Sparse Reconstruction as Noncoherent Decoding

Phil Schniter

CTW – May 2009

Joint work with Lee Potter, Subhojit Som, and Justin Ziniel

The Sparse Reconstruction Problem:

From the M-length observation

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{e},$$

where

$$\begin{aligned} \boldsymbol{A} &\quad \text{is known and} \\ \boldsymbol{e} &\quad \text{is AWGN,} \end{aligned}$$

we desire to estimate the N-length signal $\boldsymbol{x}$, which is

1. *under-determined*: $\boldsymbol{x}$ has $N > M$ coefficients, and
2. *sparse*: $\boldsymbol{x}$ has $K < M$ non-zero coefficients (K unknown).

Sparse Reconstruction as Optimization in $\mathbb{R}^N$:

Many techniques treat sparse reconstruction as optimization over $\boldsymbol{x} \in \mathbb{R}^N$:

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{x}\|_1 \text{ s.t. } \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \le \epsilon \qquad \text{Basis Pursuit}$$

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 \text{ s.t. } \|\boldsymbol{x}\|_1 \le t \qquad \text{Lasso}$$

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + \sigma^2\tau\|\boldsymbol{x}\|_1 \qquad \text{GPSR}$$

$$= \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} p(\boldsymbol{x}|\boldsymbol{y}) \text{ s.t. } \begin{cases} p(\boldsymbol{x}) \propto e^{-\tau\|\boldsymbol{x}\|_1} \\ p(\boldsymbol{e}) \propto e^{-\|\boldsymbol{x}\|_2^2/\sigma^2} \end{cases} \qquad \text{Laplacian MAP}$$

$$\hat{\boldsymbol{x}} = \arg\min_{\boldsymbol{x}\in\mathbb{R}^N} p(\boldsymbol{x}|\boldsymbol{y}, \hat{\boldsymbol{\alpha}}_{\text{ML}}, \hat{\beta}_{\text{ML}}) \text{ s.t. } \begin{cases} \boldsymbol{x}|\boldsymbol{\alpha} \sim \text{indep } \mathcal{N}(0, \alpha_n^{-1}) \\ \boldsymbol{\alpha} \sim \text{iid } \Gamma(0,0) \\ \boldsymbol{e}|\beta \sim \mathcal{N}(\boldsymbol{0}, \beta^{-1}\boldsymbol{I}) \\ \beta \sim \Gamma(0,0) \end{cases} \qquad \text{RVM}$$

Sparse Reconstruction via Model Selection:

For true active-coefficient indices S_0, we can write

$$\boldsymbol{y} = \boldsymbol{A}_{S_0}\boldsymbol{x}_{S_0} + \boldsymbol{e}.$$

This motivates two-step sparse reconstruction procedures such as

1) $\hat{S}_{\mathsf{MAP}} = \arg\max_{S\in\mathbb{S}} p(S|\boldsymbol{y})$ "MAP model selection"
2) $\hat{\boldsymbol{x}}_{\mathsf{LS}|\hat{S}_{\mathsf{MAP}}} = (\boldsymbol{A}^T_{\hat{S}_{\mathsf{MAP}}}\boldsymbol{A}_{\hat{S}_{\mathsf{MAP}}})^{-1}\boldsymbol{A}^T_{\hat{S}_{\mathsf{MAP}}}\boldsymbol{y}$ "conditional LS estimation"

and

1) $\hat{\mathcal{S}}_\tau = \{S \in \mathbb{S} : p(S|\boldsymbol{y}) > \tau\}$ "soft model selection"
2) $\hat{\boldsymbol{x}}_{\mathsf{MMSE}} \approx \sum_{S\in\hat{\mathcal{S}}_\tau} p(S|\boldsymbol{y})\,\hat{\boldsymbol{x}}_{\mathsf{MMSE}|S}$ "MMSE estimation"

where $\mathbb{S}$ denotes the set of admissible models S. (known $K \Rightarrow$ restricted $\mathbb{S}$.)

We now show that the *model selection*
is closely related to *noncoherent decoding*. . .

Noncoherent Decoding:

Consider observations $\boldsymbol{y} \in \mathbb{R}^M$, channel $\boldsymbol{h} \in \mathbb{R}^K$, and codeword matrix $\boldsymbol{B}_i$:

$$\boldsymbol{y} \;=\; \boldsymbol{B}_i \boldsymbol{h} + \boldsymbol{e}, \quad i \in \{1, \dots, J\}.$$

In *noncoherent decoding*, we attempt to infer the codeword index i from $\boldsymbol{y}$ without knowing the channel state $\boldsymbol{h}$.

Sometimes we assume known channel statistics

$$\boldsymbol{h} \;\sim\; \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{R}) \text{ with } \begin{cases} \boldsymbol{\mu} = \boldsymbol{0} & \text{for Rayleigh fading} \\ \boldsymbol{\mu} \neq \boldsymbol{0} & \text{for Ricean fading} \end{cases}$$

and noise statistics $\boldsymbol{e} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

Examples of vectorized model $\boldsymbol{y} = \boldsymbol{B}_i \boldsymbol{h} + \boldsymbol{e}$:

1. MIMO flat-fading:

$$\boldsymbol{Y} = \boldsymbol{C}_i \boldsymbol{H} + \boldsymbol{E} \quad \text{for} \quad \boldsymbol{H} \in \mathbb{C}^{N_t \times N_r}, \boldsymbol{C}_i \in \mathbb{C}^{L \times N_t}, \boldsymbol{Y} \in \mathbb{C}^{L \times N_r}$$

$$\Rightarrow \quad \text{vec}(\boldsymbol{Y}) \;=\; (\boldsymbol{I}_{N_r} \otimes \boldsymbol{C}_{\boldsymbol{i}}) \,\text{vec}(\boldsymbol{H}) + \text{vec}(\boldsymbol{E})$$

2. SISO with ISI:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} b_0^{(i)} & & \\ b_1^{(i)} & b_0^{(i)} & \\ b_2^{(i)} & b_1^{(i)} & b_0^{(i)} \\ & b_2^{(i)} & b_1^{(i)} \\ & & b_2^{(i)} \end{bmatrix} \begin{bmatrix} h_0 \\ h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

3. SISO with TV-ISI:

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \\ y_4 \end{bmatrix} = \begin{bmatrix} b_0^{(i)} & & & & & & & & \\ & b_0^{(i)} & & b_1^{(i)} & & & & & \\ & & b_0^{(i)} & & b_1^{(i)} & & b_2^{(i)} & & \\ & & & & & b_1^{(i)} & & b_2^{(i)} & \\ & & & & & & & & b_2^{(i)} \end{bmatrix} \begin{bmatrix} h_{0,0} \\ h_{1,0} \\ h_{2,0} \\ h_{0,1} \\ h_{1,1} \\ h_{2,1} \\ h_{0,2} \\ h_{1,2} \\ h_{2,2} \end{bmatrix} + \begin{bmatrix} e_0 \\ e_1 \\ e_2 \\ e_3 \\ e_4 \end{bmatrix}$$

Noncoherent Decoding as Model Selection:

Notice that we can rewrite

$$\boldsymbol{y} \;=\; \boldsymbol{B}_i \boldsymbol{h} + \boldsymbol{e}, \quad i \in \{1, \ldots, J\}$$

as the familiar sparse reconstruction problem:

$$\boldsymbol{y} \;=\; \underbrace{\begin{bmatrix} \boldsymbol{B}_1 \cdots \boldsymbol{B}_i \cdots \boldsymbol{B}_J \end{bmatrix}}_{\boldsymbol{A}} \underbrace{\begin{bmatrix} \mathbf{0} \\ \vdots \\ \boldsymbol{h} \\ \vdots \\ \mathbf{0} \end{bmatrix}}_{\boldsymbol{x}} + \boldsymbol{e}$$

for K-sparse $\boldsymbol{x} \in \mathbb{R}^{JK}$. Thus

noncoherent decoding $\Leftrightarrow$ *model selection*
under $\mathbb{S} = \big\{(1, \ldots, K), (K{+}1, \ldots, 2K), \cdots, (KJ{-}K{+}1, \ldots, KJ)\big\}$.

Noncoherent Decoding – Typical Approaches:

Known channel/noise statistics, non-equal codeword priors:

$$\begin{aligned}\hat{i}_{\mathsf{MAP}} &= \arg\max_i p(i|\boldsymbol{y}) \\ &= \arg\max_i \big\{ \ln p(\boldsymbol{y}|i) + \ln p(i) \big\} \qquad \text{where } p(\boldsymbol{y}|i) = \textstyle\int p(\boldsymbol{y}|i,\boldsymbol{h})\, p(\boldsymbol{h})\, d\boldsymbol{h} \\ \hat{\mathcal{I}}_\tau &= \{ i : \ln p(\boldsymbol{y}|i) + \ln p(i) > \ln \tau \} \qquad \text{...soft decoding}\end{aligned}$$

Known channel/noise statistics, equal codeword priors:

$$\hat{i}_{\mathsf{ML}} = \arg\max_i p(\boldsymbol{y}|i)$$

Unknown channel/noise statistics:

$$\begin{aligned}\hat{i}_{\mathsf{GLRT}} &= \arg\max_i p(\boldsymbol{y}|i, \hat{\boldsymbol{h}}_{\mathsf{ML}|i}) \quad \text{where } \hat{\boldsymbol{h}}_{\mathsf{ML}|i} = \boldsymbol{B}_i^+ \boldsymbol{y} \\ &= \arg\min_i \boldsymbol{y}^T \boldsymbol{\Pi}^\perp_{\boldsymbol{B}_i} \boldsymbol{y}.\end{aligned}$$

Model Selection – Typical Approaches:

Known signal/noise statistics, non-equal model priors:

$$\hat{S}_{\mathsf{MAP}} = \arg\max_{S\in\mathbb{S}} p(S|\boldsymbol{y})$$

$$= \arg\max_{S\in\mathbb{S}} \big\{ \ln p(\boldsymbol{y}|S) + \ln p(S) \big\} \text{ where } p(\boldsymbol{y}|S) = \textstyle\int p(\boldsymbol{y}|S,\boldsymbol{x}_S)\, p(\boldsymbol{x}_S)\, d\boldsymbol{x}_S$$

$$\hat{\mathcal{S}}_{\tau} = \{S : \ln p(\boldsymbol{y}|S) + \ln p(S) > \ln\tau\}$$ ...Bayesian model averaging

Known signal/noise statistics, equal model priors:

$$\hat{S}_{\mathsf{ML}} = \arg\max_{S\in\mathbb{S}} p(\boldsymbol{y}|S)$$

Unknown signal/noise statistics:

$$\hat{S}_{\mathsf{GLRT}} = \arg\max_{S\in\mathbb{S}} p(\boldsymbol{y}|S,\hat{\boldsymbol{x}}_{\mathsf{LS}|S}) \quad \text{where } \hat{\boldsymbol{x}}_{\mathsf{LS}|S} = \boldsymbol{A}_S^{+}\boldsymbol{y}$$

$$= \arg\min_{S\in\mathbb{S}} \boldsymbol{y}^T \boldsymbol{\Pi}_{\boldsymbol{A}_S}^{\perp} \boldsymbol{y}$$ ...fails for nested $\mathbb{S}$!

Leveraging the Connection – PWEP Analysis:

- Pair-wise error probability (PWEP) of noncoherent decoding, e.g.,

$$P_{j|i} \;=\; \Pr\left\{\, p(\boldsymbol{y}|j) > p(\boldsymbol{y}|i) \mid i \right\} \quad \text{for ML}$$

  has been thoroughly studied.

- The results apply directly to model selection under the constraint

$$\mathbb{S} = \big\{(1,\dots,K), (K+1,\dots,2K), \cdots, (KJ-K+1,\dots,KJ)\big\}.$$

  Note: Since this $\mathbb{S}$ is non-nested, can use GLRT.

- PWEP results can be extended to cover the case of "unrestricted" $\mathbb{S}$, where $|\mathbb{S}| = 2^N$.

Model Selection via "Generalized Information Criteria":

For general $\mathbb{S}$, model selection often takes the form

$$\hat{S} \;=\; \arg\min_{S\in\mathbb{S}} \Big\{ \tfrac{1}{\sigma^2} \big\| \boldsymbol{y} - \boldsymbol{A}_S \hat{\boldsymbol{x}}_{\mathsf{LS}|S} \big\|_2^2 + \eta |S| \Big\}.$$

This includes "information theoretic" model-order selection criteria, e.g.,

$\eta_{\mathsf{AIC}} = 2$	Akiake's information criterion
$\eta_{\mathsf{BIC}} = \ln M$	Bayesian information criterion
$\eta_{\mathsf{RIC}} = 2 \ln N$	Risk inflation criterion

as well as MAP model selection under the Zellner/iid-Bernoulli model:

$$\eta_{\mathsf{MAP}} = \tfrac{\gamma+1}{\gamma} \ln \big((1{+}\gamma)(\tfrac{1-\lambda}{\lambda})^2 \big) \qquad \text{for} \begin{cases} \text{unrestricted } \mathbb{S} \text{ (i.e., } |\mathbb{S}| = 2^N) \\ p(S) = \lambda^{|S|}(1-\lambda)^{(N-|S|)} \\ \boldsymbol{x}_S \sim \mathcal{N}\big(\boldsymbol{0}, \gamma\sigma^2 (\boldsymbol{A}_S^T \boldsymbol{A}_S)^{-1}\big). \end{cases}$$

PWEP of Model Selection:

Lemma 1 *For generic* $\mathbb{S}$*, the PWEP of*

$$\hat{S} \;=\; \arg\min_{S\in\mathbb{S}} \left\{ \tfrac{1}{\sigma^2}\big\|\boldsymbol{y} - \boldsymbol{A}_S \hat{\boldsymbol{x}}_{\mathsf{LS}|S}\big\|_2^2 + \eta|S| \right\} \quad \textit{under} \quad \boldsymbol{x}_S|S \sim \mathcal{N}(\boldsymbol{0}, \gamma\sigma^2 \boldsymbol{I}_{|S|})$$

has the upper bound (tight as $\gamma \to \infty$*):*

$$P_{\hat{S}|S} \;\le\; (\alpha_{\hat{S},S}\gamma)^{-K_\mathsf{m}} C_{K_\mathsf{m},K_\mathsf{f}}(\eta),$$

where K_m *and* K_f *denote the # of missed and false-alarm coefficients, and*

$$C_{K_\mathsf{m},K_\mathsf{f}}(\eta) \;=\; \begin{cases} e^{(K_\mathsf{m}-K_\mathsf{f})\eta} \displaystyle\sum_{k=0}^{K_\mathsf{f}-1} \frac{(K_\mathsf{f}-K_\mathsf{m})^k \eta^k}{k!} \binom{K_\mathsf{m}+K_\mathsf{f}-1-k}{K_\mathsf{m}} & K_\mathsf{m} \le K_\mathsf{f}, \\ \displaystyle\sum_{k=0}^{K_\mathsf{m}} \frac{(K_\mathsf{m}-K_\mathsf{f})^k \eta^k}{k!} \binom{K_\mathsf{m}+K_\mathsf{f}-1-k}{K_\mathsf{f}-1} & K_\mathsf{m} > K_\mathsf{f}. \end{cases}$$

$$\alpha_{\hat{S},S} \;=\; \lambda_{\min}(\boldsymbol{A}_\mathsf{m}^T \boldsymbol{\Pi}_{\boldsymbol{A}_{\hat{S}}}^{\perp} \boldsymbol{A}_\mathsf{m}) \qquad \ldots \textit{Restricted Isometry Property}$$

(An extension of Brehler & Varanasi TIT 2001.)

Performance Guarantees for MAP Model Selection:

Assuming that $\boldsymbol{A}$ has unit-norm columns and satisfies a Restricted Isometry Property (RIP), we've recently shown that the following properties hold *with high probability* for reasonably small constants K_1, K_2, K_3, K_4:

1. The energy of the missed signal coefficients is upper bounded by $K_1 M \sigma_e^2$.
2. No active coefficients are missed when $|\mu| > 4\sigma_1 + K_2\sqrt{M}\sigma_e^2$.
3. No coefficients are falsely detected when $|\mu| > K_3\sqrt{M}\sigma_1 + K_4\sqrt{M}\sigma_e^2$.

Leveraging the Connection – A Sparse-Reconstruction Algorithm:

Optimal model selection under known statistics and non-equal priors is

$$\hat{S}_{\text{MAP}} = \arg\max_{S\in\mathbb{S}} p(S|\boldsymbol{y}) = \arg\min_{S\in\mathbb{S}} \big\{ -\ln p(\boldsymbol{y}|S) - \ln p(S) \big\}$$

where, for $\boldsymbol{x}_S \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{R})$,

$$-\ln p(\boldsymbol{y}|S) = \tfrac{1}{\sigma^2}\big\|\boldsymbol{y} - \boldsymbol{A}_S\hat{\boldsymbol{x}}_{\text{MMSE}|S}\big\|_2^2 + \big\|\hat{\boldsymbol{x}}_{\text{MMSE}|S} - \boldsymbol{\mu}\big\|_{\boldsymbol{R}^{-1}}^2 + \ln\big|\boldsymbol{A}_S\boldsymbol{R}\boldsymbol{A}_S^T + \sigma^2\boldsymbol{I}\big| + C$$

As in *soft noncoherent decoding*, can use *tree search* to find the set of models $\hat{\mathcal{S}}$ with significant posterior probability. The "per-survivor" nuisance parameter estimates $\{\hat{\boldsymbol{x}}_{\text{MMSE}|S}\}_{S\in\hat{\mathcal{S}}}$ can then be combined for MMSE estimation:

$$\hat{\boldsymbol{x}}_{\text{MMSE}} \approx \sum_{S\in\hat{\mathcal{S}}} p(S|\boldsymbol{y})\,\hat{\boldsymbol{x}}_{\text{MMSE}|S}$$

...Bayesian model averaging.

Using $\mathcal{O}(MNK)$-complexity tree search, "Fast Bayesian Matching Pursuit" yields

near-optimal performance with OMP-like complexity.

Numerical Experiments — "Compressible" Signal:

Setup:

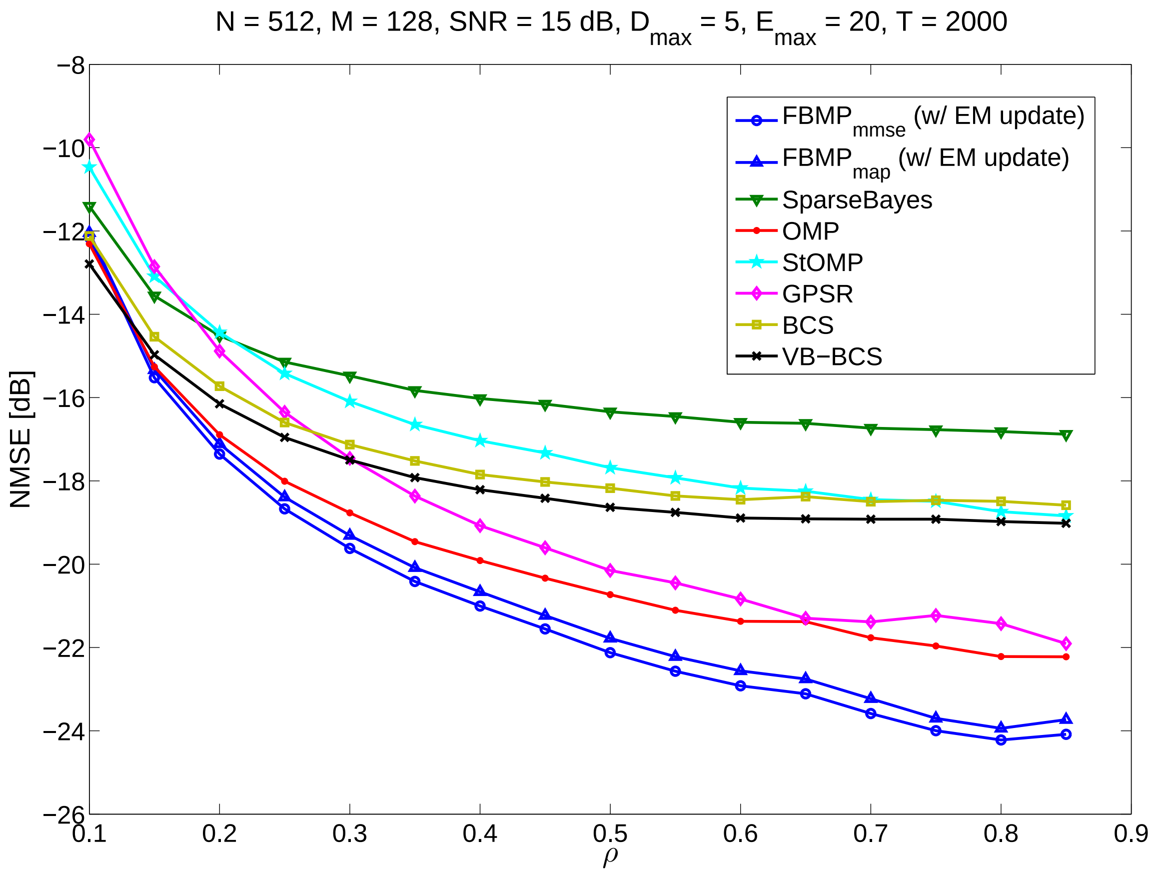
$$
\begin{aligned}
N &= 512 \\
M &= 128 \\
\boldsymbol{A} &: \text{ i.i.d. } \mathcal{N}(0,1) \quad \text{with columns scaled to unit norm} \\
\boldsymbol{x} &: \text{ shuffled } x_n = e^{-\rho n} \quad \text{with sparsity } \rho \in (0,1) \\
\text{SNR} &= 15\text{dB}
\end{aligned}
$$

Algorithms:

OMP	Tropp & Gilbert
StOMP	Donoho, Tsaig, Drori & Starck
GPSR-Basic	Figueiredo, Nowak & Wright ($\min_{\boldsymbol{x}} \|\boldsymbol{y}-\boldsymbol{A}\boldsymbol{x}\|_2^2 + \tau\|\boldsymbol{x}\|_1$)
SparseBayes	Wipf & Rao (RVM)
BCS	Ji & Carin (RVM)
VB-BCS	Ji & Carin (RVM)
FBMP	Schniter, Potter & Ziniel (BMA)

Performance: NMSE $\triangleq$ Avg $\left\{ \dfrac{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2}{\|\boldsymbol{x}\|_2^2} \right\}$ over 2500 random trials.

NMSE versus decay rate ρ:



FBMP outperformed GPSR and OMP by 2 dB and others by much more.

Note: The signal priors favor GPSR!

The Relevance Vector Machine (RVM):

The RVM is an alternate Bayesian approach to sparse reconstruction:

- For coefficient activity, RVM uses continuous "precisions" $\boldsymbol{\alpha} \in (\mathbb{R}^+)^N$:

$$\boldsymbol{x}|\boldsymbol{\alpha} \sim \text{independent } \mathcal{N}(0, \alpha_n^{-1}) \quad \text{and} \quad \boldsymbol{\alpha} \sim \text{iid } \Gamma(0,0)$$
$$\boldsymbol{e}|\beta \sim \mathcal{N}(\boldsymbol{0}, \beta^{-1}\boldsymbol{I}) \quad \text{and} \quad \beta \sim \Gamma(0,0)$$

- The RVM's gamma hyperpriors lead to the convenient posterior

$$p(\boldsymbol{x}|\boldsymbol{y},\boldsymbol{\alpha},\beta) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}) \quad \text{for} \quad \begin{cases} \bar{\boldsymbol{\mu}} = \beta\boldsymbol{\Sigma}\boldsymbol{A}^T\boldsymbol{y} \\ \bar{\boldsymbol{\Sigma}} = \big(\beta\boldsymbol{A}^T\boldsymbol{A} + \mathcal{D}(\boldsymbol{\alpha})\big)^{-1} \end{cases}$$

  and thus $\hat{\boldsymbol{x}}_{\text{MMSE}} = \bar{\boldsymbol{\mu}}$.

- The EM algorithm can be used to estimate $\{\boldsymbol{\alpha}, \beta\}$ jointly with $\{\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}}\}$. Can implement with an $\mathcal{O}(NK^2)$ recursion after an $O(N^2M)$ initialization.

[1] Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Res.*, 2001.

[2] Wipf and Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Processing*, 2004.

[3] Ji, Xue, and Carin, "Bayesian Compressive Sensing," *IEEE Trans. Signal Processing*, 2008.

Bayesian Model Averaging versus the Relevance Vector Machine:

- Both are Bayesian approaches to sparse parameter estimation.
- For coefficient activity, RVM uses the continuous parameterization $\boldsymbol{\alpha}$, while BMA uses the discrete parameterization S.
- Implementations have roughly the same complexity (recall that FBMP is $\mathcal{O}(NMK)$ and RVM has $\mathcal{O}(NK^2)$ recursion plus $\mathcal{O}(N^2M)$ initialization).
- Upon termination, the RVM posterior is Gaussian

  $$p(\boldsymbol{x}|\boldsymbol{y}) \sim \mathcal{N}(\bar{\boldsymbol{\mu}}, \bar{\boldsymbol{\Sigma}})$$

  whereas the BMA posterior is a Gaussian mixture:

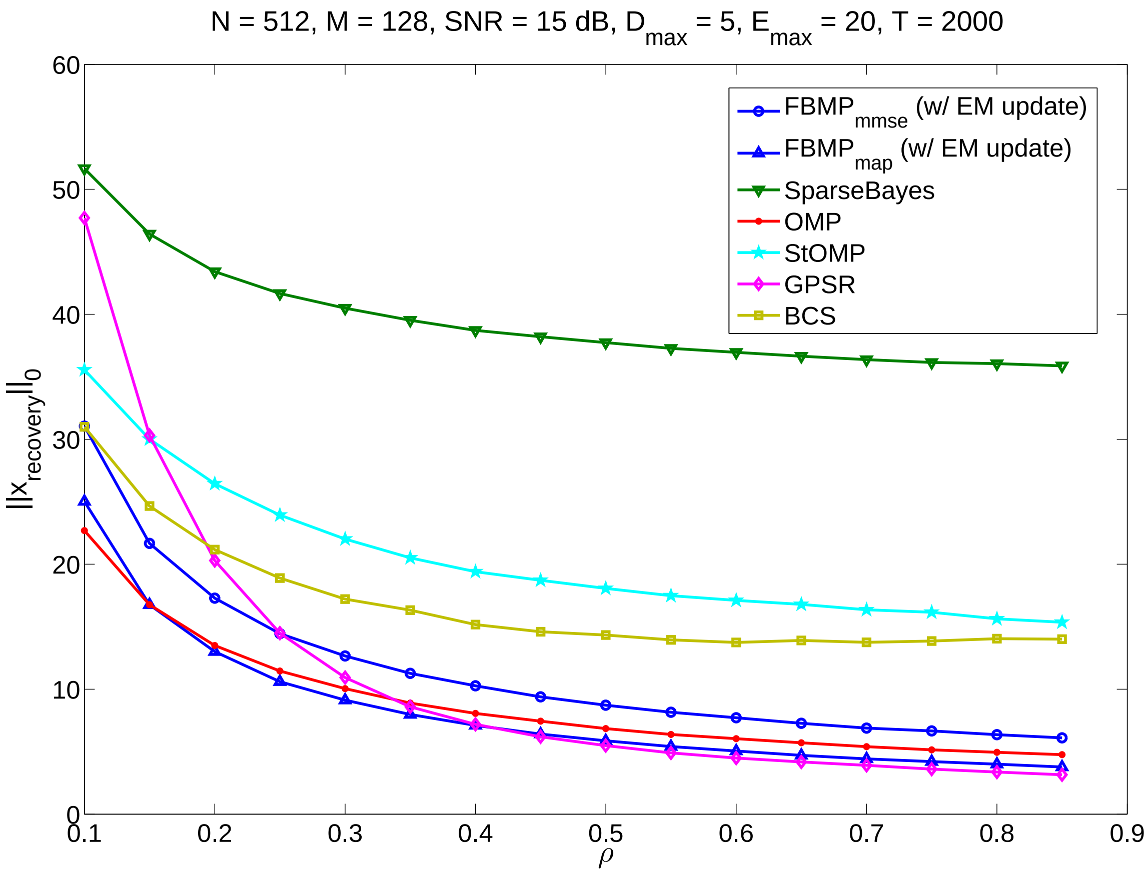
  $$p(\boldsymbol{x}|\boldsymbol{y}) \sim \sum_S \mathcal{N}\big(\hat{\boldsymbol{x}}_{\mathsf{MMSE}|S}, \boldsymbol{\Sigma}_S\big)\ p(S|\boldsymbol{y}).$$

  Thus, the BMA posterior can be more informative.
- Simulation results show advantages of BMA over RVM.

Conclusions:

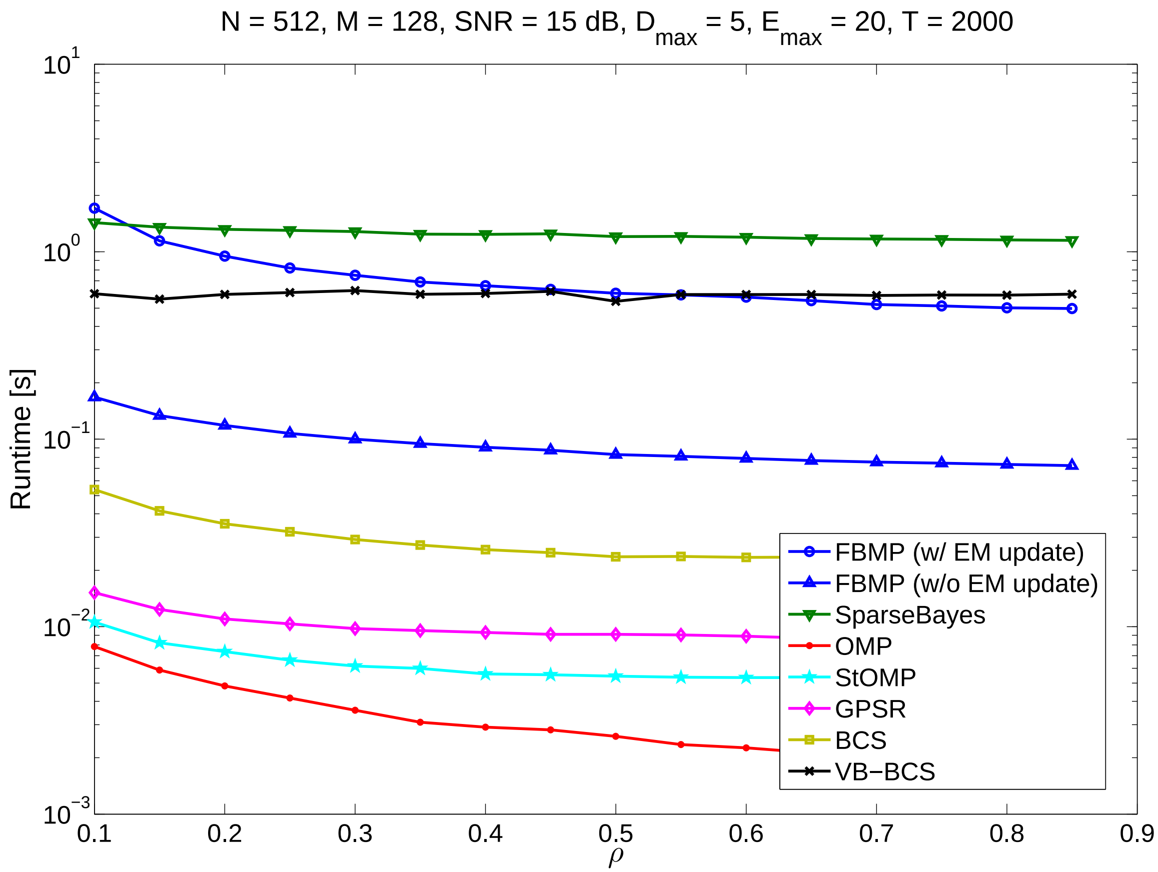
- Sparse reconstruction can be viewed as (discrete) model selection followed by (continuous) parameter estimation.
- Noncoherent decoding is (discrete) codeword selection under (continuous) nuisance parameters.
- Noncoherent decoding becomes equivalent to sparse reconstruction under a particular admissible model set $\mathbb{S}$.
- Noncoherent decoding techniques can be exploited for sparse reconstruction:
  - PWEP analyses for noncoherent decoding can be extended to yield PWEP analyses for model selection under general $\mathbb{S}$.
  - Noncoherent decoding algorithms based on soft tree search inspire low-complexity near-optimal sparse reconstruction algorithms like Fast Bayesian Matching Pursuit.

Sparsity of estimate versus decay rate ρ:



The estimates returned by FBMP are among the sparsest.

Runtime versus decay rate ρ:



FBMP (without EM iterations) is on par with other Bayesian algorithms, and a bit slower than other matching pursuit and convex programming algorithms.