Fast Bayesian Matching Pursuit

Philip Schniter, Lee C. Potter, and Justin Ziniel
Dept. of ECE, The Ohio State University, Columbus, OH 43210.
e-mail: schniter@ece.osu.edu, potter@ece.osu.edu, ziniel.1@osu.edu.

***Abstract*—A low-complexity recursive procedure is presented for minimum mean squared error (MMSE) estimation in linear regression models. A Gaussian mixture is chosen as the prior on the unknown parameter vector. The algorithm returns both an approximate MMSE estimate of the parameter vector and a set of high posterior probability mixing parameters. Emphasis is given to the case of a sparse parameter vector. Numerical simulations demonstrate estimation performance and illustrate the distinctions between MMSE estimation and MAP model selection. The set of high probability mixing parameters not only provides MAP basis selection, but also yields relative probabilities that reveal potential ambiguity in the sparse model.**[1]

I. Introduction

Sparse linear regression is a topic of long standing interest in statistics and signal processing. The linear regression model is

$$\boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\nu}, \tag{1}$$

with unknown parameter vector $\boldsymbol{x}$, unit norm columns in the regressor matrix $\boldsymbol{A}$, and additive noise $\boldsymbol{\nu}$. We provide a brief, and necessarily incomplete, survey of existing approaches, with an emphasis on themes relevant to the proposed estimator.

Algorithmic approaches have been proposed over several decades, providing greedy heuristic solutions. Examples include CLEAN [1], iteratively re-weighted least-squares [2], and orthogonal matched pursuit (OMP) [3]. Tropp and Gilbert [4], for example, provide sufficient conditions on the sparsity of $\boldsymbol{x}$ and correlation among columns of $\boldsymbol{A}$ such that the greedy OMP provides correct model selection with high probability in the noiseless measurement case.

In addition to greedy approaches, penalized least-squares solutions for $\boldsymbol{x}$ have likewise been presented in the past four decades. In this class of approaches, parameters are found via the optimization

$$\hat{\boldsymbol{x}} \;=\; \arg\min_{\boldsymbol{x}} \;\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 + \lambda\|\boldsymbol{x}\|_p^p, \tag{2}$$

or, equivalently for some $\epsilon > 0$

$$\hat{\boldsymbol{x}} \;=\; \arg\min_{\boldsymbol{x}} \;\|\boldsymbol{x}\|_p \;\text{ s.t. }\; \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 < \epsilon. \tag{3}$$

Ridge regression [5] (Tikhonov regularization) adopts $p = 2$, while basis pursuit [6] and Lasso [7] use $p = 1$. Equation (2) has been widely adopted, for example in radar imaging [8], image reconstruction [9], [10], and elsewhere [11], [12]. Elegant recent results by several authors [13]–[15] have demonstrated sufficient conditions on $\boldsymbol{A}$, $\boldsymbol{\nu}$, and sparsity of $\boldsymbol{x}$ such that the convex problem in (3) for $p = 1$ provides the unique solution to the non-convex task

$$\min \|\boldsymbol{x}\|_0 \;\text{ s.t. }\; \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 < \epsilon. \tag{4}$$

These proofs have validated the widespread use of (2)-(3), providing a deeper understanding, spurring a resurgent interest, and promoting the interpretation as "compressive sensing." The large class of methods adopting (2) may be interpreted as implicitly seeking the Bayesian MAP estimate of x under a sparsity inducing prior

$$p(\boldsymbol{x}) \;\propto\; \exp\big\{-\tfrac{\lambda}{2}\|\boldsymbol{x}\|_p^p\big\}. \tag{5}$$

The method of sparse Bayesian learning [16], [17] explicitly adopts a Bayesian framework with x_i independent, zero-mean, Gaussian with unknown variance σ_i^2. The unknown variances are given the Gamma conjugate prior, and an expectation-maximization (EM) iteration computes a MAP estimate of $\boldsymbol{x}$.

In the literature, primary focus is placed on the detection of the few significant entries of the sparse $\boldsymbol{x}$—a task alternatively known as model selection or basis selection. In contrast, we adopt a minimum mean-squared error (MMSE) estimation formulation and focus on accurately inferring $\boldsymbol{x}$ from the noisy observations, $\boldsymbol{y}$. The MMSE estimation approach was likewise adopted in a crisp exposition by Larsson and Selén [18]; as they noted, the Bayesian formulation requires *a priori* assumptions that are explicitly stated and admit unambiguous interpretation. (We specifically identify similarities to [18] in Section V.)

As a byproduct of approximating the proposed MMSE estimation algorithm, we also provide exact ratios of posterior probabilities for a set of high probability solutions to the detection problem. These relative probabilities serve to reveal potential ambiguity among multiple models, due to low signal-to-noise ratio and/or significant correlation among columns in the regressor matrix, $\boldsymbol{A}$.

The remainder of the paper is organized as follows. In Section II, we state the signal model and explicitly identify the assumed priors. In Section III, we describe our proposed technique, and in Section IV we investigate its performance numerically. In Section V, we discuss our findings, and in Section VI we conclude.

[1]This work was supported by the National Science Foundation under Grant 0237037, the Office of Naval Research grant N00014-07-1-0209, the Wright Brothers Institute grant GRT00009715, and AFOSR under award FA9550-06-1-0324.

II. Signal Model

Consider observing $\boldsymbol{y} \in \mathbb{R}^M$, a noisy linear combination of the parameters in $\boldsymbol{x} \in \mathbb{R}^N$:

$$\boldsymbol{y} \;=\; \boldsymbol{A}\boldsymbol{x} + \boldsymbol{\nu}, \tag{6}$$

where the noise $\boldsymbol{\nu}$ is assumed to be white Gaussian with variance σ^2, i.e., $\boldsymbol{\nu} \sim \mathcal{N}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M)$, and the columns of $\boldsymbol{A}$ are taken to be unit-norm. Our focus is on the under-determined case (i.e., $N \gg M$) with a suitably *sparse* parameter vector $\boldsymbol{x}$ (i.e., $\|\boldsymbol{x}\|_0 \ll N$).

To model sparsity, we assume that the parameters are generated from a Gaussian mixture density:

$$\boldsymbol{x}|\boldsymbol{s} \;\sim\; \mathcal{N}(\boldsymbol{0}, \boldsymbol{R}(\boldsymbol{s})), \tag{7}$$

where the covariance $\boldsymbol{R}(\boldsymbol{s})$ is determined by a discrete random vector $\boldsymbol{s} = [s_0, \ldots, s_{N-1}]^T$ of mixture parameters. For simplicity, we take $\boldsymbol{R}(\boldsymbol{s})$ to be diagonal with $[\boldsymbol{R}(\boldsymbol{s})]_{n,n} = \sigma^2_{s_n}$, implying that $\{x_n|s_n\}_{n=0}^{N-1}$ are independent with $x_n|s_n \;\sim \mathcal{N}(0, \sigma^2_{s_n})$. Also for simplicity, we assume that the mixture parameters $\{s_n\}_{n=0}^{N-1}$ are[2] Bernoulli(p_1). To model sparse $\boldsymbol{x}$, we choose $\sigma_0^2 = 0$ and $p_1 \ll 1$.

From the model assumptions it can be seen that

$$\begin{bmatrix}\boldsymbol{y}\\ \boldsymbol{x}\end{bmatrix}\Bigg| \boldsymbol{s} \;\sim\; \mathcal{N}\left(\boldsymbol{0}, \begin{bmatrix}\boldsymbol{\Phi}(\boldsymbol{s}) & \boldsymbol{A}\boldsymbol{R}(\boldsymbol{s})\\ \boldsymbol{R}(\boldsymbol{s})\boldsymbol{A}^T & \boldsymbol{R}(\boldsymbol{s})\end{bmatrix}\right), \tag{8}$$

where

$$\boldsymbol{\Phi}(\boldsymbol{s}) \;:=\; \boldsymbol{A}\boldsymbol{R}(\boldsymbol{s})\boldsymbol{A}^T + \sigma^2 \boldsymbol{I}_M. \tag{9}$$

III. Estimation of Basis and Parameters

In this section, we propose an efficient search procedure to find the most probable basis configurations along with their respective posterior probabilities. These posteriors can then be used to compute an MMSE estimate of the sparse parameters $\boldsymbol{x}$.

A. Basis Selection Metric

As a consequence of the model described in Section II, the nonzero locations in $\boldsymbol{s}$ specify which of the basis elements (i.e., columns of $\boldsymbol{A}$) are "active." Thus, basis selection reduces to estimation of $\boldsymbol{s}$. Because we have adopted a probabilistic model for $\{\boldsymbol{s}, \boldsymbol{y}\}$, we can not only compute *which* of the basis configurations are most likely, but also *how likely* these bases are. The latter is accomplished through the estimation of dominant posteriors $p(\boldsymbol{s}|\boldsymbol{y})$.

The posterior can be written, via Bayes rule, as

$$p(\boldsymbol{s}|\boldsymbol{y}) \;=\; \frac{p(\boldsymbol{y}|\boldsymbol{s})p(\boldsymbol{s})}{\sum_{\boldsymbol{s}' \in \mathcal{S}} p(\boldsymbol{y}|\boldsymbol{s}')p(\boldsymbol{s}')}, \tag{10}$$

where $\mathcal{S} = \{0,1\}^N$, which shows that estimating $\{p(\boldsymbol{s}|\boldsymbol{y})\}_{\boldsymbol{s}\in\mathcal{S}}$ reduces to estimating $\{p(\boldsymbol{y}|\boldsymbol{s})p(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$. The size of $\mathcal{S}$ makes it impractical to compute $\{p(\boldsymbol{s}|\boldsymbol{y})\}$ or $\{p(\boldsymbol{y}|\boldsymbol{s})p(\boldsymbol{s})\}$ for all $\boldsymbol{s} \in \mathcal{S}$. However, the set $\mathcal{S}_\star$ responsible for the *dominant* posteriors can be quite small and therefore practical to compute. Working in the log domain, we find

$$\begin{aligned}
\mu(\boldsymbol{s}) \;&:=\; \ln p(\boldsymbol{y}|\boldsymbol{s})p(\boldsymbol{s}) && (11)\\
&=\; \ln p(\boldsymbol{y}|\boldsymbol{s}) + \sum_{n=0}^{N-1} \ln p(s_n) && (12)\\
&=\; \ln p(\boldsymbol{y}|\boldsymbol{s}) + \|\boldsymbol{s}\|_0 \ln p_1 + (N - \|\boldsymbol{s}\|_0)\ln(1-p_1) && (13)\\
&=\; \ln p(\boldsymbol{y}|\boldsymbol{s}) + \|\boldsymbol{s}\|_0 \ln \tfrac{p_1}{1-p_1} + N\ln(1-p_1) && (14)\\
&=\; -\tfrac{M}{2}\ln 2\pi - \tfrac{1}{2}\ln\det(\boldsymbol{\Phi}(\boldsymbol{s})) - \tfrac{1}{2}\boldsymbol{y}^T\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{y}\\
&\quad + \|\boldsymbol{s}\|_0 \ln \tfrac{p_1}{1-p_1} + N\ln(1-p_1). && (15)
\end{aligned}$$

We will refer to $\mu(\boldsymbol{s})$ as the *basis selection metric*.

B. MMSE Parameter Estimation

For applications in which the identification of the most probable basis is the primary objective, the sparse coefficients $\boldsymbol{x}$ can be regarded as nuisance parameters. In other applications, however, estimation of $\boldsymbol{x}$ is the primary goal.

The MMSE estimate of $\boldsymbol{x}$ from $\boldsymbol{y}$ is

$$\hat{\boldsymbol{x}}_{\mathsf{mmse}} \;:=\; \mathrm{E}\{\boldsymbol{x}|\boldsymbol{y}\} \;=\; \sum_{\boldsymbol{s}\in\mathcal{S}} p(\boldsymbol{s}|\boldsymbol{y})\,\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\} \tag{16}$$

where from (8) it is straightforward (e.g., [19, p. 155]) to obtain

$$\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\} \;=\; \boldsymbol{R}(\boldsymbol{s})\boldsymbol{A}^T\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{y}. \tag{17}$$

Although exact evaluation of (16) involves a summation over 2^N terms, which may be computationally infeasible, the MMSE estimate can be closely approximated using only the dominant posteriors:

$$\hat{\boldsymbol{x}}_{\mathsf{ammse}} \;:=\; \sum_{\boldsymbol{s}\in\mathcal{S}_\star} p(\boldsymbol{s}|\boldsymbol{y})\,\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\}. \tag{18}$$

Likewise, the covariance of the corresponding estimation error can be closely approximated as

$$\begin{aligned}
\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y}\} \;\approx\; \sum_{\boldsymbol{s}\in\mathcal{S}_\star} p(\boldsymbol{s}|\boldsymbol{y})\big[&\mathrm{Cov}\left\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\right\} + (\hat{\boldsymbol{x}}_{\mathsf{ammse}} -\\
&\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\})(\hat{\boldsymbol{x}}_{\mathsf{ammse}} - \mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\})^T\big] && (19)\\
\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\} \;=\; \boldsymbol{R}(\boldsymbol{s}) - \boldsymbol{R}(\boldsymbol{s})\boldsymbol{A}^T&\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{A}\boldsymbol{R}(\boldsymbol{s}). && (20)
\end{aligned}$$

Note that, in evaluating (18)-(20), the primary challenge becomes that of obtaining $p(\boldsymbol{s}|\boldsymbol{y})$ and $\boldsymbol{\Phi}(\boldsymbol{s})^{-1}$ for each $\boldsymbol{s} \in \mathcal{S}_\star$. In the sequel, we propose a fast algorithm to search for the dominant basis configurations $\mathcal{S}_\star$ that, as a byproduct, also generates $p(\boldsymbol{s}|\boldsymbol{y})$ and $\boldsymbol{\Phi}(\boldsymbol{s})^{-1}$ for each of the $\boldsymbol{s}$ returned by the search.

C. Bayesian Matching Pursuit

We now describe an efficient means of determining $\mathcal{S}_\star$, the set of mixture parameters $\boldsymbol{s}$ yielding the dominant values of $p(\boldsymbol{s}|\boldsymbol{y})$, or, equivalently, the dominant values of $\mu(\boldsymbol{s})$. First we present a prosaic description of the search heuristic; the detailed algorithm will be specified in Section III-E.

The search starts with $\boldsymbol{s} = \boldsymbol{0}$ and first "turns on" one mixture parameter at a time, yielding a set of N binary vectors $\boldsymbol{s}$ which

[2] In other words, s_n is binary with $\Pr\{s_n = 1\} = p_1$ and $\Pr\{s_n = 0\} = 1 - p_1$.

we refer to as $\mathcal{S}^{(1)}$. The metrics $\mu(\boldsymbol{s})$ for the N vectors in $\mathcal{S}^{(1)}$ are then computed, and the elements of $\mathcal{S}^{(1)}$ with the D largest metrics are collected in $\mathcal{S}_\star^{(1)}$. For each candidate in $\mathcal{S}_\star^{(1)}$, all locations of a second active mixture parameter are then considered, yielding $(N-1)+(N-2)+\cdots+(N-D) = ND - \frac{D(D+1)}{2}$ unique binary vectors to store in $\mathcal{S}^{(2)}$. The metrics $\mu(\boldsymbol{s})$ for all vectors in $\mathcal{S}^{(2)}$ are then computed, and the elements of $\mathcal{S}^{(2)}$ with the D largest metrics are collected in $\mathcal{S}_\star^{(2)}$. Then, for each candidate vector in $\mathcal{S}_\star^{(2)}$, all possibilities of a third active mixture parameter are considered, and those with the D largest metrics are stored in $\mathcal{S}_\star^{(3)}$. The process continues until $\mathcal{S}_\star^{(P)}$ is computed, where P can be chosen[3] to make $\Pr(\|\boldsymbol{s}\|_0 > P)$ sufficiently small.[4] Note that $\mathcal{S}_\star^{(P)}$ constitutes the algorithm's final estimate of $\mathcal{S}_\star$. Henceforth we denote this final estimate by $\hat{\mathcal{S}}_\star$.

D. Fast Metric Update

For use with the aforementioned Bayesian matching pursuit (BMP) algorithm, we propose a fast metric update which computes the change in $\mu(\cdot)$ that results from the activation of a single mixture parameter. More precisely, if we denote by $\boldsymbol{s}_n$ the vector identical to $\boldsymbol{s}$ except for the n^{th} coefficient, which is active in $\boldsymbol{s}_n$ but inactive in $\boldsymbol{s}$ (i.e., $[\boldsymbol{s}_n]_n = 1$ and $[\boldsymbol{s}]_n = 0$), then we seek an efficient method of computing $\Delta_n(\boldsymbol{s}) := \mu(\boldsymbol{s}_n) - \mu(\boldsymbol{s})$. Note that the metric at the root node (i.e., $\boldsymbol{s} = \boldsymbol{0}$) is

$$\begin{aligned}\mu(\boldsymbol{0}) \;=\; & -\tfrac{M}{2}\ln 2\pi - \tfrac{M}{2}\ln\sigma^2 - \tfrac{1}{2\sigma^2}\|\boldsymbol{y}\|_2^2 \\ & + N\ln(1-p_1) \end{aligned} \tag{21}$$

via (15) and the fact that $\boldsymbol{\Phi}(\boldsymbol{0}) = \sigma^2\boldsymbol{I}_M$.

To derive the fast metric update, we start with the property that, for any n and $\boldsymbol{s}$,

$$\boldsymbol{\Phi}(\boldsymbol{s}_n) \;=\; \boldsymbol{\Phi}(\boldsymbol{s}) + \sigma_1^2\boldsymbol{a}_n\boldsymbol{a}_n^T, \tag{22}$$

from which the matrix inversion lemma implies

$$\boldsymbol{\Phi}(\boldsymbol{s}_n)^{-1} \;=\; \boldsymbol{\Phi}(\boldsymbol{s})^{-1} - \sigma_1^2\beta_n\boldsymbol{b}_n\boldsymbol{b}_n^T \tag{23}$$

$$\boldsymbol{b}_n \;:=\; \boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{a}_n \tag{24}$$

$$\beta_n \;:=\; \left(1 + \sigma_1^2\boldsymbol{a}_n^T\boldsymbol{b}_n\right)^{-1}. \tag{25}$$

Equations (22)-(25) then imply

$$\begin{aligned}\boldsymbol{y}^T\boldsymbol{\Phi}(\boldsymbol{s}_n)^{-1}\boldsymbol{y} \;&=\; \boldsymbol{y}^T\left(\boldsymbol{\Phi}(\boldsymbol{s})^{-1} - \sigma_1^2\beta_n\boldsymbol{b}_n\boldsymbol{b}_n^T\right)\boldsymbol{y} && (26)\\ &=\; \boldsymbol{y}^T\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{y} - \sigma_1^2\beta_n(\boldsymbol{y}^T\boldsymbol{b}_n)^2 && (27)\\ \ln\det(\boldsymbol{\Phi}(\boldsymbol{s}_n)) \;&=\; \ln\det\left(\boldsymbol{\Phi}(\boldsymbol{s}) + \sigma_1^2\boldsymbol{a}_n\boldsymbol{a}_n^T\right) && (28)\\ &=\; \ln\left[\left(1 + \sigma_1^2\boldsymbol{a}_n^T\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{a}_n\right)\det\left(\boldsymbol{\Phi}(\boldsymbol{s})\right)\right] \\ &=\; \ln\det(\boldsymbol{\Phi}(\boldsymbol{s})) - \ln\beta_n && (29)\\ \|\boldsymbol{s}_n\|_0\ln\tfrac{p_1}{1-p_1} \;&=\; (\|\boldsymbol{s}\|_0 + 1)\ln\tfrac{p_1}{1-p_1} && (30)\\ &=\; \|\boldsymbol{s}\|_0\ln\tfrac{p_1}{1-p_1} + \ln\tfrac{p_1}{1-p_1}, && (31)\end{aligned}$$

which, combined with (15), yield

$$\mu(\boldsymbol{s}_n) \;=\; \mu(\boldsymbol{s}) + \underbrace{\tfrac{1}{2}\ln\beta_n + \tfrac{\sigma_1^2}{2}\beta_n(\boldsymbol{y}^T\boldsymbol{b}_n)^2 + \ln\tfrac{p_1}{1-p_1}}_{\Delta_n(\boldsymbol{s})}. \tag{32}$$

In summary, $\Delta_n(\boldsymbol{s})$ in (32) quantifies the change in our basis selection metric $\mu(\cdot)$ due to the activation of the n^{th} tap of $\boldsymbol{s}$.

E. Fast Bayesian Matching Pursuit

Notice that the cost of computing $\{\beta_n\}_{n=0}^{N-1}$ via (24)-(25) is $\mathcal{O}(NM^2)$ if standard matrix multiplication is used. As we now describe, the complexity of this operation can be made linear in M by exploiting the structure of $\boldsymbol{\Phi}(\boldsymbol{s})^{-1}$.

Say that $\boldsymbol{t} = [t_1, t_2, \ldots, t_p]^T$ contains the indices of active elements in $\boldsymbol{s}$. Then, from (23),

$$\boldsymbol{\Phi}(\boldsymbol{s})^{-1} \;=\; \tfrac{1}{\sigma^2}\boldsymbol{I}_M - \sigma_1^2\textstyle\sum_{i=1}^p \beta^{(i)}\boldsymbol{b}^{(i)}\boldsymbol{b}^{(i)T}, \tag{33}$$

where $\boldsymbol{b}^{(i)}$ and $\beta^{(i)}$ denote the values of $\boldsymbol{b}$ and β generated while activating index t_i in the mixture parameter vector defined by the active indices $[t_1, \ldots, t_{i-1}]$. From (24), we are required to compute

$$\boldsymbol{b}_n \;=\; \tfrac{1}{\sigma^2}\boldsymbol{a}_n - \sigma_1^2\textstyle\sum_{i=1}^p \beta^{(i)}\boldsymbol{b}^{(i)}\underbrace{\boldsymbol{b}^{(i)T}\boldsymbol{a}_n}_{:=\, c_n^{(i)}} \tag{34}$$

when activating the n^{th} tap in $\boldsymbol{s}$. The key observation is that the coefficients $\{c_n^{(i)}\}_{n=0}^{N-1}$ need only be computed once, i.e., when index t_i is activated. Furthermore, $\{c_n^{(i)}\}_{n=0}^{N-1}$ only need to be computed for *surviving* indices t_i. These tricks form the foundation of the Fast Bayesian Matching Pursuit (FBMP) algorithm outlined in Table I. From the table, it is straightforward to verify that the number of multiplications required by the algorithm is $\mathcal{O}(NMPD)$.

IV. Numerical Experiments

A. FBMP Behavior

Numerical experiments were conducted to investigate the performance of FBMP from Table I for various values of model and algorithmic parameters, and the results are reported in Figs. 1-5. Unless otherwise noted, the experiments used $N = 256$, $M = 64$, $\mathsf{SNR} = 15$ dB, $p_1 = 0.04$, and $P = \left\lceil \mathrm{erfc}^{-1}(2P_0)\sqrt{2Np_1(1-p_1)} + Np_1 \right\rceil$ where $P_0 = 0.00005$ is the target value of $\Pr\{\|\boldsymbol{s}\|_0 > P\}$ as suggested in Section III-C. Here we use $\mathsf{SNR} := \frac{\sigma_1^2 p_1 N}{\sigma^2 M}$, as motivated by the unit-norm assumption on the columns of $\boldsymbol{A}$. The plots represent an average of 200 independent model realizations. For each realization of $\boldsymbol{A}$, an i.i.d. normal matrix was drawn and then scaled to make each of its columns unit-norm. Note that the average number of active coefficients $\mathrm{E}\{\|\boldsymbol{x}\|_0\} = p_1 N$ is approximately equal to 10 when $p_1 = 0.04$ and $N = 256$. When referring to the "normalized mean-squared error" (NMSE) of an estimate $\hat{\boldsymbol{x}}$, we mean $\mathrm{E}\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2/\|\boldsymbol{x}\|_2^2\}$.

In Fig. 1, we plot NMSE versus observation length M for FBMP under several values of the search parameter D. Recall that D effects a tradeoff between search accuracy and search complexity (the latter of which is expected to grow linearly

[3] One could also determine the stopping parameter P adaptively.

[4] Notice that $\|\boldsymbol{s}\|_0$ is Binomial(N, p_1) distribution. When $Np_1 > 5$, it is common to use the approximation $\|\boldsymbol{s}\|_0 \sim \mathcal{N}\big(Np_1, Np_1(1-p_1)\big)$, in which case $\Pr(\|\boldsymbol{s}\|_0 > P) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{P - Np_1}{\sqrt{2Np_1(1-p_1)}}\right)$.

$\mu_{0,1} = -\frac{M}{2}\ln 2\pi - \frac{M}{2}\ln\sigma^2 - \frac{1}{2\sigma^2}\|\boldsymbol{y}\|_2^2 + N\ln(1-p_1)$;
for $n = 1:N$,
  $\tilde{\boldsymbol{b}}_{1,n} = \sigma^{-2}\boldsymbol{a}_n$;
  $\tilde{\beta}_{1,n} = \left(1 + \sigma_1^2 \boldsymbol{a}_n^T \tilde{\boldsymbol{b}}_{1,n}\right)^{-1}$;
  $\tilde{\mu}_{1,n} = \mu_{0,1} + \frac{1}{2}\log\tilde{\beta}_{1,n} + \frac{\sigma_1^2}{2}\tilde{\beta}_{1,n}(\boldsymbol{y}^T\tilde{\boldsymbol{b}}_{1,n})^2 + \log\frac{p_1}{1-p_1}$;
end
for $q = 1:D$,
  $n_\star = n$ corresponding to q^{th} largest $\tilde{\mu}_{1,n}$;
  $\mu_{1,q} = \tilde{\mu}_{1,n_\star}$;
  $\boldsymbol{b}_{1,q}^{(1)} = \tilde{\boldsymbol{b}}_{1,n_\star}$; $\boldsymbol{c}_{1,q}^{(1)} = \boldsymbol{A}^T\boldsymbol{b}_{1,q}^{(1)}$; $\beta_{1,q}^{(1)} = \tilde{\beta}_{1,n_\star}$; $t_{1,q}^{(1)} = n_\star$;
end
for $p = 2:P$,
  for $d = 1:D$,
    for $n = 1:N$,
      $\tilde{\boldsymbol{b}}_{d,n} = \sigma^{-2}\boldsymbol{a}_n - \sum_{i=1}^{p-1}\boldsymbol{b}_{p-1,d}^{(i)}\beta_{p-1,d}^{(i)}\left[\boldsymbol{c}_{p-1,d}^{(i)}\right]_n$;
      $\tilde{\beta}_{d,n} = \left(1 + \sigma_1^2\boldsymbol{a}_n^T\tilde{\boldsymbol{b}}_{d,n}\right)^{-1}$;
      $\tilde{\mu}_{d,n} = \mu_{p-1,d} + \frac{1}{2}\log\tilde{\beta}_{d,n} + \frac{\sigma_1^2}{2}\tilde{\beta}_{d,n}(\boldsymbol{y}^T\tilde{\boldsymbol{b}}_{d,n})^2 + \log\frac{p_1}{1-p_1}$;
      if $n \in \{t_{p-1,d}^{(i)}\}_{i=1}^{p-1}$ then $\tilde{\mu}_{d,n} = -\infty$;
    end
  end
  for $q = 1:D$,
    $\{d_\star, n_\star\} = \{d, n\}$ corresponding to q^{th} largest $\tilde{\mu}_{d,n}$;
    $\mu_{p,q} = \tilde{\mu}_{d_\star,n_\star}$;
    $\boldsymbol{b}_{p,q}^{(p)} = \tilde{\boldsymbol{b}}_{d_\star,n_\star}$; $\boldsymbol{c}_{p,q}^{(p)} = \boldsymbol{A}^T\boldsymbol{b}_{p,q}^{(p)}$; $\beta_{p,q}^{(p)} = \tilde{\beta}_{d_\star,n_\star}$; $t_{p,q}^{(p)} = n_\star$;
    for $i = 1:p-1$,
      $\boldsymbol{b}_{p,q}^{(i)} = \boldsymbol{b}_{p-1,q}^{(i)}$; $\boldsymbol{c}_{p,q}^{(i)} = \boldsymbol{c}_{p-1,q}^{(i)}$; $\beta_{p,q}^{(i)} = \beta_{p-1,q}^{(i)}$; $t_{p,q}^{(i)} = t_{p-1,q}^{(i)}$;
    end
  end
end

TABLE I
FAST BAYESIAN MATCHING PURSUIT

in D). There we see that NMSE performance improves as M gets larger, i.e., as the average number of unknown parameters per measurement $\frac{p_1 N}{M}$ decreases. For $D = 1$ (i.e., the simplest possible search), Fig. 1 shows a "knee" in the curve at $M = 64$ (i.e., $\frac{p_1 N}{M} = 0.16$) below which NMSE degrades quickly. By increasing search complexity D, the knee shifts so that the FBMP is robust to a wider range of M (e.g., $M = 48$ or $\frac{p_1 N}{M} = 0.21$ when $D = 5$). The benefits of increased D diminish quickly for $D > 5$, however.

In Fig. 2, we plot the number of active basis elements missing from FBMP's estimate of the MAP basis configuration:

$$\hat{\boldsymbol{s}}_{\mathsf{map}} := \underset{s \in \hat{\mathcal{S}}_\star}{\operatorname{argmax}}\, p(\boldsymbol{s}|\boldsymbol{y}). \tag{35}$$

In particular, the traces in Fig. 2 show number-of-misses versus observation length M for FBMP under several values of search parameter D. Because the number-of-misses in Fig. 2 closely parallel the NMSEs in Fig. 1, we conjecture that the sub-optimality of FBMP's greedy search is to blame for the relatively large NMSE values that occur when $M < 64$ (i.e., when $\frac{p_1 N}{M} > 0.16$).

In Fig. 3, we plot NMSE versus $p_1 N$, the expected number of active coefficients, for FBMP under several values of search parameter D. There we see that NMSE performance quickly degrades as $p_1 N$ increases above $p_1 N = 10$ (i.e., above $\frac{p_1 N}{M} = 0.16$), mirroring the results in Figs. 1-2. As in Fig. 1, when $\frac{p_1 N}{M} > 0.16$, increasing D from 1 to 10 can yield an NMSE improvement of 3 dB. When $\frac{p_1 N}{M} \leq 0.16$, however, $D = 1$ appears to suffice.

In Fig. 4, we plot NMSE versus SNR for FBMP under several values of search parameter D (where $\{M, p_1\}$ correspond to the aforementioned breakpoints in the NMSE-vs-M and NMSE-vs-$p_1 N$ curves). Figure 4 shows a satisfying linear relationship between NMSE and SNR (in dB). As expected, the effect of increasing D from 1 to 10 is negligible because $\frac{p_1 N}{M} = 0.16$; a more significant effect would be expected if $\frac{p_1 N}{M}$ had been larger.

In Fig. 5, we plot NMSE versus SNR for two FBMP-derived estimates: the (approximate) MMSE estimate $\hat{\boldsymbol{x}}_{\mathsf{ammse}}$ from (18) and the quasi-MAP estimate $\hat{\boldsymbol{x}}_{\mathsf{amap}}$ from (36):

$$\hat{\boldsymbol{x}}_{\mathsf{amap}} := \mathrm{E}\{\boldsymbol{x}|\boldsymbol{y}, \hat{\boldsymbol{s}}_{\mathsf{map}}\}. \tag{36}$$

Whereas $\hat{\boldsymbol{x}}_{\mathsf{ammse}}$ is the *average* of the conditional MMSE estimates $\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{s}\}$ over $\boldsymbol{s} \in \hat{\mathcal{S}}_\star$, the estimate $\hat{\boldsymbol{x}}_{\mathsf{amap}}$ is MMSE conditioned on (FBMP's estimate of) the MAP basis-configuration $\hat{\boldsymbol{s}}_{\mathsf{map}}$. In terms of average NMSE, Fig. 5 demonstrates that $\hat{\boldsymbol{x}}_{\mathsf{ammse}}$ are about 1 dB better than $\hat{\boldsymbol{x}}_{\mathsf{map}}$ at SNR ≤ 10 dB and about 0.5 dB better at SNR > 10 dB. The improvement reflects the advantage of allowing for *uncertainty* in the estimated basis.

Finally, in Fig. 6, we plot average FBMP runtime versus search parameter D. As expected from the algorithmic description in Table I, the runtime scales linearly in D.

B. Comparison To Other Algorithms

In Figs. 7-8 we compare FBMP to several other popular sparse estimation algorithms, including SparseBayes [16], OMP [4], StOMP [20], GPSR-Basic [21], and BCS [22]. Unless otherwise noted, the model parameters were set at $N = 512$, $M = 128$, $p_1 = 0.04$, and $\sigma^2 = 0.001$ (which corresponds to SNR $= 19$ dB at the nominal values of N, M, and p_1). Our plots represent an average of 100 independent model realizations.

For FBMP, we used *non*-optimized MATLAB code (which we plan to optimize in the near future), and unless otherwise noted used $D = 5$ and the same P specified in Section IV-A. For the other algorithms, we used the publicly available implementations that were found at the web-sites listed in our bibliography. The algorithmic parameters were chosen largely in accordance with suggested values provided by the authors of the software, or in accordance with values used in examples that accompanied the algorithms. The SparseBayes algorithm was tested with the initial hyper-parameter set to $\alpha = 1$. StOMP was tested using the "False Alarm Control" thresholding strategy, with the thresholding parameter set to $\frac{M}{NQ}(1 - \frac{1}{M}\|\boldsymbol{x}\|_0)$, where the default number of iterations, $Q = 10$, was used. The ℓ_1-penalty in the GPSR algorithm was chosen as $\tau = 0.1\|\boldsymbol{A}^H\boldsymbol{y}\|_\infty$, and the MSE kept for comparison purposes was the smaller of the MSEs of the un-debiased and debiased estimates. The BCS algorithm was tested with the "Adaptive CS" option turned off.

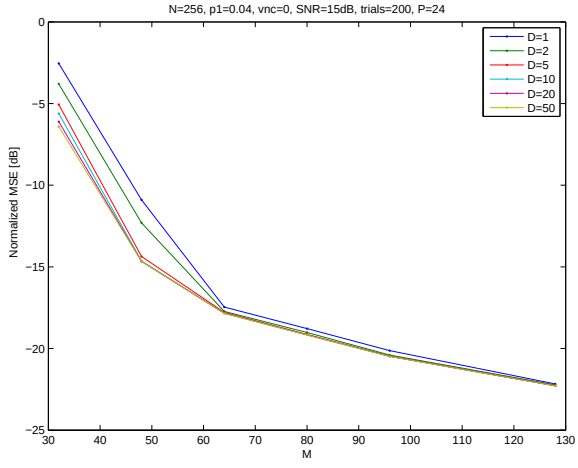

Fig. 1. Normalized MSE versus observation length M for FBMP under several values of search parameter D. From top to bottom, the traces reflect increasing D. (See the graph title for configuration.)

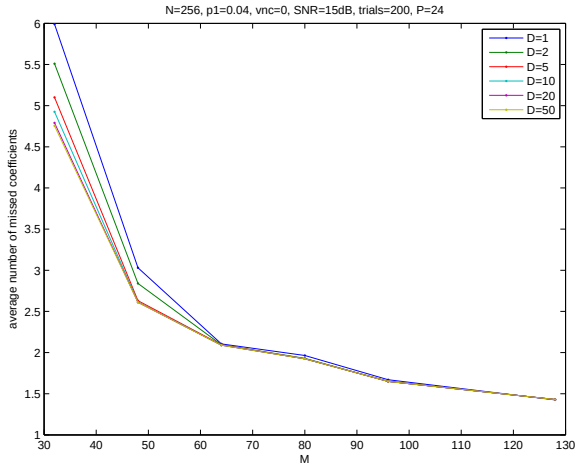

Fig. 2. Number of active coefficients in $\boldsymbol{x}$ missing from FBMP's estimate of the most probably basis configuration versus observation length M and under several values of search parameter D. From top to bottom, the traces reflect increasing D. (See the graph title for configuration.)

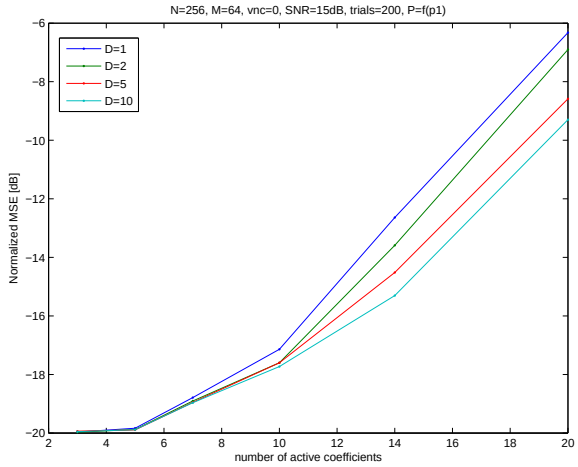

Fig. 3. Normalized MSE versus number of active coefficients (i.e., $\|\boldsymbol{x}\|_0$) for FBMP under several values of search parameter D. From top to bottom, the traces reflect increasing D. (See the graph title for configuration.)

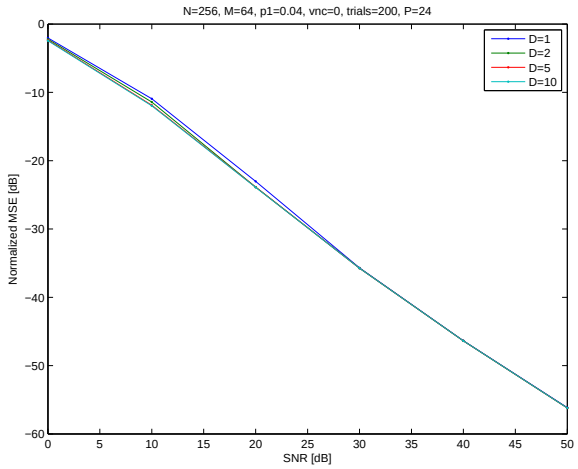

Fig. 4. Normalized MSE versus SNR for FBMP under several values of search parameter D. From top to bottom, the traces reflect increasing D. (See the graph title for configuration.)

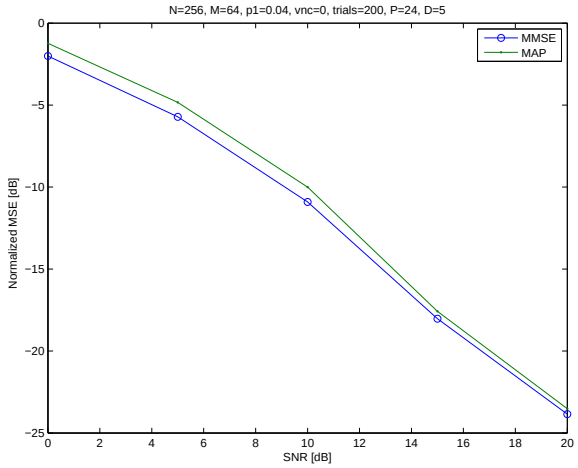

Fig. 5. Normalized MSE versus SNR of FBMP-returned MMSE estimate $\hat{\boldsymbol{x}}_{\mathsf{ammse}}$ and quasi-MAP estimate $\hat{\boldsymbol{x}}_{\mathsf{amap}}$. (See the graph title for configuration.)

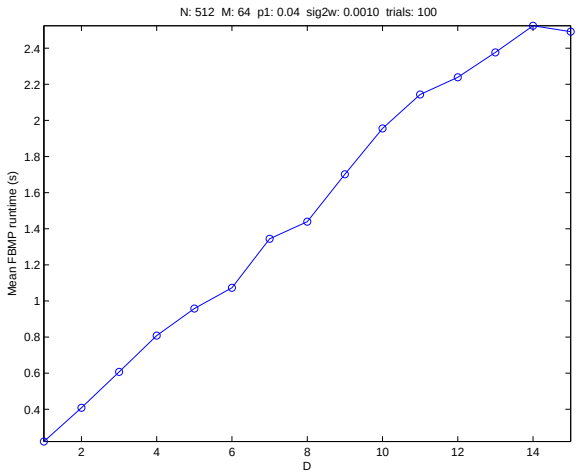

Fig. 6. Average FBMP runtime versus search parameter D. (See the graph title for configuration.)

In Fig. 7 we plot NMSE versus observation length M (at $\sigma^2 = 0.001$) for the various sparse estimation algorithms. There we see that FBMP achieved significantly lower NMSE than the other algorithms over the examined range of M. In particular, it outperformed BCS by approximately 3 dB, it outperformed OMP by 3 dB at small M and 10 dB at large M, and it outperformed the other algorithms by even more. In Fig. 8 we plot NMSE versus SNR (at $M = 128$) for the various sparse estimation algorithms. Again, the NMSEs achieved by FBMP were significantly lower than those achieved by the other algorithms. At SNR $= 22$ dB, FBMP outperformed BCS by approximately 3 dB and the other algorithms by > 9 dB; at SNR $= 15$ dB, FBMP outperformed all other algorithms by > 6 dB; and, at SNR $= 3$ dB, FBMP outperformed GPSR by approximately 1 dB and the other algorithms by > 5 dB.

Finally, in Fig. 9, we plot average runtime versus observation length M for the various sparse estimation algorithms. For FBMP, we used $D = 1$. Figure 9 shows that FBMP is about an order of magnitude faster than SparseBayes, on the same order of complexity as BCS, and about an order of magnitude slower than OMP, StOMP, and GPSR. We anticipate that optimized FBMP code will yield improved runtimes.

V. Discussion

The Bayesian framework provides a report on the confidence of estimates of both the coefficients $\boldsymbol{x}$ and the basis configuration $\boldsymbol{s}$. In particular, the basis selection metric $\mu(\boldsymbol{s})$ provides a posterior confidence label for a candidate basis configuration $\boldsymbol{s}$, in addition to providing the MMSE estimate $\hat{\boldsymbol{x}}_{\text{mmse}}$ through (16). Specifically, from (10), we can write the posterior probability of basis configuration $\boldsymbol{s}$ as

$$p(\boldsymbol{s}|\boldsymbol{y}) = \frac{\exp\{\mu(\boldsymbol{s})\}}{\sum_{\boldsymbol{s}'\in\mathcal{S}}\exp\{\mu(\boldsymbol{s}')\}} \approx \frac{\exp\{\mu(\boldsymbol{s})\}}{\sum_{\boldsymbol{s}'\in\mathcal{S}_\star}\exp\{\mu(\boldsymbol{s}')\}}, \quad (37)$$

where the approximation in (37) includes only the basis configurations $\mathcal{S}_\star \subset \mathcal{S}$ that account for the dominant values of $\exp\{\mu(\boldsymbol{s})\}$. Likewise, (19) provides an approximate error covariance for the MMSE estimate $\hat{\boldsymbol{x}}_{\text{mmse}}$. These posterior confidence values reflect the ambiguity inherently present in the sparse inference problem—an ambiguity especially evident when the SNR is low and/or the correlation among the columns of $\boldsymbol{A}$ is high.

Standard errors for estimated $\hat{\boldsymbol{x}}$ are largely absent in the compressive sensing literature. Exceptions are found in [7], [22] which give the error covariance for the simple linear problem conditioned on *perfect knowledge of the active basis elements*. As noted by Tibshirani [7], such a measure of posterior uncertainty has dubious value, because "a difficulty with this formula is that it gives an estimated variance of 0 for predictors with" $s_i = 0$. In this light, we expect certain advantages for algorithms that consider the active basis as implicitly uncertain.

A Gaussian mixture model similar to that in Section II was likewise adopted by Larsson and Selén [18], who also constructed the MMSE estimate in the manner of (18) but with an $\mathcal{S}_\star$ that contains exactly one sequence $\boldsymbol{s}$ for each Hamming

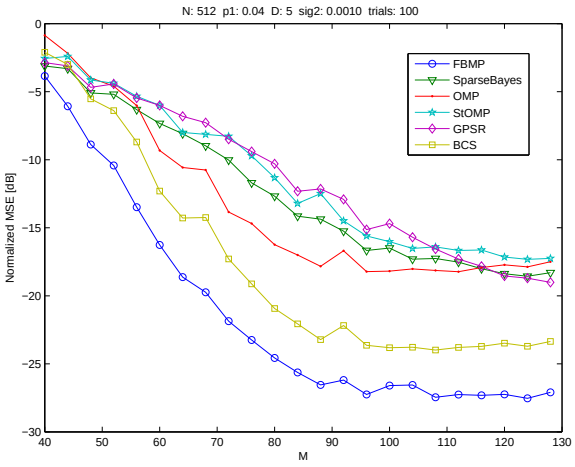

Fig. 7. Normalized MSE versus observation length M for several algorithms. (See the graph title for configuration.)

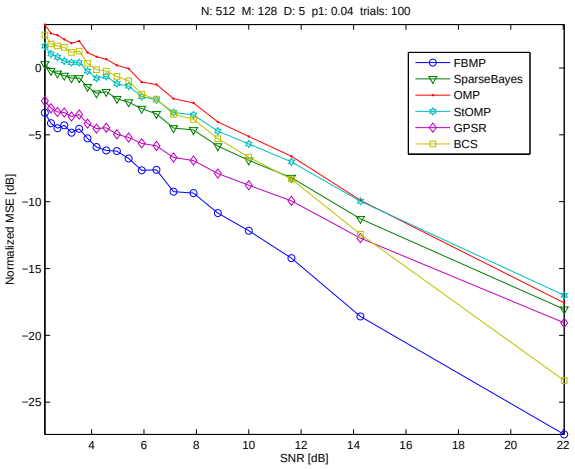

Fig. 8. Normalized MSE versus SNR for several algorithms. (See the graph title for configuration.)

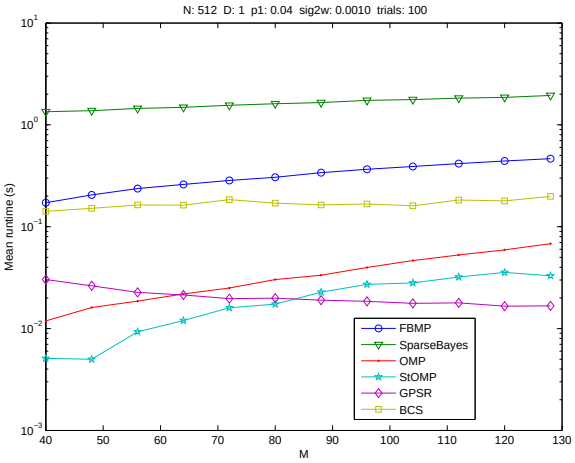

Fig. 9. Average runtime versus observation length M for several algorithms. (See the graph title for configuration.)

weight 0 to N. They proposed to find these $\boldsymbol{s}$ via greedy deflation, i.e., starting with an all-active basis configuration and recursively deactivating one element at a time. Thus, the $D = 1$ version of the BMP heuristic from Section III-C recalls the heuristic of [18], but in reverse. Note, however, that the *fast* $D = 1$ BMP presented in Section III-E has a complexity of only $\mathcal{O}(NMP)$, in comparison to $\mathcal{O}(N^3M^2)$ for the technique in [18]. Given the typically large values of N encountered in practice, the complexity of FBMP can be several orders of magnitude lower than that of [18].

As a caveat, we should emphasize that our greedy FBMP search returns only $\hat{\mathcal{S}}_\star$, an *estimate* of the dominant subset $\mathcal{S}_\star$, along with the values of $\mu(\boldsymbol{s})$ for $\boldsymbol{s} \in \hat{\mathcal{S}}_\star$. Thus, while the values $\mu(\boldsymbol{s})$ returned by FBMP can be used to compute exact *ratios* between the posterior probabilities of the configurations in $\hat{\mathcal{S}}_\star$, the *absolute* posteriors of these configurations (as approximated by (37) with $\hat{\mathcal{S}}_\star$ in place of $\mathcal{S}_\star$) will only be accurate when $\hat{\mathcal{S}}_\star$ indeed contains $\mathcal{S}_\star$. For example, if FBMP somehow missed the MAP configuration $\hat{\boldsymbol{s}}_{\mathsf{map}}$ (i.e., $\hat{\boldsymbol{s}}_{\mathsf{map}} \notin \hat{\mathcal{S}}_\star$), then we would expect a large discrepancy between $\sum_{\boldsymbol{s}' \in \mathcal{S}_\star} \exp\{\mu(\boldsymbol{s}')\}$ and $\sum_{\boldsymbol{s}' \in \hat{\mathcal{S}}_\star} \exp\{\mu(\boldsymbol{s}')\}$ which would in turn corrupt the FBMP estimates of $p(\boldsymbol{s}|\boldsymbol{y})$ and $\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y}\}$. Fortunately, the proposed greedy FBMP basis search seems to perform quite well, as least for $\frac{p_1 N}{M} \leq 0.16$ (as suggested by Fig. 2).

Although the model in Section II assumed that each x_i is generated according to a binary mixture of zero-mean Gaussians, one can imagine extending the model to, e.g., a mixture of finitely many Gaussians with non-zero means. In this case, one would need to generalize the BMP search heuristic of Section III-C to handle several types of active coefficient (e.g., one for each allowed mean).

VI. Conclusion

In this paper, we proposed an algorithm for joint basis selection and sparse parameter estimation which we call fast Bayesian matching pursuit (FBMP). In brief, FBMP models each unknown coefficient x_i as either inactive or active (with prior probability p_1), where an i.i.d. Gaussian distribution (with zero mean and variance σ_1^2) is assigned to the values of active coefficients. The observation $\boldsymbol{y}$ is then modeled as an AWGN-corrupted version of the unknown coefficients that has been mixed by a known matrix $\boldsymbol{A}$. FBMP navigates through the tree of active/inactive configurations $\mathcal{S}$ with the goal of finding the configurations with dominant posterior probability, $\mathcal{S}_\star$. The search is controlled by a parameter D which effects a tradeoff between complexity and accuracy. Numerical experiments suggest that the estimates returned by FBMP outperform (in normalized MSE) those of other popular algorithms (e.g., SparseBayes, OMP, StOMP, GPSR-Basic, BCS) by several dB in typical situations.

We plan, in the near future, to extend FBMP to the case where the active coefficients are complex Gaussian with non-zero means chosen from a finite set according to some prior probabilities. An example of the non-zero mean sparse signal model can be found in electron paramagnetic resonance imaging [23], where micro-liter particulate probes are inserted into a tumor and fill less than 0.25% volume in the field of view. The fabrication of the paramagnetic signal probes results in variable shape, size and electron spin density, giving rise to a non-zero-mean and nearly Gaussian distribution of signal strength in a very few active voxels.

References

[1] J. H ogbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astrophys. J. Suppl. Ser*, vol. 15, pp. 417–426, 1974.

[2] H. Lee, D. Sullivan, and T. Huang, "Improvement of discrete band-limited signal extrapolation by iterative subspace modification," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 1569 – 1572, 1987.

[3] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad., "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in *Proc. 27th Ann. Asilomar Conf. Signals, Systems, and Computers*, 1993.

[4] J. Tropp and A. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Information Theory*, vol. 53, pp. 4655–4666, 2007. (software available at `http://sparselab.stanford.edu/`).

[5] A. E. Hoerl and R. W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–67, 1970.

[6] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[7] R. Tibshirani, "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B*, vol. 58, no. 1, pp. 267 – 288, 1996.

[8] M. Çetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation based on nonquadratic regularization," *IEEE Trans. Image Process.*, vol. 10, pp. 623–631, 2001.

[9] C. Bouman and K. Sauer, "A generalized Gaussian image model for edge-preserving MAP estimation," *IEEE Trans. Image Process.*, vol. 2, pp. 296–310, 1993.

[10] A. H. Delaney and Y. Bresler, "A fast and accurate Fourier algorithm for iterative parallel-beam tomography," *IEEE Trans. Image Process.*, vol. 5, pp. 740–753, May 1996.

[11] S. Levy and P. K. Fullagar, "Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution," *Geophysics*, vol. 46, no. 9, pp. 1235–1243, 1981.

[12] B. D. Rao and K. Kreutz-Delgado, "An affine scaling methodology for best basis selection," *IEEE Trans. Signal Processing*, vol. 47, pp. 187–200, 1999.

[13] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Information Theory*, vol. 52, no. 1, pp. 6–18, 2006.

[14] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signal," *IEEE Trans. Info. Theory*, vol. 51, no. 3, pp. 1030–1051, 2006.

[15] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.

[16] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Machine Learning Res.*, vol. 1, pp. 211–244, 2001. (software available at `http://www.miketipping.com/index.php?page=rvm`).

[17] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selectionn," *IEEE Trans. Signal Process.*, vol. 52, pp. 2153 – 2164, 2004.

[18] E. Larsson and Y. Selén, "Linear regression with a sparse parameter vector," *IEEE Trans. Signal Process.*, vol. 55, pp. 451 – 460, 2007.

[19] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer, 2nd ed., 1994.

[20] D. L. Donoho, Y. Tsaig, I. Drori, and J.-C. Starck, "Sparse solution of underdetermined linear equations by stagewise orthogonal matching pursuit," Tech. Rep. 2006-02, Dept. of Statistics, Stanford University, Stanford, CA, 2006. (software available at `http://sparselab.stanford.edu/`).

[21] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007. (software available at `http://www.lx.it.pt/~mtf/GPSR/`).

[22] S. Ji and L. Carin, "Bayesian compressive sensing and projection optimization," in *Proc. 24th Intern. Conf. Machine Learning (ICML)*, 2007. (software available at `http://www.ece.duke.edu/~shji/BCS.html`).

[23] P. Kuppusamy, "EPR spectroscopy in biology and medicine," *Antioxid. Redox Signa*, vol. 6, pp. 583–585, 2004.