FAST BAYESIAN COMPRESSIVE SENSING USING LAPLACE PRIORS

S. Derin Babacan[1]*, Rafael Molina*[2]*, Aggelos K. Katsaggelos*[1]

[1]Department of Electrical Engineering
and Computer Science
Northwestern University, Evanston, IL 60208, USA
sdb@northwestern.edu, aggk@eecs.northwestern.edu

[2]Departamento de Ciencias
de la Computación e I.A.
Universidad de Granada, 18071 Granada, Spain
rms@decsai.ugr.es

ABSTRACT

In this paper we model the components of the compressive sensing (CS) problem using the Bayesian framework by utilizing a hierarchical form of the Laplace prior to model sparsity of the unknown signal. This signal prior includes some of the existing models as special cases and achieves a high degree of sparsity. We develop a constructive (greedy) algorithm resulting from this formulation where necessary parameters are estimated solely from the observation and therefore no user-intervention is needed. We provide experimental results with synthetic 1D signals and images, and compare with the state-of-the-art CS reconstruction algorithms demonstrating the superior performance of the proposed approach.

Index Terms— Bayesian methods, compressive sensing, inverse problems, sparse Bayesian learning, relevance vector machine (RVM).

1. INTRODUCTION

Compressive sensing (or sampling) has become a very active research area in recent years due to its interesting theoretical nature and its practical utility in a wide range of applications. Let $\mathbf{w}$ be an $N \times 1$ sparse signal, i.e., most of its coefficients are zero. Consider the following acquisition system

$$\mathbf{y} = \Phi\mathbf{w} + \mathbf{n}, \tag{1}$$

where $M \times 1$ linear measurements $\mathbf{y}$ of the original unknown signal $\mathbf{w}$ are taken with an $M \times N$ measurement matrix $\Phi = [\phi_1, \phi_2, \ldots, \phi_N]$, with $M \ll N$ and $\mathbf{n}$ representing the acquisition noise. According to the theory of compressive sensing when the number of measurements is small compared to the number of signal coefficients ($M \ll N$), under certain conditions the original signal $\mathbf{w}$ can be reconstructed very accurately by utilizing appropriate reconstruction algorithms [1, 2].

A common formulation of the reconstruction algorithm is obtained by exploiting the sparsity of $\mathbf{w}$ and regularizing the inverse problem by constraining the l_p norm of $\mathbf{w}$, $\| \mathbf{w} \|_p$ with $0 \leq p \leq 1$, that is,

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \left\{ \| \mathbf{y} - \Phi\mathbf{w} \|_2^2 + \tau \| \mathbf{w} \|_p \right\}. \tag{2}$$

The case with $p = 0$, where $\| \mathbf{w} \|_0$ is the number of nonzero coefficients in $\mathbf{w}$, results in an optimization problem which is NP-hard. A more common formulation is obtained with the use of the l_1 norm ($p = 1$). A number of methods have been proposed to solve the CS reconstruction problems defined in (2), most of which are examples of energy minimization methods, including linear programming algorithms [3, 4] and constructive (greedy) algorithms [5, 6].

The CS reconstruction problem can also be formulated in a Bayesian framework [7, 8]. In [7], the *relevance vector machine* (RVM) proposed in [9] is adapted to the CS problem. Independent Laplace priors are utilized for each coefficient in an expectation-propagation framework in [8]. However, the resulting algorithm is complicated to implement, and all required parameters are not estimated, but rather left as parameters to be tuned.

In this paper, we also formulate the CS reconstruction problem from a Bayesian perspective. We utilize a Bayesian model for the CS problem and propose the use of Laplace priors on the basis coefficients in a hierarchical manner. As will be shown, our formulation includes the RVM formulation [9] as a special case, but results in smaller reconstruction errors while imposing sparsity to a higher extent. We derive an efficient greedy constructive algorithm resulting from our formulation. The proposed algorithm is fully automated since all required model parameters are estimated along with the unknown signal coefficients $\mathbf{w}$. This is in contrast to most of the proposed methods in the literature which include a number of parameters to be tuned specifically to the data, which is a cumbersome process. We will demonstrate with experimental results that despite being fully automated, the proposed algorithm provides competitive and even higher reconstruction performance than state-of-the-art methods.

The rest of this paper is organized as follows: We present the hierarchical Bayesian modeling of the CS problem in Section 2. In Section 3 we apply the evidence procedure to the CS problem and propose a constructive reconstruction algorithm. We present experimental results in Section 4 and conclusions are drawn in Section 5.

2. BAYESIAN MODELING

In Bayesian modeling, all unknowns are treated as stochastic quantities with assigned probability distributions. The unknown signal $\mathbf{w}$ is assigned a *prior* distribution $\mathrm{p}(\mathbf{w}|\boldsymbol{\gamma})$, which models our knowledge on the nature of $\mathbf{w}$. The observation $\mathbf{y}$ is also a random process with *conditional* distribution $\mathrm{p}(\mathbf{y}|\mathbf{w}, \beta)$, where $\beta = 1/\sigma^2$ is the inverse noise variance. These distributions depend on the model parameters $\boldsymbol{\gamma}$ and β, which are called *hyperparameters*, and additional prior distributions, called *hyperpriors*, are assigned to them.

2.1. Observation (Noise) Model

The observation noise is independent and Gaussian with zero mean and variance equal to β^{-1}, that is, with (1),

$$\mathrm{p}(\mathbf{y}|\mathbf{w}, \beta) = \mathcal{N}(\mathbf{y}|\Phi\mathbf{w}, \beta^{-1}) \tag{3}$$

This work was supported in part by the "Comisión Nacional de Ciencia y Tecnologa" under contract TIC2007-65533 and the Spanish research programme Consolider Ingenio 2010: MIPRCV (CSD2007-00018).

A Gamma prior can be placed on β to estimate it. Unfortunately, this method cannot be used in practice with greedy algorithms since the reconstruction and, therefore, the estimation of β using this method are unreliable at early iterations. Due to the under-determined nature of the compressive sensing problem, once the estimate of β is very far from its true value, the reconstruction quality is also significantly affected. Therefore, we fix the estimate of this parameter in the beginning of the algorithm using $\beta = 0.01 \parallel \mathbf{y} \parallel_2^2$ inspired by [4, 7].

2.2. Signal Model

The l_1 regularization formulation in (2) with $p = 1$ is equivalent to using a Laplace prior on the coefficients $\mathbf{w}$, that is,

$$\mathrm{p}(\mathbf{w}|\lambda) = \frac{\lambda}{2}\exp(-\lambda|\mathbf{w}|) \tag{4}$$

and using a *maximum a posteriori* (MAP) formulation with (3) and (4) for $\tau = \lambda/\beta$. However, this formulation of the Laplace prior does not allow for a tractable Bayesian analysis, since it is not conjugate to the conditional distribution in (3). To alleviate this, hierarchical priors are employed.

As the first stage of a hierarchical model, the following prior is employed on $\mathbf{w}$

$$\mathrm{p}(\mathbf{w}|\boldsymbol{\gamma}) = \prod_{i=1}^{N} \mathcal{N}(w_i|0, \gamma_i), \tag{5}$$

where $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_N)$. In the second stage, we use the following hyperpriors on γ_i

$$\mathrm{p}(\gamma_i|\lambda) = \Gamma(\gamma_i|1, \lambda/2) = \frac{\lambda}{2}\exp\left(-\frac{\lambda\gamma_i}{2}\right),\ \gamma_i \geq 0,\ \lambda \geq 0 \tag{6}$$

and finally, we model λ as the realization of a Jeffrey's hyperprior

$$\mathrm{p}(\lambda) \propto \frac{1}{\lambda}. \tag{7}$$

Based on the above we have [10]

$$\begin{aligned}\mathrm{p}(\mathbf{w}|\lambda) &= \int \mathrm{p}(\mathbf{w}|\boldsymbol{\gamma})\mathrm{p}(\boldsymbol{\gamma}|\lambda)d\boldsymbol{\gamma} = \prod_i \int \mathrm{p}(w_i|\gamma_i)\mathrm{p}(\gamma_i|\lambda)d\gamma_i \\ &= \frac{\lambda^{N/2}}{2^N}\exp\left(-\sqrt{\lambda}\sum_i |w_i|\right).\end{aligned} \tag{8}$$

The proposed modeling constitutes a three-stage hierarchical form. The first two stages (5) and (6) of this hierarchical prior result in a Laplace distribution $\mathrm{p}(\mathbf{w}|\lambda)$ in (8), and the last stage (7) is embedded to calculate λ.

An alternative Bayesian formulation applicable to the CS problem is the *relevance vector machine* (RVM) [7, 9] (or *sparse Bayesian learning* (SBL) [11]), where separate Gaussian priors employed on the entries of $\mathbf{w}$. As explained in [8], compared to this formulation, Laplace priors enforce the sparsity constraint more heavily by distributing the posterior mass more on the axes so that signal coefficients close to zero are preferred. Furthermore, the Laplace prior is also the prior that promotes sparsity to the largest extent while being log-concave. The log-concavity provides the very useful advantage of eliminating local-minima since it leads to unimodal posterior distributions [8].

By combining the stages of the hierarchical Bayesian model, the joint distribution can finally be defined as $\mathrm{p}(\mathbf{w}, \boldsymbol{\gamma}, \lambda, \beta, \mathbf{y}) = \mathrm{p}(\mathbf{y}|\mathbf{w}, \beta)\mathrm{p}(\mathbf{w}|\boldsymbol{\gamma})\mathrm{p}(\boldsymbol{\gamma}|\lambda)\mathrm{p}(\lambda)$, where $\mathrm{p}(\mathbf{y}|\mathbf{w}, \beta)$, $\mathrm{p}(\mathbf{w}|\boldsymbol{\gamma})$, $\mathrm{p}(\boldsymbol{\gamma}|\lambda)$ and $\mathrm{p}(\lambda)$ are defined in (3), (5), (6), and (7), respectively.

3. BAYESIAN INFERENCE

In this paper we utilize the evidence procedure (type-II maximum likelihood approach) [12] to perform Bayesian inference. Our inference procedure is based on the following decomposition

$$\mathrm{p}(\mathbf{w}, \boldsymbol{\gamma}, \lambda, \beta \mid \mathbf{y}) = \mathrm{p}(\mathbf{w}|\mathbf{y}, \boldsymbol{\gamma}, \beta, \lambda)\ \mathrm{p}(\boldsymbol{\gamma}, \beta, \lambda|\mathbf{y}). \tag{9}$$

Since $\mathrm{p}(\mathbf{w}|\mathbf{y}, \boldsymbol{\gamma}, \beta, \lambda) \propto \mathrm{p}(\mathbf{w}, \mathbf{y}, \boldsymbol{\gamma}, \beta, \lambda)$, then $\mathrm{p}(\mathbf{w}|\mathbf{y}, \boldsymbol{\gamma}, \beta, \lambda)$ is found to be a multivariate Gaussian distribution $\mathcal{N}(\mathbf{w}|\mu, \Sigma)$ with parameters

$$\Sigma = \left[\beta\Phi^T\Phi + \Lambda\right]^{-1}, \tag{10}$$

$$\mu = \Sigma\ \beta\Phi^T\mathbf{y}, \tag{11}$$

with $\Lambda = \mathrm{diag}(1/\gamma_i)$.

We now utilize $\mathrm{p}(\boldsymbol{\gamma}, \beta, \lambda|\mathbf{y})$ in (9) to estimate the hyperparameters. In the type-II maximum likelihood procedure we represent $\mathrm{p}(\boldsymbol{\gamma}, \beta, \lambda|\mathbf{y})$ by a degenerate distribution where the distribution is replaced by a delta function at its mode, where we assume that this posterior distribution is sharply peaked around its mode [12]. Then, using $\mathrm{p}(\boldsymbol{\gamma}, \beta, \lambda|\mathbf{y}) = \frac{\mathrm{p}(\mathbf{y}, \boldsymbol{\gamma}, \beta, \lambda)}{\mathrm{p}(\mathbf{y})} \propto \mathrm{p}(\mathbf{y}, \boldsymbol{\gamma}, \beta, \lambda)$, we estimate the hyperparameters by the maxima of the joint distribution $\mathrm{p}(\mathbf{y}, \boldsymbol{\gamma}, \beta, \lambda)$, or equivalently its logarithm, which is obtained by

$$\begin{aligned}\mathcal{L} &= \log \int \mathrm{p}(\mathbf{y}|\mathbf{w}, \beta)\ \mathrm{p}(\mathbf{w}|\boldsymbol{\gamma})\ \mathrm{p}(\boldsymbol{\gamma}|\lambda)\ \mathrm{p}(\lambda)\ d\mathbf{w} \\ &= -\frac{1}{2}\log|\mathbf{C}| - \frac{1}{2}\mathbf{y}^t\mathbf{C}^{-1}\mathbf{y} + (N-1)\log\lambda - \frac{\lambda}{2}\sum_i \gamma_i\end{aligned} \tag{12}$$

with $\mathbf{C} = \left(\beta^{-1}\mathbf{I} + \Phi\Lambda^{-1}\Phi^t\right)$. Maximizing $\mathcal{L}$ with respect to λ results in the following update equation

$$\lambda = \frac{N-1}{\sum_i \gamma_i/2}. \tag{13}$$

Similarly, the hyperparameters $\boldsymbol{\gamma}$ can be found by maximizing $\mathcal{L}$.

To promote sparsity and to decrease the computational requirements, only a single γ_i will be updated at each iteration of the algorithm instead of updating the whole vector $\boldsymbol{\gamma}$. The update of a single γ_i can be found by maximizing $\mathcal{L}$ when all components of $\boldsymbol{\gamma}$ except γ_i are kept fixed. Let us express matrix $\mathbf{C}$ in (12) as

$$\mathbf{C} = \beta^{-1}\mathbf{I} + \sum_{j\neq i}\gamma_j\phi_j\phi_j^t + \gamma_i\phi_i\phi_i^t = \mathbf{C}_{-i} + \gamma_i\phi_i\phi_i^t, \tag{14}$$

where $\mathbf{C}_{-i}$ denotes that the contribution of the i^{th} basis is not included. Using the Woodbury identity in (14) we obtain

$$\mathbf{C}^{-1} = \mathbf{C}_{-i}^{-1} - \frac{\mathbf{C}_{-i}^{-1}\phi_i\phi_i^t\mathbf{C}_{-i}^{-1}}{1/\gamma_i + \phi_i^t\mathbf{C}_{-i}^{-1}\phi_i} \tag{15}$$

and using the determinant identity we obtain

$$|\mathbf{C}| = |\mathbf{C}_{-i}||1 + \gamma_i\,\phi_i^t\mathbf{C}_{-i}^{-1}\phi_i|\,. \tag{16}$$

Substituting the last two equations in (12) and treating $\mathcal{L}$ as a function of $\boldsymbol{\gamma}$ only, we obtain

$$\begin{aligned}\mathcal{L}(\boldsymbol{\gamma}) &= -\frac{1}{2}\big[\log|\mathbf{C}_{-i}| + \mathbf{y}^t\mathbf{C}_{-i}^{-1}\mathbf{y} + \frac{\lambda}{2}\sum_{j\neq i}\gamma_j\big] \\ &\quad + \frac{1}{2}\big[\log\frac{1}{1+\gamma_i\,s_i} + \frac{q_i^2\,\gamma_i}{1+\gamma_i\,s_i} - \lambda\gamma_i\big] \\ &= \mathcal{L}(\boldsymbol{\gamma}_{-i}) + \frac{1}{2}\big[\log\frac{1}{1+\gamma_i\,s_i} + \frac{q_i^2\,\gamma_i}{1+\gamma_i\,s_i} - \lambda\gamma_i\big] \\ &= \mathcal{L}(\boldsymbol{\gamma}_{-i}) + l(\gamma_i)\end{aligned} \tag{17}$$

where q_i and s_i are defined as

$$s_i = \phi_i^t \mathbf{C}_{-i}^{-1} \phi_i, \;\; q_i = \phi_i^t \mathbf{C}_{-i}^{-1} \mathbf{y} \tag{18}$$

This way the terms related to a single hyperparameter γ_i are separated from others. The maximum of $\mathcal{L}(\boldsymbol{\gamma})$, when all components of $\boldsymbol{\gamma}$ except γ_i are kept fixed, is therefore equal to the maximum of $l(\gamma_i)$, and it is found by taking its derivative with respect to γ_i and setting it equal to zero. The optimal value found in this manner is given by (the derivation is omitted)

$$\gamma_i = \begin{cases} \frac{-s_i(s_i+2\lambda)+s_i\sqrt{(s_i+2\lambda)^2-4\lambda(s_i-q_i^2+\lambda)}}{2\lambda s_i^2} & \text{if } q_i^2 - s_i > \lambda \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

Note that in the case of $\gamma_i = 0$, the corresponding basis ϕ_i is pruned out from the model and μ_i is set equal to zero. Therefore (19) provides a systematic method of deciding which basis vectors should be included in the model and which should be excluded.

Utilizing (19), we obtain an iterative procedure by starting with an "empty" model ($\boldsymbol{\gamma} = 0$) and iteratively adding/updating/deleting components. The procedure is summarized below in Algorithm 1.

Algorithm 1 Fast Laplace

1: INPUTS: Φ, $\mathbf{y}$
2: OUTPUTS: $\mathbf{w}$, Σ, γ
3: Initialize all $\gamma_i = 0$, $\lambda = 0$
4: **while** convergence criterion not met **do**
5: Choose a γ_i (or equivalently choose a basis vector ϕ_i)
6: Update γ_i using (19)
7: Update Σ and μ
8: Update s_i, q_i
9: Update λ using (13)
10: **end while**

At step 5 of the algorithm, the candidate γ_i is selected by calculating each γ_i and choosing the one that results in the greatest increase in $\mathcal{L}(\boldsymbol{\gamma})$ in (17).

Note that unlike other constructive (or greedy) methods such as OMP [5], StOMP [6], included basis vectors can also be deleted once they are determined to be irrelevant. This is a powerful feature of the algorithm, since errors in the beginning of the reconstruction process can be fixed in later stages by effectively pruning out irrelevant basis vectors which can drive the algorithm away from the optimal result.

We complete this section by comparing the variance estimates provided by the RVM with the ones provided by the proposed method in terms of sparsity. As we have seen the estimate provided by the modeling using the Laplace distribution is given by the maxima of $l(\gamma_i)$ in (17) whereas the estimate γ_i^{RVM} in the RVM framework corresponds to the case with $\lambda = 0$ [13]. The difference $\gamma_i^{\text{RVM}} - \gamma_i^{\text{L}}$ can be expressed as

$$\gamma_i^{\text{RVM}} - \gamma_i^{\text{L}} = \begin{cases} 0 & \text{if } q_i^2 - s_i < 0 \\ \gamma_i^{\text{RVM}} & \text{if } 0 \le q_i^2 - s_i < \lambda \\ \frac{s_i^2+2\lambda q_i^2-s_i\sqrt{s_i^2+4\lambda q_i^2}}{2\lambda s_i^2} > 0 & \text{if } q_i^2 - s_i \ge \lambda \end{cases} \tag{20}$$

Therefore, the estimates γ_i^{L} using the Laplace prior are always smaller than the estimates γ_i^{RVM} of the relevance vector machine. Note also that compared to RVM more components will possibly be pruned out from the model when $\lambda > 0$, since the cardinality of the set $\{w_i\}$ for which $q_i^2 - s_i > \lambda$ is smaller than that of the set $\{w_i\}$ for which $q_i^2 - s_i > 0$. These observations imply that the solution obtained by the proposed method is at least as sparse as the one provided by the RVM. This will also be shown empirically in Section 4.

4. EXPERIMENTAL RESULTS

In this section we present experimental results with both one-dimensional (1D) synthetic signals and 2D images to demonstrate the performance of the proposed method. For the 1D synthetic signal reconstruction experiments, we generated four different types of signals of length 512, where 20 coefficients at random locations of the signals are drawn from five different probability distributions, and the rest of the coefficients are set equal to zero. The nonzero coefficients of the sparse signals are realizations of the following five distributions: 1) Uniform ± 1 random spikes, 2) zero-mean unit variance Gaussian, 3) unit variance Laplace, and 4) Student's t with 3 degrees of freedom. As the measurement matrix Φ we chose a uniform spherical ensemble, where the columns ϕ_i are uniformly distributed on the sphere R^N. In the experiments we vary the number of measurements M from 40 to 120 in steps of 5. We added zero mean white Gaussian noise with standard deviation 0.03 to the observations.

The reconstruction error is calculated as $\| \hat{\mathbf{w}} - \mathbf{w} \|_2^2 \; / \; \| \mathbf{w} \|_2^2$, where $\hat{\mathbf{w}}$ and $\mathbf{w}$ are the estimated and true coefficient vectors, respectively. The criterion $\| \mathcal{L}(\boldsymbol{\gamma}^k) - \mathcal{L}(\boldsymbol{\gamma}^{k-1}) \|^2 \; / \; \| \mathcal{L}(\boldsymbol{\gamma}^{k-1}) \|^2 < 10^{-8}$ is used to terminate the iterative procedure.

We compare the proposed method (denoted by Laplace) with the algorithms BCS [7], BP [3], OMP [5], StOMP with CFAR thresholding (denoted by FAR) [6], and GPSR [4]. For all algorithms, their MATLAB implementations in the corresponding websites are used. The required algorithm parameters are set according to their default setups.

Average reconstruction errors of 100 runs are shown in Fig. (1) for all types of signals. It is clear that the proposed algorithm outperforms all other methods in terms of reconstruction error except for the first signal, for which it provides the second-best performance after BP. However, BP results in worse performance than other methods for the rest of the signals. The poor performance of GPSR can be explained by assuming that the default selection of algorithm parameters is poor for these types of signals, and its performance is expected to increase if the optimal parameters can be found by trial-and-error. Note however that both BCS and the proposed method do not require parameter tuning. Despite this fact, note that the proposed method provides the best overall performance among all methods.

In the second set of experiments, we present a comparison between the proposed method and the algorithms BP, BCS, and StOMP with CFAR and CFDR thresholding on a widely used experimental setup, namely the multiscale CS reconstruction [14] of the Mondrian image. We adapted the same test parameters as in the SparseLab package (http://sparselab.stanford.edu), where the number of samples is $N = 4096$, the number of measurements is $M = 2073$, and the measurement matrices are drawn from a uniform spherical distribution. The multiscale CS scheme is applied on the wavelet transform of the image with a "symmlet8" wavelet with the coarsest scale 4 and finest scale 6. The parameters of the algorithms BP, CFAR and CFDR are chosen as in the SparseLab package. As in the previous experiment, the parameters of BCS and the proposed method are solely estimated from the measurements.

Since the measurement matrices are random, the experiment is repeated 100 times and their average is reported. Average recon-

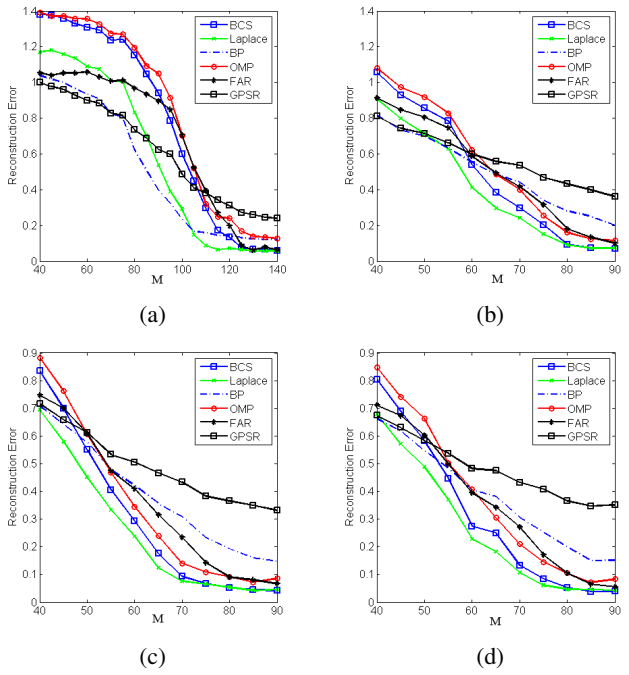

Fig. 1. Number of measurements M vs reconstruction error with noisy observations for different algorithms. (a) Uniform spikes ± 1; Nonuniform spikes drawn from (b) zero mean unit variance Gaussian, (c) unit variance Laplace, (d) Student's t with 3 degrees of freedom. In (b), (c), and (d) values corresponding to $M > 90$ are not shown for clarity as the error rates converged.

struction errors, running times and the number of nonzero components in the reconstructed images are shown in Table 1, where "Linear" denotes linear reconstruction with $M = 4096$ measurements and represents the best reconstruction performance that can be achieved. It is clear that although BCS and Laplace have nearly the same error rate, Laplace is faster and the reconstructed image is sparser. In fact, Laplace provides the sparsest reconstructed image among all methods. The CFDR method, although it is the fastest, has the worst reconstruction error, and the BP method, although it has the best reconstruction error, has the largest computation time. Laplace and CFAR are clearly the methods that should be preferred, having near-best reconstruction errors and smallest computation times, where CFAR being slightly faster and Laplace having slightly lower reconstruction error.

5. CONCLUSIONS

In this paper we formulated the compressing sensing problem from a Bayesian perspective, and presented a framework to simultaneously model and estimate the sparse signal coefficients. We proposed the use of a hierarchical form of Laplace priors on signal coefficients. We have shown that the relevance vector machine is a special case of our formulation, and that our hierarchical prior modeling provides solutions with a higher degree of sparsity and lower reconstruction errors. We presented a constructive (greedy) algorithm resulting from our framework which updates the signal coefficients sequentially in order to achieve low computation times and efficiency in practical problems. The proposed algorithm automatically estimates the model parameters solely from the observation and does not require user intervention unlike most existing methods. We demonstrated that overall, the proposed algorithm results in higher performance than most state-of-the-art algorithms.

Table 1. Average reconstruction errors, running times and number of nonzero components for multi-scale CS reconstruction of the *Mondrian* image.

	# Nonzeros	Time	Error
Linear	4096	-	0.13325
BP	4096	78.254	0.13933
CFAR	1139.2	13.88	0.14971
CFDR	2177.3	7.86	0.20867
BCS	1174.2	18.343	0.1443
Laplace	1078.7	15.372	0.1451

6. REFERENCES

[1] E.J. Candes, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[3] 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, 1999.

[4] M. Figueiredo, R. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Trans. on Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, Dec 2007.

[5] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inform. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

[6] D. L. Donoho and Y. Tsaig, "Sparse solution of underdetermined linear equations. by stagewise orthogonal matching pursuit," *Preprint*, March 2006.

[7] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Processing*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[8] M. W. Seeger and H. Nickisch, "Compressed sensing and Bayesian experimental design," in *ICML*, 2008, pp. 912–919.

[9] M. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, pp. 211–244, 2001.

[10] M. Figueiredo, "Adaptive sparseness for supervised learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1150–1159, 2003.

[11] D.P. Wipf and B.D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Processing*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.

[12] C. M. Bishop, *Pattern Recognition and Machine Learning*, Springer-Verlag, 2006.

[13] M. Tipping and A. Faul, "Fast marginal likelihood maximisation for sparse Bayesian models," in *9th International Workshop on AI and Statistics*, 2003.

[14] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Process.*, vol. 86, no. 3, pp. 549–571, 2006.