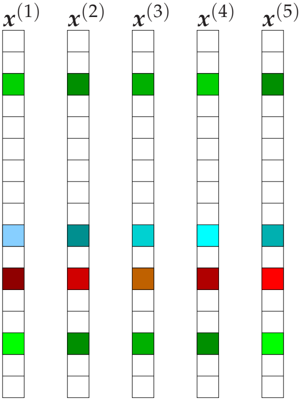
Our Proposed Method
- Belief Propagation-Based Inference
- EM Parameter Learning

Empirical Study

Conclusion

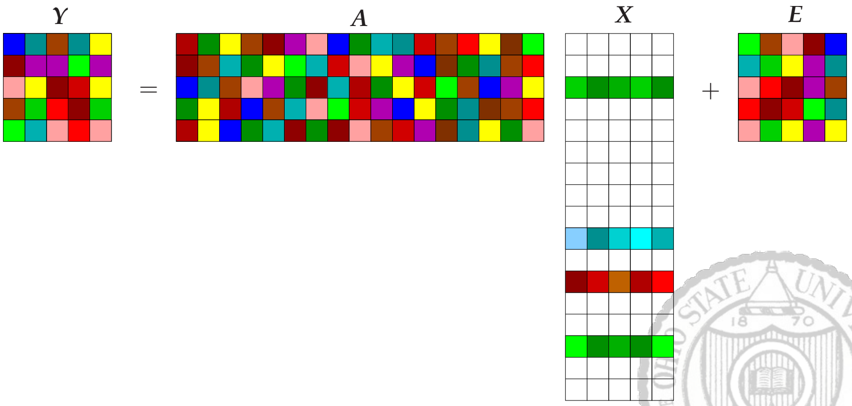
The Multiple Measurement Vector (MMV) Problem

Consider a time-series of sparse, temporally correlated signal vectors *that share a common support*...



The Multiple Measurement Vector (MMV) Problem

...observed through a noisy linear measurement process, $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{E}$.



Applications: Magnetoencephalogaphy, direction-of-arrival estimation, parallel MRI,...

Existing methods

- Greedy pursuit
  - M-BMP, M-OMP, M-ORMP [Cotter et al., '05]
  - S-OMP [Tropp et al., '06]
  - Subspace-augmented MUSIC* [Lee et al., '10]
- Mixed-norm (ℓ_1/ℓ_2) minimization
  - M-FOCUSS [Cotter et al., '05]
  - RX-penalty, RX-error [Tropp et al., '06]
  - JLZA [Hyder and Mahata, '10]
  - tMFOCUSS* [Zhang and Rao, '11a]
- Bayesian MMV
  - M-SBL [Wipf and Rao, '07]
  - JSSR-MP [Shedthikere and Chockalingam, '11]
  - T-MSBL*, T-SBL* [Zhang and Rao, '11b]
- Block-sparse single measurement vector
  - [Eldar and Mishali, '09]
  - bSBL [Zhang and Rao, '11b]

* = Accounts for temporal correlation in amplitudes

Comparing Different Approaches

Approach	**Speed**	**Performance**
Greedy	Fast	Fair
Mixed-norm	Okay	Good
Bayesian	Slow	Great

Why Bayesian?

- Modeling assumptions are made explicit
- Model parameters have meaningful interpretations
- Principled parameter learning
- Soft inference

Comparing Different Approaches

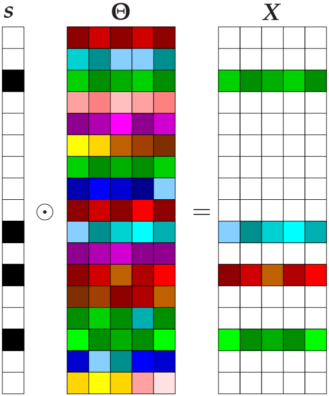
Approach	Speed	Performance
Greedy	Fast	Fair
Mixed-norm	Okay	Good
Bayesian	Slow	Great

Why Bayesian?

- Modeling assumptions are made explicit
- Model parameters have meaningful interpretations
- Principled parameter learning
- Soft inference

A Model of Sparse Time-Evolving Signals

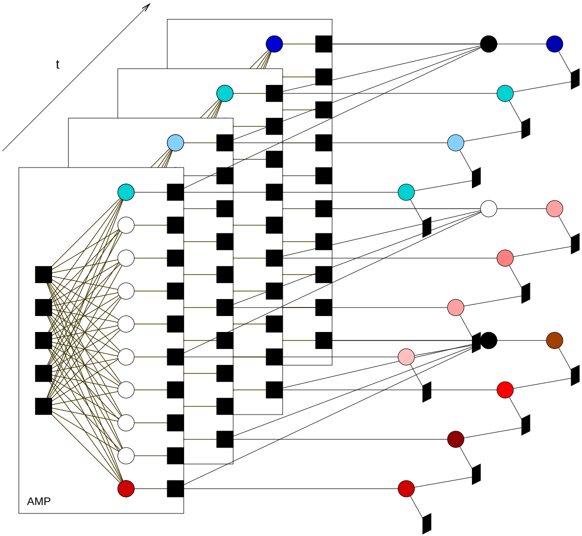
We write: $x_n^{(t)} = s_n^{(t)} \cdot \theta_n^{(t)}$ for $s_n^{(t)} \in \{0,1\}$ and $\theta_n^{(t)} \sim \mathcal{CN}(\zeta, \sigma^2)$.



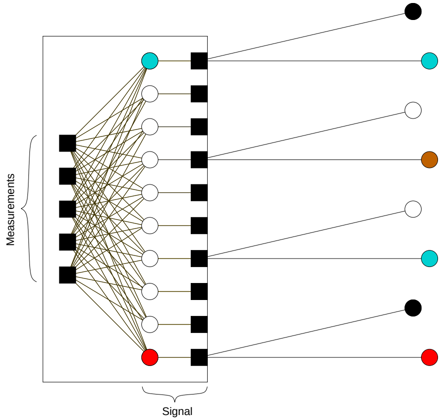
Amplitude Evolution

Treat $\{\theta_n^{(t)}\}_{t=1}^{T}$ as a Gauss-Markov process: $\theta_n^{(t)} = (1-\alpha)\theta_n^{(t-1)} + \alpha w_n^{(t)}$, where $w_n^{(t)} \sim \mathcal{CN}(0,\rho)$, and α conrols the correlation in the random process.

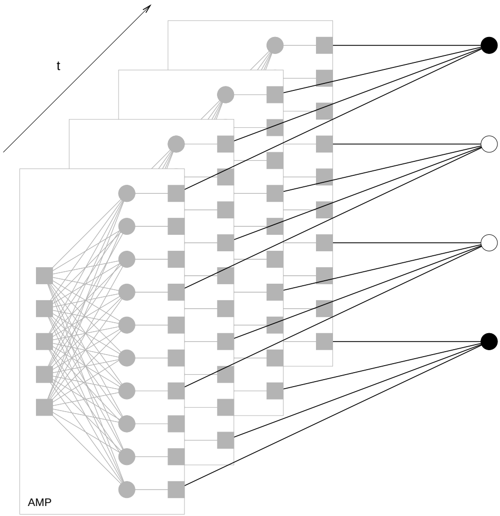
The Factor Graph Representation



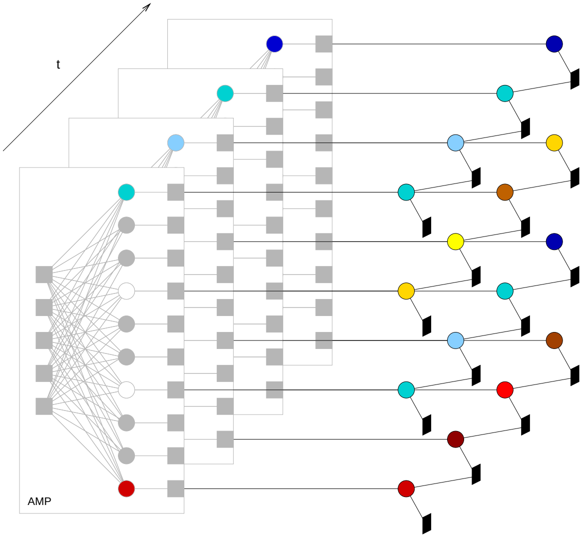
The Factor Graph Representation: Single Timestep



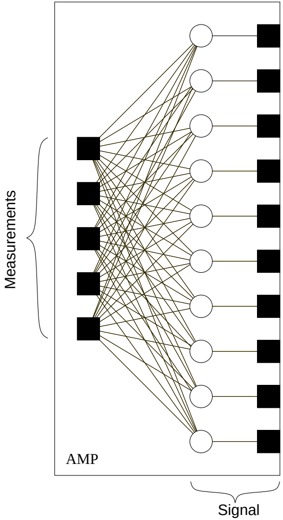
The Factor Graph Representation: Support Variables



The Factor Graph Representation: Amplitude Variables



Approximate Message Passing (AMP)

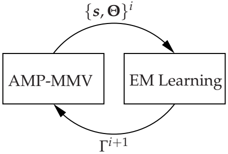


- Standard belief propagation is intractable here
- *Simplification*: Approximate message passing (AMP), [Donoho, Maleki, and Montanari, '09, '10]
- Marginal for $x_n^{(t)}$: *Bernoulli-Gaussian* - $(1-\pi_n^{(t)})\delta(x_n^{(t)}) + \pi_n^{(t)}\mathcal{CN}(x_n^{(t)};\xi_n^{(t)},\psi_n^{(t)})$
- As $M,N \to \infty$, AMP behavior described precisely by state evolution $\to$ MMSE-optimal estimates [Bayati and Montanari, '10]

# of messages exchanged: $\mathcal{O}(N)$
Complexity per iteration: $\mathcal{O}(MN)$ **(matrix-vector product)**

Parameter Learning via Expectation-Maximization

- Signal model governed by a number of parameters: $\Gamma \triangleq \{\lambda, \zeta, \sigma^2, \alpha, \rho, \sigma_e^2\}$
- Parameters can be tuned automatically from the data using an expectation-maximization (EM) algorithm



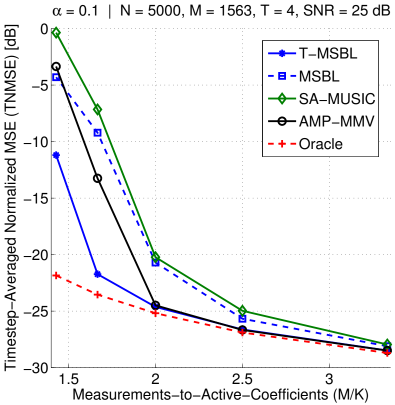
- Finds local maximizer of $p(\boldsymbol{Y}|\Gamma)$
- EM parameter estimation fits naturally into the existing message passing procedure
  - The E-step of the EM algorithm makes use of quantities available for free as a byproduct of AMP-MMV!

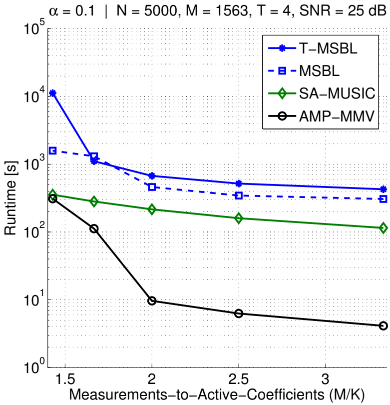
Empirical Study: Setup

- AMP-MMV w/ EM parameter learning was compared against 3 powerful MMV algorithms, and an oracle-aided MMSE bound (support-aware Kalman smoother)
    - *Bayesian*: MSBL and T-MSBL* [Zhang and Rao, '11b]
    - *Greedy*: Subspace-augmented MUSIC (SA-MUSIC*) [Lee et al., '10]
- Signals generated according to signal model; i.i.d. Gaussian $\boldsymbol{A}$ matrices; AWGN corrupting noise

* = Accounts for temporal correlation in amplitudes

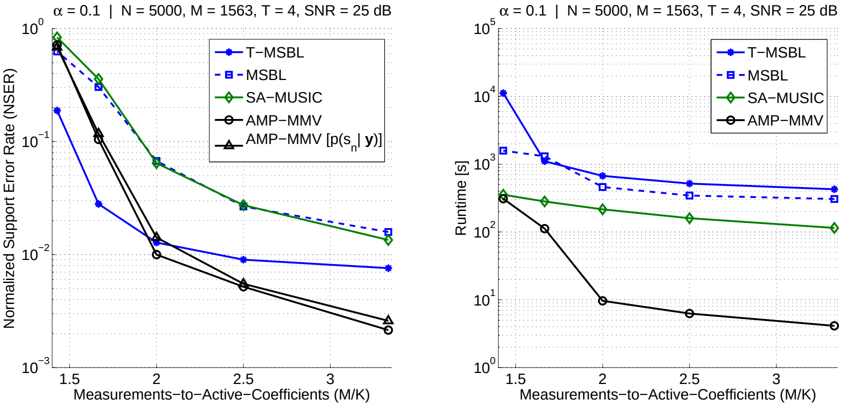
Empirical Study: MSE vs. Normalized Sparsity Rate





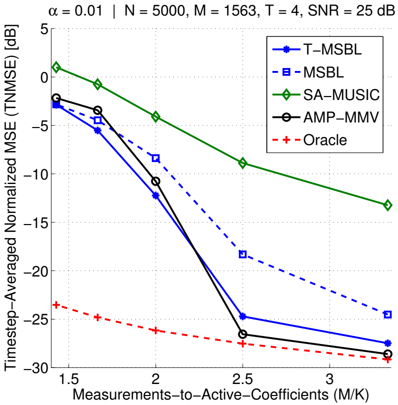
Correlation: $1 - \alpha = 0.90$

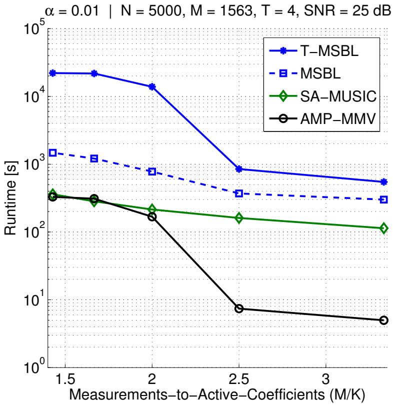
Empirical Study: NSER vs. Normalized Sparsity Rate



Correlation: $1 - \alpha = 0.90$

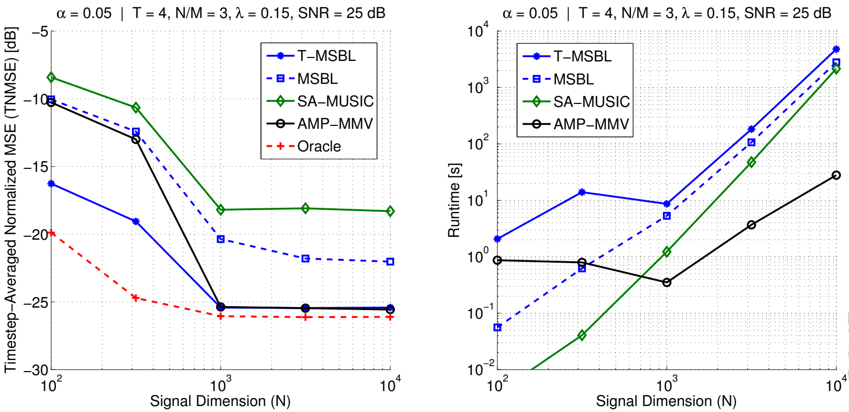
Empirical Study: MSE vs. Normalized Sparsity Rate





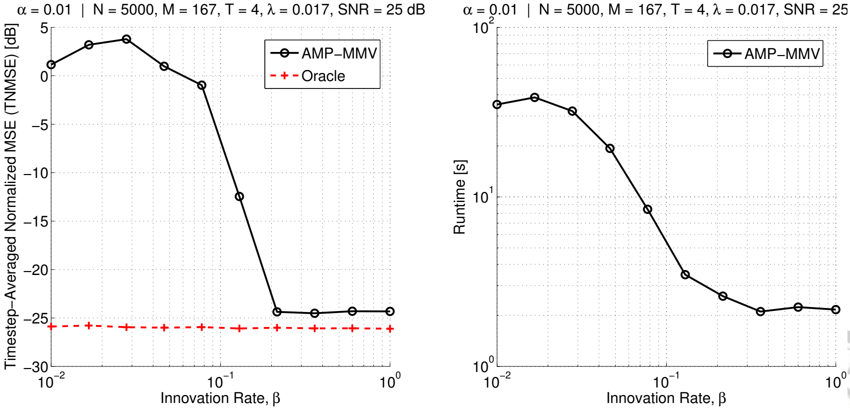
Correlation: $1 - \alpha = 0.99$

Empirical Study: MSE vs. Signal Dimension



Correlation: $1 - \alpha = 0.95$

Empirical Study: MSE vs. Measurement Innovation



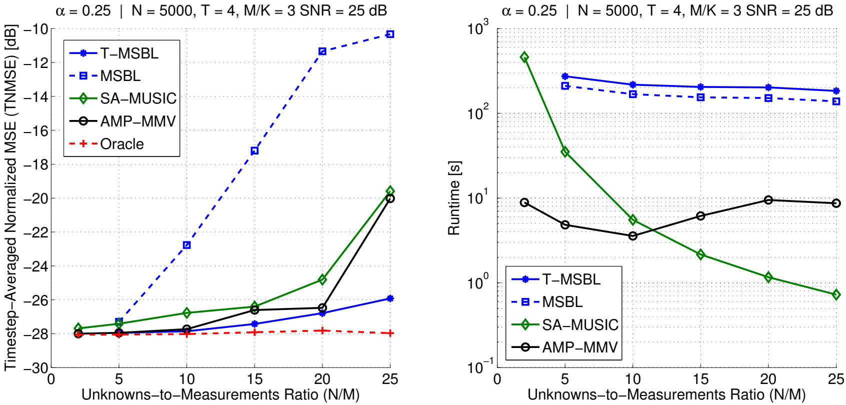
Time-varying measurement matrix: $A^{(t)} = (1-\beta)A^{(t-1)} + \beta W^{(t)}$

Correlation: $1-\alpha = 0.99$ | Undersampling Rate (N/M): 30 | Normalized Sparsity (M/K): 2

Conclusion

- AMP-MMV
  - Works with temporally correlated signal amplitudes
  - Performance rivals an oracle-aided MMSE bound (support aware Kalman smoother) over a wide range of problems
  - Computational complexity scales *linearly* in all problem dimensions
- EM parameter learning
  - Principled method of learning signal model parameters
  - Closed-form updates using outputs of AMP-MMV
- Empirical study
  - Two orders-of-magnitude improvement in runtime
  - Major gains possible from matrix diversity

Empirical Study: MSE vs. Undersampling Rate



Correlation: $1 - \alpha = 0.75$