HHS Public Access

Author manuscript

Int Workshop Comput Adv Multisens Adapt Process. Author manuscript; available in PMC 2019 November 22.

Published in final edited form as:

Int Workshop Comput Adv Multisens Adapt Process. 2017 December ; 2017: . doi:10.1109/CAMSAP.2017.8313115.

Bi-Linear Modeling of Manifold-Data Geometry for Dynamic-MRI Recovery

Konstantinos Slavakis, **Gaurav N. Shetty**, **Abhishek Bose**, **Ukash Nakarmi**, **Leslie Ying**

Dept. of Electrical Engineering, University at Buffalo, The State University of New York, Buffalo, NY 14226-2500, USA

Konstantinos Slavakis: kslavaki@buffalo.edu; Gaurav N. Shetty: gauravna@buffalo.edu; Abhishek Bose: abose3@buffalo.edu; Ukash Nakarmi: ukashnak@buffalo.edu; Leslie Ying: leiying@buffalo.edu

Abstract

This paper establishes a modeling framework for data located onto or close to (unknown) smooth manifolds, embedded in Euclidean spaces, and considers its application to dynamic magnetic resonance imaging (dMRI). The framework comprises several modules: First, a set of landmark points is identified to describe concisely a data cloud formed by highly under-sampled dMRI data, and second, low-dimensional renditions of the landmark points are computed. Searching for the linear operator that decompresses low-dimensional data to high-dimensional ones, and for those combinations of landmark points which approximate the manifold data by affine patches, leads to a bi-linear model of the dMRI data, cognizant of the intrinsic data geometry. Preliminary numerical tests on synthetically generated dMRI phantoms, and comparisons with state-of-the-art reconstruction techniques, underline the rich potential of the proposed method for the recovery of highly under-sampled dMRI data.

I. INTRODUCTION

Current medical research and diagnosis rely heavily on *magnetic resonance imaging* (MRI); a non-invasive and non-ionizing technology for high fidelity visualization of anatomical structures and physiological functions [9]. Collecting a sufficiently large number of MRI data to guarantee high-quality image reconstruction is an inherently slow process due to patient discomfort as well as to physical and physiological constraints imposed on the scanning speed [9]. Such obstacles arise prominently in *dynamic* (d)MRI, where data acquisition needs to abide also by the inescapable constraints imposed by the monitored dynamical process, *e.g.*, a beating heart [8].

Under-sampling the dMRI data is an efficient way to speed up scanning times, at the price of aliasing effects in the reconstructed images. To surmount aliasing given the limited number of data, spatio-temporal correlations and any available prior knowledge regarding the dMRI mechanism have to be taken into account. Along these lines, compressed sensing [6], [11], [15] and the principal-component-analysis (PCA) or low-rank-approximation paradigm [5], [7], [10], [12], [26] are modeling frameworks that have shown a rich potential in incorporating successfully prior information and data dependencies into dMRI-recovery algorithms. For example, [26] extracts a low-dimensional subspace, via the singular-value-decomposition of a highly under-sampled data matrix, prior to forming a convex inverse

problem whose solution yields high-quality reconstructed images. Elaborate dictionary learning (DL) techniques have also contributed largely to the dMRI literature [1], [14], [23], [24], [27]. Remarkably, in all of the previous DL techniques, etre is no utilization of the (presumably low-dimensional) data geometry, other than the popular algebraic tool of matrix factorization.

Driven by the recently successful machine-learning paradigm, manifold-learning techniques have been also implemented in dMRI recovery [13], [16], [22], [28]. Local linear embedding (LLE) [18] is used in [22] to achieve dimensionality reduction of the high-dimensional dMRI data, prior to the application of a reconstruction algorithm. Study [16] formulates a convex minimization recovery task which is penalized by a Laplacian-matrix-based quadratic term that quantifies all the available knowledge about the underlying smooth data manifold. Further, to exploit any potential non-linear dependencies between data, [28] maps the observed data to even higher dimensional functional spaces, via kernel mappings, and solves an inverse problem to recover the dependencies in the original input space. Very recently, a joint manifold-learning and sparsity-cognizant framework has been introduced in [13]. Prior to forming a convex minimization recovery task, the intrinsic low-dimensionality geometry of the data is learned from their highly under-sampled renditions via an LLE-motivated framework [19]. Extensive experimentation has showed that learning, first, any (potentially non-linear) data geometry helps the solutions of convex inverse problems to produce higher quality reconstruction images than state-of-the-art PCA-based techniques [13].

The present study takes [13] a step forward into explicitly modeling the low-dimensional dMRI data geometry. Built again on an unknown smooth-manifold hypothesis for the data geometry, a small set of data representatives, called *landmark* points, are extracted from the under-sampled dMRI data cloud. These landmark points provide a concise description of the observed data, promoting low-dimensional and parsimonious data representations as well as efficient data storage. Motivated by the concept of tangent spaces of smooth manifolds, data are approximated via affine combinations of the extracted landmark points. Similarly to [13] and along the lines of [19], a dimensionality-reduction module is applied to compute the low-dimensional rendition, or, “compressed” versions of the landmark points. Having the previous affine combinations, as well as a linear *decompression* operator, as unknowns of the data-modeling hypothesis, a *bi-linear* model is derived [*cf.* (2)], cognizant of the underlying low-dimensional nature of the observed dMRI data. Exploiting also the fact that dMRI data usually capture periodic time series, *e.g.*, a beating heart, a non-convex, bi-linear recovery task, penalized by terms which account for the periodicity of the recorded time series, is formulated to achieve high-fidelity data reconstruction. Subsequently, recently developed convex and non-convex minimization techniques are employed to solve the resultant recovery tasks. Preliminary numerical tests on synthetically generated high-dimensional dMRI phantom data and comparisons against state-of-the-art methodologies underline the rich potential of the method for high-quality data recovery.

II. PROPOSED FRAMEWORK

MRI data are observed at the discrete *k-space,* or frequency domain, which admits complex-valued data and spans an area of size $N_p \times N_f$, where N_p stands for the number of phase-encoding lines and N_f for the number of frequency-encoding ones [9]. In dMRI, an additional dimension is added to the MRI k-space to account for the time horizon (the axis vertical on the paper in Fig. 1a), resulting in the augmented (k,t)-space. In other words, the dMRI (k,t)-space can be viewed as the N_{fr}-fold Cartesian product of the ($N_p \times N_f$)-sized MRI k-space, where N_{fr} represents the number of collected MRI frames over time. K-space data $\mathcal{Y}_j \in \mathbb{C}^{N_p \times N_f}$ ($\mathbb{C}$ is the set of all complex-valued numbers), $j \in \{1, \ldots, N_{fr}\}$, can be also viewed as the two-dimensional (discrete) Fourier transform of the $N_p \times N_f$ image-domain data $\mathcal{X}_j$, *i.e.*, $\mathcal{Y}_j = \mathcal{F}(\mathcal{X}_j)$ Without any loss of generality, this study assumes that the "low-frequency" part of $\mathcal{Y}_j$ is located around the center of the $N_p \times N_f$ area. Availability of a large number of dMRI data is infeasible in practice; the (k,t)-space of Fig. 1a is usually severely under-sampled [8]. To extract useful information from (k,t)-space data, the present framework assumes *full* availability of a part of the low-frequency region of the (k,t)-space. For simplicity and illustration reasons, it is assumed that a small number $\nu (\ll N_p)$ of phase-encoding lines of (coined "navigator") data are always available (the gray-colored area in Fig. 1a). These navigator data will be exploited to learn the intrinsic low-dimensional structure of the dMRI data. Other than the navigator data, highly (pseudo-randomly) under-sampled (k,t)-space data are considered to be also available.

To facilitate processing and data representations, the (k,t)-space data are vectorized. More specifically, the $\mathrm{vec}(\mathcal{Y}_j)$ operation stacks one column of $\mathcal{Y}_j$ below the other to yield the complex-valued $N_p N_f \times 1$ vector $\mathrm{y}_j := \mathrm{vec}(\mathcal{Y}_j)$, $\forall j \in \{1, \ldots, N_{fr}\}$. To avoid notation clutter, $\mathcal{F}$ stands also for the two-dimensional (discrete) Fourier transform even when applied to vectorized versions of image frames: $\mathcal{F}[\mathrm{vec}(\mathcal{X}_j)] := \mathrm{vec}[\mathcal{F}(\mathcal{X}_j)] = \mathrm{vec}(\mathcal{Y}_j)$. All vectorized k-space frames are gathered in the $N_p N_f \times N_{fr}$ matrix $Y := [\mathrm{y}_1, \mathrm{y}_2, \ldots, \mathrm{y}_{N_{fr}}]$, so that the vectorized original image-domain data are $\mathbf{X} := \mathcal{F}^{-1}(\mathbf{Y})$. In a similar way, per jth k-space frame, the "navigator" data yield a $\nu N_f \times 1$ vector $\mathrm{y}_j^{\mathrm{nav}}$ and thus the $\nu N_f \times N_{fr}$ matrix $\mathbf{Y}_{\mathrm{nav}} := [\mathrm{y}_1^{\mathrm{nav}}, \mathrm{y}_2^{\mathrm{nav}}, \ldots, \mathrm{y}_{N_{fr}}^{\mathrm{nav}}]$. Clearly, $\mathbf{Y}_{\mathrm{nav}} = \mathbf{\Omega Y}$, where $\mathbf{\Omega}$ is a submatrix of the identity matrix $\mathbf{I}_{N_p N_f}$ that selects the rows of $\mathbf{Y}$ which correspond to the navigator data. Recovering $\mathbf{Y}$ from its partial $\mathbf{Y}_{\mathrm{nav}}$ is viable only under errors, *e.g.*, an approximation can be obtained via $\mathbf{\Omega}^\dagger \mathbf{Y}_{\mathrm{nav}}$, where $\mathbf{\Omega}^\dagger$ stands for the Moore-Penrose pseudoinverse of $\mathbf{\Omega}$. The following modeling hypothesis generalizes the previous argument.

Assumption 1. There exist an $N_p N_f \times \nu N_f$ matrix $\mathbf{G}_1$ and an $N_p N_f \times N_{fr}$ matrix $\mathbf{E}_1$, which gathers all approximation errors, such that (s.t.) $\mathbf{Y} = \mathbf{G}_1 \mathbf{Y}_{\mathrm{nav}} + \mathbf{E}_1$.

A. Landmark points

The high-dimensional navigator data $\left\{\mathrm{y}_j^{\mathrm{nav}}\right\}_{j=1}^{N_{\mathrm{fr}}}$ carry useful information about spatio-temporal dependencies in the (k,t)-space. To promote parsimonious data representations, especially in cases where N_{fr} attains large values, it is desirable to extract a subset $\left\{\ell_k\right\}_{k=1}^{N_1} \subset \left\{\mathbf{y}_j^{\mathrm{nav}}\right\}_{j=1}^{N_{\mathrm{fr}}} (N_1 \leq N_{\mathrm{fr}})$, called *landmark* points, which provide a "concise description," in a user-defined sense, of the data cloud $\left\{\mathrm{y}_j^{\mathrm{nav}}\right\}_{j=1}^{N_{\mathrm{fr}}}$. To this end, the following modeling assumption imposes structure on $\left\{\mathrm{y}_j^{\mathrm{nav}}\right\}_{j=1}^{N_{\mathrm{fr}}}$.

Assumption 2. Data $\left\{\mathrm{y}_j^{\mathrm{nav}}\right\}_{j=1}^{N_{\mathrm{fr}}}$ lie on a smooth low-dimensional manifold $\mathscr{M}$ [21] embedded in the high-dimensional Euclidean space $\mathbb{C}^{\nu N_{\mathrm{f}}}$ (*cf.* Fig. 1b).

For example, the most commonly met case of a smooth manifold in theory and practice is a linear subspace. Based on As. 2 and the concept of the tangent space of a smooth manifold, it is conceivable that neighboring landmark points cooperate affinely to describe vector $\mathbf{y}_j^{\mathrm{nav}}$ (the gray-colored area in Fig. 1b depicts *all* possible affine combinations of $\left\{\ell_{k_1}, \ell_{k_2}, \ell_{k_3}\right\}$). Upon defining the $\nu N_{\mathrm{f}} \times N_{\mathrm{l}}$ matrix $\mathbf{\Lambda} := \left[\ell_1, \ell_2, \ldots, \ell_{N_1}\right]$, it is assumed that there exists an $N_{\mathrm{l}} \times 1$ vector $\mathbf{b}_j$ that renders the approximation error $\left\|\mathrm{y}_j^{\mathrm{nav}} - \Lambda \mathrm{b}_j\right\|$ small, where $\left\|\cdot\right\|$ denotes the standard Euclidean norm of space $\mathbb{C}^{\nu N_{\mathrm{f}}}$. Since affine combinations are desirable, $\mathbf{b}_j$ is constrained to satisfy $1_{N_1}^{\top} \mathrm{b}_j = 1$, where 1_{N_1} stands for the all-one $N_{\mathrm{l}} \times 1$ vector and superscipt $\top$ denotes vector/matrix transposition. Moreover, motivated by the low-dimensional nature of $\mathscr{M}$, according to As. 2, it is envisioned that only a few landmark points cooperate into representing $\mathrm{y}_j^{\mathrm{nav}}$, *i.e.*, $\mathbf{b}_j$ is sparse. The previous arguments are summarized into the following modeling hypothesis.

Assumption 3. There exist a sparse $N_{\mathrm{l}} \times N_{\mathrm{fr}}$ matrix $\mathbf{B}$, with $\mathbf{1}_{N_1}^{\top} \mathbf{B} = \mathbf{1}_{N_{\mathrm{fr}}}^{\top}$, and a $\nu N_{\mathrm{f}} \times N_{\mathrm{fr}}$ matrix $\mathbf{E}_2$, which gathers approximation errors, s.t. $\mathbf{Y}_{\mathrm{nav}} = \mathbf{\Lambda B} + \mathbf{E}_2$.

The previous modeling assumption holds true in the prototypical case where $\mathscr{M}$ is a linear subspace. In such a case, any linearly independent subset of $\left\{y_j^{\mathrm{nav}}\right\}_{j=1}^{N_{\mathrm{fr}}}$, which spans the column (range) space of $\mathbf{Y}_{\mathrm{nav}}$, can be selected as the columns of $\mathbf{\Lambda}$. Notice also that under

such a choice for $\mathbf{\Lambda}$, the coefficient matrix $\mathbf{B}$ satisfies $1_{N_1}^\top \mathbf{B} = 1_{N_{\text{fr}}}^\top$, since the column (range) space of $\mathbf{Y}_{\text{nav}}$, being a linear subspace, is also the affine hull [17] of the columns of $\mathbf{\Lambda}$.

Although several strategies may be implemented to identify the landmark points $\mathbf{\Lambda}$ of a data cloud that lies onto or close to a general manifold $\mathscr{M}$, a greedy optimization methodology, introduced in [2], is adopted here. In short, at every step of the algorithm, a landmark point is selected from $\left\{\mathbf{y}_j^{\text{nav}}\right\}_{j=1}^{N_{\text{fr}}}$ by maximizing, over all *un-selected* $\left\{\mathbf{y}_j^{\text{nav}}\right\}_{j=1}^{N_{\text{fr}}}$, the minimum distance to the landmark points which have been already selected up to the previous step of the algorithm. The algorithm of [2] scores a computational complexity of order $\mathcal{O}(N_1 N_{\text{fr}})$, which is naturally heavier than that of the naive random-selection algorithm that chooses $\{\ell_k\}_{k=1}^{N_1}$ randomly from $\left\{\mathbf{y}_j^{\text{nav}}\right\}_{j=1}^{N_{\text{fr}}}$.

B. Reducing the dimension of the landmark points

The landmark points Λ, obtained in the previous section, are still high-dimensional. To meet restrictions imposed by finite computational resources, it is desirable to reduce the dimensionality of $\mathbf{\Lambda}$. To this end, the methodology of [19], which is motivated by [3], [18], is employed. The approach comprises two steps.

1) Given $\mathbf{\Lambda}$ and a user-defined $\lambda_W > 0$ solve the convex minimization task

$$\begin{aligned} &\min_{\mathbf{W} \in \mathbb{C}^{N_1 \times N_1}} \left\|\mathbf{\Lambda} - \mathbf{\Lambda} W\right\|_{\text{F}}^2 + \lambda_W \left\|\mathbf{W}\right\|_1 \\ &\text{s.to } \mathbf{1}_{N_1}^\top \mathbf{W} = \mathbf{1}_{N_1}^\top \text{ and } \text{diag}(\mathbf{W}) = 0, \end{aligned} \tag{1}$$

where $\left\|\cdot\right\|_{\text{F}}$ stands for the Frobenius norm of a matrix. A few comments on the choice of the loss function and constraints follow. Since $\{\ell_k\}_{k=1}^{N_1}$ lie on the manifold $\mathscr{M}$, then according to As. 2, Fig. 1b and the discussion regarding $\mathbf{y}_j^{\text{nav}}$, any point taken from $\{\ell_k\}_{k=1}^{N_1}$ may be faithfully approximated by an affine combination of the rest of the landmark points. In other words, there exists a matrix $\mathbf{W}$ s.t. $\mathbf{\Lambda} \approx \mathbf{\Lambda W}$. With $\mathbf{1}_{N_1}^\top \mathbf{W} = \mathbf{1}_{N_1}^\top$ manifesting the previous desire for affine combinations, the constraint $\text{diag}(\mathbf{W}) = \mathbf{0}$ is used to exclude the trivial solution of the identity matrix $\mathbf{I}_{N_1}$ for $\mathbf{W}$. Task (1) is an affinely constrained composite convex minimization task, and, hence, the framework of [20] can be employed to solve it, due to the flexibility by which [20] deals with affine constraints when compared with state-of-the-art convex optimization techniques.

2) Once $\mathbf{W}$ has been obtained from the previous step and for a user-defined integer number $d \ll N_{\mathrm{p}} N_{\mathrm{f}}$, solve

$$\min_{\breve{\mathbf{\Lambda}} \in \mathbb{C}^{d \times N_1}} \left\| \breve{\mathbf{\Lambda}} - \breve{\mathbf{\Lambda}} \mathbf{W} \right\|_{\mathrm{F}}^2 \text{s.to} \breve{\mathbf{\Lambda}} \breve{\mathbf{\Lambda}}^* = \mathbf{I}_d,$$

where the constraint $\breve{\mathbf{\Lambda}} \breve{\mathbf{\Lambda}}^* = \mathbf{I}_d$ is used to exclude the trivial solution of $\breve{\mathbf{\Lambda}} = 0$. It is not difficult to verify that the solution of the previous task is nothing but the complex conjugate transpose of the matrix (denoted by the * superscipt) which comprises the d minimal eigenvectors of $\left(\mathbf{I}_{N_1} - \mathbf{W}\right)\left(\mathbf{I}_{N_1} - \mathbf{W}\right)^*$.

The following hypothesis establishes a linear relation between $\mathbf{\Lambda}$ and its low-dimensional rendition $\breve{\mathbf{\Lambda}}$.

Assumption 4. There exist an $N_{\mathrm{p}} N_{\mathrm{f}} \times d$ matrix $\mathbf{G}_3$ and an $N_{\mathrm{p}} N_{\mathrm{f}} \times N_1$ matrix $\mathbf{E}_3$, which gathers all approximation errors, so that $\mathbf{\Lambda} = \mathbf{G}_3 \breve{\mathbf{\Lambda}} + \mathbf{E}_3$.

Matrix $\mathbf{G}_3$ can be viewed as the "decompression" operator which reconstructs the "full" $\mathbf{\Lambda}$ from its low-dimensional representation $\breve{\mathbf{\Lambda}}$.

C. Data recovery task

Putting modeling assumptions 1, 3 and 4 together, it can be verified that there exist matrices $\mathbf{G}$ and $\mathbf{E}$ s.t. $\mathbf{Y} = \mathbf{G}\breve{\mathbf{\Lambda}}\mathbf{B} + \mathbf{E}$. Upon defining $\mathbf{U} := \mathscr{F}^{-1}(\mathbf{G})$, and since $\mathbf{G}\breve{\mathbf{\Lambda}}\mathbf{B} = \mathscr{F}(\mathbf{U})\breve{\mathbf{\Lambda}}\mathbf{B} = \mathscr{F}(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B})$ due to the definition of $\mathscr{F}$ (*cf.* Sec. II), the following *bi-linear* model between $\mathbf{Y}$ and the unknowns (U, B) is established:

$$\mathbf{Y} = \mathscr{F}(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B}) + \mathbf{E}. \tag{2}$$

Bi-linearity means that whenever one of the block of variables (**U**, **B**) is fixed to a specific value, then the dependence between $\mathbf{Y}$ and the other block of variables is linear (modulo the error $\mathbf{E}$ term). Interestingly, the linearity of the inverse Fourier transform $\mathscr{F}^{-1}$ suggests that the previous modeling hypothesis holds true also in the image domain: $\mathscr{F}^{-1}(\mathbf{Y}) = \mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B} + \mathscr{F}^{-1}(\mathbf{E})$.

In practice, only few (k,t)-space data are known. To explicitly take account of the limited number of data, a (linear) sampling operator $\mathscr{S}(\cdot)$ is introduced where $\mathscr{S}(\mathbf{Y})$ keeps only those entries of $\mathbf{Y}$ that are collected by the measurement system. It is also often in dMRI that image frames capture a periodic process, *e.g.*, heart movement, other than the static background. In other words, it is reasonable to assume that the one-dimensional Fourier transform $\mathscr{F}_t$ of the time profile of every MR pixel, *i.e.*, every row of the matrix $\mathscr{F}_t(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B})$, is a sparse vector. Since the DC Fourier coefficients do not provide any useful information on periodic time series, the linear operator $\mathscr{P}(\cdot)$ is also applied to $\mathscr{P}[\mathscr{F}_t(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B})]$ to remove the

first column of $\mathscr{F}_t(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B})$, which gathers the DC Fourier coefficients of all the MR-pixel time profiles.

Algorithm 1 Solving the bi-linear recovery task of(3)

Input: Available are data $\mathcal{S}(\mathbf{Y})$. These include the navigator $\mathbf{Y}_{\text{nav}}$ ones. Choose parameters $\lambda_1, \lambda_2, \lambda_3, C_U, \tau_U, \tau_B > 0$, as well as $\varsigma \in (0, 1)$ and $\gamma_0 \in (0, 1]$.

1: Compute landmark points $\mathbf{\Lambda}$ from the columns of $\mathbf{Y}_{\text{nav}}$ according to Sec. II-A.

2: Compute the low-dimensional rendition $\breve{\mathbf{\Lambda}}$ of $\mathbf{\Lambda}$ according to Sec. II-B.

3: Arbitrarily fix $(\mathbf{U}_0, \mathbf{B}_0, \mathbf{Z}_0)$ and set $n = 0$.

4: **while** $n \geq 0$ **do**

5: Available are $(\mathbf{U}_n, \mathbf{B}_n, \mathbf{Z}_n)$ and γ_n.

6: Let $\gamma_{n+1} := \gamma_n(1 - \zeta\gamma_n)$.

7: Obtain $\widehat{\mathbf{U}}_n$ via (4), $\widehat{\mathbf{B}}_n$ via (5), and the (i, j)th entry of $\widehat{\mathbf{Z}}_n$, $\forall(i, j)$, via the following soft-thresholding rule:

$$[\widehat{\mathbf{Z}}_n]_{ij} := \left[\mathscr{P}\mathscr{F}_t(\mathbf{U}_n\breve{\mathbf{\Lambda}}\mathbf{B}_n\right]_{ij} \cdot \left(1 - \frac{\lambda_2/\lambda_1}{\max\left\{\lambda_2/\lambda_1, \left|\left[\mathscr{P}\mathscr{F}_t(\mathbf{U}_n\breve{\mathbf{\Lambda}}\mathbf{B}_n\right]_{ij}\right|\right\}}\right).$$

8: Update

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

9: Set n equal to $n + 1$ and go to step 5.

10: **end while**

Output: Extract the limit points $\mathbf{U}_*$ and $\mathbf{B}_*$ of sequences $(\mathbf{U}_n)_n$ and $(\mathbf{B}_n)_n$, respectively, and recover the dMRI data by the estimate $\widehat{\mathbf{X}} := \mathbf{U}_*\breve{\mathbf{\Lambda}}\mathbf{B}_*$.

To summarize, given parameters $\lambda_1, \lambda_2, \lambda_3, C_U > 0$, the bi-linear data-recovery task is formulated as

$$\begin{aligned} &\min_{(\mathbf{U},\mathbf{B},\mathbf{Z})} \frac{1}{2}\left\|\mathcal{S}(\mathbf{Y}) - \mathcal{S}\mathscr{F}(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B})\right\|_{\text{F}}^2 + \frac{\lambda_1}{2}\left\|\mathbf{Z} - \mathscr{P}\mathscr{F}_t(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B})\right\|_{\text{F}}^2 + \lambda_2\|\mathbf{Z}\|_1 + \lambda_3\|\mathbf{B}\|_1 \\ &\text{s.to}\left\|\mathbf{U}\mathbf{e}_i\right\| \leq C_U, \forall i \in \{1, \ldots, d\} \text{ and } \mathbf{1}_{N_1}^\top\mathbf{B} = \mathbf{1}_{N_{\text{fr}}}^\top, \end{aligned} \tag{3}$$

where $\mathbf{e}_i$ denotes the ith column of the identity matrix $\mathbf{I}_d$ and the auxiliary variable $\mathbf{Z}$ is used to incorporate the sparsity of $\mathscr{P}[\mathscr{F}_t(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B})]$ into the design. Notice that the C_U bound is used to prevent unbounded solutions for $\mathbf{U}$ due to the bi-linearity in $\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B}$. To solve the non-convex task (3), the successive-convex-approximation framework of [4] is employed and presented in a concise form in Alg. 1.

Step 7 of Alg. 1 comprises convex minimization sub-tasks. More specifically, at every step of the algorithm, given $(\mathbf{U}_n, \mathbf{B}_n, \mathbf{Z}_n)$, the following estimates are required (for $\tau_U, \tau_B > 0$):

$$\begin{aligned} \widehat{\mathbf{U}}_n = \arg\min_{\mathbf{U}} &\frac{1}{2}\left\|\mathcal{S}(\mathbf{Y}) - \mathcal{S}\mathscr{F}(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B}_n)\right\|_{\text{F}}^2 + \frac{\tau_U}{2}\left\|\mathbf{U} - \mathbf{U}_n\right\|_{\text{F}}^2 + \frac{\lambda_1}{2}\left\|\mathbf{Z}_n - \mathscr{P}\mathscr{F}_t(\mathbf{U}\breve{\mathbf{\Lambda}}\mathbf{B}_n)\right\|_{\text{F}}^2 \\ &\text{s.to}\left\|\mathbf{U}\mathbf{e}_i\right\| \leq C_U, \quad \forall i \in \{1, \ldots, d\}, \end{aligned} \tag{4}$$

and

$$\begin{aligned} \widehat{\mathbf{B}}_n \in \arg\min_{\mathbf{B}} &\frac{1}{2}\left\|\mathcal{S}(\mathbf{Y}) - \mathcal{S}\mathscr{F}(\mathbf{U}_n\breve{\mathbf{\Lambda}}\mathbf{B})\right\|_{\text{F}}^2 + \frac{\tau_B}{2}\left\|\mathbf{B} - \mathbf{B}_n\right\|_{\text{F}}^2 + \frac{\lambda_1}{2}\left\|\mathbf{Z}_n - \mathscr{P}\mathscr{F}_t(\mathbf{U}_n\breve{\mathbf{\Lambda}}\mathbf{B})\right\|_{\text{F}}^2 + \lambda_3\|\mathbf{B}\|_1 \\ &\text{s.to } \mathbf{1}_{N_1}^\top\mathbf{B} = \mathbf{1}_{N_{\text{fr}}}^\top. \end{aligned} \tag{5}$$

Both (4) and (5) can be viewed as affinely constrained composite convex minimization tasks, so that [20] can be used to solve them. From a computational complexity perspective, it is important to point out that the proposed scheme relies on minimization sub-tasks for three block variables, whose computational complexities depend on the solver adopted for the optimization. It is also worth noting that those three minimization sub-tasks are independent of each other, hence they can be solved in parallel. Details on the implementation issues of parallel optimization techniques, which speed up execution time, and of [20], as the optimization module which solves the minimization sub-tasks, are deferred to an upcoming journal publication.

III. NUMERICAL TESTS

Following [13], an MRXCAT phantom [25], based on extended cardiac torso (XCAT), was used to generate breath-hold cardiac cine data of size $(N_{\mathrm{p}}, N_{\mathrm{f}}, N_{\mathrm{fr}}) = (408, 408, 360)$. Data are dynamic, since they include 15 cardiac cycles and 24 cardiac phases, and under-sampled: **i**) A number of $\nu := 4$ navigator lines are considered (*cf.* Fig. 1); **ii**) Only 8 samples, picked randomly, out of the N_{fr} -length time profile of each k-space pixel are considered. The locations of the chosen samples are distributed independently over all time profiles. Overall, the under-sampling factor turns out to be (approximately) equal to 32. The proposed method is tested vs. the PCA-based PS-sparse scheme [26] and [13]. Validation is performed via the normalized-root-mean-square error $\mathrm{NRMSE} := \|\mathbf{X} - \widehat{\mathbf{X}}\|_{\mathrm{F}} / \|\mathbf{X}\|_{\mathrm{F}}$ where $\mathbf{X}$ is the original image-domain data (*cf.* Sec. II) and $\widehat{\mathbf{X}}$ stands for an estimate of $\mathbf{X}$, computed via any of the employed reconstruction schemes. Parameters of each method are tuned to achieve the best possible NRMSE performance. With regards to the proposed method, the dimensionality of the compressed data is $d := 4$, while the number of landmark points is set to $N_{\mathrm{l}} := 60$.

Figs. 2 and 3 depict results of the numerical tests for all employed algorithms. Fig. 2 demonstrates reconstruction results for image frame #351 of the dMRI data. Further, Fig. 3 compares the NMRSE errors across all 360 frames. It can be observed that **i**) the errors for the proposed scheme are uniform across all 360 frames, as opposed to the fluctuations observed for the other two schemes, and **ii**) the proposed scheme achieves the lowest NMRSE across all frames. Quantitative results on the whole data set (408 × 408 × 360) show an NRMSE of 0.051 for the proposed scheme, which is lower in comparison to the value of 0.081 for PS-Sparse [26] and 0.079 for [13].

IV. CONCLUSIONS AND THE ROAD AHEAD

A modeling framework of manifold data and its application to dMRI were introduced. Given a cloud of highly under-sampled dMRI data, landmark points were computed to capture the underlying low-dimensional manifold-data geometry. Low-dimensional renditions of the landmark points were also computed to promote parsimonious representations and efficient storage of data. Overall, the dMRI reconstruction task was viewed as an inverse problem where the data fidelity is modeled via a bi-linear term, and the objective function is penalized by a term which quantifies the a-priori information on the periodicity of the observed time series. Preliminary numerical tests on synthetically generated phantom dMRI data underline the rich potential of the proposed framework against state-of-the-art dMRI

reconstruction techniques. On-going research includes tests on real data, comparisons with additional state-of-the-art techniques and several extensions of the advocated model, which will be presented during CAMSAP 2017.

ACKNOWLEDGMENT

The authors would like to thank the Institute of Biomedical Eng., ETH Zurich, and the Cardiovascular Magnetic Resonance Society for making the data, used in Sec. III, available online. Moreover, this work was partially supported by the NSF grants 1514056, 1718796 and the NIH grant R21EB020861.

REFERENCES

[1]. Awate SP and DiBella EVR, “Spatiotemporal dictionary learning for undersampled dynamic MRI reconstruction via joint frame-based and dictionary-based sparsity,” in Proc. Intern. Symp. Biomedical Imag, 2012, pp. 318–321.

[2]. De Silva V and Tenenbaum JB, “Sparse multidimensional scaling using landmark points,” Stanford University, Tech. Rep, 2004.

[3]. Elhamifar E and Vidal R, “Sparse manifold clustering and embedding,” in Proc. NIPS, Granada: Spain, 12 2011.

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

[5]. Jung H, Ye JC, and Kim EY, “Improved k-t BLAST and k-t SENSE using FOCUSS,” Physics in Medicine and Biology, vol. 52, no. 11, pp. 3201–3226, 2007. [PubMed: 17505098]

[6]. Liang D, DiBella EVR, Chen R-R, and Ying L, “k-t ISD: Dynamic cardiac MR imaging using compressed sensing with iterative support detection,” Magnetic Resonance in Medicine, vol. 68, no. 1, pp. 41–53, 2012. [PubMed: 22113706]

[7]. Liang Z-P, “Spatiotemporal imaging with partially separable functions,” in Proc. Intern. Symp. Biomedical Imag, 2007, pp. 988–991.

[8]. Liang Z-P and Lauterbur PC, “An efficient method for dynamic magnetic resonance imaging,” IEEE Trans. Medical Imag, vol. 13, no. 4, pp. 677–686, 1994.

[9]. Liang Z-P, Principles of Magnetic Resonance Imaging: A Signal Processing Perspective IEEE Press, 2000.

[10]. Lingala S, Hu Y, DiBella EVR, and Jacob M, “Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR,” IEEE Trans. Medical Imag, vol. 30, no. 5, pp. 1042–1054, 2011.

[11]. Lustig M, Donoho D, and Pauly JM, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” Magnetic Resonance in Medicine, vol. 58, no. 6, pp. 1182–1195, 2007. [PubMed: 17969013]

[12]. Mardani M, Ying L, Scutari G, Slavakis K, and Giannakis GB, “Dynamic MRI using subspace tensor tracking,” in Proc. Engineering in Medicine & Biology Conf, Chicago: USA, 8 2014.

[13]. Nakarmi U, Slavakis K, Lyu J, and Ying L, “M-MRI: A manifold-based framework to highly accelerated dynamic magnetic resonance imaging,” in Proc. Intern. Symp. Biomedical Imag, Melbourne: Australia, 4 2017.

[14]. Nakarmi U, Zhou Y, Lyu J, Slavakis K, and Ying L, “Accelerating dynamic magnetic resonance imaging by nonlinear sparse coding,” in Proc. Intern. Symp. Biomedical Imag, 2016, pp. 510–513.

[15]. Otazo R, Kim D, Axel L, and Sodickson DK, “Combination of compressed sensing and parallel imaging for highly accelerated first-pass cardiac perfusion MRI,” Magnetic Resonance in Medicine, vol. 64, no. 3, pp. 767–776, 2010. [PubMed: 20535813]

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

[17]. Rockafellar RT, Convex Analysis Princeton, NJ: Princeton University Press, 1970.

[18]. Saul LK and Roweis ST, “Think globally, fit locally: Unsupervised learning of low dimensional manifolds,” J. Machine Learning Research, vol. 4, pp. 119–155, 2003.

[19]. Slavakis K, Giannakis GB, and Leus G, “Robust sparse embedding and reconstruction via dictionary learning,” in Proc. Conf. Inform. Sciences and Systems, Baltimore: USA, 3 2013.

[20]. Slavakis K and Yamada I, “Fejér-monotone hybrid steepest descent method for affinely constrained and composite convex minimization tasks,” arXiv e-prints, 2017 [Online]. Available: http://arxiv.org/abs/1608.02500

[21]. Tu LW, An Introduction to Manifolds New York: Springer, 2008.

[22]. Usman M, Atkinson D, Kolbitsch C, Schaeffter T, and Prieto C, “Manifold learning based ECG-free free-breathing cardiac CINE MRI,” J. Magnetic Resonance Imag, vol. 41, no. 6, pp. 1521–1527, 2015.

[23]. Wang Y, Cao N, Liu Z, and Zhang Y, “Real-time dynamic MRI using parallel dictionary learning and dynamic total variation,” Neurocomputing, vol. 238, pp. 410–419, 2017.

[24]. Wang Y, Zhou Y, and Ying L, “Undersampled dynamic magnetic resonance imaging using patch-based spatiotemporal dictionaries,” in Proc. Intern. Symp. Biomedical Imag, 2013, pp. 294–297.

[25]. Wissmann L, Santelli C, Segars WP, and Kozerke S, “MRXCAT: Realistic numerical phantoms for cardiovascular magnetic resonance,” J. Cardiovascular Magnetic Resonance, vol. 16, no. 1, p. 63, 2014.

[26]. Zhao B, Haldar JP, Christodoulou AG, and Liang Z-P, “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.

[27]. Zhong W, Li D, Wang L, and Zhang M, “Low-rank plus sparse reconstruction using dictionary learning for 3D-MRI,” in Proc. Intern. Congress Image and Signal Process., BioMedical Eng. and Informatics, 2016, pp. 1407–1411.

[28]. Zhou Y, Shi C, Ren F, Lyu J, Liang D, and Ying L, “Accelerating MR parameter mapping using nonlinear manifold learning and supervised pre-imaging,” in Proc. Intern. Symp. Biomedical Imag, 2015, pp. 897–900.

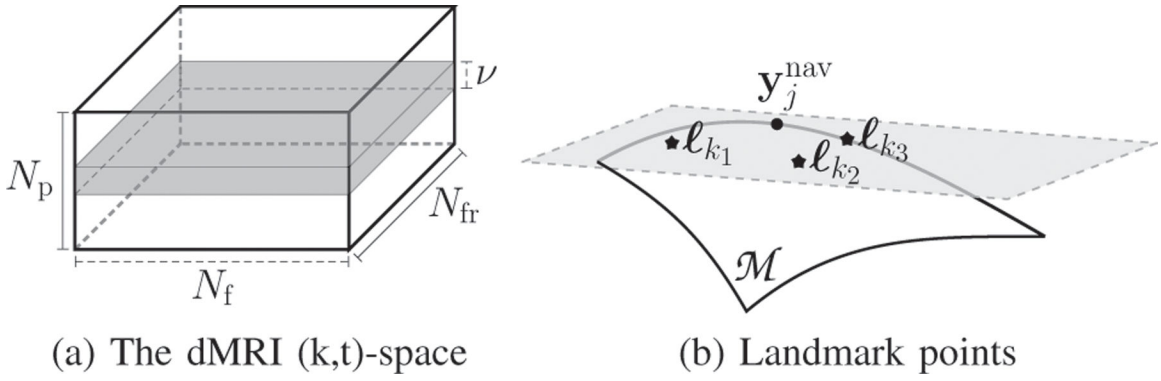

Fig. 1.

(a) The $N_{\mathrm{p}} \times N_{\mathrm{f}} \times N_{\mathrm{fr}}$ (k,t)-space. Data geometry is learned from the "navigator" data, *i.e.*, the data which lie in the gray-colored $\nu \times N_{\mathrm{f}} \times N_{\mathrm{fr}}$ area of the (k,t)-space ($\nu \ll N_{\mathrm{p}}$). (b) Landmark points $\left\{\ell_{k_i}\right\}_{i=1}^{3}$ are affinely combined to describe $\mathbf{y}_j^{\mathrm{nav}}$. All possible affine combinations of $\left\{\ell_{k_i}\right\}_{i=1}^{3}$ are denoted by the gray-colored area.

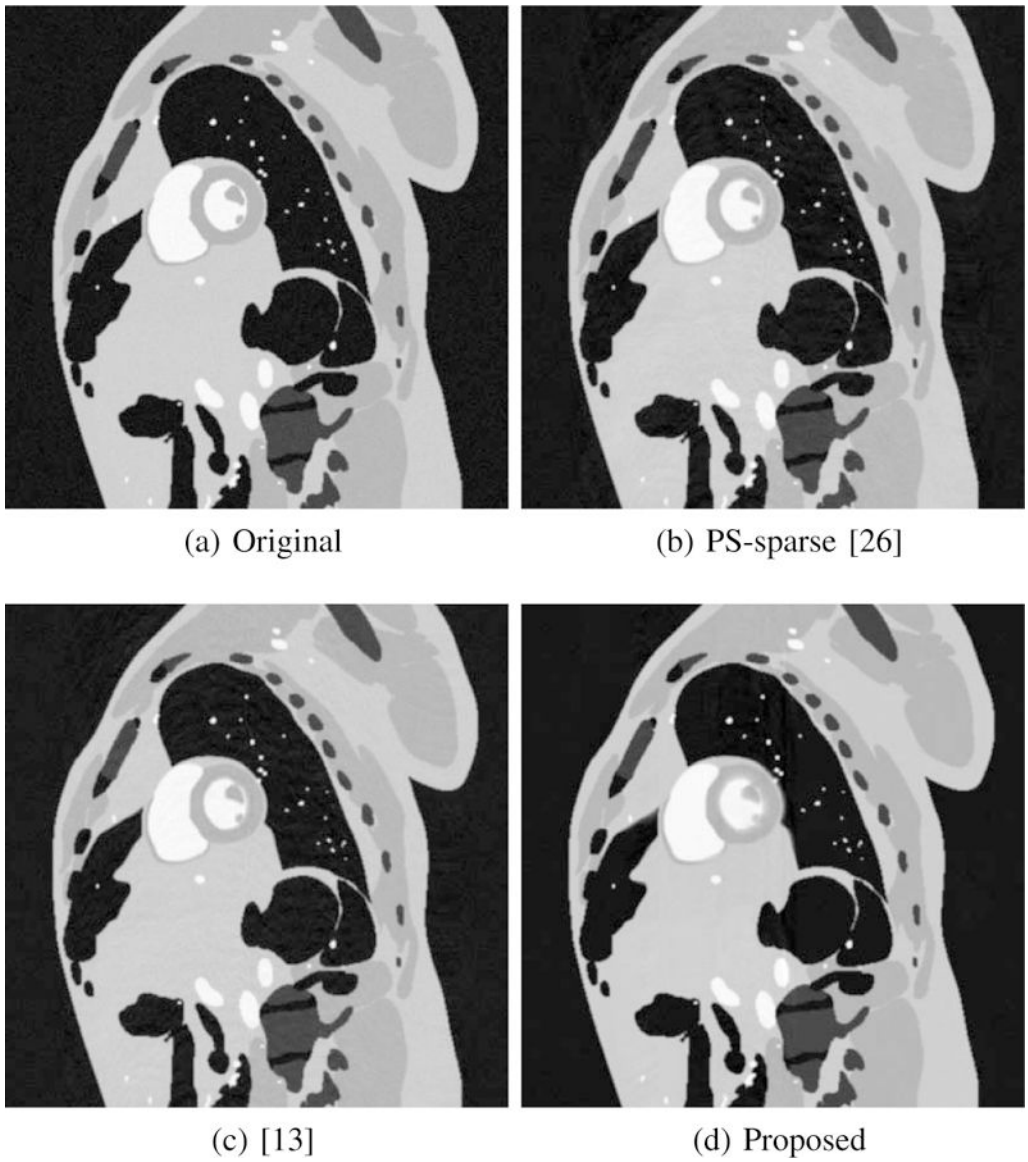

Fig. 2.

Original frame #351, of size 408 × 408, and its reconstructed versions. NRMSE values per method are: 0.059 for PS-sparse [26], 0.06 for [13] and 0.052 for the proposed method.

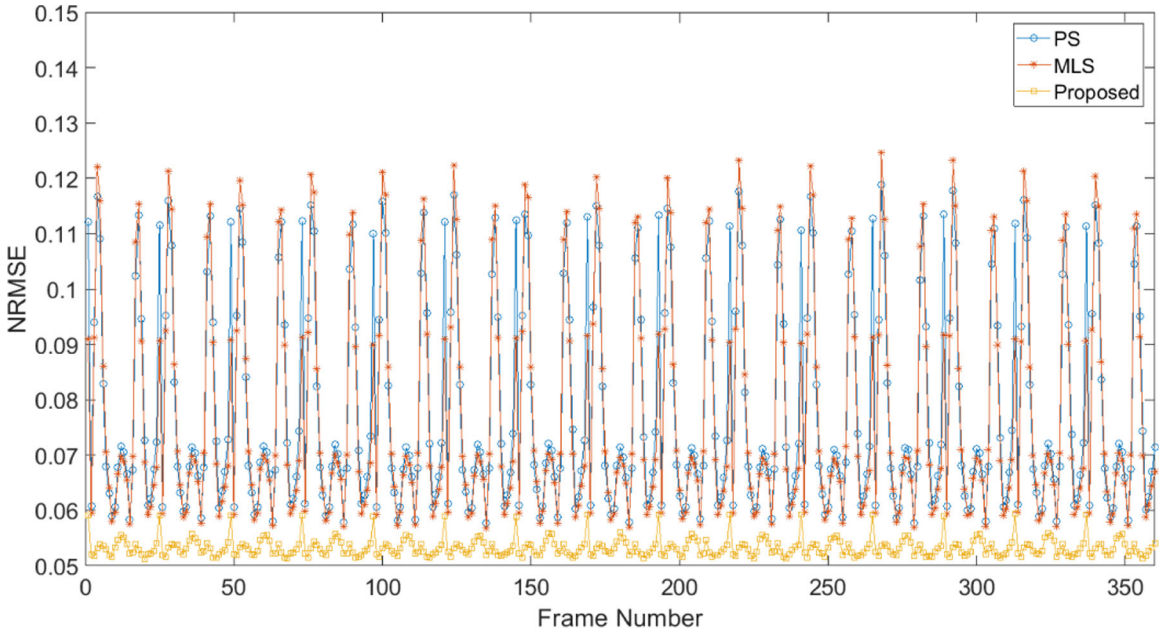

Fig. 3.

Frame-wise NRMSE: Comparing the proposed scheme (yellow curve), [13] (blue curve, ‘MLS’) and [26] (red curve, ‘PS’) over all 360 MRXCAT frames of size 408 × 408. The proposed method achieves the lowest NRMSE uniformly across all frames.