

A FAST ALGORITHM FOR THE BAYESIAN ADAPTIVE LASSO

Athanasios A. Rontogiannis, Konstantinos E. Themelis and Konstantinos D. Koutroumbas

Institute for Space Applications and Remote Sensing,
National Observatory of Athens,
152 36, Penteli, Greece
E-mail: {themelis, tronto, koutroum}@noa.gr

ABSTRACT

This paper presents a novel hierarchical Bayesian model which allows to reconstruct sparse signals using a set of linear measurements corrupted by Gaussian noise. The proposed model can be considered as the Bayesian counterpart of the adaptive lasso criterion. A fast iterative algorithm, which is based on the type-II maximum likelihood methodology, is properly adjusted to conduct Bayesian inference on the unknown model parameters. The performance of the proposed hierarchical Bayesian approach is illustrated on the reconstruction of both sparse synthetic data, as well as real images. Experimental results show the improved performance of the proposed approach, when compared to state-of-the-art Bayesian compressive sensing algorithms.

Index Terms— Bayesian compressive sensing, adaptive lasso, sparse linear regression, hierarchical Bayesian analysis.

1. INTRODUCTION

Compressive sensing (CS) has gained considerable attention in the signal processing community over the recent years. In the standard framework of CS, the goal is to recover a sparse signal $\mathbf{w} \in R^N$ from a set of M linear, noisy measurements

$$\mathbf{y} = \Phi \mathbf{w} + \mathbf{n}, \quad (1)$$

where $\Phi = [\phi_1, \phi_2, \dots, \phi_N]$ is a $M \times N$ design matrix, $M \ll N$, and \mathbf{n} is the additive noise. The existence of fewer measurements than parameters in (1) results in an infinite number of possible solutions for \mathbf{w} . However, utilizing the prior knowledge that \mathbf{w} is sparse, CS theory dictates that in principle, one may recover the exact solution of \mathbf{w} with high probability [1]. Mathematically speaking, under certain conditions, a sparse solution for \mathbf{w} can be recovered by considering the following ℓ_1 norm minimization problem

$$\min \|\mathbf{w}\|_1 \text{ subject to } \|\mathbf{y} - \Phi \mathbf{w}\|_2^2 \leq \epsilon, \quad (2)$$

which retains the sparsity of the solution [1]. This problem, also known as the lasso [2] is a convex optimization problem and can be efficiently solved in polynomial time.

Recently, a number of Bayesian type methods for solving (2) have been proposed, in which the sparsity promoting Laplace distribution is utilized. In [3], a hierarchical Bayes interpretation of the Laplacian prior for \mathbf{w} is introduced and an expectation maximization (EM) algorithm is used to estimate the unknown model parameters. The framework of Bayesian CS is analytically established in [4], and a hierarchical model based on the concept of the relevance vector machine (RVM) [5] is adopted. Then, a fast, sub-optimal marginal likelihood maximization algorithm, proposed in [6], is adjusted to perform Bayesian inference. The same efficient algorithm is also employed and properly adjusted, under a suitable hierarchical formulation of the Laplace prior in [7].

In this paper, we extend the hierarchical Bayesian model of [7, 8] to incorporate an independent Laplace prior for each coefficient of the sparse signal vector \mathbf{w} . Our motivation is to establish the Bayesian analogue of the adaptive lasso criterion, recently proposed in [9]. It is known that the assignment of different penalization weights to different entries of the sparse signal vector \mathbf{w} can lead to a consistent estimator for \mathbf{w} , as opposed to the ℓ_1 minimization of the original lasso [2]. To perform Bayesian inference, the fast technique proposed in [6] is properly adjusted to our model. This is based on a type-II maximum likelihood approach, where a fast, iterative, coordinate-type algorithm is employed for the maximization of the marginal log-likelihood function. Experimental results on simulated data, as well as an image restoration example illustrate the performance improvement offered by the proposed estimator, when compared to other related Bayesian methods.

This paper is organized as follows. The proposed hierarchical Bayesian model is described in Section 2. Section 3 presents the fast type-II maximum likelihood algorithm. Simulation results are presented in Section 4. Finally, conclusions are provided in Section 5.

2. BAYESIAN MODELING

In this section, a novel hierarchical Bayesian model is presented and its analogy to the Bayesian adaptive lasso criterion

is proven. In the following, we assume that the additive noise in (1) is independent of \mathbf{w} and has a zero-mean Gaussian distribution, $\mathbf{n} \sim \mathcal{N}(\mathbf{n}|0, \beta^{-1}\mathbf{I}_M)$. This leads to a multivariate Gaussian likelihood for the measurement \mathbf{y} ,

$$\begin{aligned} p(\mathbf{y}|\mathbf{w}, \beta) &= \mathcal{N}(\mathbf{y}|\Phi\mathbf{w}, \beta^{-1}\mathbf{I}_M) \\ &= (2\pi)^{-\frac{M}{2}} \beta^{\frac{M}{2}} \exp\left[-\frac{\beta}{2}\|\mathbf{y} - \Phi\mathbf{w}\|_2^2\right], \end{aligned} \quad (3)$$

where \mathbf{I}_M is the $M \times M$ identity matrix and $\beta > 0$ is the noise precision parameter.

2.1. The Bayesian adaptive lasso

As noted in [9], the ℓ_1 norm in (2) penalizes more heavily the larger signal components rather than the smaller ones, and as a result, can lead to suboptimal solutions. According to [9], different weighting coefficients for the ℓ_1 penalty can be assigned to different entries of the sparse signal vector \mathbf{w} , in order to improve estimation accuracy. Using Lagrangian arguments and utilizing a weighted ℓ_1 norm, the *adaptive* lasso optimization problem is expressed as

$$\mathbf{w}_{adl\text{asso}} = \arg \min_{\mathbf{w}} \left\{ \|\mathbf{y} - \Phi\mathbf{w}\|_2^2 + \sum_{i=1}^N \mu_i |w_i| \right\}, \quad (4)$$

where $\mu_i > 0$, $i = 1, \dots, N$. Note that, the solutions of the lasso (2) and the adaptive lasso (4) will be both sparse, yet different. Therefore, it is expected that suitable selection of the penalization parameters can result in better estimation performance. In the following, we develop a hierarchical Bayesian model which is analogue to the adaptive lasso criterion, by utilizing N independent Laplace priors.

Assuming independence among the entries w_i 's of \mathbf{w} , as in [4, 7, 8], the prior of \mathbf{w} is expressed as a multivariate zero-mean Gaussian distribution, i.e.,

$$\begin{aligned} p(\mathbf{w}|\boldsymbol{\gamma}) &= \prod_{i=1}^N \mathcal{N}(w_i|0, \gamma_i) \\ &= (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Lambda}|^{\frac{1}{2}} \exp\left[-\frac{1}{2}\mathbf{w}^T \boldsymbol{\Lambda} \mathbf{w}\right] = \mathcal{N}(\mathbf{w}|0, \boldsymbol{\Lambda}^{-1}), \end{aligned} \quad (5)$$

where $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_N]^T$, $\gamma_i \geq 0$ and $\boldsymbol{\Lambda}^{-1} = \text{diag}(\boldsymbol{\gamma})$. Moreover, a conjugate Gamma distribution is assigned to β ,

$$p(\beta; \kappa, \theta) = \Gamma(\beta|\kappa, \theta) = \frac{\theta^\kappa}{\Gamma(\kappa)} \beta^{\kappa-1} \exp[-\theta\beta], \quad (6)$$

An effective way to construct a sparsity-promoting Laplace distribution for \mathbf{w} is to introduce an exponential prior over the variance hyperparameter vector $\boldsymbol{\gamma}$ [3]. In this paper, N

independent exponential distributions are utilized for the hyperparameters γ_i , as in [10], i.e.,

$$\begin{aligned} p(\boldsymbol{\gamma}|\boldsymbol{\lambda}) &= \prod_{i=1}^N p(\gamma_i|\lambda_i) = \prod_{i=1}^N \left[\frac{\lambda_i}{2} \exp\left[-\frac{\lambda_i}{2}\gamma_i\right] \right] \\ &= \left(\frac{1}{2}\right)^N |\boldsymbol{\Psi}| \exp\left[-\frac{1}{2}\sum_{i=1}^N \lambda_i \gamma_i\right], \end{aligned} \quad (7)$$

where $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \dots, \lambda_N]^T$, $\lambda_i > 0$ and $\boldsymbol{\Psi} = \text{diag}(\boldsymbol{\lambda})$. The proposed hierarchical model is completed by assigning a conjugate Gamma prior over the sparsity-controlling hyperparameter vector $\boldsymbol{\lambda}$ given by

$$\begin{aligned} p(\boldsymbol{\lambda}; \rho, \delta) &= \prod_{i=1}^N \frac{\delta^\rho}{\Gamma(\rho)} \lambda_i^{\rho-1} \exp[-\delta\lambda_i] \\ &= \left(\frac{\delta^\rho}{\Gamma(\rho)}\right)^N |\boldsymbol{\Psi}|^{\rho-1} \exp\left[-\delta\sum_{i=1}^N \lambda_i\right], \end{aligned} \quad (8)$$

where ρ and δ are two hyperparameters not treated as random variables, with $\rho \geq 0$ and $\delta \geq 0$. Let us now highlight the relation of the proposed hierarchical Bayesian model defined by (5), (6), (7), and (8), with the adaptive lasso. Utilizing (5), (7) and marginalizing out the parameter $\boldsymbol{\gamma}$, a multivariate Laplace distribution arises as a prior for \mathbf{w} ,

$$\begin{aligned} p(\mathbf{w}|\boldsymbol{\lambda}) &= \prod_{i=1}^N \int p(w_i|\gamma_i) p(\gamma_i|\lambda_i) d\gamma_i \\ &= 2^{-N} |\boldsymbol{\Psi}|^{\frac{1}{2}} \exp\left[-\sum_{i=1}^N \sqrt{\lambda_i} |w_i|\right]. \end{aligned} \quad (9)$$

It is then straightforward to verify that the MAP estimator of \mathbf{w} under the prior in (9), is expressed as

$$\begin{aligned} \mathbf{w}_{\text{MAP}} &= \arg \min_{\mathbf{w}} \{-\log[p(\mathbf{y}|\mathbf{w}, \beta)p(\mathbf{w}|\boldsymbol{\lambda})]\} \\ &= \arg \min_{\mathbf{w}} \left\{ \|\mathbf{y} - \Phi\mathbf{w}\|_2^2 + \frac{2}{\beta} \sum_{i=1}^N \sqrt{\lambda_i} |w_i| \right\}, \end{aligned} \quad (10)$$

i.e., it coincides with the solution of the adaptive lasso for $\mu_i = \frac{2}{\beta} \sqrt{\lambda_i}$.

3. BAYESIAN INFERENCE

As it is common in practice, Bayesian inference is based on the posterior distribution of the model parameters $\mathbf{w}, \beta, \boldsymbol{\gamma}$, and $\boldsymbol{\lambda}$, which can be expressed as

$$p(\mathbf{w}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \beta|\mathbf{y}) = \frac{p(\mathbf{w}, \boldsymbol{\gamma}, \boldsymbol{\lambda}, \beta, \mathbf{y})}{p(\mathbf{y})}. \quad (11)$$

However, it is rather difficult to compute this posterior distribution analytically. One way to perform Bayesian inference is to resort to a type-II maximum likelihood approach, as described below.

3.1. Type-II maximum likelihood approach

In a type-II maximum likelihood approach the objective is to maximize the marginal likelihood, obtained by integrating out \mathbf{w} , with respect to the unknown model parameters. The posterior distribution in (11) can be factored as $p(\mathbf{w}, \gamma, \boldsymbol{\lambda}, \beta | \mathbf{y}) = p(\mathbf{w} | \gamma, \boldsymbol{\lambda}, \beta, \mathbf{y}) p(\gamma, \boldsymbol{\lambda}, \beta | \mathbf{y})$. Then, the distribution $p(\mathbf{w} | \gamma, \boldsymbol{\lambda}, \beta, \mathbf{y})$ is easily shown to be the following multivariate Gaussian,

$$\begin{aligned} p(\mathbf{w} | \mathbf{y}, \gamma, \boldsymbol{\lambda}, \beta) &= \mathcal{N}(\mathbf{w} | \boldsymbol{\mu}, \boldsymbol{\Sigma}) \\ &= (2\pi)^{-\frac{N}{2}} |\boldsymbol{\Sigma}|^{-\frac{1}{2}} \exp \left[-\frac{(\mathbf{w} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{w} - \boldsymbol{\mu})}{2} \right], \end{aligned} \quad (12)$$

with parameters

$$\boldsymbol{\Sigma} = \left[\beta \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \boldsymbol{\Lambda} \right]^{-1} \quad \text{and} \quad \boldsymbol{\mu} = \beta \boldsymbol{\Sigma} \boldsymbol{\Phi}^T \mathbf{y}. \quad (13)$$

In the sequel, the model parameters β, γ , and $\boldsymbol{\lambda}$ are individually selected to maximize their joint posterior distribution $p(\gamma, \boldsymbol{\lambda}, \beta | \mathbf{y})$. From Bayes' law, $p(\gamma, \boldsymbol{\lambda}, \beta | \mathbf{y}) = p(\gamma, \boldsymbol{\lambda}, \beta, \mathbf{y}) / p(\mathbf{y}) \propto p(\gamma, \boldsymbol{\lambda}, \beta, \mathbf{y})$. It is therefore sufficient to maximize $p(\gamma, \boldsymbol{\lambda}, \beta, \mathbf{y})$, which is obtained by integrating out \mathbf{w} from the joint $p(\mathbf{y}, \mathbf{w}, \gamma, \boldsymbol{\lambda}, \beta)$, i.e.,

$$\begin{aligned} p(\gamma, \boldsymbol{\lambda}, \beta, \mathbf{y}) &= \int p(\mathbf{w}, \gamma, \boldsymbol{\lambda}, \beta, \mathbf{y}) d\mathbf{w} \\ &= (2\pi)^{-\frac{M}{2}} |\mathbf{C}|^{-\frac{1}{2}} \exp \left[-\frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} \right] p(\gamma | \boldsymbol{\lambda}) p(\boldsymbol{\lambda}) p(\beta), \end{aligned} \quad (14)$$

where $\mathbf{C} = \beta^{-1} \mathbf{I}_M + \boldsymbol{\Phi} \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}^T$. Equivalently, we can maximize the logarithm of (14), denoted as \mathcal{L} , with respect to $\boldsymbol{\lambda}$, γ , and β . After some algebraic manipulations similar to those reported in [7], \mathcal{L} can be expressed as

$$\begin{aligned} \mathcal{L} &= -\frac{M}{2} \log(2\pi) - \frac{1}{2} \log |\mathbf{C}| - \frac{1}{2} \mathbf{y}^T \mathbf{C}^{-1} \mathbf{y} \\ &\quad + N \log \frac{1}{2} + \rho \log |\boldsymbol{\Psi}| - \frac{1}{2} \sum_{i=1}^N \lambda_i \gamma_i + N \log \frac{\delta^\rho}{\Gamma(\rho)} \\ &\quad - \delta \sum_{i=1}^N \lambda_i + \log \frac{\theta^\kappa}{\Gamma(\kappa)} + (\kappa - 1) \log \beta - \theta \beta \end{aligned} \quad (15)$$

Taking the partial derivatives of \mathcal{L} , with respect to the parameters γ , $\boldsymbol{\lambda}$, and β , and equating them to zero, the following expressions are obtained

$$\frac{\partial \mathcal{L}}{\partial \gamma_i} = 0 \Rightarrow \gamma_i = -\frac{1}{2\lambda_i} + \sqrt{\frac{1}{4\lambda_i^2} + \frac{\langle w_i^2 \rangle}{\lambda_i}} \quad (16)$$

$$\frac{\partial \mathcal{L}}{\partial \lambda_i} = 0 \Rightarrow \lambda_i = \frac{\rho}{\frac{1}{2}\gamma_i + \delta} \quad (17)$$

$$\frac{\partial \mathcal{L}}{\partial \beta} = 0 \Rightarrow \beta = \frac{M + 2\kappa - 2}{\|\mathbf{y} - \boldsymbol{\Phi} \boldsymbol{\mu}\|_2^2 + 2\theta} \quad (18)$$

$$\frac{\partial \mathcal{L}}{\partial \rho} = 0 \Rightarrow \log |\boldsymbol{\Psi}| + N \log \delta - N \psi(\rho) = 0 \quad (19)$$

$$\frac{\partial \mathcal{L}}{\partial \delta} = 0 \Rightarrow \delta = \frac{\rho N}{\sum_{i=1}^N \lambda_i}, \quad (20)$$

where $\psi(\cdot)$ is the digamma function, and $\langle w_i^2 \rangle = \mu_i^2 + \boldsymbol{\Sigma}_{ii}$, with μ_i being the i th diagonal element of $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}_{ii}$ the i th diagonal element of $\boldsymbol{\Sigma}$. Due to the dependencies among the model parameters, equations (13) and (16) - (18) form an iterative updating scheme. However, this scheme is problematic in practice, since (13) requires the inversion of a $N \times N$ matrix, which is both computationally demanding and susceptible to numerical errors. By properly modifying the sub-optimum scheme proposed in [6], a numerically robust, fast Bayesian Adaptive Lasso (Fast-BALa) algorithm is proposed in the following section.

3.2. Fast suboptimal solution

In this section we develop a fast sub-optimal technique (note the connection with the analysis presented in [7] - interested readers are encouraged to refer to the analysis of [7] for details) to maximize the marginal log-likelihood \mathcal{L} . In the following analysis, note that the importance of the parameter γ lies in the fact that setting $\gamma_i = 0$ is equivalent to pruning the i th variable out of the model, i.e., $w_i = 0$ (see Eq. (5)).

Exploiting the diagonal form of $\boldsymbol{\Lambda}$, matrix \mathbf{C} in (14) is written in a form that is convenient for analyzing the dependence of \mathcal{L} on a single parameter,

$$\mathbf{C} = \overbrace{\beta^{-1} \mathbf{I}_M}^{\mathbf{C}_{-i}} + \sum_{j \neq i} \gamma_j \boldsymbol{\phi}_j \boldsymbol{\phi}_j^T + \gamma_i \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T \quad (21)$$

Using the matrix inversion and determinant lemmas, (21) allows us to write

$$\mathbf{C}^{-1} = \mathbf{C}_{-i}^{-1} - \frac{\mathbf{C}_{-i}^{-1} \boldsymbol{\phi}_i \boldsymbol{\phi}_i^T \mathbf{C}_{-i}^{-1}}{\frac{1}{\gamma_i} + \boldsymbol{\phi}_i^T \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}_i} \quad (22)$$

$$|\mathbf{C}| = |1 + \gamma_i \boldsymbol{\phi}_i^T \mathbf{C}_{-i}^{-1} \boldsymbol{\phi}_i| |\mathbf{C}_{-i}^{-1}|. \quad (23)$$

Thus, the summation of the terms of \mathcal{L} that depend on γ , denoted as $\mathcal{L}(\gamma)$, becomes

$$\begin{aligned} \mathcal{L}(\gamma) &= -\frac{1}{2} \left[\log |\mathbf{C}_{-i}^{-1}| + \mathbf{y}^T \mathbf{C}_{-i}^{-1} \mathbf{y} + \sum_{j \neq i} \lambda_j \gamma_j \right] \\ &\quad + \frac{1}{2} \left[\log \frac{1}{1 + \gamma_i s_i} + \frac{q_i^2 \gamma_i}{1 + \gamma_i s_i} - \lambda_i \gamma_i \right] = \mathcal{L}(\gamma_{-i}) + l(\gamma_i), \end{aligned} \quad (24)$$

Table 1. The Fast-BALa algorithm

```

Input  $\Phi, \mathbf{y}, \kappa, \theta, \rho, \delta$ .
Set  $\beta = 0.01 \|\mathbf{y}\|_2^2$ .
Initialize  $\gamma^{(0)} = \mathbf{0}, \lambda^{(0)} = \mathbf{1}$ .
for  $n = 1, 2, \dots$  do
    Compute all possible  $\gamma_i^{(n)}$ 's using (26).
    Choose the  $\gamma_i$  with the maximum
     $d^{(n)} = l(\gamma_i^{(n)}) - l(\gamma_i^{(n-1)})$ 
    if  $d^{(n)}$  is sufficiently small, break, endif
    if  $(q_i^{(n)})^2 - s_i^{(n)} > \lambda_i^{(n)}$  and  $\gamma_i^{(n-1)} = 0$ 
        Add  $\gamma_i$  to the model.
    else if  $(q_i^{(n)})^2 - s_i^{(n)} > \lambda_i^{(n)}$  and  $\gamma_i^{(n-1)} > 0$ 
        Keep  $\gamma_i^{(n)}$  as the updated value of  $\gamma_i$ .
    else if  $(q_i^{(n)})^2 - s_i^{(n)} < \lambda_i^{(n)}$ 
        Prune  $\gamma_i$  from the model.
    endif
    Update  $\boldsymbol{\mu}$  and  $\boldsymbol{\Sigma}$  using (13).
    Update  $s_i$  and  $q_i$  using (29).
    Update  $\boldsymbol{\lambda}$  using (17).
    Update  $\rho$  and  $\delta$  using (19) and (20).
endfor

```

where s_i and q_i are defined as $s_i = \phi_i^T \mathbf{C}_{-i}^{-1} \phi_i$ and $q_i = \phi_i^T \mathbf{C}_{-i}^{-1} \mathbf{y}$. It is interesting to see that the term $\mathcal{L}(\gamma_{-i})$ does not depend on the single parameter γ_i . Thus the partial derivative of $\mathcal{L}(\boldsymbol{\gamma})$ with respect to γ_i can be computed as

$$\begin{aligned} \frac{\partial \mathcal{L}(\boldsymbol{\gamma})}{\partial \gamma_i} &= \frac{\partial l(\gamma_i)}{\partial \gamma_i} = \frac{1}{2} \left[-\frac{s_i}{1 + \gamma_i s_i} + \frac{q_i^2}{(1 + \gamma_i s_i)^2} - \lambda_i \right] \\ &= -\frac{1}{2} \left[\frac{\gamma_i^2 (\lambda_i s_i^2) + \gamma_i (s_i^2 + 2\lambda_i s_i) + (s_i - q_i^2 + \lambda_i)}{(1 + \gamma_i s_i)^2} \right]. \end{aligned} \quad (25)$$

Computing the roots of the numerator polynomial in (25), γ_i is estimated as ([7])

$$\gamma_i = \frac{-s_i (s_i + 2\lambda_i) + s_i \sqrt{(s_i + 2\lambda_i)^2 - 4\lambda_i (s_i - q_i^2 + \lambda_i)}}{2\lambda_i s_i^2} \quad (26)$$

if $q_i^2 - s_i > \lambda$ and 0 otherwise. The updating of γ_i using (26) is followed by the updating of $\boldsymbol{\mu}$, $\boldsymbol{\Sigma}$, $\boldsymbol{\lambda}$, s_i , and q_i . Based on the analysis in [6], the updating of s_i and q_i , which has to be performed for all N model variables, is made as follows,

$$S_i = \beta \phi_i^T \phi_i - \phi_i^T \beta \Phi \boldsymbol{\Sigma} \Phi^T \beta \phi_i \quad (27)$$

$$Q_i = \beta \phi_i^T \mathbf{y} - \phi_i^T \beta \Phi \boldsymbol{\Sigma} \Phi^T \beta \mathbf{y}, \quad (28)$$

$$s_i = \frac{S_i}{1 - \gamma_i S_i}, \quad q_i = \frac{Q_i}{1 - \gamma_i S_i} \quad (29)$$

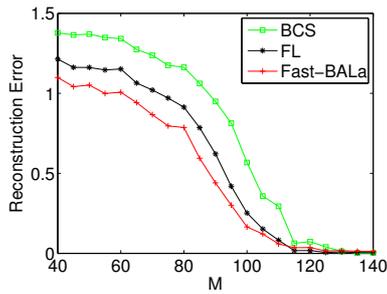
The resulting fast Bayesian Adaptive Lasso algorithm is summarized in Table 1. In its simplest form, i.e., when all λ_i 's are

identical and are updated using a single equation, Fast-BALa is similar to the fast Laplace (FL) algorithm presented in [7]. Updating each λ_i independently using (17) improves estimation accuracy. It should be noted that FL solves the conventional lasso problem (2). Moreover, the updating of matrix $\boldsymbol{\Sigma}$ takes into account only the selected variables at the current iteration, which, in practice, are significantly fewer than N . The parameter β is not included in the iterative scheme; rather, it is set equal to $\beta = 0.01 \|\mathbf{y}\|_2^2$, as in [7]. A powerful feature of the algorithm is its ability to remove basis vectors from the model that have been selected at early stages, which is not the case for greedy algorithms, e.g., the OMP algorithm.

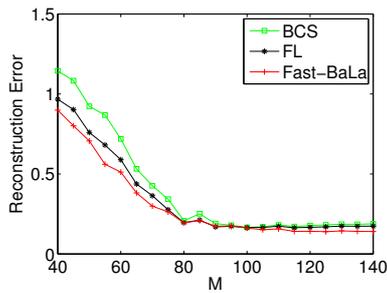
4. SIMULATION RESULTS

Simulations are conducted following the experimental settings of [4, 7]. We compare the Fast-BALa algorithm with the Bayesian compressive sensing (BCS) algorithm of [4], and FL of [7]. In all experiments, we evaluate the reconstruction error as $\|\mathbf{w} - \hat{\mathbf{w}}_{est}\|_2 / \|\mathbf{w}\|_2$, where $\hat{\mathbf{w}}_{est}$ is the estimated coefficients vector. The first example considers a signal of length $N = 512$ that contains $T = 20$ non-zero coefficients at random locations. As in [7], the columns of Φ are uniformly distributed on the sphere R^N . The non-zero coefficients are generated by employing: (a) ± 1 uniform spikes, (b) the standard Gaussian distribution, $\mathcal{N}(0, 1)$. The number of measurements M varies between 40 and 120 with a step size of 5, and the average reconstruction error is computed for 100 signal realizations. Zero-mean Gaussian noise with standard deviation $\sigma = 0.03$ is added to the model and the results are shown in Fig. 1. As expected, the reconstruction error reduces as the number of measurements increases. It can be seen from Fig. 1 that the proposed Fast-BALa algorithm has the best performance, especially for lower values of M . It is interesting to note that the only difference among the three methods lies on the updating of the sparsity controlling parameter $\boldsymbol{\lambda}$; in BCS $\boldsymbol{\lambda} = \mathbf{0}$, in FL all λ_i 's are equal and are updated using a single equation, while in Fast-BALa λ_i 's are independent and are updated using (17). The average computation time per sample for the three methods is 0.0271s for BCS, 0.0872s for FL, and 0.1713s for Fast-BALa.

Next, the proposed Fast-BALa algorithm is used for the sparse representation of a 512×512 Mondrian image. As in [4, 7], a multiscale CS scheme is applied on the wavelet transform of the image using the ‘‘symmlet 8’’ wavelet with the coarsest scale 4 and finest scale 6. The number of wavelet samples is $N = 4096$, the number of measurements is $M = 2713$, and the measurement matrices are drawn from a uniform spherical distribution. Fig. 2(a) displays the linear reconstruction of the image, using all the $M = 2713$ measurements, which is the best possible reconstruction. The corresponding reconstructions of the BCS, FL, and Fast-BALa are displayed in Figs. 2(b), 2(c), and 2(d), respectively. The number of the nonzero coefficients for BCS, FL, and Fast-BALa



(a) \pm uniform spikes



(b) Standard Gaussian distribution

Fig. 1. Comparison of Bayesian CS algorithms

algorithms are 751, 802, and 911 respectively. As demonstrated in Fig. 2, Fast-BALa offers the best reconstruction performance, although it is slightly more computationally demanding.

5. CONCLUSION

In this paper, we have provided a Bayesian perspective over the adaptive lasso criterion. The non-uniformly weighted ℓ_1 penalty of the adaptive lasso was attained by employing N independent Laplace priors. An efficient algorithm was then properly modified to perform Bayesian inference. Experimental results showed that the proposed method achieves better performance than other state-of-the-art Bayesian CS algorithms, at the cost of a small additional complexity.

6. REFERENCES

- [1] E.J. Candes and M.B. Wakin, "An introduction to compressive sampling," *IEEE SP Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [2] R. Tibshirani, "Regression shrinkage and selection via the Lasso," *Journal of the Royal Statistical Society*, vol. 58, no. 1, pp. 267–288, 1996.
- [3] M.A.T. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1150–1159, Sept. 2003.

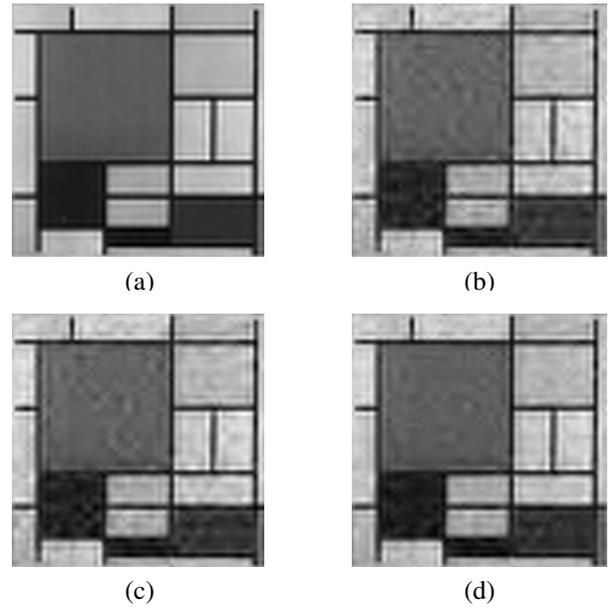


Fig. 2. Mondrian image reconstruction using the algorithms: (a) linear sparse reconstruction (error = 0.13325), (b) BCS (error = 0.14979, time = 11.7836s), (c) FL (error = 0.15094, time = 15.6063s), (d) Fast-BALa (error = 0.14402, time = 26.5266s).

- [4] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.
- [5] M.E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, 2001.
- [6] M. E. Tipping and A. C. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *Proc. of the 9th Int. Workshop on AI and Stats*, 2003, pp. 3–6.
- [7] S.D. Babacan, R. Molina, and A.K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.
- [8] T. Park and G. Casella, "The Bayesian Lasso," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 681–686, June 2008.
- [9] H. Zou, "The adaptive Lasso and its oracle properties," *Journal of the American Statistical Association*, vol. 101, pp. 1418–1429, Dec. 2006.
- [10] K.E. Themelis, A.A. Rontogiannis, and K.D. Koutroumbas, "A novel hierarchical Bayesian approach for sparse semi-supervised hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. 60, no. 2, pp. 585–599, Feb. 2012.