Sparse Reconstruction for RADAR

Lee C. Potter, Philip Schniter, and Justin Ziniel

Dept. of Electrical & Computer Engineering, The Ohio State University, Columbus, OH, US

ABSTRACT

Imaging is not itself a system goal, but is rather a means to support inference tasks. For data processing with linearized signal models, we seek to report all high-probability interpretations of the data and to report confidence labels in the form of posterior probabilities. A low-complexity recursive procedure is presented for Bayesian estimation in linear regression models. A Gaussian mixture is chosen as the prior on the unknown parameter vector. The algorithm returns both a set of high posterior probability mixing parameters and an approximate minimum mean squared error (MMSE) estimate of the parameter vector. Emphasis is given to the case of a sparse parameter vector. Numerical simulations demonstrate estimation performance and illustrate the distinctions between MMSE estimation and maximum *a posteriori* probability (MAP) model selection. The proposed tree-search algorithm provides exact ratios of posterior probabilities for a set of high probability solutions to the sparse reconstruction problem. These relative probabilities serve to reveal potential ambiguity among multiple candidate solutions that are ambiguous due to low signal-to-noise ratio and/or significant correlation among columns in the super-resolving regressor matrix.

Keywords: sparse reconstruction, radar imaging, Gaussian mixture models

1. INTRODUCTION

1.1 High frequency scattering

At high frequencies, the scattering response of an object can be well approximated as a sum of responses from individual scattering centers.[1] These scatters provide a concise, yet physically relevant, description of the object. Additionally, a parametric representation of the sum of reflectors offers a low dimensional manifold for inversion and compression of phase history data. For example, an approximate physical optics model for a set of canonical reflectors[2,3] parameterizes responses as a functions of frequency, aspect, elevation and polarization. Fitting the models to phase history data is a nonlinear regression task, with the accompanying challenges of computational cost, model selection (order selection and reflector type detection), and local minima in the non-convex optimization. Yet, the nonlinear regression can provide a compressed, informative, super-resolved interpretation of the scattering data. For example, Fig. 1 shows a graphical representation of parameters estimated for vehicle in the 640 MHz bandwidth Gotcha public release data.[4] The HH and VV polarization images from eight circular passes are jointly and coherently processed by automated model fitting. Estimated parameters for 3D position, orientation, length and polarization (even versus odd bounce) are displayed as icons in the figure. For visualization, a CAD model of the actual vehicle is superimposed in the figure.[5] The eight passes provide only a sparse sampling across less than 2 degrees elevation, and hence present high side lobes and aliasing in backprojection or Fourier 3D imaging.

1.2 Sparse linear regression

An alternative to the nonlinear regression is to recast the inverse scattering problem as a linear regression. The unknown coefficients are taken as the reflections coefficients at discretely sampled 3D locations, and a regressor matrix, $\boldsymbol{A}$, is typically formed by sampling some parametric model (such as an ideal point scattering model). Thus, the linear model is given by

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

Send correspondence to LCP, potter.36@osu.edu.

Figure 1. Nonlinear regression yields a parsimonious interpretation of 8-pass circular SAR data from the Gotcha Public Release. Icons depict the estimated parameters describing 3D position, orientation, length and polarization (even versus odd bounce) for individual scattering centers. A CAD model is superimposed for visualization.

with unknown parameter vector $\boldsymbol{x}$, unit norm columns in the regressor matrix $\boldsymbol{A}$, and additive noise $\boldsymbol{\nu}$. The high-frequency approximation motivates a preference towards a sparse solution in which few coefficients in $\boldsymbol{x}$ are non-zero.

Sparse linear regression is a topic of long standing interest in statistics, geophysics, and signal processing. Algorithmic approaches have been proposed over several decades. One class of algorithms adopts a greedy search heuristic. Examples include CLEAN,[6] projection pursuit,[7] and orthogonal matched pursuit (OMP).[8] Tropp and Gilbert,[9] for example, provide sufficient conditions on the sparseness 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. A second class of recursive algorithms solves a sequence of iteratively re-weighted linear least-squares (IRLS) problems;[10,11] recent results have likewise established sufficient conditions such that the sequence converges to the sparsest solution for a given residual energy.[12,13]

The method of sparse Bayesian learning[14,15] 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 addition to greedy, IRLS and Bayesian learning approaches, penalized least-squares solutions for $\boldsymbol{x}$ have likewise been presented in the past four decades.[16] In this class of approaches, parameters are found via the optimization

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

or, equivalently for some $\epsilon > 0$

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

Ridge regression[17] (Tikhonov regularization) adopts $p = 2$, while basis pursuit[18] and Lasso[19] use $p = 1$. Equation (2) has been widely adopted, for example in radar imaging,[20] image reconstruction,[21,22] and elsewhere.[23,24] The large class of methods adopting (2) may be interpreted as implicitly seeking the Bayesian MAP estimate of x under a sparsity inducing prior

$$p(\boldsymbol{x}) \quad \propto \quad \exp\left\{ -\tfrac{\lambda}{2}\|\boldsymbol{x}\|_p^p \right\}. \tag{4}$$

Elegant recent results by several authors[25–27] 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

$$\min \|\boldsymbol{x}\|_0 \ \text{ s.t. } \ \|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{y}\|_2^2 < \epsilon. \tag{5}$$

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 sufficient conditions restrict the sparseness of $\boldsymbol{x}$ and the correlation among the columns in the regressor matrix, $\boldsymbol{A}$. Example conditions are the restricted isometry property[27] (RIP) and the mutual coherence.[26] The low correlation required among columns of $\boldsymbol{A}$ is incompatible with super-resolution at any SNR.

The constructive design of experiments consistent with RIP (design of $\boldsymbol{A}$) remains open.[28] But the compressive sensing hypotheses are met with high probability by draws from classes of random matrices. In this sense, compressive sensing trades the NP-hard ℓ_0 sparsest solution task for an intractable experiment design, then uses randomization for experiment design. Randomization has been used for over 40 years in array processing for low side-lobe responses.[29,30] Thus, compressive sensing theorems offer an invitation to randomized sampling.

For radar imaging, traditional phase history sampling and point scattering models result in an $\boldsymbol{A}$ matrix that has high correlation among columns; for example, the maximum correlation between columns of $\boldsymbol{A}$ is 0.38 for MSTAR image chips. And, the aperture randomization used in magnetic resonance k-space sampling[30] would require random transport of an antenna across the full viewing hemisphere. For temporal sampling, phase jitter and sample-hold bandwidth limit SNR for randomized sampling. Thus, application of the randomization concept to radar imaging remains an open opportunity and challenge.

In the sparse reconstruction and compressive sensing 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 addition, an estimate of $\boldsymbol{x}$ is also sought. Further, a single solution is reported, and typically without a report of uncertainty.

1.3 A third way

For data processing with linearized signal models, we adopt an alternative approach that embraces ambiguity among competing sparse interpretations of the noisy data. Linearization offers the potential to avoid the model order and computational difficulties of nonlinear regression. Our approach is motivated by the reality that reconstruction of $\boldsymbol{x}$ (i.e., imaging) is not itself a system goal, but is rather a means to support inference tasks. Therefore, we seek to report all high-probability interpretations of the data, and to report confidence labels in the form of posterior probabilities. An analogy is found in digital communications, where capacity-achieving codes require soft-decisions on bits; the individual bits are not the goals, but are rather probabilistic evidence for decoding the message. Thus, we present a detection procedure that provides exact ratios of posterior probabilities for a set of high probability solutions to the sparse reconstruction problem. These relative probabilities serve to reveal uncertainty among multiple candidate solutions that are ambiguous due to low signal-to-noise ratio and/or significant correlation among columns in the (super-resolved) regressor matrix, $\boldsymbol{A}$. An additional output from the approach is a minimum mean-squared error (MMSE) estimator for $\boldsymbol{x}$ from the noisy observations, $\boldsymbol{y}$. The MMSE estimation approach was likewise adopted in a crisp exposition by Larsson and Selén;[31] as they noted, the Bayesian formulation requires *a priori* assumptions that are explicitly stated and admit precise interpretation. (We specifically identify similarities to[31] in Section 5.)

The remainder of the paper is organized as follows. In Section 2, we state the signal model and explicitly identify the assumed priors. In Section 3, we describe our proposed technique, and in Section 4 we investigate its performance numerically. In Section 5, we discuss our findings, and in Section 6 we conclude.

2. SIGNAL MODEL

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

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

where the noise $\boldsymbol{\nu}$ is assumed to be white circular Gaussian with variance σ^2, i.e., $\boldsymbol{\nu} \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_M)$, 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 sparseness, we assume that the parameters are generated from a circular Gaussian mixture density:

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

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 $x_n|s_n \sim \mathcal{CN}(0, \sigma^2_{s_n})$. Also for simplicity, we assume that the mixture parameters $\{s_n\}_{n=0}^{N-1}$ are* Bernoulli(p_1). To model sparseness in $\boldsymbol{x}$, we choose $\sigma_0^2 = 0$ and $p_1 \ll 1$.

From the model assumptions it can be seen that

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

where

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

3. 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 sparse parameters $\boldsymbol{x}$.

3.1 Basis Selection Metric

As a consequence of the model described in Section 2, the locations of the nonzero elements in $\boldsymbol{s}$ specify the set of "active" basis elements, i.e., columns of $\boldsymbol{A}$. 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 probable, but also *how probable* 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

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

where $\mathcal{S} = \{0,1\}^N$, which shows that the estimation of $\{p(\boldsymbol{s}|\boldsymbol{y})\}_{\boldsymbol{s}\in\mathcal{S}}$ reduces to the estimation of $\{p(\boldsymbol{y}|\boldsymbol{s})p(\boldsymbol{s})\}_{\boldsymbol{s}\in\mathcal{S}}$. Although 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 $\boldsymbol{s} \in \mathcal{S}$, the subset $\mathcal{S}_\star \subset \mathcal{S}$ that includes the *dominant* posteriors can be quite small and therefore practical to compute. Working in the log domain, we find

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

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

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

3.2 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

$$\hat{\boldsymbol{x}}_{\mathsf{mmse}} \;:=\; \mathrm{E}\{\boldsymbol{x}|\boldsymbol{y}\} \;=\; \sum_{\boldsymbol{s}\in\mathcal{S}} p(\boldsymbol{s}|\boldsymbol{y})\,\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\} \tag{15}$$

where from (8) it is straightforward to obtain

$$\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\} \;=\; \boldsymbol{R}(\boldsymbol{s})\boldsymbol{A}^H\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{y}. \tag{16}$$

Approximating (15) with a summation over only the dominant basis configurations $\mathcal{S}_\star$ yields

$$\hat{\boldsymbol{x}}_{\mathsf{ammse}} \;:=\; \sum_{\boldsymbol{s}\in\mathcal{S}_\star} p(\boldsymbol{s}|\boldsymbol{y})\,\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\}. \tag{17}$$

The posterior probability is a Gaussian mixture, and hence possibly mutli-modal. The second moment is the conditional covariance $\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y}\}$ and can be closely approximated as

$$\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y}\} \;\approx\; \sum_{\boldsymbol{s}\in\mathcal{S}_\star} p(\boldsymbol{s}|\boldsymbol{y})\left\{\mathrm{Cov}\left\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\right\} + (\hat{\boldsymbol{x}}_{\mathsf{ammse}}-\hat{\boldsymbol{x}}_{\mathsf{sy}})(\hat{\boldsymbol{x}}_{\mathsf{ammse}}-\hat{\boldsymbol{x}}_{\mathsf{sy}})^H\right\} \tag{18}$$

$$\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\} \;=\; \boldsymbol{R}(\boldsymbol{s}) - \boldsymbol{R}(\boldsymbol{s})\boldsymbol{A}^H\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{A}\boldsymbol{R}(\boldsymbol{s}). \tag{19}$$

Here, $\hat{\boldsymbol{x}}_{\mathsf{sy}} = \mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\}$ from Eqn. 16. Note that, in evaluating (15) and (18), 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 $\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\}$ and $\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}\}$ for each of the configurations $\boldsymbol{s}$ identified by the search.

3.3 Bayesian Matching Pursuit

We next consider the estimation of the set of dominant basis configurations $\mathcal{S}_\star$. Here we present a prosaic description of the proposed search heuristic. Later, in Section 3.5, we detail a fast algorithm.

The proposed 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 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 unique 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 so that[†] $\Pr(\|\boldsymbol{s}\|_0 > P)$ is sufficiently small.[‡] The algorithm's final estimate of $\mathcal{S}_\star$ is then $\hat{\mathcal{S}}_\star := \bigcup_{i=0}^{P}\mathcal{S}_\star^{(i)}$.

The proposed algorithm, which we refer to as Bayesian matching pursuit (BMP), can be interpreted as an application of Cotter and Rao's "MP:K" algorithm[32] to maximization of the Bayesian metric $\mu(\boldsymbol{s})$.

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

[‡]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 $\Pr(\|\boldsymbol{s}\|_0 > P) = \frac{1}{2}\,\mathrm{erfc}\left(\frac{P-Np_1}{\sqrt{2Np_1(1-p_1)}}\right)$.

3.4 Fast Metric Update

To accelerate the aforementioned BMP algorithm, we propose a fast metric update which computes the change in $\mu(\cdot)$ that results from the activation of a single basis element. More precisely, if we denote by $\boldsymbol{s}_n$ the vector identical to $\boldsymbol{s}$ except for the n^{th} element, 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

$$\mu(\boldsymbol{0}) \;=\; -N\ln\pi - M\ln\sigma^2 - \tfrac{1}{\sigma^2}\|\boldsymbol{y}\|_2^2 + N\ln(1-p_1) \tag{20}$$

via (14) 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 $\boldsymbol{s}$,

$$\boldsymbol{\Phi}(\boldsymbol{s}_n) \;=\; \boldsymbol{\Phi}(\boldsymbol{s}) + \sigma_1^2\boldsymbol{a}_n\boldsymbol{a}_n^H, \tag{21}$$

from which the matrix inversion lemma implies

$$\begin{aligned}
\boldsymbol{\Phi}(\boldsymbol{s}_n)^{-1} &= \boldsymbol{\Phi}(\boldsymbol{s})^{-1} - \sigma_1^2\beta_n\boldsymbol{b}_n\boldsymbol{b}_n^H && (22)\\
\boldsymbol{b}_n &:= \boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{a}_n && (23)\\
\beta_n &:= \left(1+\sigma_1^2\boldsymbol{a}_n^H\boldsymbol{b}_n\right)^{-1}. && (24)
\end{aligned}$$

Equations (21)-(24) then imply

$$\begin{aligned}
\boldsymbol{y}^H\boldsymbol{\Phi}(\boldsymbol{s}_n)^{-1}\boldsymbol{y} &= \boldsymbol{y}^H\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{y} - \sigma_1^2\beta_n|\boldsymbol{y}^H\boldsymbol{b}_n|^2 && (25)\\
\ln\det(\boldsymbol{\Phi}(\boldsymbol{s}_n)) &= \ln\det\left(\boldsymbol{\Phi}(\boldsymbol{s}) + \sigma_1^2\boldsymbol{a}_n\boldsymbol{a}_n^H\right) && (26)\\
&= \ln\left[\left(1+\sigma_1^2\boldsymbol{a}_n^H\boldsymbol{\Phi}(\boldsymbol{s})^{-1}\boldsymbol{a}_n\right)\det\left(\boldsymbol{\Phi}(\boldsymbol{s})\right)\right]\\
&= \ln\det(\boldsymbol{\Phi}(\boldsymbol{s})) - \ln\beta_n && (27)\\
\|\boldsymbol{s}_n\|_0\ln\tfrac{p_1}{1-p_1} &= (\|\boldsymbol{s}\|_0+1)\ln\tfrac{p_1}{1-p_1} && (28)\\
&= \|\boldsymbol{s}\|_0\ln\tfrac{p_1}{1-p_1} + \ln\tfrac{p_1}{1-p_1}, && (29)
\end{aligned}$$

which, in conjunction with (14), yield

$$\mu(\boldsymbol{s}_n) \;=\; \mu(\boldsymbol{s}) + \underbrace{\ln\beta_n + \sigma_1^2\beta_n|\boldsymbol{y}^H\boldsymbol{b}_n|^2 + \ln\tfrac{p_1}{1-p_1}}_{\Delta_n(\boldsymbol{s})}. \tag{30}$$

In summary, $\Delta_n(\boldsymbol{s})$ in (30) quantifies the change in our basis selection metric $\mu(\cdot)$ due to the activation of the n^{th} tap of $\boldsymbol{s}$.

3.5 Fast Bayesian Matching Pursuit

Notice that the parameters $\{\beta_n\}_{n=0}^{N-1}$, which are essential for the metric exploration step (30), require $\mathcal{O}(NM^2)$ operations to compute using (23)-(24) with standard matrix multiplication. As described next, the structure of $\boldsymbol{\Phi}(\boldsymbol{s})^{-1}$ can be exploited to make this complexity in M.

Consider the $(i+1)^{th}$ stage of the sequential search described in Section 3.3, where the task is to evaluate the metrics corresponding to all valid single-tap extensions of a surviving hypothesis $\boldsymbol{s}^{(i)}$, where $\|\boldsymbol{s}^{(i)}\| = i$. Let us denote the (complete) set of these extensions by $\{\boldsymbol{s}_n^{(i+1)}\}_{n=0}^{N-1}$, remembering that particular extensions (corresponding, e.g., to re-activation of already-active taps) may be invalid. Say that, at the i^{th} stage, the quantities $\{\boldsymbol{b}_n^{(i)}\}_{n=0}^{N-1}$ and $\{\beta_n^{(i)}\}_{n=0}^{N-1}$ were computed in order to estimate the metrics $\{\mu(\boldsymbol{s}_n^{(i)})\}_{n=0}^{N-1}$ via (30), from which the surviving configuration $\boldsymbol{s}^{(i)}$ was chosen as

$$\boldsymbol{s}^{(i)} = \boldsymbol{s}_{n_\star}^{(i)} \;\text{ for }\; n_\star = \operatorname*{argmax}_n \mu(\boldsymbol{s}_n^{(i)}).$$

Then, to generate $\{\boldsymbol{b}_n^{(i+1)}\}_{n=0}^{N-1}$ and $\{\beta_n^{(i+1)}\}_{n=0}^{N-1}$, from which the metrics $\{\mu(\boldsymbol{s}_n^{(i+1)})\}_{n=0}^{N-1}$ can be evaluated using (30), equations (21) and (23) suggest

$$\begin{aligned}
\boldsymbol{b}_n^{(i+1)} &= \boldsymbol{\Phi}(\boldsymbol{s}^{(i)})^{-1}\boldsymbol{a}_n && (31)\\
&= \left[\boldsymbol{\Phi}(\boldsymbol{s}^{(i-1)})^{-1} - \sigma_1^2\beta_{n_\star}^{(i)}\boldsymbol{b}_{n_\star}^{(i)}\boldsymbol{b}_{n_\star}^{(i)H}\right]\boldsymbol{a}_n && (32)\\
&= \boldsymbol{b}_n^{(i)} - \sigma_1^2\beta_{n_\star}^{(i)}\boldsymbol{b}_{n_\star}^{(i)}\boldsymbol{b}_{n_\star}^{(i)H}\boldsymbol{a}_n && (33)\\
\beta_n^{(i+1)} &= \left(1+\sigma_1^2\boldsymbol{a}_n^H\boldsymbol{b}_n^{(i+1)}\right)^{-1}, && (34)
\end{aligned}$$

which requires $\mathcal{O}(NM)$ operations in total.

Going further, if we define $\boldsymbol{B}^{(i+1)} := [\boldsymbol{b}_0^{(i+1)},\ldots,\boldsymbol{b}_{N-1}^{(i+1)}]$ and notice that $\boldsymbol{B}^{(i+1)} = \boldsymbol{\Phi}(\boldsymbol{s}^{(i)})^{-1}\boldsymbol{A}$, then we can compute the $\boldsymbol{s}^{(i)}$-conditional MMSE estimate, and its covariance, as

$$\begin{aligned}
\mathrm{E}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}^{(i)}\} &= \boldsymbol{R}(\boldsymbol{s}^{(i)})\boldsymbol{B}^{(i+1)H}\boldsymbol{y} && (35)\\
\mathrm{Cov}\{\boldsymbol{x}|\boldsymbol{y},\boldsymbol{s}^{(i)}\} &= \left[\boldsymbol{I}_N - \boldsymbol{R}(\boldsymbol{s}^{(i)})\boldsymbol{B}^{(i+1)H}\boldsymbol{A}\right]\boldsymbol{R}(\boldsymbol{s}^{(i)}), && (36)
\end{aligned}$$

using (15), (18), and the fact that $\boldsymbol{\Phi}(\boldsymbol{s}^{(i)})$ is Hermitian. Because $\boldsymbol{R}(\boldsymbol{s}^{(i)})\boldsymbol{B}^{(i+1)H}$ has only i nonzero rows and $\boldsymbol{A}\boldsymbol{R}(\boldsymbol{s}^{(i)})$ has only i nonzero columns, (35) and (36) can be computed using only $\mathcal{O}(iN)$ and $\mathcal{O}(i^2N)$ multiplications, respectively.

The recursive metric update (30), via recursive computation of $\{\boldsymbol{b}_n^{(i+1)}\}_{n=0}^{N-1}$ from (33), forms the foundation of the Fast Bayesian Matching Pursuit (FBMP) algorithm detailed in Table 1. From the table, it is straightforward to verify that the number of multiplications FBMP required to compute the set of dominant basis configurations $\hat{\mathcal{S}}_\star$, as well as the corresponding conditional means and covariances, is $\mathcal{O}(NMPD)$.

4. NUMERICAL EXPERIMENTS

4.1 FBMP Behavior

Numerical experiments were conducted to investigate the performance of FBMP from Table 1 for various values of model and algorithmic parameters, and the results are reported in Figs. 2-4. 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 3.3. 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 $\mathrm{E}\{\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2^2/\|\boldsymbol{x}\|_2^2\}$.

Sparse reconstruction with Gaussian ensembles asymptotically requires measurements numbering[33]

$$M > 2K\log(N-K) + K + 1 \qquad (37)$$

Here, we have $K = p_1 N$ is the (average) number of active coefficients. We note that the approximate inequality $M > 2K\log(N/K)$ is satisfied for $2p_1N/M < -1/\log(p_1)$. Our nominal parameters yield $2p_1N/M = 0.32$ and $-1/\log(p_1) = 0.31$, and below we see threshold behavior consistent with the asymptotic results.

In Fig. 2(a), 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 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. 2(a) 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.

$\mu^{(0,1)} = -N \ln 2\pi - M \ln \sigma^2 - \frac{1}{\sigma^2}\|\boldsymbol{y}\|_2^2 + N \ln(1-p_1)$;
$\boldsymbol{n}^{(0,0)} = [\,]$;
for $n = 0 : N-1$,
  $\boldsymbol{b}_n^{(1,1)} = \frac{1}{\sigma^2}\boldsymbol{a}_n$;
  $\beta_n = (1 + \sigma_1^2 \boldsymbol{a}_n^H \boldsymbol{b}_n^{(1,1)})^{-1}$;
  $\mu_n = \mu^{(0,1)} + \log \beta_n + \sigma_1^2 \beta_n |\boldsymbol{y}^H \boldsymbol{b}_n^{(1,1)}|^2 + \log \frac{p_1}{1-p_1}$;
end
for $q = 1 : D$,
  $n_\star = n$ indexing q^{th} largest element in $\{\mu_n\}$;
  $\mu^{(1,q)} = \mu_{n_\star}$;
  $\boldsymbol{n}^{(1,q)} = n_\star$;
  for $n = 0 : N-1$,
    $\boldsymbol{b}_n^{(2,q)} = \boldsymbol{b}_n^{(1,1)} + \sigma_1^2 \beta_{n_\star} \boldsymbol{b}_{n_\star}^{(1,1)} \boldsymbol{b}_{n_\star}^{(1,1)H} \boldsymbol{a}_n$;
  end
end
for $i = 2 : P$,
  for $d = 1 : D$,
    for $n = 0 : N-1$,
      $\beta_{d,n} = (1 + \sigma_1^2 \boldsymbol{a}_n^H \boldsymbol{b}_n^{(i,d)})^{-1}$;
      $\mu_{d,n} = \mu^{(i-1,d)} + \log \beta_{d,n} + \sigma_1^2 \beta_{d,n} |\boldsymbol{y}^H \boldsymbol{b}_n^{(i,d)}|^2 + \log \frac{p_1}{1-p_1}$;
      if $n \in \boldsymbol{n}^{(i-1,d)}$ then $\mu_{d,n} = -\infty$;
    end
  end
  for $q = 1 : D$,
    $(d_\star, n_\star) = (d, n)$ indexing q^{th} largest unique element in $\{\mu_{d,n}\}$;
    $\mu^{(i,q)} = \mu_{d_\star, n_\star}$;
    $\boldsymbol{n}^{(i,q)} = [\boldsymbol{n}^{(i,d_\star)}, n_\star]$;
    for $n = 0 : N-1$,
      $\boldsymbol{b}_n^{(i+1,q)} = \boldsymbol{b}_n^{(i,d_\star)} + \sigma_1^2 \beta_{d_\star,n_\star} \boldsymbol{b}_{n_\star}^{(i,d_\star)} \boldsymbol{b}_{n_\star}^{(i,d_\star)H} \boldsymbol{a}_n$;
    end
    $\hat{\boldsymbol{x}}^{(i,q)} = \sigma_1^2 \sum_{j=1}^{i} \boldsymbol{\delta}_{[\boldsymbol{n}^{(i,q)}]_j} \boldsymbol{b}_{[\boldsymbol{n}^{(i,q)}]_j}^{(i+1,q)T} \boldsymbol{y}$;
    $\hat{\boldsymbol{\Sigma}}^{(i,q)} = \sigma_1^2 \sum_{j=1}^{i} \sum_{k=1}^{i} \boldsymbol{\delta}_{[\boldsymbol{n}^{(i,q)}]_j} \big[\delta_{[\boldsymbol{n}^{(i,q)}]_j - [\boldsymbol{n}^{(i,q)}]_k}$
      $-\sigma_1^2 \boldsymbol{b}_{[\boldsymbol{n}^{(i,q)}]_j}^{(i+1,q)T} \boldsymbol{a}_{[\boldsymbol{n}^{(i,q)}]_k}\big] \boldsymbol{\delta}_{[\boldsymbol{n}^{(i,q)}]_k}^T$;
  end
end

Table 1. Fast Bayesian Matching Pursuit

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

$$\hat{\boldsymbol{s}}_{\mathsf{map}} \;:=\; \underset{s\in\hat{\mathcal{S}}_\star}{\operatorname{argmax}}\, p(\boldsymbol{s}|\boldsymbol{y}). \tag{38}$$

In particular, the traces in Fig. 2(b) show number-of-misses versus observation length M for FBMP under several values of search parameter D. Because the number-of-misses closely parallel the NMSE values, 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(a), 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 Fig. 2. As in Fig. 2, 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. 3(b), 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). Fig. 3(b) 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. 4(a), we plot NMSE versus SNR for two FBMP-derived estimates: the (approximate) MMSE estimate $\hat{\boldsymbol{x}}_{\mathsf{ammse}}$ from (17) and the quasi-MAP estimate $\hat{\boldsymbol{x}}_{\mathsf{amap}}$ from (39):

$$\hat{\boldsymbol{x}}_{\mathsf{amap}} \;:=\; \mathrm{E}\{\boldsymbol{x}|\boldsymbol{y}, \hat{\boldsymbol{s}}_{\mathsf{map}}\}. \tag{39}$$

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. 4(a) demonstrates that $\hat{\boldsymbol{x}}_{\mathsf{ammse}}$ are about 1 dB better than $\hat{\boldsymbol{x}}_{\mathsf{map}}$ at $\mathsf{SNR} \leq 10$ dB and about 0.5 dB better at $\mathsf{SNR} > 10$ dB. The improvement reflects the advantage of allowing for *uncertainty* in the estimated basis.

Finally, in Fig. 4(b), we plot average FBMP runtime versus search parameter D. As expected from the algorithmic description in Table 1, the runtime scales linearly in D.

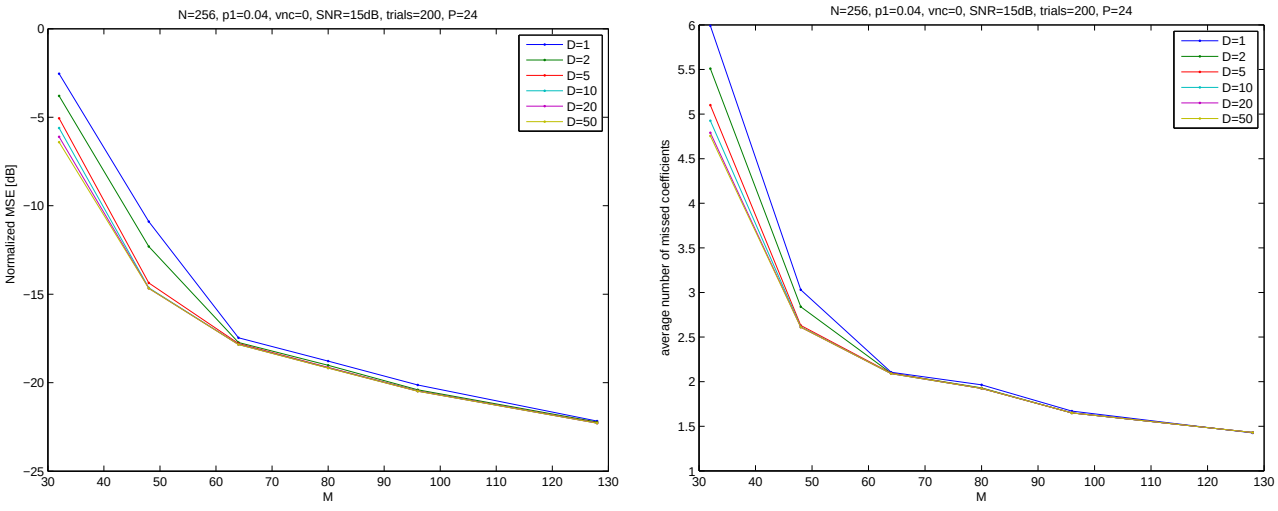

Figure 2. (a) Normalized MSE for FBMP versus observation length M. (b) Number of active coefficients in $\boldsymbol{x}$ missing from FBMP's estimate of the most probable basis configuration versus observation length M. In each subfigure, from top to bottom, the traces reflect increasing D. (See the graph title for configuration.)

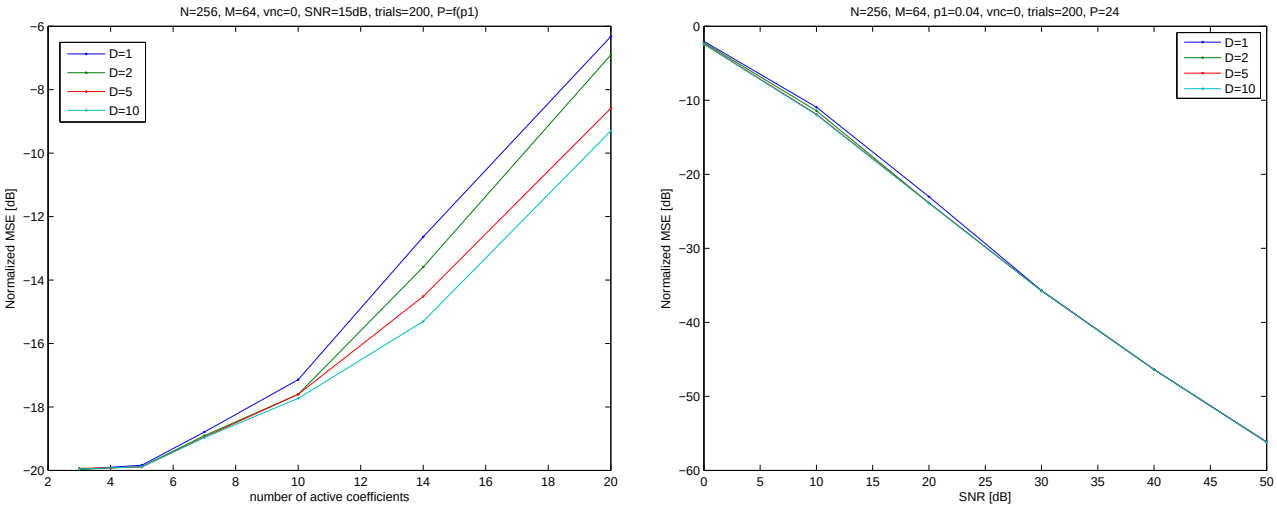

Figure 3. (a) Normalized MSE versus average number of active coefficients (i.e., $p_1 N$) for FBMP. (b) Normalized MSE versus SNR for FBMP. In each subfigure, from top to bottom, the traces reflect increasing D. (See the graph title for configuration.)

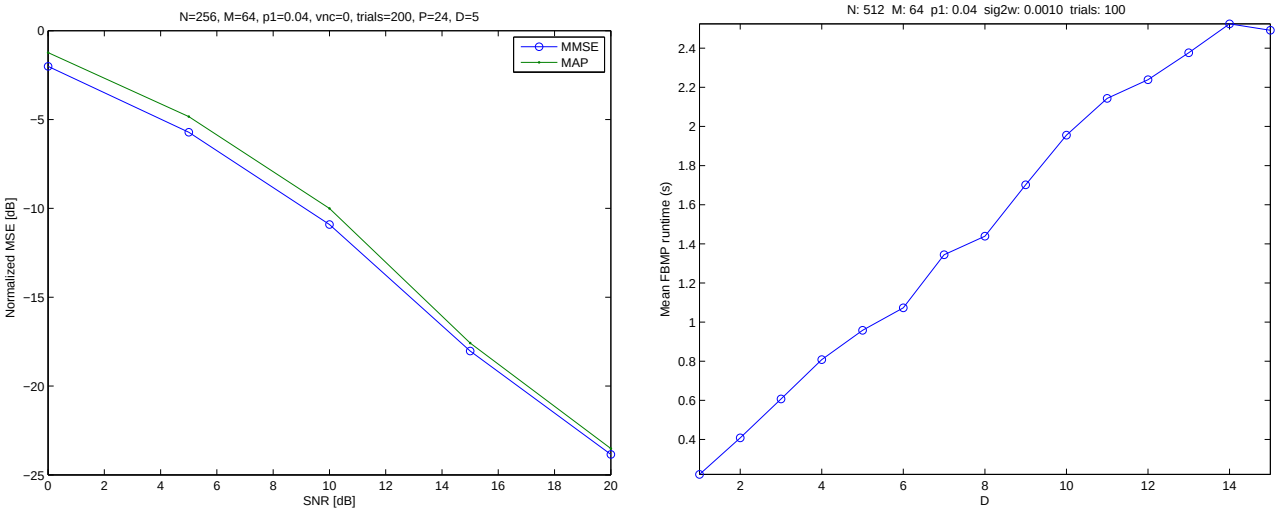

Figure 4. (a) Normalized MSE versus SNR of FBMP-returned MMSE estimate $\hat{\boldsymbol{x}}_{\mathsf{ammse}}$ and quasi-MAP estimate $\hat{\boldsymbol{x}}_{\mathsf{amap}}$. (b) Average FBMP runtime versus search parameter D. (See the graph title for configuration.)

4.2 Comparison To Other Algorithms

In Figs. 5-7 we compare FBMP to several other popular sparse estimation algorithms, including SparseBayes,[14] OMP,[9] StOMP,[34] GPSR-Basic,[35] and BCS.[36] 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 4.1. 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.

In Fig. 5 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 for $M > 64$. In particular, for $M > 85$ it outperformed BCS by approximately 4 dB and outperformed OMP by 6 dB. In Fig. 6 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 > 5 dB; and, at SNR $= 3$ dB, FBMP outperformed GPSR by approximately 1 dB and the other algorithms by > 5 dB.

Finally, in Fig. 7, we plot average runtime versus observation length M for the various sparse estimation algorithms. For FBMP, we used $D = 1$. Fig. 7 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.

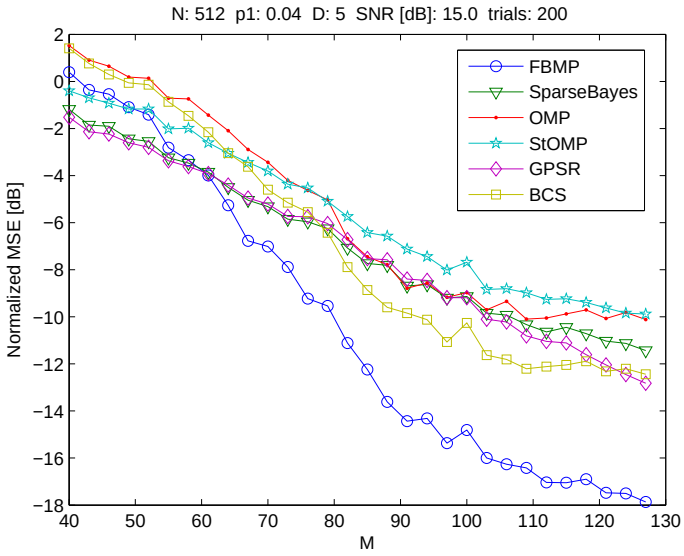

Figure 5. Normalized MSE versus observation length M for several algorithms. (See the graph title for configuration.)

5. 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 $\hat{\boldsymbol{x}}_{\mathsf{mmse}}$ through (15). Specifically, from (10), we can write the posterior probability of basis configuration $\boldsymbol{s}$ as

$$p(\boldsymbol{s}|\boldsymbol{y}) \quad = \quad \frac{\exp\{\mu(\boldsymbol{s})\}}{\sum_{\boldsymbol{s}' \in \mathcal{S}} \exp\{\mu(\boldsymbol{s}')\}} \approx \frac{\exp\{\mu(\boldsymbol{s})\}}{\sum_{\boldsymbol{s}' \in \mathcal{S}_\star} \exp\{\mu(\boldsymbol{s}')\}}, \tag{40}$$

where the approximation in (40) includes only the basis configurations $\mathcal{S}_\star \subset \mathcal{S}$ that account for the dominant values of $\exp\{\mu(\boldsymbol{s})\}$. Likewise, $p(\boldsymbol{x}|y)$ provides an approximate density function describing the uncertainty in

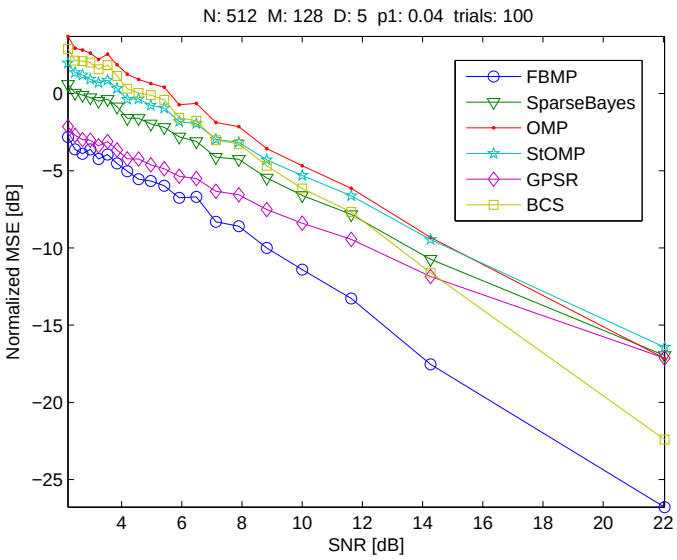

Figure 6. Normalized MSE versus SNR for several algorithms. (See the graph title for configuration.)

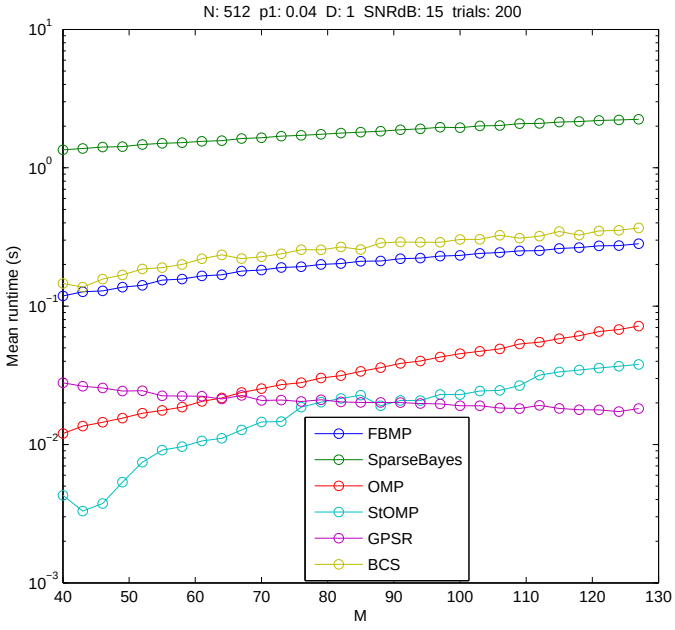

Figure 7. Average runtime versus observation length M for several algorithms. (See the graph title for configuration.)

resolving $\boldsymbol{x}$ with the noisy observations, $\boldsymbol{y}$

$$p(\boldsymbol{x}|\boldsymbol{y}) \approx \sum_{\boldsymbol{s}' \in \mathcal{S}_\star} p(\boldsymbol{s}|\boldsymbol{y}) p(\boldsymbol{x}|\boldsymbol{y}, \boldsymbol{s}) \tag{41}$$

The posterior density function reflects the multi-modal 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.

In contrast, confidence labels for estimated $\hat{\boldsymbol{x}}$ are largely absent in the compressive sensing and sparse linear regression literature. Exceptions are found in[19, 36] which give the error covariance for the simple linear problem conditioned on *perfect knowledge of the active basis elements*. As noted by Tibshirani,[19] 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" $s_i = 0$. The Gaussian confidence intervals in[19, 36] are constructed by assuming that $\hat{\boldsymbol{s}}_{\mathsf{amap}}$ is the true sequence of active coefficients. However, in our simulations the true $\boldsymbol{s}$ typically does not rank first in $p(\boldsymbol{s}|\boldsymbol{y})$, and instead usually ranks between second and tenth (or lower). 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 2 was likewise adopted by Larsson and Selén,[31] who also constructed the MMSE estimate in the manner of (17) but with an $\mathcal{S}_\star$ that contains exactly one sequence $\boldsymbol{s}$ for each Hamming 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 3.3 recalls the heuristic of,[31] but in reverse. Note, however, that the *fast* $D = 1$ BMP presented in Section 3.5 has a complexity of only $\mathcal{O}(NMP)$, in comparison to $\mathcal{O}(N^3M^2)$ for the technique in.[31] Given the typically large values of N encountered in practice, the complexity of FBMP can be several orders of magnitude lower than that of.[31]

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 (40) 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 $\frac{p_1 N}{M} \leq -1/\log(p_1)$.

Although the model in Section 2 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 3.3 to handle several types of active coefficient (e.g., one for each allowed mean).

6. CONCLUSION

At high frequencies, radar imaging can be viewed as sparse reconstruction from noisy observations on a sparse aperture. On the one hand, nonlinear regression techniques can provide super-resolution and posterior confidence labels, but such techniques may suffer from model selection, model mismatch, and computational cost. On the other hand, sparse linear regression techniques are computationally inexpensive and without model selection difficulties; however, the linear techniques provide neither super-resolution of high SNR reflectors nor posterior confidence labels.

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.

Thus, we presented a detection procedure that provides exact ratios of posterior probabilities for a set of high probability solutions to the sparse reconstruction problem. These relative probabilities serve to reveal potential ambiguity among multiple candidate solutions that are ambiguous due to low signal-to-noise ratio and/or significant correlation among columns in the (super-resolved) regressor matrix. The explicit management

of ambiguity and uncertainty is essential for applications in which the estimated detections, $\hat{\boldsymbol{s}}$, and estimated coefficients, $\hat{\boldsymbol{x}}$, are not final products, but rather are statistics for use in making inference from the noisy observations, $\boldsymbol{y}$.

We plan, in the near future, to extend FBMP to the case where the active coefficients have non-zero means chosen from a finite set according to some prior probabilities. Additionally, we are developing a large-scale version of FBMP in which $M \times M$ matrices are not explicitly maintained and the operator $\boldsymbol{A}$ is only given implicitly. An example of the non-zero mean sparse signal model can be found in electron paramagnetic resonance imaging,[37] 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.

ACKNOWLEDGMENTS

LCP was supported by AFOSR award FA9550-06-1-0324; PS was supported by National Science Foundation Grant 0237037 and the Office of Naval Research grant N00014-07-1-0209. Cited work by S. Sharma was supported by the SET Corporation. A preliminary version of Fast Bayes Matched Pursuits was presented at the Workshop on Information Theory and Applications (ITA), La Jolla, CA, January 2008.

REFERENCES

[1] Keller, J. B., "Geometrical theory of diffraction," *J. Opt. Soc. Amer.*, 116–130 (1962).

[2] Gerry, M. J., Potter, L. C., Gupta, I. J., and van der Merwe, A., "A parametric model for synthetic aperture radar measurements," *IEEE Transactions on Antennas and Propagation* **47**, 1179–1188 (1999).

[3] Jackson, J. A., Rigling, B. D., and Moses, R. L., "Canonical scattering feature models for 3D and bistatic SAR," *IEEE Trans. Aerospace and Electronic Systems* (in review).

[4] Casteel, Jr., C. H., Gorham, L. A., Minardi, M. J., Scarborough, S. M., Naidu, K. D., and Majumder, U. K., "A challenge problem for 2D/3D imaging of targets from a volumetric data set in an urban environment," in *Algorithms for Synthetic Aperture Radar Imagery XIV*, Zelnio, E. G. and Garber, F. D., eds., *Proc. SPIE* **6568**, 65680D (2007).

[5] Sharma, S., *Three Dimensional Parameter Estimation from Sparse, Multipass Synthetic Aperture Radar*, Master's thesis, The Ohio State University (2007).

[6] Hogbom, J., "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astrophys. J. Suppl. Ser* **15**, 417–426 (1974).

[7] Huber, P. J., "Projection pursuit," *The Annals of Statistics* **13**, 435–475 (1985).

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

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

[10] Lawson, C. L., "Contributions to the theory of linear least maximum approximations," PhD dissertation, UCLA (1961).

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

[12] Chartrand and Yin, ICASSP 2008, Las Vegas, NV, to appear.

[13] R. DeVore, "Performance of iterated least squares," IMA Seminar, June 2007.

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

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

[16] Taylor, H. L., Banks, S. C., and McCoy, J. F., "Deconvolution with the ℓ_1 norm," *Geophysics* **44**, 39–52 (1979).

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

[18] Chen, S. S., Donoho, D. L., and Saunders, M. A., "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing* **20**(1), 33–61 (1998).

[19] Tibshirani, R., "Regression shrinkage and selection via the lasso," *J. R. Statist. Soc. B* **58**(1), 267 – 288 (1996).

[20] Çetin, M. and Karl, W. C., "Feature-enhanced synthetic aperture radar image formation based on non-quadratic regularization," *IEEE Trans. Image Process.* **10**, 623–631 (2001).

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

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

[23] Levy, S. and Fullagar, P. K., "Reconstruction of a sparse spike train from a portion of its spectrum and application to high-resolution deconvolution," *Geophysics* **46**(9), 1235–1243 (1981).

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

[25] Donoho, D. L., Elad, M., and Temlyakov, V. N., "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Information Theory* **52**(1), 6–18 (2006).

[26] Tropp, J. A., "Just relax: Convex programming methods for identifying sparse signal," *IEEE Trans. Info. Theory* **51**(3), 1030–1051 (2006).

[27] Candès, E., Romberg, J., and Tao, T., "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics* **59**(8), 1207–1223 (2006).

[28] Tao, T., "Open question: deterministic UUP matrices," http://terrytao.wordpress.com/2007/07/02/open-question-deterministic-uup-matrices (July 2007).

[29] Lo, Y. T., "A mathematical theory of antenna arrays with randomly spaced elements," *IEEE Transactions on Antennas and Propagation* **12**, 257–268 (1964).

[30] Marseille, G. J., de Beer, R., Fuderer, M., Mehlkopf, A. F., and van Ormondt, D., "Nonuniform phase-encode distributions for MRI scan time reduction," *J. Magn. Reson.* **111**, 70–75 (1996).

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

[32] Cotter, S. F. and Rao, B. D., "Application of tree-based searches to matching pursuit," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 3933–3936, Salt Lake City, UT (2001).

[33] Wainwright, M. J., "Sharp thresholds for noisy and high-dimensional recovery of sparsity using ℓ_1-constrained quadratic programming," Tech. Rep. 709, Dept. of Statistics, UC-Berkeley (May 2006).

[34] Donoho, D. L., Tsaig, Y., Drori, I., and Starck, J.-C., "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/`).

[35] Figueiredo, M. A. T., Nowak, R. D., and Wright, S. J., "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing* **1**(4), 586–597 (2007). (software available at `http://www.lx.it.pt/~mtf/GPSR/`).

[36] Ji, S. and Carin, L., "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`).

[37] Kuppusamy, P., "EPR spectroscopy in biology and medicine," *Antioxid. Redox Signal* **6**, 583–585 (2004).