A SPARSE BAYESIAN LEARNING FOR HIGHLY ACCELERATED DYNAMIC MRI

Hong Jung, and Jong Chul Ye

Bio-Imaging & Signal Processing Lab., Dept. of Bio and Brain engineering
Korea Adv. Inst. of Science & Technology (KAIST)
373-1 Guseong-dong Yuseong-gu, Daejon 305-701, Republic of Korea

ABSTRACT

In dynamic MRI, spatio-temporal resolution is a very important issue. Recently, compressed sensing approach has become a highly attracted imaging technique since it enables accelerated acquistion without aliasing artifacts. Our group has proposed an l_1-norm based compressed sensing dynamic MRI called k-t FOCUSS, which outperforms existing methods. However, it is known that the restrictive conditions for l_1 exact reconstruction usually cost more measurements than l_0 minimization. In this paper, we adopts a sparse Bayesian learning approach to improve k-t FOCUSS and achieve l_0 solution. We demonstrated the improved image quality using *in vivo* cardiac cine imaging.

Index Terms— Dynamic MRI, Compressed sensing, l_1 minimization, Sparse Bayesian learning

1. INTRODUCTION

In dynamic MRI for imaging moving objects, simultaneous improvement of spatial and temporal resolution is highly required. Conventional MR imaging methods rely on Nyquist sampling theory, which imposes a significant limitation in improving spatio-temporal resolution. In cardiac cine imaging, to resolve this limitation, ECG gating is used which enables data acquisitions during multiple heartbeats. However, breath-holding is required to avoid motion artifacts. Therefore, accelerated data acquisition is really essential for its success.

Recently, sparse approximation of unknown signals has become a main research interest, since it provides a clue to break Nyquist sampling limit. Since dynamic MRI can be easily sparsified by exploiting temporal redundancies, it is a very important application of such sparse approximation.

According to [1], when the unknown signal is very sparse, by minimizing l_0-quasi norm, exact reconstruction is possible from very small number of measurements. However, since l_0 minimization is NP-hard, it is computationally infeasible. Instead, the emerging sampling theory, "compressed sensing" [2], tells us that l_0 minimization solution can be obtained by solving an l_1 convex optimization with additional over-sampling factors. There have been a large number of researches applying the compressed sensing theory to MRI [3, 4, 5]. Among them, k-t FOCUSS [4, 5] is one of the successful approaches for dynamic MRI. k-t FOCUSS achieves an l_1 minimization by iteratively solving a reweighted l_2 minimization problem. Furthermore, by applying motion estimation and compensation (ME/MC) to make the residual signal sparse, more accurate reconstruction can be achieved [5].

Even though k-t FOCUSS has been successful in dynamic MRI, there is more room for improvements. Recall that l_1 minimization requires more number of measurements than l_0 minimization to achieve l_0/l_1 equivalence. From this aspect, non-convex l_p $(0 < p < 1)$ norm minimization has been recently investigated in MRI [6]. However, since l_p minimization is not a convex optimization, the number of local minima increases combinatorially. Meanwhile, there have been Bayesian approaches that interpret the l_0 or l_1 minimization as a priori distribution of the unknown solution. Under the Bayesian framework, the compressed sensing can be explained as maximum a posteriori (MAP) estimation. Furthermore, rather than explicitly specifying priors, by empirically updating the sparse information for each pixel, sparse Bayesian learning (SBL) is shown to avoid many local minima of l_0 cost function and achieves the l_0 solution directly [7]. Furthermore, for multiple measurement vector case, SBL is the only algorithm up to now that is guaranteed to achieve the "global" l_0 optimal solution under some restricted conditions [7].

Therefore, this paper employs the SBL to improve k-t FOCUSS. It was found out that FOCUSS and SBL are very closely related to each other, so the modification of k-t FOCUSS is minimal. However, in achieving the l_0 minimization, we found that many small components are forced to be zero if the signal is not perfectly sparse but rather compressible. This implies that unlike the k-t FOCUSS, the spatial domain sparsifying transform is essential to reduce the artifacts in SBL. Hence, we apply a phase encoding directional finite difference to sparsify the coefficients. Since neighboring pixels usually have similar motion characteristic, finite difference along the phase encoding direction allows large parts of

This work was supported by the Korea Science and Engineering Foundation, Grant No:2009-0081089.

coefficients to be zero. The proposed method showed superior performance over k-t FOCUSS at large down sampling ratio.

2. THEORY

2.1. k-t FOCUSS and MAP estimation

From compressed sensing perspective, the sparse approximation for dynamic MR imaging problem can be stated as follows:

$$\min ||\boldsymbol{\rho}||_1, \quad \text{subject to } ||\boldsymbol{\upsilon} - \mathbf{F}\boldsymbol{\rho}||_2 \leq \epsilon. \tag{1}$$

Let the down-sampled k-space data be stacked in a vector $\boldsymbol{\upsilon}$ with size of N. Then, the sensing matrix $\mathbf{F} \in \mathbb{C}^{N\times M}(N < M)$ can be represented as $\mathbf{F} = \mathbf{F}_x\mathbf{F}_t$, where $\mathbf{F}_x$ and $\mathbf{F}_t$ represent down sampled Fourier transform along phase encoding direction-x and Fourier transform along temporal direction-t, respectively. Since Fourier transform along temporal direction sparsely represents dynamic MR images especially when objects have periodic motions, $\boldsymbol{\rho}$ can be sparsely represented with a M-dimensional stacked vector of unknown x-f spectrum.

The solution of Eq (1) is achieved using a reweighted l_2 minimization in k-t FOCUSS [4] by iteratively solving the following equation:

$$\hat{\boldsymbol{\rho}}_{n+1} = \boldsymbol{\Theta}_n\mathbf{F}^H\left(\mathbf{F}\boldsymbol{\Theta}_n\mathbf{F}^H + \lambda\mathbf{I}\right)^{-1}\boldsymbol{\upsilon}, \tag{2}$$

where $\boldsymbol{\Theta}_n$ is updated at each iteration by

$$\boldsymbol{\Theta}_{n+1} = diag([|\hat{\boldsymbol{\rho}}_{n+1}(1)|^{2-p}, \cdots, |\hat{\boldsymbol{\rho}}_{n+1}(M)|^{2-p}]). \tag{3}$$

When $p = 1$, k-t FOCUSS solves an l_1 minimization problem. Choosing an arbitrary value in $(0, 1)$, k-t FOCUSS is also known to achieve the l_p minimization, but have many local minima [7].

Interestingly, the derivation of the above solution can be recast in Bayesian terms by applying an $\exp[-(\cdot)]$ transform to Eq. (1) as explained in [7]. This derives a Gaussian likelihood function $p(\boldsymbol{\upsilon}|\boldsymbol{\rho})$ and a priori distribution $p(\boldsymbol{\rho})$ as follows:

$$\begin{aligned} p(\boldsymbol{\upsilon}|\boldsymbol{\rho}) &\propto \exp[-\frac{1}{\lambda}||\boldsymbol{\upsilon} - \mathbf{F}\boldsymbol{\rho}||_2^2] \\ p(\boldsymbol{\rho}) &\propto \exp[-||\boldsymbol{\rho}||_p]. \end{aligned} \tag{4}$$

A priori distribution for Eq. (1) corresponds to $p = 1$. Then, Eq. (1) can be interpreted as a MAP estimation as follows:

$$\begin{aligned} \hat{\boldsymbol{\rho}} &= \arg\max_{\boldsymbol{\rho}} \ p(\boldsymbol{\upsilon}|\boldsymbol{\rho})p(\boldsymbol{\rho}) \\ &= \arg\max_{\boldsymbol{\rho}} \ p(\boldsymbol{\rho}|\boldsymbol{\upsilon}). \end{aligned} \tag{5}$$

Then, using the expectation maximization (EM) algorithm with a set of latent variables $\boldsymbol{\theta}_n = [\theta_n(1), \cdots, \theta_n(M)]$ related to $\boldsymbol{\rho}$, Eq. (5) can be solved as follows:

$$\begin{aligned} \text{E-step} &: \ \boldsymbol{\theta}_n(i) = |\hat{\boldsymbol{\rho}}_n(i)|^{2-p}, \quad \forall i = 1, \cdots, M \\ \text{M-step} &: \ \hat{\boldsymbol{\rho}}_{n+1} = \boldsymbol{\Theta}_n\mathbf{F}^H\left(\mathbf{F}\boldsymbol{\Theta}_n\mathbf{F}^H + \lambda\mathbf{I}\right)^{-1}\boldsymbol{\upsilon}, \end{aligned} \tag{6}$$

where $\boldsymbol{\Theta}_n = diag(\boldsymbol{\theta}_n)$. The solution and update equation is exactly same with k-t FOCUSS.

As $p(\boldsymbol{\rho})$ is a priori distribution of unknown $\boldsymbol{\rho}$, as $p \to 0$ in Eq. (4), $p(\boldsymbol{\rho})$ has a higher probability at $||\boldsymbol{\rho}||_p \to 0$. This means that sparse solution has a higher likelihood. However, if $p < 1$, the optimization is not convex so that the number of local minima combinatorially increases. For example, if the desired solution has $D \leq N$ non-zero elements without noise, l_p minimization has $\binom{M}{N} - \binom{M-D}{N-D} + 1$ local minima [7]. Therefore, we cannot guarantee that the l_p $(0 < p < 1)$ minimization always solves a sparser solution than l_1 minimization.

2.2. SBL: Empirical Bayes

To address this problem, SBL [7] uses an empirical prior, which is a flexible priori distribution dependent on a set of unknown hyperparameters that must be estimated form the data. More specifically, instead of specifying l_p norm, SBL imposes sparsity for each unknown pixel by assuming zero mean Gaussian distribution:

$$p(\boldsymbol{\rho}(i); \boldsymbol{\theta}(i)) = \mathcal{N}(0, \boldsymbol{\theta}(i)), \tag{7}$$

where $\boldsymbol{\theta}(i)$ is a hyperparameter corresponding to an unknown variance of $\boldsymbol{\rho}(i)$. By combining all of the priors, a full prior can be obtained as follows:

$$p(\boldsymbol{\rho}; \boldsymbol{\theta}) = \prod_{i=1}^{M} p(\boldsymbol{\rho}(i); \boldsymbol{\theta}(i)). \tag{8}$$

Then, the posterior density of $\boldsymbol{\rho}$ is obtained with

$$p(\boldsymbol{\rho}|\boldsymbol{\upsilon}; \boldsymbol{\theta}) = \frac{p(\boldsymbol{\rho}, \boldsymbol{\upsilon}; \boldsymbol{\theta})}{\int p(\boldsymbol{\rho}, \boldsymbol{\upsilon}; \boldsymbol{\theta})d\boldsymbol{\rho}} = \mathcal{N}(\hat{\boldsymbol{\rho}}, \boldsymbol{\Sigma}), \tag{9}$$

with mean and covariance given by:

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\Theta}\mathbf{F}^H(\lambda\mathbf{I} + \mathbf{F}\boldsymbol{\Theta}\mathbf{F}^H)^{-1}\boldsymbol{\upsilon}, \tag{10}$$

$$\boldsymbol{\Sigma} = \boldsymbol{\Theta} - \boldsymbol{\Theta}\mathbf{F}^H(\lambda\mathbf{I} + \mathbf{F}\boldsymbol{\Theta}\mathbf{F}^H)^{-1}\mathbf{F}\boldsymbol{\Theta}, \tag{11}$$

where $\boldsymbol{\Theta} = diag(\boldsymbol{\theta})$. From Eq. (10), when $\boldsymbol{\theta}(i) = 0$, $\hat{\boldsymbol{\rho}}(i)$ is zero with probability of 1 as desired. Therefore, sparsity of $\boldsymbol{\rho}$ is determined by the sparsity of hyperparmeters $\boldsymbol{\theta}$. In other words, estimating the hyperparamters $\boldsymbol{\theta}$ is equal to a model selection of a priori distribution. In this context, estimating sparse hyperparameters is the only problem we have to address so that the unknown signal $\boldsymbol{\rho}$ can be integrated out. Then, the maximum likelihood of $\boldsymbol{\theta}$ can be calculated by minimizing the following cost function:

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= -2\log\int p(\boldsymbol{\upsilon}|\boldsymbol{\rho})p(\boldsymbol{\rho};\boldsymbol{\theta})d\boldsymbol{\rho} = -2\log p(\boldsymbol{\upsilon};\boldsymbol{\theta}) \\ &= \log|\lambda\mathbf{I}+\mathbf{F}\boldsymbol{\Theta}\mathbf{F}^H| + \boldsymbol{\upsilon}^H(\lambda\mathbf{I}+\mathbf{F}\boldsymbol{\Theta}\mathbf{F}^H)^{-1}\boldsymbol{\upsilon}.\end{aligned} \quad (12)$$

Taking a derivative of Eq. (12) with respect to $\boldsymbol{\theta}$, the minimizer can be achieved with the following fixed point iteration:

$$\boldsymbol{\theta}_n(i) = \frac{|\hat{\boldsymbol{\rho}}_n(i)|^2}{1-\boldsymbol{\theta}_{n-1}(i)^{-1}\boldsymbol{\Sigma}(i,i)}, \quad \forall i = 1,\cdots,M, \quad (13)$$

where $\boldsymbol{\rho}_n$ and $\boldsymbol{\Sigma}(i,i)$ are updated by mean and diagonal components of variance of Eq. (9) as given in Eq. (10) and (11), respectively. Finally, the optimal solution $\hat{\boldsymbol{\rho}}$ is achieved by Eq. (10) when $\boldsymbol{\theta}_n$ converges. Even if SBL does not appear to directly minimize l_0 norm of unknown signals $\boldsymbol{\rho}$, it is proved that the globally optimal solution of SBL is equal to the solution of l_0 minimization of $\boldsymbol{\rho}$ [7].

Interestingly, the only difference between MAP approach and SBL comes from the update equation of $\boldsymbol{\theta}_n$. Furthermore, if the denominator of Eq. (13) is a constant, the update rule of SBL becomes exactly same with that of k-t FOCUSS with $p = 0$. However, surprisingly, this enables the size of local minima set of SBL to be much smaller than that of MAP estimation [7]. More specifically, [7] showed that the set of local minima of SBL is a subset of basic feasible solution that is equal to the set of local minima of l_0 minimization under the so-called unique representation property condition. Furthermore, the size of local minima set of SBL is significantly shrunk by following necessary condition for local minima of SBL. Intuitively, if we found an optimal sparse solution $\hat{\boldsymbol{\rho}}$ ($||\hat{\boldsymbol{\rho}}||_0 = D_0 \leq N$) and a set of columns $\tilde{\mathbf{F}} \in \mathbb{C}^{N\times D_0}$ in $\mathbf{F}$ corresponding to the non-zero components of $\hat{\boldsymbol{\rho}}$, the similarity between the measurements $\boldsymbol{\upsilon}$ and a column $\mathbf{x}$ should be minimal for $\forall\mathbf{x} \in \tilde{\mathbf{F}}^c$, where $\tilde{\mathbf{F}}^c$ represents a complement set of $\tilde{\mathbf{F}}$ in $\mathbf{F}$. The similarity can be indirectly checked by comparing the solutions of $\boldsymbol{\rho}_* = \tilde{\mathbf{F}}^{-1}\boldsymbol{\upsilon}$ and $\mathbf{v}_* = \tilde{\mathbf{F}}^{-1}\mathbf{x}$. Interestingly, SBL cost function Eq. (12) measures the similarity between $\boldsymbol{\rho}_*$ and $\mathbf{v}_*$; and then the necessary condition for local minima of SBL can be derived as follows [7]:

$$\sum_{i\neq j}\frac{\Re(\mathbf{v}_*(i))\Re(\mathbf{v}_*(j))}{\Re(\boldsymbol{\rho}_*(i))\Re(\boldsymbol{\rho}_*(j))} \leq 0, \quad \forall\mathbf{x} \in \tilde{\mathbf{F}}^c, \quad (14)$$

where $\Re$ indicates real parts. Eq. (14) implies that the sign patterns of $\mathbf{v}_*$ and $\boldsymbol{\rho}_*$ should be very different at local minimizers of SBL. Eq. (14) is the additional constraint which filters out the most of the local minimizers in the case of MAP approach. For detailed proof, refer to [7]. Accordingly, SBL can achieve sparse solution with much higher probability than conventional MAP estimation.

3. IMPLEMENTAL ISSUES

To successfully employ SBL into dynamic MRI, we have to consider the characteristic of dynamic MRI. Unlike the case where we have derived the optimality of SBL, it is not easy to make the most of pixels become exact zero. More specifically, sparsifying transform such as Fourier transform along temporal direction makes the spectrum sparse, but not the image along phase encoding direction. In order to resolve this problem, a discrete integrator $\mathbf{K}_x$ was applied along phase encoding direction. Then, the sensing matrix $\mathbf{F}$ is redefined as $\mathbf{F} = \mathbf{F}_x\mathbf{F}_t\mathbf{K}_x$ and the unknown coefficients correspond to finite difference. This finite difference scheme allows many of pixels to approach to zero. In contrast, we observed that this kind of spatial sparsifying transfrom was not necessary in k-t FOCUSS especially solving l_1 minimization, as will be discussed later. This different behavior comes from the update rules of SBL and k-t FOCUSS. k-t FOCUSS updates $\boldsymbol{\Theta}$ with power factor of 1 in Eq. (3), while SBL updates $\boldsymbol{\Theta}$ with power factor of 2 in Eq. (13). This induces pruning process relatively slowly in k-t FOCUSS at $p = 1$. Therefore, as done in SBL, at $p = 0$, a discrete integrator can also improve the performance of k-t FOCUSS but the improvement was minimized as shown in Figure 1. Note that the improvements of SBL over k-t FOCUSS are derived from the denominator of Eq. (13) that is not a constant and adapted from the covariance update.

4. EXPERIMENTAL RESULTS

We have acquired 25 frames of full k-space data from a cardiac cine of a patient at a 1.5 T Philips scanner. The field of view (FOV) was $345.00\times270.00mm^2$, and the matrix size for scanning was 256×220, which corresponds to 256 samples in frequency encoding and 220 phase encoding steps. From these full data sets, we only used 8-fold or 13-fold randomly down sampled data for reconstructions.

First, in Figure 1, we compared the performance of k-t FOCUSS and k-t SBL at 8-fold down sampling. As previously mentioned, in k-t FOCUSS at $p = 1$, a discrete integrator does not improve image qualities. In both cases with and without a discrete integrator, heart motions are blurred as indicated with arrows. Meanwhile, in k-t FOCUSS at $p = 0$ and k-t SBL, when a discrete integrator is not used, line pattern artifacts are observed. However, using a discrete integrator, the x-t images can be correctly recovered with finer heart motions. Furthermore, in MSE plots, it is confirmed that k-t SBL outperforms k-t FOCUSS. At all of the time points, k-t SBL provides the smallest errors.

In Figure 2, the reconstructed images from k-t FOCUSS ($p = 1$) and k-t SBL are compared at 13-fold down sampling. k-t SBL reconstruction shows finer image structures. MSE plot also shows the smaller error in k-t SBL at all of the time points.

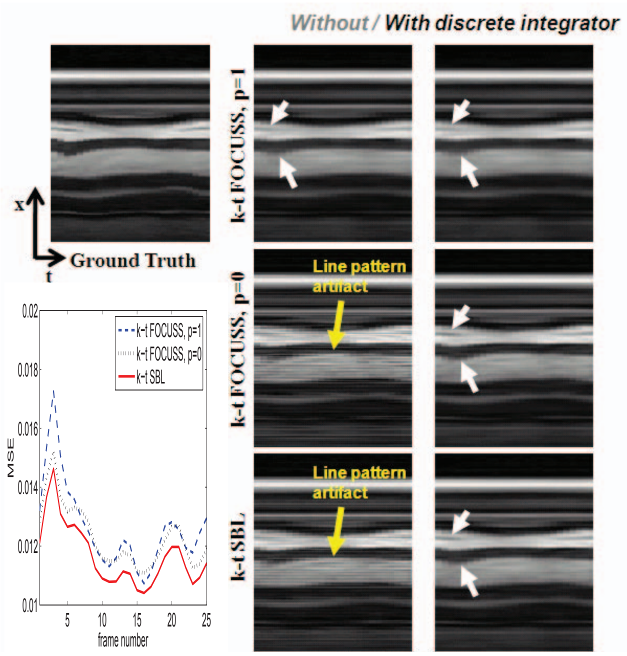

Fig. 1. Reconstructed x-t images and MSE plot for k-t FOCUSS and k-t SBL at 8-fold down sampling.

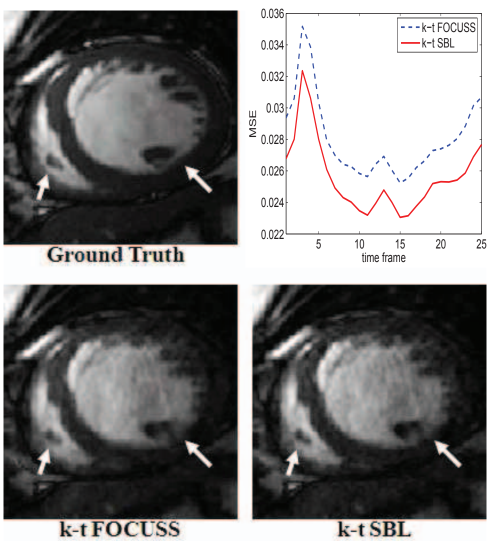

Fig. 2. Reconstructed images and MSE plots for k-t FOCUSS and k-t SBL at 13-fold down sampling.

5. CONCLUSION

This paper described an optimal dynamic MR imaging algorithm derived from an empirical Bayesian method. Instead of specifying l_p $(p \leq 1)$ norm, assuming a flexible prior which can be updated from measured data, more accurate sparse solution could be obtained. Furthermore, incorporating a discrete integrator into our algorithm, the sparse solution could accurately represent dynamic MR images.

6. REFERENCES

[1] D. L. Donoho and M. Elad, "Optimally sparse representation in general (non-orthogonal) 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.

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

[3] M. Lustig, D. Donoho, and J.M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn. Reson. Med*, vol. 58, no. 6, pp. 1182, 2007.

[4] Hong Jung, Jong Chul Ye, and Eung Yeop Kim, "Improved k-t BLAST and k-t SENSE using FOCUSS," *Physics in Medicine and Biology*, vol. 52, no. 11, pp. 3201–3226, June 2007.

[5] Hong Jung, Kyunghyun Sung, Krishna S. Nayak, Eung Yeop Kim, and Jong Chul Ye, "k-t FOCUSS: a general compressed sensing framwork for high resolution dynamic mri," *Magnetic Resonance in Medicine*, vol. 61, pp. 103–116, Jan 2009.

[6] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic 0-minimization," *IEEE Trans. Medical Imaging*, vol. 28, no. 1, pp. 106–121, 2009.

[7] David Paul Wipf, *Bayesian Methods for Finding Sparse Representations*, Ph.d. dissertation, University of California, Sand Diego, San Diego, CA, 2006.