Kernel Bi-Linear Modeling for Reconstructing Data on Manifolds: The Dynamic-MRI Case

Gaurav N. Shetty*, Konstantinos Slavakis†, Ukash Nakarmi‡, Gesualdo Scutari§, and Leslie Ying¶
Depts. of *†¶Electrical Engineering and ¶Biomedical Engineering, University at Buffalo, Buffalo, NY 14260, USA
‡Depts. of Electrical Engineering and Radiology, Stanford University, Stanford, CA 94305, USA
§Dept. of Industrial Engineering, Purdue University, West Lafayette, IN 47907-2023, USA
Emails: {*gauravna, †kslavaki, ¶leiying }@buffalo.edu, ‡nakarmi@stanford.edu, §gscutari@purdue.edu

***Abstract*—This paper establishes a kernel-based framework for reconstructing data on manifolds, tailored to fit the dynamic-(d)MRI-data recovery problem. The proposed methodology exploits simple tangent-space geometries of manifolds in reproducing kernel Hilbert spaces, and follows classical kernel-approximation arguments to form the data-recovery task as a bi-linear inverse problem. Departing from mainstream approaches, the proposed methodology uses no training data, employs no graph Laplacian matrix to penalize the optimization task, uses no costly (kernel) preimaging step to map feature points back to the input space, and utilizes complex-valued kernel functions to account for k-space data. The framework is validated on synthetically generated dMRI data, where comparisons against state-of-the-art schemes highlight the rich potential of the proposed approach in data-recovery problems.**

***Index Terms*—Manifold, kernel, signal recovery, dynamic MRI, low rank, sparsity, dimensionality reduction.**

I. Introduction

High-quality and artifact-free image reconstruction is a perennial task in a plethora of imaging technologies, such as dMRI; a high-fidelity visualization technology with widespread applications in cardiac and neuro-imaging [1], [2]. Due to various physiological constraints, MR scanners are usually slow in data acquisition and unable to keep up with organ motions or fluid flow [1]. Moreover, there is a constant need to speed up the data collection process to make it inexpensive and to cause less patient discomfort. The unpleasant end result is the availability of a limited number of measurements/data, which leads in turn to distorted and aliasing image reconstructions [1], [2].

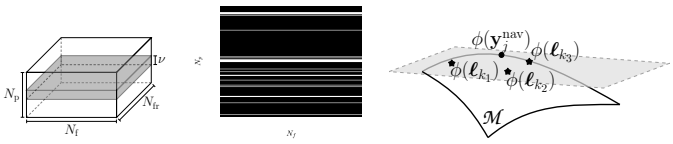
Naturally, a lot of research focuses on reconstruction algorithms that yield high-fidelity image series given the usually inadequate number of measurements from MR scanners. Recent years have seen a paradigm shift to artificial-intelligence (AI) approaches in image reconstruction, such as convolutional neural networks (CNNs) [3], [4] and generative adversarial networks (GANs) [5], [6]. However, these techniques rely on the availability of large number of training data and a time consuming training phase preceding the reconstruction task.

In contrast to AI approaches, this paper focuses on methods that do not use any training data, but explore and exploit instead latent geometries of the observed data. Among such methods, a mainstream approach is compressed sensing [7]–[10]. For instance, [9] estimates first a temporal basis from image series via singular-value decomposition, followed by the estimation of a spatial basis through a sparsity inducing convex optimization task.

Manifold-learning techniques, identifying intrinsic manifold geometries beneath raw data for data recovery, have found applications in face recognition [11], [12], as well as dMRI [13]. Use of local SVD/PCA in data modeling has gathered also interest in applications not necessarily related to MRI [14]. A graph Laplacian matrix, capturing relations within data clouds, is used as a regularizer in a convex recovery problem in [13]. Motivated by the success of kernel methods [15] in extending linear models to non-linear counterparts, the MRI community has also seen the introduction of kernel methods in MRI data recovery tasks [16]–[18]. Notwithstanding, such kernel-based approaches require explicit and often costly (kernel) preimaging steps, and leave the complex-valued nature of the k-space data underused.

The present study builds on [19] to establish its generalization towards manifold geometries in reproducing kernel Hilbert spaces (RKHSs). RKHSs help in better understanding the relations between the input data/measurements by projecting it to a higher dimensional feature space but still using the existing linear algorithms by merely expressing these algorithms in terms of inner products. To promote a computationally efficient scheme, a small set of landmark points is chosen to describe concisely the data cloud in the input space, and then mapped non-linearly to the RKHS feature space. Driven by the continuity of the mapping (kernel) function, it is hypothesized that the mapped feature points lie on or close to a smooth unknown and low-dimensional manifold. Departing from mainstream manifold-learning approaches, where the graph Laplacian operator is used as a regularizer in an inverse problem, e.g., [13], this study relies heavily on the concept of tangent spaces of smooth manifolds to establish a local and data-adaptive method of revealing affine relations among data points. Unlike [19] these affine relations are revealed in the feature space which amount to non-linear operations in the original input space, yielding thus better data approximations according to kernel-method arguments [15], [20]. Costly preimaging operations to solve for vectors in the original input space are avoided via a bi-linear model-

This work was supported in part by the NSF grant 1718796.



(a) (b) (c)

Fig. 1. (a) The $N_{\rm p} \times N_{\rm f} \times N_{\rm fr}$ (k,t)-space. "Navigator data" comprise the gray-colored $\nu \times N_{\rm f} \times N_{\rm fr}$ area of the (k,t)-space. (b) 1-D Cartesian sampling trajectory. (c) In the feature space $\mathcal{H}_\kappa$, mapped landmark points $\{\boldsymbol{\phi}(\boldsymbol{\ell}_{k_i})\}_{i=1}^3$ are affinely combined to describe $\boldsymbol{\phi}(\mathbf{y}_j^{\rm nav})$. All possible affine combinations of $\{\boldsymbol{\phi}(\boldsymbol{\ell}_{k_i})\}_{i=1}^3$ are represented by the gray-coloured area.

ing approach. The framework employs also a dimensionality reduction module to impose a low-rank structure to the data. Furthermore, terms capitalizing on the periodicity that exists often in dMRI are also introduced to achieve high-fidelity image reconstructions. Finally, recently developed convex and non-convex minimization techniques are employed to solve the resultant recovery tasks. It is worth stressing here the points that distinguish the proposed framework from prior art: **i**) No training data are used; **ii**) in contrast to mainstream manifold-learning techniques, no graph Laplacian matrix is used to penalize the optimization task; **iii**) bi-linear terms are used to approximate data; **iv**) no costly (kernel) preimaging step is necessary to map feature points back to the input space; and **v**) complex-valued kernel functions are defined to account for the k-space data. Preliminary numerical tests on synthetically generated dMRI phantom data and comparisons against state-of-the-art methodologies underline the rich potential of the method for high-quality data recovery.

II. DMRI-Data Description & Representation

An MR image $\boldsymbol{\mathcal{X}} \in \mathbb{C}^{N_{\rm p} \times N_{\rm f}}$ ($\mathbb{C}$ is a set of all complex numbers) is observed via $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{N_{\rm p} \times N_{\rm f}}$ at discrete k-space (or frequency domain) locations, spanning an area of $N_{\rm p} \times N_{\rm f}$ where $N_{\rm p}$ and $N_{\rm f}$ stand for number of phase and frequency encoding lines, respectively [1]. Measurements $\boldsymbol{\mathcal{Y}}$ can be viewed as the two-dimensional (2D) discrete Fourier transform of $\boldsymbol{\mathcal{X}}$: $\boldsymbol{\mathcal{Y}} = \mathcal{F}(\boldsymbol{\mathcal{X}})$ [1]. Without any loss of generality, it is assumed that the "low-frequency" part of $\boldsymbol{\mathcal{Y}}$ is located around the center of the $N_{\rm p} \times N_{\rm f}$ area. In dMRI, a temporal dimension is added to yield $\boldsymbol{\mathcal{Y}} \in \mathbb{C}^{N_{\rm p} \times N_{\rm f} \times N_{\rm fr}}$ where $N_{\rm fr}$ represents the number of frames collected over time (Fig. 1a), resulting in the augmented (k,t)-space. The (k,t)-space can be seen as the collection of k-space measurements $\boldsymbol{\mathcal{Y}}_i = \mathcal{F}(\boldsymbol{\mathcal{X}}_j)$, $\forall j \in \{1, \ldots, N_{\rm fr}\}$.

In practice, due to physical and physiological phenomena, k-space measurements are missing, causing severe *under-sampling* of the k-space, which in turn results in distorted, artifact-induced reconstructions of $\boldsymbol{\mathcal{X}}_j = \mathcal{F}^{-1}(\boldsymbol{\mathcal{Y}}_j)$ [1]. The present framework's objective is to reconstruct images from limited k-space measurements but free of aliasing effects and distortions. These limited k-space measurements are usually made along predefined trajectories for efficient acquisition of MR data [2]. One such trajectory is the one-dimensional (1D) Cartesian sampling, exhibited in Fig. 1b, where measurements are recorded only along the "white" lines and the "black" space is assumed to be zero-filled. To this end, the proposed framework considers the availability of heavily sampled low-frequency part at the center of the k-space data, coined *navigator* data which are usually a small number $\nu (\ll N_{\rm p})$ of phase encoded lines ("gray-coloured region" in Fig. 1a), and the highly (pseudo-randomly) undersampled high frequency region as in [16], [17], [19], [21].

To facilitate processing, (k,t)-space data are vectorised. More specifically, $\mathrm{vec}(\boldsymbol{\mathcal{Y}}_j)$ operation stacks one column of $\boldsymbol{\mathcal{Y}}_j$ below the other to form a $N_{\rm k} \times 1$ vector $\mathbf{y}_j = \mathrm{vec}(\boldsymbol{\mathcal{Y}}_j)$, where $N_{\rm k} := N_{\rm p} N_{\rm f}$. To avoid notation clutter, $\mathcal{F}$ still denotes the 2D discrete Fourier transform even when applied to vectorized versions of image frames: $\mathcal{F}[\mathrm{vec}(\boldsymbol{\mathcal{X}}_j)] := \mathrm{vec}[\mathcal{F}(\boldsymbol{\mathcal{X}}_j)] = \mathrm{vec}(\boldsymbol{\mathcal{Y}}_j)$. All vectorized k-space frames are gathered in the $N_{\rm k} \times N_{\rm fr}$ matrix $\mathbf{Y} := [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_{N_{\rm fr}}]$ so that the vectorized image-domain data are $\mathbf{X} := \mathcal{F}^{-1}(\mathbf{Y}) := [\mathcal{F}^{-1}(\mathbf{y}_1), \mathcal{F}^{-1}(\mathbf{y}_2), \ldots, \mathcal{F}^{-1}(\mathbf{y}_{N_{\rm fr}})]$. The navigator data of the jth k-space frame (*cf.* Fig. 1a), are gathered into a $\nu N_{\rm f} \times 1$ vector $\mathbf{y}_j^{\rm nav}$. All navigator data comprise the $\nu N_{\rm f} \times N_{\rm fr}$ matrix $\mathbf{Y}_{\rm nav} := [\mathbf{y}_1^{\rm nav}, \mathbf{y}_2^{\rm nav}, \ldots, \mathbf{y}_{N_{\rm fr}}^{\rm nav}]$.

III. Data Modeling

The navigator data $\mathbf{Y}_{\rm nav}$ carries useful spatio-temporal information in the (k,t) space. However, $N_{\rm fr}$ tends to be usually large and to promote parsimonious data representations, it is desirable to extract a subset $\{\boldsymbol{\ell}_k\}_{k=1}^{N_\ell} \subset \{\mathbf{y}_j^{\rm nav}\}_{j=1}^{N_{\rm fr}}$ (coined *landmark points*; $N_\ell \leq N_{\rm fr}$) which concisely describes the data cloud $\{\mathbf{y}_j^{\rm nav}\}_{j=1}^{N_{\rm fr}}$. Stacking $\{\boldsymbol{\ell}_k\}_{k=1}^{N_\ell}$ as columns in a $\nu N_{\rm f} \times N_{\rm fr}$ matrix, $\boldsymbol{\Lambda} := [\boldsymbol{\ell}_1, \boldsymbol{\ell}_2, \ldots, \boldsymbol{\ell}_{N_\ell}]$ is defined. To reveal potential non-linear spatio-temporal dependencies in $\mathbf{Y}_{\rm nav}$, embedding into a high dimensional (potentially infinite) complex RKHS $\mathcal{H}_\kappa$ *(feature space)* is pursued. The feature map $\boldsymbol{\phi} : \mathbb{C}^{\nu N_{\rm f}} \to \mathcal{H}_\kappa : \mathbf{y}_j^{\rm nav} \mapsto \boldsymbol{\phi}(\mathbf{y}_j^{\rm nav}) := \kappa(\cdot, \mathbf{y}_j^{\rm nav})$ is defined via the reproducing kernel $\kappa(\cdot, \cdot)$ [15]. The celebrated *kernel trick*, $\kappa(\mathbf{y}_i^{\rm nav}, \mathbf{y}_j^{\rm nav}) = \langle \boldsymbol{\phi}(\mathbf{y}_i^{\rm nav}) \mid \boldsymbol{\phi}(\mathbf{y}_j^{\rm nav}) \rangle =: \boldsymbol{\phi}(\mathbf{y}_i^{\rm nav})^{\mathsf{H}} \boldsymbol{\phi}(\mathbf{y}_j^{\rm nav})$, facilitates processing [15], where $\langle \cdot \mid \cdot \rangle$ is the inner product associated with $\mathcal{H}_\kappa$, and H, which denotes the Hermitian transposition of vectors/matrices, is used to simplify notations.

All landmark points, which are mapped to $\mathcal{H}_\kappa$, can be stacked into matrix $\boldsymbol{\Phi}(\boldsymbol{\Lambda}) := [\boldsymbol{\phi}(\boldsymbol{\ell}_1), \boldsymbol{\phi}(\boldsymbol{\ell}_2), \ldots, \boldsymbol{\phi}(\boldsymbol{\ell}_{N_\ell})]$. The kernel matrix $\mathbf{K} = [K_{ij} := \kappa(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j)]$ can be written concisely as $\mathbf{K} := \boldsymbol{\Phi}(\boldsymbol{\Lambda})^{\mathsf{H}} \boldsymbol{\Phi}(\boldsymbol{\Lambda})$. A range of kernel functions is available for complex RKHS, such as the Gaussian $\kappa(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j) = \exp(-\gamma \|\boldsymbol{\ell}_i - \boldsymbol{\ell}_j^*\|^2)$ ($*$ represents vector/matrix conjugation) and the polynomial kernel $\kappa(\boldsymbol{\ell}_i, \boldsymbol{\ell}_j) = (\boldsymbol{\ell}_i^{\mathsf{H}} \boldsymbol{\ell}_j + c)^r$ [22]. The following assumption imposes a structure on $\boldsymbol{\phi}(\mathbf{y}_j^{\rm nav})$; a condition often met in manifold-learning approaches [23].

Assumption 1. Data $\{\boldsymbol{\phi}(\mathbf{y}_j^{\rm nav})\}_{j=1}^{N_{\rm fr}}$ lie on a smooth low-dimensional manifold $\mathcal{M}$ [24] in the high dimensional RKHS $\mathcal{H}_\kappa$ (*cf.* Fig. 1c).

Following As. 1, the concept of tangent spaces of smooth manifolds can be employed to approximate each $\boldsymbol{\phi}(\mathbf{y}_j^{\rm nav})$ as an affine combination of its neighbouring points in the feature space. Motivated by Fig. 1c, where $\mathbf{y}_j^{\rm nav}$ is described by the "gray coloured" area depicting all possible combinations of

$\{\boldsymbol{\phi}(\boldsymbol{\ell}_{k_1}), \boldsymbol{\phi}(\boldsymbol{\ell}_{k_2}), \boldsymbol{\phi}(\boldsymbol{\ell}_{k_3})\}$, the existence of an $N_\ell \times 1$ vector $\boldsymbol{\beta}_j \in \mathbb{C}^{N_\ell}$ that approximates $\boldsymbol{\phi}(\mathbf{y}_j^{\text{nav}}) \approx \boldsymbol{\Phi}(\boldsymbol{\Lambda})\boldsymbol{\beta}_j$ is assumed. Since affine combinations are desired, constraint $\mathbf{1}_{N_\ell}^\intercal \boldsymbol{\beta}_j = 1$ is imposed where $\mathbf{1}_{N_\ell}$ is an all-one $N_\ell \times 1$ vector and $\intercal$ denotes vector/matrix transposition. Since $\boldsymbol{\phi}(\mathbf{y}_j^{\text{nav}})$ is approximated by points in its close vicinity, $\boldsymbol{\beta}_j$ can be considered sparse. Motivated also by standard kernel-approximation arguments [20], a measurement y of the k-space data vector $\mathbf{y}_j$ is modeled here via $y \approx \langle \sum_{i=1}^{N_\ell} \alpha_i \boldsymbol{\phi}(\boldsymbol{\ell}_i) \mid \boldsymbol{\phi}(\mathbf{y}_j^{\text{nav}}) \rangle = \langle \boldsymbol{\Phi}(\boldsymbol{\Lambda})\boldsymbol{\alpha} \mid \boldsymbol{\phi}(\mathbf{y}_j^{\text{nav}}) \rangle$ for some $\boldsymbol{\alpha} \in \mathbb{C}^{N_\ell}$. The previous arguments are put together in $y \approx \langle \boldsymbol{\Phi}(\boldsymbol{\Lambda})\boldsymbol{\alpha} \mid \boldsymbol{\Phi}(\boldsymbol{\Lambda})\boldsymbol{\beta}_j \rangle = \boldsymbol{\alpha}^{\mathsf{H}} \mathbf{K} \boldsymbol{\beta}_j$ s.t. $\mathbf{1}_{N_\ell}^\top \boldsymbol{\beta}_j = 1$. The previous model is summarized into the following hypothesis.

Assumption 2. There exist an $N_{\text{k}} \times N_\ell$ matrix $\mathbf{A}_1$, $N_\ell \times N_{\text{fr}}$ sparse matrix B, and an $N_{\text{k}} \times N_{\text{fr}}$ matrix $\mathbf{E}_1$, gathering all approximation errors, s.t. $\mathbf{Y} = \mathbf{A}_1 \mathbf{K} \mathbf{B} + \mathbf{E}_1$.

The dimensions of the kernel matrix $\mathbf{K}$ is contingent on the number of landmark points chosen and therefore, $\mathbf{K}$ can still be high dimensional depending on the physical characteristics of the acquired data. To impose a low-rank structure in As. 2, and to meet the restrictions imposed by limited computational resources, dimensionality reduction on $\mathbf{K}$ is desirable. To this end, the methodology of [26] is applied. Since $\{\boldsymbol{\phi}(\ell)_j\}_{j=1}^{N_\ell}$ lie on smooth manifold $\mathcal{M}$, and following the earlier discussion, there exists a matrix $\mathbf{W} \in \mathbb{C}^{N_\ell \times N_\ell}$ s.t. $\boldsymbol{\Phi}(\boldsymbol{\Lambda}) \approx \boldsymbol{\Phi}(\boldsymbol{\Lambda})\mathbf{W}$, with $\mathbf{1}_{N_\ell}^\intercal \mathbf{W} = \mathbf{1}_{N_\ell}^\intercal$ (which manifests the desire for affine combinations) and $\text{diag}(\mathbf{W}) = \mathbf{0}$ (which excludes the trivial solution of $\mathbf{W} = \mathbf{I}_{N_\ell}$). To use the kernel trick, $\boldsymbol{\Phi}(\boldsymbol{\Lambda})^{\mathsf{H}} \boldsymbol{\Phi}(\boldsymbol{\Lambda}) \approx \boldsymbol{\Phi}(\boldsymbol{\Lambda})^{\mathsf{H}} \boldsymbol{\Phi}(\boldsymbol{\Lambda}) \mathbf{W} \Rightarrow \mathbf{K} \approx \mathbf{K}\mathbf{W}$. To summarize this step, given $\mathbf{K}$ and $\lambda_W > 0$, solve $\min_{\mathbf{W}} \|\mathbf{K} - \mathbf{K}\mathbf{W}\|_{\text{F}}^2 + \lambda_W \|\mathbf{W}\|_1$ s.to $\mathbf{1}_{N_\ell}^\intercal \mathbf{W} = \mathbf{1}_{N_\ell}^\intercal$ and $\text{diag}(\mathbf{W}) = \mathbf{0}$. This affinely constrained convex optimization task can be solved by [27]. Once $\mathbf{W}$ is determined, for a user-defined integer $d \ll N_\ell$, solve $\min_{\check{\mathbf{K}} \in \mathbb{C}^{d \times N_\ell}} \|\mathbf{K} - \check{\mathbf{K}}\mathbf{W}\|_{\text{F}}^2$ s.to $\check{\mathbf{K}}\check{\mathbf{K}}^{\mathsf{H}} = \mathbf{I}_d$ to reduce the dimension of $\mathbf{K}$ to d while preserving the underlying manifold geometry in the original high-dimensional space. The constraint $\check{\mathbf{K}}\check{\mathbf{K}}^{\mathsf{H}} = \mathbf{I}_d$ is used to exclude the trivial solution of $\check{\mathbf{K}} = \mathbf{0}$. The solution of the previous task is the Hermitian transpose of the matrix comprising the d minimal eigen-vectors of $(\mathbf{I}_{N_\ell} - \mathbf{W})(\mathbf{I}_{N_\ell} - \mathbf{W})^{\mathsf{H}}$. Other methods for reducing the dimension of $\mathbf{K}$, such as extracting a random subset of its rows, are reserved for a future journal publication due to lack of space. The previous discussion raises the need to unfold the kernel matrix $\check{\mathbf{K}}$ to its original dimensions. The following "decompression" hypothesis establishes a linear relation between $\mathbf{K}$ and $\check{\mathbf{K}}$.

Assumption 3. There exist an $N_\ell \times d$ matrix $\mathbf{A}_2$ and an $N_\ell \times N_\ell$ matrix $\mathbf{E}_2$, gathering all approximation errors, s.t. $\mathbf{K} = \mathbf{A}_2 \check{\mathbf{K}} + \mathbf{E}_2$.

Algorithm 1 Recovering dMRI data

Input: Available are $\mathcal{S}(\mathbf{Y})$ and navigator data $\mathbf{Y}_{\text{nav}}$. Choose $\lambda_1, \lambda_2, \lambda_3, C_D, \tau_D, \tau_B > 0$, $\zeta \in (0,1)$ and $\gamma_0 \in (0,1]$.

Output: Extract the limit points $\mathbf{D}_*$ and $\mathbf{B}_*$ of sequences $(\mathbf{D}_n)_n$ and $(\mathbf{B}_n)_n$, respectively, and recover the dMRI data as $\hat{\mathbf{X}} := \mathbf{D}_* \check{\mathbf{K}} \mathbf{B}_*$.

1: Identify landmark points $\boldsymbol{\Lambda}$ from the columns of $\mathbf{Y}_{\text{nav}}$ according to [25].
2: Compute the kernel matrix $\mathbf{K}$ from landmark points $\boldsymbol{\Lambda}$.
3: Compute the "compressed" $\check{\mathbf{K}}$ (*cf.* Sec. III).
4: Arbitrarily fix $(\mathbf{D}_0, \mathbf{B}_0, \mathbf{Z}_0)$ and set $n = 0$.
5: **while** $n \geq 0$ **do**
6: Available are $(\mathbf{D}_n, \mathbf{B}_n, \mathbf{Z}_n)$ and γ_n.
7: Let $\gamma_{n+1} := \gamma_n(1 - \zeta\gamma_n)$.
8: Obtain $\hat{\mathbf{D}}_n$ of (2a) and $\hat{\mathbf{B}}_n$ of (2b), and the (i,j)th entry of $\hat{\mathbf{Z}}_n$, $\forall (i,j)$, via $\mathbf{A}_{\text{aux}} := \mathcal{F}_t(\mathbf{D}_n \check{\mathbf{K}} \mathbf{B}_n)$ and the following soft-thresholding rule:

$$[\hat{\mathbf{Z}}_n]_{ij} := [\mathbf{A}_{\text{aux}}]_{ij} \left(1 - \frac{\lambda_3/\lambda_1}{\max\{\lambda_3/\lambda_1, |[\mathbf{A}_{\text{aux}}]_{ij}|\}}\right).$$

9: Update

$$(\mathbf{D}_{n+1}, \mathbf{B}_{n+1}, \mathbf{Z}_{n+1}) := (1 - \gamma_{n+1})(\mathbf{D}_n, \mathbf{B}_n, \mathbf{Z}_n) + \gamma_{n+1}(\hat{\mathbf{D}}_n, \hat{\mathbf{B}}_n, \hat{\mathbf{Z}}_n).$$

10: Set n equal to $n+1$ and go to step 5.
11: **end while**

IV. The Non-Convex Inverse Problem and its Algorithmic Solution

By As. 2 and 3, the (k,t)-space measurements $\mathbf{Y}$ can be modeled as $\mathbf{Y} = \mathbf{A}\check{\mathbf{K}}\mathbf{B} + \mathbf{E}$. Upon defining, $\mathbf{D} := \mathcal{F}^{-1}(\mathbf{A})$, $\mathbf{A}\check{\mathbf{K}}\mathbf{B} = \mathcal{F}(\mathbf{D})\check{\mathbf{K}}\mathbf{B}$, and by the virtue of the linearity of $\mathcal{F}$, $\mathbf{A}\check{\mathbf{K}}\mathbf{B} = \mathcal{F}(\mathbf{D}\check{\mathbf{K}}\mathbf{B})$. This establishes the following bi-linear relation between $\mathbf{Y}$ and the unknowns $(\mathbf{D}, \mathbf{B})$: $\mathbf{Y} = \mathcal{F}(\mathbf{D}\check{\mathbf{K}}\mathbf{B}) + \mathbf{E}$. This bi-linear relation holds true also in the image domain: $\mathbf{X} = \mathcal{F}^{-1}(\mathbf{Y}) = \mathbf{D}\check{\mathbf{K}}\mathbf{B} + \mathcal{F}^{-1}(\mathbf{E})$.

As discussed in Sec. II, only a few measurements are available in the (k,t)-space. To account for missing entries, a sampling operator $\mathcal{S}(\cdot)$ is introduced to emulate sampling trajectories. The operation $\mathcal{S}(\mathbf{Y})$ retrospectively under-samples $\mathbf{Y}$ retaining the sampled k-space locations and filling the other locations with zeros. Given the positive real-valued parameters $(\lambda_1, \lambda_2, \lambda_3, C_D)$, the following inverse problem is formulated:

$$\begin{aligned} \min_{(\mathbf{D},\mathbf{B},\mathbf{Z})} \; & \overbrace{\tfrac{1}{2}\left\|\mathcal{S}(\mathbf{Y}) - \mathcal{S}\mathcal{F}(\mathbf{D}\check{\mathbf{K}}\mathbf{B})\right\|_{\text{F}}^2}^{\text{T1}} + \overbrace{\tfrac{\lambda_1}{2}\left\|\mathbf{Z} - \mathcal{F}_t(\mathbf{D}\check{\mathbf{K}}\mathbf{B})\right\|_{\text{F}}^2}^{\text{T2}} \\ & + \underbrace{\lambda_2 \|\mathbf{B}\|_1}_{\text{T3}} + \underbrace{\lambda_3 \|\mathbf{Z}\|_1}_{\text{T4}} \\ \text{s.to} \; & \underbrace{\|\mathbf{D}\mathbf{e}_i\| \leq C_D, \; \forall i \in \{1, \ldots, d\}}_{\text{C1}}; \; \underbrace{\mathbf{1}_{N_\ell}^\intercal \mathbf{B} = \mathbf{1}_{N_{\text{fr}}}^\intercal}_{\text{C2}}; \\ & \mathbf{D} \in \mathbb{C}^{N_{\text{k}} \times d}; \mathbf{B} \in \mathbb{C}^{N_\ell \times N_{\text{fr}}}; \mathbf{Z} \in \mathbb{C}^{N_{\text{k}} \times N_{\text{fr}}}, \end{aligned} \tag{1}$$

where $\mathbf{e}_i$ denotes the ith column of the identity matrix $\mathbf{I}_d$. In addition to the data fit term T1, (1) contains terms T2 and T4 where the auxiliary variable $\mathbf{Z}$ is introduced to capture the periodic process, (*e.g.,* beating heart motion) by imposing sparsity on $\mathcal{F}_t(\mathbf{D}\check{\mathbf{K}}\mathbf{B})$. $\mathcal{F}_t(\cdot)$ performs 1D Fourier Transform of the $1 \times \text{N}_{\text{fr}}$ time profile of every single pixel. The term T3 imposes the sparsity constraint and C2 accounts for the affine

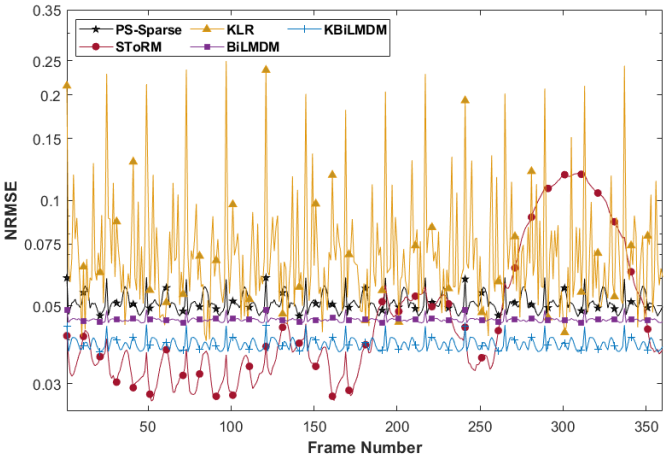


Fig. 2. Frame-wise NRMSE of PS-Sparse ($0.056 \pm 3.2 \times 10^{-3}$), SToRM ($0.0521 \pm 2.6 \times 10^{-2}$), KLR ($0.0753 \pm 3.6 \times 10^{-2}$), BiLMDM ($0.0461 \pm 7.6 \times 10^{-4}$), KiBLMDM ($0.0395 \pm 1.6 \times 10^{-3}$). The previous numerical values indicate mean and standard deviation of the NRMSE across 360 frames.

constraint on $\mathbf{B}$ as in As. 2. Term C1 prevents the unbounded solution of $\mathbf{D}$ due to the scaling ambiguity in $\mathbf{D}\check{\mathbf{K}}\mathbf{B}$.

To solve (1), the successive-convex-approximation framework of [28] is employed and is presented in Alg. 1. The study [28] is employed for its ability to decompose task (1) into a series of smaller subtasks which can be parallelised for faster implementation. Convergence to a stationary solution of (1) is guaranteed [28]. Step 8 of Alg. 1 calls for solving convex minimization sub-tasks (2a) and (2b). Given $\tau_U, \tau_B > 0$, for $(\mathbf{D}_n, \mathbf{B}_n, \mathbf{Z}_n)$ at every iteration of the algorithm, the following estimates are required:

$$\hat{\mathbf{D}}_n \in \arg\min_{\mathbf{D}} \tfrac{1}{2}\left\|\mathcal{S}(\mathbf{Y}) - \mathcal{SF}(\mathbf{D}\check{\mathbf{K}}\mathbf{B}_n)\right\|_{\mathrm{F}}^2 + \tfrac{\tau_D}{2}\|\mathbf{D} - \mathbf{D}_n\|_{\mathrm{F}}^2 + \tfrac{\lambda_1}{2}\left\|\mathbf{Z}_n - \mathcal{F}_t(\mathbf{D}\check{\mathbf{K}}\mathbf{B}_n)\right\|_{\mathrm{F}}^2 \quad \text{s.to} \quad \|\mathbf{D}\mathbf{e}_i\| \leq C_D,\ \forall i \in \{1, \ldots, d\}\,. \tag{2a}$$

$$\hat{\mathbf{B}}_n \in \arg\min_{\mathbf{B}} \tfrac{1}{2}\left\|\mathcal{S}(\mathbf{Y}) - \mathcal{SF}(\mathbf{D}_n\check{\mathbf{K}}\mathbf{B})\right\|_{\mathrm{F}}^2 + \tfrac{\tau_B}{2}\|\mathbf{B} - \mathbf{B}_n\|_{\mathrm{F}}^2 + \tfrac{\lambda_1}{2}\left\|\mathbf{Z}_n - \mathcal{F}_t(\mathbf{D}_n\check{\mathbf{K}}\mathbf{B})\right\|_{\mathrm{F}}^2 + \lambda_3\|\mathbf{B}\|_1 \quad \text{s.to} \quad \mathbf{1}_{N_\ell}^{\mathsf{T}}\mathbf{B} = \mathbf{1}_{N_{\mathrm{fr}}}^{\mathsf{T}}\,. \tag{2b}$$

Tasks in (2) can be viewed as affinely constrained composite convex minimization tasks, hence allowing the use of [27]. Alg. (1) requires tuning hyperparameters, chosen empirically to minimize NRMSE values for datasets where the ground truth is known. The effect of these hyperparameters is explored and a range of values is reached. The algorithmic and mathematical details of (2) and details on hyperparameter values are reserved for an upcoming journal publication.

V. Numerical Results

The proposed framework is validated on the magnetic-resonance extended cardiac-torso (MRXCAT) cine phantom [29]. Further investigation via testing on other datasets is beyond the scope of this paper due to lack of space. The extended cardiac torso (XCAT) framework was used to generate a breath-hold cardiac cine data of spatial size $(N_{\mathrm{p}}, N_{\mathrm{f}}) = (408, 408)$ with $N_{\mathrm{fr}} = 360$ number of frames. The generated data has a spatial resolution of 1.56×1.56 mm^2 for a field-of-view (FOV) of 400×400 mm^2. The phantom data consists of 15 cardiac cycles and 24 cardiac phases. The data is retrospectively undersampled under 1D Cartesian sampling strategy, as in Fig. 1b, to emulate an undersampling rate [defined by $N_{\mathrm{k}}N_{\mathrm{fr}}/(\#\text{ of acquired voxels})$] of 8. Normalised root mean square error (NRMSE), defined as $\|\mathbf{X}-\hat{\mathbf{X}}\|_{\mathrm{F}}/\|\mathbf{X}\|_{\mathrm{F}}$, is used to measure the quality of reconstruction, with $\mathbf{X}$ being the fully sampled original MR image and $\hat{\mathbf{X}}$ is the reconstructed image from various schemes.

The performance of KBiLMDM is tested against various state-of-the-art techniques, namely, the partially separable sparsity aware model (PS-Sparse) [9], smoothness regularization on manifolds (SToRM) [13], kernel low rank (KLR) [17], and the bi-linear modeling of data manifolds (BiLMDM) [19]. KBiLMDM achieves the least NRMSE value of 0.03891 as compared to PS-Sparse's 0.0556, SToRM's 0.0503, KLR's 0.0744, BiLMDM's 0.0456. KBiLMDM shows less deviation in NRMSE than KLR, SToRM and PS-Sparse while consistently scoring the least NRMSE as compared to BiLMDM, KLR and PS-Sparse across all frames. A further qualitative analysis of the reconstructions, as in Fig. 3 provides evidence that KBiLMDM outperforms the other aforementioned state-of-the-art techniques. KLR and SToRM exhibit severe aliasing effects as marked in Fig. 3, and the error maps of PS-Sparse provides evidence that reconstructions along the edges are not upto the mark, while there are discrepancies in the contrast as compared to the ground truth. On the other hand, KBiLMDM produces sharper, aliasing-free, distortion-free dMRI time series. A comparison of the error maps between BiLMDM and KBiLMDM shows improvements in the reconstructions due to the introduction of kernels. Details on choice of kernels and comparison with schemes like FRIST-MRI [10] are left for presentation and upcoming journal publications.

VI. Conclusion & the Road Ahead

A kernel-based framework for reconstructing data on manifolds was proposed and tailored to fit the dMRI-data recovery problem. The proposed methodology exploited simple tangent-space geometries of manifolds and followed classical kernel-approximation arguments to form a bi-linear inverse estimation problem. Departing from state-of-the-art approaches, **i**) no training data were used; **ii**) no graph Laplacian matrix was employed to penalize the inverse problem; **iii**) no costly (kernel) preimaging step was necessary to map feature points back to the input space; and **iv**) complex-valued kernel functions were defined to account for k-space data. Preliminary numerical tests and qualitative investigation of reconstructed images on retrospectively undersampled synthetic MR images showed the rich potential of the proposed framework against several state-of-the-art techniques. With regards to the road ahead, the incorporation of training data into the framework is currently under investigation. On-going research includes also modeling extensions, further numerical tests on synthetic and real dMRI data, which will be reported during the conference.

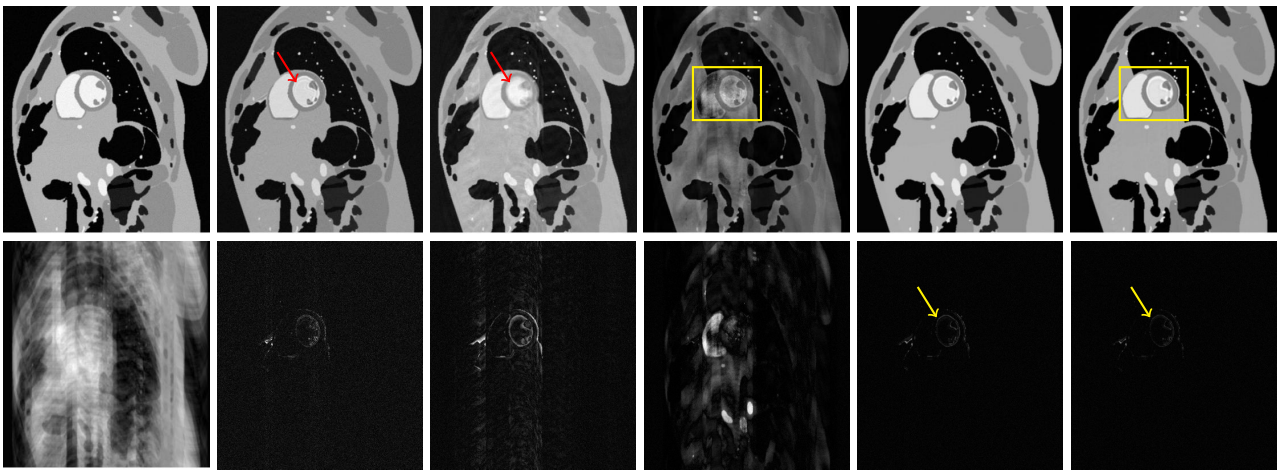


Fig. 3. Spatial Results for frame 49 of MRXCAT cardiac cine phantom retrospectively undersampled at acceleration rate of 8. Left to Right: Gold standard (top row); Undersampled Image (bottom row), PS-Sparse (0.0556), SToRM (0.0503), KLR (0.0744), BiLMDM (0.0456), KBiLMDM (0.0389). The numerical values indicate the NRMSE. Top row: Spatial frames generated from aforementioned reconstructed schemes; bottom row: Corressponding error maps.

REFERENCES

[1] Z.-P. Liang and P. C. Lauterbur, *Principles of Magnetic Resonance Imaging: A Signal Processing Perspective*. IEEE Press, 2000.

[2] ——, "An efficient method for dynamic magnetic resonance imaging," *IEEE Trans. Medical Imag.*, vol. 13, no. 4, pp. 677–686, 1994.

[3] D. Liang, J. Cheng, Z. Ke, and L. Ying, "Deep magnetic resonance image reconstruction: Inverse problems meet neural networks," *IEEE Signal Processing Magazine*, vol. 37, no. 1, pp. 141–151, 2020.

[4] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, "A deep cascade of convolutional neural networks for dynamic MR image reconstruction," *IEEE Trans. Medical Imag.*, vol. 37, no. 2, pp. 491–503, 2018.

[5] M. Mardani, E. Gong, J. Y. Cheng, S. S. Vasanawala, G. Zaharchuk, L. Xing, and J. M. Pauly, "Deep generative adversarial neural networks for compressive sensing MRI," *IEEE Trans. Medical Imag.*, vol. 38, no. 1, pp. 167–179, 2018.

[6] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, "Learning a variational network for reconstruction of accelerated MRI data," *Magnetic Resonance in Medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.

[7] M. Lustig, J. M. Santos, D. L. Donoho, and J. M. Pauly, "k-t SPARSE: High frame rate dynamic MRI exploiting spatio-temporal sparsity," in *Proc. ISMRM*, vol. 2420, 2006.

[8] H. Jung, J. C. Ye, and E. Y. Kim, "Improved k-t BLAST and k-t SENSE using FOCUSS," *Physics in Medicine and Biology*, vol. 52, no. 11, pp. 3201–3226, 2007.

[9] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled (k,t)-space data with joint partial separability and sparsity constraints," *IEEE Trans. Medical Imag.*, vol. 31, no. 9, pp. 1809–1820, 2012.

[10] B. Wen, S. Ravishankar, and Y. Bresler, "FRIST—Flipping and rotation invariant sparsifying transform learning and applications," *Inverse Problems*, vol. 33, no. 7, p. 074007, 2017.

[11] J. Jiang, R. Hu, Z. Han, Z. Wang, T. Lu, and J. Chen, "Locality-constraint iterative neighbor embedding for face hallucination," in *2013 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2013, pp. 1–6.

[12] P. Zhang, X. You, W. Ou, C. P. Chen, and Y.-m. Cheung, "Sparse discriminative multi-manifold embedding for one-sample face identification," *Pattern Recognition*, vol. 52, pp. 249–259, 2016.

[13] S. Poddar and M. Jacob, "Dynamic MRI using smoothness regularization on manifolds (SToRM)," *IEEE Trans. Medical Imag.*, vol. 35, no. 4, pp. 1106–1115, 2016.

[14] A. V. Little, M. Maggioni, and L. Rosasco, "Multiscale geometric methods for data sets I: Multiscale SVD, noise and curvature," *Applied and Computational Harmonic Analysis*, vol. 43, no. 3, pp. 504–567, 2017.

[15] B. Scholkopf and A. J. Smola, *Learning With Kernels: Support Vector Machines, Regularization, Optimization, and Beyond*. MIT press, 2001.

[16] S. Poddar and M. Jacob, "Recovery of noisy points on bandlimited surfaces: Kernel methods re-explained," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 4024–4028.

[17] U. Nakarmi, Y. Wang, J. Lyu, D. Liang, and L. Ying, "A kernel-based low-rank (KLR) model for low-dimensional manifold recovery in highly accelerated dynamic MRI," *IEEE Trans. Medical Imag.*, vol. 36, no. 11, pp. 2297–2307, 2017.

[18] O. Arif, H. Afzal, H. Abbas, M. F. Amjad, J. Wan, and R. Nawaz, "Accelerated dynamic MRI using kernel-based low rank constraint," *Journal of Medical Systems*, vol. 43, no. 8, p. 271, 2019.

[19] G. N. Shetty, K. Slavakis, A. Bose, U. Nakarmi, G. Scutari, and L. Ying, "Bi-linear modeling of data manifolds for dynamic-MRI recovery," *IEEE Trans. Medical Imag.*, 2019. [Online]. Available: https://doi.org/10.1109/TMI.2019.2934125

[20] M. P. Wand and M. C. Jones, *Kernel Smoothing*. Chapman and Hall/CRC, 1994.

[21] U. Nakarmi, K. Slavakis, and L. Ying, "MLS: Joint manifold-learning and sparsity-aware framework for highly accelerated dynamic magnetic resonance imaging," in *Proc. ISBI*, 2018, pp. 1213–1216.

[22] R. Boloix-Tortosa, F. J. Payán-Somet, E. Arias-de Reyna, and J. J. Murillo-Fuentes, "Complex kernels for proper complex-valued signals: A review," in *2015 23rd European Signal Processing Conference (EUSIPCO)*. EURASIP, 2015, pp. 2371–2375.

[23] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *J. Machine Learning Research*, vol. 4, pp. 119–155, 2003.

[24] L. W. Tu, *An Introduction to Manifolds*. New York: Springer, 2008.

[25] V. De Silva and J. B. Tenenbaum, "Sparse multidimensional scaling using landmark points," Stanford University, Tech. Rep., 2004.

[26] K. Slavakis, G. B. Giannakis, and G. Leus, "Robust sparse embedding and reconstruction via dictionary learning," in *Proc. CISS*, Baltimore: USA, Mar. 2013.

[27] K. Slavakis and I. Yamada, "Fejér-monotone hybrid steepest descent method for affinely constrained and composite convex minimization tasks," *Optimization*, vol. 67, no. 11, pp. 1963–2001, 2018.

[28] F. Facchinei, G. Scutari, and S. Sagratella, "Parallel selective algorithms for nonconvex big data optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1874–1889, 2015.

[29] L. Wissmann, C. Santelli, W. P. Segars, and S. Kozerke, "MRXCAT: Realistic numerical phantoms for cardiovascular magnetic resonance," *J. Cardiovascular Magnetic Resonance*, vol. 16, no. 1, p. 63, 2014.