Clustered Compressed Sensing in fMRI Data Analysis Using a Bayesian Framework

Solomon A. Tesfamicael and Faraz Barzideh

Abstract—**This paper provides a Bayesian method of analyzing functional magnetic resonance imaging (fMRI) data. Usually fMRI signals are noisy and need efficient algorithms to estimate or detect the signals accurately. Using a Bayesian frame work we have used two different priors: sparsity and clusterdness in the fMRI data by using a general linear model (GLM), which is used as a main tool in fMRI studies. This enhances the effectiveness of the model to help analyze the data better. So in this work we have built the Bayesian framework needed first. Then, we have applied our analysis on synthetic data that we made and that are well known, and the results show that clustered compressive sampling has given better results compared to the using of only sparse prior and/ or to the analysis without considering the two priors. Later we have applied it on fMRI data and the results are much better in terms of signal to noise ratio (SNR) and intensity of images.**

Index Terms—**Bayesian framework, sparse prior, clustered prior, posterior, MAP, compressive sensing, LASSO, clustered LASSO, GLM, fMRI data.**

I. Introduction

Functional magnetic resonance imaging (fMRI), a non-invasive technique of brain mapping, has become crucial in the study of brain activity. It measures brain activity as a function of local vasodilatations, which are characterized by hemodynamic response functions (HRF) [1], [2]. FMRI is based on the blood oxygenated level development (BOLD) signal, since neuronal activation is accompanied by a change in the ratio of the oxyhemoglobin to deoxyhemoglobin in the part of the brain that is related to a particular activity [1]. The purpose of fMRI data analysis is to estimate, detect or predict the BOLD signals, which are basically weak, from the noisy data so that one can identify the activated voxels in the brain regions in contrast to the non-activated ones as a function of some stimulus or task.

The methods for analyzing fMRI data can generally be divided into two groups. The first group is model based methods where one assumes models, such as the General linear model (GLM). Traditionally, GLM has been used as a statistical tool to analyze fMRI data [3]. It uses prior knowledge and assumes a canonical hemodynamic response function (HRF) to model a BOLD signal. GLM has been implemented using a classical inference method called Statistical Parametric Mapping (SPM) [3]. But one of the limitations of GLM is that it ignores correlation between voxels in time and space, generally present in fMRI data. The second group of methods of fMRI data analysis is data dependent and assumes no model of HRF. Instead these methods try to consider the correlation in time and/or space of the fMRI data. Among them are the principal component analysis (PCA) [4]-[6], independent component analysis (ICA) [7], [8] and clustering [9]-[18].

In order to estimate the parameters of GLM one can use either the classical or the Bayesian approach [19]-[25]. The latter have been applied to fMRI data inference [26], [27] and it is also possible to compare both as in [28], [29]. Bayesian approach is our main focus in this paper.

Under Bayesian inference it is possible to assume different priors to signals under considerations. Taking a sparse prior as the first assumption on the fMRI signals we incorporate a compressive sampling or compressive sensing framework in our inference on the fMRI data.

Compressive sensing (CS) is a new methodology to capture signals at lower rate than the Nyquist sampling rate when the signals are sparse or sparse in some domain [30]-[33]. In [34] the authors have used the sparsity of magnetic resonance imaging (MRI) signals and showed that this can be exploited to significantly reduce scan time, or alternatively, improve the resolution of MR imagery. In addition, the sparsity concept under a Bayesian framework has been used to determine the design matrix (dimension reduction in space) for fMRI data analysis [35], [36].

Besides the sparse priors, clusters (or structures on the patterns of sparsity) can also be considered as additional priors [37]. Our contribution in this work is to use the Bayesian framework and incorporate two different priors in order to analyze fMRI data and compare different algorithms. Specifically, we assume a clustered prior in addition to the sparse prior in order to remedy the limitations of GLM.

The paper is organized as follows. In Section II we present our basis for analysis and inference that is the Bayesian framework, and based on that we show how the sparse and clustered priors are used for the linear model presented in the paper. Section III shows our results using synthetic and real fMRI data, and Section IV presents conclusion and future work.

II. Bayesian Framework for Sparse and Clustered Signals

Beginning with a given vector of measurements and measurement matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$, assuming noisy measurement with $\boldsymbol{w} \in \mathbb{R}^{M}$ being i.i.d. Gaussian random

Manuscript received October 3, 2013; revised December 9, 2013. This work was supported in part by the Sør-Trondlag University College (HIST) and the department of Electronics and Telecommunication (IET) at Norwegian University of Science and Technology (NTNU).

S. A. Tesfamicael is with Sør-Trondlag University College (HIST-ALT). He is also with the department of Electronics and Telecommunication (IET) at the Norwegian University of Science and Technology (NTNU), Trondheim, Norway (e-mail: solomont@hist.no, tesfamic@iet.ntnu.no).

F. Barzideh is with the Norwegian University of Science and Technology (NTNU), Trondheim, Norway (e-mail: faraz.barzideh@gmail.com).

variables with zero mean and covariance matrix $\sigma^2\mathbf{I}$, estimating the sparse vector $\boldsymbol{x} \in \mathbb{R}^N$ is the problem that we are considering given the linear model

$$\boldsymbol{y} = \boldsymbol{A}\boldsymbol{x} + \boldsymbol{w}. \tag{1}$$

Here $N \gg M$ and $N \gg k$, where k is the number of non-zero entries in $\boldsymbol{x}$. Note that (1) has the same shape as the GLM used in fMRI data analysis [26].

Various methods for estimating $\mathbf{x}$ may be used. We have the least square (LS) estimator in which no prior information is applied:

$$\hat{\boldsymbol{x}} = \left(\boldsymbol{A}^T\boldsymbol{A}\right)^{-1}\boldsymbol{A}^T\boldsymbol{y}, \tag{2}$$

which performs very badly for the CS estimation problem we are considering. Another approach to estimate $\boldsymbol{x}$ is via the solution of the unconstrained optimization problem

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + u\, f(\boldsymbol{x}), \tag{3}$$

where $uf(\boldsymbol{x})$ is a regularizing term, for some non-negative u. If $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_p$, emphasis is made on a solution with LP norm, and $\|x\|_p$ is denotes a penalizing norm. When $p = 2$, we get

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + u\|\boldsymbol{x}\|_2 . \tag{4}$$

This is penalizing the least square error by the L2 norm and this performs badly as well, since it does not introduce sparsity into the problem. When $p = 0$, we get the L0 norm, which is defined as

$$\|\boldsymbol{x}\|_0 = k \equiv \neq \{i \in \{1, 2, 3, \ldots, N\} | x_i \neq 0\} \tag{5}$$

the number of the non-zero entries of x, which actually is a partial norm since it does not satisfy the triangle inequality property, but can be treated as norm by defining it as in [38], and get the L0 norm regularizing estimator

$$\hat{\boldsymbol{x}} = min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + u\|\boldsymbol{x}\|_0 , \tag{6}$$

which gives the best solution for the problem at hand since it favors sparsity in $\boldsymbol{x}$. Nonetheless, it is an NP-hard combinatorial problem. Instead, it has been a practice that one approximates it using L1 penalizing norm to get the estimator

$$\hat{\boldsymbol{x}} = \min_{\boldsymbol{x} \in \mathbb{R}^N} \frac{1}{2} \|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}\|_2^2 + u\|\boldsymbol{x}\|_1 , \tag{7}$$

which is a convex approximation to the L0 penalizing solution (6). These estimators, (4), (6) and (7), can equivalently be presented as solutions to constrained optimization problem [30]-[33], and in the CS literature there are many different types of algorithms to implement them. A very popular one is the L1 penalized L2 minimization called LASSO (Least Absolute Shrinkage and Selection Operator), which we later will present it in Bayesian framework. So first we present what a Bayesian approach is and come back to the problem at hand.

A. Bayesian Framework

Under Bayesian inference consider two random variables $\boldsymbol{x}$ and $\boldsymbol{y}$ with probability density function (pdf) $p(\boldsymbol{x})$ and $p(\boldsymbol{y})$, respectively. The product rule gives us

$$p(\boldsymbol{x}, \boldsymbol{y}) = p(\boldsymbol{x}|\boldsymbol{y})p(\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}) \tag{8}$$

and Bayes' theorem gives

$$p(\boldsymbol{x}|\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})/p(\boldsymbol{y}), \tag{9}$$

where $p(\boldsymbol{y}|\boldsymbol{x})$ is the likelihood function derived from the measurements $\boldsymbol{y}$, $p(\boldsymbol{x})$ is the prior distribution for $\boldsymbol{x}$ and $p(\boldsymbol{x}|\boldsymbol{y})$ is the posterior distribution of $\boldsymbol{x}$ given $\boldsymbol{y}$. However, for the estimation of $\boldsymbol{x}$ the denominator, which is called the evidence, has no relevance. Hence, the posterior is proportional to the product of the Likelihood function and the prior density:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x}). \tag{10}$$

Equation (10) is called Updating Rule [19], [20], in which the data allows us to update our prior views about $\boldsymbol{x}$. And as a result we get the posterior which combines both the data and non-data information of $\boldsymbol{x}$. As an example for a binomial trials let's have a beta distribution as a prior, gamma distribution for the data, as a result we get posterior distribution which is also a beta distribution. Fig. 1 shows that the posterior density is taller and narrower than the prior density. It therefore favors strongly a smaller range of $\boldsymbol{x}$ values, reflecting the fact that we now have more information.

Further, the Maximum a posterior (MAP), $\hat{\boldsymbol{x}}_{MP}$, is defined as

$$\begin{aligned}\hat{\boldsymbol{x}}_{MP} &= arg\, max_{\boldsymbol{x}} \frac{p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})}{\int_{\tilde{\boldsymbol{x}}} p(\boldsymbol{y}|\tilde{\boldsymbol{x}})p(\tilde{\boldsymbol{x}})d\tilde{\boldsymbol{x}}} \\ &= \arg max_{\boldsymbol{x}}\, p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})\end{aligned} \tag{11}$$

MAP is related to Fisher's methods of Maximum Likelihood Estimation (MLE), $\hat{\boldsymbol{x}}_{ML}$:

$$\hat{\boldsymbol{x}}_{ML} = arg\, max_{\boldsymbol{x}}\, p(\boldsymbol{y}|\boldsymbol{x}). \tag{12}$$

As we can see it from (11) and (12) the difference between MAP and MLE is the prior distribution. The former can be considered as a regularized form of the latter.

Inference based on the posterior may not be easy task since it involves multiple integrations, which may be cumbersome to perform. However, it can be computed in several ways: Numerical optimization (like Conjugate gradient method, Newton method), modification of an expectation-maximization algorithm and others. To proceed further, we will to assume prior distributions on $\boldsymbol{x}$.

B. Sparse Prior

The estimators of $\boldsymbol{x}$ resulting from (3) for the sparse problem we consider in this paper given by, (4), (6), and (7), can be presented as a maximum a posteriori (MAP) estimator under the Bayesian framework as in [38].

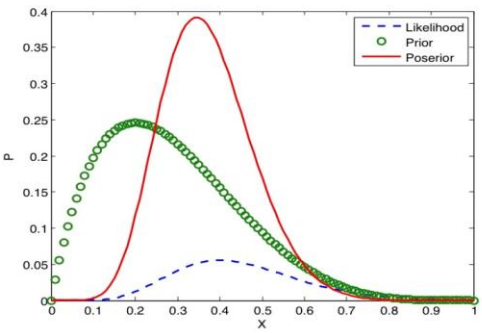

Fig. 1. Figure showing the updating rule: The posterior synthesizes and compromises by favoring values between the maximum of the prior density and likelihood. The prior we had is challenged to shift by the arrival of little amount of data.

We show this by defining a prior probability distribution for $\boldsymbol{x}$ on the form

$$p(\boldsymbol{x}) = \frac{exp(-u\, f(x))}{\int_{x\in\mathbb{R}^N} exp(-u\, f(x))dx} \tag{13}$$

where the regularizing function $f:\chi \rightarrow \mathbb{R}$ is some scalar valued, non-negative function with $\chi \subseteq \mathbb{R}$ which can be expanded to a vector argument by

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} f(x_i) \tag{14}$$

such that for sufficiently large u, $\int_{\mathbf{x}\in\mathbb{R}^N} \exp\left(-u\, f(\boldsymbol{x})\right) d\boldsymbol{x}$ is finite. Further, let the assumed variance of the noise be given by

$$\sigma^2 = \lambda/u, \tag{15}$$

where λ is system parameter which can be taken as $\lambda = \sigma^2 u$. Note that the prior, (13), is defined in such a way that it can incorporate the different estimators considered above by assuming different penalizing terms via $f(\boldsymbol{x})$ [38].

Since the pdf of the noise $\boldsymbol{w}$ is Gaussian, the likelihood function of $\boldsymbol{y}$ given $\boldsymbol{x}$ is given by

$$p_{\mathbf{y}|\mathbf{x}}(\boldsymbol{y}|\boldsymbol{x}) = \frac{1}{(2\pi\sigma)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\left\|\boldsymbol{y} - \boldsymbol{Ax}\right\|_2^2\right) \tag{16}$$

Together with (9) and (13), this now gives

$$p_{\mathbf{x}|\mathbf{y}}(\boldsymbol{x}|\boldsymbol{y}; A) = \frac{\exp\left(-u\left(\frac{1}{2}\|y - Ax\|_2^2 + \lambda f(x)\right)\right)}{(2\pi\sigma)^2 \int_{\mathbf{x}\in\mathbb{R}^N} \exp\left(-u\left(\frac{1}{2\lambda}\|y - Ax\|_2^2 + \lambda f(x)\right)\right)d\mathbf{x}}. \tag{17}$$

(17) is identical to the MAP estimator (11) is now found as

$$\hat{\boldsymbol{x}}_{MP} = arg\, min_{x\in\mathbb{R}^N} \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{Ax}\right\|_2^2 + \lambda f(\boldsymbol{x}). \tag{18}$$

(18) is identical to (3). Now, as we choose different regularizing function which enforces sparsity into the vector $\mathbf{x}$, we get different estimators listed below [38].

- Linear Estimators: when $f\,(\boldsymbol{x}) = \|\boldsymbol{x}\|_2^2$, (18) reduces to

$$\hat{\boldsymbol{x}}_{Linear} = \boldsymbol{A}^T(\boldsymbol{AA}^T + \lambda\boldsymbol{I})^{-1}\boldsymbol{y}, \tag{19}$$

which is the LMMSE estimator. But we ignore this estimator in our analysis since the following two estimators are more interesting for CS problems.

- LASSO Estimator: when $f\,(\boldsymbol{x}) = \|\boldsymbol{x}\|_1$, we get the LASSO estimator and (18) becomes,

$$\hat{\boldsymbol{x}}_{LASSO} = arg\, min_{x\in\mathbb{R}^N} \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{Ax}\right\|_2^2 + \lambda\|\boldsymbol{x}\|_1, \tag{20}$$

C. Which Is the Same as (II.6).

- Zero-Norm regularization estimator: when $f(\boldsymbol{x}) = \|\boldsymbol{x}\|_0$, we get zero norm regularization estimator (6) and (18) becomes

$$\hat{\boldsymbol{x}}_{Zero-Norm} = arg\, min_{x\in\mathbb{R}^N} \frac{1}{2}\left\|\boldsymbol{y} - \boldsymbol{Ax}\right\|_2^2 + \lambda\|\boldsymbol{x}\|_0. \tag{21}$$

As mentioned earlier, this is the best solution for estimation of the sparse vector $\boldsymbol{x}$, but is NP-complete. The worst approximation for the sparse problem considered is the L2- regularization solution given by (19). However, the best approximation is given by the equation (20) and its equivalent forms. We have used some of the algorithms in literature in our simulation which are considered as equivalent to this approximations such as Bayesian Compressive sensing (BCS) [39] and L1-norm regularized least-squares (L1-LS) [30]-[33].

D. Clustering Prior

Building on the Bayesian philosophy, we can further assume another prior distribution for clustering. The entries of the sparse vector $\boldsymbol{x}$ may have some structure that can be represented using distributions. In [37] a hierarchical Bayesian generative model for sparse signals is found in which they have applied full Bayesian analysis by assuming prior distributions to each parameter appearing in the analysis. We follow a different approach. Instead we use another penalizing parameter γ to represent clusterdness in the data. For that we define the clustering using the distance between the entries of the sparse vector x by

$$D \equiv \sum_{i=2}^{N} |x_i - x_{i-1}|, \tag{22}$$

and we use a regularizing parameter γ. Hence, we define the clustering prior to be

$$q(\mathbf{x}) = \frac{\exp(-\gamma\, D(\mathbf{x}))}{\int_{\mathbf{x}\in\mathbb{R}^N} \exp(-D\, f(\mathbf{x}))d\mathbf{x}} \ . \tag{23}$$

The new posterior evolving this prior under the Bayesian framework is proportional to the product of the three pdf's:

$$p(\boldsymbol{x}|\boldsymbol{y}) \propto p(\boldsymbol{y}|\boldsymbol{x})p(\boldsymbol{x})q(\boldsymbol{x}). \tag{24}$$

By similar arguments as used in section II.B, we arrive at the Clustered LASSO estimator

$$\hat{\boldsymbol{x}}_{Clu-LASSO} = \arg\min_{x\in\mathbb{R}^N} \frac{1}{2}\|\boldsymbol{y} - \boldsymbol{Ax}\|_2^2 + \lambda\|\boldsymbol{x}\|_1 + \gamma\sum_{i=2}^{N}|x_i - x_{i-1}| , \tag{25}$$

where λ, γ are our tuning parameters for the sparsity in $\boldsymbol{x}$ and the way the entries are clustered, respectively.

III. Results

A. Synthetic Data

In order to demonstrate the performance of the analysis presented in the paper we have used synthetic data in which we know the sparsity and the structure of the data so that we may incorporate these priors in the simulations. Unless otherwise stated we have used $\sigma^2 = 1, u = 0.01, \lambda = 0.01$, and random matrix A with Gaussian entries with variance σ^2, in our simulations.

We consider first a random clustered signal with $k = 4$, $N = 300$, $M = 200$, $k/N = 0.14$, $k/M = 0.28$, and $M/N = 0.66$. As we see it from Fig. 2, clustering has a large effect on the reconstruction and we get high SNR for Clustered LASSO about 29dB, which is better than the lasso, 13dB. Further, as we see in Fig. 3, we also compare the different algorithms used for the L_p regularization algorithms and find that the LMMSE (where L2 regularization is used) performs badly compared to the families of algorithms which uses L1 regularization, such as BCS, L1-LS and LASSO. Using $k = 4$, $N = 300$, $M = 130$, $k/N = 0.2$ and, $k/M = 0.32308$, $M/N = 0.43333$. we have the following result as shown in Table I.

TABLE I: PERFORMANCE COMPARISON

Algorithms	SNR in dB
LS	0.004179
LMMSE	-0.1466
BCS	31.7758
L1-LS	20.9524
LASSO	22.206
Clustered LASSO	37.6402

Before we apply the analysis presented above on fMRI data we used two other synthetic data sets. The first one is a two level image both sparse and clustered in space domain. We vectorize it in order to analyze it in our framework. The second one is a known medical related image Shepp-Logan phantom, it is also sparse and clustered. The results are shown in Fig. 4 and Fig. 5, respectively. We see the impact of our priors: sparse and clustered, where clustering has better effect than only using sparse prior when the signal has such structure in both figures. And we used most of the algorithms mentioned above to compare their performance. In addition, we have plotted in Fig. 6 and Fig. 7, their SNR against the amount of sparsity k/N and measurement ratio M/N, respectively. In the simulation for Fig. 4 we used $N = 200$, $M = 100$, $\sigma^2 = 0.1$, $k = 122$ and $M/N = 0.5$, whereas for Fig. 5 we used $N = 200$, $M = 92$, $\sigma^2 = 0.1$, $k = 80$ and $M/N = 0.46$. As we can see from the figures, the performance of LASSO and Clustered LASSO are much better than the LMMSE or LS.

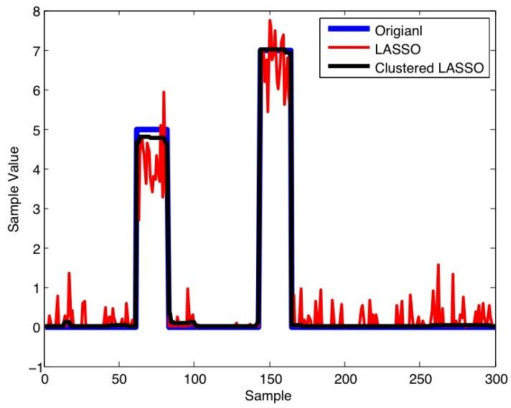

Fig. 2: Comparison of a signal reconstruction using sparse prior only and using additional clustered prior.

In addition we see another property, in Fig. 6 and Fig. 7, which is called phase transition, which shows the limits that one can under sample a sparse signal and yet reconstruct it accurately [40] - [42]. From the plots one also can see the phase transition property of the different algorithms used in CS in relation to the ratio of sparsity (k/N) and amount of measurement (M/N).

B. Sparse and Clustered fMRI Data Analysis

The raw data which is found by convolving a stimulus/task function and a hemodynamic response function (HRF) is not directly analyzed in our statistical analysis method. In general, fMRI data undergoes preprocessing in order to improve the signal quality [3], [37], [43]. This preprocessing include: artifact detection, baseline correction, movement correction, and image restoration. After these preprocessing the statistical analysis comes. At least this how it is done in many literatures mentioned in this paper. We have the advantage of using real fMRI data from the fMRI group in Trondheim and for the time being we took the preprocessed data from them and we have applied the analysis above on these data sets.

For a given measurement vector of fMRI signal $y \in R^M$ and a measurement matrix $\boldsymbol{A} \in R^{M\times N}$, and a norma distributed error vector $\boldsymbol{w} \in \mathrm{R}^{\mathrm{M}}$, using the GLM given by equation (1) as [3]. Since GLM do not account to the activations in the brain part which are not linearly correlated with the measurement matrix we invoke the sparsity and the special structures in the activation regions in the brain as a response to a stimulus or activity.

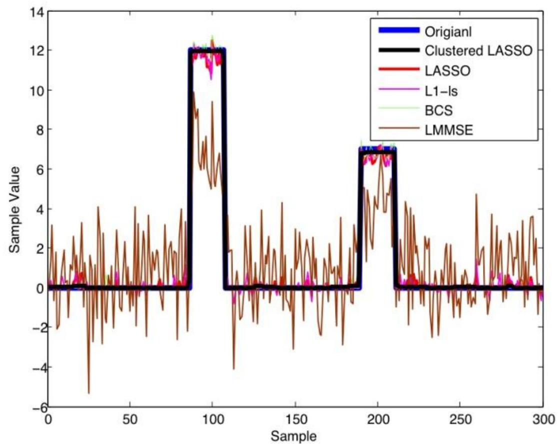

Fig. 3. Comparison of a signal reconstruction using sparse prior only and using additional clustered prior.

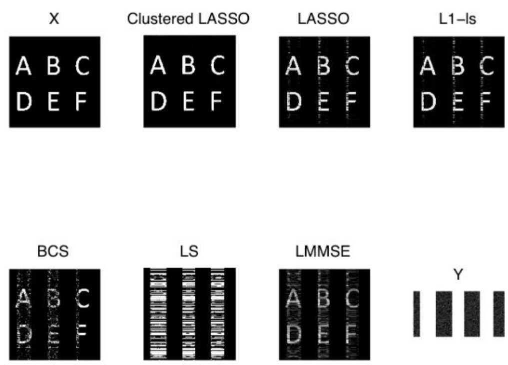

Fig. 4. Application of sparse and cluster prior on a synthetic data reconstruction.

From the fMRI data we took many slices and we saw how these data sets are sparse in the Fourier domain. This is shown in Fig. 8. We observed the whole data in this domain for the whole brain image. They all share the characteristics we have based our analysis, i.e., sparsity and clustredness. Then we took some slices which are connective in the slice order and we took different N and k on these slices. We can see the numbers at the top of Fig. 9 and Fig. 10, in which the two numbers represent k and N, respectively. In Fig. 9 we have higher sparsity, whereas in Fig. 10, we have a lower sparsity. In here, we used $N = 80, M = 2k$, k is given in the image, $\sigma^2 = 0.1$ and $\lambda = 0.1$. And as we can see these groups vary in intensity and quality. And in both cases we have the clustered LASSO and also LASSO which are much better than the others. In these figures, as in section III.A, we have used the different algorithms like L1-LS, BCS, LASSO, and clustered LASSO. Since the data has sparse representation in k-space, we invoke our framework in the paper, that is, we used sparse and cluster priors. The results show a significant improvement using the Bayesian framework analysis. In fMRI results are compared using image intensity and that gives a good ground for a health practitioner to observe and decide in accordance to the available information. The more one have prior knowledge on how the brain regions work in human beings or pets the better priors that one incorporate to analyze the data. So it this is open problem for researchers in the future, in general.

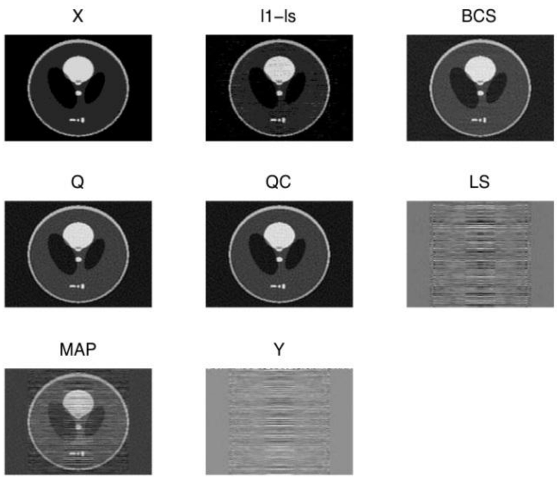

Fig. 5. Application of sparse and cluster prior on a synthetic data reconstruction: phantom.

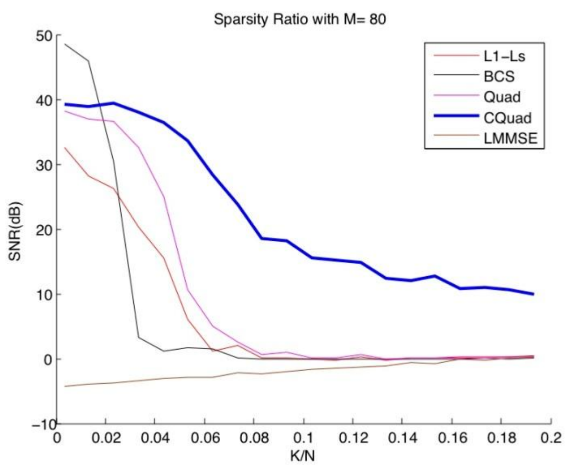

Fig. 6. Performance comparison of different algorithms including LASSO and Clustered LASSO using sparsity ratio, M/N, versus SNR.

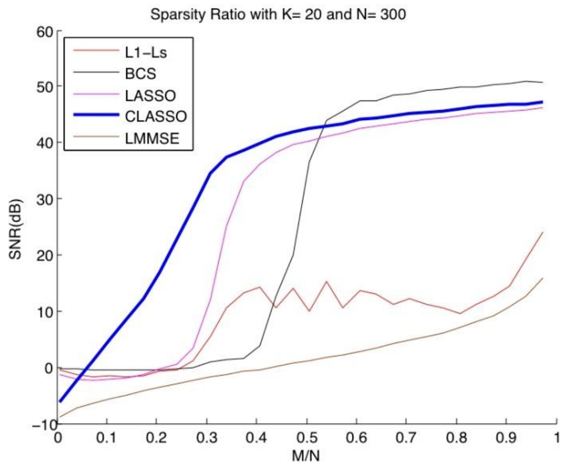

Fig. 7. Performance comparison of different algorithms in- cluding LASSO and Clustered LASSO using measurement ratio, *M/N*, versus SNR.

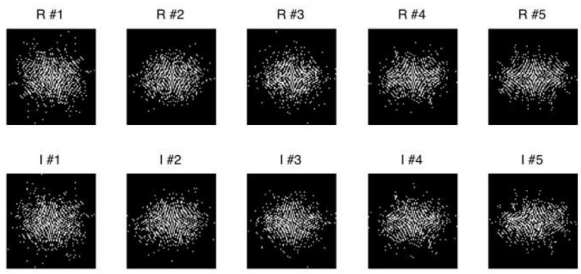

Fig. 8. The five column images represent the real and imaginary part of the Fourier transform representation of the data set we have chosen to present further, which in general shows that the fMRI data have sparse and clustered representation.

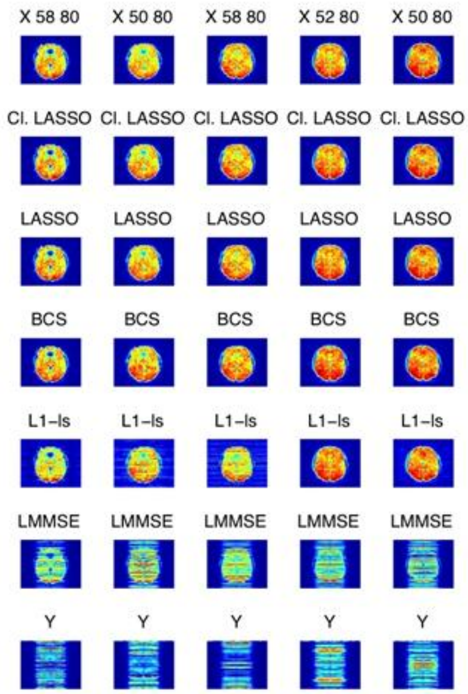

Fig. 9. Application of sparse and cluster prior on a fMRI data Analysis $N = 80$ and $k > 50$.

IV. CONCLUSIONS

In this paper, a Bayesian way of analyzing data is presented. Our emphasis in this work is to incorporate prior information in the analysis of brain data that describes the activation of the brain regions as a function of some stimulus or activity. The more information we have about how the brain regions are connected and related the more our priors will be effective to help the analysis of brain activated regions connected to some stimulus or activity. We are hoping that in our further work we will be working more on data's which have some special structure describing some function of the body and the connectivity of the regions in the brain, showing some structures. So that our work contribution on fMRI data analysis will have much more effect than we have presented here.

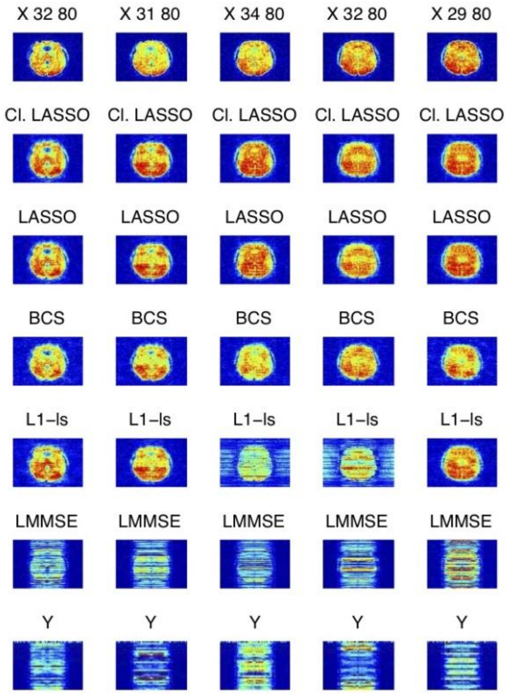

Fig. 10. Application of sparse and cluster prior on a fMRI data analysis: $N = 80$ and $k < 50$.

In addition, in this work we have shown comparison of the different reconstruction algorithms performance for different sparsity ratio and amount of measurement versus SNR as shown in Fig. 6 and Fig. 7, respectively. From some of the plots, one also can see the phase transition property of the different algorithms used in CS in relation to the ratio of sparsity (k/N) and amount of measurement (M/N).

ACKNOWLEGMENT

First, we are grateful to Ralf R. Muller and Lars Lundheim for interesting discussions and insightful suggestions. In addition we would like to express our deep gratitude for members of Trondheim fMRI group at the Norwegian University of Science and Technology (NTNU) for sharing us real fMRI data.

REFERENCES

[1] S. Ogawa, R. S. Menon, D. W. Tank, S. G. Kim, H. Merkle, J. M. Ellermann, and K. Ugurbil, "Functional brain mapping by blood oxygenation level- dependent contrast magnetic resonance imaging. A comparison of signal characteristics with a biophysical model," *Biophysical Journal*, vol. 64, no. 3, pp. 803-812, 1993.

[2] R. Mekle, A. F. Laine, G. M. Perera, and R. DeLaPaz, "Activation detection in fMRI data via multiscale singularity detection," in *Proc. of SPIE, the International Society for Optical Engineering*, 2000, vol. 4119, no. 2, pp. 615-625.

[3] K. J. Friston, A. P. Holmes, K. J. Worsley, J. P. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, no. 4, pp. 189-210, 1994.

[4] R. Baumgartner, L. Rynera, W. Richter, R. Summers, M. Jarmasz, and R. Somorjai, "Comparison of two exploratory data analysis methods for fMRI: Fuzzy clustering vs. principal component analysis," *Magnetic Resonance Imaging*, vol. 18, issue 1, pp. 89-94, January 2000.

[5] L. K. Hansen, J. Larsen, F. Å. Nielsen *et al*., "Generalizable patterns in neuroimaging: How many principal components?" *NeuroImage*, vol. 9, issue 5, pp. 534-544, May 1999.

[6] R. Viviani, G. Grön, and M. Spitzer, "Functional principal component analysis of fmri data," *Human Brain Mapping*, vol. 24, no. 2, pp. 109-129, 2005.

[7] M. J. McKeown and T. J. Sejnowski, "Independent component analysis of fMRI data: examining the assumptions," *Human Brain Mapping*, vol. 6, issue 5-6, pp. 368-372, 1998.

[8] V. D. Calhoun and T. Adali, "Unmixing fMRI with independent component analysis," *Engineering in Medicine and Biology Magazine*, vol. 25, no. 2, pp. 79-90, 2006.

[9] E. Dimitriadou, M. Barth, C. Windischberger, K. Hornik, and E. Moser, "A quantitative comparison of functional MRI cluster analysis," *Artificial Intelligence in Medicine*, vol. 31, no. 1, pp. 57-71, 2004.

[10] B. T. Yeo and W. Ou. Clustering fMRI Time Series. [Online]. Available: http://people.csail.mit.edu/ythomas/6867fMRI.pdf

[11] D. Wang, L. Shi, D. S. Yeung, P. A. Heng, T. T. Wong, and E. C. Tsang, "Support vector clustering for brain activation detection," *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2005*, Springer Berlin Heidelberg, vol. 3749, pp. 572-579, 2005.

[12] H. Chen, H. Yuan, D. Yao, L. Chen, and W. Chen, "An integrated neighborhood correlation and hierarchical clustering approach of functional MRI," *IEEE Transactions on Biomedical Engineering*, vol. 53, no. 3, pp. 452-458, 2006.

[13] M. L. Seghier, K. J. Friston, and C. J. Price, "Detecting subject-specific activations using fuzzy clustering," *Neuroimage*, vol. 36, no. 3, pp. 594-605, 2007.

[14] J. Ye, N. A. Lazar, and Y. Li, "Geostatistical analysis in clustering fMRI time series," *Statistics in Medicine*, vol. 28, no. 19, pp. 2490-2508, 2009.

[15] D. Liu, W. Lu, and N. Zhong, "Clustering of fMRI data using affinity propagation," *Brain Informatics*, Berlin Heidelberg: Springer, vol. 6334, pp. 399-406, 2010.

[16] J. Zhang, D. Li, H. Chen, and F. Fang, "Analysis of activity in fMRI data using affinity propagation clustering," *Computer Methods in Biomechanics and Biomedical Engineering*, vol. 14, no. 3, pp. 271-281, 2011.

[17] J. Ye, N. A. Lazar, and Y. Li, "Sparse geostatistical analysis in clustering fMRI time series," *Journal of Neuroscience Methods*, vol. 199, no. 2, pp. 336-345, 2011.

[18] V. Michel, A. Gramfort, G. Varoquaux *et al*., "A supervised clustering approach for fMRI-based inference of brain states," *Pattern Recognition*, vol. 45, no. 6, pp. 2041-2049, 2012.

[19] E. T. Jaynes, *Probability Theory:The Logic of Science*, ISBN: 0521592712, UK: Cambridge University Press, 2003.

[20] D. J. C. MacKay, *Information Theory, Inference and Learning Algorithms*, UK: Cambridge University Press, ISBN: 978-0-521-64298-9, 2003.

[21] A. O' Hagan, J. Forster, and M. G. Kendall, *Kendall's Advanced Theory of statistics, volume 2B, Bayesian Inference*, London: Arnold, 2004.

[22] J. O. Berger, "Bayesian and conditional frequentist hypothesis testing and model selection," *VIII C. L. A. P. E. M.*, La Habana, Cuba, November 2001.

[23] B. Efron, "Modern science and the bayesian-frequentist controversy," *Division of Biostatistics*, Stanford University, January 2005.

[24] J. Aldrich, "RA fisher on Bayes and Bayes' theorem," *Bayesian Analysis*, vol. 3, no. 1, pp. 161-170, 2008.

[25] E. Moreno and F. J. Girón, "On the frequentist and Bayesian approaches to hypothesis testing," *SORT*, vol. 30, no. 1, pp. 3-28, 2006.

[26] J. Kershaw, B. A. Ardekani, and I. Kanno, "Application of Bayesian inference to fMRI data analysis," *IEEE Transactions on Medical Imaging*, vol. 18, no. 12, pp. 1138-1153, 1999.

[27] W. Penny, S. Kiebel, and K. Friston, "Variational Bayesian inference for fMRI time series," *NeuroImage*, vol. 19, no. 3, pp. 727-741, 2003.

[28] K. J. Friston, W. Penny, C. Phillips *et al*., "Classical and Bayesian inference in neuroimaging: theory," *NeuroImage*, vol. 16, no. 2, pp. 465-483, 2002.

[29] M. A. Mohamed, F. Abou-Chadi, and B. K. Ouda, "Analysis of fmri data using classical and bayesian approaches: A comparative study," in *Proc. World Congress on Medical Physics and Biomedical Engineering*, 2006, vol. 14, pp. 924-931.
[30] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Transactions on Information Theory*, vol. 51, no. 12, pp. 4203-4215, 2005.
[31] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289-1306, 2006.
[32] E. Cand'es, J. Romberg, and T. Tao, "Robust uncertainty principles: Ex- act signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489509, Feb. 2006.
[33] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489-509, 2006.
[34] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic resonance in medicine*, vol. 58, no. 6, pp. 1182-1195, 2007.
[35] H. Luo and S. Puthusserypady, "A sparse Bayesian method for determination of flexible design matrix for fMRI data analysis," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 52, no. 12, pp. 2699-2706, 2005.
[36] V. P. Oikonomou, E. E. Tripoliti, and D. I. Fotiadis, "A sparse variational Bayesian approach for fMRI data analysis," in *Proc. 8th IEEE International Conference on BioInformatics and BioEngineering*, 2008, pp. 1-6.
[37] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for clustered sparse signals," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, 2011, pp. 3948-3951.
[38] S. Rangan, A. K. Fletcher, and V. K. Goyal. (2009). Asymptotic analysis of MAP estimation via the replica method and ap- plications to compressed sensing. [Online]. Available: http://adsabs.harvard.edu/abs/2009arXiv0906.3234R.
[39] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal processing*, vol. 56, no. 6, pp. 2346-2356, June 2008.
[40] J. D Blanchard and A. Thompson, "Pushing the rip phase transition in compressed sensing," in *Proc. European Signal Processing Conf.*, 2010, pp. 412-416.
[41] D. L. Donoho, A. Maleki, and A. Montanari, "The noise-sensitivity phase transition in compressed sensing," *IEEE Transactions on Information Theory*, vol. 57, no. 10, pp. 6920-6941, 2011.
[42] Y. Wu and S. Verdú, "Optimal phase transitions in compressed sensing," *IEEE Transactions on Information Theory*, vol. 58, issue 10, Oct. 2012.
[43] M. Behroozi, M. R. Daliri, and H. Boyaci," Statistical Analysis Methods for the fMRI Data," *Basic and Clinical Neuroscience*, vol. 2, no. 4, pp. 67-74, 2011.

Solomon A. Tesfamicael was born in Gonder, Ethiopia in 1973. He received his bachelor degree in Mathematics from Bahirdar university, Bahirdar, Ethiopia in 2000 and master degree in Coding theory from the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway in 2004. He began his Phd studies in signal processing in sept, 2009 at the department of Electronics and Telecomunications at NTNU.

He is currently working as lecturer at the Sør-Trondlag university collge (HiST), in Trondheim, Norway. He has worked as mathematics teacher for secondary schools and as assistant lecturer at the Engineering faculity in Bahirdar university in Ethiopia. In addition, he has a pedagogical studies both in Ethiopia and in Norway. He has authored and co-authtored papers related to compressed sensing (CS) and multiple input and multiple out put (MIMO) systems. His research interest are signal processing, Compresses sensing, multiuser communication (MIMO, CDMA), statistical mechanical methods like replica method, functional magnetic resonance imaging (fMRI) and mathematic education.

Faraz Barzideh was born in Tehran, Iran in 1987. He received his bachelor degree from Zanjan University, Zanjan, Iran, in the field of Electronic Engineering in 2011 and finished his master degree in the field of Medical Signal Processing and Imaging from Norwegian University of Science and Technology (NTNU) in 2013. His research interests are medical signal and image processing especially in MRI and also compressive sensing.