

Accelerated dynamic MRI using learned representations

Anthony G. Christodoulou, *Member, IEEE*, and Sajan Goud Lingala, *Member, IEEE*

I. INTRODUCTION

Dynamic magnetic resonance imaging (MRI) can be used to image a wide range of dynamic processes within the body, including the motion of internal organs, tissue-level nuclear magnetic resonance (NMR) relaxation, and dynamic contrast enhancement (DCE) of contrast agents. The ability of MRI to safely provide unique soft-tissue contrast and comprehensive functional information has made dynamic MRI central to a number of imaging exams for cardiac, interventional, vocal tract, cancer, and gastrointestinal applications, among others. Unfortunately, MRI is a notoriously slow imaging modality due to fundamental physical and physiological limitations. These limitations result in tradeoffs between spatial and temporal resolutions, spatial coverage, and signal-to-noise ratio, and have made dynamic MRI a challenging technical goal.

The past decade experienced a surge of several compressed sensing (CS) MRI approaches (e.g., [1], [2]) which have made substantial headway toward addressing this challenge. At its core, CS-MRI exploits redundancy of images in known predetermined transform domains (e.g. Fourier, wavelet). However, one drawback to this approach has been that fixed transforms often do not provide the most efficient representation of images, thereby limiting the maximum achievable acceleration. To address these shortcomings, several adaptive, learning-based methods have been proposed. These schemes have revolutionized dynamic MRI for many applications, offering exciting new capabilities in biomedical imaging.

Here we offer a unified view of several different approaches to dynamic imaging using learned representations, focusing on the signal processing aspects which make each class of learning methods so powerful. These methods include low-rank methods [3]–[5], blind compressed sensing methods [6], higher-order multidynamic methods [7]–[10], explicit motion estimation and

A.G.C. is with the Biomedical Imaging Research Institute, Cedars-Sinai Medical Center. S.G.L. is with the Roy J. Carver Department of Biomedical Engineering, University of Iowa.

compensated recovery methods [11], [12], manifold regularized recovery methods [13], [14], as well as deep learning approaches [15]–[17].

II. PROBLEM FORMULATION

We consider a complex image $\tilde{x}(\mathbf{r}, \tau_1(t), \tau_2(t), \dots, \tau_L(t))$ that is a scalar function of spatial location $\mathbf{r} = [x, y, z]^T$ and $L \geq 1$ time-varying independent variables $\{\tau_\ell(t)\}_{\ell=1}^L$, each representing a different physical or physiological dynamic process such as motion, NMR relaxation, or simply the passage of time t . The dynamic MR signal from this image can be observed from noisy, spatially-encoded discrete measurements in (\mathbf{k}, t) -space:

$$b_c(\mathbf{k}_m, t_n) = \int_{\mathbf{r}} x(\mathbf{r}, t_n) s_c(\mathbf{r}, t_n) \exp(-j\mathbf{k}_m^T \mathbf{r}) d\mathbf{r} + \eta_c(\mathbf{k}_m, t_n) \quad (1)$$

$$x(\mathbf{r}, t) = \tilde{x}(\mathbf{r}, \tau_1(t), \tau_2(t), \dots, \tau_L(t)), \quad (2)$$

where $s_c(\mathbf{r}, t_n)$ denotes the coil sensitivity profile of the c^{th} channel (which may vary with time), (\mathbf{k}_m, t_n) indicates the $(m, n)^{\text{th}}$ sampling location in (\mathbf{k}, t) -space, and η denotes additive complex-valued Gaussian noise. Specific dynamic processes can be targeted by holding unwanted τ 's constant during signal acquisition. For example, the ℓ th dynamic process can be isolated by enforcing $\tau_q(t) = \kappa_q \forall q \neq \ell$ during acquisition for some set of constant values $\{\kappa_q\}_{q \neq \ell}$. Practically speaking, this means either physically holding $\{\tau_q(t)\}_{q \neq \ell}$ constant (e.g., pausing respiratory motion by breath holding) or waiting to acquire data only when the condition $\tau_q(t) = \kappa_q \forall q \neq \ell$ is met (e.g., removing the appearance of cardiac motion by synchronizing data acquisition to a monitored electrocardiogram (ECG) signal). The decision of which dynamic processes to image and which to suppress is application-dependent.

The expression in Eq. (1) can be more compactly written in matrix-vector form as $\mathbf{b} = \mathcal{A}(\mathbf{X}) + \eta$, where the matrix \mathbf{X} has elements $[\mathbf{X}]_{mn} = x(\mathbf{r}_m, t_n)$, and the linear operator \mathcal{A} models the coil sensitivity encoding as well as Fourier encoding on an arbitrary sampling trajectory (e.g. Cartesian, or non-Cartesian). The goal of accelerated dynamic MR image reconstruction is typically to recover $x(\mathbf{r}, t)$ from undersampled (\mathbf{k}, t) -space measurements through the use of low-dimensional learned representations of the underlying image. Recovery of $\tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_L)$ is also possible, as will be discussed in Section IV.

III. SPATIOTEMPORAL DECOMPOSITION MODELS

The general partial separability (PS) model proposed by Liang [3] can be used to interpret several fixed or learned spatiotemporal decomposition models in a unified manner. The PS model utilizes a global signal model to represent the signal time profile at each voxel ($x(\mathbf{r}, t)$) as:

$$x(\mathbf{r}, t) = \sum_{i=1}^R x_i(\mathbf{r}, t) = \sum_{i=1}^R u_i(\mathbf{r})v_i(t); \quad (3)$$

The above decomposition implies that $x_i(\mathbf{r}, t)$ are separable functions of \mathbf{r} , and t ; i.e., that they can be factored as a product of the spatial model coefficients $u_i(\mathbf{r})$ and the temporal basis functions $v_i(t)$. Here R denotes the total number of basis functions (or the model order). Eq. (3) implies that $x(\mathbf{r}, t)$ can be efficiently represented in either of two low-dimensional feature spaces, $\text{span}(\{u_i(\mathbf{r})\}_{i=1}^R)$ or $\text{span}(\{v_i(t)\}_{i=1}^R)$, providing an avenue for representation learning to recover $x(\mathbf{r}, t)$ from undersampled measurements.

Both $\text{span}(\{u_i(\mathbf{r})\}_{i=1}^R)$ and $\text{span}(\{v_i(t)\}_{i=1}^R)$ are R -dimensional linear subspaces (i.e., linear manifolds intersecting the origin), so we adopt a linear algebraic framework to understand (3). In matrix form, (3) states that the dynamic Casorati matrix $\mathbf{X} \in \mathbb{C}^{M \times N}$ can be decomposed as a product of a spatial coefficient matrix $\mathbf{U} \in \mathbb{C}^{M \times R}$ and a matrix $\mathbf{V} \in \mathbb{C}^{R \times N}$ containing temporal functions:

$$\underbrace{\begin{pmatrix} x(\mathbf{r}_1, t_1) & \dots & x(\mathbf{r}_1, t_N) \\ x(\mathbf{r}_2, t_1) & \dots & x(\mathbf{r}_2, t_N) \\ \dots & \dots & \dots \\ x(\mathbf{r}_M, t_1) & \dots & x(\mathbf{r}_M, t_N) \end{pmatrix}}_{\mathbf{X}_{M \times N}} = \underbrace{\begin{pmatrix} u_1(\mathbf{r}_1) & \dots & u_R(\mathbf{r}_1) \\ u_1(\mathbf{r}_2) & \dots & u_R(\mathbf{r}_2) \\ \dots & \dots & \dots \\ u_1(\mathbf{r}_M) & \dots & u_R(\mathbf{r}_M) \end{pmatrix}}_{\mathbf{U}_{M \times R}} \underbrace{\begin{pmatrix} v_1(t_1) & \dots & v_1(t_N) \\ v_2(t_1) & \dots & v_2(t_N) \\ \dots & \dots & \dots \\ v_R(t_1) & \dots & v_R(t_N) \end{pmatrix}}_{\mathbf{V}_{R \times N}}. \quad (4)$$

Here M represents the total number of voxels, and N denotes the number of time frames. Based on constraints imposed on the matrices \mathbf{U} , \mathbf{V} , and the model order R , the general PS model in (4) can help classify various models, as laid out in Figure (1).

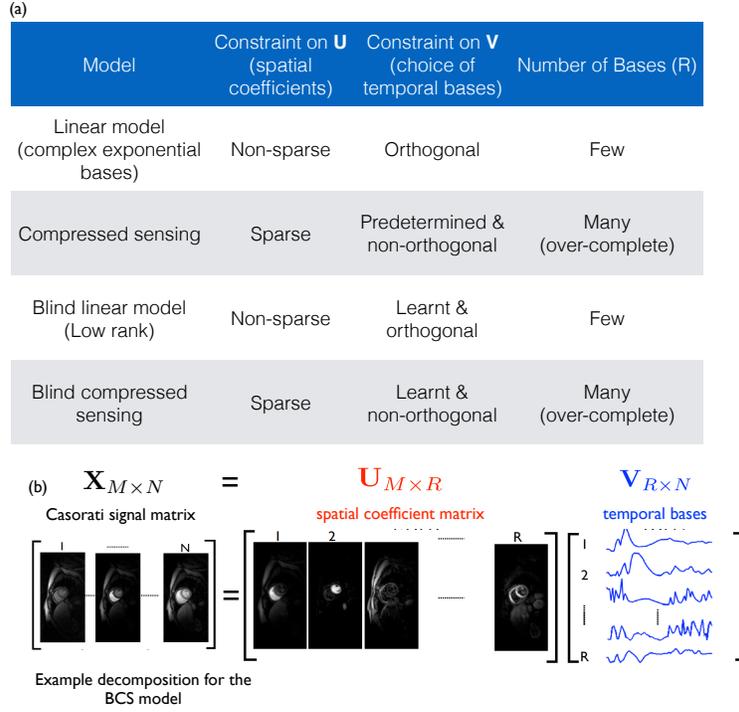


Fig. 1. The generalized PS model provides a unified view of various spatiotemporal decomposition models. (a) shows the various models based on choices of constraints on the spatial coefficient matrix \mathbf{U} and the temporal matrix \mathbf{V} , and the model order R . (b) shows an example pictorial decomposition of \mathbf{U} and \mathbf{V} for the blind CS model on a myocardial perfusion dataset.

A. Linear models with complex exponential bases

The earliest adaptive model in dynamic MRI termed as dynamic imaging by model estimation (DIME) was proposed by Liang et al. [18]. DIME focuses on the recovery of a quasi-periodic dynamic signal as a linear combination of a limited number of complex exponential temporal basis functions. Therefore the generalized PS model in (3), (4) reduces to:

$$x(\mathbf{r}, t) = \sum_{i=1}^R u_i(\mathbf{r}) e^{j2\pi f_i t}; \quad R < N \quad (5)$$

The frequencies $\{f_i\}_{i=1}^R$ can be obtained from training data during the scan, typically comprising of a low spatial but high temporal resolution dynamic data. In a second step, the coefficients $\{u_i(\mathbf{r})\}_{i=1}^R$ or \mathbf{U} are determined by method of least squares fitting to the acquired k - t data (b) as:

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \|\mathcal{A}(\mathbf{U}\mathbf{V}) - \mathbf{b}\|_2^2; \quad V_{in} = e^{j2\pi f_i t_n}; \quad (6)$$

In addition, to improve the conditioning of (6), it is common to include l_2 regularization on \mathbf{U} .

The above idea was extended by several researchers, where the problem was reformulated as a spatial-spectral filtering problem, and the goal was to design k - t lattice sampling patterns to minimize coherent aliasing in the spatial-spectral space (e.g. [19]). These schemes were particularly developed to model quasi-periodic temporal patterns as observed with cardiac cine MRI, task based functional MRI.

B. Compressed sensing

Compressed sensing (CS) under the synthesis basis representation can be interpreted by the generalized PS model as having sparse coefficients \mathbf{U} of a fixed temporal basis \mathbf{V} with $R \geq N$, where \mathbf{V} is determined *a priori*. The temporal basis $\{v_i(t)\}_{i=1}^R$ belongs to an over-complete dictionary that is chosen either off-the-shelf (e.g. wavelets, complex exponential bases), or learnt from a physical model describing the evolution of the time series (e.g. via dictionary learning [20]). CS assumes the model coefficients $u_i(\mathbf{r})$ to be sparse and estimate them from the measured (\mathbf{k}, t) -space data via the following minimization:

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \|\mathcal{A}(\mathbf{UV}) - \mathbf{b}\|_2^2 + \lambda \|\text{vec}(\mathbf{U})\|_1; \text{ for pre-determined } \mathbf{V} \text{ and } R \geq N. \quad (7)$$

An alternative to Eq. (7) is to use an analysis formulation [1]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2^2 + \lambda \|\psi(\mathbf{X})\|_1, \quad (8)$$

where ψ is an appropriate sparsity inducing operator. Note that Eqs. (7) and (8) produce the same reconstruction when $\psi(\mathbf{X}) = \mathbf{XV}^{-1}$. CS eliminates the need for training scans, and instead relies on incoherent sampling. The performance of CS is heavily dependent on the specific dictionary or sparsifying operator. For example, Fourier dictionaries may be ideally suited for breath-held cine applications due to the pseudo-periodicity of cardiac motion. However for applications involving complex temporal dynamics (e.g. free breathing, contrast enhanced MRI), the use of Fourier dictionaries may be sub-optimal (e.g. as demonstrated in [5]).

C. Blind linear models or low rank models

Blind linear models can be thought of as a generalization of the adaptive DIME model to non-periodic dynamic datasets; instead of assuming the temporal basis functions in \mathbf{V} to be

periodic exponentials, they are derived from data, such as from the PCA, SVD, KLT, or other decompositions. These models fall under the umbrella of low rank models as they capitalize the rank redundancy of the Casorati matrix \mathbf{X} . Several models such as k - t PCA [21], a variant of PS [3] utilize a two step strategy to reconstruct $x(\mathbf{r}, t)$. The temporal bases $v_i(t)$ are estimated from low spatial resolution but high temporal resolution data. Subsequently, the spatial coefficients $u_i(\mathbf{r})$ are estimated by fitting the model in (3) to the acquired k - t data in a least squares sense, similar to (6).

One limitation of the above two step strategy is that it requires sufficient training data to reliably approximate the true PCA/SVD/KLT basis functions. It requires a tradeoff between the acquisition time spent to collect the training data and the time spent to collect the under-sampled high spatial frequency measurements.

To address this, single step recovery schemes have been proposed that jointly estimates both the spatial weights and temporal bases from the under-sampled data itself [5]. The reconstruction problem can be reformulated as a rank constrained optimization:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2^2 \text{ such that } \text{rank}(\mathbf{X}) \leq R < N. \quad (9)$$

Several researchers have proposed alternate regularizers as a surrogate to the rank including the convex nuclear norm ($p = 1$), and the non-convex Schatten p -norm ($p < 1$). In addition, methods that jointly exploit low rank and transform sparsity of dynamic time series have been developed. For example, the method of k - t SLR (joint exploitation of finite difference sparsity and low-rankness) can be formulated as [5]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2^2 + \lambda_1 \underbrace{\|\mathbf{X}\|_p^p}_{\text{Schatten } p \text{ norm}} + \lambda_2 \underbrace{\left\| \sqrt{|\nabla_x(\mathbf{X})|^2 + |\nabla_y(\mathbf{X})|^2 + |\nabla_t(\mathbf{X})|^2} \right\|_1}_{\text{spatiotemporal total variation}}; \quad (10)$$

The two step recovery scheme of the blind linear model has also been improved by utilizing additional sparsity constraints [22]. These improve the recovery of the $u_i(\mathbf{r})$ problem for Eq. (3) as:

$$\hat{\mathbf{U}} = \arg \min_{\mathbf{U}} \left\| \mathcal{A}(\mathbf{U} \underbrace{\mathbf{V}}_{\text{from PCA/KLT/SVD}}) - \mathbf{b} \right\|_2^2 + \lambda \underbrace{\|\psi(\mathbf{UV})\|_1}_{\text{sparsity regularizer}}; \quad (11)$$

where the choice of ψ in (11) is motivated by the DMRI application. It is worth noting that since \mathbf{V} is explicitly estimated in the two step recovery scheme, it has advantages of using less memory to store \mathbf{U} and \mathbf{V} as opposed to storing the entire \mathbf{X} matrix—which is a key contributing factor when translated to the multidynamic setting, as we will discuss in Section IV.

D. Blind Compressed Sensing

The blind compressed sensing (BCS) model [6] share similarities with CS and blind linear models. Similar to CS, the voxel intensity profiles are modeled as a sparse linear combination of basis functions in a dictionary. However, instead of assuming a fixed dictionary \mathbf{V} , BCS estimates the dictionary from the undersampled measurements itself. BCS contrasts with blind linear models by assuming sparsity of the coefficients \mathbf{U} , and by using $R \geq N$ temporal basis functions, which are not necessarily orthogonal.

BCS reconstruction involves the joint estimation of $u_i(\mathbf{r})$, and $v_i(t)$ in Eq.(3) from under sampled k - t measurements (\mathbf{b}) with a sparsity constraint on $u_i(\mathbf{r})$, and a dictionary constraint on $v_i(t)$, which is required to avoid scale ambiguity in the product of $u_i(\mathbf{r})v_i(t)$. Choices of sparsity and dictionary constraints can range from utilizing the convex l_1 norm, or the non-convex l_0 , l_p ; ($0 < p < 1$) norms on $u_i(\mathbf{r})$, and unit column norm or Frobenius norm constraints on the dictionary. An example BCS reconstruction with l_1 coefficient sparsity, and unit column norm dictionary constraint can be formulated as the following constrained optimization problem:

$$\{\hat{\mathbf{U}}, \hat{\mathbf{V}}\} = \arg \min_{\mathbf{U}, \mathbf{V}} \|\mathcal{A}(\mathbf{UV}) - \mathbf{b}\|_2^2 + \lambda \|\text{vec}(\mathbf{U})\|_1 \text{ such that } \|v_i(t)\|_2^2 < 1 \forall i = 1, 2, \dots, R \quad (12)$$

It is worth noting that various other formulations of BCS scheme have been proposed, which cannot be described by the generalized PS model. For example, [23] considers overlapping patches of \mathbf{X} in the \mathbf{r} and t dimensions, and exploits sparsity in a learnt spatiotemporal over-complete dictionary from the patches. Convolutional sparse coding assumes \mathbf{X} to be a superposition of sparse feature images convolved with a collection of filters, where both the feature images and the filters are learnt from the under-sampled data [24].

IV. MULTIDYNAMIC MODELS

An alternative to imaging $x(\mathbf{r}, t)$ is to perform *multidynamic imaging* [7]–[10], which preserves the multiple independent variables $\{\tau_\ell\}_{\ell=1}^L$ and applies multidimensional signal modeling to image

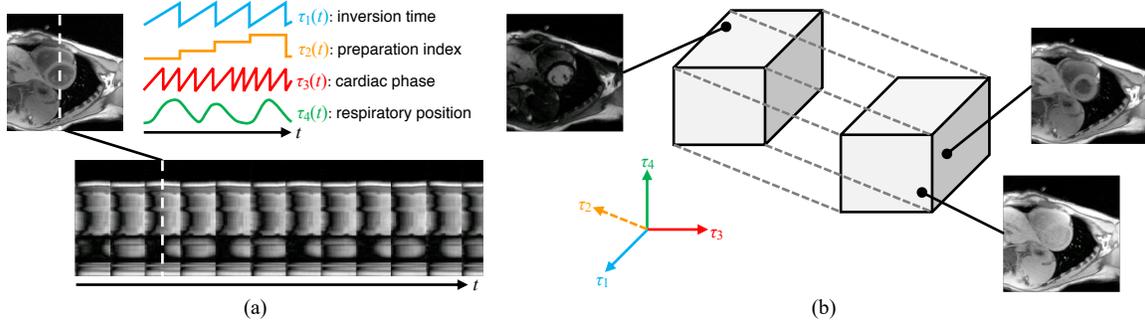


Fig. 2. Illustration of the multidynamic imaging concept. Here, four dynamic processes occur during the scan: T_1 relaxation (a function of inversion time), T_2 relaxation (a function of T_2 preparation module duration), cardiac motion, and respiratory motion. (a) All four processes overlap in the real-time dynamic image $x(\mathbf{r}, t)$, complicating the analysis of any individual dynamic process. (b) Multidynamic imaging separates these processes as different dimensions of $\tilde{x}(\mathbf{r}, \tau_1, \tau_2, \tau_3, \tau_4)$.

the underlying $\tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_L)$. This image \tilde{x} is said to have L “time dimensions”.

Multidynamic imaging of $\tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_L)$ has several benefits over imaging $x(\mathbf{r}, t)$. First, it is no longer necessary to enforce $\tau_q(t) = \kappa_q \forall q \neq \ell$ during acquisition in order to isolate a single τ_ℓ : the desired dynamic image can simply be retroactively extracted as a temporal slice of \tilde{x} , i.e., $\tilde{x}(\mathbf{r}, \kappa_1, \kappa_2, \dots, \tau_\ell, \dots, \kappa_L)$. Second, if multiple τ ’s are of interest, then they can be imaged without collapsing them into a single time dimension t , where they would otherwise confound each other during analysis. This allows multipurpose imaging such as cine imaging and T_1 mapping in the same scan, or multiparameter mapping (e.g., T_1 - T_2 mapping). Finally, this also has additional benefits for image analysis: when multiple processes are quantified from the same scan, there is no need for image registration. This is especially relevant for image fusion of parameter maps acquired in separate breath-holds, as the patient typically does not reproduce their exact respiratory position in successive breath-holds.

Imaging the entire $\tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_L)$ presents its own challenges, of course. These primarily stem from the *curse of dimensionality*, wherein the size of \tilde{x} grows geometrically with the number of dimensions. Let $\{\mathbf{r}_m\}_{m=1}^M$ be the set of M voxels in the image, and $\{\tau_{\ell, n_\ell}\}_{n_\ell=1}^{N_\ell}$ be a set of N_ℓ discrete values along the ℓ th time dimension τ_ℓ . The number of elements in $\tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_L)$ is therefore $M \prod_{\ell=1}^L N_\ell$, revealing geometric growth as L increases. As the number of elements in \tilde{x} grows, so do sampling requirements (i.e., scan time) and memory/storage requirements, presenting major practical challenges.

Fortunately, this challenge is balanced by the *blessing of dimensionality*, wherein signals in

high-dimensional spaces are often highly structured—and can be very efficiently represented with appropriate signal modeling [25]. This provides an opportunity for multidimensional extensions to Eq. (3) to represent \tilde{x} with fewer degrees of freedom than $M \prod_{\ell=1}^L N_\ell$, thereby reducing scan time and/or storage requirements. By explicitly separating sources of image dynamics into multiple time dimensions, each dynamic in \tilde{x} can be afforded its own temporal model.

To interpret various multidynamic imaging models in a unified manner, we revisit Eq. (3) in the context of multidynamic imaging:

$$\tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_L) = \sum_{i=1}^R u_i(\mathbf{r}) \tilde{v}_i(\tau_1, \tau_2, \dots, \tau_L), \quad (13)$$

which simply replaces the $\{v_i(t)\}_{i=1}^R$ with multidimensional functions $\{\tilde{v}_i(\tau_1, \tau_2, \dots, \tau_L)\}_{i=1}^R$. We notate \tilde{x} in multidimensional array/tensor form [26] as the $(L+1)$ -way tensor $\mathcal{X} \in \mathbb{C}^{M \times N_1 \times N_2 \times \dots \times N_L}$ with elements $[\mathcal{X}]_{mn_1 n_2 \dots n_L} = \tilde{x}(\mathbf{r}_m, \tau_{1, n_1}, \tau_{2, n_2}, \dots, \tau_{L, n_L})$. In this form, Eq. (13) becomes

$$\mathcal{X} = \mathcal{V} \times_1 \mathbf{U}, \quad (14)$$

where $\mathcal{V} \in \mathbb{C}^{R \times N_1 \times \dots \times N_L}$ has elements $[\mathcal{V}]_{in_1 n_2 \dots n_L} = \tilde{v}_i(\tau_{1, n_1}, \tau_{2, n_2}, \dots, \tau_{L, n_L})$ and where the definition of \mathbf{U} is unchanged from Eq. (4). As before, the constraints imposed on \mathbf{U} and \mathcal{V} and the value of R help to classify various multidynamic image models.

Multidimensional compressed sensing

Multidimensional compressed sensing for multidynamic imaging was first described by Feng et al. as the XD-GRASP method [7], which demonstrated cardiac- and respiratory-resolved imaging of the heart as well as respiratory- and DCE-resolved abdominal imaging. Cheng et al. later expanded upon the concept with the XD flow method [9], calculating multiple reconstructions with different combinations of cardiac, respiratory, flow and DCE time dimensions from one scan.

This class of methods is characterized by a large predetermined basis \mathcal{V} and sparse \mathbf{U} . The advantage of this model over its single-dynamic counterpart is that \tilde{x} is potentially more compressible along an individual time dimension τ_ℓ than along t . For example, we can reasonably expect the difference image along τ_ℓ ,

$$\nabla_{\tau_\ell} \{\tilde{x}\} = \tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_\ell, \dots, \tau_L) - \tilde{x}(\mathbf{r}, \tau_1, \tau_2, \dots, \tau_\ell - \Delta\tau_\ell, \dots, \tau_L), \quad (15)$$

to be sparser than the difference between successive real-time images,

$$\begin{aligned}\nabla_t\{x\} &= x(\mathbf{r}, t) - x(\mathbf{r}, t - \Delta t) \\ &= \tilde{x}(\mathbf{r}, \tau_1(t), \tau_2(t), \dots, \tau_L(t)) - \tilde{x}(\mathbf{r}, \tau_1(t - \Delta t), \tau_2(t - \Delta t), \dots, \tau_L(t - \Delta t));\end{aligned}\quad (16)$$

this is because Eq. (15) describes the change from one dynamic effect, whereas Eq (16) describes simultaneous changes from multiple dynamics. Image reconstruction is then possible according to

$$\hat{\mathcal{X}} = \arg \min_{\mathcal{X}} \|\mathcal{A}(\mathcal{X}) - \mathbf{b}\|_2^2 + \sum_{i=2}^{L+1} \lambda_i \|\text{vec}(\Psi_i \mathbf{X}_{(i)})\|_1, \quad (17)$$

where $\mathbf{X}_{(i)}$ is the i th unfolding of \mathcal{X} , such that the rows of $\mathbf{X}_{(1)} \in \mathbb{C}^{M \times \prod_{\ell=1}^L N_\ell}$ index \mathbf{r} , the rows of $\mathbf{X}_{(i)} \in \mathbb{C}^{N_{i-1} \times M \times \prod_{\ell \neq i-1} N_\ell}$ index τ_{i-1} for $i > 1$, and the sparsifying transform Ψ_i operates along τ_{i-1} . This recovers an image tensor \mathcal{X} which is compressible, thereby reducing scan time; however, algorithms to solve Eq. (17) generally do not explicitly store and operate upon \mathcal{X} in compressed form, so memory issues related to the curse of dimensionality still remain.

An important consideration in multidynamic sampling is that $\{\tau_\ell(t)\}_{\ell=1}^L$ must be known in order to establish the mapping between t and $(\tau_1, \tau_2, \dots, \tau_L)$. Any $\tau_\ell(t)$ that describes a sequence parameter is already known *a priori*, as it is a part of the pulse sequence design. However, any $\tau_\ell(t)$ that are physiological indices such as cardiac phase (i.e., the time point in the cardiac cycle) and respiratory position cannot be known ahead of time. These physiological timings need to be either monitored using ECG recordings or respiratory navigators, or alternatively, inferred from the acquired data (i.e., “self-gated”). One of the primary benefits of multidynamic imaging is freedom from motion monitoring, so both XD-GRASP and XD flow are self-gated methods, for which the physiological $\tau_\ell(t)$ ’s are calculated from a subset of \mathbf{b} referred to as the self-gating signal. Self-gating signals are collected throughout the entire scan at high temporal sampling rate from a limited subset of \mathbf{k} -space, similar to the subject-specific training data to learn \mathbf{V} used by some of the methods described in Section III. For multidimensional CS, the self-gating signals only serve to define any unknown timings $\{\tau_\ell(t)\}_{\ell=1}^L$ and have no bearing on \mathcal{V} .

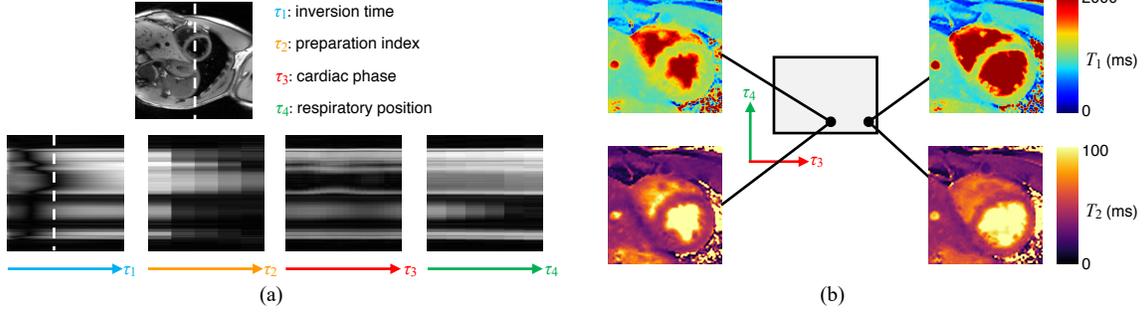


Fig. 3. Some capabilities of multidynamic imaging, as performed using MR Multitasking [10]. (a) The multidynamic image $\tilde{\mathbf{x}}(\mathbf{r}, \tau_1, \tau_2, \tau_3, \tau_4)$ can be sliced along any of its four time dimensions to retrospectively isolate individual dynamic processes. Note that by performing cardiac- and respiratory-resolved imaging, the example here did not require ECG gating or subject breath holding. (b) Because both relaxation processes T_1 and T_2 are imaged in the same motion-resolved scan, T_1 and T_2 maps are naturally co-registered and are available at any combination of cardiac phase and respiratory position.

Low-rank tensor imaging

Just as in conventional single-dynamic imaging, a complementary alternative to CS which features learned representations is blind linear modeling—or in the context of multidynamic imaging, *low-rank tensor* (LRT) imaging. He et al. described the Low-Rank Tensor with Explicit Subspace (LRTES) method [8] for multishell diffusion imaging, static T_1 - T_2^* mapping, and 2D J-resolved spectroscopic imaging. The MR Multitasking framework [10] extended the concept to handle motion-resolved imaging, performing motion-resolved quantitative imaging such as non-ECG, free-breathing T_1 - T_2 mapping in the heart and cardiac- and DCE-resolved T_1 mapping for quantitative myocardial perfusion.

LRT imaging is characterized by learning a \mathcal{V} for which $R < \min(M, \prod_{\ell=1}^L N_\ell)$, and where \mathbf{U} is not necessarily sparse. Compared to low-rank imaging with one time dimension, the LRT model places additional constraints on the structure of the temporal basis, specifically that the \mathcal{V} itself can be factored, i.e., that each $\tilde{v}_i(\tau_1, \tau_2, \dots, \tau_L)$ can be further decomposed. Various tensor decompositions are available to model \mathcal{V} , e.g., the canonical decomposition

$$\tilde{v}_i(\tau_1, \tau_2, \dots, \tau_L) = v_{1,i}(\tau_1)v_{2,i}(\tau_2) \dots v_{N,i}(\tau_L), \quad (18)$$

where $\{v_{\ell,i}(\tau_\ell)\}_{i=1}^R$ spans the subspace for the ℓ th time dimension; or the Tucker decomposition

$$\tilde{v}_i(\tau_1, \tau_2, \dots, \tau_L) = \sum_{i_1=1}^{R_1} \sum_{i_2=1}^{R_2} \dots \sum_{i_L=1}^{R_L} c_{ii_1i_2\dots i_L} v_{1,i_1}(\tau_1)v_{2,i_2}(\tau_2) \dots v_{N,i_L}(\tau_L), \quad (19)$$

where $\{v_{\ell, i_\ell}(\tau_\ell)\}_{i_\ell=1}^{R_\ell}$ is the basis for the ℓ th time dimension. The core tensor $\mathcal{C} \in \mathbb{C}^{R \times R_1 \times \dots \times R_L}$ has elements $[\mathcal{C}]_{i_1 i_2 \dots i_L} = c_{i_1 i_2 \dots i_L}$ which determine the weight assigned to each basis function, similar to the role of the singular values in matrix decompositions. We also note the option of local [27] or patch-based tensor modeling [28], [29], which may have benefits related to nonlinear manifold modeling (which will be discussed further in Section VI). For a comprehensive description of differences between various tensor decompositions, the interested reader is referred to [26]. The rest of this section will assume global modeling using Eq. (19), which permits a lower R than the canonical decomposition and is the decomposition employed by both LRTES and by MR Multitasking. In global Tucker tensor form, the temporal model in Eqs. (14) and (19) becomes

$$\begin{aligned}\mathcal{V} &= \mathcal{C} \times_2 \mathbf{V}_1 \times_3 \mathbf{V}_2 \times_4 \dots \times_{L+1} \mathbf{V}_L \\ \mathcal{X} &= \mathcal{V} \times_1 \mathbf{U} = \mathcal{C} \times_1 \mathbf{U} \times_2 \mathbf{V}_1 \times_3 \mathbf{V}_2 \times_4 \dots \times_{L+1} \mathbf{V}_L,\end{aligned}$$

where the elements of $\mathbf{V}_\ell \in \mathbb{C}^{N_\ell \times R_\ell}$ are $[\mathbf{V}_\ell]_{ij} = v_{\ell, j}(\tau_{\ell, i})$.

There are various approaches to learning a LRT representation from sparsely sampled data. One approach is to implicitly impose the LRT model by penalizing the nuclear norm of each unfolding of \mathcal{X} , i.e.

$$\hat{\mathcal{X}} = \arg \min_{\mathcal{X}} \|\mathcal{A}(\mathcal{X}) - \mathbf{b}\|_2^2 + \lambda \sum_{i=1}^{L+1} \|\mathbf{X}_{(i)}\|_* \quad (20)$$

Similarly to Eq. (17), this addresses scan time by recovering an image tensor \mathcal{X} which is compressible (this time by tensor factorization), without explicitly storing \mathcal{X} in factorized form. To directly address the effect of curse of dimensionality on memory and storage requirements, it is also possible to recover the memory-efficient individual factors of the LRT model \mathbf{U} , \mathcal{C} and $\{\mathbf{V}_\ell\}_{\ell=1}^L$ without ever calculating their product \mathcal{X} , in the sense of

$$\{\hat{\mathcal{C}}, \hat{\mathbf{U}}, \{\hat{\mathbf{V}}_\ell\}_{\ell=1}^L\} = \arg \min_{\mathcal{C}, \mathbf{U}, \{\mathbf{V}_\ell\}_{\ell=1}^L} \|\mathcal{A}(\mathcal{C} \times_1 \mathbf{U} \times_2 \mathbf{V}_1 \times_3 \mathbf{V}_2 \times_4 \dots \times_{L+1} \mathbf{V}_L) - \mathbf{b}\|_2^2. \quad (21)$$

Rather than jointly optimizing the cost function in Eq. (21) over all variables, it is often convenient to first learn representations for the dynamic processes from subject-specific training data, as in the LRTES method. This method applies when $(\mathbf{k}, \tau_1, \tau_2, \dots, \tau_L)$ -space sampling can be prospectively controlled (e.g., when all of the $\{\tau_\ell(t)\}_{\ell=1}^L$ represent sequence parameters). In

this scenario, sampling is designed to collect L subsets of training data, the ℓ th set of which densely samples across τ_ℓ within a limited region of $(\mathbf{k}, \{\tau_q\}_{q \neq \ell})$ -space. Each \mathbf{V}_ℓ is then learned by computing the SVD of each training data subset, before jointly recovering \mathcal{C} and \mathbf{U} from the remaining $(\mathbf{k}, \tau_1, \tau_2, \dots, \tau_L)$ -space data.

MR Multitasking offers a solution for scenarios where $(\mathbf{k}, \tau_1, \tau_2, \dots, \tau_L)$ -space sampling cannot be fully controlled (e.g., when some of the $\{\tau_\ell(t)\}_{\ell=1}^L$ index physiological processes such as motion). There, self-gating data are used to learn $\{\tau_\ell(t)\}_{\ell=1}^L$ before being mapped into $(\mathbf{k}, \tau_1, \tau_2, \dots, \tau_L)$ -space to serve as subject-specific training data. The training data typically only sparsely sample $(\mathbf{k}, \tau_1, \tau_2, \dots, \tau_L)$ -space, so Multitasking performs temporal feature extraction by learning \mathcal{V} from the whole training data after they have been completed using nuclear norm constraints, similar to Eq. (20). Because the training data are only collected in a limited region of \mathbf{k} -space, the training data completion problem has far more modest memory and storage requirements than the entire image completion problem in Eq. (20). After having learned the feature space $\text{span}(\{\tilde{v}_i(\tau_1, \tau_2, \dots, \tau_L)\}_{i=1}^R)$, the image coordinates in this space (e.g., the \mathbf{U}) are then recovered identically to the subspace-constrained low-rank matrix image reconstruction problem with $v_i(t) = \tilde{v}_i(\tau_1(t), \tau_2(t), \dots, \tau_L(t))$.

V. EXPLICIT MOTION ESTIMATION AND COMPENSATION MODELS

Several methods based on compensating the inter-frame object and/or subject motion have been proposed to improve the performance of accelerated dynamic MRI models. The general idea is to jointly estimate the deformation parameters and the dynamic dataset by imposing compactness priors on the motion corrected time series rather than the original time series. An example optimization criterion for an explicit motion estimation, and compensated dynamic MRI recovery scheme can be written as [11], [12]:

$$\{\hat{\mathbf{X}}, \hat{\theta}(\mathbf{r}, t)\} = \arg \min_{\mathbf{X}, \theta(\mathbf{r}, t)} \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2^2 + \lambda_1 \underbrace{\psi(\mathcal{T}_\theta(\mathbf{X}))}_{\text{regularization on motion compensated time series}} + \lambda_2 \underbrace{\phi(\theta(\mathbf{r}, t))}_{\text{regularization on motion maps}}; \quad (22)$$

where, \mathcal{T}_θ denotes the deformation operator, parametrized by motion field parameters $\theta(\mathbf{r}, t)$. The motion field can be modeled as rigid body motion or non-rigid deformations. ϕ is a regularizer on the motion field such as a smooth spatial filter to penalize large unrealistic deformations.

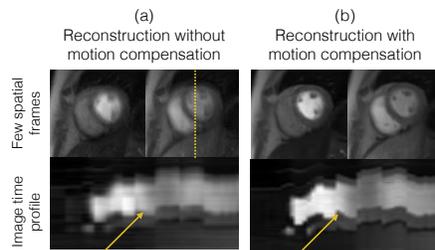


Fig. 4. Demonstration of explicit motion estimation and compensation models: This example shows a rate 4 prospectively acquired myocardial perfusion data with heavy breathing. Performing CS (via a patch based low rank regularizer) shows considerable motion artifacts in (a). With motion compensation, the reconstruction is robust to these artifacts and demonstrate improved temporal fidelity. Figure courtesy of Xiao Chen, Siemens Healthineers.

ψ specifies arbitrary choices of compactness priors on the motion compensated time series including the spatial-spectral sparsity prior, spatiotemporal finite difference sparsity prior, patch-based and/or global low rank priors.

The optimization in (22) is challenging to solve due its non-linearity, and non-convexity. Typical approaches include alternating between reconstruction and motion estimation by starting with reasonably good initial estimates of the reconstruction and the motion maps. In practice, strategies that correct for motion in a coarse to fine stage have shown to be robust to undesirable local minima. In a similar realm, other efficient methods based on variable splitting and continuation rules have been proposed to decouple the problem in (22) to simpler problems, and gradually update the complexity. Figure (4) shows an example motion compensated reconstruction scheme applied to the recovery of free breathing cardiac perfusion datasets from 4 fold under-sampled data. While the explicit motion estimation and compensation are extremely powerful in reducing motion artifacts, they are practically challenged by long computation times, non-convexity, tuning of several parameters (eg. reconstruction and motion estimation regularization parameters).

VI. MANIFOLD MODELS

The manifold structure of data has been widely used for the visualization of the structure of complex datasets. Non-linear dimensionality reduction or manifold embedding methods assume that the data are points on a low dimensional manifold (smooth surface) in higher dimensional space. If the manifold is of low-enough dimensionality, the data can be compactly visualized in low-dimensional space. For example, if the points are on a curve in 3-D, one can associate each point to a point on a straight line; non-linear dimensionality reduction methods such as ISO-MAP generates a nonlinear mapping between each point on the 3-D curve to a point on a

line [30]. Emerging research shows that the manifold structure of dynamic MRI data can be used to regularize the recovery of dynamic MRI data from under sampled measurements.

In applications such as free breathing and ungated cine MRI, the dynamic images can be assumed to non-linear functions of two parameters: the cardiac phase and respiratory phase. Images with similar cardiac and respiratory phase are expected to be similar; this property is used in gating methods. Hence, the images in these applications can be assumed to be points on a smooth low-dimensional manifold with high ambient dimension, where the dimension is the total number of pixels in each image. Figure (5) demonstrates this concept on a simulated free breathing cardiac cine phantom, where similar but distant time frames are mapped to lie in close proximity on the smooth manifold.

Manifold regularization exploits the similarity of image frames in terms of proximity of points on the smooth manifold. The reconstruction formulation can be written as [13]:

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2^2 + \lambda \sum_{i=1}^N \sum_{j=1}^N (\sqrt{W_{ij}} \|\mathbf{x}_i - \mathbf{x}_j\|_p)^p \quad (23)$$

where the matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ contains weighting factors that determine the degree of similarity between \mathbf{x}_i and \mathbf{x}_j , the i th and j th columns (time frames) of \mathbf{X} ; and $\|\cdot\|_p^p$ indicates the l_p norm; $p \leq 2$ for the image differences. These weights are inversely proportional to the distance between the corresponding points on the manifold. For a pairs of points closer to each other, W_{ij} has a high value, while for pairs of points that are farther from each other, W_{ij} has a low value. One strategy to estimate these weights is to do so from navigator signals $y(\mathbf{k}, t)$ as:

$$W_{ij} = \begin{cases} e^{-\frac{\|\mathbf{y}_i - \mathbf{y}_j\|_2^2}{\sigma^2}} & \text{if } \|\mathbf{y}_i - \mathbf{y}_j\|_2^2 < 2\sigma \\ 0 & \text{else} \end{cases} \quad (24)$$

where σ is a parameter that is dependent on the maximum curvature on the manifold. The acquisition scheme to collect the navigator data $y(\mathbf{k}, t)$ requires sampling of k-space at the same location for each time frame. Typically, the navigator data are interspersed with the measurement data simultaneously. For example, in free breathing cardiac cine MRI [13], 3–4 radial spokes are acquired as navigator k-space lines for each frame, while the remaining spokes are acquired according to a classical golden angle radial scheme. This can be extended using other navigator sampling schemes (e.g. with spirals).

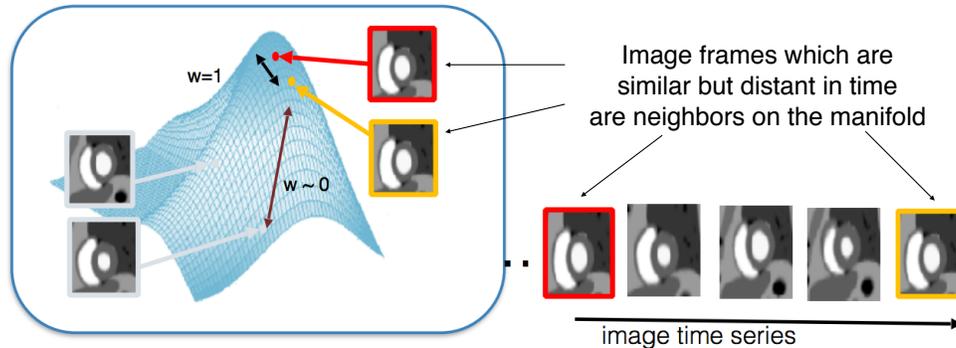


Fig. 5. Manifold structure of dynamic images. In this example, a free breathing ungated cardiac cine phantom is considered for demonstration. These image frames can be considered to live as points on a smooth 2D manifold in a higher dimensional space. Note, that in the sequence of image frames on the right, there exists frames that are similar but distant in time (i.e., of the similar cardiac and respiratory phase; see the red and yellow squares). On the manifold, these frames are mapped to points that are neighbors to each other. Similarly, dissimilar frames correspond to points that are far away from each other on the manifold. Weighting factors that determine the degree of similarity between a pair of frames are typically estimated via navigator signals, and the reconstruction is formulated as a penalized optimization (as in 23).

Other manifold regularizations have been proposed. For example, in localized linear embedding [31], each image (i.e., a point on a manifold) is closely approximated by the weighted combination of other images in the dataset (i.e., neighbors on the manifold). The weights are estimated from navigator data similar to STORM.

Inspired by the success of non-local means in denoising [32], patch based regularization for implicit motion compensated recovery has been proposed in dynamic MRI recovery [33]. Small spatial patches are defined in an image frame, and similar patches are searched in a localized spatiotemporal neighborhood in subsequent time frames. Redundancies amongst the similar patches are then exploited. The recovery can be posed as a regularized reconstruction scheme, where the non-local regularization penalty is an unweighted sum of distances between image patch pairs in the dynamic dataset.

Kernel based PCA approaches have been recently proposed which also fall under the umbrella of manifold regularization methods [14]. Kernel PCA maps the low-dimensional dynamic signal to a high dimensional feature space via a nonlinear mapping. Then, it exploits the low-rank structure in the high dimensional feature space, and lastly maps the reconstruction from the feature space to the signal space. Recently, STORM has been shown to be interpreted as a kernel PCA approach by exploiting a union of curves model [34].

VII. DEEP LEARNING MODELS

Deep learning, representation learning via multiple layers of hierarchical non-linear representation modules (e.g., neural networks), is a hugely active research area in many fields of science and engineering today [35]. Deep learning networks are capable of solving complex inverse problems non-iteratively (or rather, with all iterations isolated to an initial training phase); as such, they offer an avenue for nonlinear image reconstruction at unprecedented calculation speeds. Neural networks are known to be highly effective at learning representations of data which lie on non-linear manifolds [36], further making deep learning a promising fit for dynamic image reconstruction. As covered elsewhere in this special issue, most deep learning MR image reconstruction networks have been designed to recover static images or individual frames of dynamic images (e.g. [37]–[39]), thereby learning spatial representations. However, explicitly dynamic methods learning spatiotemporal representations have recently started to appear in the literature as well, e.g. [15]–[17], presenting an exciting new avenue for accelerated dynamic MRI.

Rather than iteratively inverting the forward problem $\mathcal{A}(\mathbf{X}) = \mathbf{b}$ to reconstruct $\hat{\mathbf{X}}$ for each individual subject, deep learning reconstruction methods learn the parameters $\hat{\theta}$ of a nonlinear forward reconstruction operator $f(\cdot; \hat{\theta})$ from a large training set. This training set $\mathcal{D}_{\text{train}}$ contains matched pairs of known ground truth images and their undersampled \mathbf{k} -space data (\mathbf{X}, \mathbf{b}) from multiple subjects, allowing the network to be trained according to the objective function:

$$\hat{\theta} = \arg \min_{\theta} \sum_{i | (\mathbf{X}_i, \mathbf{b}_i) \in \mathcal{D}_{\text{train}}} \mathcal{L}(\mathbf{X}_i, f(\mathbf{b}_i; \theta)), \quad (25)$$

where the loss function $\mathcal{L}(\mathbf{X}_i, f(\mathbf{b}_i; \theta))$ compares the i th reconstructed image $f(\mathbf{b}_i; \theta)$ to the corresponding labeled training image \mathbf{X}_i . In the context of dynamic MR image reconstruction, the L2 loss function has been the main loss function explored: $\hat{\theta} = \arg \min_{\theta} \sum_{i | (\mathbf{X}_i, \mathbf{b}_i) \in \mathcal{D}_{\text{train}}} \|\mathbf{X}_i - f(\mathbf{b}_i; \theta)\|_2^2$. It should be noted that in practice, \mathbf{b} is optionally transformed into image space by an initial pre-processing step such as the adjoint operation $\mathcal{A}^*(\mathbf{b})$ before being passed through the network. Here we absorb pre-processing steps into f as a means of simplifying notation.

After training by Eq. (25), future images can then be reconstructed by a single pass through the deployed network, $\hat{\mathbf{X}} = f(\mathbf{b}; \hat{\theta})$, or by performing data-consistent reconstruction incorporating

$f(\mathbf{b}; \hat{\theta})$ as a prior on $\hat{\mathbf{X}}$ [15], [40], as in

$$\hat{\mathbf{X}} = \arg \min_{\mathbf{X}} \|\mathcal{A}(\mathbf{X}) - \mathbf{b}\|_2^2 + \lambda \|\mathbf{X} - f(\mathbf{b}; \hat{\theta})\|_F^2. \quad (26)$$

A two-step process of 1) applying f to remove noise and artifacts then 2) enforcing data consistency via Eq. (26) mirrors individual iterations of optimization algorithms used for iterative nonlinear image reconstruction. As a result, deep learning reconstruction pipelines which repeat this two-step process multiple times in series are analogous to “unrolling” iterative algorithms.

Deep learning networks for imaging are generally structured so that early layers of f encode the input within the nonlinear feature space learned during training, and that later layers subsequently decode the desired $\hat{\mathbf{X}}$. In the context of dynamic MRI, this can be understood as replacing handcrafted model choices (e.g., whether \mathbf{U} is sparse, what size of \mathbf{V} , whether to perform local or global modeling) to some degree with automatic modeling. These choices do not entirely disappear, as they are reflected in the choice of specific network architecture; for example, the hierarchy of layers are usually designed to permit multiscale modeling, but the specific $\hat{\theta}$ learned during training would ultimately determine the balance between global and local modeling. Most deep learning methods for dynamic MRI to date have been based on convolutional neural network (CNN) architectures, such as those proposed by Schlemper et al. as a cascaded CNN with data sharing from previous time points [15], by Qin et al. in a recurrent CNN architecture [16], and by Biswas et al. [17] as an unfolded CNN with an additional spatiotemporal manifold constraint added to Eq. (26).

VIII. OUTLOOK

Representation learning has already made important strides in accelerating dynamic MRI; with continued development, it has the potential to change the very structure of clinical MRI exams. Today, MRI exams comprise a series of different scans, each of which is acquired with a different contrast weighting or which interrogates a different dynamic process. All of these scans are reconstructed independently and are typically stored for later analysis.

Multidynamic imaging has the potential to replace this series of independent scans with comprehensive, integrated single-scan examinations. Currently, the time between scans is used for technologists to make imaging decisions such as contrast-weighting selection or ECG trigger

window definition, as well as for patients to recover from breath holds. Imaging multiple contrast weightings for a variety of motion states would shift these imaging decisions to after image reconstruction and would eliminate breath holds. This would go beyond simply reducing the amount of data being acquired, by also eliminating delays *between* acquisitions. A motion-resolved, multicontrast paradigm would also be a major step toward a fully-quantitative multiparameter exam for objective, comprehensive tissue characterization. There are many signal processing opportunities along the way towards making this new paradigm a reality. Multidynamic CS has so far only explored the use of temporal finite difference sparsifying transforms (i.e., temporal TV), and LRT imaging has primarily been limited to the Tucker decomposition. Major questions in optimal sampling design remain, particularly on how to prospectively design sampling when physiological timings are unpredictable, i.e., when the patient’s physiology partially controls the sampling pattern. There are also open opportunities to extend other decompositions and models to their multidynamic counterparts, e.g., low-rank+sparse [41] tensor models, multidynamic BCS models, and multidynamic manifold models. The richness (and massive size) of data will also require faster image reconstruction and will necessitate new modes of viewing and analyzing high-dimensional images, all of which are potential areas to incorporate more advanced representation learning approaches such as deep learning.

As the newest class of methods described here, deep learning has perhaps the largest number of open technical problems and opportunities. Deep learning in dynamic MRI is currently dependent on supervised learning with huge sets of highly-representative training data, which limits the ability to enable new modes of imaging where labels are not yet available. This suggests the importance of developing unsupervised reconstruction networks or methods of generating realistic, representative training data (e.g., self-supervised learning where training images are generated by a secondary network). There are opportunities to explore more advanced loss functions such as with generative adversarial networks wherein \mathcal{L} itself output by a network, as well as opportunities to go beyond (\mathbf{r}, t) -space networks and explore (\mathbf{k}, t) -space or hybrid (\mathbf{k}, t) -to- (\mathbf{r}, t) networks. The current reliance of deep learning on GPU training, where there are stricter memory limitations than CPU processing presents challenges for reconstructing long image sequences or dynamic images with three spatial dimensions, and creates further opportunity to explore low-resource architectures. The excellent performance of deep networks for image analysis tasks, although not

covered in this article, also suggest the opportunity to build combined reconstruction and analysis networks which can directly make clinically useful inferences directly from input (\mathbf{k}, t) -space data, changing the imaging pipeline entirely.

REFERENCES

- [1] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magn Reson Med*, vol. 58, no. 6, pp. 1182–1195, Dec 2007.
- [2] H. Jung, K. Sung, K. Nayak, E. Kim, and J. Ye, "k-t FOCUSS: A general compressed sensing framework for high resolution dynamic MRI," *Magn Reson Med*, vol. 61, no. 1, pp. 103–116, Jan 2008.
- [3] Z.-P. Liang, "Spatiotemporal imaging with partially separable functions," in *Proc IEEE Int Symp Biomed Imaging*, 2007, pp. 988–991.
- [4] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Kim, "k-t PCA: Temporally constrained k-t BLAST reconstruction using principal component analysis," *Magn Reson Med*, vol. 62, no. 3, pp. 706–716, Sep 2009.
- [5] S. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans Med Imaging*, vol. 30, no. 5, pp. 1042–1054, May 2011.
- [6] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic mri," *IEEE transactions on medical imaging*, vol. 32, no. 6, pp. 1132–1145, 2013.
- [7] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "XD-GRASP: Golden-angle radial MRI with reconstruction of extra motion-state dimensions using compressed sensing," *Magn Reson Med*, vol. 75, no. 2, pp. 775–88, Feb 2016.
- [8] J. He, Q. Liu, A. G. Christodoulou, C. Ma, F. Lam, and Z.-P. Liang, "Accelerated high-dimensional MR imaging with sparse sampling using low-rank tensors," *IEEE Trans Med Imaging*, vol. 35, no. 9, pp. 2119–29, Sep 2016.
- [9] J. Y. Cheng, T. Zhang, M. T. Alley, M. Uecker, M. Lustig, J. M. Pauly, and S. S. Vasanawala, "Comprehensive multi-dimensional MRI for the simultaneous assessment of cardiopulmonary anatomy and physiology," *Sci Rep*, vol. 7, no. 1, p. 5330, Jul 2017.
- [10] A. G. Christodoulou, J. L. Shaw, C. Nguyen, Q. Yang, Y. Xie, N. Wang, and D. Li, "Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging," *Nature Biomed Eng*, vol. 2, no. 4, pp. 215–226, Apr 2018.
- [11] X. Chen, M. Salerno, Y. Yang, and F. H. Epstein, "Motion-compensated compressed sensing for dynamic contrast-enhanced MRI using regional spatiotemporal sparsity and region tracking: Block low-rank sparsity with motion-guidance (BLOSM)," *Magn Reson Med*, vol. 72, no. 4, pp. 1028–1038, 2014.
- [12] S. Goud Lingala, E. DiBella, and M. Jacob, "Deformation corrected compressed sensing (DC-CS): a novel framework for accelerated dynamic MRI," *IEEE Trans Med Imaging*, vol. 34, no. 1, pp. 72–85, 2015.
- [13] S. Poddar and M. Jacob, "Dynamic MRI using smoothness regularization on manifolds (SToRM)," *IEEE Trans Med Imaging*, vol. 35, no. 4, pp. 1106–1115, 2016.
- [14] 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 Med Imaging*, vol. 36, no. 11, pp. 2297–2307, 2017.
- [15] 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 Med Imaging*, vol. 37, no. 2, pp. 491–503, Feb 2018.
- [16] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, "Convolutional recurrent neural networks for dynamic MR image reconstruction," *IEEE Trans Med Imaging*, vol. 38, no. 1, pp. 280–290, Jan 2019.
- [17] S. Biswas, H. K. Aggarwal, and M. Jacob, "Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM," *Magn Reson Med*, 2019, in press.
- [18] Z.-P. Liang, H. Jiang, C. Hess, and P. Lauterbur, "Dynamic imaging by model estimation," *Int J Imaging Syst Technol*, vol. 8, no. 6, pp. 551–557, 1997.
- [19] N. Aggarwal and Y. Bresler, "Patient-adapted reconstruction and acquisition dynamic imaging method (paradigm) for MRI," *Inverse Probl*, vol. 24, no. 4, p. 045015, 2008.
- [20] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. S negas, and A. Mertins, "Compressed sensing reconstruction for magnetic resonance parameter mapping," *Magn Reson Med*, vol. 64, no. 4, pp. 1114–1120, 2010.
- [21] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t PCA: Temporally constrained k-t BLAST reconstruction using principal component analysis," *Magn Reson Med*, vol. 62, no. 3, pp. 706–716, 2009.
- [22] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled (\mathbf{k}, t) -space data with joint partial separability and sparsity constraints," *IEEE Trans Med Imaging*, vol. 31, no. 9, pp. 1809–1820, 2012.

- [23] Y. Wang and L. Ying, "Compressed sensing dynamic cardiac cine MRI using learned spatiotemporal dictionary," *IEEE Trans Biomed Eng*, vol. 61, no. 4, pp. 1109–1120, 2013.
- [24] T. M. Quan and W.-K. Jeong, "Compressed sensing dynamic mri reconstruction using gpu-accelerated 3d convolutional sparse coding," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2016, pp. 484–492.
- [25] D. L. Donoho, "High-dimensional data analysis: The curses and blessings of dimensionality," in *AMS Math Chall 21st Century*, 2000.
- [26] T. G. Kolda and B. W. Bader, "Tensor decompositions and applications," *SIAM Rev*, vol. 51, no. 3, pp. 455–500, 2009.
- [27] B. Yaman, S. Weingärtner, N. Kargas, N. D. Sidiropoulos, and M. Akçakaya, "Locally low-rank tensor regularization for high-resolution quantitative dynamic MRI," in *IEEE Int Workshop Comput Adv Multisens Adapt Process*, 2017.
- [28] M. Akçakaya, T. A. Basha, B. Goddu, L. A. Goepfert, K. V. Kissinger, V. Tarokh, W. J. Manning, and R. Nezafat, "Low-dimensional-structure self-learning and thresholding: regularization beyond compressed sensing for MRI reconstruction," *Magn Reson Med*, vol. 66, no. 3, pp. 756–67, Sep 2011.
- [29] A. Bustin, G. Lima da Cruz, O. Jaubert, K. Lopez, R. M. Botnar, and C. Prieto, "High-dimensionality undersampled patch-based reconstruction (HD-PROST) for accelerated multi-contrast MRI," *Magn Reson Med*, vol. 81, no. 6, pp. 3705–3719, Jun 2019.
- [30] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.
- [31] U. Nakarmi, K. Slavakis, and L. Ying, "MLS: Joint manifold-learning and sparsity-aware framework for highly accelerated dynamic magnetic resonance imaging," in *Proc IEEE Int Symp Biomed Imaging*, 2018, pp. 1213–1216.
- [32] M. Maggioni, G. Boracchi, A. Foi, and K. Egiazarian, "Video denoising, deblocking, and enhancement through separable 4-D nonlocal spatiotemporal transforms," *IEEE Trans Image Process*, vol. 21, no. 9, pp. 3952–3966, 2012.
- [33] Y. Q. Mohsin, S. G. Lingala, E. DiBella, and M. Jacob, "Accelerated dynamic MRI using patch regularization for implicit motion compensation," *Magn Reson Med*, vol. 77, no. 3, pp. 1238–1248, 2017.
- [34] S. Poddar, Y. Mohsin, D. Ansah, B. Thattaliyath, R. Ashwath, and M. Jacob, "Manifold recovery using kernel low-rank regularization: Application to dynamic imaging," *IEEE Trans Comput Imaging*, 2019, in press.
- [35] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–44, May 2015.
- [36] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans Pattern Anal Mach Intell*, vol. 35, no. 8, pp. 1798–828, Aug 2013.
- [37] S. Wang, Z. Su, L. Ying, X. Peng, S. Zhu, F. Liang, D. Feng, and D. Liang, "Accelerating magnetic resonance imaging via deep learning," in *Proc IEEE Int Symp Biomed Imaging*, 2016, pp. 514–517.
- [38] B. Zhu, J. Z. Liu, S. F. Cauley, B. R. Rosen, and M. S. Rosen, "Image reconstruction by domain-transform manifold learning," *Nature*, vol. 555, no. 7697, pp. 487–492, Mar 2018.
- [39] 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," *Magn Reson Med*, vol. 79, no. 6, pp. 3055–3071, Jun 2018.
- [40] H. K. Aggarwal, M. P. Mani, and M. Jacob, "MoDL: Model-based deep learning architecture for inverse problems," *IEEE Trans Med Imaging*, vol. 38, no. 2, pp. 394–405, Feb 2019.
- [41] R. Ramb, M. Zenge, L. Feng, M. Muckley, C. Forman, L. Axel, D. Sodickson, and R. Otazo, "Low-rank plus sparse tensor reconstruction for high-dimensional cardiac MRI," in *Proc Int Soc Magn Reson Med*, vol. 25, 2017, p. 1199.