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
Emails: {kslavaki,gauravna,abose3,ukashnak,leiyng}@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 coarsely 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 decomposes 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 insurmountable 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, there 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 decomposition 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 $N_t$-fold Cartesian product of the $(N_p \times N_f)$-
provide a “concise description,” in a user-defined sense, of the data cloud \( \{y_{j}^{\text{new}}\}_{j=1}^{N_0} \). To this end, the following modeling assumption imposes structure on \( \{y_{j}^{\text{new}}\}_{j=1}^{N_0} \).

**Assumption 2.** Data \( \{y_{j}^{\text{new}}\}_{j=1}^{N_0} \) lie on a smooth low-dimensional manifold \( \mathcal{M} \) [21] embedded in the high-dimensional Euclidean space \( \mathbb{C}^{\nu N_0} \) (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 \( y_{j}^{\text{new}} \) (the gray-colored area in Fig. 1b depicts all possible affine combinations of \( \{\ell_{k,1}, \ell_{k,2}, \ell_{k,3}\} \)). Upon defining the \( \nu N_0 \times N_1 \) matrix \( \mathbf{A} := [\ell_1, \ell_2, \ldots, \ell_{N_1}] \), it is assumed that there exists an \( N_1 \times 1 \) vector \( b_j \) that renders the approximation error \( \|y_{j}^{\text{new}} - \mathbf{A}b_j\| \) small, where \( \|\cdot\| \) denotes the standard Euclidean norm of space \( \mathbb{C}^{\nu N_0} \). Since affine combinations are desirable, \( b_j \) is constrained to satisfy \( 1^\top N_1 b_j = 1 \), where \( N_1 \) stands for the all-one \( N_1 \times 1 \) vector and superscript \( \top \) denotes vector/matrix transposition. Moreover, motivated by the low-dimensional nature of \( \mathcal{M} \), according to As. 2, it is envisioned that only a few landmark points cooperate into representing \( y_{j}^{\text{new}} \), i.e., \( b_j \) is sparse. The previous arguments are summarized into the following modeling hypothesis.

**Assumption 3.** There exist a sparse \( N_1 \times N_0 \) matrix \( \mathbf{B} \), with \( 1^\top N_1 B = 1_{N_0} \), and a \( \nu N_0 \times N_1 \) matrix \( \mathbf{E}_2 \), which gathers approximation errors, s.t. \( Y_{\text{nav}} = AB + E_2 \).

The previous modeling assumption holds true in the prototypical case where \( \mathcal{M} \) is a linear subspace. In such a case, any linearly independent subset of \( \{y_{j}^{\text{new}}\}_{j=1}^{N_0} \), which spans the column (range) space of \( Y_{\text{nav}} \), can be selected as the columns of \( \mathbf{A} \). Notice also that under such a choice for \( \mathbf{A} \), the coefficient matrix \( \mathbf{B} \) satisfies \( 1^\top N_1 B = 1_{N_0} \), since the column (range) space of \( Y_{\text{nav}} \), being a linear subspace, is also the affine hull [17] of the columns of \( \mathbf{A} \).

Although several strategies may be implemented to identify the landmark points \( \mathbf{A} \) of a data cloud that lies onto or close to a general manifold \( \mathcal{M} \), a greedy optimization methodology, introduced in [2], is adopted here. In short, at every step of the algorithm, a landmark point is selected from \( \{y_{j}^{\text{new}}\}_{j=1}^{N_0} \) by maximizing, over all un-selected \( \{y_{j}^{\text{new}}\}_{j=1}^{N_0} \), 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 \( O(N_1 N_0) \), which is naturally heavier than that of the naive random-selection algorithm that chooses \( \{\ell_{k,1}\}_{k=1}^{N_1} \) randomly from \( \{y_{j}^{\text{new}}\}_{j=1}^{N_0} \).

**B. Reducing the dimension of the landmark points**

The landmark points \( \mathbf{A} \), 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 \( \mathbf{A} \). To this end, the methodology of [19], which is motivated by [3], [18], is employed. The approach comprises two steps.

1) Given \( \mathbf{A} \) and a user-defined \( \lambda_W \), solve the convex minimization task

\[ \min_{W \in \mathbb{C}^{\nu N_0 \times N_1}} \|A - AW\|_F^2 + \lambda_W \|W\|_1 \]

s.t. \( 1^\top N_1 W = 1_{N_0} \) and \( \text{diag}(W) = 0 \). \( (1) \)

where \( \|\cdot\|_F \) stands for the Frobenius norm of a matrix. A few comments on the choice of the loss function and constraints...
follow. Since \( \{\xi_k\}_{k=1}^{N_1} \) lie on the manifold \( \mathcal{M} \), then according to As. 2, Fig. 1b and the discussion regarding \( Y_{\text{raw}} \), any point taken from \( \{\xi_k\}_{k=1}^{N_1} \) may be faithfully approximated by an affine combination of the rest of the landmark points. In other words, there exists a matrix \( W \) s.t. \( \Lambda \approx \Lambda W \). With \( \mathbf{1}_{N_1} W = \mathbf{1}_{N_1} \), manifesting the previous desire for affine combinations, the constraint \( \text{diag}(W) = 0 \) is used to exclude the trivial solution of the identity matrix \( \mathbf{1}_{N_1} \) for \( W \). 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 \( W \) has been obtained from the previous step and for a user-defined integer number \( d \ll N_p N_t \), solve

\[
\min_{\Lambda \in \mathbb{C}^{d \times N_1}} \| \Lambda - \hat{\Lambda} W \|_F^2 \quad \text{s.t.} \quad \Lambda^* \Lambda = I_d,
\]

where the constraint \( \Lambda^* \Lambda = I_d \) is used to exclude the trivial solution of \( \Lambda = 0 \). 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 * superscript) which comprises the \( d \) minimal eigenvectors of \((\mathbf{1}_{N_1} - W)(\mathbf{1}_{N_1} - W)^*\).

The following hypothesis establishes a linear relation between \( \Lambda \) and its low-dimensional rendition \( \hat{\Lambda} \).

**Assumption 4.** There exist an \( N_p N_1 \times d \) matrix \( G_3 \) and an \( N_p N_1 \times N_1 \) matrix \( E_3 \), which gathers all approximation errors, so that \( \Lambda = G_3 \hat{\Lambda} + E_3 \).

Matrix \( G_3 \) can be viewed as the “decomposition” operator which reconstructs the “full” \( \Lambda \) from its low-dimensional representation \( \hat{\Lambda} \).

### C. Data recovery task

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

\[
Y = F(UAB) + E. \tag{2}
\]

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

In practice, only few \((k,t)\)-space data are known. To explicitly take account of the limited number of data, a (linear) sampling operator \( S(\cdot) \) is introduced where \( S(Y) \) keeps only those entries of \( Y \) 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 \( F_1 \) of the time profile of every MR pixel, i.e., every row of the matrix \( F_1(UAB) \), is a sparse vector. Since the DC Fourier coefficients do not provide any useful information on periodic time series, the linear operator \( P(\cdot) \) is also applied to \( P[F_1(UAB)] \) to remove the first column of \( F_1(UAB) \), 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 \( S(Y) \). These include the navigator \( Y_{\text{raw}} \) one. Choose parameters \( \lambda_1, \lambda_2, \lambda_3, C_U, \tau_U, \tau_B > 0 \), as well as \( \zeta \in \{0, 1\} \) and \( \gamma_0 \in \{0, 1\} \).

1. Compute landmark points \( \hat{\Lambda} \) from the columns of \( Y_{\text{raw}} \) according to Sec. II-A.
2. Compute the low-dimensional rendition \( \hat{\Lambda} \) of \( \Lambda \) according to Sec. II-B.
3. Arbitrarily fix \((U_0, B_0, Z_0)\) and set \( n = 0 \).
4. while \( n \geq 0 \) do
5. Available are \((U_0, B_0, Z_0)\) and \( \gamma_0 \).
6. Let \( \gamma_{n+1} := \gamma_0(1 - \zeta \gamma_n) \).
7. Obtain \( \hat{U}_n \) via (4), \( \hat{B}_n \) via (5), and the \((i, j)\)th entry of \( \hat{Z}_n \), \( \forall (i,j) \), via the following soft-thresholding rule:

\[
[Z_n]_{ij} := \max\left\{ \lambda_2/\lambda_1, \frac{1}{\|PF_1(U_0\hat{A}B_0)\|_1} \right\}.
\]
8. Update \( (U_{n+1}, B_{n+1}, Z_{n+1}) := (1 - \gamma_{n+1})(U_n, B_n, Z_n) \). + \( \gamma_{n+1}(U_0, B_0, Z_0) \).
9. Set \( n \) equal to \( n + 1 \) and go to step 5.
10: end while

**Output:** Extract the limit points \( U_* \) and \( B_* \) of sequences \( (U_n) \) and \( (B_n) \), respectively, and recover the dMRI data by the estimate \( \hat{X} := U_*\hat{A}B_* \).

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

\[
\min_{(U, B, Z)} \frac{1}{2} \| S(Y) - SF(U\hat{A}B) \|^2_F + \frac{\lambda_2}{2} \| Z - PF_1(U\hat{A}B) \|^2_F + \lambda_2 \| Z \|_1 + \lambda_3 \| B \|_1 \tag{3}
\]

s.t. \( \| U e_i \| \leq C_U, \forall i \in \{1, \ldots, d\} \) and \( 1_{N_1}^T \hat{B} = 1_{N_1}^T \).

Step 7 of Alg. 1 comprises convex minimization sub-tasks. More specifically, at every step of the algorithm, given \((U_0, B_0, Z_0)\), the following estimates are required for \( \tau_U, \tau_B > 0 \):

\[
\hat{U}_n = \arg \min_U \frac{1}{2} \| S(Y) - SF(U\hat{A}B_n) \|^2_F + \frac{\tau_U}{2} \| U - U_n \|^2_F
\]

\[
+ \frac{\lambda_2}{2} \| Z_n - PF_1(U\hat{A}B_n) \|^2_F \tag{4}
\]

s.t. \( \| U e_i \| \leq C_U, \forall i \in \{1, \ldots, d\} \),

and

\[
\hat{B}_n = \arg \min_B \frac{1}{2} \| S(Y) - SF(U_0\hat{A}B) \|^2_F + \frac{\tau_B}{2} \| B - B_n \|^2_F
\]

\[
+ \frac{\lambda_2}{2} \| Z_n - PF_1(U_0\hat{A}B) \|^2_F + \lambda_3 \| B \|_1 \tag{5}
\]

s.t. \( 1_{N_1}^T \hat{B} = 1_{N_1}^T \).
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 $(N_p, N_t, N_h) = (408, 408, 360)$. Data are dynamic, since they include 15 cardiac cycles and 24 cardiac phases, and undersampled: i) A number of $\nu := 4$ navigator lines are considered (cf. Fig. 1); ii) Only 8 samples, picked randomly, out of the $N_h$-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 $\text{NRMSE} := \| \hat{X} - X \|_F / \| X \|_F$, where $X$ is the original image-domain data (cf. Sec. II) and $\hat{X}$ stands for an estimate of $X$, 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 $N_l := 60$.

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 $(408 \times 408 \times 360)$ 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.

![Original frame #351 of size 408 x 408, and its reconstructed versions.](image1)

**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.](image2)

**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.