The Surprising Effectiveness of Bayesian Sparse Signal Recovery

Chandra R. Murthy
Dept. of ECE
Indian Institute of Science

cmurthy@iisc.ac.in

Outline

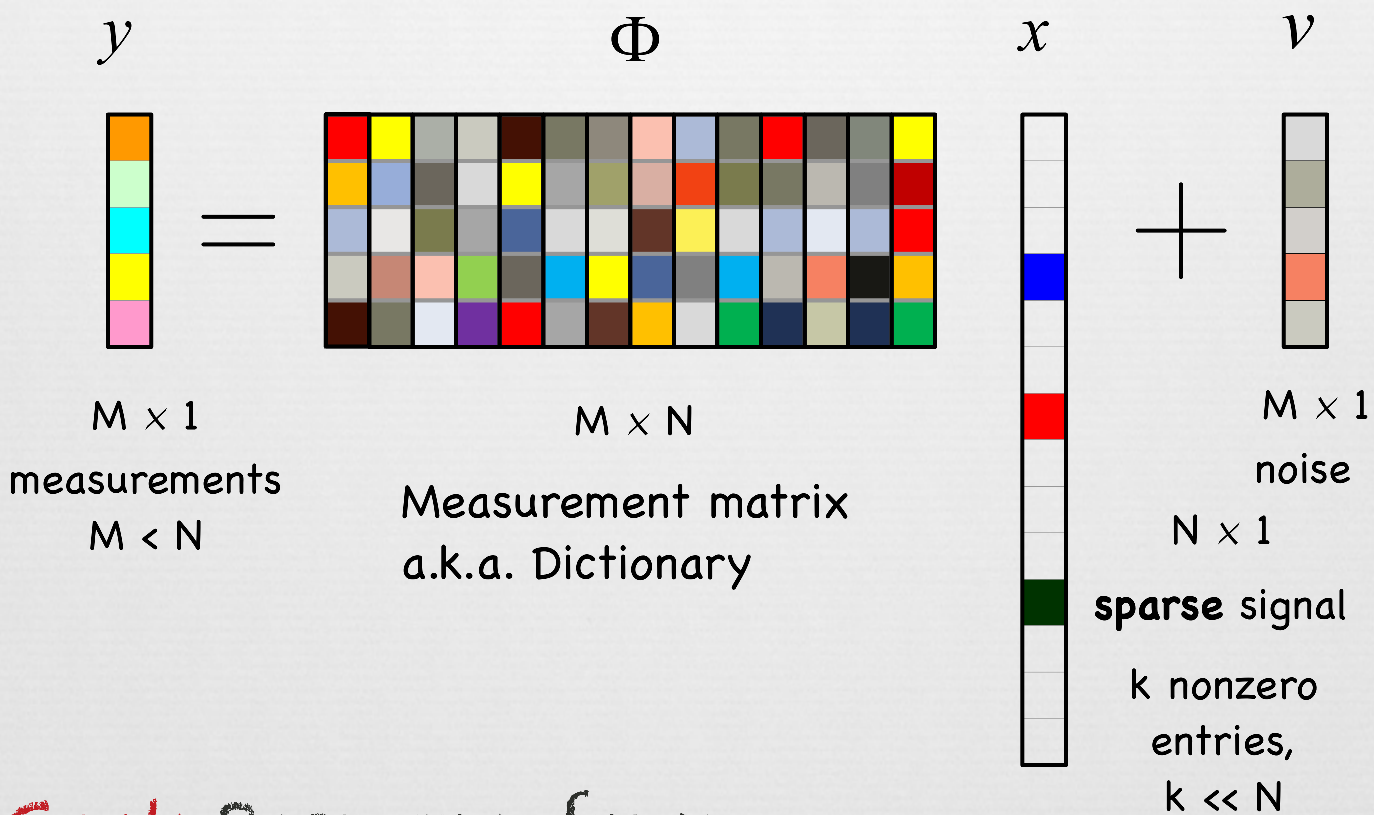
- Background and motivation
- Sparse Bayesian learning
- Joint-sparse recovery
  - Support recovery guarantees
- Extensions and new algorithms
- Applications in communication systems

Part 1: Setting the Stage

Motivation and background

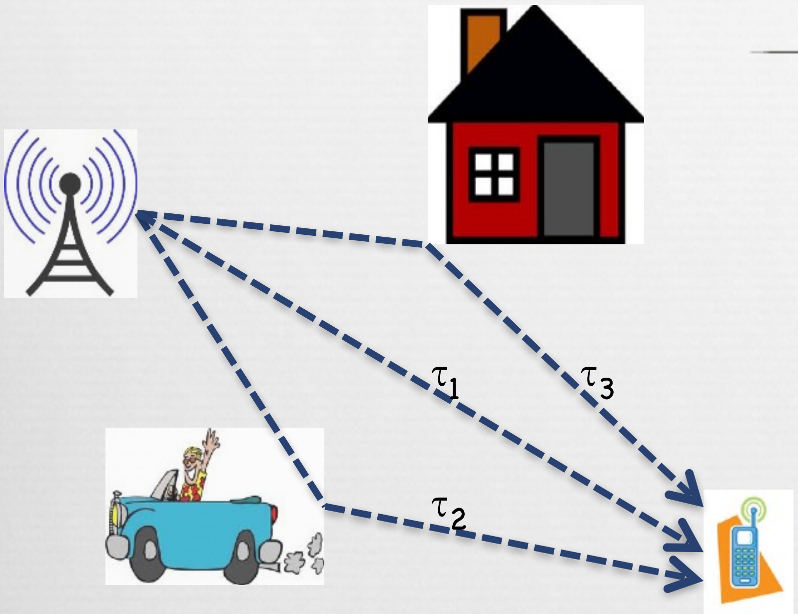
Basic results

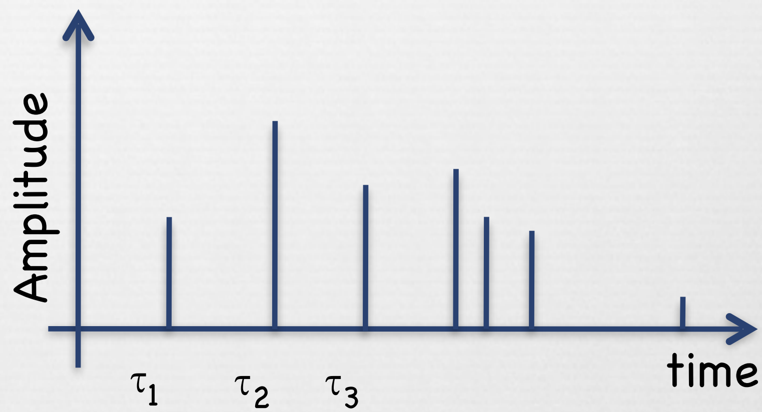
Sparse Signal Recovery



- Goal: Recover x from y
- M << N: infinitely many solutions

Wireless Channel Estimation





- Wireless channels exhibit multipath
  - Naturally sparse in the lag-domain
  - Need to estimate both support & channel
- Channel equalization & data detection

Compressed Sensing

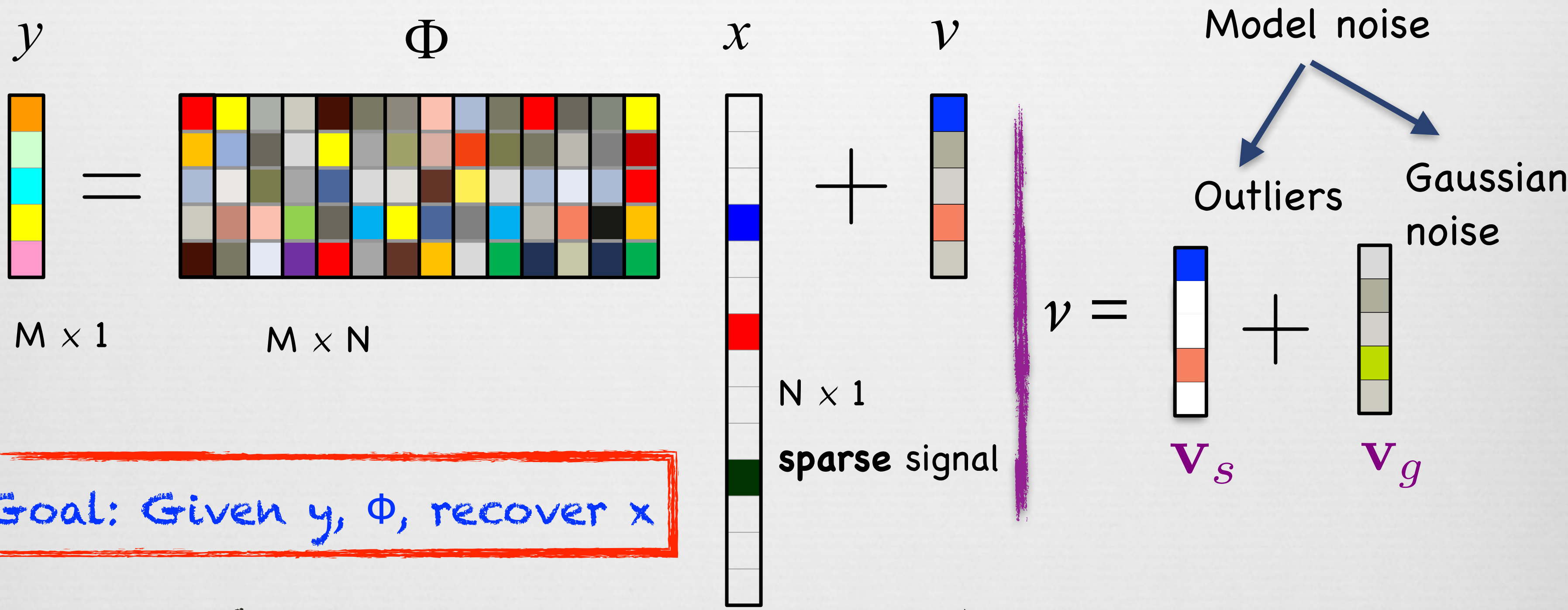
- Deals with three main questions:
  - Design of sensing matrices

$$\Phi_{M\times N} = \mathbf{A}_{M\times N}\Psi_{N\times N}$$

Sparsifying Basis

  - Guarantees for recovery
  - Computationally efficient algorithms
- This talk: New algorithms and guarantees for sparse signal recovery!

Robust Linear Regression: Underdetermined Case



Goal: Given y, Φ, recover x

- Transform into an overcomplete problem:

$$\mathbf{Y} = \mathbf{\Phi x} + \mathbf{\Psi v}_s + \mathbf{v}_g, \text{ where } \mathbf{\Psi} = \mathbf{I}$$

$$\text{or } \mathbf{Y} = [\mathbf{\Phi}, \mathbf{\Psi}] \begin{bmatrix} \mathbf{x} \\ \mathbf{v}_s \end{bmatrix} + \mathbf{v}_g$$

Sparse recovery algos are now applicable!

Robust Linear Regression: Overdetermined Case

- Measurement model:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{E} + \mathbf{e}$$

$M \times N;$ $M \geq N$ — Outliers; sparse — Noise

- Use SVD: $\mathbf{A} = \mathbf{U}_1 \Sigma \mathbf{V}_1^T;\ \mathbf{U}_2^T \mathbf{A} = \mathbf{0}$
- Processed measurements:

$$\tilde{\mathbf{y}} = \mathbf{U}_2^T \mathbf{y} = \mathbf{U}_2^T \mathbf{E} + \mathbf{U}_2^T \mathbf{e}$$

- Can now directly apply sparse signal recovery algorithms to estimate and remove outliers!

The Problem

- Noiseless case: Given y and Φ, solve

$$\min \|\mathbf{x}\|_0 \text{ subject to } \mathbf{y} = \mathbf{\Phi x}$$

- Noisy case: solve

$$\min \|\mathbf{x}\|_0 \text{ subject to } \|\mathbf{y} - \mathbf{\Phi x}\|_2 \leq \beta$$

- l_0 norm minimization
  - Combinatorial complexity
  - Not robust to noise

Breakthrough 1: The Null Space Property

- Underdetermined systems: $y = \Phi x$; Φ is M x N, $M < N$, x is k-sparse:
  - Infinitely many solutions, but …
  - Unique soln. if nullspace of Φ has no "sparse" vectors [Donoho, Elad '02]
    - Recovery of all k-sparse x: $M \geq 2k$ is nec. & suff.
  - Given x, unique soln. with high probability, if $M \geq k+1$ [Bresler; Wakin etc]
- Thus: Sub-Nyquist sampling (compression) possible:
  - When we restrict to sparse signals
  - And sample in an "appropriate" basis Φ

Breakthrough 2: Just Relax!

- l_1 min. instead of l_0 min.

$$\min \|\mathbf{x}\|_1 \text{ subject to } \mathbf{y} = \mathbf{\Phi x}$$

  - Convex optimization problem; linear program
- Same solution as l_0 minimization!
  - If the measurement matrix is random
  - Use slightly larger number of measurements
  - Robust to measurement noise $M \approx k \log\left(\frac{N}{k}\right) \ll N$
- See [Donoho; Candes, Romberg, Tao etc]

Breakthrough 3: Recovery Guarantees

- Noisy measurements: $y = \Phi x + v$
- We solve $(P_{1,\eta}): \min_{\mathbf{z}} \|\mathbf{z}\|_1$ subject to $\|\mathbf{\Phi z} - \mathbf{y}\|_2 \leq \eta$
- Robust NSP: Φ satisfies RNSP(k) if, $\forall$ S of cardinality k,

$$\|\mathbf{w}_S\|_1 \leq \rho \|\mathbf{w}_{S^c}\|_1 + \tau \|\Phi \mathbf{w}\|_1 \ \forall \ \mathbf{w}$$

- **Result:** If Φ satisfies RNSP(k), a sol. x^* of $(P_{1,\eta})$ with $\|\mathbf{v}\|_2 \leq \eta$ and $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{v}$ satisfies

Error in approximating x by a k-sparse vector

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \leq 2\left(\frac{1+\rho}{1-\rho}\right) \sigma_k(\mathbf{x})_1 + \frac{4\tau}{1-\rho} \eta$$

Noise

Analysis of BP: stable and robust recovery

- Theorem: Suppose $2k^{th}$ Restricted Isometry Constant of Φ satisfies

$$\delta_{2k} < \frac{4}{\sqrt{41}} \approx 0.6246$$

then for x,y with $\|Ax-y\|_2 \leq \eta$, the sol x^* of

$$\min_{\mathbf{z}} \|\mathbf{z}\|_1 \text{ subject to } \|\mathbf{\Phi z} - \mathbf{y}\|_2 \leq \eta$$

satisfies

$$\|\mathbf{x} - \mathbf{x}^*\|_1 \leq C\sigma_k(\mathbf{x})_1 + D\sqrt{k}\eta$$

$$\|\mathbf{x} - \mathbf{x}^*\|_2 \leq \frac{C}{\sqrt{k}}\sigma_k(\mathbf{x})_1 + D\eta$$

where C, D >0 depend only on δ_{2k}

Breakthrough 4: Recovery Algorithms

- Greedy algorithms:
  - Matching pursuit [Mallat, Zhang; Cotter, Rao]
  - Orthogonal matching pursuit [Tropp 03]
  - CoSAMP [Needell, Tropp]
- Relaxation based methods (minimize diversity meas.):
  - Basis pursuit (l-p, with p=1) [Chen et al.]
  - Lasso (BPDN) [Tibshirani]
  - Dantzig selector [Candes, Tao]
  - Homotopy based methods (e.g., LARS) [Garrigues et al. 09]
  - FOCUSS (l-p, with p < 1) [Gordonitsky et al.]
- Iterative methods:
  - Basic/Iterative hard thresholding
  - Hard thresholding pursuit

Recovery guarantees exist for most of these algorithms! See [Rauhut & Foucart]

Limitations of Greed & Relaxation

- Performance of BP and OMP depend on Φ
  - Poor performance when conditions are violated
  - Hard to relate estimation error to the dictionary
  - Correlated dictionary: disrupts l_0-l_1 equivalence
  - BP: performance independent of nonzero coeffs [Malioutov et al. 2004]
    - Cannot improve when situation is favorable
  - OMP: performance highly sensitive to magnitudes of nonzero coefficients
    - Poor performance with unit magnitudes
- Other issues:
  - Estimating embedded parameters, exploiting additional structure when available

To Recap

- Sparse signal recovery
  - Basic problem, breakthroughs in CS
  - Algorithms
  - Guarantees
- Limitations
  - Scaling/shrinkage
  - Correlated dictionary
  - Embedded parameters

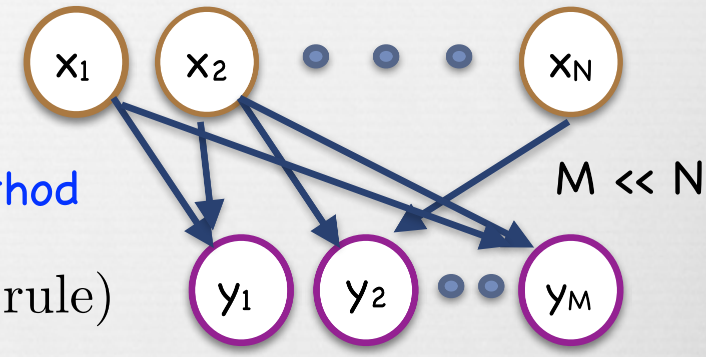
Part 2: Dont Relax!

A time and place for nonconvex methods?

Bayesian Methods

- MAP estimation (Type I):
  - Also a regression problem with sparsity promoting penalties (e.g., l_p-norm)
  - l_1-min (BP/LASSO) is a special case
- Hierarchical Bayesian methods (Type II):
  - Iterative reweighted l_1 [Candes et al. 2008]
  - Iterative reweighted l_2 [Chartrand & Yin 2008]
  - EM-based SBL [Tipping, 2001], [Wipf, Rao 2007]
  - AMP [Schniter 2008], [Rangan 2011]

MAP Estimation



$$\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$$ ← Type-I method

$$= \arg\min_{\mathbf{x}} -\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}) \quad \text{(Bayes' rule)}$$

$$= \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^{N} g(|x_i|)$$ ← Separable prior

- For sparse solutions, $g(|x_i|)$ should be a concave, nondecreasing function
  - Example: $g(|x_i|) = |x_i|^p$, $p \leq 1$
  - Lasso is a special case: p=1
- Any local min. of the MAP estmn problem has at most M nonzeros [Rao et al., 99]

The Optimization Problem

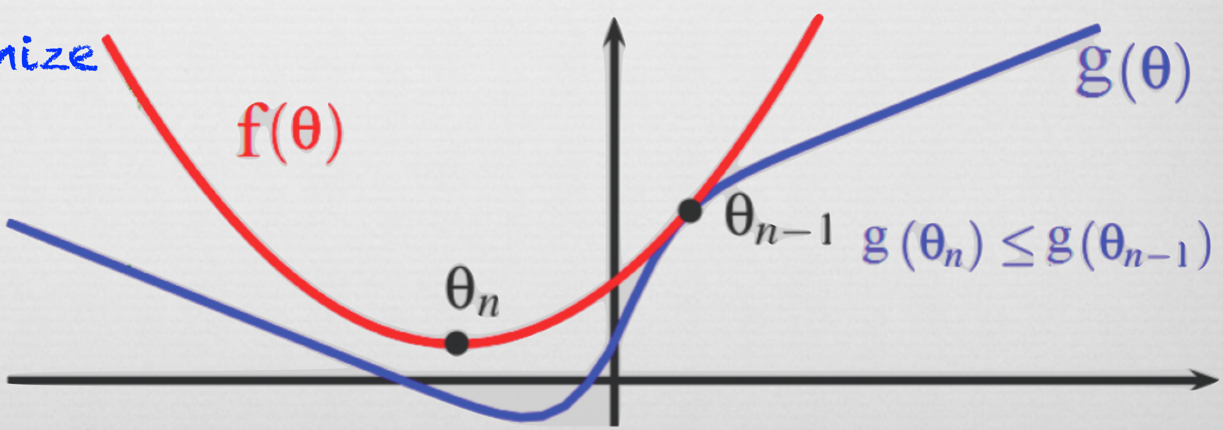
- To solve

$$\arg\min_{\mathbf{x}} G(\mathbf{x}) \triangleq \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^{N} g(|x_i|)$$

  - g(x) concave, monotonically ↑ in |x|
  - G(x) convex + concave
- Many options for g(x) to promote sparsity
- Many options for solving the optz. problem

Sparsity-Promoting Penalties

- Concave penalty fns. promote sparsity
  - $g(x) = \log(x^2 + \varepsilon),\ \varepsilon > 0$ [Chartrand & Yin 2008]
  - $g(x) = \log(|x| + \varepsilon),\ \varepsilon > 0$ [Candes et al. 2008]
  - $g(x) = |x|^p,\ 0 < p < 1$ [Rao et al., 99]
- A general approach: majorize-minimize



-

Majorization-Minimization Approach

- Find an upper bound $g(x) \leq f(x|x^{(m)})$
  - Equality at $x = x^{(m)}$, convenient for opt.
- Step 1: Optimize

$$\arg\min_{\mathbf{x}} F\left(\mathbf{x}|\mathbf{x}^{(m)}\right) \triangleq \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + \lambda \sum_{i=1}^{N} f\left(|x_i||x_i^{(m)}\right)$$

- Step 2: Set m <- m+1, update $f(x|x^{(m)})$, iterate
- Works because

$$G(x^{(m+1)}) \leq F(x^{(m+1)}|x^{(m)}) \leq F(x^{(m)}|x^{(m)}) = G(x^{(m)})$$

Iterative Reweighted l_1

- Concavity in $|x|$: $g(x) \le g'(x^{(m)})(x-x^{(m)}) + g(x^{(m)})$
  - Equality at x = x(m), linear in x
- Iterative reweighted l_1: [Candes et al. 08]
  - Init: m = 0, $x^{(m)}$ = something convenient
  - Iterate:
    - Optimize

$$\mathbf{x}^{(m+1)} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda \sum_{i=1}^{N} g'(x_i^{(m)})|x_i|$$

    - m <- m+1, update $g'(x_i^{(m)})$
  - Until convergence

Weighted l_1 minimization

Iterative Reweighted l_2

- g(x) concave in x²: $g(x) \leq \left(\frac{\partial g(\sqrt{x^2})}{\partial (x^2)} \bigg|_{x=x_0} \right) (x^2 - x_0^2) + g(x_0)$
- Optimization problem
$$\mathbf{x}^{(m+1)} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \lambda \sum_{i=1}^{N} w_i^{(m)} |x_i|^2$$
$$\|\mathbf{W}_m^{-\frac{1}{2}} \mathbf{x}\|_2^2$$
- Iterative reweighted l2 [Chartrand et al. 08]
  - Init: m = 0, $x^{(m)}$ = something convenient
  - Iterate:
    - Compute $\mathbf{x}^{(m+1)} = \mathbf{W}_m \Phi^T \left(\lambda \mathbf{I} + \Phi \mathbf{W}_m \Phi^T\right)^{-1} \mathbf{y}$
    - m <- m+1, update W_m
  - Until convergence

An Example

- Suppose g(x) = log (|x| + ε), ε > 0
  - Concave in |x|, x^2
- Iterative reweighted l1

$$g'\left(x_i^{(m)}\right) = \left[\left|x_i^{(m)}\right| + \epsilon\right]^{-1}$$

- Iterative reweighted l2

$$w_i^{(m)} = \left[\left(x_i^{(m)}\right)^2 + \epsilon\left|x_i^{(m)}\right|\right]^{-1}$$

Limitations of MAP

- Many local minima $O({}^{N}C_{M})$
  - May get stuck at a local minimum
- MAP only guarantees max $p(x = x_0|y)$
  - Probability mass, rather than mode, may be more relevant for continuous random vars
  - Perhaps posterior mean $E(x|y)$?
- Even with the true prior, MAP estimators do not minimize MSE: so MSE may be high!
  - In fact, using "true" statistics often does not lead to the lowest MSE!

To Recap

- Bayesian estimation
  - Basic MAP estimation
  - Majorization-minimization approach
  - Iterative reweighted algorithms
- Limitations
  - Many local minima
  - Posterior mean vs. posterior mode

Part 3: Sparse Bayesian Learning

Use lots of priors and pick the best one!

Point of Departure: Alternative Prior

- Gaussian Scaled Mixtures (GSM)

$$\mathbf{x} = \gamma G;\ G \sim \mathcal{N}(\mathbf{g}; 0, 1)$$

$$p(\mathbf{x}) = \int p(\mathbf{x}|\gamma)p(\gamma)\mathrm{d}\gamma = \int \mathcal{N}(\mathbf{x}; 0, \gamma)p(\gamma)\mathrm{d}\gamma$$

  - γ: +ve random variable, indep. of G
    - Spike-and-slab model if γ = Bern(0,1)
  - Most priors can be expressed using GSM (incl. ones with concave g) [Palmer et al., 2006]

Examples

- Laplacian density
  - We use: $$p(\gamma) = \frac{a^2}{2} \exp\left(-\frac{a^2}{2}\gamma\right), \gamma \geq 0$$
  - And get: $$p(x_i; a) = \frac{a}{2} \exp(-a|x_i|)$$
  - Which leads to the familiar LASSO problem
- Student's t distribution
  - We use: gamma distribution
  - And get:

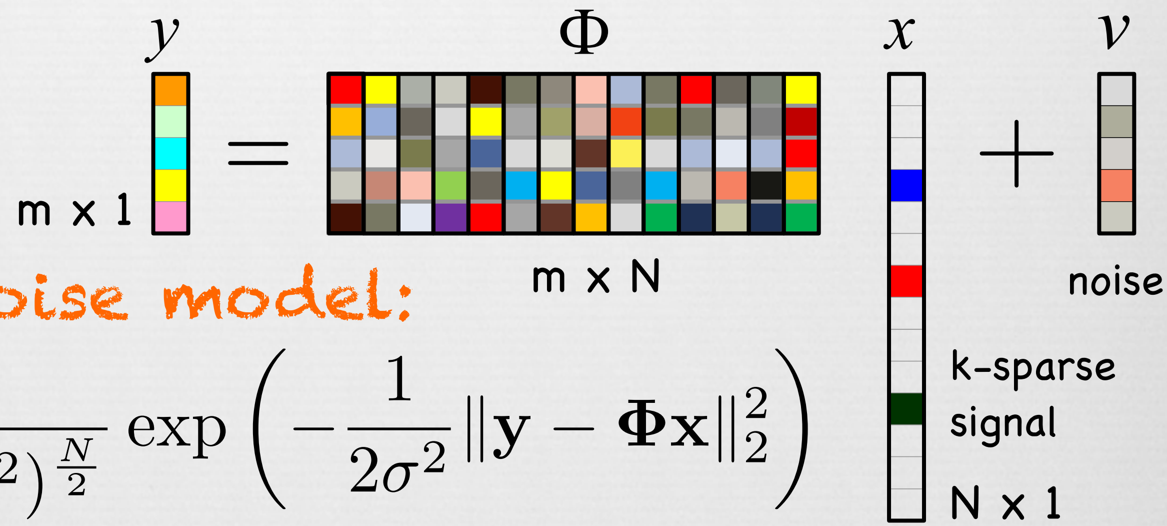
$$p(x_i; a, b) = \frac{b^a \Gamma(a + 1/2)}{\sqrt{2\pi}\Gamma(a)} \frac{1}{(b + x_i^2/2)^{a+1/2}}$$

Examples

- Generalized Gaussian
  - We use: positive alpha-stable density of order p/2
  - And get: $$p(x_i;p) = \frac{1}{2\Gamma\left(1+\frac{1}{p}\right)} \exp(-|x_i|^p)$$
- Generalized logistic distribution
  - We use: A scale mixing density related to the Kolmogorov Smirnoff distance
  - And get: $$p(x_i;\alpha) = \frac{\Gamma(2\alpha)}{\Gamma(\alpha)^2} \frac{\exp(-\alpha|x_i|)}{(1+\exp(-|x_i|))^{2\alpha}}$$

Sparse Bayesian Learning

- Canonical model



- Gaussian noise model:

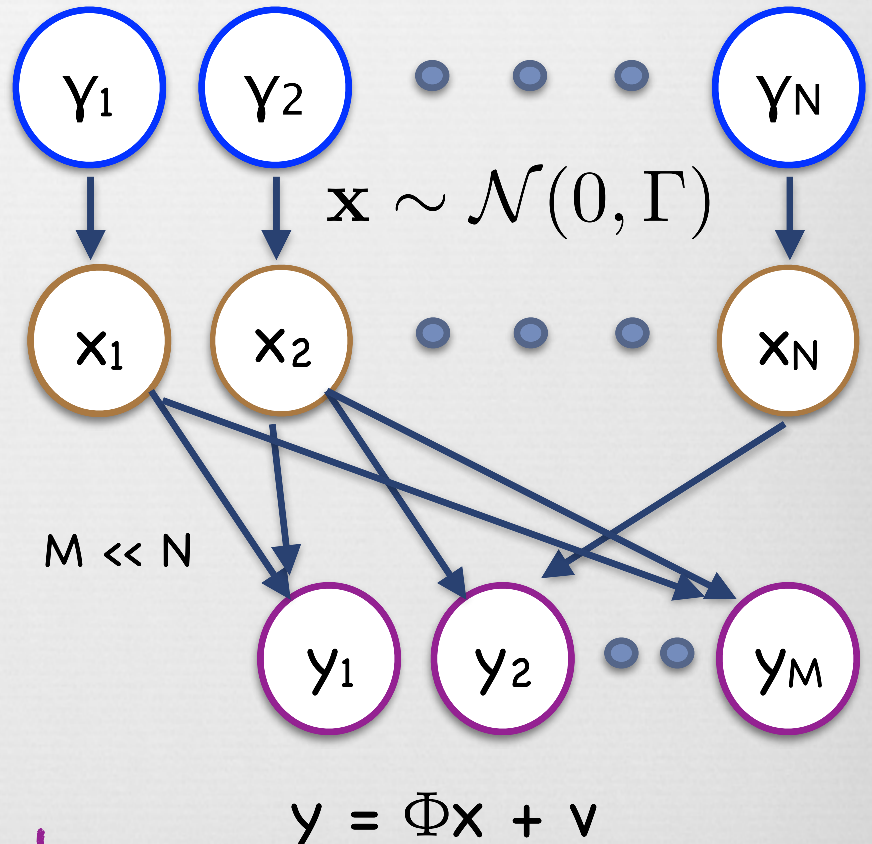
$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{\Phi x}\|_2^2\right)$$

- Parameterized Gaussian prior:

$$p(x_i;\gamma_i) = \frac{1}{\sqrt{2\pi\gamma_i}} \exp\left(-\frac{x_i^2}{2\gamma_i}\right), \ \gamma_i \geq 0$$

Graphical Model

- Markov chain (graphical model): $\gamma \rightarrow x \rightarrow y$
- $p(x;\gamma)$ Gaussian - leads to tractable algorithms
- Given γ, $p(x|y;\gamma)$ is Gaussian: easy to find point estimates
- But we don't know γ
- When in doubt, approximate! Find $p(\mathbf{x}|\mathbf{y};\hat{\gamma})$ instead



Approach

- First, estimate hyperparameters: $\hat{\gamma} = \arg\max_{\gamma} p(\gamma|\mathbf{y})$
  - γ : deterministic and unknown, or random with hyperprior distbn.
- Then, find posterior distribution p(x|y;$\hat{\gamma}$)

$$p(\mathbf{x}|\mathbf{y};\hat{\gamma}) = \mathcal{N}(\mu_x, \Sigma_x)$$

$$\mu_x = \hat{\Gamma}\Phi^T \left(\Phi\hat{\Gamma}\Phi^T + \lambda\mathbf{I}\right)^{-1}\mathbf{y}$$

$$\Sigma_x = \hat{\Gamma} - \hat{\Gamma}\Phi^T \left(\Phi\Gamma\Phi^T + \lambda\mathbf{I}\right)^{-1}\Phi\hat{\Gamma}$$

- For point estimates: e.g., posterior mean: $\mathbb{E}(\mathbf{x}|\mathbf{y};\hat{\gamma})$

Estimating the Hyperparameters

- Estimate γ_i from the data: Type-II ML

$$\mathcal{L}(\Gamma) = \log p(\mathbf{y};\Gamma) = \log \int p(\mathbf{y}|\mathbf{x};\Gamma)p(\mathbf{x};\Gamma)\mathrm{d}\mathbf{x}$$

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{x} + \mathbf{v} \qquad p(\mathbf{y};\Gamma) = \mathcal{N}\left(0, \underbrace{\sigma^2\mathbf{I} + \Phi\Gamma\Phi^T}_{\Sigma_{\mathbf{y}}}\right)$$

- SBL cost function:

$$\mathcal{L}(\Gamma) \propto -\log\det(\Sigma_{\mathbf{y}}) - \mathbf{y}^T\Sigma_{\mathbf{y}}^{-1}\mathbf{y}$$

Optimization via EM

- Log likelihood of the complete data

$$-\log p(\mathbf{y},\mathbf{x};\gamma) = \frac{\|\mathbf{y}-\Phi\mathbf{x}\|_2^2}{2\sigma^2} + \frac{1}{2}\left[\sum_{i=1}^{N}\frac{x_i^2}{\gamma_i} + \log\gamma_i\right]$$

$-\log p(\mathbf{y}|\mathbf{x};\gamma)$ — indep. of γ

$-\log p(\mathbf{x};\gamma)$ — func. of γ

- E-Step: compute "Q-function"

$$Q\left(\Gamma|\Gamma^{(t)}\right) = \mathbb{E}_{\mathbf{x}|\mathbf{y};\Gamma^{(t)}}\left[-\log p(\mathbf{y},\mathbf{x};\Gamma)\right]$$

$$\doteq \sum_{i=1}^{N}\frac{\mathbb{E}(x_i^2|\mathbf{y};\Gamma^{(t)})}{\gamma_i} + \log\gamma_i$$

from previous iteration

- Easy to compute: $p(x_i|\mathbf{y};\Gamma^{(t)})$ is Gaussian

The EM Iterations

- E-step (continued): $p\left(\mathbf{x}|\mathbf{y};\Gamma^{(t)}\right) = \mathcal{N}(\mu, \Sigma)$

$$\Sigma = \left(\sigma^{-2}\Phi^T\Phi + \left(\Gamma^{(t)}\right)^{-1}\right)^{-1} \qquad \mu = \sigma^{-2}\Sigma\Phi^T\mathbf{y}$$

- M-step: maximize $Q(\Gamma|\Gamma^{(t)})$ given posteriors gathered in the E-step:

$$\mathbb{E}(x_i^2|\mathbf{y};\Gamma^{(t)})$$

$$\Gamma^{(t+1)} = \arg\max_{\gamma_i \geq 0} Q\left(\Gamma|\Gamma^{(t)}\right) = \mathrm{diag}\left(\mu_i^2 + \Sigma_{ii}\right)$$

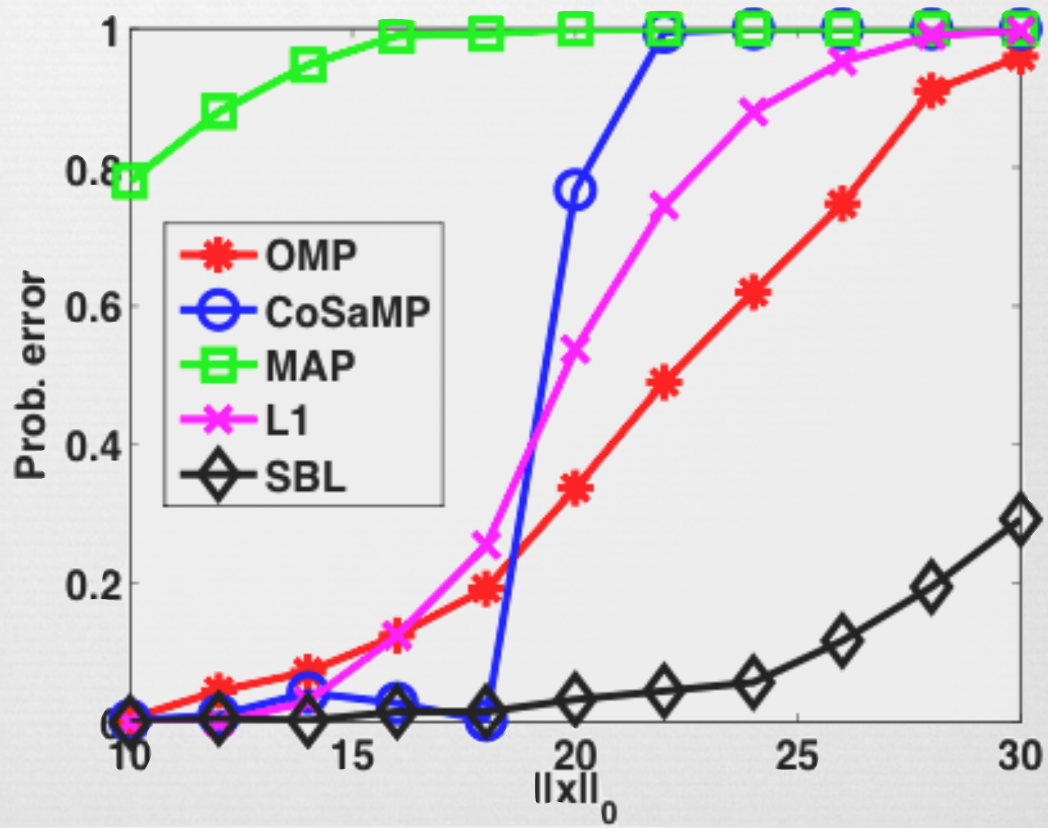
- Component-wise updates

The SBL Algorithm

1. Initialize $\Gamma = I$
2. Compute
$$\Sigma = \left(\sigma^{-2}\Phi^T\Phi + \left(\Gamma^{(t)}\right)^{-1}\right)^{-1}$$
$$\mu = \sigma^{-2}\Sigma\Phi^T\mathbf{y}$$
3. Update $\Gamma^{(t+1)} = \mathrm{diag}\left(\mu_i^2 + \Sigma_{ii}\right)$
4. Repeat steps 2 and 3
5. Output μ after convergence

Empirical Example

- Generate random 50 x 100 matrix A
- Generate sparse vector x_0
- Compute $y = Ax_0$
- Solve for x_0, average over 1000 trials
- Repeat for different sparsity values



Highly scaled nonzero entries

Convergence

- Convergence guaranteed to a fixed pt. of L from any initialization (property of EM)
- The global min of L occurs at the sparsest solution in the noiseless case [Wipf et al. 04]
- All local minima occur at sparse solutions in the noisy case [Wipf et al. 04]
- Strictly better than MAP estmn. with a factorial prior [Wipf and Nagarajan 09]
  - Always has fewer local minima
  - Global min. at global optimum of l_0 min.

Other Options

- McKay updates [Tipping, 2001]
  - Set gradient of SBL cost = 0
  - Faster convergence than EM
- Greedy approach:
  - Update hyperparams one at a time [Tipping & Faul, 2003]
    - Closed-form update for each hyperparam
    - Fast, but can get trapped in a local min.
  - Fast Bayesian matching pursuit [Schniter et al., 08]

Other Options

- Use dual-form of SBL. Cost function:

$$\mathbf{x}_{\text{opt}} = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi x}\|_2^2 + \sigma^2 g_{\text{SBL}}(\mathbf{x})$$

$$g_{\text{SBL}}(\mathbf{x}) \triangleq \min_{\gamma \geq 0} \mathbf{x}^T \Gamma^{-1} \mathbf{x} + \log\det\left(\sigma^2 \mathbf{I} + \Phi \Gamma \Phi^T\right)$$

  - Facilitates iterative reweighted l_1 and l_2 algorithms [Wipf and Nagarajan, 09]

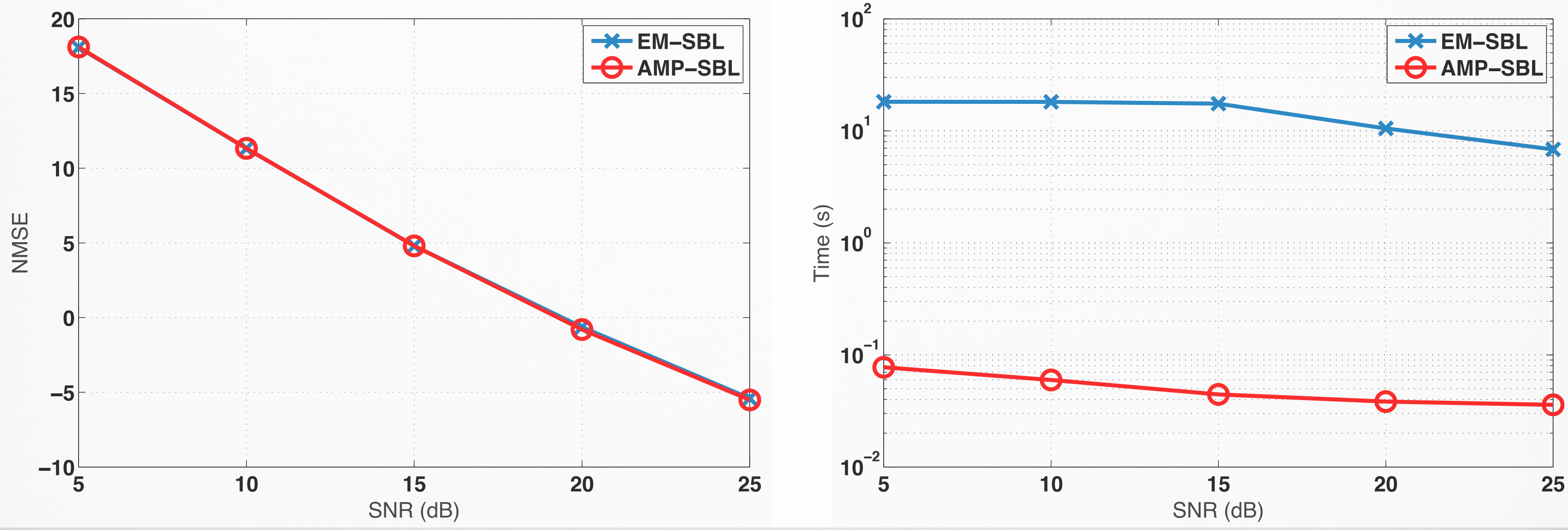
- Replace E-step with an approx. posterior computation: AMP-SBL [Al-Shoukairi and Rao 14]

Approximate Message Passing

- AMP [Donoho, Maleki, Montanari 09]:
  - Uses loopy belief propagation + Gaussian approximations to solve LASSO
  - Key advantage: low complexity
- In SBL:
  - All Gaussian PDFs: approximation is not necessary
  - Only need to track means and variances
  - Can replace computationally expensive E-step with the AMP based iterations

Empirical Example

- N = 200, M = 100, K = 20, Gaussian measurement matrix



[Al-Shakouri et al., 14]

Advantages of SBL

- Averaging over x: fewer minima in p(y;γ)
- Get an estimate of the error in recovery
- Allows for "exact inference"
- Versatile: γ can also be used to tie several params. together - easier to estimate
- Useful extensions: incorporate structure
  - Intra/inter-vector correlation
    - SBL allows the use of Kalman framework
  - Block/cluster sparsity
  - Colored noise (rank-deficient cov.)

To Recap

- Sparse Bayesian learning
  - Sparse vector recovery via estimating hyperparameters
  - Expectation-maximization iterations
  - Convergence properties
  - Alternative implementations
- Limitations
  - Computational complexity
    - More recent algos overcome this
  - Slow convergence
    - Fast versions exist, but without the same convergence guarantees

Part 4: Joint Sparse Signal Recovery

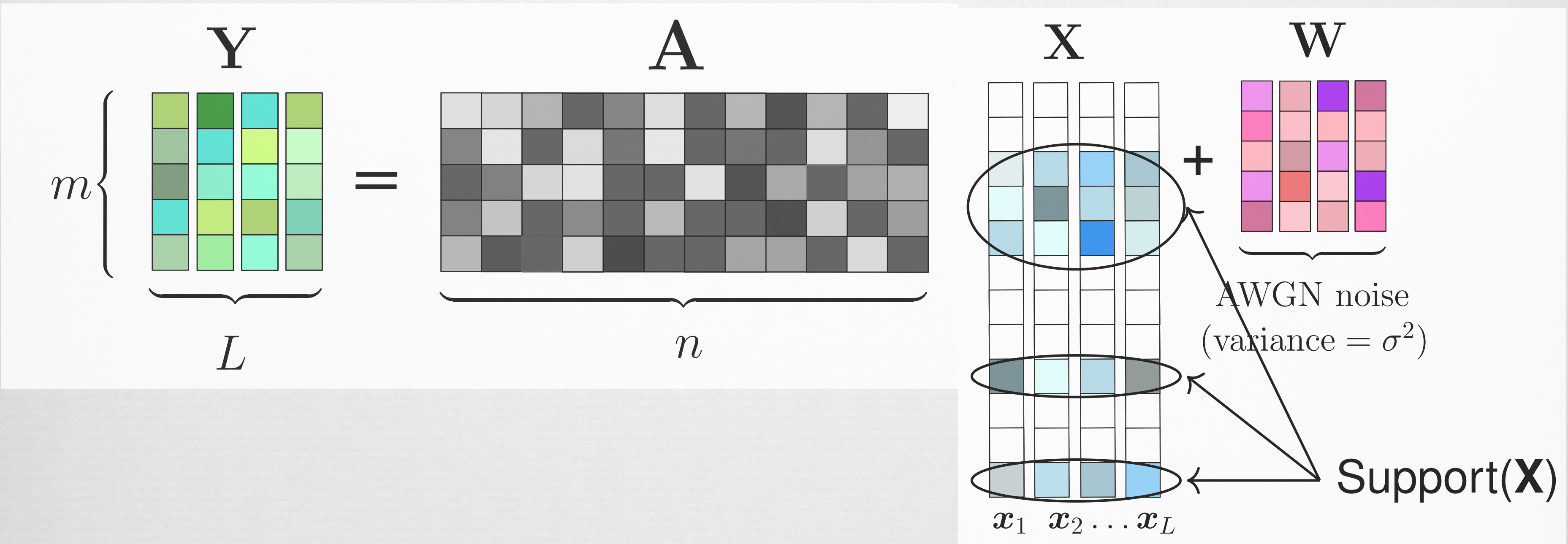
One important variant of the sparse recovery problem

Outline

- The joint sparse signal recovery problem
- l_0 bound in support recovery
- Sparse Bayesian Learning (the MMV version)
  - Performance guarantees and new insights
- New algorithms
  - Covariance matching framework
- New theoretical results
  - Restricted isometry of Khatri-Rao product

Joint Sparse Recovery Problem

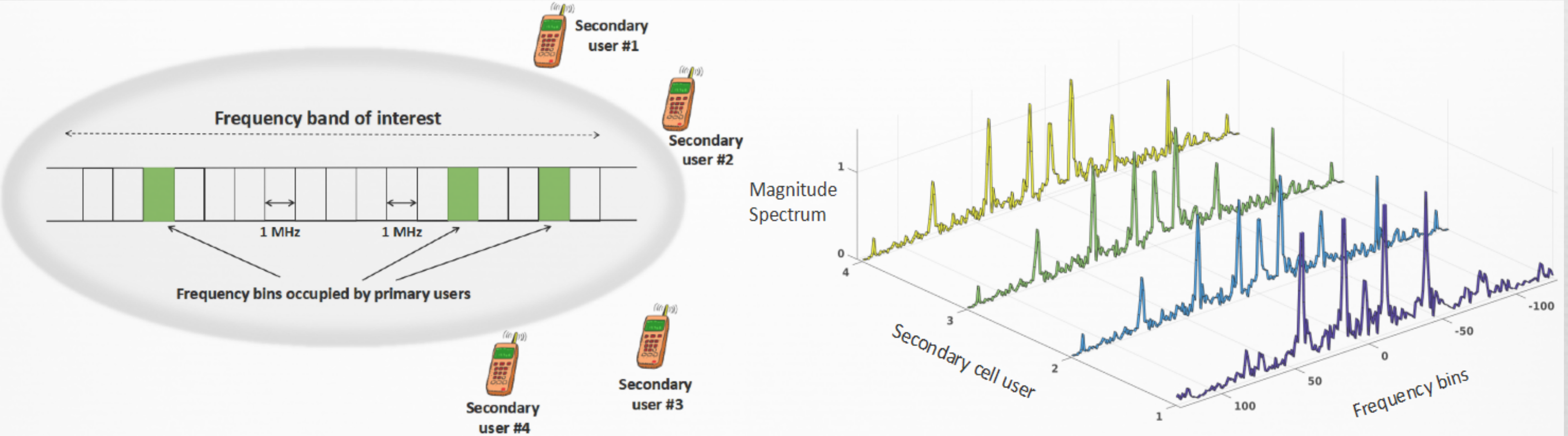
Observation model



Let k = number of nonzero rows in X.

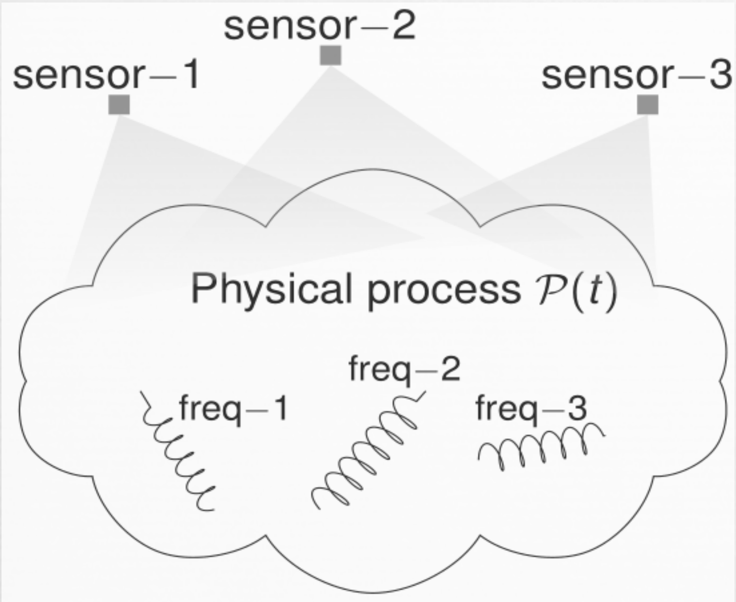
Want to recover X or support(X) from the Multiple Measurement Vectors (MMVs) Y

Wideband Spectrum Sensing



- Magnitude spectrum across secondary cell users is approximately jointly sparse.
- Exploit structure to improve accuracy.

Multi-sensor Data



- Signals acquired by different sensors have overlapping subspaces

⇩

- Approximate as different linear combinations of the same elementary signals

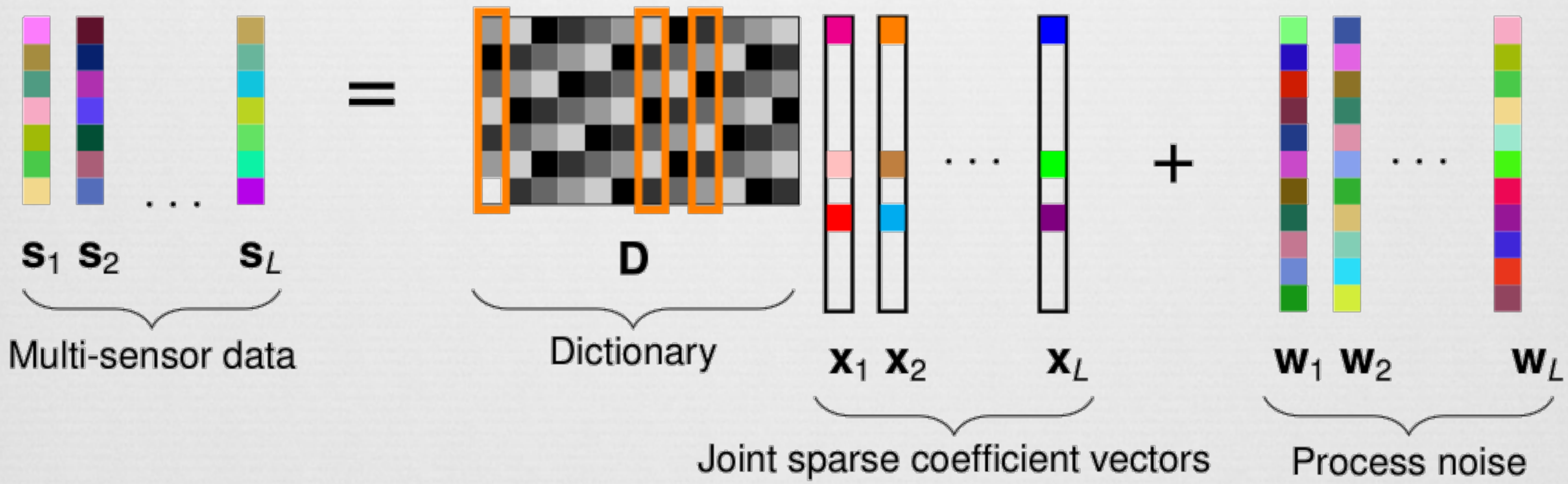
Why do subspaces overlap?

Commonality of physical process being sensed

Overlapping sensing regions

Generative Model for Multi-sensor Data

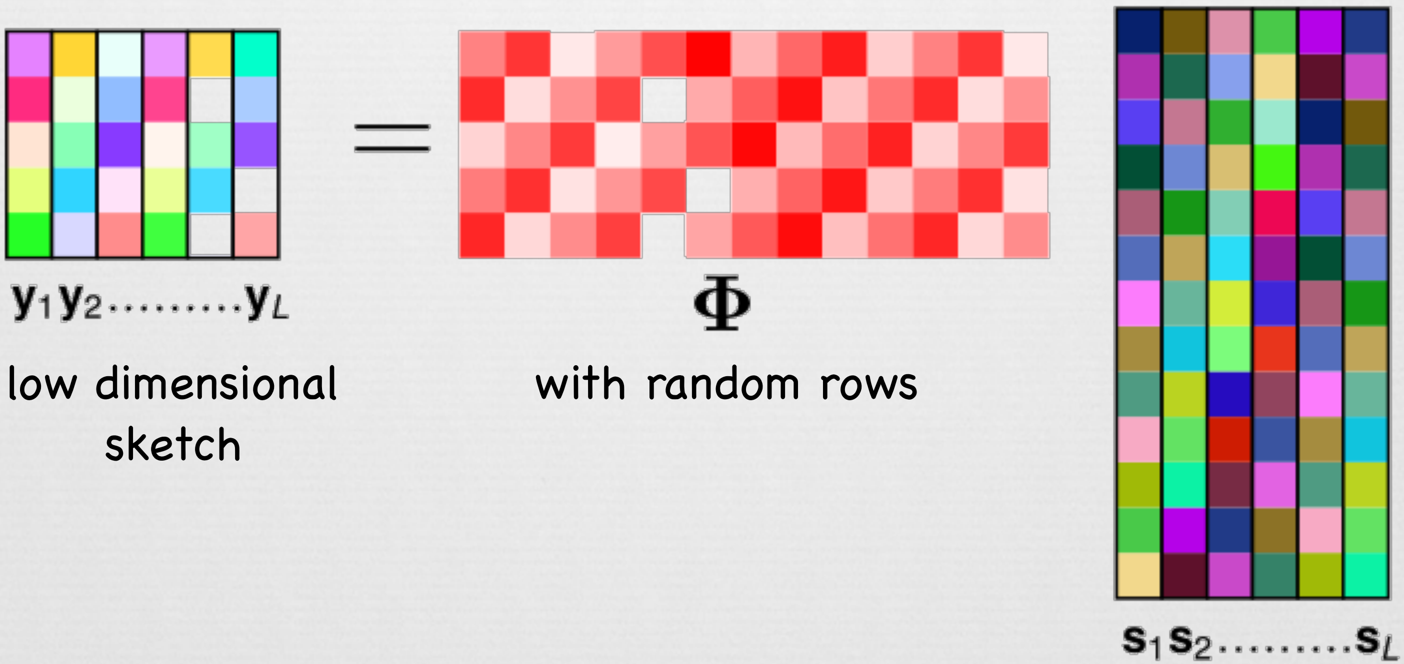
- Simultaneous sparse approximation (SSA) [Tropp 04]



Overlapping subspace = column-space(D_s).

Compression of Multi-sensor Data

- Linear encoder

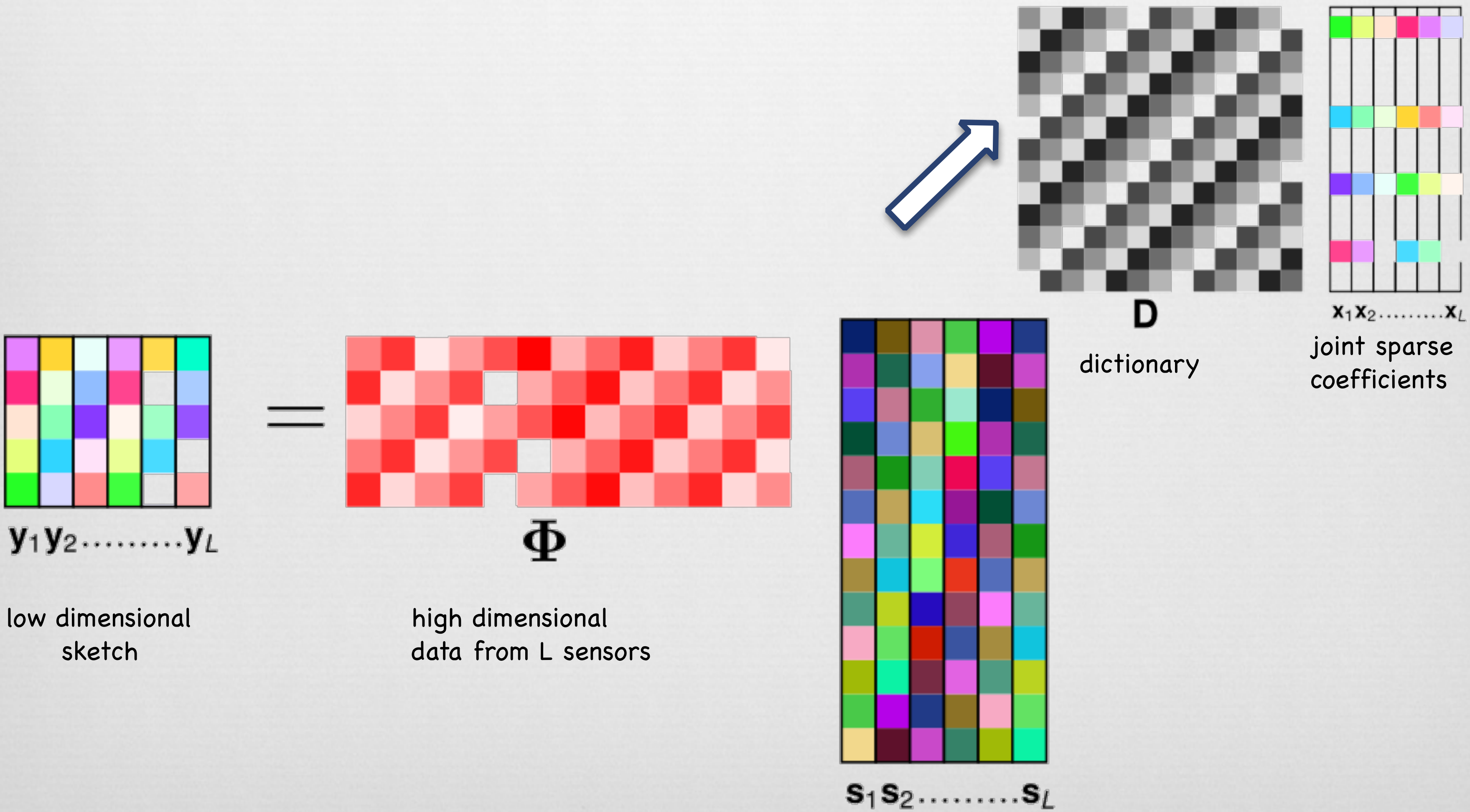


low dimensional sketch

with random rows

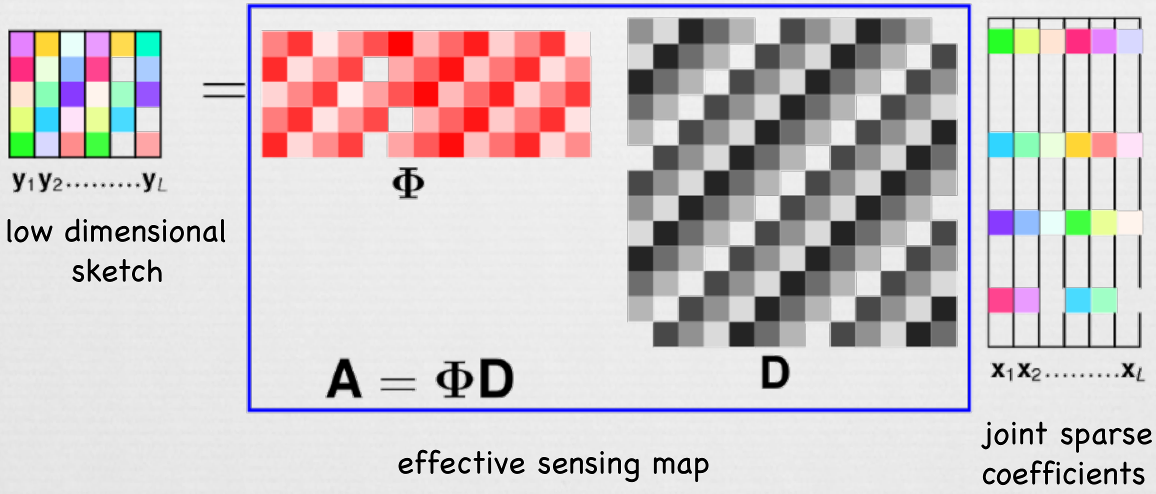
high dimensional data from L sensors

Two-Stage Decoder



Two-Stage Decoder

- Two stage decoder

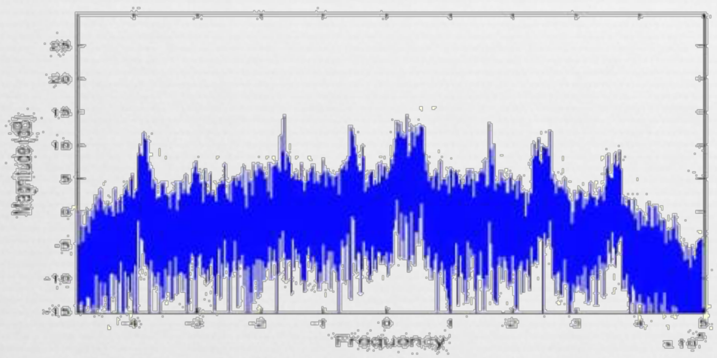


Stage 1: Recover joint-sparse X_{est} from sketch Y

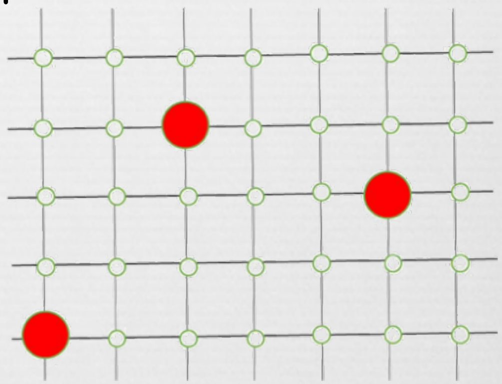
Stage 2: $S_{est} = D\ X_{est}$

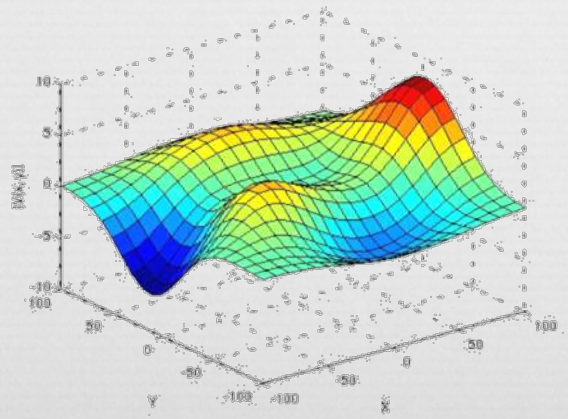
Support Recovery is also Important

Wideband spectrum sensing



Sparse event localization





Subspace filtering
by projecting to
common signal subspace

L_0 Bound

- Canonical L_0 problem:

$$(\mathbf{L_0}): \quad \min_{\mathbf{X}} \mathcal{R}(\mathbf{X}) \quad \text{subject to } \mathbf{Y} = \mathbf{\Phi X}.$$

$$\mathcal{R}(\mathbf{X}) = \text{No. of nonzero rows in } \mathbf{X}$$

- Unique k-sparse solution if [Chen & Huo, 06]

$$k < \frac{\text{Spark}(\mathbf{\Phi}) - 1 + \text{Rank}(\mathbf{Y})}{2} \qquad (\ell_0 - \text{bound})$$

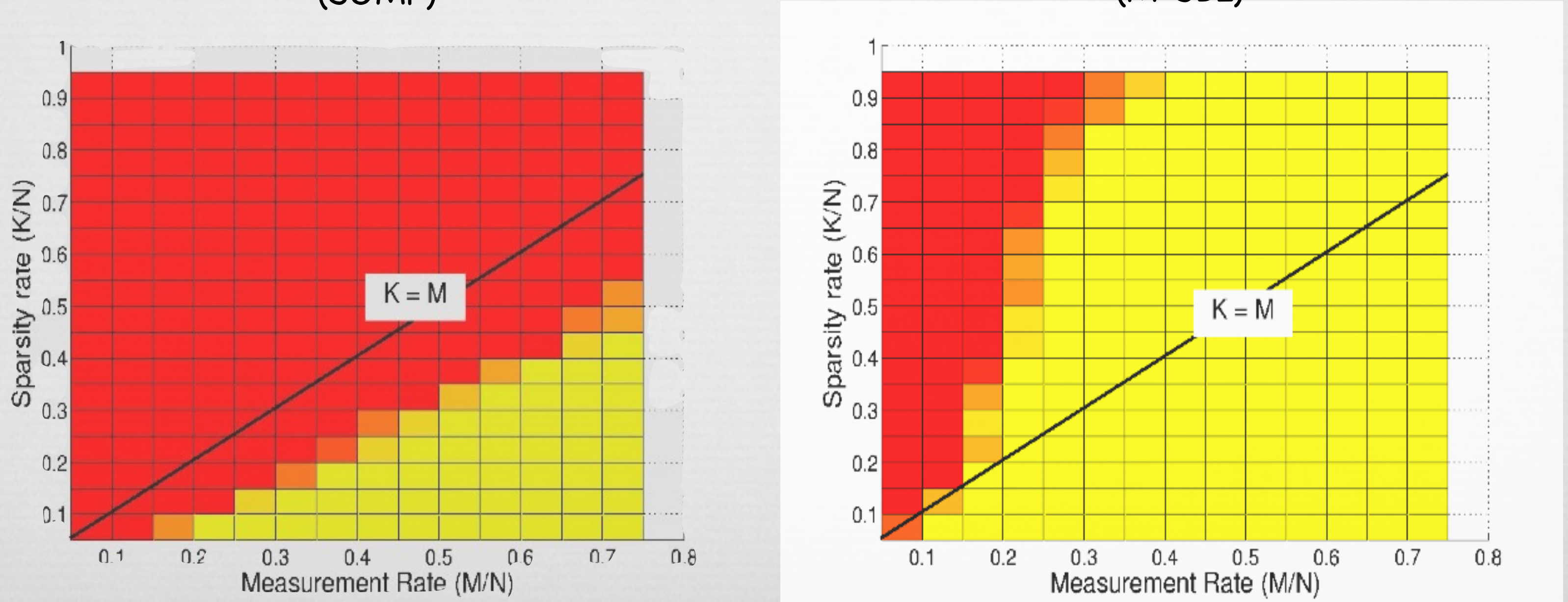
  - Spark($\mathbf{\Phi}$) = min. num. of lin. dep. cols in $\mathbf{\Phi}$.
  - L_0 bound on num. of nonzero rows: k < m.

Support Recovery Beyond l_0 Bound

- Supports of size k > m are recoverable!
- Key idea: Use correlation-aware priors

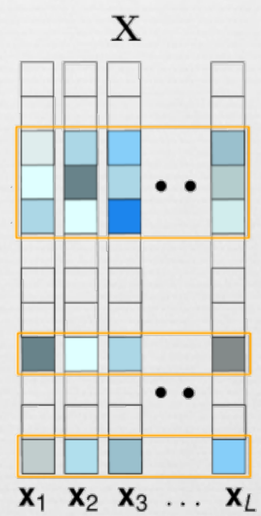
Support recovery phase transition

Simultaneous Orthogonal Matching Pursuit (SOMP) | Sparse Bayesian Learning (M-SBL)



Correlation Awareness

- Latent structure within joint sparse vectors!



Zero intra-vector correlation

$$\mathbb{E}\left[\mathbf{x}_j(i)\mathbf{x}_j(k)\right] \approx 0$$

Correlation-aware prior [Pal & Vaidyanathan, 15]

$$\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(\mathbf{0}, \Gamma), \qquad \Gamma = \mathrm{diag}(\gamma).$$

MSBL-Sparse Bayesian Learning using MMVs

- Observation model: $\mathbf{Y} = \mathbf{\Phi X} + \mathbf{W}$
- Correlation-aware prior: $\mathbf{x}_j \overset{i.i.d.}{\sim} \mathcal{N}(0, \mathbf{\Gamma}),\ \mathbf{\Gamma} = \mathrm{diag}(\gamma)$
  - Common $\mathbf{\Gamma}$ enforces same support in columns of X.
  - Gaussian MMVs:

$$\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2\mathbf{I} + \mathbf{\Phi\Gamma\Phi}^T)$$

- M-SBL algorithm: $\hat{\gamma} = \underset{\gamma \in \mathbb{R}_+^n}{\mathrm{argmax}}\ \log p(\mathbf{Y}; \gamma)$
  - Nonconvex objective
  - Solved via Expectation Maximization (EM)
  - Estimated support = support($\hat{\gamma}$).

The M-SBL Algo

- Cost function

$$p(\mathbf{Y};\gamma) = \int p(\mathbf{Y},\mathbf{X};\gamma)\mathrm{d}\mathbf{X} = \prod_{j=1}^{L} \int p(\mathbf{y}_j|\mathbf{x}_j)p(\mathbf{x}_j;\gamma)\mathrm{d}\mathbf{x}_j$$

- Key point: γ couples the sparsity pattern across xj
  - Fewer parameters to estimate: N << (N x L)
- EM iterations

$$\text{E-step: } Q(\gamma|\gamma^k) = \mathbb{E}_{\mathbf{X}|\mathbf{Y},\gamma^k}\left[\log p(\mathbf{Y},\mathbf{X};\gamma)\right]$$

$$\text{M-step: } \gamma^{k+1} = \arg\max_{\gamma\in\mathbb{R}_+^N} Q(\gamma|\gamma^k)$$

- Posterior distbn.: $p\left(\mathbf{x}_j|\mathbf{y}_j;\gamma^k\right) \sim \mathcal{N}\left(\mu_j^{k+1},\Sigma_j^{k+1}\right)$

E & M Steps

- E Step:

$$\Sigma_j^{k+1} = \Gamma^k - \Gamma^k \Phi_j^T \left(\sigma_j^2 \mathbf{I}_M + \Phi_j \Gamma^k \Phi_j^T\right)^{-1} \Phi_j \Gamma^k$$

$$\mu_j^{k+1} = \sigma_j^{-2} \Sigma_j^{k+1} \Phi_j^T \mathbf{y}_j$$

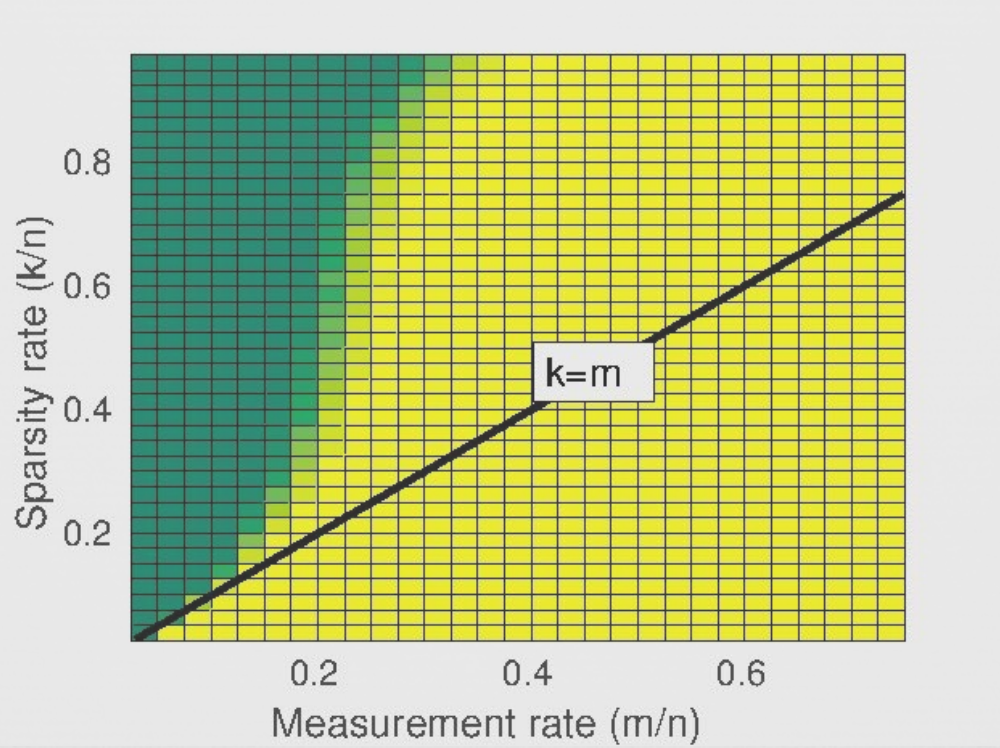
- M Step:

$$\gamma^{k+1}(i) = \frac{1}{L} \sum_{j=1}^{L} \mu_j^{k+1}(i)^2 + \Sigma_j^{k+1}(i,i)$$

  - Average of the individual estimates of γ_i across measurements

Performance of MSBL

Support recovery phase transition



n=200,
L=400,
SNR=20 dB

Recoverable support size k grows as $O(m^2)$!

Part 5: Performance Guarantees for Sparse Bayesian Learning

Sufficient conditions for support recovery by M-SBL

Sufficient Conditions for Support Recovery in SBL

- Single measurement vector ($L = 1$)
- Noiseless observations
- Result: SBL correctly recovers the support for all $1 \le k < \text{spark}(\Phi) - 1$
  - spark: min. number of lin. dep. cols
  - Usually, in CS, $\text{spark}(\Phi) = m + 1$
- For l1 recovery, $1 \le k \le O(m / \log N)$

Sufficient Conditions for Support Recovery in MSBL

- Suppose $x_1, x_2, \ldots, x_L$ have common support S^* of size k. Nonzero entries i.i.d. zero mean Gaussian with variance in $[\gamma_{\min}, \gamma_{\max}]$. Then,

$$\mathbb{P}\left(\text{supp}(\hat{\gamma}) \neq \mathcal{S}^*\right) \leq \exp\left(-\frac{\eta L}{8}\right)$$

- Under Conditions 1 & 2 (next slide)

[Khanna and M., CoRR abs/1703.04930 (2017)]

Conditions 1 & 2

- Condition 1: The self Khatri-Rao product matrix $\mathbf{\Phi} \odot \mathbf{\Phi}$ has a strictly positive minimum singular value

- Condition 2:

$$\eta \geq \inf_{\gamma \in \Theta_K} \frac{\|(\mathbf{\Phi} \odot \mathbf{\Phi})(\gamma - \gamma^*)\|_2^2}{\left(\sigma^2 + \gamma_{\max}\|\mathbf{\Phi}_{\mathcal{S}\cup\mathcal{S}^*}^T \mathbf{\Phi}_{\mathcal{S}\cup\mathcal{S}^*}\|_2\right)^2}$$

Proof Outline

- Error event $$\mathcal{E}_{\mathcal{S}^*} = \bigcup_{\mathcal{S}\in\mathcal{S}_K\setminus\{\mathcal{S}^*\}} \left\{ \max_{\boldsymbol{\gamma}\in\Theta(\mathcal{S})} \mathcal{L}(\mathbf{Y};\boldsymbol{\gamma}) \geq \max_{\boldsymbol{\gamma}\in\Theta(\mathcal{S}^*)} \mathcal{L}(\mathbf{Y};\boldsymbol{\gamma}) \right\}$$
- Apply union bound, replace first max. by a finite sized union over an epsilon net
- Use a large deviation property on the likelihood function to bound error prob.
$$\mathbb{P}(\mathcal{E}_{\mathcal{S}^*}) \leq \sum_{\mathcal{S}\in\mathcal{S}_K\setminus\mathcal{S}^*} \sum_{\boldsymbol{\gamma}\in\Theta^\epsilon(\mathcal{S})} \exp\left(-L\mathcal{D}^*_{1/2}/4\right)$$
- Lower bound the worst case exponent
$$\mathcal{D}^*_{1/2} \geq \eta$$
- Hence
$$\mathbb{P}(\mathcal{E}_{\mathcal{S}^*}) \leq \exp\left(-L\left(\frac{\eta}{4} - \frac{\log|\mathcal{S}_K|}{L} - \frac{\log(\max_{\mathcal{S}\in\mathcal{S}_K}|\Theta^\epsilon(\mathcal{S})|)}{L}\right)\right)$$
- Bound the cardinality of the epsilon net
  - Lipschitz const. of the log likelihood

When will Condition 1 hold?

- Theorem (paraphrased): Suppose $\mathbf{\Phi}$ has real i.i.d. Gaussian(0,1) entries. Then,

$$\mathbb{P}\left(\delta_k\left(\frac{\mathbf{\Phi}}{\sqrt{m}}\odot\frac{\mathbf{\Phi}}{\sqrt{m}}\right)\geq\delta\right)\leq 5n^{-2(\beta-1)}$$

for all β > 1, provided

$$m\geq 4c_3\beta\left(\frac{k\log n}{\delta}\right)$$

If A has i.i.d. Gaussian entries,

$$\mathbb{P}\left(\delta_k\left(\frac{\mathbf{A}}{\sqrt{m}}\right)>\delta\right)\leq\frac{1}{n^\alpha}$$

provided

$$m\geq\frac{c}{\delta^2}(k+\alpha)\log n$$

[Foucart & Rauhut, Thm. 9.27]

Proof Outline

- Point of departure

$$\delta_k\left(\frac{\mathbf{\Phi}}{\sqrt{m}}\odot\frac{\mathbf{\Phi}}{\sqrt{m}}\right)=\sup_{\mathbf{z}\in\mathbb{R}^n,\|\mathbf{z}\|_2=1,\|\mathbf{z}\|_0\le k}\left|\left\|\left(\frac{\mathbf{\Phi}}{\sqrt{m}}\odot\frac{\mathbf{\Phi}}{\sqrt{m}}\right)\mathbf{z}\right\|_2^2-1\right|$$

- Union bound Pr(RHS ≥ δ). Use

$$(\mathbf{\Phi}\odot\mathbf{\Phi})^T(\mathbf{\Phi}\odot\mathbf{\Phi})=\mathbf{\Phi}^T\mathbf{\Phi}\circ\mathbf{\Phi}^T\mathbf{\Phi}$$

- Then use Gaussian tail bounds and the Hanson Wright inequality:

$$\mathbb{P}\left\{\left|\mathbf{x}^T\mathbf{A}\mathbf{x}-\mathbb{E}\mathbf{x}^T\mathbf{A}\mathbf{x}\right|>t\right\}\le 2\exp\left[-c\min\left(\frac{t^2}{K^4\,||\mathbf{A}||_{HS}^2},\frac{t}{K^2\,||\mathbf{A}||}\right)\right]$$

Implications

- In MSBL, for a fixed n, we get perfect support recovery w.h.p., with noisy meas. & even if k > m
  - For $1 \leq k \leq \text{krank}(\mathbf{\Phi} \odot \mathbf{\Phi})/2$
  - $\text{krank}(\mathbf{\Phi} \odot \mathbf{\Phi})$: largest p s.t. any p cols of $\mathbf{\Phi} \odot \mathbf{\Phi}$ are linearly independent
  - For suitable[1] $\mathbf{\Phi}$, $\text{krank}(\mathbf{\Phi} \odot \mathbf{\Phi}) = \mathcal{O}(m^2)$
  - Sparsity level up to O(m²) is potentially recoverable!

[1] P. Pal and P. P. Vaidyanathan, "Correlation-Aware Techniques for Sparse Support Recovery", SSP Workshop, 2012.

New Interpretation of M-SBL Cost Function

- M-SBL cost:

$$-\log p(\mathbf{Y};\gamma) = -\sum_{j=1}^{L} \log \mathcal{N}\left(\mathbf{y}_j; 0, \sigma^2 \mathbf{I}_m + \mathbf{\Phi}\Gamma\mathbf{\Phi}^T\right)$$

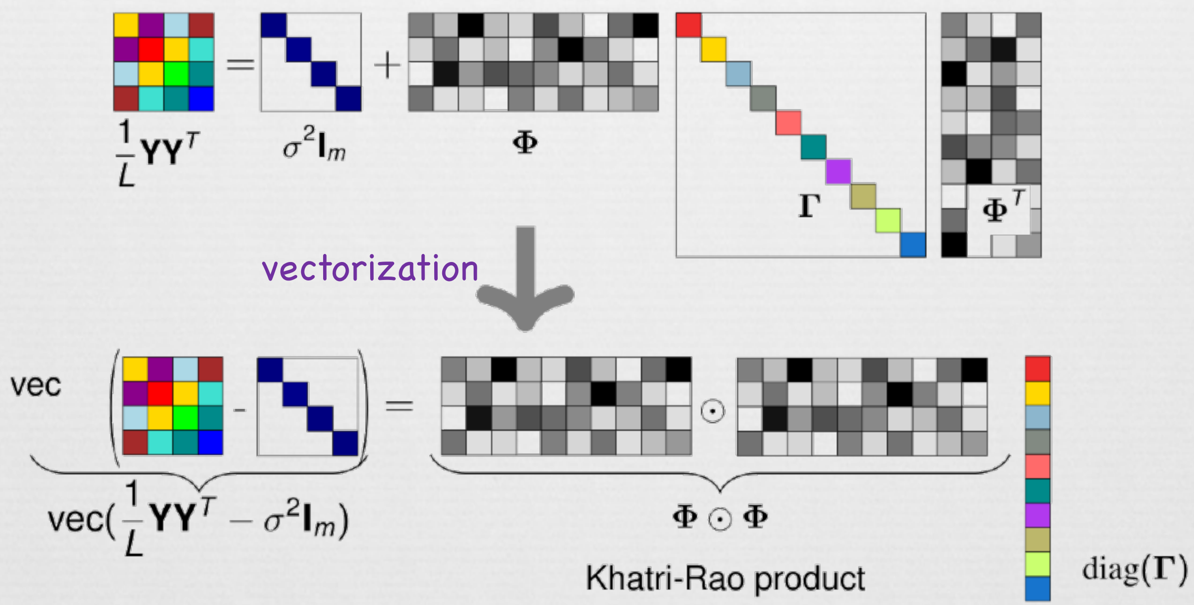
$$\propto \log|\sigma^2\mathbf{I}_m + \mathbf{\Phi}\Gamma\mathbf{\Phi}^T| + \mathrm{tr}\left(\left(\sigma^2\mathbf{I}_m + \mathbf{\Phi}\Gamma\mathbf{\Phi}^T\right)^{-1}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right)\right)$$

$$\propto \mathcal{D}^{\text{Bregman}}_{-\log\det}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T, \sigma^2\mathbf{I}_m + \mathbf{\Phi}\Gamma\mathbf{\Phi}^T\right) + \text{const. terms}$$

- Motivates covariance matching based approaches to sparse recovery
- Can we use some other divergence?

A Closer Look at Covariance Matching

- Covariance matching constraints: $\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \approx \sigma^2\mathbf{I}_m + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T$

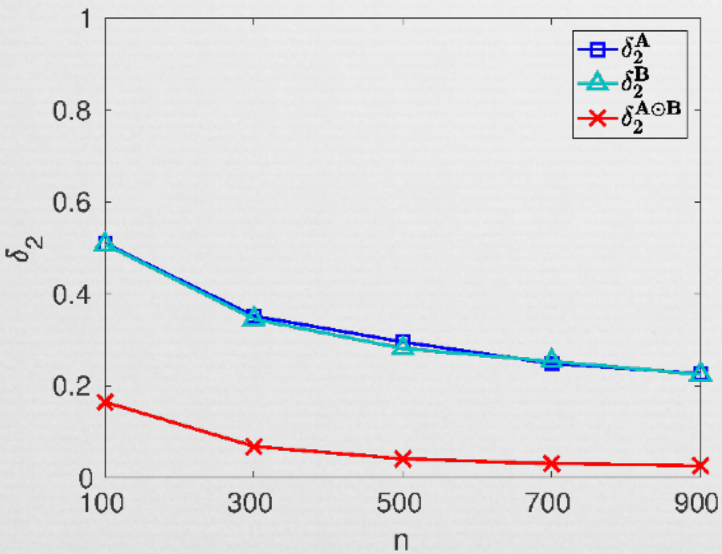


- Restricted isometry of $\mathbf{\Phi} \odot \mathbf{\Phi}$ ensures robust recovery of sparse γ
- Also, $\gamma \geq 0$. Want a non-negative null space property

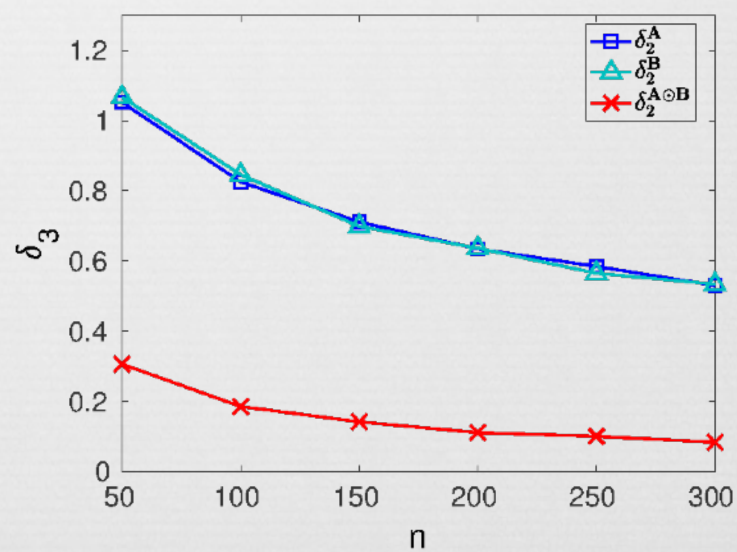
Empirical Study

- $A_{ij}, B_{ij} \sim N(0, 1/m)$ & $m = 0.5n$

2-RIC



3-RIC



RIP improved by taking Khatri-Rao product!

Deterministic RIC bound for Khatri-Rao product

- Theorem: [Khanna & M., TSP 2018]
  For $m \times n$ sized matrices $\mathbf{A}$ and $\mathbf{B}$ with unit norm columns,

$$\delta_k\left(\mathbf{A} \odot \mathbf{B}\right) \;\leq\; \left[\max\left(\delta_k(\mathbf{A}), \delta_k(\mathbf{B})\right)\right]^2$$

- Proof technique:
  - $(\mathbf{A} \odot \mathbf{B})^T (\mathbf{A} \odot \mathbf{B}) = \mathbf{A}^T\mathbf{A} \circ \mathbf{B}^T\mathbf{B}$
  - Kantorovitch matrix inequalities

- Useful corollary: $\delta_k(\mathbf{A} \odot \mathbf{A}) \leq (\delta_k(\mathbf{A}))^2 < \delta_k(\mathbf{A})$ RIP improves!

Part 6: New Algorithms

Covariance matching is the key!

Covariance Matching framework

- Observation Model:

$$\mathbf{Y} = \mathbf{\Phi}\mathbf{X} + \mathbf{W}$$

Correlation-aware Gaussian prior

$$\mathbf{x}_j \sim \mathcal{N}(0, \mathrm{diag}(\gamma))$$

$$\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2\mathbf{I}_m + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T)$$

- Principle:

Parametrized covariance matrix

$$\hat{\gamma} = \arg\min_{\gamma \in \mathbb{R}_+^n} \mathbf{dist}\left(\frac{1}{L}\mathbf{Y}\mathbf{Y}^T, \sigma^2\mathbf{I}_m + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T\right)$$

Empirical covariance matrix

Support estimate = Support($\hat{\gamma}$)

Covariance Matching Algorithm 1

- Distance = Frobenius norm
- Algorithm = CO-LASSO [Pal & Vaidyanathan, 15]

$$\hat{\gamma} = \arg\min_{\gamma \in \mathbb{R}^n_+} \left|\left|\left| \frac{1}{L}\mathbf{Y}\mathbf{Y}^T - (\sigma^2\mathbf{I} + \boldsymbol{\Phi}\boldsymbol{\Gamma}\boldsymbol{\Phi}^T) \right|\right|\right|_F^2 + \lambda \|\gamma\|_1$$

  - Convex objective
  - High memory requirements (to store $\Phi \odot \Phi$)

Covariance Matching Algorithm 2

- Distance = Log-Det Bregman Matrix Divergence
- Algorithm = M-SBL [Wipf & Rao, 07]

$$\hat{\gamma} = \arg\min_{\gamma \in \mathbb{R}_+^n} \log \left| \sigma^2 \mathbf{I} + \mathbf{\Phi} \mathbf{\Gamma} \mathbf{\Phi}^T \right| + \operatorname{tr}\left(\left(\sigma^2 \mathbf{I} + \mathbf{\Phi} \mathbf{\Gamma} \mathbf{\Phi}^T \right)^{-1} \left(\frac{1}{L} \mathbf{Y} \mathbf{Y}^T \right) \right)$$

- Nonconvex objective (optimize using EM)
- Slow convergence
- Best performance

Covariance Matching Algorithm 3

- Distance = Log-Det Jensen Difference
- Algorithm = Rényi Divergence based Covariance Matching Pursuit (RD-CMP) [Khanna & M., 17]
  - A new correlation-aware prior

$$\mathbf{x}_j \sim \mathcal{N}(0, \gamma \mathrm{diag}(\mathbf{1}_S))$$

Covariance matrix parameterized by support S

  - Induces Gaussian distributed MMVs

$$\mathbf{y}_j \sim \mathcal{N}(0, \sigma^2 \mathbf{I}_m + \gamma \mathbf{\Phi}_S \mathbf{\Phi}_S^T)$$

  - Covariance matching with α-Rényi divergence

$$\hat{\mathcal{S}} = \underset{\mathcal{S} \subseteq [n]}{\mathrm{argmin}}\, \mathcal{D}_\alpha \left(\mathcal{N}\left(0, \frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right), \mathcal{N}\left(0, \sigma^2 \mathbf{I}_m + \gamma \mathbf{\Phi}_S \mathbf{\Phi}_S^T\right) \right)$$

Rényi Divergence based Covariance Matching Pursuit

- Covariance matching using α-Rényi Divergence

$$\hat{\mathcal{S}} = \underset{\mathcal{S}\subseteq[n]}{\operatorname{argmin}}\ \mathcal{D}_\alpha\left(\mathcal{N}\left(0, \frac{1}{L}\mathbf{Y}\mathbf{Y}^T\right), \mathcal{N}\left(0, \sigma^2\mathbf{I}_m + \gamma\boldsymbol{\Phi}_\mathcal{S}\boldsymbol{\Phi}_\mathcal{S}^T\right)\right)$$

$$\hat{\mathcal{S}} = \underset{\mathcal{S}\subseteq[n]}{\operatorname{argmin}}\ \underbrace{\log\left|(1-\alpha)\frac{1}{L}\mathbf{Y}\mathbf{Y}^T + \alpha\left(\sigma^2\mathbf{I} + \gamma\boldsymbol{\Phi}_\mathcal{S}\boldsymbol{\Phi}_\mathcal{S}^T\right)\right|}_{f(\mathcal{S}),\ \text{submodular in } \mathcal{S}} - \underbrace{\alpha\log\left|\sigma^2\mathbf{I} + \gamma\boldsymbol{\Phi}_\mathcal{S}\boldsymbol{\Phi}_\mathcal{S}^T\right|}_{g(\mathcal{S}),\ \text{submodular in } \mathcal{S}}$$

  - Generalizes M-SBL cost
  - RD-CMP objective = difference of submodular funcs.

Submodular functions: a primer

- $\mathcal{V}$ = ground set of elements
- Set func. $f : \mathcal{V} \rightarrow \mathbb{R}_+$ submodular if, for $\mathcal{S} \subseteq \mathcal{T} \subseteq \mathcal{V}$,
  - f is Monotonic: $f(\mathcal{S}) \leq f(\mathcal{T})$
  - f follows Law of Diminishing Returns

$$f(\mathcal{T} \cup \{a\}) - f(\mathcal{T}) \leq f(\mathcal{S} \cup \{a\}) - f(\mathcal{S}) \quad \forall a \in \mathcal{V} \backslash \mathcal{T}$$

- Examples:
  - Rank of a matrix
  - Joint entropy
  - $f(\mathcal{S}) = \log \left| \mathbf{A} + \gamma \mathbf{B}_{\mathcal{S}} \mathbf{B}_{\mathcal{S}}^T \right|$ is submodular in S for A, γ >0

Submodular functions: a primer

- Maximizing a submodular function subject to cardinality constraints is easy!
    - A Greedy algorithm will maximize a submodular function f to within $\left(1-\frac{1}{e}\right) f_{opt}$
- Tight modular upper bnd for submodular f

$$f(\mathcal{S}) \leq h^f_{\mathcal{X}}(\mathcal{S}) \triangleq f(\mathcal{X}) - \sum_{j \in \mathcal{X} \setminus \mathcal{S}} \left[f(\mathcal{X}) - f(\mathcal{X} \setminus \{j\}) \right] + \sum_{j \in \mathcal{S} \setminus \mathcal{X}} \left[f(j) - f(\phi) \right]$$

Rényi Divergence based Covariance Matching Pursuit

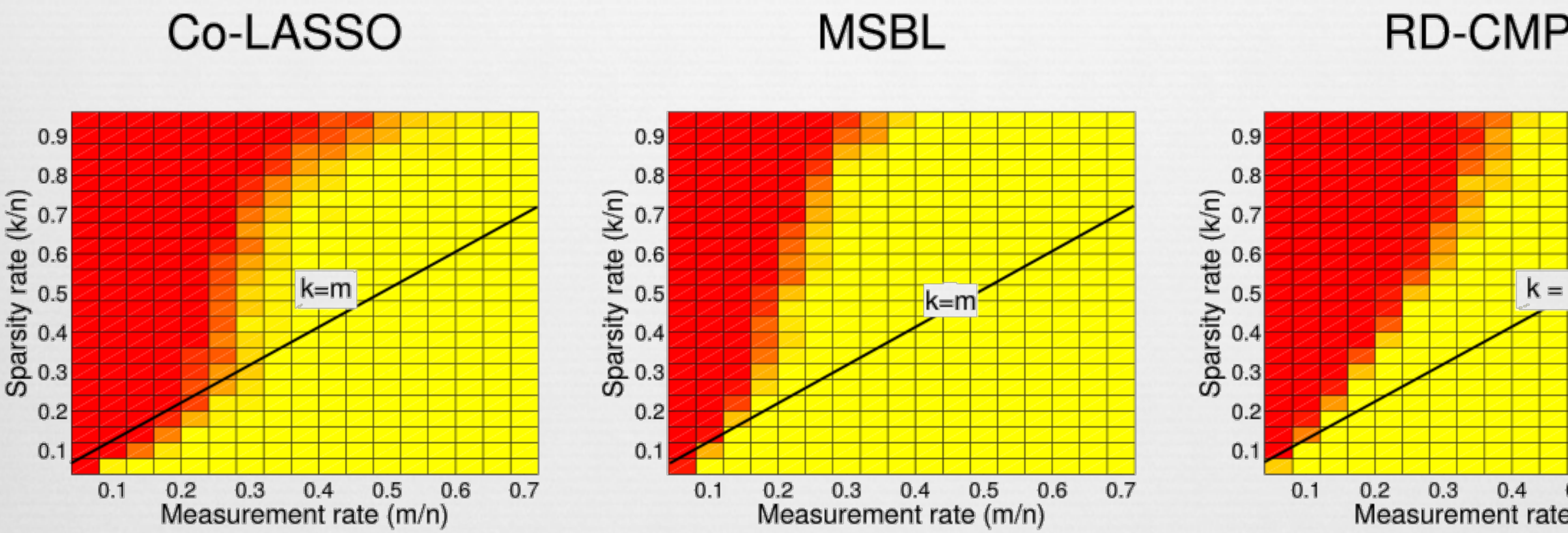
- RD-CMP objective

$$\hat{\mathcal{S}} = \underset{\mathcal{S}\subseteq[n]}{\operatorname{argmin}} \underbrace{\log\left|(1-\alpha)\frac{1}{L}\mathbf{Y}\mathbf{Y}^T + \alpha\left(\sigma^2\mathbf{I} + \gamma\mathbf{\Phi}_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}}^T\right)\right|}_{f(\mathcal{S}),\ \text{submodular in } \mathcal{S}} - \alpha \underbrace{\log\left|\sigma^2\mathbf{I} + \gamma\mathbf{\Phi}_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}}^T\right|}_{g(\mathcal{S}),\ \text{submodular in } \mathcal{S}}$$

- Find $\hat{S}$ via majorization-minimization
  - Majorization
    - Replace 1st log-det term f(S) by its modular upper bound $h_{S_t}(S)$: tight at S_t
  - Minimization

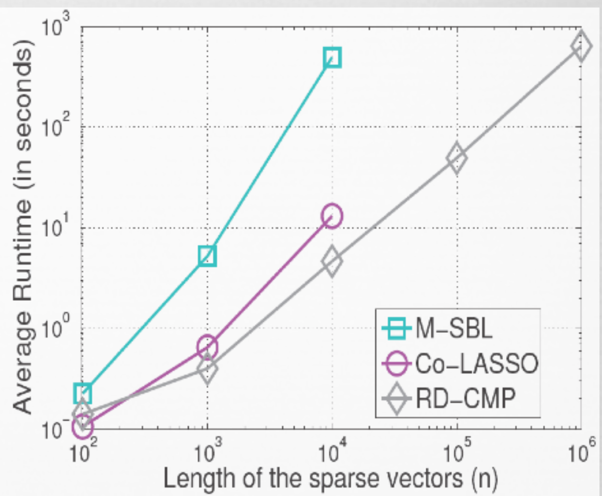
$$\mathcal{S}_{t+1} = \underset{\mathcal{S}\subseteq[n]}{\arg\min} \underbrace{h_{\mathcal{S}_t}(\mathcal{S}) - \alpha\log\left|\sigma^2\mathbf{I} + \gamma\mathbf{\Phi}_{\mathcal{S}}\mathbf{\Phi}_{\mathcal{S}}^T\right|}_{\text{Supermodular func., minimized by greedy search}}$$

RD-CMP Performance



SNR = 10 dB; n = 200; L = 200

SNR = 10 dB; k = 50 log n
m = 0.75 k, mL = 50 k log n

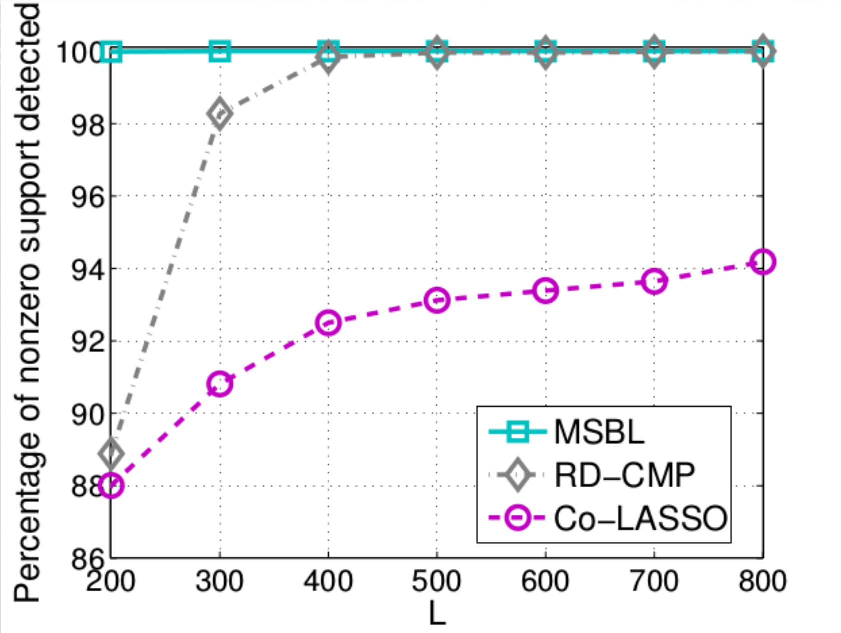


RD-CMP currently the FASTEST
covariance matching based MMV solver!

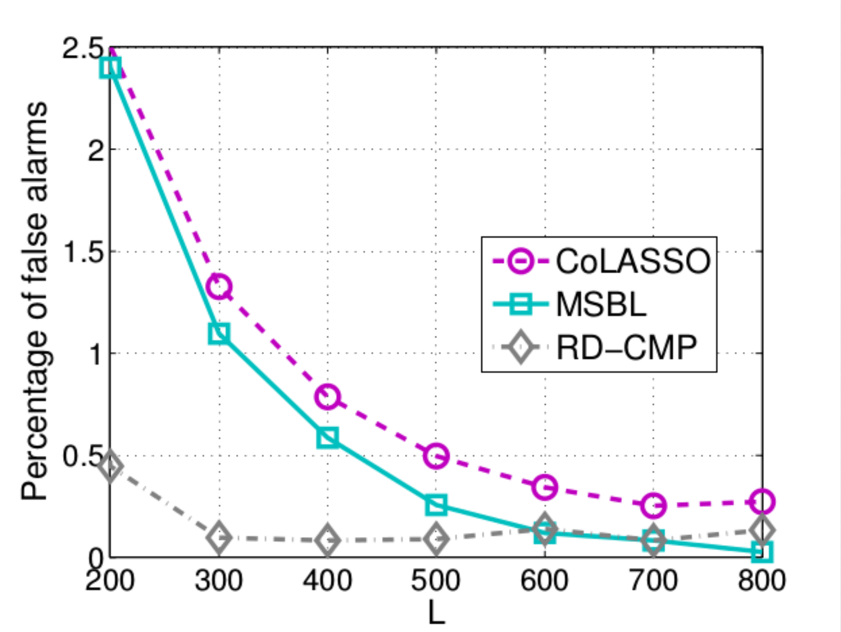
RD-CMP Performance

SNR=20 dB, n=500, k=200, m =100

Support detection rate



Support false alarm rate



RD-CMP performs better than Co-LASSO but slightly worse than M-SBL

Support Recovery via Non-negative Parameter Estimation

- Covariance matching principle:

  find sparse, nonnegative diagonal $\mathbf{\Gamma}$ s.t.

$$\frac{1}{L}\mathbf{Y}\mathbf{Y}^T \approx \sigma^2\mathbf{I}_m + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T$$

- Example: Recover $\mathbf{\Gamma}$ via nonnegative LASSO

$$\hat{\gamma} = \arg\min_{\gamma\in\mathbb{R}^n_+} \left|\left|\left|\frac{1}{L}\mathbf{Y}\mathbf{Y}^T - (\sigma^2\mathbf{I} + \mathbf{\Phi}\mathbf{\Gamma}\mathbf{\Phi}^T)\right|\right|\right|_F^2 + \lambda\,||\gamma||_1$$

Non-negative Least Squares Based Approach

- We write $\hat{\Sigma} = \frac{1}{L}\mathbf{Y}\mathbf{Y}^T = \Sigma + \mathbf{E}$
- In the noiseless case, $\Sigma = \mathbf{\Phi}\Gamma\mathbf{\Phi}^T$
  - Noisy case: L. Ramesh and M. ICASSP 18
- Vectorize: $\mathbf{r} \triangleq \mathrm{vec}(\hat{\mathbf{\Sigma}}) = (\mathbf{\Phi} \odot \mathbf{\Phi})\gamma + \mathbf{e}$
- Model e as Gaussian distributed with mean zero and covariance

$$\mathbf{W} \triangleq \mathrm{cov}(\mathbf{e}) = \frac{1}{L}(\mathbf{\Phi} \otimes \mathbf{\Phi})(\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}})\mathbf{B}(\Gamma^{\frac{1}{2}} \otimes \Gamma^{\frac{1}{2}})(\mathbf{\Phi} \otimes \mathbf{\Phi})^\top$$

- Here, B = cov(zz^T) where z is N(0, I_n). Can be found in closed form

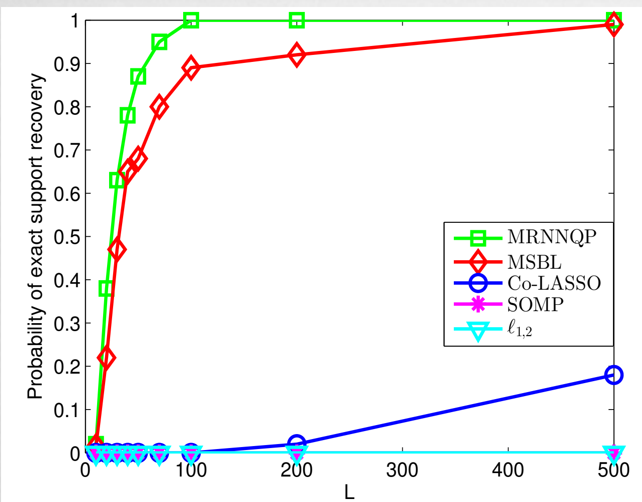
ML estimation of γ

- We seek to solve: $\gamma_{\mathrm{ML}} = \arg\max_{\gamma \geq 0} p(\mathbf{r}; \gamma)$
- Optimization problem:

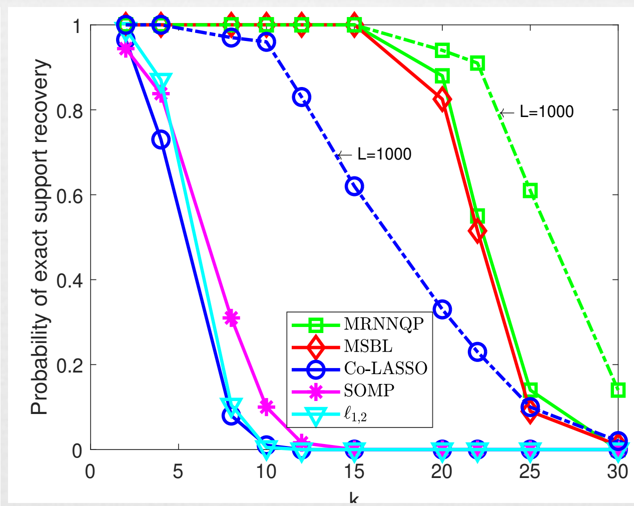
$$\gamma_{\mathrm{ML}} = \arg\min_{\gamma \geq 0} \; \log |\mathbf{W}| + (\mathbf{r} - \mathbf{A}\gamma)^\top \mathbf{W}^{-1} (\mathbf{r} - \mathbf{A}\gamma)$$

- For a given W, it is a non-negative weighted least-squares problem
  - Can solve using non-negative quadratic programming (NNQP)
- Reweighted minimization procedure: (1) Compute W, (2) solve an NNQP; iterate

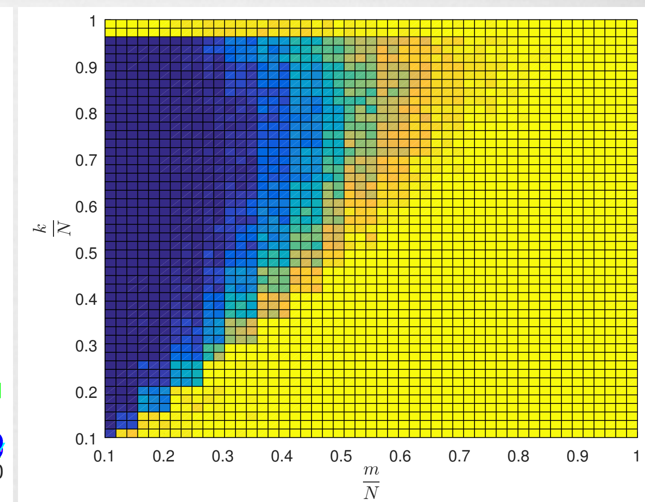
Performance



N = 40, M = 20, k = 25



N = 70, M = 20, L = 50, 1000



N = 20, L = 200

Non-negativity Encourages Sparsity!

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0 \quad (P_{0^+})$$
$$\text{s. t. } \mathbf{y} = \Phi\mathbf{x},\ \mathbf{x} \geq 0$$

$$\min_{\mathbf{x}} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \quad \text{(NNQP)}$$
$$\text{s. t. } \mathbf{x} \geq 0$$

- When do the two yield the same soln?
- [Uniqueness]: x_0 k-sparse & $y = \Phi x_0$. Then, x_0 is unique sol. to (P_{0^+}) iff every $\mathbf{v} \in \ker(\Phi)\backslash\{\mathbf{0}\}$ has at least (k+1) +ve or (k+1) -ve entries
- [Recoverability via NNQP]: x_0 is exactly recovered by NNQP if every $\mathbf{v} \in \ker(\Phi)\backslash\{\mathbf{0}\}$ has at least (k+1) +ve and (k+1) -ve entries

Non-Negative Parameter Estimation Framework

- Minimize a convex loss $L : S_+^n \to \mathbb{R}_+$
- [Tsuda, Rätsch and Warmuth, Matrix Exponentiated Gradient Updates for On-line Learning and Bregman Projection, 2005]

$$\mathbf{W}_{t+1} = \arg\min_{\mathbf{W} \in S_+^n} \mathcal{D}_F(\mathbf{W}, \mathbf{W}_t) + \eta L(\mathbf{W})$$

Proximal term — Tradeoff parameter — Loss

- Proximal term $\mathcal{D}_F(\mathbf{W}, \mathbf{W}_t)$: Bregman matrix divergence

Bregman Matrix Divergence

- Bregman matrix divergence

$$\mathcal{D}_F(\mathbf{W}, \mathbf{W}_t) = F(\mathbf{W}) - \underbrace{F(\mathbf{W}_t) - \operatorname{tr}\left(f(\mathbf{W}_t)^T(\mathbf{W} - \mathbf{W}_t)\right)}_{\text{first order approx. of } F(\mathbf{W}) \text{ around } \mathbf{W}_t},$$

  - $f = \nabla F$
  - Seed function F: strictly convex and differentiable

- Choice of F
  - F(W) = -log det (W)
  - F(W) = tr(W logW)

Matrix Exponentiated Gradient (MEG) Updates

$$\mathbf{W}_{t+1} = \arg\min_{\mathbf{W} \in S_+^n} \; \mathcal{D}_F(\mathbf{W}, \mathbf{W}_t) + \eta L(\mathbf{W})$$

Proximal term

Tradeoff parameter

Loss

MEG update: (zero gradient condition)

$$\mathbf{W}_{t+1} = f^{-1}\left(f(\mathbf{W}_t) - \eta \nabla L(\mathbf{W}_t)\right), \quad \text{where } f = \nabla F.$$

MEG Updates

- $F(\mathbf{W}) = -\log\det\mathbf{W}$

$$\mathcal{D}_F(\mathbf{W}, \mathbf{W}_t) = \log\frac{|\mathbf{W}_t|}{|\mathbf{W}|} + \operatorname{tr}\left(\mathbf{W}_t^{-1}\mathbf{W}\right) - n$$

Log-Det Bregman Matrix divergence

MEG update: $\mathbf{W}_{t+1} = \left((\mathbf{W}_t)^{-1} + \eta\nabla L(\mathbf{W}_t)\right)^{-1}$

- $F(\mathbf{W}) = \operatorname{tr}(\mathbf{W}\log\mathbf{W})$

$$\mathcal{D}_F(\mathbf{W}, \mathbf{W}_t) = \operatorname{tr}\left(\mathbf{W}\log\mathbf{W} - \mathbf{W}\log\mathbf{W}_t - \mathbf{W} + \mathbf{W}_t\right)$$

Von-Neumann Matrix divergence

MEG update: $\mathbf{W}_{t+1} = \exp\left(\log\mathbf{W}_t - \eta\left(\nabla L(\mathbf{W}_t)\right)\right)$

MEG Updates for Covariance Matching

- Loss $$L(\Gamma): \left|\left|\left|\mathbf{R_{yy}} - \Phi\Gamma\Phi^T\right|\right|\right|_F^2 + \lambda\,||\gamma||_1$$
- Parameter space: all p.s.d. diagonal matrices
- Log-Det divergence based MEG update

$$\gamma_{t+1}(i) = \gamma_t(i)\left(\frac{1}{1 + 2\eta\gamma_t(i)\left[\phi_i^T\left(\Phi\Gamma\Phi^T - \mathbf{R_{yy}}\right)\phi_i + \lambda\right]}\right), \quad \forall i \in [n]$$

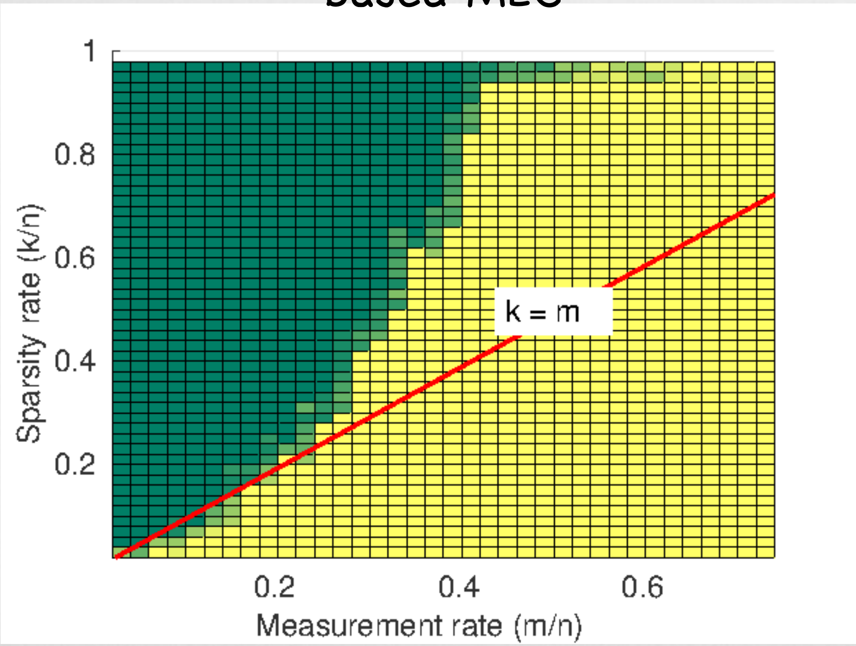
- Von-Neumann divergence based MEG update $\gamma_{t+1}(i) = \gamma_t(i) \cdot e^{-2\eta\left[\phi_i^T\left(\Phi\Gamma\Phi^T - \mathbf{R_{yy}}\right)\phi_i + \lambda\right]}, \quad \forall i \in [n].$

Performance

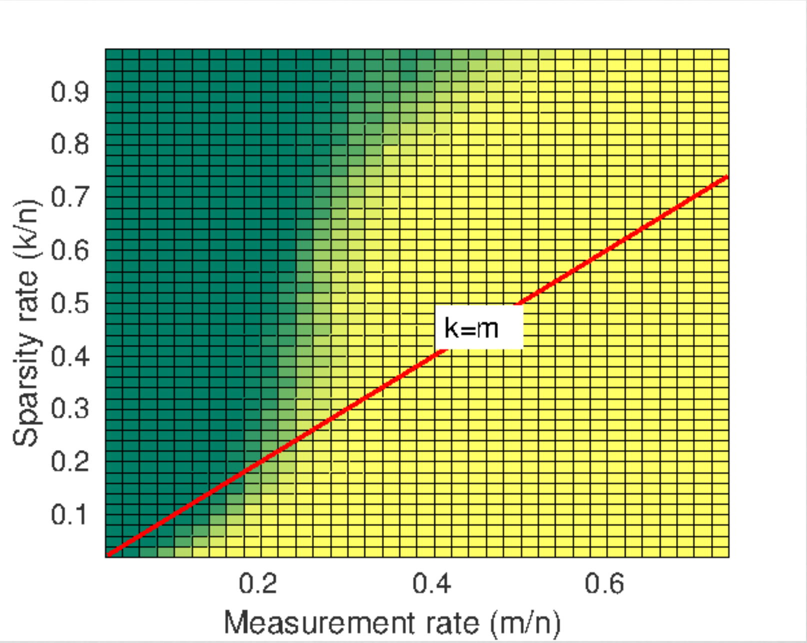
Support recovery phase transition

(n = 200, L = 400, SNR=20dB, λ = 0.25, η = 0.5)

Log-Det divergence based MEG



Von-Neumann divergence based MEG

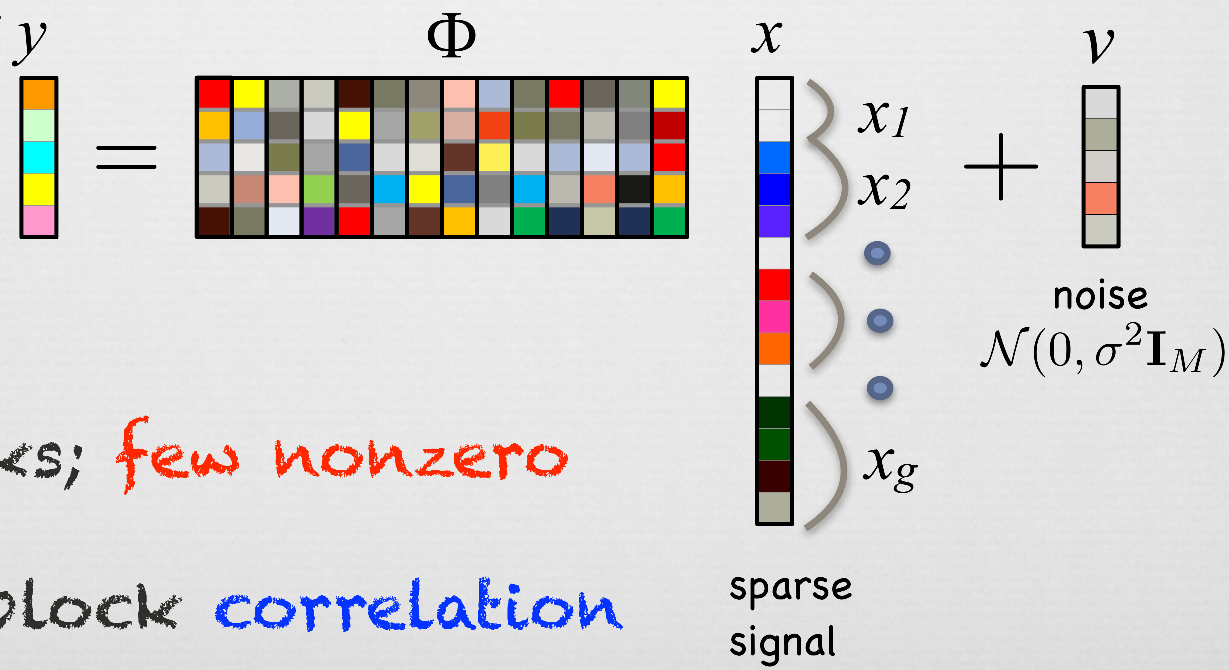


Part 7: Other Extensions

1. Cluster-sparsity, inter-vector correlation
2. Online sparse signal recovery
3. Distributed sparse signal recovery

Block Sparsity & Intra-Block Correlation

- Intra-vector correlation is often present, and is important to model & exploit



- g blocks; few nonzero
- Intra-block correlation

Block-Sparse Bayesian Learning Framework

- Measurement model: $\mathbf{y} = \Phi\mathbf{x} + \mathbf{v}$

$$\mathbf{x} = [\underbrace{x_1, \ldots x_{d_1}}_{\mathbf{x}_1^T}, \ldots, \underbrace{x_{d_{g-1}+1}, \ldots x_{d_g}}_{\mathbf{x}_g^T}]^T$$

- Parameterized prior

$$p(\mathbf{x}_i; \gamma_i, \mathbf{B}_i) \sim \mathcal{N}(0, \gamma_i \mathbf{B}_i),\ i = 1, 2, \ldots, g$$

  - γ_i controls sparsity
  - $\mathbf{B}_i$ controls intra-block correlation

Optimization Problem

- Posterior distribution

$$p\left(\mathbf{x}|\mathbf{y};\sigma^2,(\gamma_i\mathbf{B}_i)_{i=1}^g\right)\sim\mathcal{N}(\mu_x,\Sigma_x)$$

- where $\mu_x=\Sigma_0\Phi^T(\sigma^2\mathbf{I}+\Phi\Sigma_0\Phi^T)^{-1}\mathbf{y}$

$$\Sigma_x=\Sigma_0-\Sigma_0\Phi^T(\sigma^2\mathbf{I}+\Phi\Sigma_0\Phi^T)^{-1}\Phi\Sigma_0$$

$$\Sigma_0=\mathrm{diag}(\gamma_1\mathbf{B}_1,\ldots,\gamma_g\mathbf{B}_g)$$

- All params. can be estimated by maximizing:

$$\mathcal{L}(\Theta)=-2\log\int p(\mathbf{y}|\mathbf{x};\sigma^2)p(\mathbf{x};\Sigma_0)\mathrm{d}\mathbf{x}$$

$$=\log\det\left(\sigma^2\mathbf{I}+\Phi\Sigma_0\Phi^T\right)+\mathbf{y}^T\left(\sigma^2\mathbf{I}+\Phi\Sigma_0\Phi^T\right)^{-1}\mathbf{y}$$

Several Options for Optimization

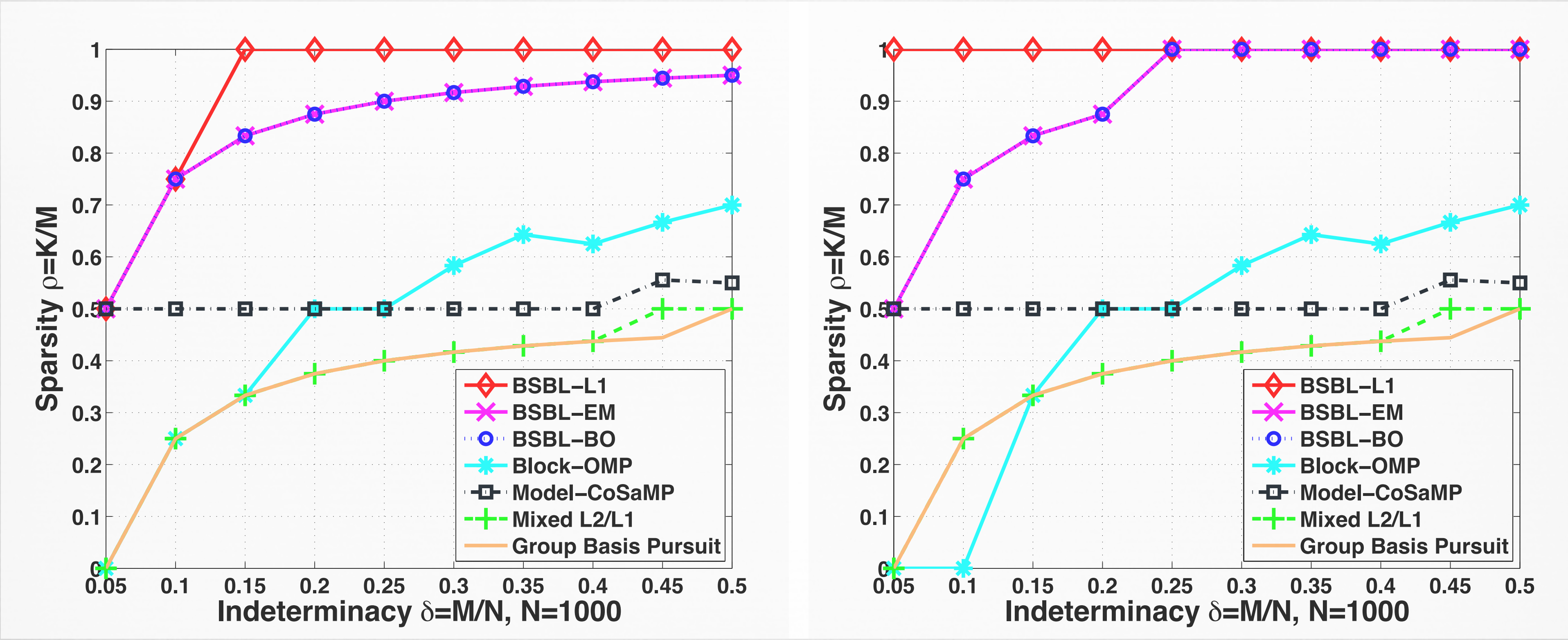
- BSBL-EM: Use expectation-maximization
- BSBL-BO: Use bounded optimization, i.e., a form of majorization-minimization
- BSBL-L1: Use a reweighted L1 procedure (special case of BSBL-BO)
- Different strategies offer a variety of performance-complexity tradeoffs

[Zhang et al. 2013]

Phase Transition

Correlation = 0

Correlation = 0.95



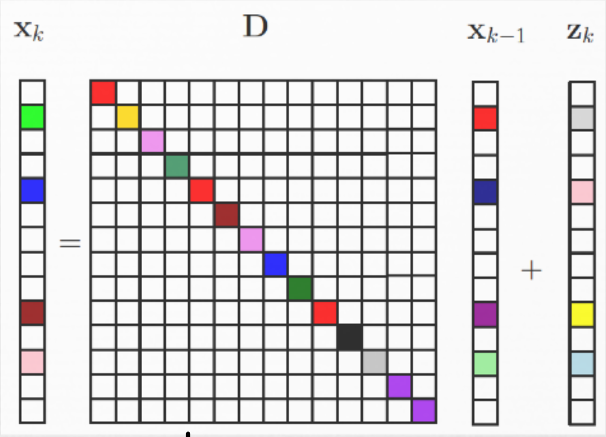
N = 1000, M = δ N, g = 40, block size = 25

Curves indicate > 99% success

[Zhang et al. 2013]

Delay-Constrained Sparse Vector Recovery

- Temporally correlated sparse vectors with a common support
- Goal: Given $y_{1:k}$ estimate $x_{k-\Delta}$
- Max. delay constraint between measurement and estimation



Offline approach:

M-step: Compute $\Gamma^{(r+1)}$ as a closed-form function of the state statistics

Mean: $\hat{\mathbf{x}}_{t|K} = \mathbb{E}\{\mathbf{x}_t|\mathbf{y}_{1:K}\}$

Autocov.: $\mathbf{P}_{t|K} = \text{cov}\{\mathbf{x}_t, \mathbf{x}_t|\mathbf{y}_{1:K}\}$

Cross-cov: $\mathbf{P}_{t,t-1|K} = \text{cov}\{\mathbf{x}_t, \mathbf{x}_{t-1}|\mathbf{y}_{1:K}\}$

E-step: Compute state statistics using fixed interval Kalman smoothing

Online Version

- Key idea: Update γ once as each input arrives and compute sparse vector

Non-iterative!

- Online recursion:

$$\gamma_k = \gamma_{k-1} + \frac{1}{k}\mathrm{Diag}\left\{\left(\mathbf{I} - \mathbf{D}^2\right)^{-1}\boxed{\mathbf{T}_{k|k+\Delta}} - \Gamma_{k-1}\right\}$$

fcn. of $\hat{\mathbf{x}}_{t|t+\Delta}, \mathbf{P}_{t|t+\Delta}, \mathbf{P}_{t,t-1|t+\Delta}$

- Two implementations of the Kalman filter:
  - fixed lag smoothing
  - sawtooth lag smoothing
- Main advantage: reduced computational and memory costs

[G. Joseph, M., TSP 2017]

Convergence Analysis

- Assume D=0 (uncorrelated sparse vecs)
- Simplified algorithm:

$$\gamma_k = \gamma_{k-1} + \frac{1}{k}\mathrm{diag}\{\mathbf{P}(\gamma_{k-1}) + \hat{\mathbf{x}}(\gamma_{k-1})\hat{\mathbf{x}}(\gamma_{k-1})^T - \Gamma_{k-1}\}$$

$$\mathbf{P}(\gamma) = \Gamma - \Gamma\mathbf{\Phi}^T\left(\mathbf{\Phi}\Gamma\mathbf{\Phi}^T + \sigma^2\mathbf{I}\right)^{-1}\mathbf{\Phi}\Gamma$$

$$\hat{\mathbf{x}}(\gamma) = \mathbf{P}(\gamma)\mathbf{\Phi}^T\mathbf{y}_k/\sigma^2$$

- Stochastic approximation recursion

$$\gamma_k = \gamma_{k-1} + \frac{1}{k}\mathbf{f}(\gamma_{k-1}) + \frac{1}{k}\mathbf{e}_k$$

Mean field function

$$\mathbf{f}(\gamma) = \mathbb{E}\left\{\mathrm{diag}\left(\mathbf{P}(\gamma) + \hat{\mathbf{x}}(\gamma)\hat{\mathbf{x}}(\gamma)^T - \Gamma\right)\right\}$$

$$\mathbf{e}_k = \mathrm{diag}\left\{\mathbf{P}(\gamma_{k-1}) + \hat{\mathbf{x}}(\gamma_{k-1})\hat{\mathbf{x}}(\gamma_{k-1})^T\right\} - \gamma_{k-1} - \mathbf{f}(\gamma_{k-1})$$

Martingale difference sequence

Convergence Result

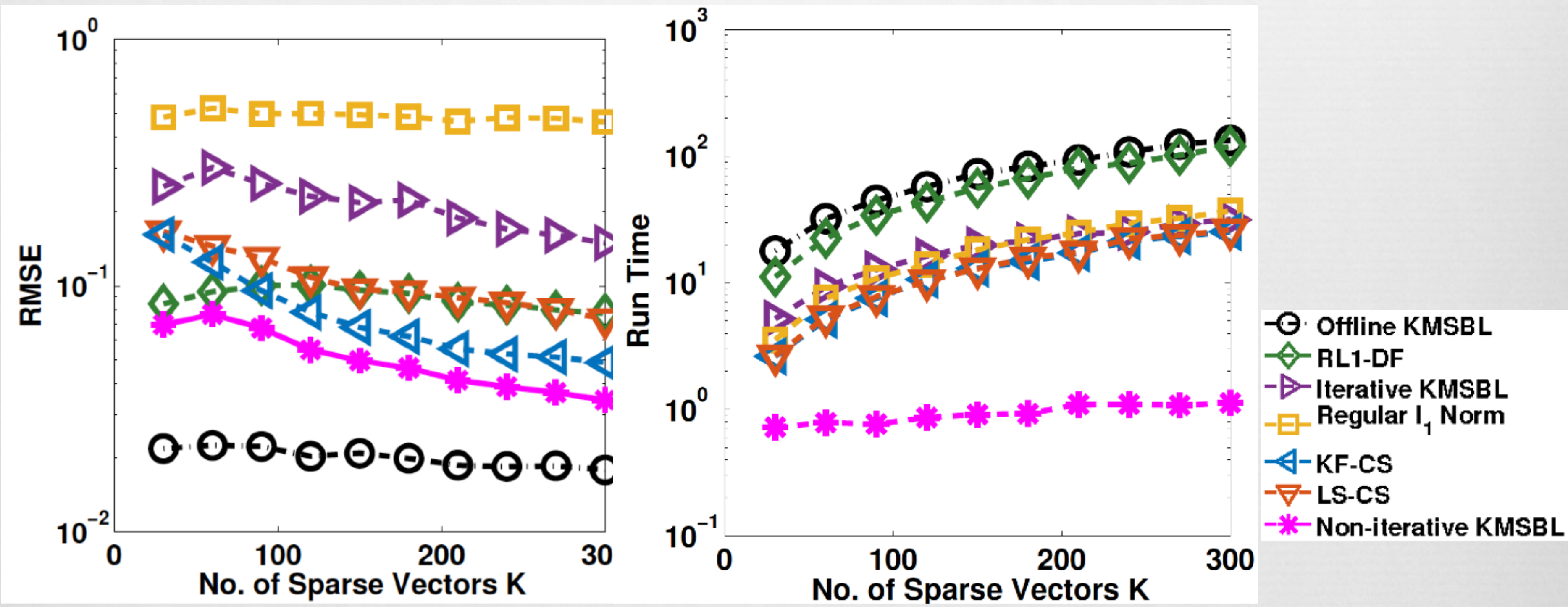
- Theorem:

  $\text{If Rank}\,\{\boldsymbol{\Phi} \odot \boldsymbol{\Phi}\} = N, \text{ then } \gamma_k \to \gamma \in \{\mathbf{0}, \gamma_{\text{opt}}\} \text{ a. } s.$

- Result independent of
  - Sparsity level
  - Initialization
  - Distribution of sparse vecs
  - Restricted isometry properties of $\boldsymbol{\Phi}$
  - Noise level

[G. Joseph, M., TSP 2017]

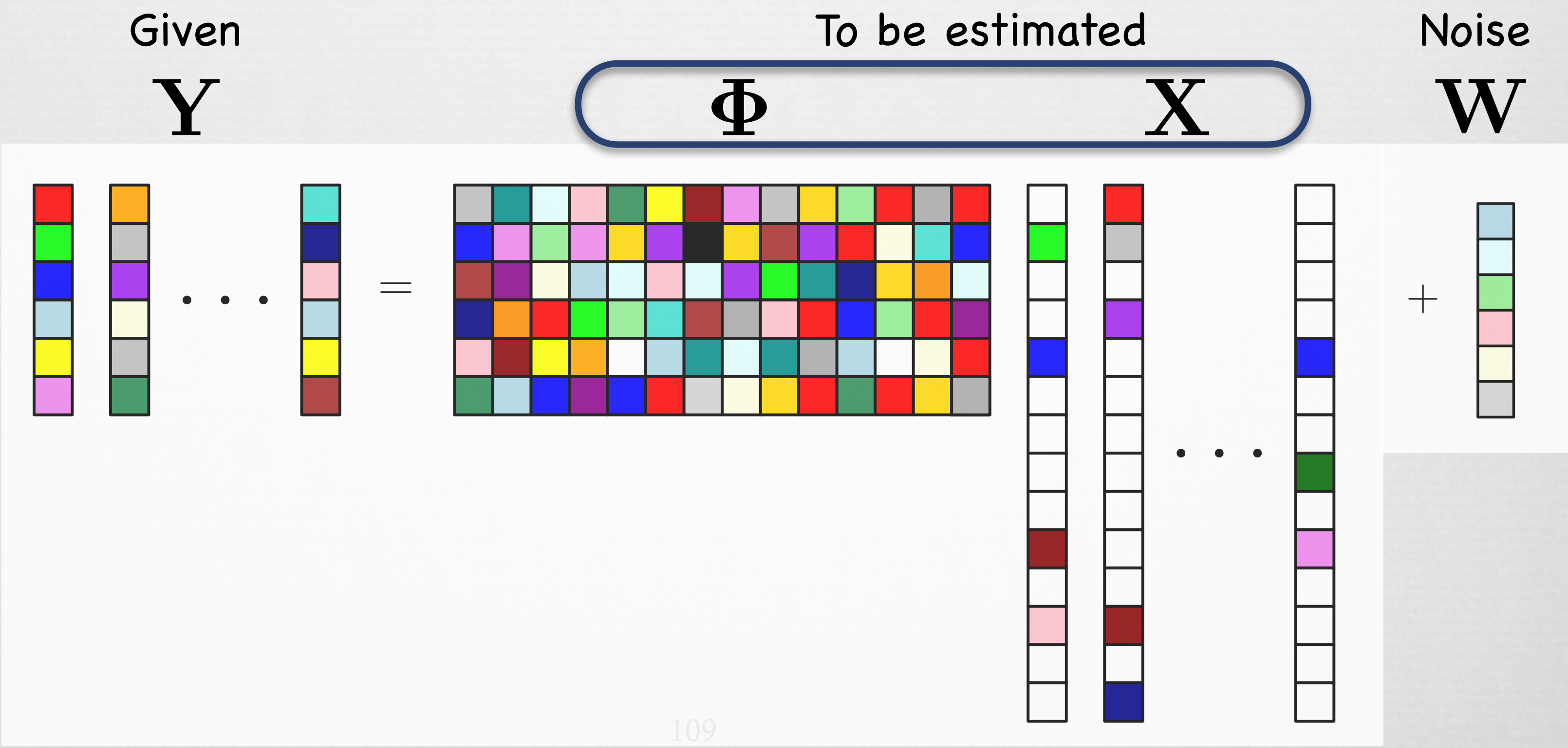
Performance



m = 20, N = 60, sparsity = 6, SNR = 20 dB

Dictionary Learning

- Matrix factorization problem:



SBL framework for DL

- Type-II ML: solve $\max_{\mathbf{\Lambda}=(\mathbf{\Phi},\mathbf{\Gamma})} -\log p\left(\mathbf{Y};\mathbf{\Lambda}\right)$
- EM procedure:
  - E-step: update statistics of X, as before
  - M-step: separable in variables $\mathbf{\Phi},\mathbf{\Gamma}$
    - Closed-form update for $\mathbf{\Gamma}$
    - Non-convex in $\mathbf{\Phi}$
    - Alternating minimization (AM): update one column of $\mathbf{\Phi}$ at a time

M-Step Analysis

- Cost function for updating $\mathbf{\Phi}$:

$$g(\mathbf{\Phi}) = -\mathrm{Tr}\left\{\mathbf{M}\mathbf{Y}^T\mathbf{\Phi}\right\} + \frac{1}{2}\mathrm{Tr}\left\{\mathbf{\Phi}(\mathbf{\Sigma} - \mathcal{D}\{\mathbf{\Sigma}\})\mathbf{\Phi}^T\right\}$$

- Proposition: The sequence of matrices generated by AM converges to a fixed point of $g(\mathbf{\Phi})$, and every fixed point is a Nash equilibrium, i.e.,

$$g\left(\Phi_1^*, \ldots, \Phi_{i-1}^*, \Phi_i^*, \Phi_{i+1}^*, \ldots, \Phi_N^*\right) \leq g\left(\Phi_1^*, \ldots, \Phi_{i-1}^*, \mathbf{a}, \Phi_{i+1}^*, \ldots, \Phi_N^*\right), \forall i \in [N]$$

for any unit norm $\mathbf{a}$

Image Denoising Example

(a) Original image
(b) Corrupted image, PSNR = 20 dB
(c) DL-SBL, PSNR = 28.96 dB, run time = 105.7 s
(d) SimCO, PSNR = 28.64 dB, run time = 58.7 s
(e) DL-MM, PSNR = 28.54 dB, run time = 98.7 s
(f) KSVD, PSNR = 28.34 dB, run time = 76.7 s
(g) SGK, PSNR = 27.44 dB, run time = 82.5 s
(h) PAU, PSNR = 27.44 dB, run time = 84.5 s
(i) MOD, PSNR = 27.42 dB, run time = 79.2 s

- 512 x 512 image “Barbara”
- Goal: remove AWGN
- Learn dictionary using 1000 8 x 8 blocks, randomly chosen
- N = 256
- Learn dictionary
- Reconstruct image using OMP

Sparsity and Linear Dynamical Systems

- System Model:

$$\mathbf{x}_{k+1} = \mathbf{D}\mathbf{x}_k + \mathbf{H}\mathbf{h}_k$$
$$\mathbf{y}_k = \mathbf{A}_k\mathbf{x}_k$$

- Want to observe, control and stabilize linear dynamical systems under sparsity constraints
  - Examples:
    - Known inputs: recover sparse state
    - Unknown sparse inputs: recover state and inputs

Applications

- Diffusion processes with sparse initialization
  - Disease/epidemic spreading
  - Pollution
  - Computer/mobile nwk virus spreading
  - Info. propagation in social networks
- Identifying the initial state of the system is critical to control the spreading
- Goal: Observability of the system when initialized with a sparse x_0

Compressed Sensing Formulation

$$\begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \dots \\ \boldsymbol{y}_{K-1} \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{A}_{(0)} \\ \boldsymbol{A}_{(1)}\boldsymbol{D} \\ \dots \\ \boldsymbol{A}_{(K-1)}\boldsymbol{D}^{K-1} \end{bmatrix}}_{\tilde{\mathbf{A}}_{(K)}} \boldsymbol{x}_0$$

- Challenges:
  - Non-identically distributed rows
  - Columns not independent
  - D can be arbitrary

Independent Subgaussian A_k

- RIP condition: The RIC δ_s of $\tilde{\mathbf{A}}_{(K)}$ satisfies $\delta_s < \delta$ for all $\delta > 1 - \lambda^{2(K-1)}$ with probability at least $1 - \epsilon$ if

$$Km\left(\delta - 1 + \lambda^{2(K-1)}\right)^2 \geq \tilde{c}\left[9\log\left(\frac{eN}{s}\right) + 2\log(2\epsilon^{-1})\right]$$

- $\lambda \leq 1$ is the ratio of the smallest to largest singular value of D
- System observable if $\delta_s <$ threshold
- Total number of meas. Km = O(s log N)!

Joint Recovery of Initial State and Sparse Inputs

$$\begin{bmatrix} \boldsymbol{y}_0 \\ \boldsymbol{y}_1 \\ \dots \\ \boldsymbol{y}_{K-1} \end{bmatrix} = \begin{bmatrix} \boldsymbol{A}_{(0)} & \boldsymbol{0} & \dots & \boldsymbol{0} \\ \boldsymbol{A}_{(1)}\boldsymbol{D} & \boldsymbol{A}_{(1)}\boldsymbol{H} & \dots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{A}_{(K-1)}\boldsymbol{D}^{K-1} & \boldsymbol{A}_{(K-1)}\boldsymbol{D}^{K-2}\boldsymbol{H} & \dots & \boldsymbol{A}_{(K-1)}\boldsymbol{H} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_0 \\ \boldsymbol{h}_1 \\ \vdots \\ \boldsymbol{h}_K \end{bmatrix}$$

- Same conditions as before, with

$$\tilde{\lambda} = 1 - \max_{i=0,1,\dots,K-1} \delta_i$$

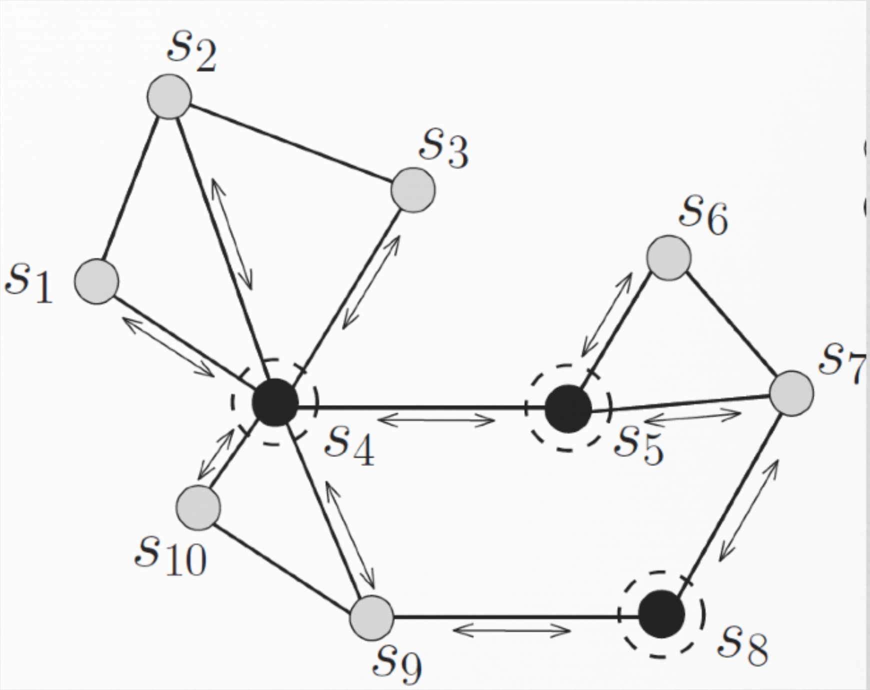
$$\delta_i = \text{RIC of } [\mathbf{D}^i \, \mathbf{D}^{i-1}\mathbf{H} \, \dots \, \mathbf{H}] \text{ of order } s$$

Key Observations

- Num. meas. sufficient for observability reduced from $O(N)$ to $O(s \log N)$
- Fewer measurements if
  - Transfer matrix D is well conditioned
  - Sparsity s is small
- Systems unobservable in the classical setting are observable under sparsity constraints!

Distributed Recovery: Learning Over a Network

- Network of L data centers
  - Node j has observation y_j
- Want to learn x_j:
  - Statistically related to y_j
- Centralized processing:
  - Optimal, but
  - Computationally demanding
- Distributed (in-network) processing:
  - Secure
  - Robust to node failures



Recap: SBL for Joint Sparse Recovery

- EM Iterations:
  - E-step:

$$\Sigma_j^{k+1} = \Gamma^k - \Gamma^k \Phi_j^T \left(\sigma_j^2 \mathbf{I}_M + \Phi_j \Gamma^k \Phi_j^T\right)^{-1} \Phi_j \Gamma^k$$

$$\mu_j^{k+1} = \sigma_j^{-2} \Sigma_j^{k+1} \Phi_j^T \mathbf{y}_j$$

    - Separable: x_j are independent given Γ
    - Can be computed locally at each node
  - M-step: not separable

$$\Gamma^{k+1} = \frac{1}{L} \sum_{j=1}^{L} \mathbf{a}_j^{(k+1)}$$

A Simple Trick

- Equivalent problems

$$\gamma^* = \frac{1}{L}\sum_{j=1}^{L} a_j \qquad \gamma^* = \arg\min_{\gamma} \sum_{j=1}^{L} |\gamma - a_j|^2$$

- For distributed implementation
  - Can now use, e.g., ADMM to solve

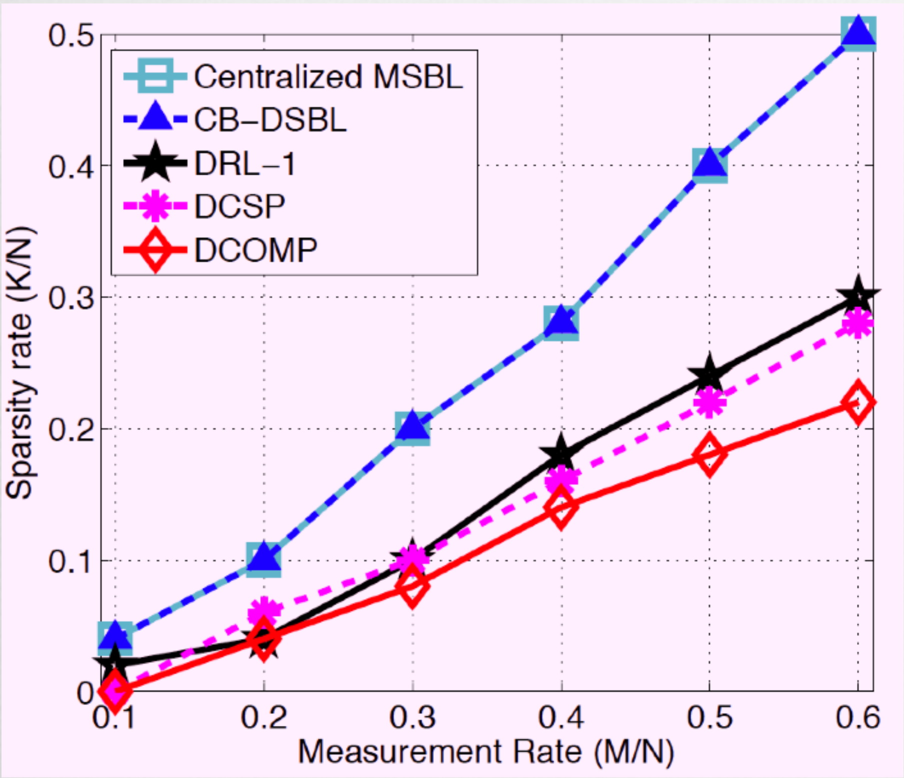
$$\underset{\gamma_j, j\in[L]}{\arg\min} \sum_{j=1}^{L} |\gamma_j - a_j|^2$$

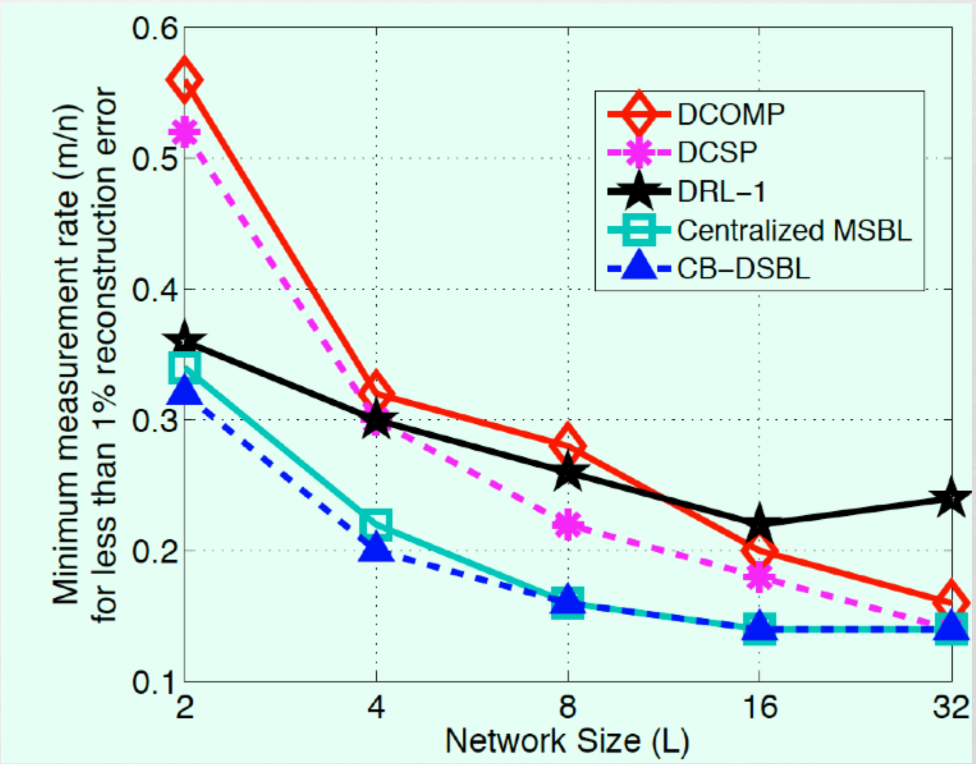
$$\text{subject to } \gamma_j = \gamma_b, b \in \mathcal{B}_j, j \in [L]$$

Can be computed locally at each node! Objective fn. separable

Bridge nodes Linear constraints

Simulation Result: NMSE Phase Transition





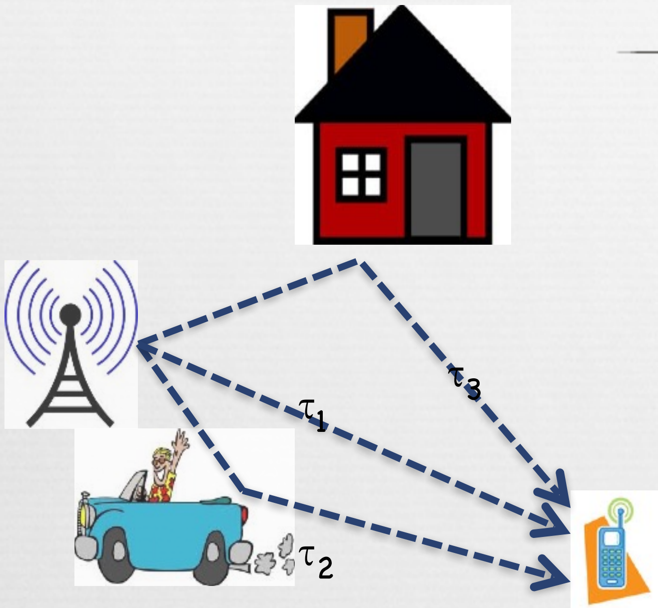
L = 5 nodes, n = 50, m = 10, 10% sparsity, SNR = 30 dB

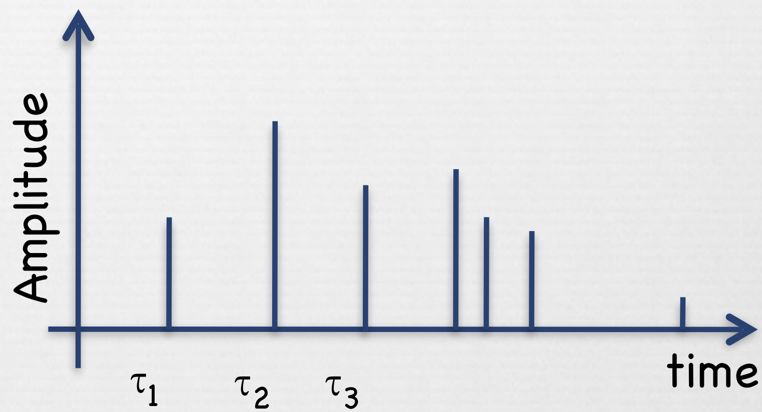
[S. Khanna, C. R. Murthy, TSIPN 2017]

Part 8: Applications

Wireless channel estimation & data detection

Wireless Channels





- Wireless channels exhibit multipath
  - Naturally sparse in the lag-domain
- Channel equalization & data detection
  - Need to estimate both support & channel

Channel Models

- Block fading channel:

  Channel constant for the duration of a block (say, K symbols), changes i.i.d. from block-to-block (classic SMV-SBL)

- Time-varying channel:

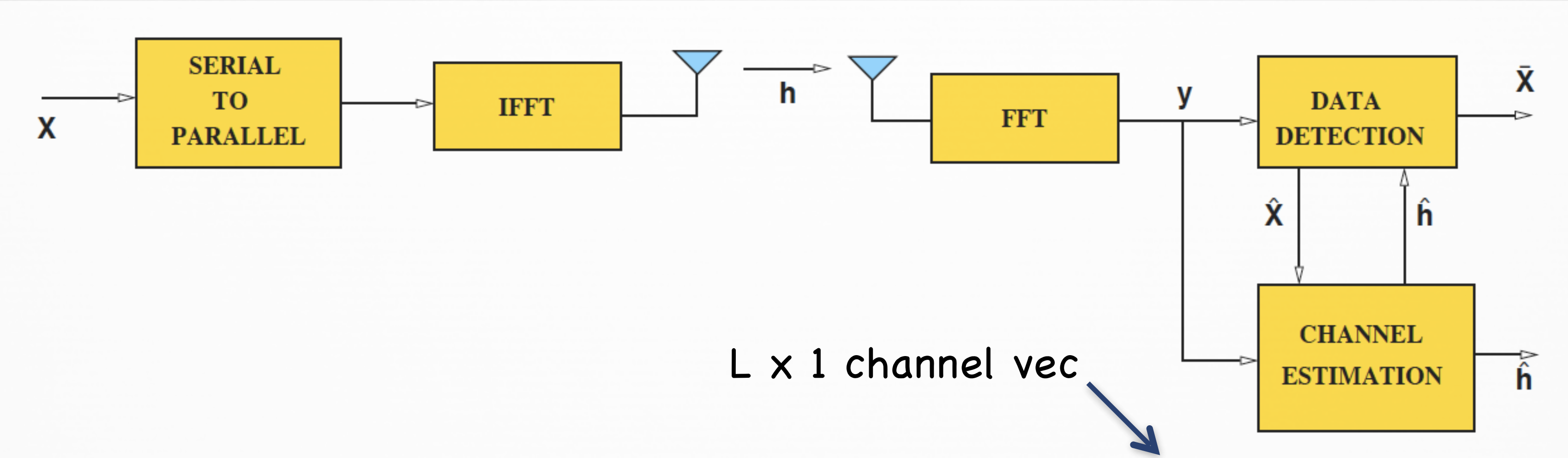
  Channel varies from symbol-to-symbol

  - Want to exploit temporal correlation and group-sparsity (MMV-SBL)

Outline

1. Block fading case:
    1. Known channel support: Joint channel estimation & data detection
    2. Unknown channel support: Channel and support estimation using pilot symbols
    3. Unknown data & support: Joint support, channel estimation & data detection
2. Time-varying case:
    1. AR model: Kalman-EM algo for joint support, channel estimation & data detn

OFDM with Block Fading Channel



L x 1 channel vec

- Received signal model $y = X F h + v$

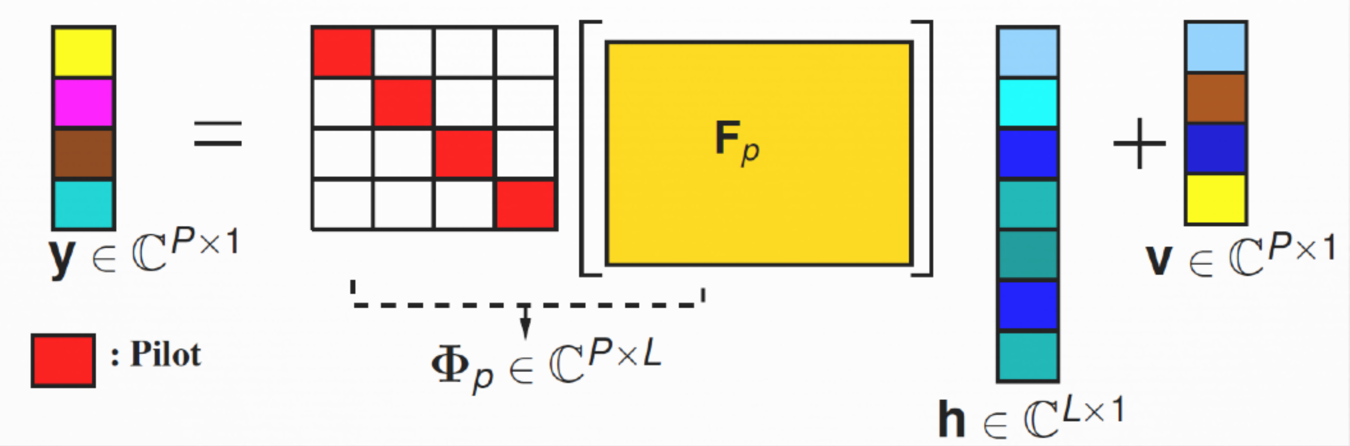
Diagonal data matrix; N x N
N: number of subcarriers

N x L DFT matrix, containing
first L cols of N x N DFT matrix
L: max channel delay spread

Noise

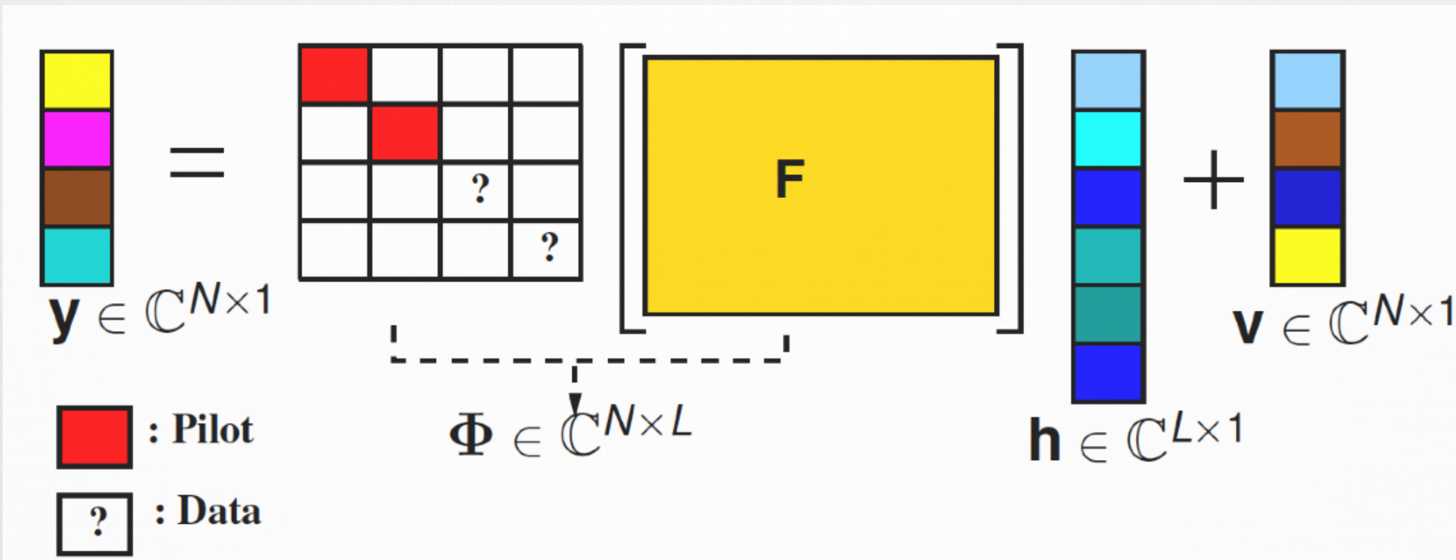
- Goal: Given y, jointly estimate X & h

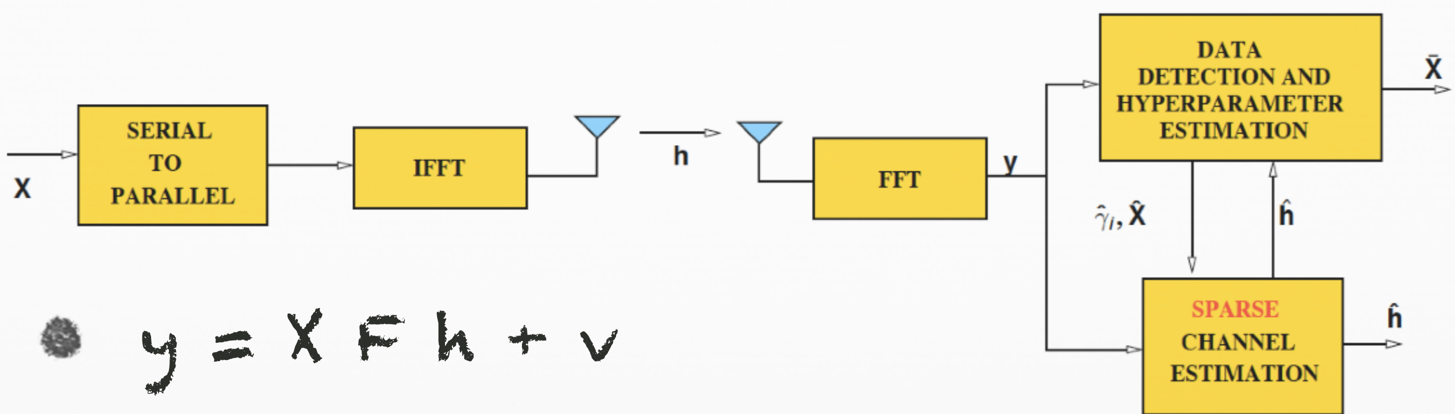
Sparse Channel Estimation from Pilot Symbols



- h sparse in time (lag) domain
- Hierarchical prior: $\mathbf{h}(i) = \mathcal{CN}(0, \gamma_i)$
  γ_i deterministic, unknown hyperparams
- Goal:
  Given y, X, estimate h & sparsity profile

Joint Channel, Support Estmn. & Data Detn.





- $y = X F h + v$

Joint Channel, Support Estmn. & Data Detn.

E–step: $E_{\mathbf{h}/\mathbf{y},\mathbf{X}^{(p)},\Gamma^{(p)}}[\log p(\mathbf{y},\mathbf{h};\mathbf{X},\Gamma)]$

M–step: $\arg\max_{\Gamma,\mathbf{X}}\{\text{E-step}\}$

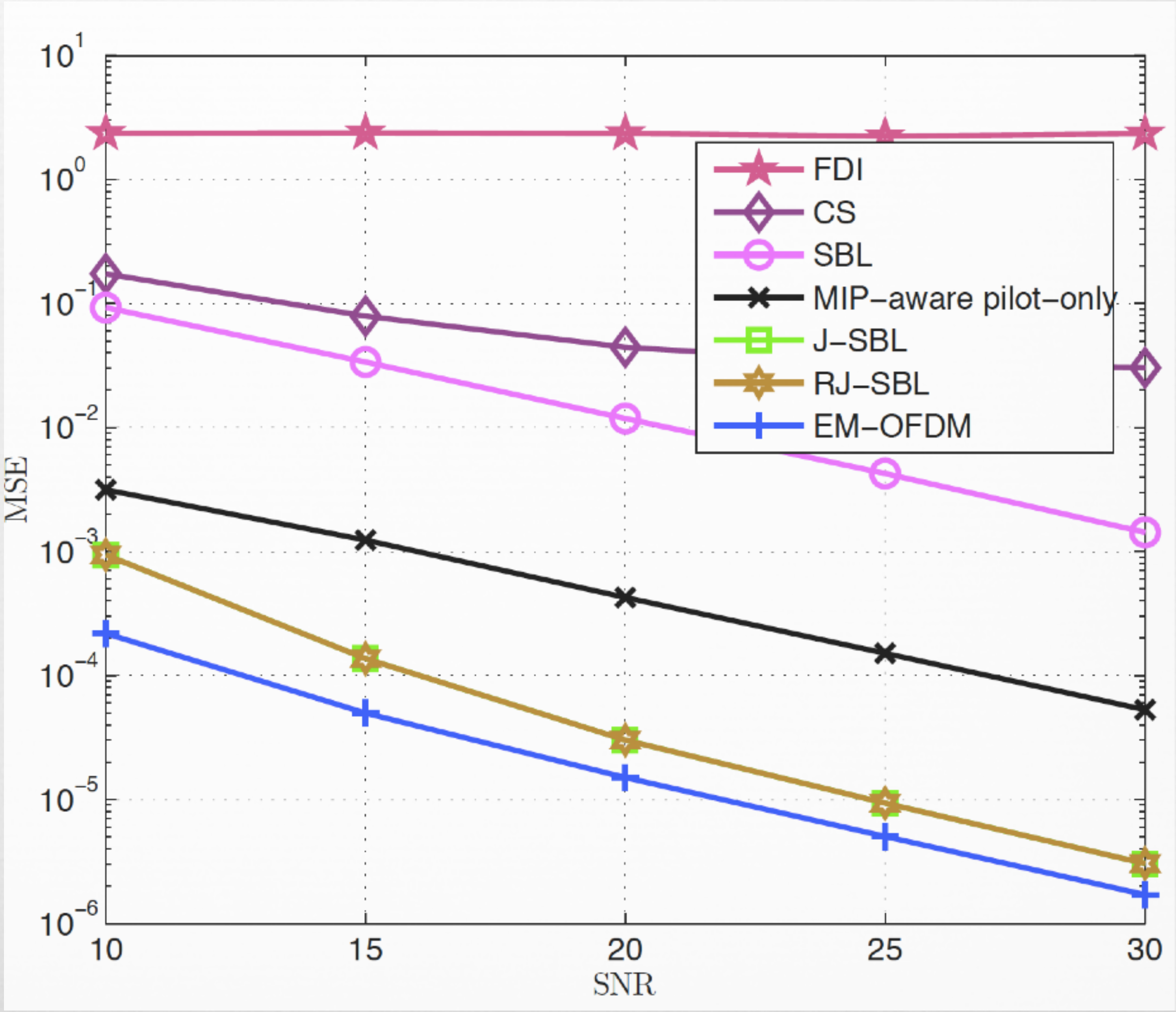
$\arg\max_{\Gamma} E_{\mathbf{h}/\mathbf{y},\mathbf{X}^{(p)},\Gamma^{(p)}}[\log p(\mathbf{h};\Gamma)]$ → Γ_{ML}

$\arg\max_{\mathbf{X}} E_{\mathbf{h}/\mathbf{y},\mathbf{X}^{(p)},\Gamma^{(p)}}[\log p(\mathbf{y}/\mathbf{h};\mathbf{X})]$ → $\mathbf{X}_{ML}$

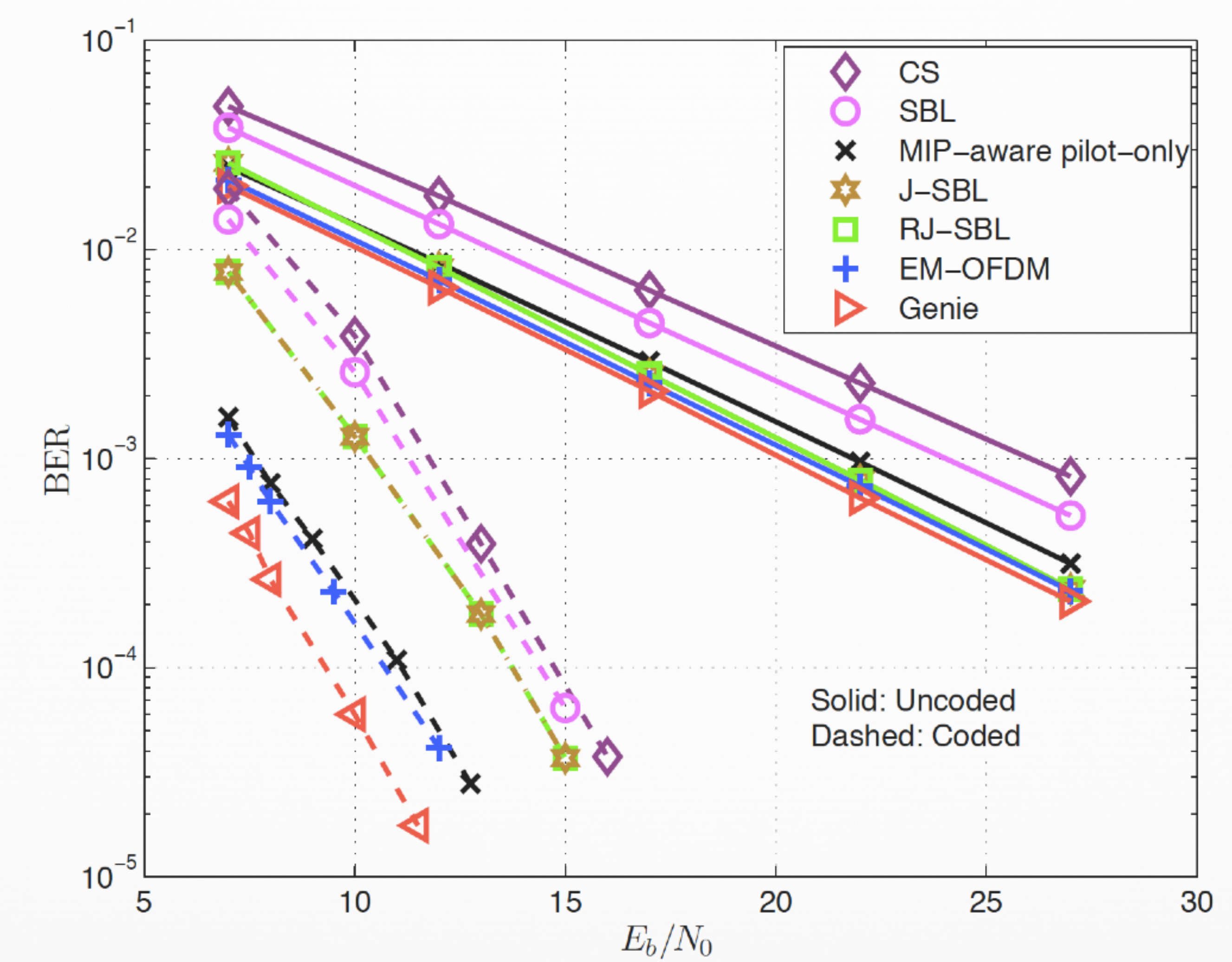
- Get h as a by-product of the E-step

Simulation Result

- OFDM system
- N=256 subcarriers,
- max delay spread L=64
- K=7 symbols/slot
- PedB PDP: 6 nonzero taps
- 44 pilot subcarriers
- Data: rate ½ turbo code, QPSK



BER Performance

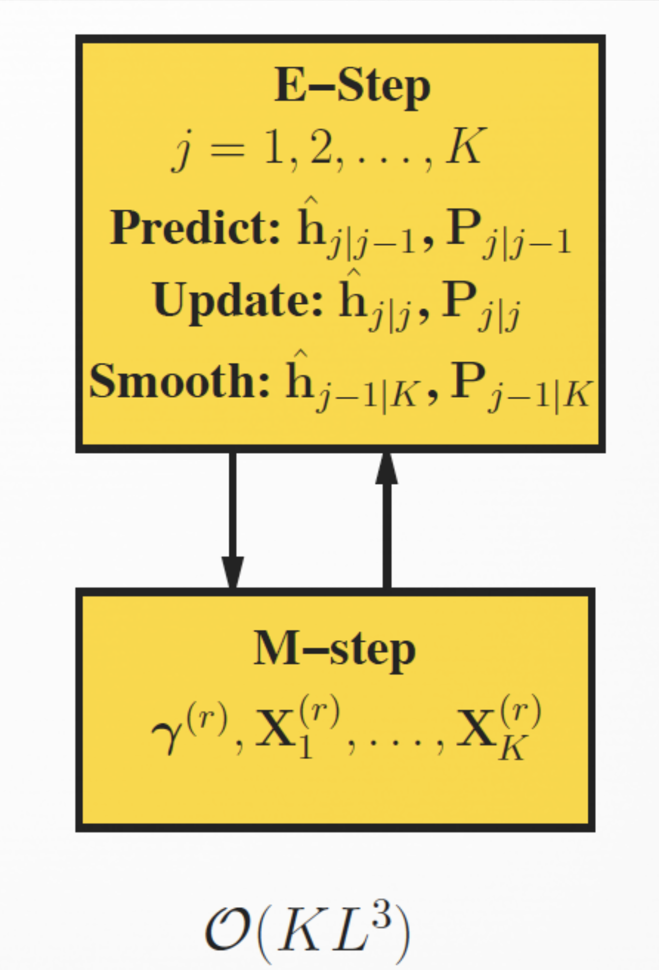


Time-Varying Channels

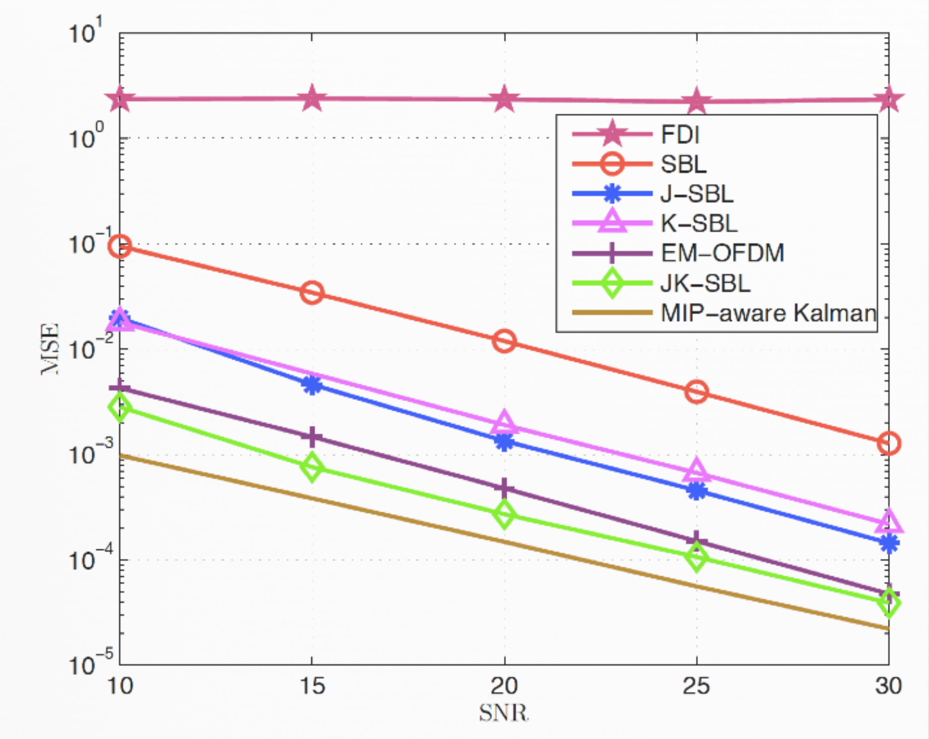
- Channel correlated from symbol-to-symbol
- AR model: $\mathbf{h}_k = \rho \mathbf{h}_{k-1} + \mathbf{u}_k$
- The factor ρ depends on the normalized doppler freq, which in turn depends on the speed of the mobile
- SBL framework can be extended to incorporate the temporal correlation

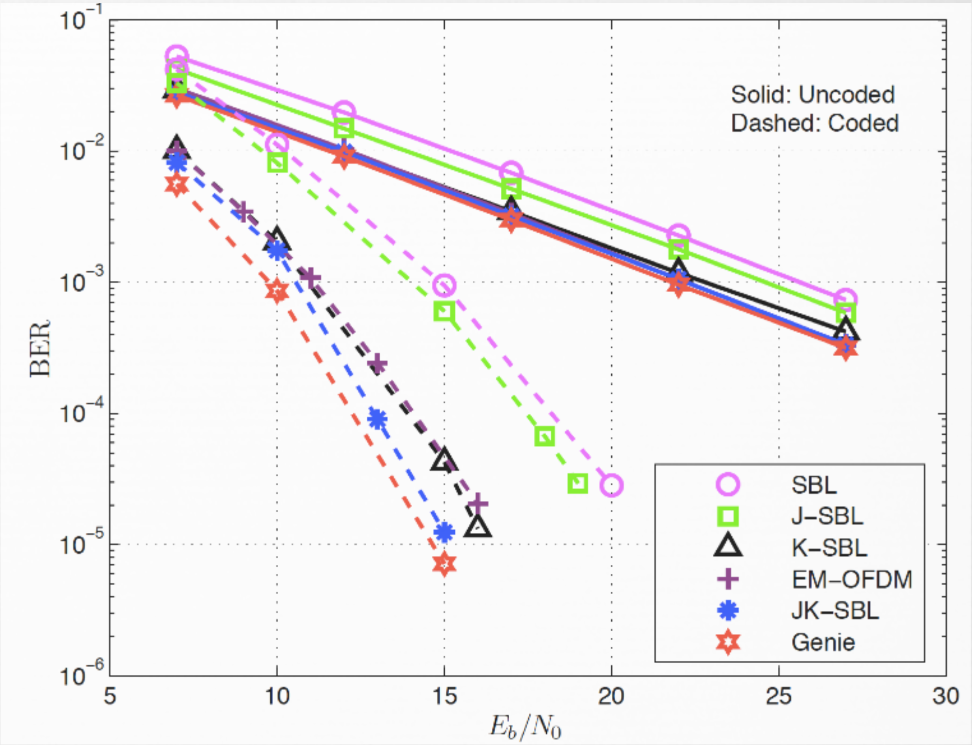
Joint Kalman SBL (JK-SBL)

- Complexity $O(KL^3)$: smaller than block-based methods $O(K^3L^3)$ [Zhang et al. 10]
  - (K = num. OFDM symbols used in joint estimation)
- In the block-fading case: get recursive, more computationally efficient versions of our algos



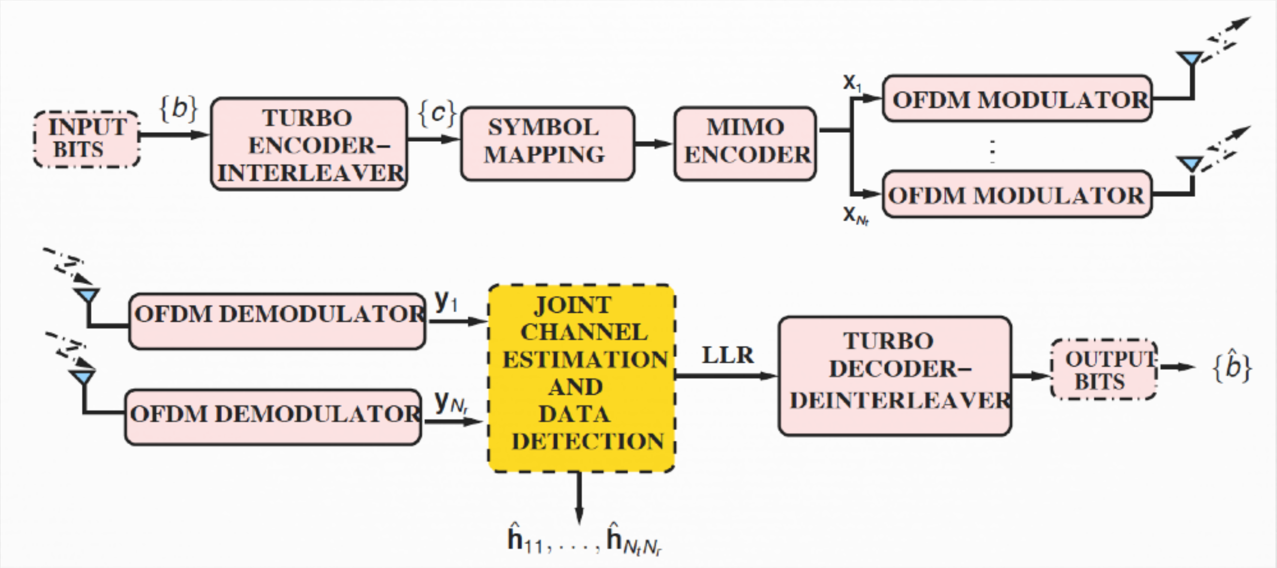
Simulation Result





- $f_d T_s = 0.001$ (slowly time-varying)

MIMO-OFDM



$$\mathbf{y}_{n_r} = \sum_{n_t=1}^{N_t} \mathbf{X}_{n_t} \mathbf{F} \mathbf{h}_{n_t n_r} + \mathbf{v}_{n_r},\ n_r = 1, \ldots, N_r$$

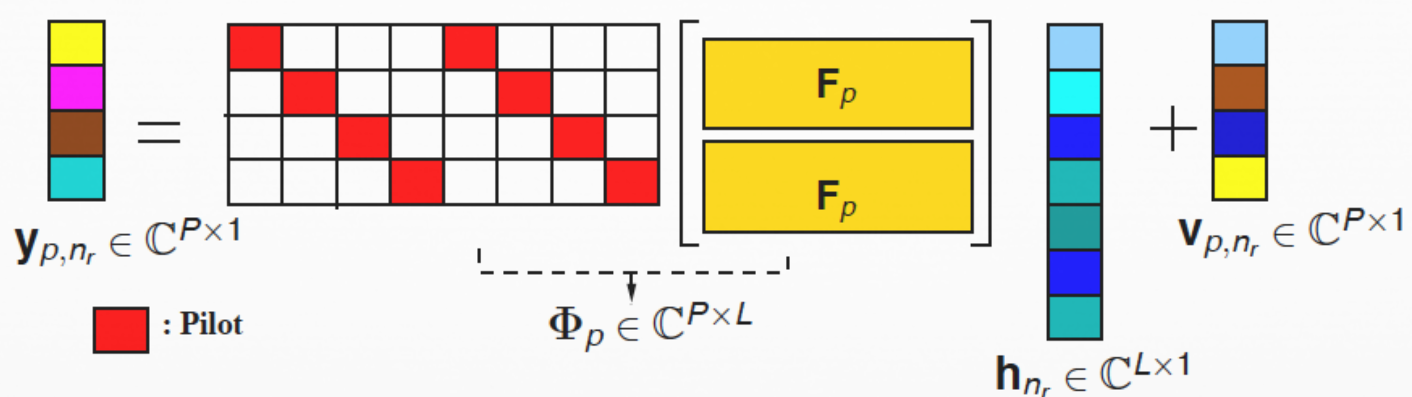
- Goal: Recover $h_1, \ldots, h_{Nr}$ from $y_1 \ldots y_{Nr}$
  - [Prasad & M., NCC 2014]

MMV Framework

- Measurement model

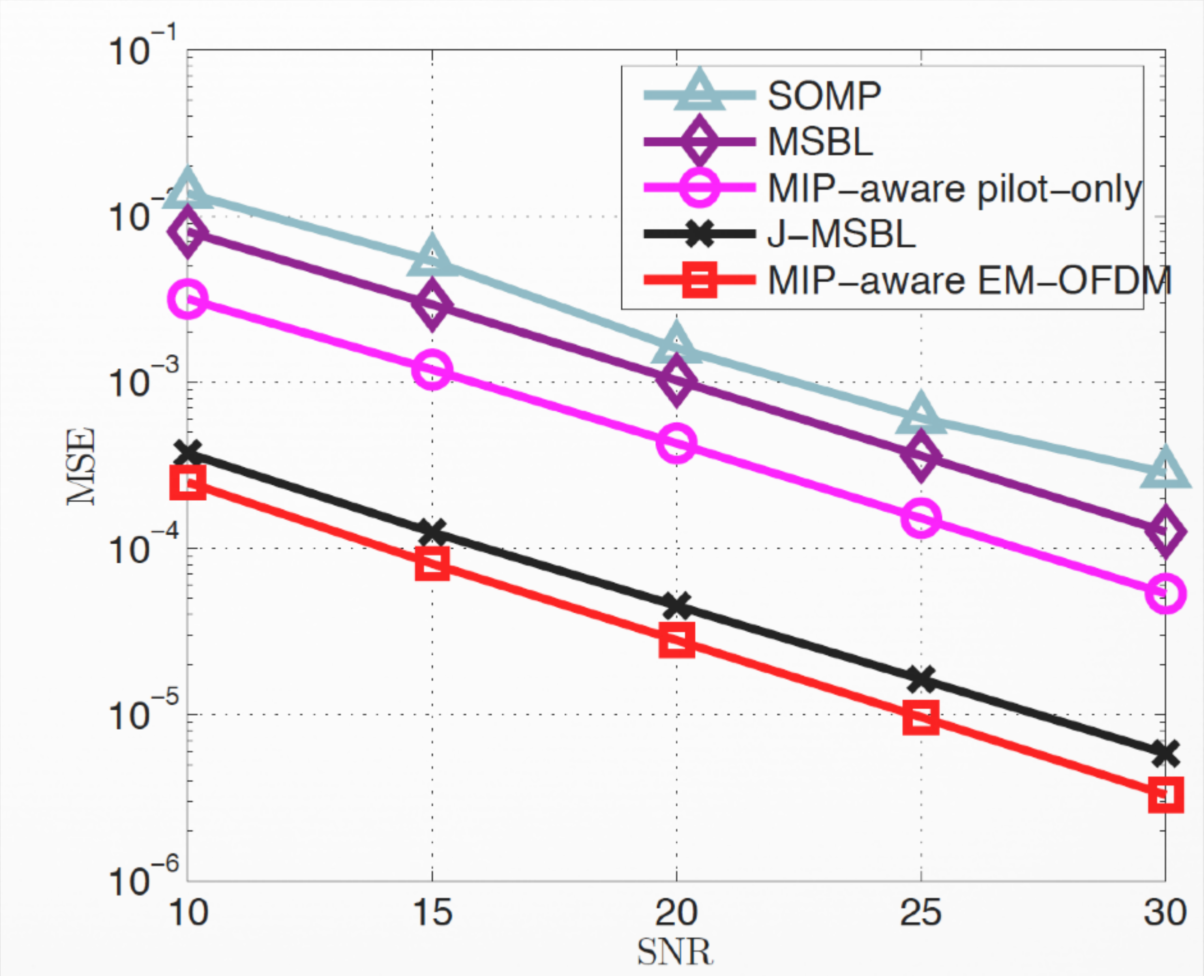
$$\underbrace{[\mathbf{y}_1,\ldots,\mathbf{y}_{N_r}]}_{\mathbf{Y}\in\mathbb{C}^{N\times N_r}} = \underbrace{\mathbf{X}(\mathbf{I}_{N_t}\otimes\mathbf{F})}_{\Phi\in\mathbb{C}^{N\times LN_t}} \underbrace{\begin{pmatrix} \mathbf{h}_{11} & \ldots & \mathbf{h}_{1N_r} \\ \vdots & \vdots & \\ \mathbf{h}_{N_t1} & \ldots & \mathbf{h}_{N_tN_r} \end{pmatrix}}_{\mathbf{H}\in\mathbb{C}^{LN_t\times N_r}} + \underbrace{[\mathbf{v}_1,\mathbf{v}_2,\ldots,\mathbf{v}_{N_r}]}_{\mathbf{V}\in\mathbb{C}^{N\times N_r}}$$

- Pilot subcarriers

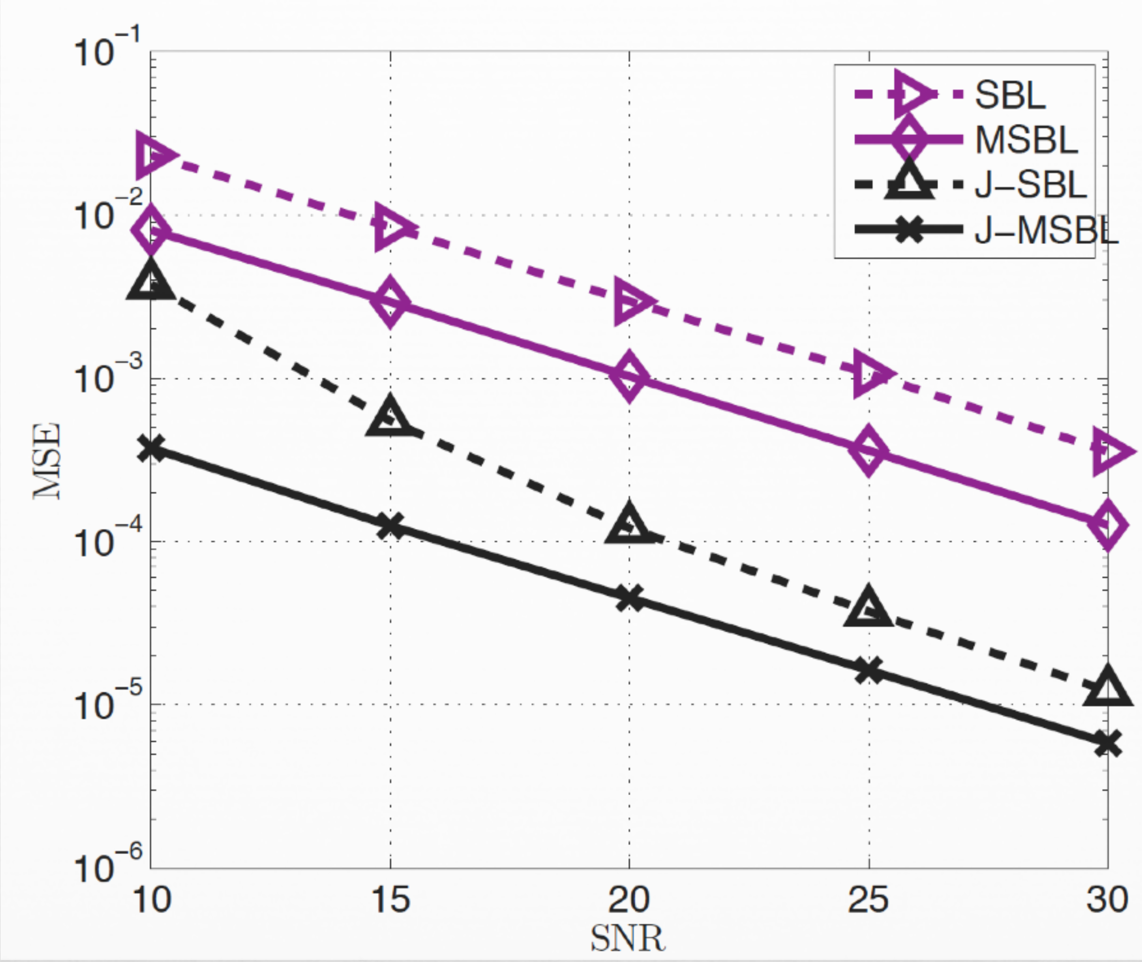


MSE Performance

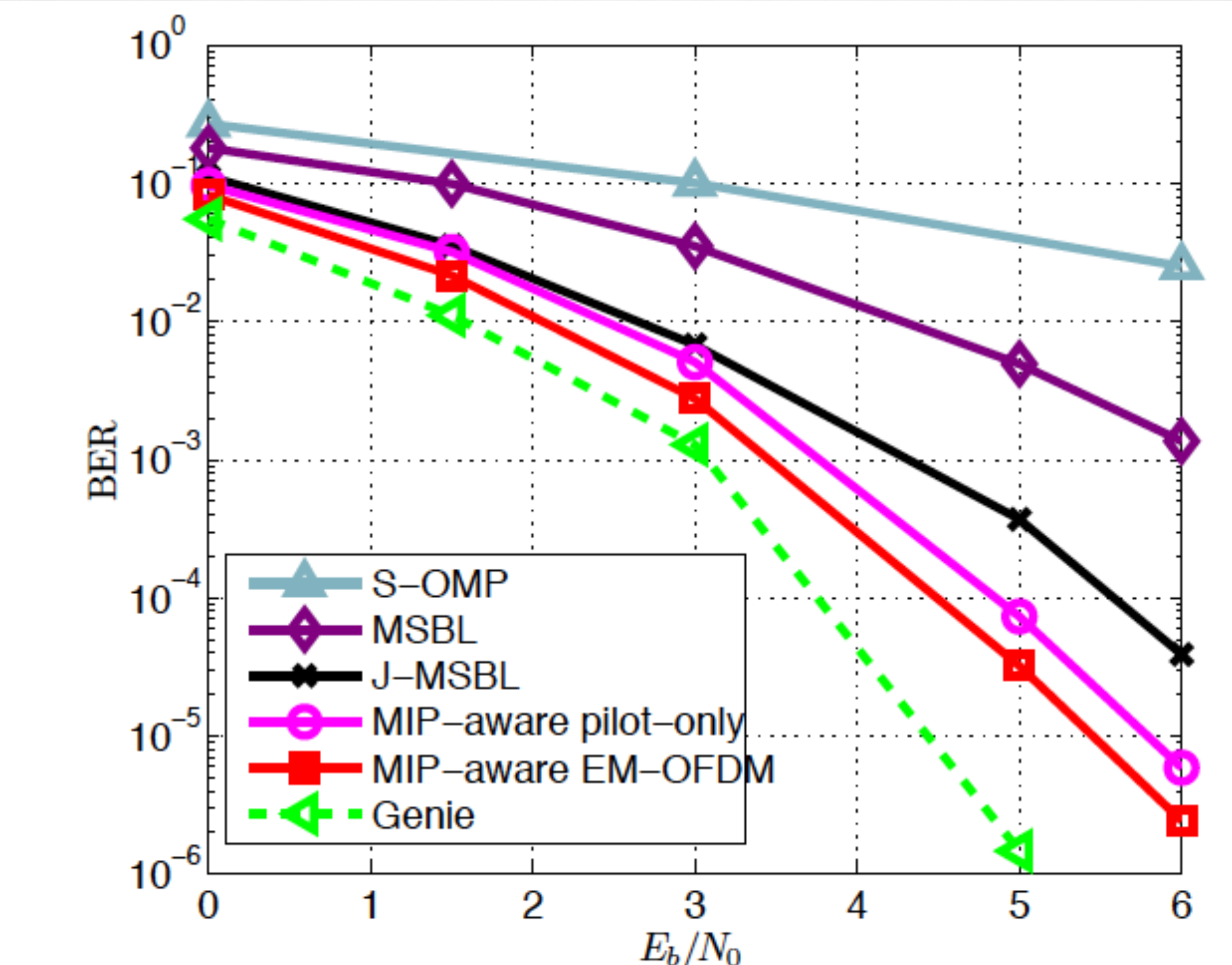
- 2 x 2 MIMO-OFDM System
- 256 subcarriers
- CP length 64
- 44 pilot subcarriers
- PedB PDP
- QPSK constellation



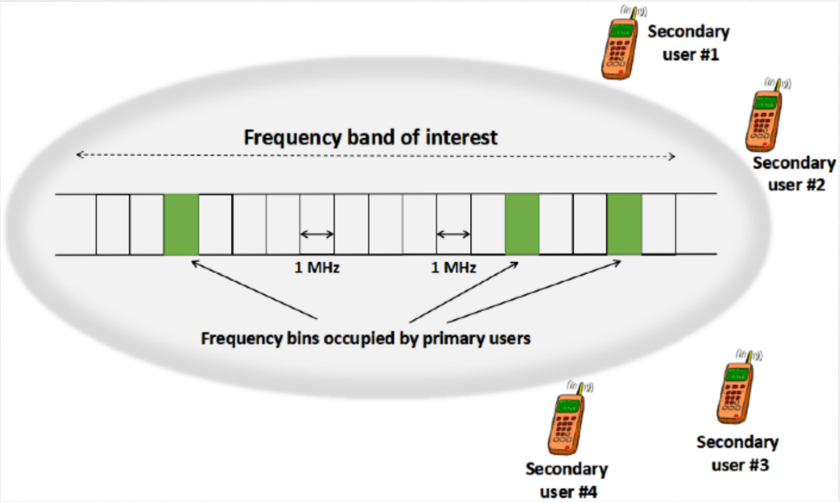
Exploiting Structure Helps!

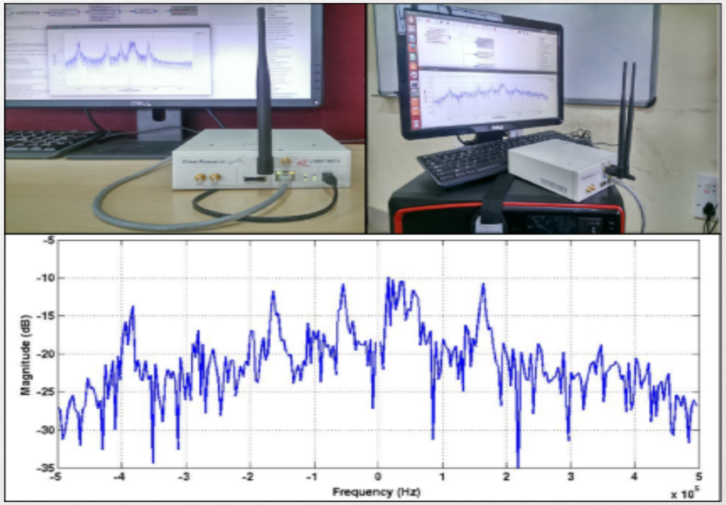


BER Performance



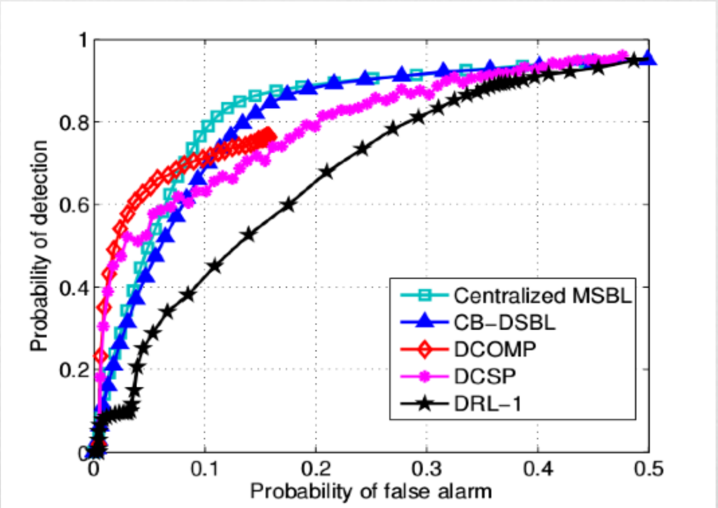
Wideband spectrum sensing using compressive measurements





Experimental setup

- No. primary users = 5
- No. secondary users = 10
- 11 of total 128 frequency subbands are in use
- SNR range: −2.4 to 7.8 dB.
- Compression ratio = 12.5



Summary

- Bayesian methods:
  - Simple updates
  - Promising performance
- Challenges:
  - Theoretical analysis
  - New algorithms
  - Novel applications
- Plenty of opportunities!

References

- J. M. Adler, B. Rao, and K. Kreutz-Delgado, **Comparison of basis selection methods**, Asilomar 1999
- S. F. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, **Sparse solutions to linear inverse problems with multiple measurement vectors**, IEEE Trans. Sig. Proc., 2005
- D. Wipf and B. Rao, **Sparse Bayesian learning for basis selection,** IEEE Trans. on Sig. Proc., 2004
- D. Wipf and B. Rao, **An empirical bayesian strategy for solving the simultaneous sparse approximation problem,** IEEE Trans. Sig. Proc., 2007
- Z. Zhang and B. Rao, **Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning,** IEEE J-STSP, 2011
- Z. Zhang and B. Rao, **Recovery of block sparse signals using the framework of block sparse bayesian learning**, ICASSP 2012
- R. Giri, B. Rao, **Type I and Type II Bayesian Methods for Sparse Signal Recovery using Scale Mixtures,** submitted, IEEE Trans. Sig. Proc., 2015

References

- R. Prasad and C. R. Murthy, **Cramér-Rao-Type Bounds for Sparse Bayesian Learning**, IEEE Transactions on Sig. Proc., vol. 61, no. 3, pp. 622-632, Mar. 2013
- R. Prasad, C. R. Murthy and B. Rao, **Joint Approximately Sparse Channel Estimation and Data Detection in OFDM Systems using Sparse Bayesian Learning**, IEEE Trans. Sig. Proc., Jul. 2014
- R. Prasad and C. R. Murthy, **Joint Approximately Group Sparse Channel Estimation and Data Detection in MIMO-OFDM Systems Using Sparse Bayesian Learning**, NCC 2014 [best paper award!]
- S. Khanna and C. R. Murthy, **Decentralized Bayesian Learning of Jointly Sparse Signals**, Globecom 2014
- V. Vinuthna, R. Prasad, and C. R. Murthy, **Sparse signal recovery in the presence of colored noise and rank-deficient noise covariance matrix: an SBL approach**, ICASSP 2015
- R. Prasad, C. R. Murthy, and B. D. Rao, **Joint Channel Estimation and Data Detection in MIMO-OFDM Systems: A Sparse Bayesian Learning Approach**, IEEE Trans. on Sig. Proc., Oct. 2015

Acknowledgements


Geethu Joseph


Saurabh Khanna


Ranjitha Prasad


Vinuthna Vinjamuri

Thank you!

Contact: cmurthy@iisc.ac.in