ACCELERATING DYNAMIC MAGNETIC RESONANCE IMAGING BY NONLINEAR SPARSE CODING

Ukash Nakarmi[1], Yihang Zhou[1], Jingyuan Lyu[1], Konstantinos Slavakis[1], Leslie Ying[1, 2]
{ukashnak, yihnagzh, jlv27, kslavaki, leiying}@buffalo.edu
Dept. of Electrical Engineering[1], Dept. of Biomedical Engineering[2],
University at Buffalo, State University of New York

ABSTRACT

Although being high-dimensional, dynamic magnetic resonance images usually lie on low-dimensional manifolds. Nonlinear models have been shown to capture well that latent low-dimensional nature of data, and can thus lead to improvements in the quality of constrained recovery algorithms. This paper advocates a novel reconstruction algorithm for dynamic magnetic resonance imaging (dMRI) based on nonlinear dictionary learned from low-spatial but high-temporal resolution images. The nonlinear dictionary is initially learned using kernel dictionary learning, and the proposed algorithm subsequently alternates between sparsity enforcement in the feature space and the data-consistency constraint in the original input space. Extensive numerical tests demonstrate that the proposed scheme is superior to popular methods that use linear dictionaries learned from the same set of training data.

Index Terms — Sparse coding, kernel dictionary learning, compressed sensing, dynamic MRI

1. INTRODUCTION

Accelerated dMRI is desirable for high-temporal and spatial resolutions [1][2]. Compressed sensing (CS) has shown potential in addressing the tradeoff between temporal and spatial resolutions in dMRI. Both sparsity [3][4][5] and/or low-rank [6][7] properties have been widely exploited as prior constraints in CS recovery. Among sparsity-cognizant MRI recovery methods, learning dictionaries have shown advantages over many fixed sparsifying transforms [8]-[12].

Lately, few works have studied kernel-based nonlinear dictionary learning (DL) [13]-[19] to capture the intrinsic nonlinear correlations in signals often neglected by classical linear models. Among these works, kernel CS [19] not only learns the dictionary but also reconstructs the signal in the feature space, and then finds the pre-image in the original space. However, as with most machine learning algorithms, a sufficient number of training data is very important for kernel CS. In MRI, only low-spatial but high-temporal resolution images (or vice versa) are available as training data. However, those data are not rich enough to capture the features of the high-spatial and temporal resolution images. As such, kernel CS faces severe obstacles when applied to dMRI.

Recent works [20]-[23] investigated applying kernel principal component analysis (PCA) to reconstruct MR or dynamic MR images and showed improvements over linear (conventional) PCA. Motivated by the success of kernel K-SVD in yielding more compact signal representations than kernel PCA [14], this paper studies the use of kernel DL in dMRI recovery. A novel dMRI reconstruction method is advocated based on nonlinear DL and kernel K-SVD. Similar to the conventional linear-dictionary-based dMRI reconstruction methods, nonlinear dictionaries are learned in a high-dimensional feature space using the high-temporal but low-spatial resolution dynamic images acquired in the original input space. Our method comprises nonlinear dictionary learning, sparsity enforcement in the feature space, and preimaging back to the input space for data consistency.

The rest of the paper is organized as follows. In section 2 the proposed method and a detailed description of its steps are presented. Section 3 provides numerical tests, and section 4 concludes the paper.

2. THEORY AND METHODS

As in CS with linear dictionary learning, we are interested in reconstructing an image series while a nonlinear dictionary is learned from training data. Since the computational complexity of DL algorithms is usually high, practical reconstruction models often enforce 'patch-level' sparsity. Data patches can be extracted along the spatial direction, the temporal direction, or both. Given the nonlinear transform $\phi:\chi\rightarrow H$, $\phi:\mathbf{x}\rightarrow\phi(\mathbf{x})$, from the input space χ to the high-dimensional feature space H, and using R_t to represent the operator for extracting the t^{th} patch, the vector of the dynamic image series $\mathbf{x}$ can be reconstructed by the following kernel-based DL task

$$\min_{\mathbf{x},\mathbf{D},\{\mathbf{c}_t\}} \quad \mathrm{TV}(\mathbf{x})+\lambda_1\left\|\mathbf{y}\text{-}\mathbf{F}_u\mathbf{x}\right\|_2^2+\lambda_2\sum_t\left\|\phi(R_t\mathbf{x})\text{-}\phi(\mathbf{D})\mathbf{c}_t\right\|_2^2$$
$$s.t.\quad \left\|\mathbf{c}_t\right\|_0\leq\tau,\ \forall l \tag{1}$$

where the first term is the total variation (TV) of the image series, $\mathbf{y}$ is the undersampled k-space data, $\mathbf{F}_u$ is the Fourier transform with undersampling, $\mathbf{c}_t$ is the sparse coefficient, and $\phi(\mathbf{D})=\left[\phi(\mathbf{d}_1),\phi(\mathbf{d}_2),\ldots,\phi(\mathbf{d}_N)\right]$ is the dictionary in the feature space. To solve the optimization problem in (1), the following three steps are taken: 1) nonlinear DL; 2) sparsity enforcement; and 3) preimaging under the data-consistency constraint.

2.1 Nonlinear dictionary learning

Let $\mathbf{P} = [\mathbf{p}_1\ \mathbf{p}_2\ , \ldots, \mathbf{p}_T]$ be a set of T training signals. Here, we use the temporal signal at a certain spatial location of a set of high-temporal, low-spatial resolution dynamic images as the training signal $\mathbf{p}_t$. Accordingly, R_t in (1) extracts the temporal signal at a particular spatial location of dMRI. Consequence of the Representer's theorem, the learned dictionary $\phi(\mathbf{D})$ should lie in the subspace spanned by the training data, the dictionary in the feature space can be represented as $\phi(\mathbf{D}) = \phi(\mathbf{P})\mathbf{B}$. In the DL step, our objective is to find $\mathbf{B}$ and $\mathbf{W}=\{\boldsymbol{\omega}_i\}$ such that

$$\min_{\mathbf{B},\mathbf{W}} \sum_t \|\phi(\mathbf{P})\text{-}\phi(\mathbf{P})\mathbf{B}\mathbf{W}\|_F^2 \ s.t.\ \ \|\boldsymbol{\omega}_t\|_0 \le \tau. \tag{2}$$

However, it is not computationally feasible to solve this problem because of two major restrictions: i) In most of the cases the nonlinear map ϕ is not explicitly defined, and ii) the dimension of feature space is prohibitively large, even infinite. Let a polynomial kernel function be $k(\mathbf{p}_i,\mathbf{p}_j) = (\langle \mathbf{p}_i,\mathbf{p}_j \rangle + c)^d = \langle \phi(\mathbf{p}_i),\phi(\mathbf{p}_j) \rangle$ and the kernel matrix $\mathbf{K}_p = \begin{pmatrix} k(\mathbf{p}_1,\mathbf{p}_2) & \cdots & k(\mathbf{p}_1,\mathbf{p}_T) \\ \vdots & \ddots & \vdots \\ k(\mathbf{p}_T,\mathbf{p}_1) & \cdots & k(\mathbf{p}_T,\mathbf{p}_T) \end{pmatrix}$. Based on Mercer's theory, (2) takes the form of

$$\|\phi(\mathbf{P}) - \phi(\mathbf{P})\mathbf{B}\mathbf{W}\|_F^2 = \mathrm{tr}\big((\mathbf{I}-\mathbf{B}\mathbf{W})^T \mathbf{K}_p (\mathbf{I}-\mathbf{B}\mathbf{W})\big), \tag{3}$$

rendering computations free from explicit transformation.

The DL step alternates between the calculation of $\mathbf{B}$ (dictionary update) and $\mathbf{W}$ (nonlinear sparse encoding).

<u>Nonlinear sparse encoding</u>: This step is similar to classical orthogonal matching pursuit (OMP) [27] in feature space termed as Kernel OMP [16]. In this step, the overcomplete dictionary $\mathbf{B}$ is fixed and our objective is to find the coefficients $\boldsymbol{\omega}_t, \forall t$, corresponding to τ columns of $\mathbf{B}$. Let I_i denote the set of selected columns of $\mathbf{B}$, $\hat{\mathbf{p}}_t^i$ be the approximation, $\boldsymbol{r}_{p_t}^i$ the residual, at i[th] iteration,

$$\mathrm{Proj}_q(\mathbf{r}_{pt}^i) = \big(\mathbf{k}_{pt} - \mathbf{K}_p (\hat{\mathbf{p}}_t^i)^T\big)\mathbf{b}_q,\ q \notin I_{i-1}, \tag{4}$$

where, $\mathbf{k}_{pt} = [k(\mathbf{p}_t,\mathbf{p}_1), k(\mathbf{p}_t,\mathbf{p}_2), \ldots, k(\mathbf{p}_t,\mathbf{p}_T)]$, at i=0, I_i=0, $\hat{\mathbf{p}}_t^0$ =0 and $\mathbf{r}_{pt}^0 = \phi(\mathbf{p}_t)$. The index set I_i is updated as $I_i = I_{i-1} \cup \arg\max_q |\mathrm{Proj}_q|$. Letting $\mathbf{B}_{Ii}$ denote the set of columns of $\mathbf{B}$ indexed by I_i, then, the coefficient vector $\boldsymbol{\omega}_t^i$ at i[th] iteration is calculated as

$$\boldsymbol{\omega}_t^i = (\mathbf{B}_{Ii}{}^T \mathbf{K}_p \mathbf{B}_{Ii})^{-1} (\mathbf{k}_{pt} \mathbf{B}_{Ii})^T. \tag{5}$$

The approximation is updated as $\hat{\mathbf{p}}_t^i = \mathbf{B}_{Ii}\boldsymbol{\omega}_t^i$. The process then repeats τ times for each of the training signal.

<u>*Dictionary update*</u>*:* The dictionary update process is carried out in a similar fashion as in K-SVD [24]. Given $\boldsymbol{W}$ from the nonlinear sparse encoding step, the dictionary approximation penalty is calculated as

$$\|\phi(\mathbf{P}) - \phi(\mathbf{P})\mathbf{B}\mathbf{W}\|_F^2 = \|\phi(\mathbf{P})\mathbf{E}_q - \phi(\mathbf{P})\mathbf{M}_q\|_F^2 \tag{6}$$

where $\mathbf{E}_q = \Big(\mathbf{I} - \sum_{j \ne q} \mathbf{b}_j \boldsymbol{\omega}^j\Big)$, $\mathbf{M}_q = \big(\mathbf{b}_q \boldsymbol{\omega}^q\big)$, $\mathbf{b}_q$ is q[th] column of $\mathbf{B}$, and $\boldsymbol{\omega}^q$ is the q[th] row of $\mathbf{W}$. Similarly as in [24], the shrinked-error and contribution matrices are estimated as $\mathbf{E}_q^R = \mathbf{E}_q \Omega_q, \mathbf{M}_q^R = \mathbf{M}_q \Omega_q$, respectively, where Ω_q is the shrinkage matrix with binary entries. Hence the penalty term in (6) is reduced to

$$\|\phi(\mathbf{P})\mathbf{E}_q - \phi(\mathbf{P})\mathbf{M}_q\|_F^2 = \|\phi(\mathbf{P})\mathbf{E}_R^q - \phi(\mathbf{P})\mathbf{b}_q \boldsymbol{\omega}_R^q\|_F^2. \tag{7}$$

Based on the principle that $\phi(\mathbf{P})\mathbf{b}_q\boldsymbol{\omega}_R^q$ is rank-1 matrix, as in KSVD, and relating the SVD of $\phi(\mathbf{P})\mathbf{E}_q$ as the Eigen-decomposition of kernel matrix as,

$$(\mathbf{E}_R^q)^T \mathbf{K}_p (\mathbf{E}_R^q) = \mathbf{V}\Delta\mathbf{V}^T, \tag{8}$$

the optimal solution $\mathbf{b}_q$, $\boldsymbol{\omega}^q$ of (7) can be obtained as, $\boldsymbol{\omega}_R^q = \sqrt{\lambda}\ \mathbf{v}_1^T$ and $\mathbf{b}_q = \big(\sqrt{\lambda}\big)^{-1} \mathbf{E}_R^q \mathbf{v}_1$, where λ is the Eigen-value corresponding to the first Eigen-vector in (8). It should be noted here that, the dictionary update is carried out for each column of $\mathbf{B}$. The sparse encoding and dictionary update process is then iterated.

2.2 Sparsity enforcement

Given the learned nonlinear dictionary $\mathbf{B}$ from Section 2.1, nonlinear sparse encoding enforces the sparsity constraints on the desired dynamic image sequence. For the desired image $\mathbf{x}$, we find the sparse representation coefficients $\mathbf{E} = \{\boldsymbol{\alpha}_t\}$, such that

$$\min_{\boldsymbol{\alpha}_t} \ \|\phi(R_t\mathbf{x}) - \phi(\mathbf{P})\mathbf{B}\boldsymbol{\alpha}_t\|_2^2 \quad s.t. \quad \|\boldsymbol{\alpha}_t\|_0 \le \tau, \forall t. \tag{9}$$

We can see that this problem is similar to the nonlinear sparse encoding step in the section 2.1. Letting $\tilde{\mathbf{x}}_t = R_t\mathbf{x}$, we modify (4) and (5) as

$$\mathrm{Proj}_q(\mathbf{r}_{xt}^i) = \big(\mathbf{k}_{xt} - \mathbf{K}_p(\hat{\mathbf{x}}_t^i)^T\big)\mathbf{b}_q,\ q \notin I_i, \tag{10}$$

$$\boldsymbol{\alpha}_t^i = (\mathbf{B}_{Ii}{}^T \mathbf{K}_p \mathbf{B}_{Ii})^{-1} (\mathbf{k}_{xt}\mathbf{B}_{Ii})^T, \tag{11}$$

where $\mathbf{k}_{xt} = [k(\mathbf{x}_t,\mathbf{p}_1), k(\mathbf{x}_t,\mathbf{p}_2), \ldots, k(\mathbf{x}_t,\mathbf{p}_T)]$ and $\hat{\mathbf{p}}_t^i \to \hat{\mathbf{x}}_t^i = \mathbf{B}_{Ii}\boldsymbol{\alpha}_t^i$, and the sparse coefficient can be computed using the same step as in nonlinear sparse encoding.

2.3 Preimaging for data consistency

From section 2.1 we computed the basis $\phi(\mathbf{P})\mathbf{B}$, which is a nonlinear function of training data from input space, and sparse coefficient α_t from section 2.2 for the sparse representation of dynamic images in features space such that the dictionary penalty term (9) is satisfied. However, from (2) and (9) it should be noted the sparse coefficients are in the feature space whereas the image data and hence the data consistency term is in the input space. So it is obligatory to project the sparsity enforced images from feature space back into the input space, the so called preimaging. Let $\mathbf{\Gamma}$ be defined such that $\mathbf{\Gamma} = \mathbf{BE}$. For a polynomial kernel function with $c \geq 0$ and $d = odd$, there exists an invertible function f_k such that, $k(\mathbf{x}_i, \mathbf{x}_j) = f_k(\langle \mathbf{x}_i, \mathbf{x}_j \rangle)$. Then the preimage $\tilde{\mathbf{z}}_n \in \chi$ of $\phi(\tilde{\mathbf{x}}_n)$ can be calculated as,

$\tilde{\mathbf{z}}_n = \sum_{i=1}^{M} \langle \tilde{\mathbf{z}}_n, \xi_i \rangle \xi_i = \sum_{i=1}^{M} f_k^{-1}\left(k(\tilde{\mathbf{z}}_n, \xi_i)\right)\xi_i$. Hence,

$$\tilde{\mathbf{z}}_n = \sum_{i=1}^{M} f_k^{-1}\left(\sum_{t=1}^{T} \gamma_t^n \; k(\mathbf{p}_t, \xi_i)\right)\xi_i \qquad (12)$$

where γ_t^n is the coefficient of $\phi(\tilde{\mathbf{x}}_n)$ on the $\phi(\mathbf{p}_t)$, given by $\gamma_t^n = \mathbf{\Gamma}(\mathrm{t}, \mathrm{n})$, and ξ_i is any orthonormal basis in the input space. Once we obtain $\tilde{\mathbf{z}}_n$ at all spatial locations, we then use the Bregman alternating direction method of multipliers [26] to alternate between the conjugate gradient and TV enforcement.

Finally step 2 and step 3 are repeated until convergence. The steps involved in the proposed method are summarized in algorithm 1.

Algorithm 1.

Step 1: **Nonlinear Dictionary Learning**.

a. Nonlinear Sparse Encoding.

For, τ times

Calculate: Sparse Coefficients $\boldsymbol{\omega}_t^i$ using (8)

Calculate i^th approximation: $\hat{\mathbf{p}}_t^i = \mathbf{B}_{I_i}\boldsymbol{\omega}_t^i$

Update column index set I_i

b. Dictionary Update.

Update: $\mathbf{b}_q = (\sqrt{\lambda})^{-1}\, \mathbf{E}_R^q \mathbf{v}_1$ and $\boldsymbol{\omega}_R^q = \sqrt{\lambda}\, \mathbf{v}_1^T$

Iterate a. and b.

Step 2: **Sparsity Enforcement**.

For each undersampled signal $\tilde{\mathbf{x}}_t$

For, i=1: τ

Calculate: $\mathrm{Proj}_q(\mathbf{r}_{xt}^i)$

Calculate: Sparse Coefficients:

$$\boldsymbol{\alpha}_t^i = (\mathbf{B}_{I_i}{}^T \mathbf{K}_p \mathbf{B}_{I_i})^{-1} (\mathbf{k}_{xt} \mathbf{B}_{I_i})^T,$$

Calculate i^th approximation:

$\hat{\mathbf{x}}_t^i = \mathbf{B}_n \boldsymbol{\alpha}_t^i$

Update column index set I_i

End

End

Step 3: **Preimaging for Data Consistency**

Calculate: $\mathbf{\Gamma} = \mathbf{BE}$

For each desired sparse signal:

Calculate preimage $\tilde{\mathbf{z}}_n$

Use ADMM to enforce conjugate gradient and TV

Iterate : Step 2 and Step 3.

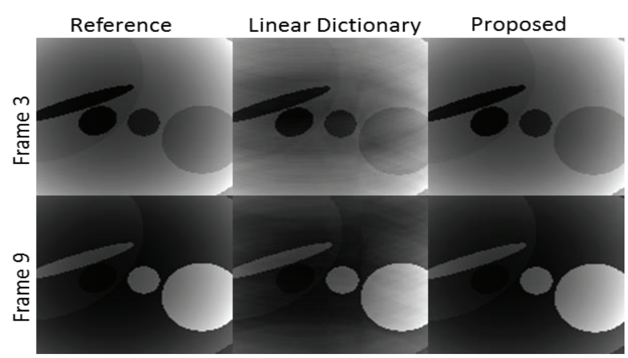

Figure 1 Phantom results comparison

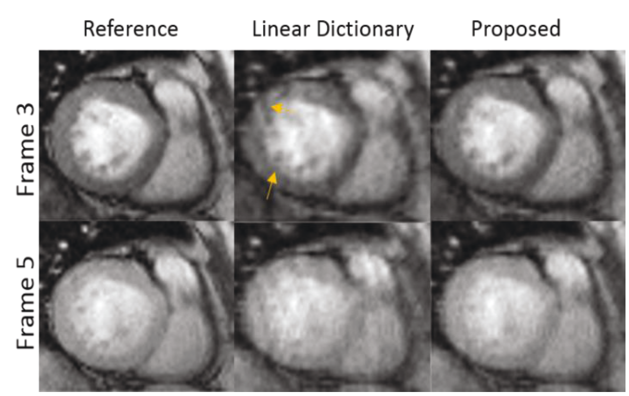

Figure 2 Reconstruction results comparison for cardiac ASL data

3. SIMULATIONS AND RESULTS

We used two data sets to evaluate the proposed method: Simulated data from a numeric liver phantom, and the in-vivo cardiac ASL data. Data matrix: 120 x120 x 25, 100x120x12; reduction factor R = 5, 3; 1–D random under sampling, c=5,
d=3; # training signals =1,500, 2,000; overcomplete dictionary size = 1,500, 2,000 were used for liver phantom and cardiac data, respectively. Figure 1 shows the simulation results for numeric liver phantom. We compare our results with linear dictionary learning method based on K-SVD. Results show that our method outperforms the conventional linear method. For the phantom results, the linear method is not able to remove the aliasing artifacts as effectively as the nonlinear method. Figure **2** shows the results for the myocardial region of the cardiac data. Due to space constraints, we present only frames 2 and 5. It can be seen that the proposed method is able to preserve more spatial structures and reduce more aliasing artifacts than the linear method as indicated by the yellow pointers.

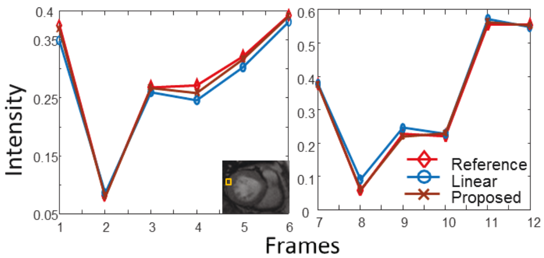

Figure 3 Temporal Intensity Curve

Kinetic information is also equally important in ASL imaging to provide the perfusion map. For better visualization we show the temporal curve of a particular ROI (shown in inset picture) from myocardium region in Figure 3. We can see that the proposed method follows the reference curve more consistently than the conventional linear method.

4. CONCLUSION

In this paper, we proposed a novel nonlinear dictionary learning method within the framework of kernel methods for dynamic MRI. The proposed method integrates the principles of kernel dictionary learning and sparse representation in the feature space to find efficient sparse bases for dynamic MR images. Numerical tests have shown promising results. At this stage, a single predefined kernel was used to capture nonlinear correlations. However, we are currently examining the more challenging case of multiple-kernel learning, as well as the combination of both linear and nonlinear kernels, to capture inherent linear correlations, intrinsic nonlinear features, and to construct a generalized nonlinear learning framework for various types of dMRI signals.

Acknowledgement:
This work is supported in part by the National Science foundation CBET-1265612, the National Institute of Health R21EB020861. The authors would also like to thank Dr. Pascal Spincemaille and Dr. Jie Zheng for the data used in this paper.

REFERENCES

[1] Z.-P. Liang; P. C. Lauterbur, "An efficient method for dynamic magnetic resonance imaging," *IEEE Trans. Medical Imaging*, vol.13, no.4, pp.677,686, Dec 1994.
[2] J. Tsao and S. Kozerke, "MRI temporal acceleration techniques," *J. Magn. Reson. Imag.,* vol. 36, pp. 543-560, 2012.
[3] M. Lustig, J. M. Santos, D. L. Donoho, et al., "k-t SPARSE: High frame rate dynamic MRI exploiting spatiotemporal sparsity," *in Proc. ISMRM*, pp. 2420 , 2006.
[4] U. Gamper, P. Boesiger and S. Kozerke, "Compressed sensing in dynamic MRI," *Magn. Reson. Med.*, vol. 59, pp. 365-373, 2008.
[5] D. Liang, E.V.R. Dibella, R.R. Chen, et al., "k-t ISD: Dynamic cardiac MR imaging using compressed sensing with iterative support detection," *Magn. Reson. Med.*, vol. 68, pp. 41-53, 2012.
[6] Zhao B, Haldar JP, Christodoulou AG, Liang Z-P. Image reconstruction from highly undersampled (**k**, *t*)-space data with joint partial separability and sparsity constraints. *IEEE Trans. Medical Imaging*. 2012;31(9):1809-1820.
[7] Lingala, S. G., Hu, Y., DiBella, E., & Jacob, M. (2011). Accelerated dynamic MRI exploiting sparsity and low-rank structure: kt SLR. *IEEE Trans. Medical Imaging,* 30(5), 1042-1054.
[8] S. Ravishankar and Y. Bresler, "MR Image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Medical Imaging,* vol.30, no.5, pp.1028,1041, May 2011.
[9] M. Doneva, P. Börnert, H. Eggers, et al., "Compressed sensing reconstruction for magnetic resonance parameter mapping," *Magn. Reson. Med.,* vol. 64, no. 4, pp. 1114-1120, 2010.
[10] S. G. Lingala and M. Jacob, "A Blind Compressive Sensing Framework for Accelerated Dynamic MRI," *Proc. 9th IEEE Int. Symp. Biomed. Imag. (ISBI),* Barcelona, pp. 1060-1063, 2012.
[11] Y. Wang and L. Ying. "Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary." *IEEE Trans. Biomed. Eng.*, vol. 61, pp. 1109-1120, 2014.
[12] Caballero, J., Price, A. N., Rueckert, D., & Hajnal, J. V. (2014). Dictionary learning and time sparsity for dynamic MR data reconstruction. *IEEE Trans. Medical Imaging, 33*(4), 979-994.
[13] B. Schlkopf and A. J. Smola, "Learning with kernels: support vector machines, regularization, optimization, and beyond." MIT Press, Boston, 2001.
[14] B. Scholkopf, A. Smola, and K.R. Muller, "Kernel principal component analysis," *Proc. 1997 ICANN LNCS*, pp. 583-588.
[15] S. Mika, B. Schölkopf, A. Smola, K.L. Müller, M. Scholz, and G. Rätsch, "Kernel PCA and de-noising in feature spaces," *Adv. Neural. Inf. Process. Syst.*, vol. 11, pp. 536-542, 1999.
[16] Nguyen, H.V.; Patel, V.M.; Nasrabadi, N.M.; Chellappa, R., "Kernel dictionary learning," in *Proc. ICASSP,* pp. 2021-2024, 25-30 March 2012.
[17] A. Golts and M. Elad, "Linearized kernel dictionary learning", arXiV: 1509.05634v1, Sept, 2015.
[18] H. Qi and S. Hughes, "Using the kernel trick in compressive sensing Accurate signal recovery from fewer measurements." *Proc. ICASSP*, pp. 3940-3943.
[19] A. F. Pourkamali and S. Hughes, "Kernel compressed sensing," *Proc. ICIP*, pp.494-498, 2013
[20] Y. Zhou, Y. Wang and L. Ying, "A kernel-based compressed sensing approach to dynamic MRI from highly undersampled data." *Proc. ISBI*, pp. 310-313, 2013.
[21] J. F. M. Schmidt and S. Kozerke, "MR image reconstruction exploiting nonlinear transforms,'' *Proc. ISMRM*, pp. 746, 2014.
[22] Y. Wang and L. Ying, "Undersampled dynamic magnetic resonance imaging using kernel principal component analysis," *Proc. EMBC,* pp.1533-1536.
[23] Nakarmi, U.; Wang, Y ; Lyu, J; Ying, L., "Dynamic magnetic resonance imaging using compressed sensing with self-learned nonlinear dictionary (NL-D)," in *Proc. ISBI,* April 2015.
[24] Aharon, M.; Elad, M.; Bruckstein, A., "k -SVD: An algorithm for designing overcomplete dictionaries for sparse representation," in *IEEE Trans. Signal Processing,* ,pp.4311-4322, Nov. 2006.
[25] K. Engan, S. O. Aase, and J. H. Hakon-Husoy, "Method of optimal directions for frame design," in *Proc. ICASSP,* vol. 5, pp. 2443–2446, 1999.
[26] T. Goldstein and S. Osher, "The split Bregman method for L1-regularized problems," *SIAM J. Imag. Sci.,* vol. 2, no.2, pp. 323-343, 2009.
[27] Y. C. Pati, R. Rezaiifar, and P. S. Krishnaprasad, "Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition," in Proc. 27th Annu. Asilomar Conf. Signals Syst. Comput., Pacific Grove, CA, USA, 1993.