

Clustered Compressive Sensing: Application on Medical Imaging

Solomon A. Tesfamicael, *IACSIT Member* and Faraz Barzideh

Abstract—This paper provides clustered compressive sensing (CCS) based image processing using Bayesian framework applied to medical images. Some images, for example like magnetic resonance images (MRI) are usually very weak due to the presence of noise and due to the weak nature of the signal itself. Compressed sensing (CS) paradigm can be applied in order to boost such signals. We applied CS paradigm via Bayesian framework. Using different sparse prior information and in addition incorporating the special structure that can be found in sparse signal, CCS can be applied to improve image processing. This is shown in the results of this paper. First, we applied our analysis on Angiogram image, then on Shepp-logan phantom and finally on another MRI image. The results show that applying the clustered compressive sensing give better results than the non-clustered version.

Index Terms—Bayesian framework, sparse prior, clustered prior, posterior, compressive sensing, LASSO, clustered LASSO.

I. INTRODUCTION

Compressive Sensing (CS) is a paradigm to capture information at lower rate than the Nyquist-Shannon sampling rate when signals are sparse in some domain [1]-[4]. The problem of limited number of samples or measurements can occur in multiple scenarios, e.g. when we have limitations on the number of data capturing devices, measurements are very expensive or slow to capture such as in magnetic resonance imaging (MRI) [5], [6]. The CS paradigm in signal processing requires three important ingredients [6]. First, the desired signal should have a sparse representation in a known transform domain, i.e., it should be compressible. If the signal is sparse spatially, for example, consider an image which is sparse in the pixels, then the transform domain can be the identity. Second, the aliasing artifacts due to under sampling should be incoherent in the transform domain. This creates a noise like structure. This measurement noise then can be modeled using white Gaussian noise. Third, a nonlinear reconstruction scheme should be used to enforce sparsity and consistency with the data [7]. Recently, this recovery using CS has been shown to be mathematically exact [2], [3]. As a signal-processing scheme, CS follows a

Manuscript received May 11, 2014; revised June 30, 2014. This work was supported in part by the Sør-Trøndelag 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-Trøndelag 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 University of Stavanger (UiS) (e-mail: faraz.barzideh@gmail.com).

similar framework: encoding, transmission/storing, and decoding.

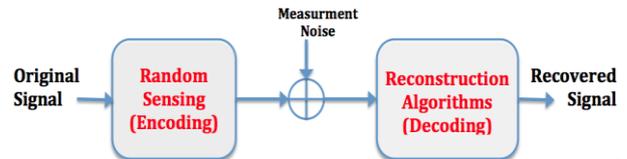


Fig. 1. Block diagram for CS based reconstruction.

Focusing on the encoding and decoding of such a system with noisy measurement the block diagram is given in Fig. 1. Encoding is done by linear projections using random sensing transformations. At the decoding side, reconstruction is done using nonlinear schemes. And there are many algorithms in addressing this issue: convex relaxations [4], [8], greedy iterative algorithms [9], iterative thresholding algorithms [10]-[12]. In this paper, the focus is merely on the convex relaxation methods. We consider a noisy measurement and applied convex relaxation algorithms for robust reconstruction. This is done using a Bayesian framework for its flexibility and robustness. Basically, it is the updated version of the recent work [13].

Therefore, this paper is organized as follows. In section II we present the CS problem and redefine it under Bayesian framework as in [13], [14]. Section III shows our results using synthetic signals, and section IV presents conclusion and future work.

II. COMPRESSED SENSING BASED RECOVERY

Beginning with a given vector of measurements $\mathbf{y} \in \mathbb{R}^M$ and measurement matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, assuming noisy measurement with $\mathbf{w} \in \mathbb{R}^M$ being i.i.d. Gaussian random variables with zero mean and covariance matrix $\sigma^2 \mathbf{I}$, estimating the sparse vector $\mathbf{x} \in \mathbb{R}^N$ is the problem that we are considering given the linear model

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}. \quad (1)$$

Here $N \gg M$ and $N \gg k$, where k is the number of non-zero entries in \mathbf{x} . Applying CS reconstructions using different algorithms we recover the estimate of the original signal \mathbf{x} , say $\hat{\mathbf{x}}$. The measurement noise is reduced simultaneously with the reconstruction of the true image data using nonlinear reconstruction schemes.

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

$$\hat{\mathbf{x}} = (\mathbf{A}^T \mathbf{A})^{-1} \mathbf{A}^T \mathbf{y}, \quad (2)$$

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

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

where $uf(\mathbf{x})$ is a regularizing term, for some non-negative u . If $f(\mathbf{x}) = \|\mathbf{x}\|_p$, as a penalizing norm. In this paper we shall consider when $p = 0, 1, \text{ and } 2$, which gives us different estimators which we define them here using Bayesian framework.

A. Bayesian Framework

Under Bayesian inference consider two random variables \mathbf{x} and \mathbf{y} with probability density function (pdf) $p(\mathbf{x})$ and $p(\mathbf{y})$, respectively. Using Bayes' theorem it is possible to show that the posterior distribution, $p(\mathbf{x}|\mathbf{y})$, is proportional to the product of the likelihood function, $p(\mathbf{y}|\mathbf{x})$, and the prior distribution, $p(\mathbf{x})$,

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}). \quad (4)$$

Equation (4) is called Updating Rule in which the data allows us to update our prior views about \mathbf{x} . And as a result we get the posterior which combines both the data and non-data information of \mathbf{x} [13], [15]-[17]. Further the Maximum a posterior (MAP), $\hat{\mathbf{x}}_{MP}$, is given by

$$\hat{\mathbf{x}}_{MP} = \arg \max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \quad (5)$$

to proceed further, we assume prior distributions on \mathbf{x} .

B. Sparse Prior

The reconstruction of \mathbf{x} resulting from (3) for the sparse problem we consider in this paper is presented as a maximum a posteriori (MAP) estimator under the Bayesian framework as in [14]. We show this by defining a prior probability distribution for \mathbf{x} of the form

$$p(\mathbf{x}) = \frac{\exp(-u f(\mathbf{x}))}{\int_{\mathbf{x} \in \mathbb{R}^N} \exp(-u f(\mathbf{x})) d\mathbf{x}} \quad (6)$$

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 $f(\mathbf{x}) = \sum_{i=1}^N f(x_i)$, such that for sufficiently large u , $\int_{\mathbf{x} \in \mathbb{R}^N} \exp(-u f(\mathbf{x})) d\mathbf{x}$ is finite. Further, the likelihood function of \mathbf{y} given \mathbf{x} is given by

$$p_{\mathbf{y}|\mathbf{x}}(\mathbf{y}|\mathbf{x}) = \frac{1}{(2\pi\sigma)^{N/2}} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{Ax}\|_2^2\right) \quad (7)$$

together with (4) and (6), this now gives the posterior

$$p_{\mathbf{x}|\mathbf{y}}(\mathbf{x}|\mathbf{y}; A) = \frac{\exp\left(-u\left(\frac{1}{2}\|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda f(\mathbf{x})\right)\right)}{(2\pi\sigma)^2 \int_{\mathbf{x} \in \mathbb{R}^N} \exp\left(-u\left(\frac{1}{2}\|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda f(\mathbf{x})\right)\right) d\mathbf{x}} \quad (8)$$

and the MAP estimator is then given by

$$\hat{\mathbf{x}}_{MP} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda f(\mathbf{x}). \quad (9)$$

As shown in [13]. Now, as we choose different regularizing function which enforces sparsity into the vector \mathbf{x} , we get different estimators listed below [16].

- 1) Linear Estimators: when $f(\mathbf{x}) = \|\mathbf{x}\|_2^2$, (9) reduces to

$$\hat{\mathbf{x}}_{Linear} = \mathbf{A}^T (\mathbf{AA}^T + \lambda \mathbf{I})^{-1} \mathbf{y} \quad (10)$$

which is the LMMSE estimator. But this estimator is not good enough for sparsity problem since it does not enforce sparsity well. Instead the following two estimators are more interesting for CS problems.

- 2) LASSO Estimator: when $f(\mathbf{x}) = \|\mathbf{x}\|_1$, we get the LASSO estimator and (9) becomes,

$$\hat{\mathbf{x}}_{LASSO} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_1. \quad (11)$$

- 3) Zero-Norm regularization estimator: when $f(\mathbf{x}) = \|\mathbf{x}\|_0$, we get zero norm regularization estimator and (9) becomes

$$\hat{\mathbf{x}}_{Zero-Norm} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_0. \quad (12)$$

This is the best solution for reconstruction of the sparse vector \mathbf{x} , but is NP-complete. The worst solution among these LP penalizing forms for the sparse problem considered is the L2- regularization solution given by (10). However, the best approximation is given by equation (11) and its equivalent forms such as L1-norm regularized least-squares (L1-LS) and others [2]-[4].

C. Clustering Prior

Building on the Bayesian philosophy, we can further assume another prior distribution for clustering. The entries of the sparse vector \mathbf{x} may have some structure that can be represented using distributions. In [18] 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

$$D(\mathbf{x}) \equiv \sum_{i=2}^N |x_i - x_{i-1}|, \quad (13)$$

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

$$q(\mathbf{x}) \propto \frac{1}{\beta} \left(\lambda \|\mathbf{x}\|_1 + \gamma D(\mathbf{x}) \right), \quad (14)$$

where β is the normalization constant to get $\int_{\mathbf{x} \in \mathbb{R}^N} q(\mathbf{x}) d\mathbf{x} = 1$. The new posterior evolving this prior under the Bayesian framework is proportional to the product of the three pdf's $p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})q(\mathbf{x})$.

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

$$\hat{\mathbf{x}}_{clu-LASSO} = \arg \min_{\mathbf{x} \in \mathbb{R}^N} \frac{1}{2} \|\mathbf{y} - \mathbf{Ax}\|_2^2 + \lambda \|\mathbf{x}\|_1$$

$$+ \gamma \sum_{i=2}^N |x_i - x_{i-1}|, \quad (14)$$

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

III. RESULTS

The main focus of this paper is to give a practical application of clustered compressed sensing. In order to verify the theory we have selected different medical related images. We used LS, LMMSE, LASSO and Clustered LASSO given by equations (2), (10), (11), and (12), respectively, to reconstruct from a noisy measurement and compare their performances too. We apply our analysis to Angiogram, phantom and then to functional MRI (fMRI) images.

A. Angiogram Image

The first one is an angiogram image taken from University Hospital Rechts der Isar, Munich, Germany [19]. Angiogram images are already sparse in the pixel representation. In general MRI images are sparse (and even clustered) in the spatial and the transformed domain. The image we took is also clustered as well. The original signal after vectorization is \mathbf{x} of length $N = 960$. By taking 746 measurements, and maximum number of non-zero elements $k = 373$, we applied the different the reconstruction schemes and the results are shown in Fig. 2 and Table I.

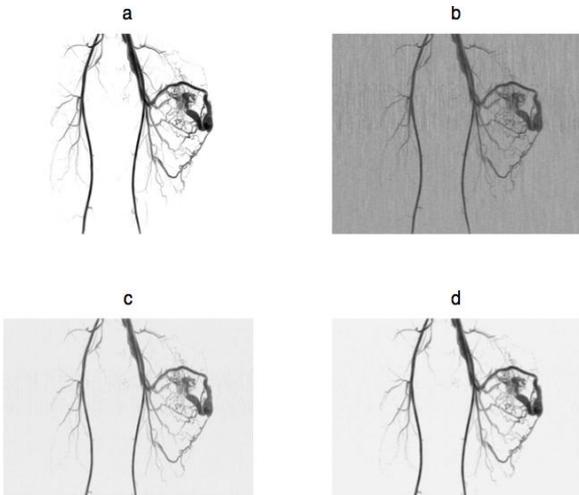


Fig. 2. Comparison of reconstruction schemes: a) original image \mathbf{x} b) LMMSE c) LASSO d) clustered LASSO.

TABLE I: PERFORMANCE COMPARISON

Algorithms	MSE in dB
LMMSE	-35.1988
LASSO	-53.6195
Clustered LASSO	-63.6889

B. Phantom Image

Consider the Shepp-Logan phantom which is not sparse in spatial domain but can be sparsified in k -space by zeroing out small coefficients. We then measured the sparsified image and added noise. The original signal after vectorization is \mathbf{x} of length $N = 200$. By taking 94 measurements, that is \mathbf{y} is of

length $M = 94$, and maximum number of non-zero elements $k = 47$, we applied different the reconstruction algorithms used above. The result shows clustered LASSO does well compared to the others as can be seen in Fig. 3 and Table II.

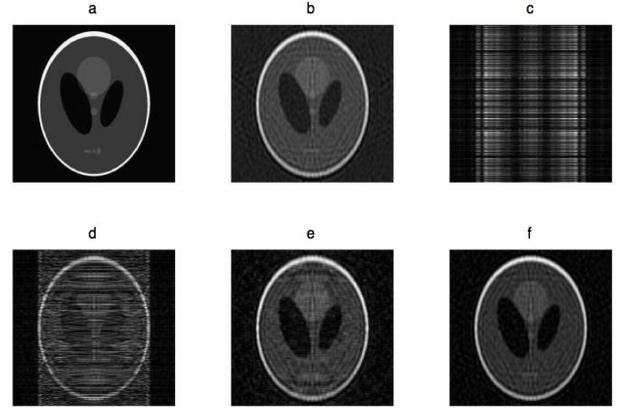


Fig. 3. Comparison of reconstruction schemes: a) original image \mathbf{x} b) sparsified image c) LS d) LMMSE e) LASSO f) clustered LASSO.

TABLE II: PERFORMANCE COMPARISON

Algorithms	MSE in dB
LS	-21.3304
LMMSE	-27.3876
LASSO	-37.9978
Clustered LASSO	-40.0068

C. fMRI Image

Another example to apply the clustered LASSO based image reconstruction using Bayesian framework to medical images is a functional MRI (fMRI) image. We took many slices of fMRI image given in Fig. 5, which is sparse with some clusterdness in the transform domain as it is shown in Fig. 4. And this gives ground to apply the framework and the procedure used here. The performance of the different reconstruction schemes is visible from Fig. 5.

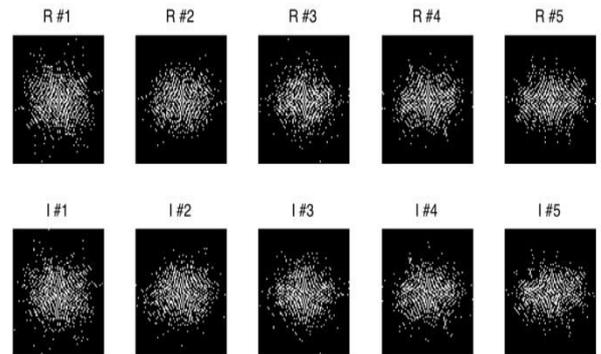


Fig. 4. The five column images represent the real and imaginary part of the Fourier transform representation of the data set we have chosen to present further, which in general shows that the fMRI data have sparse and clustered representation.

In addition, for a synthetic data we have compared the different recovery techniques by using pick signal to noise ratio (PSNR) and mean square error (MSE) versus measurement ratio (M/N) and the results are shown in Fig. 6 and Fig. 7, respectively. Generally reconstruction using LASSO is much better than LS and LMMSE algorithms for the sparse reconstruction problem.

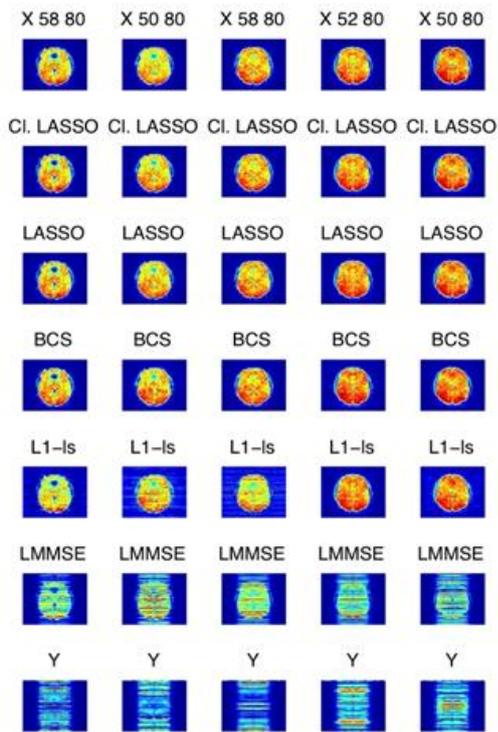


Fig. 5. Application of sparse and cluster prior on a fMRI data analysis $N=80$ and $k>50$.

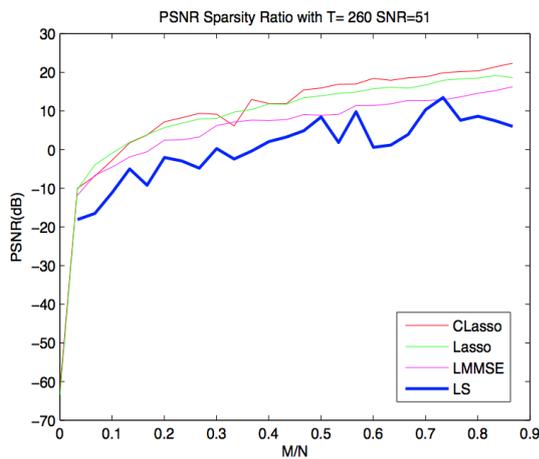


Fig. 6. Comparison of reconstruction schemes using PSNR versus measurement ratio M/N .

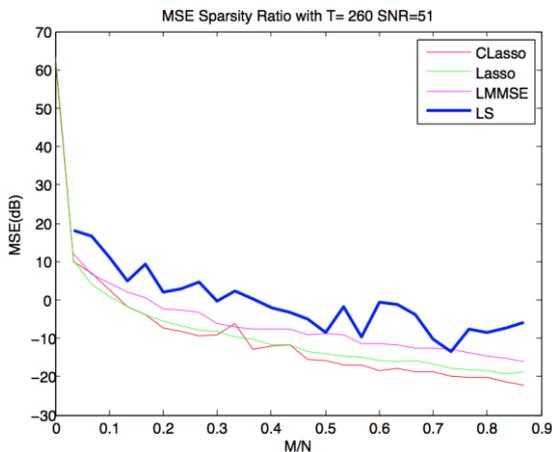


Fig. 7. Comparison of reconstruction schemes using MSE versus measurement ratio M/N .

Further, clustered LASSO outperforms LASSO since it uses more accurate information about the structure of the

sparsity. Finally, we see the impact of sparsity ratio on the performances of the reconstruction schemes in Fig. 8 using reconstruction ratio versus sparsity ratio, k/N .

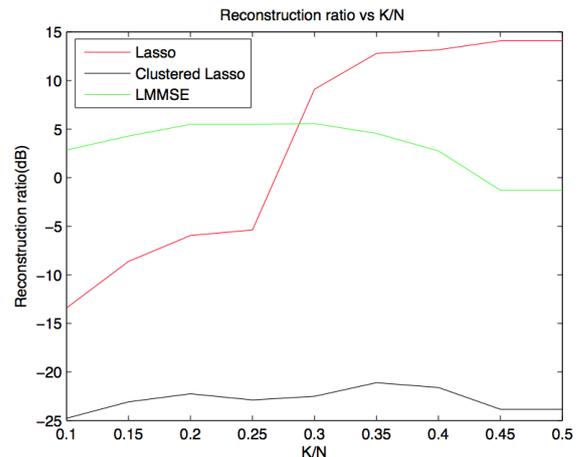


Fig. 8. Comparison of different reconstruction schemes using the metric reconstruction ratio versus sparsity ratio k/N .

IV. CONCLUSIONS

In this paper clustered compressive sensing using Bayesian framework is presented. Our emphasis in this work is to incorporate prior information's like sparseness and clusteredness in the reconstruction of signals from fewer measurements. And apply it on different medical related images. Clustered LASSO recovery does well in terms of PSNR and MSE than LASSO (using only sparse prior), LMMSE and LS. In addition, in this work we have shown comparison of the different reconstruction algorithms performance for different amount of measurement ratio versus PSNR and MSE. In addition sparsity ratio versus reconstruction ratio is provided to see hoe the schemes behave with the amount of sparsity. For future work we plan to apply different forms of clustering depending on the prior information's of images or geometry of clusteredness.

REFERENCES

- [1] D. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, no. 4, 2006, p. 12891306.
- [2] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inform. Theory*, vol. 51, no. 12, December 2005.
- [3] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, pp. 489509, Feb. 2006.
- [4] E. J. Cande's and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inf. Theory*, vol. 52, pp. 54065425, Dec. 2006.
- [5] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *IEEE Trans.*, vol. 58, issue 6, pp. 1182-1195, 2007.
- [6] S. Qaisar, R. M. Bilal, W. Iqbal, M. Naureen, and S. Lee, "Compressive sensing: From theory to applications, a survey," *Journal of Communications and Networks*, vol. 15, no. 5, pp. 443-456, Oct. 2013.
- [7] J. Claerbout and F. Muir, "Robust modeling with Erratic data," *Geophysics*, vol. 38, no. 5, pp. 826-844, 1973.
- [8] E. Candes and B. Recht, "Exact matrix completion via convex optimization," *Foundations of Computational Mathematics*, vol. 9, no. 6, pp. 717- 772, 2009.
- [9] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Transactions on signal processing*, vol. 41, no. 12, pp. 3397- 3415, 1993.
- [10] R. Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society, Series B*, vol. 58, pp. 267-288, 1994.

- [11] D. Donoho, "Denoising by soft thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613-627, 2002.
- [12] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Applied and Computational Harmonic Analysis*, vol. 27, no. 8, pp. 265274, Nov. 2009.
- [13] S. A. Tesfamicael and F. Barzideh, "Clustered compressed sensing in fMRI data analysis using a Bayesian framework," *International Journal of Information and Electronics Engineering*, vol. 4, no. 2, pp. 74-80, 2014.
- [14] S. Rangan, A. K. Fletcher, and V. K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," arXiv:0906.3234v1, 2009.
- [15] A. O' Hagen and J. Forster, "Bayesian Inference," in *Kendall's Advanced Theory of statistics*, vol. 2B, Arnold, a member of the Hodder Headline Group, 2004.
- [16] E. T. Jaynes, "Probability Theory," in *The Logic of Science*, Cambridge University Press, 2003.
- [17] D. J. C. Mackay, "Information Theory," in *Inference, and Learning Algorithms*, Cambridge University Press, 2003.
- [18] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for clustered sparse signals," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2011.
- [19] Magnetic-resonance-imaging. [Online]. Available: <http://usa.healthcare.siemens.com/magnetic-resonance-imaging/0-35-to-1-5t-mri-scanner/magnetom-avanto/use>



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 sept. 2009 at the Department of Electronics and Telecommunications at NTNU.

He is currently working as lecturer at the Sør-Trøndrag University College (HiST), in Trondheim, Norway. He has worked as mathematics teacher for secondary schools and as assistant lecturer at the Engineering faculty in Bahirdar university in Ethiopia. In addition, he has a pedagogical studies both in Ethiopia and in Norway. He has authored and co-authored 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 and started his PhD study in University of Stavanger (UiS) in 2014.

His research interests are medical signal and image processing especially in MRI and also compressive sensing and dictionary learning.