

# SPARSE IMAGE RECONSTRUCTION USING SPARSE PRIORS

Michael Ting, Raviv Raich, and 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. We study the usage of sparse priors in the empirical Bayes framework: it permits the selection of the hyperparameters of the prior in a data-driven fashion. Three sparse image reconstruction methods are proposed. A simulation study was performed using a binary-valued image and a Gaussian point spread function. In the range of signal to noise ratios considered, the proposed methods had better performance than sparse Bayesian learning (SBL).

**Index Terms**— sparse image reconstruction, empirical Bayes, Stein's unbiased risk estimator, sparse Bayesian learning, LASSO estimator

## 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 naturally, for example, in radioastronomy and molecular imaging. As well, a non-sparse image might have a sparse representation in some appropriate domain. For example, an image composed of several constant-valued areas will, upon spatial differentiation, become sparse. We consider the model where the observation is a linear transformation of the image, and corrupted by additive white Gaussian noise (AWGN).

There are several existing methods that address the sparse image reconstruction problem. The first is sparse Bayesian learning (SBL) [1]. The second existing method is the estimator formed by maximizing the penalized likelihood criterion with a  $l_1$  norm 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; it is also known as the LASSO estimator [3]. The L1 estimator can be regarded as the maximum a posteriori (MAP) solution when an i.i.d. Laplacian prior is used.

We seek to use sparse priors in the empirical Bayes framework to derive sparse image reconstruction methods. This framework permits the selection of the hyperparameters of the prior in a data-driven

fashion. It therefore has the element of adaptability: this is critical, as different images will have different sparsity levels and their non-zero values will have different histograms. Three reconstruction methods are proposed. Firstly, we propose using Stein's unbiased risk estimator (SURE) [4] to select the hyperparameter for the L1 estimator. The other two methods rely on the sparse prior used in the empirical Bayes denoising (EBD) method of [5], which is a weighted average of a Laplacian p.d.f. and an atom at zero (LAZE). Marginal maximum likelihood (MML) and maximum a posteriori (MAP) were used to learn the hyperparameter for these two other methods. A simulation study was conducted comparing the three proposed methods to SBL. For the range of signal to noise ratios (SNR) considered, the proposed methods have better performance than SBL.

## 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, take  $\underline{y} \in \mathbb{R}^N$ . Let  $\underline{\theta}$  be the parameters of interest (e.g., the original image) that one would like to estimate from  $\underline{y}$ . Again, without loss of generality, let  $\underline{\theta} \in \mathbb{R}^M$ .

Consider the conditional p.d.f. of  $\underline{y}$  given  $\underline{\theta}$ , 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  $\theta_i$  are zero. In this paper, a linear model for  $\underline{y}$  given by

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

is considered, where:  $\mathcal{N}(\cdot; \underline{\mu}, \underline{\Sigma})$  is the Gaussian density with mean  $\underline{\mu}$  and covariance matrix  $\underline{\Sigma}$ ; and  $\mathbf{H} \in \mathbb{R}^{N \times M}$ . The problem we consider is as follows. Suppose that  $\underline{y}$ ,  $\mathbf{H}$ ,  $\sigma$  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 written as

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

where  $\underline{y}' \triangleq \mathbf{H}^\dagger \underline{y}$ ,  $\mathbf{H}^\dagger \triangleq (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T$  is the pseudoinverse of  $\mathbf{H}$ , and  $\underline{w}' \triangleq \mathbf{H}^\dagger \underline{w}$  is coloured Gaussian noise. In this case, (1) would be equivalent to denoising a sparse  $\underline{\theta}$  in coloured Gaussian noise. When  $\mathbf{H}$  is an orthonormal matrix, the elements of  $\underline{w}'$  would be i.i.d. Gaussian. The problem is then reduced to a sparse denoising problem. In this case, by re-labelling  $\underline{y}'$  and  $\underline{w}'$  as  $\underline{y}$  and  $\underline{w}$  respectively, we can suppose without loss of generality that  $\mathbf{H} = \mathbf{I}$ .

\*This work was partially supported by the DARPA Mosaic program under ARO contract DAAD19-02-C-0055, and by the ARO MURI grant W911NF-05-1-0403.

### 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 [5]. 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

$$\hat{\underline{\theta}} = \underset{\underline{\theta}}{\operatorname{argmax}} (\log p(\underline{y}|\underline{\theta}) - \operatorname{pen}(\underline{\theta})), \quad (3)$$

where  $\operatorname{pen}(\underline{\theta})$  is a suitable 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 [6] 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

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

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

The norm  $\|\cdot\|$  without a subscript denotes the  $l_2$  norm. Equation (4) can be regarded as a deconvolution step (indeed, it is a Landweber iteration) and (5) as a denoising step. The iterations can be more succinctly written as

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

where  $\mathcal{D}(\cdot)$  is a denoising operation that depends on the form of  $\operatorname{pen}(\cdot)$ .

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

Several priors have been used to model a sparse  $\underline{\theta}$ . It is known that the  $l_1$  norm penalty in the MPL framework induces sparsity in  $\hat{\underline{\theta}}$  [2]. This corresponds to the  $\theta_i$ s being i.i.d. Laplacian r.v.s

$$\theta_i \stackrel{\text{i.i.d.}}{\sim} p_l(\theta_i; a), \quad p_l(x; a) \triangleq \frac{1}{2} a e^{-a|x|} \quad (7)$$

Even through the Laplacian density is a sparsifying prior, it does not explicitly model the assumption that the preponderance of  $\theta_i$ s are zero. In [5], this natural extension was applied. The following sparse prior for  $\underline{\theta}$  was considered

$$\theta_i \stackrel{\text{i.i.d.}}{\sim} (1-w)\delta(\theta_i) + wp_l(\theta_i; a), \quad (8)$$

where  $\delta(\cdot)$  is the Dirac delta function. Recall that density (8) is called the LAZE prior. In SBL, the  $\theta_i$ s were taken to be independent but *not* identically distributed Gaussian r.v.s [1].

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

One would not normally think of the Gaussian density as a sparsifying prior. However, when  $\gamma_i = 0$ , the  $i$ th element of the posterior mean  $E[\underline{\theta}|\underline{y}, \underline{\gamma}]$  is 0, where  $\underline{\gamma} \triangleq (\gamma_1, \dots, \gamma_M)^T$ .

### 5. ESTIMATION OF THE HYPERPARAMETER AND ORIGINAL IMAGE

The tuning parameters in the prior density, e.g.,  $a, w$  in (8), are not known a priori. Indeed, 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 tuning parameters of the prior on  $\underline{\theta}$  shall be called the *hyperparameters*. Let  $\underline{\phi}$  be the vector of hyperparameters.

We adopt a data-driven, empirical, approach to estimating the hyperparameter. This paper will discuss three ways.

**MML** One computes  $p(\underline{y}|\underline{\phi}) = \int p(\underline{y}, \underline{\theta}|\underline{\phi}) d\underline{\theta}$ , and  $\hat{\underline{\phi}} = \underset{\underline{\phi}}{\operatorname{argmax}} p(\underline{y}|\underline{\phi})$ . 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 prior (8) when  $\mathbf{H} = \mathbf{I}$  and for (9) with arbitrary  $\mathbf{H}$ .

**MAP** The variables  $\underline{\theta}$  and  $\underline{\phi}$  can be jointly estimated in the MAP framework as  $\hat{\underline{\theta}}, \hat{\underline{\phi}} = \underset{\underline{\theta}, \underline{\phi}}{\operatorname{argmax}} \log p(\underline{\theta}|\underline{y}, \underline{\phi})$ .

**SURE** Define the risk as  $R \triangleq E_Y [\|\mathbf{H}\underline{\theta} - \mathbf{H}\hat{\underline{\theta}}\|^2]$ . In order to compute the risk,  $\underline{\theta}$  is required. Stein's unbiased risk estimate  $\hat{R}$  is an unbiased estimate of  $R$  [4]. Select  $\underline{\phi}$  that minimizes  $\hat{R}$ , subject to perhaps a non-negativity criteria for each  $\phi_i$ . A more ideal risk would be  $E_Y [\|\underline{\theta} - \hat{\underline{\theta}}\|^2]$ ; however, one would not be able to compute SURE for this risk function for a general  $\mathbf{H}$ .

Once  $\underline{\phi}$  is learned, an estimate of  $\underline{\theta}$  can be formed, if it has not been jointly estimated with  $\underline{\phi}$ . In the EBD denoising method of [5], the posterior median is used to compute an estimate of  $\underline{\theta}$ . As the prior on  $\underline{\theta}$  and the noise  $w$  are i.i.d., denoising of the  $M$  elements of  $\underline{\theta}$  can be done on an element-by-element basis. When  $p(\underline{\theta})$  has the form (8) and  $\mathbf{H} = \mathbf{I}$ , the posterior median is a *thresholding rule* [5]. Let  $T_1(\cdot; \underline{\phi}, \sigma) : \mathbb{R} \rightarrow \mathbb{R}$  denote the posterior median (thresholding rule) induced by (8) when  $\mathbf{H} = \mathbf{I}$ . 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. Lastly, the MAP/MPL framework of (3) can be employed as well to form  $\hat{\underline{\theta}}$ .

### 6. PROPOSED RECONSTRUCTION METHODS

Three methods for sparse image reconstruction are proposed. The first will use the EBD denoising method of [5] as  $\mathcal{D}(\cdot)$  in (6). This will be referred to as EBD-LAZE. Note that EBD-LAZE is not an EM implementation. Instead, it is an ad-hoc formulation that uses EBD as a sparse denoising operator. The iteration for EBD-LAZE at the  $n$ th step is:

1. Compute  $\hat{\underline{z}}^{(n)}$  according to (4).
2. Find  $\hat{\underline{\phi}}^{(n)} = \underset{\underline{\phi}}{\operatorname{argmax}} p(\hat{\underline{z}}^{(n)}|\underline{\phi})$ , using (8) for  $p(\underline{\theta}|\underline{\phi})$ .
3. Set  $\hat{\underline{\theta}}_i^{(n+1)} = T_1(\hat{z}_i^{(n)}; \hat{\phi}_i^{(n)}, \alpha)$ , for  $i = 1, \dots, M$ .

The second method will use the discrete-continuous version of the LAZE sparse prior, 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

$\theta_i = I_i \tilde{\theta}_i$ ,  $1 \leq i \leq M$ .  $I_i$  is taken to be a Bernoulli r.v. with parameter  $p = w$ , i.e.,  $I_i = 1$  with probability  $w$  and 0 with probability  $1 - w$ . The r.v.  $\tilde{\theta}_i$  has the conditional p.d.f.

$$p(\tilde{\theta}_i | I_i) = \begin{cases} g(\tilde{\theta}_i) & I_i = 0 \\ p_i(\tilde{\theta}_i; a) & I_i = 1 \end{cases} \quad (10)$$

and  $(\tilde{\theta}_i, I_i)$  are assumed to be i.i.d. for  $1 \leq i \leq M$ . The density  $g(\cdot)$  is an arbitrary density independent of  $a$ . The estimation of  $\underline{\phi}$  and  $\underline{\theta}$  occurs jointly in the MAP framework. This method will be called MAP-LAZE. The optimality criterion is

$$\hat{\underline{\theta}}, \hat{\underline{I}}, \hat{\underline{\phi}} = \underset{\underline{\theta}, \underline{I}, \underline{\phi}}{\operatorname{argmax}} \log p(\tilde{\underline{\theta}}, \underline{I} | \underline{y}, \underline{\phi}). \quad (11)$$

The optimization of (11) is done using block coordinate-wise maximization. The maximizing  $(\hat{\underline{\theta}}, \hat{\underline{I}}, \hat{\underline{\phi}})$  is obtained by alternately (i) maximizing  $\underline{\phi}$  while holding  $(\hat{\underline{\theta}}, \hat{\underline{I}})$  fixed, and (ii) maximizing  $(\hat{\underline{\theta}}, \hat{\underline{I}})$  while holding  $\underline{\phi}$  fixed. The maximization in step (i) is solvable in closed form as

$$\hat{a} = \|\hat{\underline{\theta}}\|_0 / \|\hat{\underline{\theta}}\|_1 \text{ and } \hat{w} = \|\hat{\underline{\theta}}\|_0 / M \quad (12)$$

where  $\|\underline{x}\|_0 \triangleq \#\{i : x_i = 0\}$  is the  $l_0$  measure. 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

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

The thresholding rule  $T_2$  is defined as follows. Let  $g^* \triangleq \sup_x g(x)$  and  $r \triangleq g^* \frac{2}{\alpha} \frac{1-w}{w}$ . The thresholding rule  $T_2(\cdot; \underline{\phi}, \sigma) : \mathbb{R} \rightarrow \mathbb{R}$  is given by

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

where  $t_1^m = a\sigma^2 + \sqrt{2\sigma^2 \log r}$  and  $t_2^m = a\sigma^2$ . For  $0 \leq r < 1$ ,  $T_2$  is the soft-thresholding function. Other optimization techniques can be used to maximize the criterion in (11). Note that  $g^*$  is a tuning parameter that must be manually set.

The third method is the MAP estimate with the Laplacian prior, but with  $\underline{\phi}$  selected by minimizing the SURE criterion. The MAP estimate can be regarded as the MPL estimate with the  $l_1$  norm penalty on  $\underline{\theta}$ , i.e., (3) with  $\operatorname{pen}(\underline{\theta}) = \beta' \|\underline{\theta}\|_1$ ,  $\beta' \geq 0$ . Equivalently, by letting  $\beta = 2\sigma^2 \beta'$ , the MPL estimate is given by

$$\hat{\underline{\theta}} = \underset{\underline{\theta}}{\operatorname{argmin}} (\|\underline{y} - \mathbf{H}\underline{\theta}\|^2 + \beta \|\underline{\theta}\|_1). \quad (14)$$

If the columns of  $\mathbf{H}$  are linearly independent, SURE for the  $l_1$  regularization criterion is equal to

$$\hat{R}(\beta) = N\sigma^2 + \|\underline{y} - \mathbf{H}\hat{\underline{\theta}}(\beta)\|^2 + 2\sigma^2 \|\hat{\underline{\theta}}(\beta)\|_0 \quad (15)$$

where the dependence of  $\hat{\underline{\theta}}$  on  $\beta$  is explicitly noted [7]. The estimator corresponding to the non-negative  $\beta$  that minimizes  $\hat{R}(\beta)$  in (15) will be called the L1-SURE estimator. A SURE expression similar to (15) was derived in the case of a diagonal  $\mathbf{H}$  and where the  $l_1$  penalty was imposed on the coefficients of a 2-d wavelet transform of  $\underline{\theta}$  [8, (10)–(11)].

Of the three methods proposed in this section, the computational complexity can be ordered from highest to lowest as: EBD-LAZE,

L1-SURE, and MAP-LAZE. EBD-LAZE requires a 2-dimensional search for  $\hat{\underline{\phi}}$  in each iteration. This can be decreased by performing the search every  $m$ th iteration, if the hyperparameter estimates are not changing that rapidly. In L1-SURE, a search in  $\beta$  for  $\beta \geq 0$  is performed to minimize the SURE criterion. Least angle regression (LARS) can be used to efficiently compute the L1 estimator [9]. The columns of  $\mathbf{H}$  must be linearly independent in order to apply LARS; however, that is already assumed by the application of (15). Finally, with MAP-LAZE, the iterations are given above in closed form, and no search is needed. The proposed methods have less computational complexity than SBL [7].

## 7. SIMULATION STUDY

The following four methods are compared in this section: EBD-LAZE, MAP-LAZE, L1-SURE, and SBL. The parameter  $\underline{\theta}$  was set to a  $32 \times 32$  binary image, i.e., the pixel values were either 0 or 1. Due to a lack of space, we will not consider other possible  $\underline{\theta}$ , e.g., non-binary images.  $\mathbf{H}$  was taken to be a square matrix, i.e.,  $M = N = 32^2$ . In particular,  $\mathbf{H}$  implemented convolution with a Gaussian point spread function (psf). The columns of  $\mathbf{H}$  are linearly independent, so the SURE expression (15) can be used, and the L1-SURE estimator can be implemented using LARS.

The four reconstruction methods were tested under four different SNR values, where the SNR is defined as  $\operatorname{SNR} \triangleq (M^{-1} \|\mathbf{H}\underline{\theta}\|^2) \sigma^{-2}$ . The four SNR values examined were: 1.5, 2, 2.5, and 3. A wider range of SNRs are examined in [7]. A sparsity level of 12 non-zero values in  $\underline{\theta}$  was considered. This corresponded to approximately 1.2% of the pixel values of  $\underline{\theta}$  being non-zero. The image  $\underline{\theta}$  is depicted in Fig. 1(a), and a realization of  $\underline{y}$  under SNR = 3 is depicted in Fig. 1(b). The MAP-LAZE reconstruction method was used with

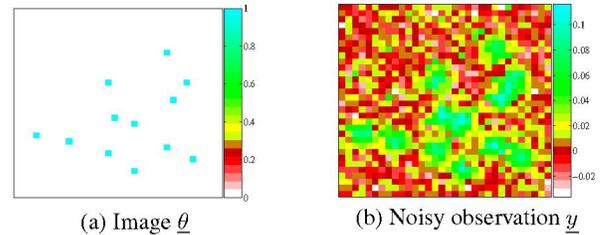


Fig. 1. Image  $\underline{\theta}$  and a realization of  $\underline{y}$  under SNR = 3.

$g^* = (\sqrt{2})^{-1}$ . Each reconstruction method was repeated with 30 realizations of the noise  $\underline{w}$  (except for MAP-LAZE under SNR = 1.5, where 100 realizations were used).

Define the reconstruction error  $\underline{e} \triangleq \underline{\theta} - \hat{\underline{\theta}}$ . The performance of the reconstruction methods is evaluated by considering (i)  $\|\underline{e}\|_2$ ; (ii)  $E_d(\underline{\theta}, \hat{\underline{\theta}}; \delta) \triangleq \sum_i |I(|\hat{\theta}_i| < \delta) - I(\theta_i = 0)|$ ; and (iii)  $\|\hat{\underline{\theta}}\|_0$ . Criterion (ii) is a detection error criterion. It measures the ability of the estimator to distinguish between the zero and non-zero values of  $\underline{\theta}$ . The value of  $\delta$  used was  $10^{-2} \|\underline{\theta}\|_\infty = 10^{-2}$ . For criteria (i) and (ii), smaller values are preferable. On the other hand, for criterion (iii), a value closer to  $\|\underline{\theta}\|_0 = 12$  is preferable.

The plot of the  $l_2$  reconstruction error is given in Fig. 2(a), the detection error criterion in Fig. 2(b), and the number of non-zero values of  $\hat{\underline{\theta}}$  in Fig. 2(c). The mean of each criterion and error bars of one standard deviation are plotted. The performance of the methods in terms of  $\|\underline{e}\|_2$  can be ordered from best to worst as: MAP-

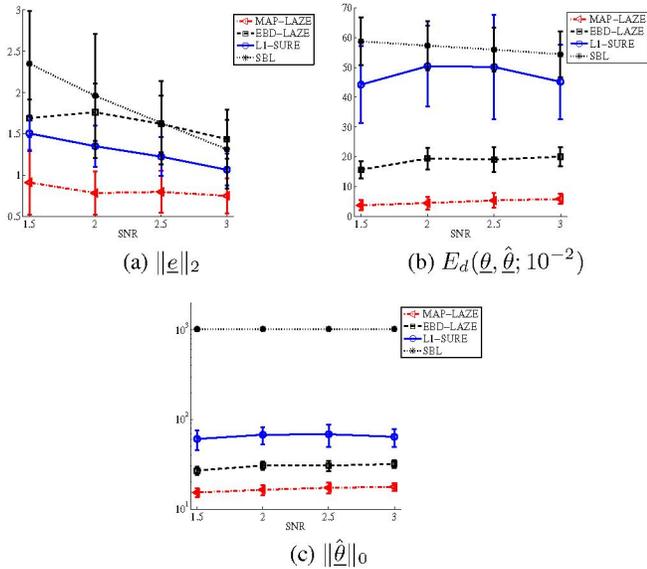


Fig. 2. Performance of the reconstruction methods.

LAZE, L1-SURE, EBD-LAZE, and SBL. In terms of the detection error criterion, EBD-LAZE has better performance than L1-SURE, giving rise to the ordering from best to worst as: MAP-LAZE, EBD-LAZE, L1-SURE, and SBL. The estimates  $\hat{\theta}$  that SBL produced in the simulations were not strictly sparse. Under the SNR values considered, SBL produced estimates  $\hat{\theta}$  such that  $\|\hat{\theta}\|_0 = 1024$ , i.e., none of the  $\hat{\theta}_i$ s were zero. In contrast, the other three methods produced estimates that were sparser by at least an order of magnitude.

Reconstructed images for one realization of the noise  $\underline{w}$  under SNR = 3 are given in Figs. 3(a)–(d). MAP-LAZE has the best reconstruction. While there is blurring around some of the non-zero pixels, there is less as compared to the other methods. EBD-LAZE’s  $\hat{\theta}$  contains noticeably more blurring, while the L1-SURE estimator more closely resembles MAP-LAZE’s  $\hat{\theta}$ . We notice that the SBL reconstructed image contains negative-valued pixels. While none of the reconstruction methods enforce positivity, there are no negative values in Figs. 3(b)–(d). SBL’s  $\hat{\theta}$  contains spurious non-zero pixels far away from the non-zero pixel locations of  $\theta$ . It was previously noted that SBL never produced a strictly sparse image. Indeed, the  $\hat{\theta}$  in Fig. 3a looks sparser than would be suggested by  $\|\hat{\theta}\|_0 = 1024$ . The reason is that many of non-zero  $\hat{\theta}_i$ s have small magnitudes that make them visually indistinguishable from zero.

## 8. CONCLUSION

Three priors were examined for the purposes of forming an empirical Bayes estimator of sparse images. It was assumed that the parameter vector  $\theta$  was distributed according to one of the following p.d.f.s: i.i.d. Laplacian, i.i.d. LAZE, independent but not identical Gaussian. The choice of prior is but one aspect that contributes to final form of the sparse image reconstruction method. The two other aspects that play an important role are: method of learning the hyperparameter and method of forming the estimator  $\hat{\theta}$ . Several possibilities were mentioned for each. Three sparse image reconstruction methods were proposed based on the i.i.d. Laplacian and i.i.d. LAZE priors. In a simulation study, these were compared to SBL, which is based

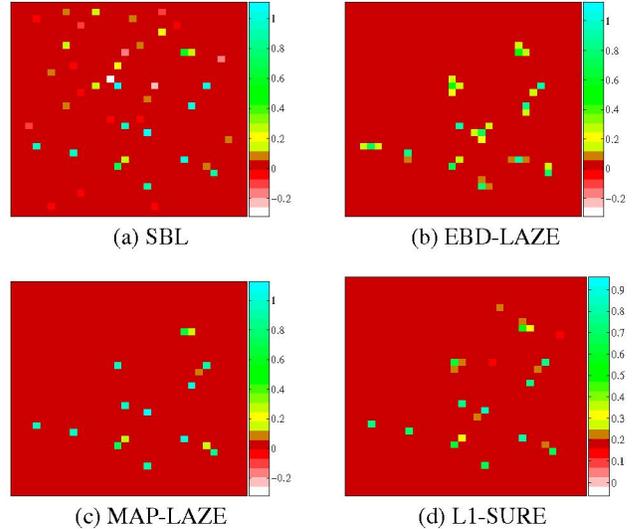


Fig. 3. Reconstructed images for a realization of  $\underline{y}$  under SNR = 3.

on the Gaussian prior. In the range of SNR = 1.5 to 3, the proposed methods demonstrated benefits over SBL. The MAP-LAZE estimator in particular had the best performance in terms of the three error criteria considered. MAP-LAZE is also the least computationally complex of the four methods; it is therefore an attractive candidate for future study. It was noted that SBL produced estimates that were not strictly sparse.

## 9. 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] R. Gribonval and M. Nielsen, “Sparse representations in unions of bases,” *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, 2003.
- [3] R. Tibshirani, “Regression shrinkage and selection via the lasso,” *Journal of the Royal Statistical Society, Series B*, vol. 58, no. 1, pp. 267–288, 1996.
- [4] C. M. Stein, “Estimation of the mean of a multivariate normal distribution,” *The Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [5] 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.
- [6] 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.
- [7] M. Ting, R. Raich, and A. O. Hero, “Empirical Bayes methods of sparse image reconstruction,” Draft of paper in preparation.
- [8] L. Ng and V. Solo, “Optical flow estimation using adaptive wavelet zeroing,” in *Proceedings of the IEEE Intl. Conf. on Image Processing*, 1999, vol. 3, pp. 722–726.
- [9] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, “Least angle regression,” *The Annals of Statistics*, vol. 32, no. 2, pp. 407–499, 2004.