SPARSE IMAGE RECONSTRUCTION USING A SPARSE PRIOR

Michael Ting, Raviv Raich, Alfred O. Hero III

Electrical Engineering and Computer Science, University of Michigan
Ann Arbor, MI 48109-2122, USA
{mting, ravivr, hero}@umich.edu

ABSTRACT

Sparse image reconstruction is of interest in the fields of radioastronomy and molecular imaging. The observation is assumed to be a linear transformation of the image, and corrupted by additive white Gaussian noise (AWGN). We propose two methods of sparse image reconstruction which use the following sparse prior: a weighted average of an atom at zero and a Laplacian probability density function. The selection of the tuning parameters in the prior is done using an empirical Bayes approach. The two methods are compared with sparse Bayesian learning (SBL) and the estimator that maximizes the penalized likelihood with a l_1 penalty on the image values. In the latter method, we propose using Stein's unbiased risk estimator (SURE) to select the regularization parameter in a data-driven fashion. A simulation study was performed. The three sparse reconstruction methods proposed had better performance than SBL. The two methods based on the sparse prior previously mentioned had equal, if not better, performance than the l_1 penalty-based estimator when the image was highly sparse.

1. INTRODUCTION

In most image reconstruction problems, the images are not directly observable. Instead, one observes a transformed version of the image, possibly corrupted by noise. In the general case, the estimation of the image can be regarded as a simultaneous deconvolution and denoising problem. Intuitively, a better reconstruction can be obtained by incorporating knowledge of the image into the reconstruction algorithm.

In this paper, the images of interest to be reconstructed are assumed to be sparse. Sparse images appear in radioastronomy and molecular imaging. In the following two scenarios, prior information on the sparse image can be used in the estimation of the image: if one knew the number of non-zero voxels in the image; or, if one knew that the image was a realization of a sparse probability density function (p.d.f.). In a less informative scenario, one only knows that the image is sparse, but not the sparsity level nor the distribution of the non-zero voxels. We consider the model where the observation is a linear transformation of the image, and corrupted by AWGN. Note that the reconstruction methods mentioned here can be used to solve the sparse denoising problem with coloured Gaussian noise.

There are several existing methods that address the sparse image reconstruction problem. The first is sparse Bayesian learning (SBL) [1]. The image voxels are assumed to be independent, zero-mean Gaussian random variables (r.v.s), each with an unknown variance. The unknown variances in the image prior are learned empirically. The p.d.f. of the observation conditioned on the prior variances can be obtained in closed form. Then, marginal maximum likelihood (MML) is used to learn the prior variances. This empirical Bayes approach is a data-driven approach. The second existing method is the estimator formed by maximizing the penalized likelihood criterion with a l_1 penalty on the image values. The aforementioned error criterion is known to promote sparsity in the estimate [2]. This estimator shall be called the L1 estimator. For the L1 estimator, one has to choose a suitable regularization parameter.

Three reconstruction methods are proposed. Firstly, we propose using SURE [3] to select the regularization parameter for the L1 estimator in a data-driven fashion. This method will be referred to as L1-SURE. The derivation of the other two methods relies on the sparse prior used in the empirical Bayes denoising (EBD) method of [4], which is a weighted average of an atom at zero and a Laplacian p.d.f. Now, the deconvolution and denoising sub-problems can be separated using the EM algorithm as mentioned in [5]. EBD can then be applied to the denoising sub-problem; this method shall be referred to as E-EBD. E-EBD is not an EM algorithm: we have casually substituted EBD to solve the denoising subproblem without any initial cost function in mind. Thirdly, the sparse prior in [4] is used in the familiar MAP framework to derive the MAP estimator. This method will be called E-MAP. A simulation study was conducted comparing SBL, L1-SURE, E-EBD, and E-MAP.

2. PROBLEM FORMULATION

Denote the observation by $\underline{y}$, which typically corresponds to a 2-dimensional or 3-dimensional array. By enumerating the elements of the array lexicographically, one can equivalently represent the image by a vector. Without loss of generality, we shall take $\underline{y} \in \mathbb{R}^N$. Let $\underline{\theta}$ be the parameters of interest (e.g., the original image) that we would like to estimate from $\underline{y}$. Without loss of generality, let $\underline{\theta} \in \mathbb{R}^M$.

Consider the conditional p.d.f. of $\underline{y}$ given θ, i.e., $p(\underline{y}|\underline{\theta})$. Suppose that we would like to estimate $\underline{\theta}$ under the condition that it is sparse, i.e., most of the values of $\underline{\theta}_i$ are zero. In this paper, we model $\underline{y}$ as the result of a linear transformation of $\underline{\theta}$ with AWGN. Specifically,

$$\underline{y} = \mathbf{H}\underline{\theta} + \underline{w}, \quad \underline{w} \sim \mathcal{N}(\underline{w}; \underline{0}, \sigma^2\mathbf{I}), \tag{1}$$

where $\mathcal{N}(\cdot; \underline{\mu}, \mathbf{\Sigma})$ is the Gaussian density with mean $\underline{\mu}$ and covariance matrix $\mathbf{\Sigma}$; and $\mathbf{H} \in \mathbb{R}^{N \times M}$. The problem we consider is as follows: suppose that $\underline{y}$, $\mathbf{H}$, σ are known and model (1) is given. Knowing that $\underline{\theta}$ is sparse, how can $\underline{\theta}$ be optimally estimated?

If $\mathbf{H}$ had full column rank, $(\mathbf{H}^T\mathbf{H})$ would be invertible, and (1) could be re-written as

$$\underline{y}' = \underline{\theta} + \underline{w}', \quad \underline{w}' \sim \mathcal{N}(\underline{w}'; \underline{0}, \sigma^2\mathbf{H}^\dagger(\mathbf{H}^\dagger)^T) \tag{2}$$

where $\underline{y}' \triangleq \mathbf{H}^\dagger\underline{y}$, $\mathbf{H}^\dagger \triangleq (\mathbf{H}^T\mathbf{H})^{-1}\mathbf{H}^T$ is the pseudo-inverse of $\mathbf{H}$, and $\underline{w}' \triangleq \mathbf{H}^\dagger\underline{n}$ is coloured Gaussian noise. In this case, (1) would be equivalent to denoising a sparse $\underline{\theta}$ in coloured Gaussian noise.

3. DENOISING AND DECONVOLUTION

Two special cases of (1) have been studied in the past. Firstly, if $\mathbf{H} = \mathbf{I}$, the estimation of $\underline{\theta}$ in (1) would be a sparse denoising problem in AWGN [4, 6]. Secondly, if $\sigma = 0$, the problem reduces to the task of finding a sparse basis representation of $\underline{y}$ in terms of the columns of $\mathbf{H}$ [1, 2]. In the general case when $\mathbf{H} \neq \mathbf{I}$ and $\sigma \neq 0$, the estimation of $\underline{\theta}$ can be regarded as a simultaneous deconvolution and denoising problem.

The deconvolution and denoising subproblems can be separated in the context of finding the maximum a posteriori (MAP)/maximum penalized likelihood (MPL) estimate of $\underline{\theta}$, which is

$$\hat{\underline{\theta}} = \text{argmax}_{\underline{\theta}}(\log p(\underline{y}|\underline{\theta}) - \text{pen}(\underline{\theta})), \tag{3}$$

where $\text{pen}(\underline{\theta})$ is an arbitrary penalty function imposed on $\underline{\theta}$. Introduce an intermediate r.v. $\underline{z}$ so that $p(\underline{y}|\underline{z}, \underline{\theta}) = p(\underline{y}|\underline{z})$ and apply the Expectation-Maximization (EM) algorithm. In EM parlance, $\underline{z}$ is called the complete data. The authors in [5] selected $\underline{z} = \underline{\theta} + \alpha\underline{w}_1$, where $\underline{w}_1$ has the p.d.f. $\mathcal{N}(\underline{w}_1; \underline{0}, \mathbf{I})$. The quantity $\alpha \in \mathbb{R}$ must satisfy $\alpha^2 \leq \sigma^2/\rho(\mathbf{H}\mathbf{H}^T)$, where $\rho(\cdot)$ is the spectral radius function. Then, $\underline{y} = \mathbf{H}\underline{z} + \underline{w}_2$, where $\underline{w}_2 \sim \mathcal{N}(\underline{w}_2; \underline{0}, \sigma^2\mathbf{I} - \alpha^2\mathbf{H}\mathbf{H}^T)$. The following notation shall be used for the iterative estimates of $\hat{\underline{\theta}}$: $\hat{\underline{\theta}}^{(n)}$ shall denote the estimate at the n-th step, for $n \geq 0$. The initial condition is $\hat{\underline{\theta}}^{(0)} = \underline{\theta}^{(0)}$. The resulting EM iterations are

$$\hat{\underline{z}}^{(n)} = \hat{\underline{\theta}}^{(n)} + \frac{\alpha^2}{\sigma^2}\mathbf{H}^T(\underline{y} - \mathbf{H}\hat{\underline{\theta}}^{(n)}) \tag{4}$$

$$\hat{\underline{\theta}}^{(n+1)} = \text{argmax}_{\underline{\theta}}\left[-\frac{1}{2\alpha^2}\|\underline{\theta} - \hat{\underline{z}}^{(n)}\|_2^2 - \text{pen}(\underline{\theta})\right]. \tag{5}$$

Eqn. (4) can be regarded as a deconvolution step and (5) as a denoising step. The iterations can be more succinctly written as

$$\hat{\underline{\theta}}^{(n+1)} = \mathcal{D}\left(\hat{\underline{\theta}}^{(n)} + \frac{\alpha^2}{\sigma^2}(\underline{y} - \mathbf{H}\hat{\underline{\theta}}^{(n)})\right), \tag{6}$$

where $\mathcal{D}(\cdot)$ is a denoising operation that depends on the form of $\text{pen}(\cdot)$. We note that (6) is a Landweber iteration followed by a denoising step.

4. SPARSE PRIORS FOR $\underline{\theta}$

Several sparse priors have been used to model a sparse $\underline{\theta}$. In [6], the prior

$$\theta_i \overset{\text{i.i.d.}}{\sim} (1-w)\delta(\theta_i) + w\mathcal{N}(\theta_i; 0, \gamma) \tag{7}$$

was used, where i.i.d. denotes independent and identically distributed, and $\delta(\cdot)$ is the Dirac delta function. More recently, in [4], the prior

$$\theta_i \overset{\text{i.i.d.}}{\sim} (1-w)\delta(\theta_i) + w\gamma(\theta_i; a), \tag{8}$$

where $\gamma(x; a) = \frac{1}{2}ae^{-a|x|}$ is the Laplacian density with shape parameter a, was employed. In SBL [1],

$$\theta_i \sim \mathcal{N}(\theta_i; 0, \gamma_i). \tag{9}$$

However, the Gaussian density is not sparse, unless $\gamma_i = 0$. Note that the θ_i's are i.i.d. in (7) and (8). On the other hand, the θ_i's are independent but not identically distributed in (9).

The tuning parameters in the prior density, e.g., a, w in (8), are not known a priori. Assuming a Bayesian perspective on $\underline{\theta}$, the prior densities mentioned above might not be the true density for $\underline{\theta}$. That is, a mismatch for $p(\underline{\theta})$ is possible. The tuning parameters are important, as they should be selected so that the *assumed* density on $\underline{\theta}$ matches the *true* density as closely as possible.

The empirical Bayes approach, which we shall adopt in this paper, dictates that the tuning parameters be learned from the observation. The tuning parameters of the prior on $\underline{\theta}$ shall be called the *hyperparameters*. Let $\underline{\phi}$ be the vector of hyperparameters. Modelling $\underline{\theta}$ by a sparse prior with hyperparameters $\underline{\phi}$ introduces a hierarchical structure to the estimation problem. In [1, 4], the hyperparameters are learned via MML. In MML, one computes $p(\underline{y}|\underline{\phi}) = \int p(\underline{y}|\underline{\theta})p(\underline{\theta}|\underline{\phi})d\underline{\theta}$, and $\hat{\underline{\phi}} = \text{argmax}_{\underline{\phi}}p(\underline{y}|\underline{\phi})$. In both EBD and SBL, MML is used to compute $\hat{\underline{\phi}}$. Conveniently, the marginal likelihood can be computed in closed form for the priors (8) and (9). MML is not the only way that $\underline{\phi}$ can be learned: another alternative is to learn $\underline{\theta}, \underline{\phi}$ jointly in the MAP framework.

Once $\underline{\phi}$ is learned, it can be used to compute an estimate of $\underline{\theta}$. In the EBD method of [4], the posterior median is used to compute an estimate of $\underline{\theta}$. Recall that $\mathbf{H} = \mathbf{I}$ in the denoising subproblem. As the prior on $\underline{\theta}$ and the noise w is i.i.d., denoising of the M elements of $\underline{\theta}$ can be done on an element-by-element basis. The posterior median when $p(\theta_i)$ has the form (8) is a *thresholding rule* [4]. Let $T_1(\cdot; \underline{\phi}, \sigma) : \mathbb{R} \to \mathbb{R}$ denote the posterior median (thresholding rule) of EBD. A thresholding rule $T(\cdot; \underline{\phi}, \sigma)$ is said to have threshold t if $T(x; \underline{\phi}, \sigma) = 0$ iff $|x| \leq t$. In the non-trivial case, the posterior median will have threshold $t > 0$. The sparsifying effect is clear: any values of the observation with magnitude less than t will be set to zero. In SBL, the posterior mean is used. Unlike EBD, SBL can be used when $\mathbf{H} \neq \mathbf{I}$: it is a method that performs simultaneous deconvolution and denoising. We shall compare our proposed methods to SBL.

5. PROPOSED RECONSTRUCTION METHODS

Three methods for sparse image reconstruction are presented in this section. The first will use the EBD method of [4] as the denoising operation $\mathcal{D}(\cdot)$ in (6). This will be referred to as E-EBD. Note that E-EBD is not an EM implementation. Instead, it is an ad-hoc formulation that uses EBD as a sparse denoising operator. The iteration for E-EBD at the n-th step involves:

1. Compute $\hat{\underline{z}}^{(n)}$ according to (4)
2. Find $\hat{\underline{\phi}}^{(n)} = \text{argmax}_{\underline{\phi}}p(\hat{\underline{z}}^{(n)}|\underline{\phi})$, where $p(\underline{\theta}|\underline{\phi})$ is given in (8)
3. Set $\hat{\underline{\theta}}_i^{(n+1)} = T_1(\underline{z}_i^{(n)}; \hat{\underline{\phi}}^{(n)}, \alpha)$, for $i = 1, \ldots, M$

The second method will use the discrete-continuous version of the sparse prior (8), as the delta function is hard to work with in the MAP setting. Define the random variables $\tilde{\theta}_i$ and I_i such that $\theta_i = I_i\tilde{\theta}_i$, $1 \leq i \leq M$. The r.v.s $\tilde{\theta}_i$ and I_i are assumed to have the density

$$\tilde{\theta}_i \overset{\text{i.i.d.}}{\sim} \gamma(\tilde{\theta}_i; a), \tag{10}$$

$$I_i = \begin{cases} 0 & \text{with probability } (1-w) \\ 1 & \text{with probability } w \end{cases} \tag{11}$$

and $\tilde{\theta}_i$, I_i are independent. The estimation of $\underline{\phi}$ and $\underline{\theta}$ occurs jointly in the MAP framework. This method will be called E-MAP. The optimality criterion is

$$\hat{\underline{\tilde{\theta}}}, \hat{\underline{I}}, \hat{\underline{\phi}} = \underset{\underline{\tilde{\theta}}, \underline{I}, \underline{\phi}}{\text{argmax}} \ \log p(\underline{y}, \underline{\tilde{\theta}}, \underline{I}|\underline{\phi}). \tag{12}$$

The optimization of (12) is done using coordinate-wise maximization [7]. The maximizing $\underline{\xi}$ is obtained by alternately (i) maximizing $\underline{\phi}$ while holding $(\underline{\tilde{\theta}}, \underline{I})$ fixed, and (ii) maximizing $(\underline{\tilde{\theta}}, \underline{I})$ while holding $\underline{\phi}$ fixed. The maximization in step (i) is solvable in closed form as $\hat{a} = \|\underline{\hat{\theta}}\|_0/\|\underline{\hat{\theta}}\|_1$ and $\hat{w} = \|\underline{\hat{\theta}}\|_0/M$, where $\|\underline{x}\|_0 \triangleq \#\{i : x_i = 0\}$. The l_0 measure is not a norm; rather, it is a counting measure. Next, the maximization in step (ii) can be obtained by applying the EM algorithm with the complete data $\underline{z} = \underline{\theta} + \alpha \underline{w}_1$. The resulting iterations are as follows:

1. Compute $\underline{\hat{z}}^{(n)}$ according to (4)
2. Set $\underline{\hat{\theta}}_i^{(n+1)} = T_2(\underline{z}_i^{(n)}; \underline{\hat{\phi}}, \alpha)$, for $i = 1, \ldots, M$

where the thresholding rule $T_2(\cdot; \underline{\phi}, \sigma) : \mathbb{R} \to \mathbb{R}$ is

$$T_2(x; \underline{\phi}, \sigma) \triangleq \begin{cases} (x - \text{sgn}(x) a\sigma^2) I(|x| \geq t_1^{\text{m}}) & 0 \leq w < \frac{1}{2} \\ (x - \text{sgn}(x) a\sigma^2) I(|x| \geq t_2^{\text{m}}) & \frac{1}{2} \leq w \leq 1 \end{cases} \tag{13}$$

$$\text{and:} \quad t_1^{\text{m}} = a\sigma^2 + \sqrt{2\sigma^2 \log \frac{1-w}{w}}, \; t_2^{\text{m}} = a\sigma^2. \tag{14}$$

For $1/2 \leq w \leq 1$, T_2 is the soft-thresholding function. We note that there might be other methods to maximize the criterion in (12).

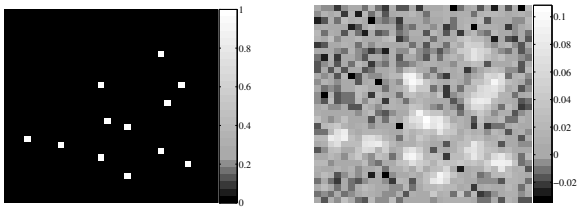
The third method is the MPL estimate with the l_1 penalty on $\underline{\theta}$, i.e., (3) with $\text{pen}(\underline{\theta}) = \beta\|\underline{\theta}\|_1$. The regularization parameter β is chosen using SURE [3]. This method will be called L1-SURE.

Of the three methods proposed in this section, the computational complexity can be ordered from highest to lowest as: E-EBD, L1-SURE, E-MAP. E-EBD requires a 2-dimensional search for $\hat{\underline{\phi}}^{(n)}$ in each iteration. In L1-SURE, a search in β is performed to minimize the SURE criterion. The SURE criterion also depends on $\underline{\hat{\theta}}(\beta)$. Finally, with E-MAP, the iterations are given above in closed form, and no miscellaneous search is needed. The memory requirement of the proposed algorithms is dominated in the general case by the storage of $\mathbf{H}$, which is of $\mathcal{O}(MN)$. Manipulations of vectors with dimensions of $\underline{y}$ and $\underline{\theta}$ involve memory of size $\mathcal{O}(\max(M, N))$. If $\mathbf{H}$ implements the convolution operator with a point spread function (psf), and the support of the psf is smaller than $\mathcal{O}(\max(M, N))$, then the overall memory requirement is $\mathcal{O}(\max(M, N))$. For more details, refer to [8].

6. SIMULATION RESULTS

The following four methods are compared in this section: E-EBD, E-MAP, SBL, and L1-SURE. The parameter $\underline{\theta}$ was set to a 32×32 binary image, i.e., the pixel values were either 0 or 1. $\mathbf{H}$ was taken to be a square matrix, i.e., $M = N$. It follows that $\underline{y}$ has the same length as $\underline{\theta}$. In particular, $\mathbf{H}$ implemented a Gaussian psf. The four reconstruction methods were tested under four different signal-to-noise ratio (SNR) values and two different sparsity levels. In this paper, the SNR is defined as $\text{SNR} \triangleq (M^{-1}\|\mathbf{H}\underline{\theta}\|_2^2)\sigma^{-2}$. The four SNR values examined were: 1.5, 2, 2.5, 3. The two different sparsity levels examined were: (1) 12 non-zero values, and (2) 36 non-zero values. These correspond to a non-zero percentage of approximately 1.2% and 3.5% respectively. The image $\underline{\theta}$ under the first sparsity level is depicted in Fig. 1(a), and the resulting $\underline{y}$ under $\text{SNR} = 3$ is depicted in Fig. 1(b).

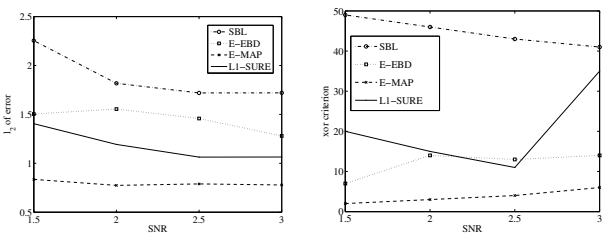
Define the reconstruction error $\delta\underline{\theta} \triangleq \underline{\hat{\theta}} - \underline{\theta}$. The performance of the reconstruction methods is evaluated by considering (i) $\|\delta\underline{\theta}\|_2$, and (ii) $\sum_i |I(|\hat{\theta}_i| < \delta) - I(\theta_i = 0)|$. The latter criterion will be referred to as the "xor error criterion". It measures the ability of the estimator to distinguish between the zero and non-zero values of $\underline{\theta}$. The value of δ in the xor error criterion will be set to $10^{-2}\|\underline{\theta}\|_\infty = 10^{-2}$.



(a) Image $\underline{\theta}$ (b) Noisy observation $\underline{y}$

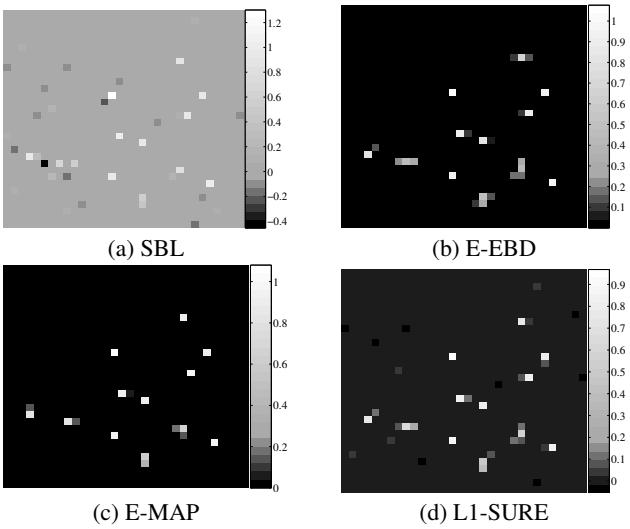
Fig. 1. Image $\underline{\theta}$ with 1.2% non-zero values and the resulting noisy observation $\underline{y}$ under $\text{SNR} = 3$.

Consider first the binary image with a non-zero percentage of 1.2%. The plot of the l_2 error for the four reconstruction methods is given in Fig. 2(a), and the xor error criterion in Fig. 2(b). Each point in the plots is the result of running a particular method on one realization of the noise w. This might explain why some curves are not monotone decreasing as the SNR increases, as one would intuitively expect. Despite the drawback of not having several noise realizations for each data point, which would reduce the variance, one can make some preliminary observations. There seems to be an ordering in the performance of the reconstruction methods that is maintained across the SNR values considered. Specifically, the performance of the methods in terms of the l_2 error can be ordered from best to worst as: E-MAP, L1-SURE, E-EBD, SBL. In terms of the xor error criterion, E-MAP has the best performance and SBL the worst. L1-SURE and E-EBD are comparable for the middle two SNR values, but the E-EBD is better for $\text{SNR} = 1.5$ and 3. The reconstructed images for $\text{SNR} = 3$ are given in Figs. 3(a)-(d). The best reconstruction is that of E-MAP, while the worst is that of SBL. In the latter, there are some negative voxels in $\underline{\hat{\theta}}$; as well, there are spurious non-zero voxels far away from the non-zero voxel locations of $\underline{\theta}$. E-EBD and L1-SURE have more blurring around the original non-zero voxel locations. The number of non-zero values in $\underline{\hat{\theta}}$ is plotted in Fig. 4 for the four reconstruction methods. The estimates $\underline{\hat{\theta}}$ that SBL produced in the simulations are not strictly sparse. Under the SNR values considered, SBL produced estimates $\underline{\hat{\theta}}$ such that $\|\underline{\hat{\theta}}\|_0 = 1024$. In contrast, the other three methods produced estimates that are sparser by at least an order of magnitude. It would be tempting to conjecture that this is due to the non-sparse Gaussian



(a) $\|\delta\underline{\theta}\|_2$ (b) xor error criterion

Fig. 2. Performance of reconstruction methods for the binary image with 1.2% non-zero values.



(a) SBL (b) E-EBD

(c) E-MAP (d) L1-SURE

Fig. 3. Reconstructed images for the binary image with 1.2% non-zero values and SNR $= 3$.

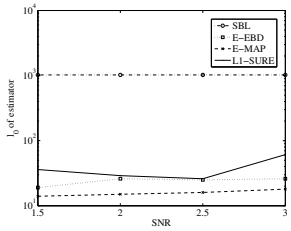
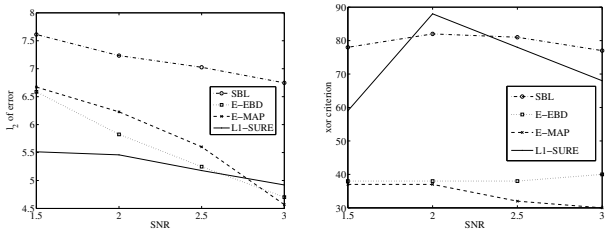


Fig. 4. Number of non-zero values in $\hat{\underline{\theta}}$ for the binary image with 1.2% non-zero values.

prior that SBL uses. However, the L1 estimator can be regarded as the MAP estimator when a Laplacian prior is placed on $\underline{\theta}$. Although the Laplacian prior is not sparse, the L1-SURE estimate is sparse. Note that if $\gamma_i = 0$ in (9), the corresponding SBL estimate $\hat{\theta}_i = 0$ w.p. 1 [1].

Next, consider the binary image with a non-zero percentage of 3.5%. The plot of the l_2 error and the xor error criterion is given in Fig. 5(a) and 5(b) respectively. Let us first focus on the l_2 reconstruc-



(a) $\|\delta\underline{\theta}\|_2$ (b) xor error criterion

Fig. 5. Performance of reconstruction methods for the binary image with 3.5% non-zero values.

tion error. At this sparsity level, the performance of the proposed reconstruction methods is not well ordered. However, they produced estimates with lower l_2 error than SBL. L1-SURE seems to perform better at low SNR than E-EBD and E-MAP. At higher SNR, L1-SURE, E-EBD, and E-MAP have comparable performance. From Fig. 5(b), one see that E-EBD and E-MAP possess a better ability to distinguish between the zero and non-zero locations of $\underline{\theta}$ as compared to SBL and L1-SURE.

7. CONCLUSION

This paper proposed two methods, E-EBD and E-MAP, of applying the sparse prior that appeared in [4], i.e. (8), to the sparse image reconstruction problem. The selection of the tuning parameters in the sparse prior was done using an empirical Bayes, i.e., data-driven, approach. The standard L1 estimator was extended to L1-SURE, where the regularization parameter was selected using SURE. L1-SURE is similar to E-EBD and E-MAP in that the regularization parameter, the frequentist analogy of the hyperparameter, is selected in a data-driven fashion.

These three methods, along with SBL, were investigated in a simulation study conducted on binary-valued $\underline{\theta}$. At the higher sparsity level, corresponding to a non-zero percentage of 1.2%, the performance of the four methods in terms of l_2 reconstruction error can be ordered from best to worst as: E-MAP, L1-SURE, E-EBD, and SBL. At the lower sparsity level, corresponding to a non-zero percentage of 3.5%, L1-SURE had a lower l_2 error than E-EBD and E-MAP for lower SNR. At the higher SNR values considered, L1-SURE, E-EBD, and E-MAP had comparable performance. In both sparsity levels, SBL had l_2 errors higher than the other three methods. In the simulation study, we noticed that the SBL estimate $\hat{\underline{\theta}}$ was never strictly sparse. In contrast, L1-SURE, E-EBD, and E-MAP produced estimates that were sparser. E-EBD and E-MAP were better able to distinguish between the zero and non-zero locations of the image than SBL and L1-SURE.

8. REFERENCES

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

[2] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via l^1 minimization," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 5, pp. 2197–2202, 2003.

[3] V. Solo, "A sure-fired way to choose smoothing parameters in ill-conditioned inverse problems," in *Proceedings of the 3rd IEEE Intl. Conf. on Image Processing*, 1996, vol. 3, pp. 89–92.

[4] I. M. Johnstone and B. W. Silverman, "Needles and straw in haystacks: Empirical Bayes estimates of possibly sparse sequences," *The Annals of Statistics*, vol. 32, no. 4, pp. 1594–1649, 2004.

[5] M. A. T. Figueiredo and R. D. Nowak, "An EM Algorithm for Wavelet-Based Image Restoration," *IEEE Trans. Image Processing*, vol. 12, no. 8, pp. 906–916, 2003.

[6] I. M. Johnstone and B. W. Silverman, "Empirical Bayes approaches to mixture problems and wavelet regression," Tech. Rep., Stanford University, 1998.

[7] J. A. Fessler, "Image Reconstruction: Algorithms and Analysis," Draft of book.

[8] M. Ting, R. Raich, and A. O. Hero, "A Bayesian approach to sparse image reconstruction," Draft of paper in preparation.