Dictionary, Structured Low-Rank, and Manifold Learning-Based Reconstruction

10

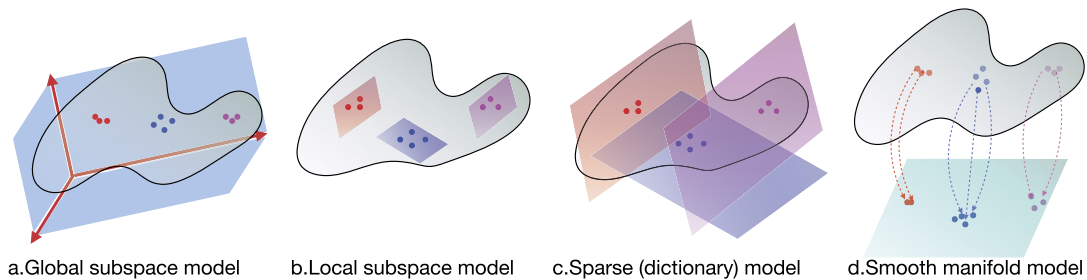**Mathews Jacob, Sajan Goud Lingala, and Merry Mani**
University of Iowa, Iowa City, IA, United States

10.1 Introduction

Efficient image representations are key to reconstructing images from fewer measurements. In particular, low-dimensional representations can represent images efficiently, enabling the robust recovery from sparse and noisy measurements. Many of the early model-based MRI approaches relied on fixed image representations that were often carefully engineered to the data. In contrast, learning the low-dimensional representations from data itself offers improved efficiency. For instance, the compressed sensing methods described in Chapter 8 may be thought of as learning-based algorithms, where the specific basis functions that are best suited to represent a given signal are chosen from a pre-engineered dictionary. The low-rank models reviewed in Chapter 9 go one step further, for instance, to make use of the redundancies of images in a time series that differ in contrasts and/or motion states. This chapter focus on *learned low-dimensional representations* that are generalizations of the low-rank methods discussed in the previous chapter. The objective is to account for image redundancies that are challenging for low-rank methods to capture.

In this chapter, we will review the generalizations of low-rank methods, which can be broadly classified as

1. approaches that capitalize the complex *nonlinear redundancies* with the datasets, which low-rank methods are not capable of capitalizing on. Sparse dictionary learning and smooth manifold representations fall into this category. Unlike low-rank methods that use energy based priors on the factors, dictionary learning methods use sparsity priors on the coefficients and energy-based priors on the dictionary atoms. The smooth manifold models use nonlinear kernel priors, allowing us to account for nonlinear redundancies in the data resulting from motion and contrast changes. Please see Fig. 10.1 for an illustration of the representations in 3D space.

2. approaches that use the above representations to *sub-parts of the image* (e.g., patches of different shape, either in the image domain or k-space) rather than the whole image. We will show that this approach enables one to capitalize on unique signal redundancies that are characteristic of each application (e.g., smoothly varying phase, exponential signal decay in time, uncalibrated parallel MRI), which are often challenging for the traditional low-rank or compressed sensing algorithms. Please see Fig. 10.2 for an illustration of the signals extracted from the images, whose redundancies are leveraged to recover the image.

a.Global subspace model   b.Local subspace model   c.Sparse (dictionary) model   d.Smooth manifold model
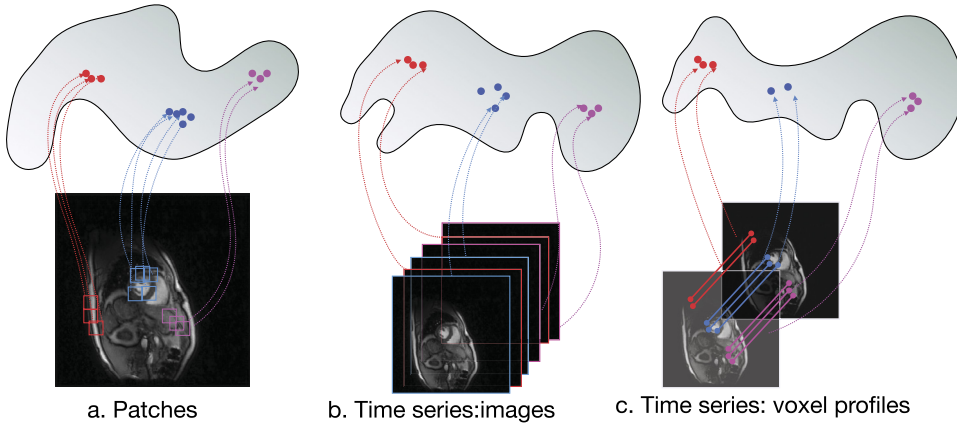
FIGURE 10.1

Overview of signals extracted from the image volumes, whose redundancy is capitalized on in this chapter: The signals (images/patches/timeprofiles) in MRI datasets have extensive redundancy as described in Fig. 10.2; they can be viewed as points lying on a manifold in high-dimensional spaces. The redundancy of the signals are captured by different adaptive models in different ways. Global subspace or global principal component analysis (PCA) models reviewed in Chapter 9 model the signals as a subspace as shown in (**a**); they learn the basis vectors indicated by the red arrows from the data, coming up with a compact representation of the space. Local subspace models cluster the data and learn a subspace for each cluster/neighborhood as shown in (**b**). Sparse dictionary learning methods learn the dictionary basis functions from the data itself; the signal space is modeled by a union of subspaces, where basis functions may be shared by local neighborhoods as shown in (**c**). The main benefit over local PCA approaches is that this approach does not need an explicit clustering step. Smooth manifold and kernel PCA models represent the signals as a smooth manifold; as shown in (**d**), they rely on a mapping that converts the nonlinear manifold to a low-dimensional subspace denoted by the plane; the structure of the low-dimensional subspace is used to recover the signals.

The chapter is organized as follows. Following a brief overview of the background in Section 10.2, we review dictionary-based methods in Section 10.3, where the sparsity of the dictionary coefficients is used to further improve the adaptation of the representation to the specific signal; these approaches use sparsity to bypass the need for clustering that is often needed in the local PCA-based methods reviewed in Chapter 9. The low-rank model in Chapter 9 is then extended to patches in image domain or k-space in Section 10.4. In particular, under specific assumptions, the resulting patch matrix/matrices are highly low-rank, which can be capitalized for acceleration as discussed in Section 10.4. We review smooth manifold models that are efficient in capturing nonlinear redundancies in the dataset in Section 10.5. An overview and broad classification of the methods reviewed in this chapter is given in Table 10.1. In this chapter, the various topics are not presented in the chronological order in which they were introduced in the MRI setting. Rather, our focus is on grouping the various methods into broader themes to facilitate easy comprehension of the links and the generalizations of the various approaches.

10.2 Background
10.2.1 Acquisition scheme

As introduced in Chapter 2, the main goal of image recovery is the estimation of the continuous domain function $\mathbf{x} : \mathbb{R}^n \to \mathbb{C}$ from a finite number of multichannel k-space measurements $s_{i,j}$, based on the

a. Patches   b. Time series:images   c. Time series: voxel profiles

FIGURE 10.2

Types of signals used by different algorithms: The MR methods reviewed in this chapter account for the re-dundancy within the image volumes, which often manifest as correlations between image sub-parts that we broadly refer to as patches. Depending upon the shape of the patches, they could be cubes in 3D/4D as shown in (**a**), images in a dynamic time series as shown in (**b**), or time profiles of pixels in a time series as shown in (**c**). The images in dynamic imaging or parameter mapping have extensive nonlocal similarity as shown in (**c**). For instance, the images in similar cardiac/respiratory phases are expected to be similar; each image in the time series may be viewed as a mapping of the cardiac/respiratory phases, which are often ac-counted by self-gating methods. Likewise, the voxel time series in image time series in (c) are also highly correlated. For instance, the time series of pixels from the same organ that experience similar motion pattern or have similar physiology (e.g., myocardium, liver) are expected to have similar intensity profiles. All of these schemes can be seen as patch based methods; the main difference is the shape of the patches. Patch based methods aim to capture the extensive similarity between patches to recover the image dataset from highly under sampled measurements. If the number of pixels in a patch is denoted by p, each of the patches can be viewed as a point in a p dimensional space. However, because of the extensive structure/redundancy between the pixel values, these signals are often localized to low-dimensional structures in this high-dimensional space as shown above.

relationship

$$s_{i,j} = \int x(\mathbf{r})c_j(\mathbf{r})\exp(-j\mathbf{k}_i^T\mathbf{r})d\mathbf{r} + \eta_{i,j}. \tag{10.1}$$

Here, $c_j(\mathbf{r})$ are the sensitivity profiles of the jth coil, and $\mathbf{k}_i$ is the location in k-space, while η denotes noise. The measurement process described in (10.1) can be compactly represented as

$$\mathbf{s} = \mathbf{E}(\mathbf{x}), \tag{10.2}$$

where $\mathbf{E}$ is the multichannel Fourier encoding operator. The function $\mathbf{x}$ may be 2D ($n = 2$) or 3D ($n = 3$), or higher-dimensional, depending on the applications. In 2D/3D + time datasets (e.g., dynamic

Table 10.1 Broad classification of methods reviewed in this chapter.

Method	Signal Type			
	Patch		Images	
	Two-step	Joint	Two-step	Joint
Low-rank image domain	BM3D [16,17]	CLEAR [19], LOST [2], HD-PROST [13] PRICE [42]	- PSF [11,70,31,30]	- k-t SLR [19,33,32]
Low-rank k-space	ESPIRIT [42] GRAPPA [20] SLR [49]	LORAKS [21,23,22] SLR [51,50,48,49] MUSSELS [41,39,40,37]	- - - -	- ALOHA [29,26,25] SLR [4,5]
Dictionary learning	Transform Learning [58,57,60]	BCS [59]	- - -	- BCS [9,8,34,35]
Manifold smoothness manifold	NLM [12] UINTA [3]	Nonlocal Regularization [68,67] Patch-STORM [43]	Kernel PCA [47,66] STORM [54] MLS [64,46] KLR-STORM [56]	Spiral-STORM [1]

imaging applications), the acquisition model correspond to

$$s_{i,j}(t) = \int x(\mathbf{r}, t) \, c_j(\mathbf{r}) \exp(-j\mathbf{k}_i^T \mathbf{r}) d\mathbf{r}. \tag{10.3}$$

In this chapter, we will use the same symbol $\mathbf{E}$ as the multichannel Fourier sampling operator in the dynamic setting. Specifically, $\mathbf{E}$ applied on the 3D/4D volume image $\mathbf{x}$ yields the vector of measurements denoted by $\mathbf{s}$. We denote $\widehat{\mathbf{x}}$ to denote the discrete Fourier coefficients of the signal on a Cartesian grid. Note that the multichannel measurements in (10.1) need not be in the Cartesian domain.

10.2.2 Manifold models of signals

We start with a brief and intuitive illustration of the manifold assumption, with the objective of connecting diverse image models used in the context of MRI. We note that an $n \times n$ image has n^2 pixels and hence can be viewed as a point in n^2 dimensional space. In the absence of any redundancies (e.g., pixels have random values), each image is a random point; the images will fill the n^2 dimensional space. However, natural images have extensive redundancies between the pixels. Hence, the images of interest often lie on low-dimensional structures (e.g., lines, union of subspaces, curves) in n^2 dimensional space.

For example, consider three pixel images whose pixel values are uniformly random. If we plot each of these three pixel images in 3D space, they fill the space. By contrast, if all of the images have their pixel values are linearly increasing with the same slope, all the three pixel images will fall on a straight line passing through the origin. When modeling larger 1D images, one may extract patches consisting of three consecutive pixels from the image, each of which can be viewed as a point on the slanted line. This approach can be viewed as a global subspace model for the image. As discussed in the next section, the patch signals can be collected into a $3 \times P$ matrix, whose columns are the patches. The cited global linear model is restrictive and may not approximate real-world 1D signals. A more

general representation is a piecewise linear model, where the slope of the signal is different at different locations. In this case, the three pixel patches from each image may not lie on a single line. Depending on the spatial location of the patch, they would lie on different straight lines; the number of straight lines would depend on the number of piecewise linear regions. This union of lines/subspaces model can be viewed as a generalization of the aforementioned global subspace model.

This representation can be generalized in many different ways to improve the efficiency and approximation power. For instance, one can use piecewise polynomial or exponential signal models, which are more efficient than the piecewise linear model. Note that the dimension of the space depends on the number of pixels in the patch (m). With higher dimensions, one can capitalize on more complex interdependencies between the pixels, beyond the piecewise models just discussed. In general, the patch signals do not fill the m-dimensional space (m being the number of pixels in the patch); they often lie on low-dimensional constructs (e.g., clusters, smooth surfaces, curves), often loosely termed as manifolds, in the m-dimensional space. This property is often referred to as the manifold assumption in machine learning [7,53].

The aforementioned idea can be extended to patches in 2D images, where the pixels within the patches may have nonlinear relationships between them, depending on the type of the image content (e.g., piecewise linear, piecewise polynomial). Depending on the image content, one may consider patches with m pixels of different sizes and shapes (see Fig. 10.2) to capture specific redundancies within the dataset. Generalizing the previous example, the patches can be viewed as points in high-dimensional space. This chapter reviews the several approaches of *learning the compact representation of data matrices*, whose columns are the signals of interest (images, patches, time-profiles of pixels in time series data) that we will capitalize on using advanced algorithms to capitalize on the unique redundancies in each application.

10.2.3 Capitalization of redundancy using structured matrices

A common approach to capitalize the redundancies within the signals of interest (e.g., images in a time-series, patches in an image, voxel profiles) is to create structured matrices from the data and use their properties to recover the images. In the general setting, one can extract patch vectors from images to form a matrix denoted by $\mathcal{T}(\mathbf{x})$:

$$\mathcal{T}(\mathbf{x}) = \begin{bmatrix} P_1(\mathbf{x}) & P_2(\mathbf{x}) & \dots & P_n(\mathbf{x}) \end{bmatrix}. \tag{10.4}$$

Here, $P_1, .., P_n$ are patch extraction operators that extract a patch from an image and convert it into a column of $\mathcal{T}(\mathbf{x})$. If each patch has m pixels and there are n patches that cover the dataset, the data matrix denoted by $\mathcal{T}(\mathbf{x})$ will be of dimension $m \times n$. The size and shape of the patches could be chosen, depending on the application, to exploit a specific property of the dataset. For instance, if the ith patch extraction operator is chosen as the ith image in a time series $P_i(x(\mathbf{r}, t)) = x(\mathbf{r}, i); i = 1, .., T$, we obtain

$$\mathcal{T}(\mathbf{x}) = \begin{bmatrix} x(\mathbf{r}, 1) & \dots & x(\mathbf{r}, T) \end{bmatrix}, \tag{10.5}$$

which is the standard Casorati matrix reviewed in Chapter 9, whose columns are the reshaped images of the time series from the dynamic data (see Fig. 10.2.(a)). Similarly, if the patch extraction operators extract the time profile of each pixel, we obtain the transpose of a Casorati matrix. These are the two extreme cases. One can choose 2D patch-extraction operators (or 3D patch-extraction operators in the

time series) to account for the correlation between patches in the dataset. In these cases, the matrix can have a block convolutional structure. See Fig. 10.6. Hence, the methods that use the low-rank property of these structured matrices are called structured low-rank (SLR) methods [65,21,38,49,23,29,26,51, 41]. One may also create a structured matrix $\mathcal{T}(\hat{\mathbf{x}})$ by choosing the patches from the discrete Fourier samples of the signal $\hat{\mathbf{x}}$. In fact, several of the structured low-rank algorithms reviewed later in the chapter rely on the low-rank property of structured matrices in the Fourier domain.

The structured matrix $\mathcal{T}(\mathbf{x})$ is often much larger in size than the original dataset $\mathbf{x}$; the operation $\mathbf{x} \to \mathcal{T}(\mathbf{x})$ of creating the structured matrix from the samples is often called as a lifting operation. We term the columns of $\mathcal{T}(\mathbf{x})$ as the signals of interest; the algorithms considered in this chapter will promote the learning and capitalization of redundancies between the columns.

10.2.4 Efficient matrix representation in terms of factors

As discussed in Chapter 9 in the context of low-rank representation, the matrix $\mathcal{T}(\mathbf{x})$ can be efficiently represented in terms of its factors as

$$\mathcal{T}(\mathbf{x}) = \mathbf{\Phi}\mathbf{W}^T, \tag{10.6}$$

where $\mathbf{\Phi}$ and $\mathbf{W}$ are the factor matrices, of size $m \times R$ and $n \times R$, respectively. In the context of low-rank matrices, R is the rank of the matrix. When the data has high redundancy, R is much smaller than m and n. Most of the algorithms choose an $R > r$, coupled with priors (e.g., ℓ_2 or ℓ_1 norms) on the factors discussed subsequently to make the recovery well-posed.

In this case, the number of free parameters in $\mathbf{\Phi}$ and $\mathbf{W}$ is often much smaller than the size of $\mathcal{T}(\mathbf{x})$. When $\mathcal{T}(\mathbf{x})$ is the Casorati matrix, the columns of $\mathbf{\Phi}$ can be viewed as the spatial factor, while that of $\mathbf{W}$ is the temporal factor. In addition to enabling the recovery from undersampled data in terms of the spatial and temporal factors, the example factor representation can also mitigate the high memory demands of directly working with $\mathcal{T}(\mathbf{x})$, by conserving space and obtain a computationally efficient algorithms. Implicit low-rank methods use the nuclear norm of $\mathcal{T}(\mathbf{x})$, denoted by $\|\mathcal{T}(\mathbf{x})\|_* = \sum_i \sigma_i [\mathcal{T}(\mathbf{x})]$, as a prior in reconstruction problems,

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \lambda \|\mathcal{T}(\mathbf{x})\|_* \tag{10.7}$$

to encourage the recovery of an $\mathbf{x}$ such that the matrix $\mathcal{T}(\mathbf{x})$ is low-rank. The nuclear norm has an alternate form [61]

$$\|\mathcal{T}(\mathbf{x})\|_* = \|\mathbf{\Phi}\|_F^2 + \|\mathbf{W}\|_F^2, \text{ where } \mathcal{T}(\mathbf{x}) = \mathbf{\Phi}\mathbf{W}^T, \tag{10.8}$$

provided R is greater than the rank of the matrix $\mathcal{T}(\mathbf{x})$. This interpretation allows one to implement an implicit low-rank method without storing the large $\mathcal{T}(\mathbf{x})$ matrix

$$\{\mathbf{\Phi}^*, \mathbf{W}^*\}, = \arg\min_{\mathbf{\Phi}, \mathbf{W}} \|\mathbf{E}(\mathbf{\Phi}\mathbf{W}^T) - \mathbf{s}\|^2 + \lambda \left(\|\mathbf{\Phi}\|_F^2 + \|\mathbf{W}\|_F^2 \right), \tag{10.9}$$

when $\mathcal{T}(\mathbf{x})$ has a Casorati form. In addition to the computational and memory efficiency, the factor interpretation opens the door to the use of other priors on the factors and the matrices, which can offer improved performance. We will now focus on how more general factorization strategies can offer improved performance, compared to the low-rank methods already mentioned.

10.3 **Dictionary learning and blind compressed sensing**

The global subspace models described in Chapter 9 enables the representation of dynamic datasets. However, when the signals that one is trying to represent is very diverse (e.g., patches in the image), the ability of the global subspace model to represent them is limited (see Fig. 10.1.a). For instance, there may be several groups of patches, each of which may possess a low-rank. However, the global subspace spanned by all the groups may be high-dimensional. Hence, state-of-the art patch-based methods such as BM3D [17,16] cluster the patches into subsets, followed by the application of the subspace model to each subset. These schemes can be viewed as the approximation of the global manifold locally by low-dimensional subspaces, as shown in Fig. 10.1.(b). A challenge for these schemes is the two-step process, involving the identification of the similar subsets, followed by low-rank modeling. These approaches are widely used in the denoising setting. Approaches, such as [2,67–69], extend this approach to image-reconstruction applications. These methods either estimate the clusters from zero-filled MRI data [2,67] or alternate between clustering and recovery of images [68,69]. More information on the alternating scheme is provided in Section 10.5.1.2 in the context of manifold recovery. Dictionary learning and blind compressing can overcome the need for this pre-clustering for the recovery of images from undersampled data.

10.3.1 **Subspace selection for each signal of interest using sparse representation**

Both dictionary learning and blind compressing schemes rely on a sparse image representation, as shown in Fig. 10.1.(c). The sparse model allows one to choose the specific basis functions needed to represent a specific column of $\mathcal{T}(\mathbf{x})$ (patch) only by allowing only a few coefficients of the representation to be nonzero. For instance, the sparse model [15] represents the signal as

$$x(\mathbf{r}) \approx \sum_{i=1}^{K} w_i \, \varphi_i(\mathbf{r}) = \underbrace{\begin{bmatrix} \varphi_1(\mathbf{r}) & \varphi_2(\mathbf{r}) & \dots \varphi_K(\mathbf{r}) \end{bmatrix}}_{\mathbf{\Phi}} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix}}_{\mathbf{w}^T}, \text{ where } \|\mathbf{w}\|_{\ell_0} \le k. \tag{10.10}$$

The matrix $\mathbf{\Phi}$ is termed as the dictionary, while its columns are the basis vectors and are often referred to as atoms. Here, $\|\mathbf{w}\|_{\ell_0}$ denotes the number of nonzero terms in the coefficient vector $\mathbf{w}$. This model allows the basis functions used to approximate each group of signals to be different and, hence, offer more compact representation of the data. Traditional compressed sensing schemes relies on pre-determined dictionaries (e.g., wavelet transform) $\mathbf{\Phi}$. Rather than using fixed dictionaries, several authors have proposed to adapt or learn the dictionaries or transforms from the data [57,34,35,58,59,8]. The adaptation of the dictionary to the data depending on the specific signal offers a quite significant reduction in the number of measurements. The learning is either performed from several fully-sampled example images [57,58,8,60] or from a single undersampled dataset [34,35,59,8] in a joint manner. The first approach is termed as pre-learning. By contrast, the joint learning of the dictionary and coefficients from the data is termed blind compressed sensing. Since the ℓ_0 norm is not convex, a common approach is to approximate it with the ℓ_1 norm that is convex [15]. Rather than employing dictionaries, the use of analysis operators $\mathbf{\Psi}$ is also a common approach.

10.3.2 Dictionary pre-learning

In pre-learning, the learning of the dictionary from a family of fully sampled signals $\mathbf{X} = [\mathbf{x}_1, .., \mathbf{x}_N]$ is posed as the optimization problem [57,58]

$$\{\mathbf{\Phi}, \mathbf{W}\} = \arg \min_{\Phi, \mathbf{W}} \|\mathbf{X} - \mathbf{\Phi}\mathbf{W}^T\|^2 + \lambda_1 \|\mathbf{W}\|_{\ell_0} + \lambda_2 \ \mathcal{R}(\mathbf{\Phi}), \tag{10.11}$$

where $R(\mathbf{\Phi})$ is a regularization penalty on the dictionary atoms. A simple choice for $\mathcal{R}(\mathbf{\Phi}) = \|\mathbf{\Phi}\|_F^2$ is the Frobenius norm, where the energy of the dictionary is restricted. Regularization penalties that encourage the dictionary to be an orthogonal transform to have a low condition number or to be a combination of orthogonal transforms have been introduced by several authors [57,34,35,58,59,8].

Note that solving for $\mathbf{W}$ from (10.11) assuming $\mathbf{\Phi}$ to be known is the traditional compressed sensing approach. Likewise, for many of the common choices of $\mathcal{R}$, the optimization of $\mathbf{\Phi}$ assuming $\mathbf{W}$ to be known, is also a simple problem. For instance, when $\mathcal{R}(\mathbf{\Phi}) = \|\mathbf{\Phi}\|_F^2$ is the Frobenius norm, the solution of $\mathbf{\Phi}$ is a quadratic problem that has an analytical solution. However, the joint optimization of $\mathbf{\Phi}$ and $\mathbf{W}$ is a nonconvex problem. Nevertheless, this problem has been well-studied by several researchers [57,34,35,58,59,8], especially when $\mathbf{\Phi}$ is an orthonormal matrix. The learning of the transform to the class of signals results in improved performance over the use of standard transforms such as wavelet transform.

#### 10.3.2.1 Dictionary pre-learning, applied to static MRI

In static imaging, a common approach is to assume the image patches of size $p \times p$ to have a sparse dictionary representation. Here, the signal matrix in (10.11) is the patch matrix $\mathbf{X} = \mathcal{T}(\mathbf{x})$ of dimension $p^2 \times n$. The ability to choose the nonzero coefficients for each patch facilitates the use of the same dictionary for the entire image. This formulation can be extended to learning from multiple images by horizontally stacking the patch matrices as $[\mathcal{T}(x_1), ..., \mathcal{T}(x_N)]$. The dictionary learning approach is thus a learning-based alternative for transformations such as wavelets or discrete cosine transform widely used in compressed sensing. It is also an alternative to the patch-based low-rank methods (e.g., BM3D) used in the image domain, where similar patches need to be clustered prior to subspace fitting. Unlike these methods, the patch dictionary-based schemes use the same dictionary for all patches; the sparsity of the coefficients allow the adaptation of the specific basis functions used in the representation to the specific patch. Once the dictionary $\mathbf{\Phi}$ is learned from multiple images, it can be used to recover images from undersampled measurements [57] as

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \ \alpha \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \|\mathcal{T}(\mathbf{x}) - \mathbf{\Phi}\mathbf{W}^T\|^2 + \lambda \|\mathbf{W}\|_{\ell_p}, \tag{10.12}$$

which is the extension of compressive sensing (CS) to image patches. The second and third term encourages the patches in the solution to be a sparse linear combination of atoms in $\mathbf{\Phi}$, while the first term encourages $\mathbf{x}$ to satisfy data consistency; the optimal solution is a compromise between the two, where the relative importance of data consistency is controlled by α.

10.3.3 Blind compressed sensing (BCS)

BCS schemes estimate the dictionary and the coefficients directly from the measured under-sampled data, rather than pre-learning the dictionary from exemplar data. By adapting the dictionary to the specific image content, these schemes can offer improved performance.
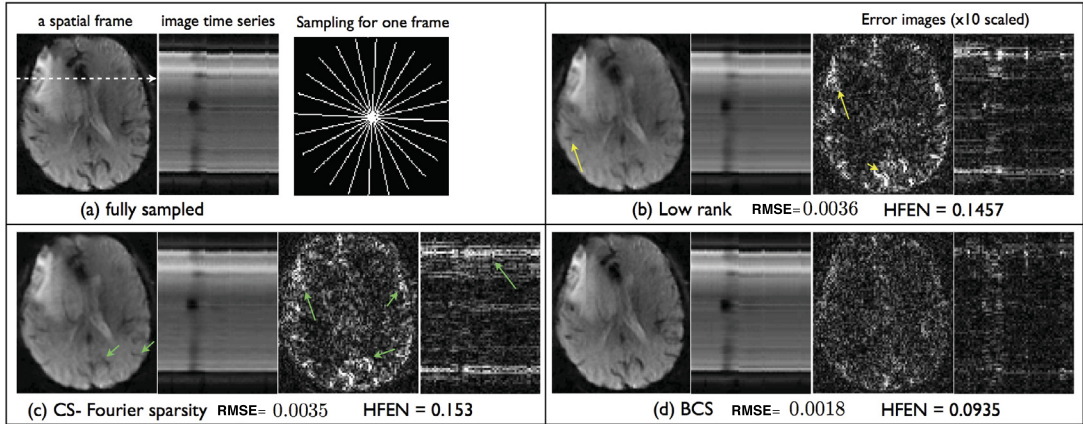
FIGURE 10.3

Comparison of low-rank, Fourier sparsity, and blind compressed sensing (BCS) reconstructions on a brain perfusion MRI dataset retrospectively undersampled at an undersampling factor of 10 fold. BCS is shown to provide superior spatial and temporal fidelity in characterizing the contrast agent temporal dynamics compared to low-rank and CS reconstructions.

10.3.3.1 *Application of BCS to dynamic MRI*

In the context of dynamic MRI, blind compressed sensing methods learn the bilinear model in (10.11) directly from undersampled data [34,35,59,8]. In the dynamic setting with the transpose of the Casorati matrix, each column of $\mathbf{X}$ corresponds to the temporal profile of a pixel. The factorization $\mathbf{X} = \boldsymbol{\Phi}\mathbf{W}^T$ amounts to expressing the temporal profiles of each pixel as a linear combination of the columns of the dictionary $\boldsymbol{\Phi}$. When $\mathbf{W}$ is sparse, the temporal profile of each pixel is expressed as the linear combination of a few atoms, which change from pixel to pixel. Please see Fig. 10.4 for the difference between low-rank and dictionary representation. Unlike the low-rank setting that uses the same basis functions at all pixels, the dictionary learning scheme is able to customize the basis functions for each pixel. In particular, the coefficients that are nonzero (and hence the temporal basis functions that are active) for the heart region with periodic oscillations may be different from that of a static region.

Since the temporal profiles of the pixels change from subject to subject, it is not practical to pre-learn the dictionary from other dataset. The dictionary $\boldsymbol{\Phi}$, which may be overcomplete, and its coefficients $\mathbf{W}$ are hence learned directly from the undersampled dataset itself as

$$\{\boldsymbol{\Phi}, \mathbf{W}\} = \arg\min_{\boldsymbol{\Phi}, \mathbf{W}} \|\mathbf{E}\left(\boldsymbol{\Phi}\mathbf{W}^T\right) - \mathbf{s}\|^2 + \lambda_1 \|\mathbf{W}\|_{\ell_0} + \lambda_2 \ \mathcal{R}(\boldsymbol{\Phi}), \qquad (10.13)$$

The use of blind compressed sensing scheme in (10.13) further improves the quality of dynamic MRI reconstructions [34,35,8] compared to linear (low-rank) models. Specifically, low-rank models use the same basis functions for the voxel profiles of each pixel. The projection of the time series to the signal subspace results in nonlocal temporal averaging [9]. Since the basis functions are the same for each pixel, the temporal point spread functions are the same for each pixel. By contrast, the active basis functions (ones corresponding to nonzero coefficients) in sparse models can potentially differ from
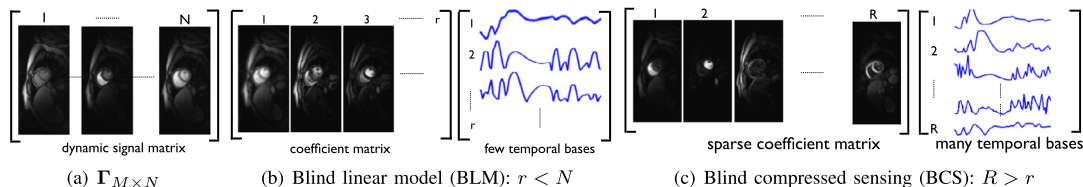
(a) $\mathbf{\Gamma}_{M \times N}$ (b) Blind linear model (BLM): $r < N$ (c) Blind compressed sensing (BCS): $R > r$

FIGURE 10.4

Comparison of blind compressed sensing (BCS) and low-rank (blind linear model) representations of dynamic imaging data: The Casorati form of the dynamic signal is shown in (**a**). The BLM and BCS decompositions are respectively shown in (**b**) and (**c**). BCS uses a large over-complete dictionary, unlike the orthogonal dictionary with few basis functions in BLM ($R > r$). Note that the coefficients/spatial weights in BCS are sparser than that of BLM. The temporal basis functions in BCS dictionary are representative of specific regions since they are not constrained to be orthogonal. For example, the first, second columns of the temporal basis functions in BCS correspond respectively to the temporal dynamics of the right and left ventricles in this myocardial perfusion data with motion. We observe that only 4-5 coefficients per pixel are sufficient to represent the dataset. This figure adapted from [35] with permission from IEEE.

pixel to pixel and hence, the temporal averaging at each pixel is different. This is a desirable feature in applications where the temporal motion patterns are drastically different from pixel to pixel depending on the organs within the field of view (e.g., heart, lung). An illustration of the blind compressed sensing (BCS) approach is shown in Fig. 10.3, where it is compared against global low-rank methods and approaches that use sparse models with fixed (Fourier) dictionary.

10.3.3.2 *Application of BCS to static imaging*

The blind compressed sensing formulation in (10.13) can also modified to the patch setting. In particular, the dictionary, the coefficient matrix $\mathbf{W}$, and the resulting image are all simultaneously learned from the measured data [59] as

$$\{\mathbf{x}, \mathbf{W}, \mathbf{\Phi}\} = \arg \min_{\mathbf{x}, \mathbf{W}, \mathbf{\Phi}} \ \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \alpha \|\mathcal{T}(\mathbf{x}) - \mathbf{\Phi}\mathbf{W}^T\|^2 + \lambda_1 \|\mathbf{W}\|_{\ell_p} + \lambda_2 \mathcal{R}(\mathbf{\Phi}). \qquad (10.14)$$

Here, the first term is the data-consistency term that measures the discrepancy of the recovered image $\mathbf{x}$ from the measurements. Ideally, we would like to have the patch matrix $\mathcal{T}(\mathbf{x})$ extracted by the image to have a compact factorization $\mathcal{T}(\mathbf{x}) = \mathbf{\Phi}\mathbf{W}^T$, where the coefficient matrix $\mathbf{W}$ is sparse and the dictionary is compact under a prior $\mathcal{R}(\mathbf{\Phi})$ such as $\|\mathbf{\Phi}\|_{\ell_2}$ or $\|\mathbf{\Phi}\|_{\ell_1}$. Rather than introducing the factorization as a constraint, the formulation in (10.14) relies on a penalty term; when $\alpha \to \infty$, the solution will satisfy $\mathcal{T}(\mathbf{x}) = \mathbf{\Phi}\mathbf{W}^T$.

10.4 Structured low-rank methods

As mentioned in Section 2.2, 2D/3D patch extraction operations can generate data matrices with block convolutional structure. The earlier methods relied on patches exclusively in the image domain. We will now review methods that exploit the similarity of patches in the Fourier domain, or equivalently

consider structured matrices $\mathcal{T}(\widehat{\mathbf{x}})$ obtained by lifting the discrete Fourier coefficients of the signal $\widehat{\mathbf{x}}$. It is interesting to note that several image properties result in extensive correlations between the k-space samples, which can be used with global low-rank regularization.

10.4.1 Low-rank structure of patch matrices in k-space

The structured low-rank methods in MRI started with the multichannel methods termed as ESPIRIT [28], simultaneous autocalibrating and k-space estimation (SAKE) [65], followed by single channel approaches termed as Low-Rank Modeling of Local k-Space Neighborhoods (LORAKS) [21,23,22], Annihilating filter based LOw-rank HAnkel matrix ALOHA [29,26,25], structured low-rank (SLR) [49,51,48] and multishot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS) [38,41]. We now briefly review some of the low-rank relationships resulting from specific signal properties.

10.4.1.1 *Low-rank relationships in multichannel MRI*

In parallel MRI schemes that acquire multichannel data, the sensitivity-weighted image data are given by

$$x_i(\mathbf{r}) = x(\mathbf{r})\, c_i(\mathbf{r}), \ i = 1, .., N_{\text{channels}}, \tag{10.15}$$

where $c_i(\mathbf{r})$ is the sensitivity weighting of the ith receiver coil. The multichannel relations specified by (10.15) results in image domain annihilation relationships [44]

$$\underbrace{x(\mathbf{r})c_1(\mathbf{r})}_{x_1(\mathbf{r})}\, c_2(\mathbf{r}) - \underbrace{x(\mathbf{r})c_2(\mathbf{r})}_{x_2(\mathbf{r})}\, c_1(\mathbf{r}) = 0. \tag{10.16}$$

One can take the Fourier transforms of both sides of the previous equation to obtain [28,65]

$$\widehat{x}_1 * \widehat{c}_2 - \widehat{x}_2 * \widehat{c}_1 = 0. \tag{10.17}$$

When the coil sensitivities c_i are smooth, one can reliably approximate them as bandlimited functions, whose Fourier support is restricted to a $p \times p$ square region. We now focus on the convolution between the signal $\mathbf{x}$ and a finite impulse response filter $\mathbf{c}$ of support $p \times p$. The convolution output at each pixel $\mathbf{r}$ can be thought of as the innerproduct between the flipped version of a $p \times p$ patch of $\mathbf{x}$, centered at $\mathbf{r}$ with $\mathbf{c}$. Thus, the convolution can be expressed in the matrix form as

$$c * x = \mathbf{c}^T\, \mathcal{T}(\mathbf{x}), \tag{10.18}$$

where $\mathcal{T}(\mathbf{x})$ is a lifted matrix, whose columns are flipped versions of $p \times p$ patches from $\mathbf{x}$. The vector $\mathbf{c}$ corresponds to a vectorized version of the filter c. With this property, one can rewrite (10.17) as

$$\mathbf{c}_2^T \mathcal{T}(\widehat{\mathbf{x}_1}) - \mathbf{c}_1^T \mathcal{T}(\widehat{\mathbf{x}_2}) = 0. \tag{10.19}$$

We note that similar annihilation relationships can be found for every pair of channels. We can compactly express these relations in the matrix form as

$$\underbrace{\begin{bmatrix} \widehat{\mathbf{c}}_2^T & -\widehat{\mathbf{c}}_1^T & 0 & \dots \\ \widehat{\mathbf{c}}_3^T & 0 & -\widehat{\mathbf{c}}_1^T & \dots \\ \vdots & \vdots & \ddots & \dots \\ \widehat{\mathbf{c}}_{N_c}^T & 0 & 0 & \dots & -\widehat{\mathbf{c}}_1^T \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \mathcal{T}(\hat{\mathbf{x}}_1) \\ \mathcal{T}(\hat{\mathbf{x}}_2) \\ \vdots \\ \mathcal{T}(\hat{\mathbf{x}}_{N_c}) \end{bmatrix}}_{\mathcal{M}(\widehat{\mathbf{X}})} = \mathbf{0}. \tag{10.20}$$

Note that each of the rows of P are null-vectors of $\mathcal{M}(\widehat{\mathbf{X}})$, which are linearly independent. Hence, the matrix $\mathcal{M}(\widehat{\mathbf{X}})$ is low-rank [28,65]. Here, $\widehat{\mathbf{X}} = [\hat{\mathbf{x}}_1, .., \hat{\mathbf{x}}_{N_c}]$ is the multichannel data in the Fourier domain. The these multichannel convolution relationships can also be rewritten as

$$\widehat{\mathbf{X}} = \underbrace{(\mathcal{I} - \mathbf{P})}_{\mathbf{G}} \mathcal{M}(\widehat{\mathbf{X}}), \tag{10.21}$$

where the operator $\mathcal{I}$ in (10.21) extracts $\widehat{\mathbf{X}}$ from $\mathcal{M}(\widehat{\mathbf{X}})$ (*i.e.*, $\mathcal{I}\{\mathcal{M}(\widehat{\mathbf{X}})\} = \widehat{\mathbf{X}}$). This relationship forms the basis of the autocalibrating parallel MRI reconstruction method, SAKE [65], which interpolates the missing k-space samples of the accelerated acquisition, based on the structured low-rank property. The SAKE relation in (10.21) can be viewed as a generalization of GRAPPA reconstruction method for multichannel MRI [20].

The formulations in (10.16)–(10.20) is general enough to be applied to settings beyond multichannel MRI. Researchers have used these relationships in several contexts, such as the calibration-less compensation of phase errors in multichannel diffusion MRI [39,37], the correction of Nyquist ghost artifacts in echo-planar imaging [41,36], and the correction of trajectory errors in radial MRI [40]. In all of these cases, different segments of k-space experience different distortions [39,41,36]. These errors can be modeled as image domain weighting (similar to the coil sensitivity weighting studied previously), allowing the use of the relationships (10.16)-(10.20), to derive the structured data matrix with the low-rank property. Fig. 10.5 shows the application of the patch low-rank idea for the reconstruction of high-resolution diffusion MRI data from multishot acquisitions, where the phase compensation of the multishot data were achieved in a calibration-less manner using patch low-rank.

10.4.1.2 *Low-rank structure resulting from finite support and smoothly varying image phase*

The LORAKS algorithms introduced in [21,23,22] makes use of the property of finite support of images to derive the patch low-rank relationships, instead of the image domain weightings. Haldar [21] showed that for images possessing finite support, i.e., the signal $x(\mathbf{r})$ is zero within a region $\mathbf{r}_i \in \Omega$, the annihilation relationships of the form

$$x(\mathbf{r}) \cdot f_i(\mathbf{r}) = 0 \tag{10.22}$$

can be derived, where $f_i(\mathbf{r})$ are functions (also referred to as "filters") that are zero at all locations except at $\mathbf{r}_i \in \Omega$. Additionally, when $f_i(\mathbf{r})$ is assumed to be smooth so that it is bandlimited in the
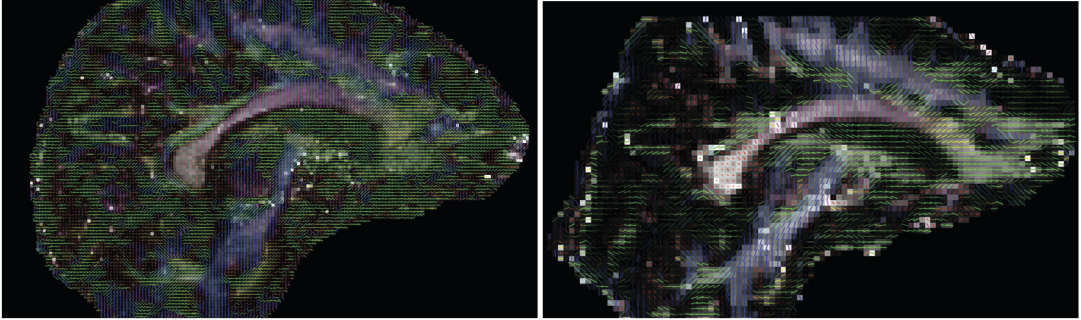
FIGURE 10.5

High-resolution diffusion MRI enabled by the patch low-rank methods in k-space. The image on the left side is acquired using a 4-shot diffusion weighted EPI scan at 1.1 mm isotropic resolution on 7 Tesla MRI with a standard clinical strength gradient of 40 mTesla/m. For comparison, a typical 2 mm-isotropic resolution single-shot diffusion MRI scan slice-matched from the same study is provided on the right. The multishot data on the left is reconstructed using the MUSSELS method that makes use of the patch low-rank in k-space and parallel imaging to recover diffusion weighted images free of phase errors. The iterative reweighted least squares implementation enables efficient reconstruction for such high-dimensional datasets.

Fourier domain, the previous multiplication relationships translate to convolution relations in k-space

$$\widehat{x}(\mathbf{r}) * \widehat{f_i}(\mathbf{r}) = 0, \tag{10.23}$$

resulting in annihilation relationships in the Fourier domain, and results in the reduction of the degrees of freedom. Typically, one can find multiple filters $f_i(\mathbf{r})$ that satisfy the relationships in (10.22). This relationship implies that

$$\mathcal{T}(\widehat{\mathbf{x}})\,\mathbf{F} = 0, \tag{10.24}$$

or equivalently $\mathcal{T}(\mathbf{x})$ is a low-rank matrix. Haldar et al. has empirically indicated that the rank of the lifted convolution matrices, $\mathcal{T}(\mathbf{x})$, corresponding to MR images indeed depends on the support of the signal. Since these patches are formed from single-channel images, it can be applied for single-channel undersampled recovery.

It is well-known that real images exhibit conjugate symmetry in k-space, resulting in annihilation relationships

$$\widehat{x}[k] - \widehat{x}[-k]^* = 0. \tag{10.25}$$

When the phase of the images are smoothly varying, [21] showed that one can construct a specialized convolution matrix using the 2D patches from the conjugate symmetric k-space samples also, that satisfy annihilation relationships and hence is low-rank. These results translate to structured low-rank algorithms that account for the patch low-rank structure.

10.4.1.3 *Low-rank structure resulting from continuous domain sparsity*

Here, we discuss the low-rank relationships for sparse continuous domain functions, which enable super-resolution reconstructions. Traditional CS schemes assume the images to be sparse on a specific grid, which may be an unrealistic assumption. Several researchers have considered the extension of CS for the super-resolution setting, where the sparse samples of the signal may not be localized to a grid [14]. Specifically, we can use an image model using impulse functions

$$x(r) = \sum_{i=1}^{R} \gamma_i \, \delta(r - r_i) \tag{10.26}$$

for sparse images, where γ_i are the weights and r_i are the location of the impulses, which are not necessarily on a uniform grid. The SLR [50] and ALOHA settings [26] extends the Fourier domain annihilation relationships discussed in the previous sections for the recovery of continuous domain sparse signals.

To see how the aforementioned sparse image model can harness the Fourier domain annihilation relationships, we first discuss a simple 1D case. The seminal work by Prony dating back to 1885 showed that 1D exponential signal of the form $\widehat{x}[k] = \alpha^k$ can be annihilated by convolution as follows [27]:

$$\widehat{x}[k] * h[k] = \underbrace{\alpha^k}_{\widehat{x}[k]} - \alpha \underbrace{\alpha^{k-1}}_{\widehat{x}[k-1]} = 0, \tag{10.27}$$

where h[k] is a two tap filter given by $[1, -\alpha]$. This theory is relevant to MR images because, when $\alpha = \exp(jr_0)$, $\widehat{x}[k] = \exp(jr_0 k)$ is the Fourier transform of a sparse signal of the form $\delta(r - r_0)$.

More generally, when the signal is a linear combination of multiple impulses at location $r_0, .., r_k$, its Fourier coefficients can be annihilated by the convolution with a $k + 1$ tap filter; the $k + 1$ tap filter is obtained by the convolution of the k two-tap filters that annihilates each of the exponentials. Here, the location of the impulses r_i are not required to be localized to a specific grid; this approach may be viewed as the continuous domain extension of discrete compressed sensing methods. The extension of this idea to two dimensions is relatively straightforward.

The convolution-based annihilation relationship in (10.27) for k impulses for the superresolution recovery can thus be compactly expressed as [27,21,49,26]

$$\mathbf{h}^T \underbrace{\begin{bmatrix} \widehat{x}[0] & \widehat{x}[1] & \ldots & \widehat{x}[N-k] \\ \widehat{x}[1] & \widehat{x}[2] & \ldots & \widehat{x}[N-k+1] \\ \vdots & \vdots & \ddots & \vdots \\ \widehat{x}[k] & \widehat{x}[k+1] & \ldots & \widehat{x}[N] \end{bmatrix}}_{\mathcal{T}(\widehat{x})} = \mathbf{0}. \tag{10.28}$$

Note that $\mathcal{T}(\widehat{\mathbf{x}})$ is a patch matrix obtained by lifting the 1D signal $\widehat{x}[n]$, which are the Fourier coefficients of $\mathbf{x}$. The columns of $\mathcal{T}(\widehat{\mathbf{x}})$ correspond to $(k + 1) \times 1$ patches in $\mathbf{x}$. (10.28) implies that the matrix $\mathcal{T}(\widehat{\mathbf{x}})$ has a null-space vector. If the number of impulses is $K' < k$, Eq. (10.28) will have $k - K' + 1$ null-space vectors [27,26]. In other words, the rank of the matrix $\mathcal{T}(\widehat{x})$ is a surrogate for the number of impulses in the signal.

Let us also briefly discuss the image domain interpretation of the filters $h[k]$. Taking the inverse Fourier transform of (10.27), we obtain

$$\hat{h}(r) \cdot x(r) = 0, \tag{10.29}$$

where $\hat{h}(r)$ and $x(r)$ the inverse Fourier transforms of $h[k]$ and $\hat{x}[k]$, respectively. Note that the convolution in (10.27) is translated into the point-by-point multiplication [49,50]. For the example considered in (10.27), we get

$$\mu(r) = 1 - \underbrace{\exp(jr_0)}_{\alpha} \exp(-jr) = 1 - \exp(-j(r - r_0)). \tag{10.30}$$

Note that this exponential is a first-order polynomial that is nonzero at all locations, except at $r = r_0$, which is the location of the impulse. Likewise, when the signal x consists of k impulses, $\mu(x)$ is a bandlimited function that is zero only on the nonzero locations of $x(r)$. This interpretation will be useful in the next section for the discussion of piecewise smooth images.

10.4.1.4 *Low-rank structure of piecewise smooth images*

We can extend the annihilation relations for sparse image models to more general settings. The gradients of piecewise constant images are often significantly sparser [49] than the support of the signal considered in [21] or sparse model assumed in [26]. Note that the Fourier transform of the gradients of the 1D signal $\widehat{\partial_{r_1} x} = jk_{r_1}\hat{x}$. Thus, replacing $\hat{x}$ by $\widehat{\partial_{r_1} f}$ will result in a matrix with a significantly smaller rank.

The generalization of the idea to multiple dimensions is not straightforward from a theoretical perspective. Specifically, the gradient of a piecewise constant image is nonzero on the edges. Unlike the sparse model considered in the 1D case, where the number of impulses is finite, the gradient cannot be modeled as the sum of a finite number of impulses, which makes the recovery of the 2D images from few measurements is ill-posed. Prony's model and the related theory is only valid when the number of impulses is finite. Nevertheless, the problem can be made well-posed by assuming the edges to be localized to the zero-sets of a 2D bandlimited function $\mu(\mathbf{r})$ [49,50]. This model amounts to stating that the piecewise constant image has smooth edge contours. In this case, we have

$$\mu(\mathbf{r}) \cdot \underbrace{\left[\partial_{r_1} x(\mathbf{r}) \quad \partial_{r_2} x(\mathbf{r}) \right]}_{\nabla x(\mathbf{r})} = 0. \tag{10.31}$$

Here, $\mu(\mathbf{r})$ is a bandlimited function that is zero at the edges of the image and nonzero elsewhere. Taking the Fourier transform on both sides, we obtain $h * \mathcal{T}_2(\mathbf{x})$, where

$$\mathcal{T}_2(\hat{\mathbf{x}}) = \left[\mathcal{T}\left(\widehat{\partial_{r_1} x}\right) \quad \mathcal{T}\left(\widehat{\partial_{r_2} x}\right) \right], \tag{10.32}$$

where $\mathcal{T}\left(\widehat{\partial_{r_1} x}\right)$ is the 2D patch matrix of the Fourier coefficients of the partial derivative of $\mathbf{x}$. The same approach can be extended to piecewise polynomials by replacing (10.32) with a matrix with more partial derivatives. The number of rows in equal to p^2, which is dependent on the size of the patch. The number of columns is equal to the number of valid patches in the images, without considering the regions outside the image. Note that, as the size of the patches increase, the number of patches and

the number of columns decrease [50]. Theoretical results show that the best performance is obtained when the matrix $\mathcal{T}_2\left(\widehat{\partial_{r_1} x}\right)$ is square shaped, which roughly corresponds to each of the patch dimensions being half the corresponding image dimensions. However, practical algorithmic considerations, such as memory and computational constraints, often force the size of the patches to be smaller.

10.4.1.5 *Low-rank relations in parameter mapping*
Many parameter mapping applications in MRI consider the imaging of a time series, where the intensity of the pixel values at the spatial location $\mathbf{r}$ and time-point n change in an exponential fashion (e.g., $\rho[\mathbf{r}, n] = \alpha[\mathbf{r}]^n + c$, where c and α are arbitrary constants). Such a signal can be annihilated by a finite difference filter, whose parameters depend on α. The same approach can be readily extended to cases where the signal is the sum of several exponentials, where the size of the filter depends on the number exponentials. For simplicity, we will consider a single exponential signal $x[n] = \alpha^n$. From (10.27), we see that such exponential signals satisfy an annihilation relationship. Thus, the matrices obtained by lifting the time series entries will be low-rank in nature. Moreover, in many cases, the parameter maps vary smoothly in space, i.e., the exponential decay in a given pixel is highly correlated to the exponential decay in the neighboring pixels. In such cases, the coefficients of the exponentials and the parameters of the exponentials, themselves, can be modeled as band-limited functions. In this case, the k-t space samples of the parameter mapping application can be annihilated by multichannel convolution relations. This property was used successfully to recover $T_1\rho$ and T_2 maps in parameter mapping [4] applications as well as B0 field inhomogeneity compensated recovery of EPI images [5].

10.4.2 **Algorithms for k-space patch low-rank methods**
In Section 4.1, we discussed several properties of the images that result in Fourier domain annihilation relationships that translate to low-rank relationships on the associated structured matrices. The main difference for each problem is the lifting operation that is used to create the structured matrix from the Fourier samples. The lifting operation depends upon the specific image property that is accounted for. Once the structured data matrix is created, the recovery of the images using the low-rank relationships is posed as an optimization problem [22,49,50,26]. We will first discuss the multichannel parallel MRI case where the recovery of the images from the undersampled measurements can be written as the unconstrained optimization

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \lambda \|\mathcal{T}(\hat{\mathbf{x}})\|_*, \tag{10.33}$$

Here, $\mathcal{T}(\hat{\mathbf{x}})$ is the lifted structured matrix formed from the multichannel convolution relationships in (10.20). This problem can be solved as a general low-rank matrix completion problem where the Hankel structure is additionally enforced. The rank minimization can be performed using singular value thresholding schemes and the optimization can be performed in an alternating manner updating the data consistency and rank-minimization. Fig. 10.6 shows a schematic of this approach where the cited problem is applied to the recovery of missing k-space samples as an interpolation problem.

The number of entries of the matrix $\mathcal{T}(\hat{\mathbf{x}})$ is several orders of magnitude larger than the size of the image, as discussed earlier. Because of this, the storage and computation of the matrix is often impossible in high-resolution and multidimensional applications. Moreover, the rank-minimization involves computing the SVD which is also computational demanding. Several algorithms were introduced to solve k-space low-rank problems similar to (10.33), as described in [22,49,50,26].
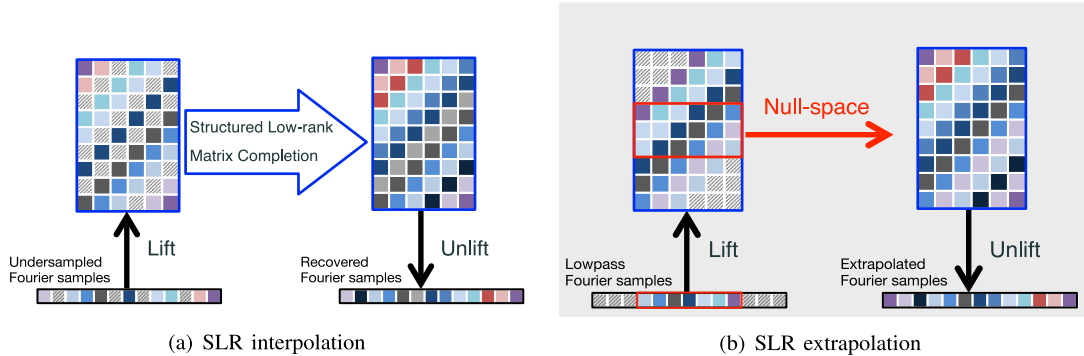
(a) SLR interpolation   (b) SLR extrapolation

FIGURE 10.6

Illustration of SLR-based interpolation and extrapolation methods in the context of 1D FRI. (**a**) In SLR interpolation, the data is acquired on a nonuniformly subsampled Fourier grid. The SLR interpolation scheme relies on a lifting of the signal samples to a Hankel matrix, which has missing entries indicated by the hashed boxes. The one-to-one relationship between the rank of a matrix and the continuous domain sparsity of the space domain signal is used to pose the recovery of missing samples as a structured low-rank matrix completion (SLRMC) problem in the lifted matrix domain. Specifically, the algorithm determines the matrix with the lowest rank that satisfies the Hankel structure and is consistent with the known matrix entries. Post-recovery, the matrix is unlifted to obtain the Fourier samples of the signal. (**b**) In SLR extrapolation problems, the low-frequency Fourier coefficients of the signal are uniformly sampled. The central fully known matrix region is used to estimate the null space of the matrix, which is used to linear-predict/extrapolate the missing high-frequency samples. The SLR algorithms that exploit the different signal structures differ only in the structure of the lifted matrix; the algorithms are essentially the same. This figure is copied from [24] with permission from IEEE.

10.4.3 Iterative reweighted least square (IRLS) algorithm

The IRLS scheme relies on the approximation of the nuclear norm [51]

$$\|\mathcal{T}_2(\hat{\mathbf{x}})\|_* \approx \|\mathbf{Q} \, \mathcal{T}(\hat{\mathbf{x}})\|^2, \tag{10.34}$$

where $\mathbf{Q}$ is a $p^2 \times p^2$ matrix

$$\mathbf{Q} = \left(\mathcal{T}(\hat{\mathbf{x}})\mathcal{T}(\hat{\mathbf{x}})^T\right)^{-1/4} \tag{10.35}$$

The main benefits of the IRLS algorithm [51] is that it significantly reduces the computational complexity and the memory demand. In addition, we use this algorithm in some of the existing SLR methods in Section 10.4.4 and to connect kernel PCA with manifold methods in Section 10.5. In the perfectly low-rank setting, $\mathbf{Q}$ can be viewed as the projection onto the null-space and hence the right-hand side of (10.34) can be viewed as the energy of the projection of $\mathcal{T}(\mathbf{x})$ on to the null-space.

Using the previous approximation and the structure of the matrix $T(\mathbf{x})$, one can solve (10.28) by alternating between (10.35) and

$$f^* = \arg \min_f \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \lambda \sum_{i=1}^{p^2} \|\widehat{\mathbf{x}} * \mathbf{q}_i\|^2, \tag{10.36}$$

where $\mathbf{q}_i$ are the columns of $\mathbf{Q}$ [51]. The main benefit of (10.36) is that this formulation does not need the computation and storage of the large matrix $\mathcal{T}_2(\mathbf{x})$, which makes it possible to apply the scheme to multidimensional high-resolution applications.

A second approach is to approximate the nuclear norm using UV factorization [26]

$$\|\mathcal{T}_2(\hat{\mathbf{x}})\|_* \approx \arg \min_{\mathbf{U}, V}, \|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2, \text{ where } \mathcal{T}_2(\hat{\mathbf{x}}) = \mathbf{U}\mathbf{V}^H, \tag{10.37}$$

as in Chapter 9. This method does not involve SVD computation, thus speeding up the minimization.

10.4.4 Algorithms that rely on calibration data

When a fully sampled center of k-space is available, a calibration based strategy can be employed for solving the minimization problem. Using the known data from the fully sampled region, the null-space filters $\mathbf{q}_i$ can be estimated. Note from (10.34) that the projection to the null-space should be as small as possible, which implies that $\mathbf{Q} \mathcal{T}(\mathbf{x}) \approx 0$. Hence, $\mathbf{Q}$ can be estimated from the central k-space regions by solving [49,28]

$$\mathbf{Q}^* = \arg \min_{\mathbf{Q}} \|\mathbf{Q} \mathcal{T}(\mathbf{k}_{\text{central}})\|^2 \text{ such that } \|\mathbf{Q}\|_F = 1 \tag{10.38}$$

using eigenvalue decomposition of $\mathcal{T}(\mathbf{k}_{\text{central}})$ corresponding to the matrix constructed from central k-space samples. Once $\mathbf{Q}$ is known, one can solve (10.36) with the knowledge of the filters, resulting in computationally efficient solution. Fig. 10.6 shows a schematic of this approach where the previous problem is applied to the recovery of missing k-space samples as an extrapolation problem.

Before we conclude the patch low-rank methods, we note several existing MRI reconstructions that are related to the patch low-rank methods previously discussed. The popular ESPIRIT reconstruction uses the null-space property to estimate the coil sensitivities using an eigen decomposition [28]. In particular, once $\mathbf{Q}$ or equivalently the signal subspace of $\mathcal{T}_2$ is obtained, it performs a pixel by pixel eigenvalue decomposition to obtain the coil sensitivities. The GRAPPA [20] approach described in Chapter 6 is also related. For instance, as the size of the patch/filter specified by p decreases, the number of columns/rows in $\mathbf{Q}$ will decrease. If there is only column, the equation $\mathbf{q} T_2(\mathbf{k}_{\text{central}}) = 0$ can be rewritten as $\mathbf{q_o} T_2(\mathbf{k_o}) = -\mathbf{q_u} T_2(\mathbf{k_u})$, obtained by partitioning the rows and columns. Here $\mathbf{k_o}$ and $\mathbf{k_u}$ are the k-space samples that can be observed and cannot be unobserved, respectively. These simplifying assumptions can translate to the GRAPPA setting, which is less general than the ESPIRIT setting.

10.5 Smooth manifold models

Smooth manifold models use nonlinear representations, which are more powerful in capturing the nonlinear relationships between signals compared to the linear counterparts. They assume the signals

(images/patches/pixel profiles) to be living on a smooth image manifold (see Fig. 10.1). Methods relying on smooth manifold models include nonlocal means [12], nonlocal regularization [42], kernel methods [45,66,47,46,64], STORM [1,43,54,10,56], and recent extensions of STORM using deep generative models [71,72]. Most of the methods rely on modeling/smoothing the signals based on their proximity on the manifold rather than in the original domain. For instance, the patches that may be far apart in space might be similar and hence close on the manifold.

We note that standard Tikhonov regularization penalizes the gradient of the image, which makes use of the fact that the intensities of the adjacent pixels are similar. Most algorithms for Tikhonov regularization rely on the Laplacian of the image. For 1D images with n pixels, the Laplacian is often approximated by the $x \times n$ finite difference matrix

$$\mathbf{L} = \begin{bmatrix} -2 & 1 & 0 & \dots & 1 \\ 1 & -2 & 1 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots \\ 1 & 0 & \dots & 1 & -2 \end{bmatrix}. \tag{10.39}$$

This matrix is block diagonal and captures the neighbor structure of the pixels in the image, thus facilitating the smoothing of the image. The sum of the off diagonal entries of each row is equal to the negative of the diagonal entry. In manifold methods, the smoothing is enabled by a custom Laplacian matrix, which captures the neighborhood structure of the images in the manifold, and is estimated from the data. The entries of the Laplacian matrix are chosen based on the proximity of the signals on the manifold (see Fig. 10.7). This approach has strong ties to the kernel methods [7] used in machine learning. In particular, one performs the smoothing in a nonlinearly transformed feature space, defined by the nonlinear mapping function $\varphi(\mathbf{x})$. The neighborhood structure or Laplacian is determined based on the distances between the nonlinearly mapped features $d_{i,j} = \|\varphi(\mathbf{x}_i) - \varphi(\mathbf{x}_j)\|^2$, rather than the conventional distance measure $\|\mathbf{x}_i - \mathbf{x}_j\|^2$. The cost functions to solve for the images using these models only depend on inner products between the features $\langle \varphi(\mathbf{x}_i), \varphi(\mathbf{x}_j) \rangle$; the *kernel-trick* that is widely used in machine learning can be used to come up with computationally efficient algorithms that does not require the explicit computation of the nonlinear mapping φ that is expensive to compute.

10.5.1 Analysis manifold methods

The recovery of a smooth multidimensional function $\mathbf{f}: \mathbb{R}^m \to \mathbb{R}^n$ has been considered in machine learning [7]. For example, when one is considering patches, $n = p^2$ is the number of pixels in the patch. Because of the redundancy of the patches, they can be viewed as a function of low-dimensional latent vectors denoted by $\mathbf{r}_k$: $\mathbf{x}_k = f(\mathbf{r}_k)$. Here, $\mathbf{f}$ is a nonlinear function, and $\mathbf{r}_k$ are as the coordinates that are unknown. We note that as the latent coordinates $\mathbf{r}$ is varying in a m-dimensional space (domain), the function values vary on a smooth surface in high-dimensional space ($n \gg m$). The recovery of the function $\mathbf{f}$ from the fully sampled data points $\mathbf{x}_k$ is posed as the Tikhonov regularized problem [7]

$$\mathbf{f}^* = \arg\min_{f, \{\mathbf{r}_k\}} \|\mathbf{f}(\mathbf{r}_k) - \mathbf{s}_k\|^2 + \eta \int_{\mathcal{M}} \|\nabla \mathbf{f}\|^2 d\mathbf{r}, \tag{10.40}$$
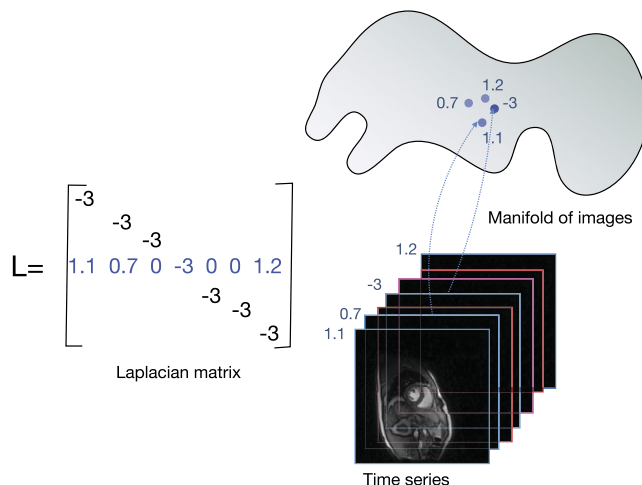
FIGURE 10.7

Illustration of the Laplacian matrix used in manifold methods. Each row of the Laplacian matrix may be thought of a second-order derivative operator in the manifold domain. In this example, the fourth row of the matrix correspond to the finite difference operator for the fourth image in the time series. Note that its neighbors on the manifold are not necessarily its temporal neighbors. The weights denote how close the images are to the fourth image in the time series. We illustrated an example with three neighbors, but the number of neighbors could be chosen arbitrarily and may vary from frame to frame.

where the second term is the smoothness of the function on $\mathcal{M}$. In the discrete setting, the regularization term is approximated as a weighted sum of differences between the points [7]

$$\mathbf{f}^* = \arg\min_{\mathbf{f},\{\mathbf{r}_j\}} \|\mathbf{f}(\mathbf{r}_k) - \mathbf{s}_k\|^2 + \eta \sum_{i=1}^{N}\sum_{i=1}^{N} \mathbf{W}_{i,j}\|\mathbf{f}(\mathbf{r}_i) - \mathbf{f}(\mathbf{r}_j)\|^2, \tag{10.41}$$

where the weights are selected based on the proximity of the points on the manifold. A simple choice of weights is specified by [7]

$$\mathbf{W}_{i,j} = \exp\left(-\frac{\|\mathbf{f}(\mathbf{r}_i) - \mathbf{f}(\mathbf{r}_j)\|^2}{\sigma^2}\right). \tag{10.42}$$

The weights capture the geometry of the manifold. Specifically, closer point pairs on $\mathcal{M}$ will have greater weights, while distant point pairs will have lesser weights. We will discuss more sophisticated approaches for the estimation of weights in Section 10.5.1.2. We note that the weighted sum can be expressed in a compact form as

$$\sum_{i=1}^{N}\sum_{i=1}^{N} \mathbf{W}_{i,j}\|\mathbf{f}(\mathbf{r}_i) - \mathbf{f}(\mathbf{r}_j)\|^2 = \text{trace}\left(\mathbf{X}\mathbf{L}\mathbf{X}^T\right).$$

Here, $\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_N \end{bmatrix} = \begin{bmatrix} \mathbf{f}(\mathbf{r}_1) & \ldots & \mathbf{f}(\mathbf{r}_N) \end{bmatrix}$ and $\mathbf{L}$ is the Laplacian matrix $\mathbf{L} = \mathbf{D} - \mathbf{W}$, which captures the structure of the manifold and $\mathbf{D}$ is a diagonal matrix $\mathbf{D} = \mathrm{diag}(\sum_j \mathbf{W}_{i,j})$. See Fig. 10.7 for an illustration of the structure of this matrix. Thus, the optimization scheme in (10.41) can also be written as

$$\mathbf{X} = \arg\min_{\mathbf{X}} \|\mathbf{X} - \mathbf{S}\|^2 + \eta\, \mathrm{trace}\left(\mathbf{X}\mathbf{L}\mathbf{X}^{\mathrm{T}}\right) \tag{10.43}$$

and $\mathbf{S} = \begin{bmatrix} \mathbf{s}_1 & \ldots & \mathbf{s}_N \end{bmatrix}$. The discrete approximation of the manifold can be viewed as a graph, where the structure of the graph is captured by the graph Laplacian $\mathbf{L}$. Signal processing on graphs is extensively studied, and the Laplacian matrix is central to most of the methods [52].

10.5.1.1 *Relationship to factor models and binning based approaches*

One can perform the eigendecomposition of the known Laplacian matrix $\mathbf{L}$ as

$$\mathbf{L} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T \tag{10.44}$$

It is well-known that the eigenfunctions (columns of $\mathbf{\Phi}$) are basis functions of functions on $\mathcal{M}$, analogous to Fourier exponentials being eigenfunctions of Laplacian operator in Euclidean space [52]. In particular, eigenfunctions corresponding to smaller eigen values of $\mathbf{L}$ correspond to smooth functions on $\mathcal{M}$; the eigenvalues are analogous to the frequency or roughness of the function. Since $\mathbf{\Phi}$ is an orthogonal basis analogous to Fourier transform, one can express the signal matrix as

$$\mathbf{X} = \mathbf{\Phi}\mathbf{W}^T, \tag{10.45}$$

where $\mathbf{W}$ can be viewed as the coefficients. Using this property, one can rewrite (10.43) as

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \|\mathbf{\Phi}\mathbf{W}^T - \mathbf{Z}\|^2 + \eta\, \underbrace{\mathrm{trace}\left(\mathbf{W}\mathbf{\Lambda}\mathbf{W}^T\right)}_{\sum_{i=1}^{N} \lambda_i \|\mathbf{w}_i\|^2} \tag{10.46}$$

Note that (10.45) is similar to the representation of the signal using dictionaries; each signal $\mathbf{k}$ is expressed as the linear combination of $\mathbf{u}_i$ with the weights specified by the kth row of $\mathbf{W}$. The weights are expected to be similar for points closer on the manifold; the active Φ basis vectors in each manifold neighborhood provides a local linear representation (similar to local PCA) on the manifold. Since the eigenvalues λ_i can be viewed as the *frequency* or the measure of roughness, one can attenuate the high-frequency components on the manifold by increasing η, thus obtaining smoother signals on the manifold. For computational efficiency, one may also truncate the representation by ignoring the basis functions corresponding to higher eigenvalues.

When the images can be grouped into r distinct clusters with minimal inter-group similarity, the Laplacian matrix can be thought of as a block diagonal matrix. In particular, the off-diagonal entries of the matrix corresponding to the images from two different clusters are zero. In this case, it is well-known that the matrix will have r zero eigenvalues. The eigenvectors $\boldsymbol{\varphi}_i$ corresponding to the zero eigenvalues will be the indicator vectors of the clusters. If the remaining eigenvalues are much higher, one can approximate (10.46) as the independent recovery of each cluster from the measured data. We note that the hard binning approach pursued in GRASP or XD-GRASP framework [18] bins the data to different clusters, followed by the recovery of the bins. Thus, the hard binning-based approaches may be viewed

as a special case of the manifold method where the inter-cluster similarities can be ignored. When the nonzero eigenvalues are not ignored in the reconstruction, the additional eigenfunctions (corresponding to nonzero eigenvalues) capture the variability of the images within each cluster. The reconstruction can thus be viewed as a local PCA approach, where each cluster is represented independently by its basis set. As discussed before, the BM3D approach pursues a similar approach, where the patches in the image are clustered into different groups.

10.5.1.2 *Estimation of manifold Laplacian*

As discussed previously, **L** captures the manifold structure. The recovery heavily depends on the specific choice of the Laplacian matrix. Several methods were introduced to estimate the Laplacian from its noisy and possibly undersampled data.

Proximity based methods: Early methods directly estimated the weight matrix based on the proximity of the function values [7,12,3,54]. For instance, they are chosen as

$$\mathbf{W}_{i,j} \approx \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\sigma_s^2}\right). \tag{10.47}$$

We note that **W** is also termed as the kernel matrix. The BM3D approach [16,17] can be viewed as a hard-clustering setting

$$\mathbf{W}_{i,j} = \begin{cases} 1 & \text{if} \quad \|\mathbf{s}_i - \mathbf{s}_j\|^2 \leq \sigma_s^2 \\ 0 & \text{else} \end{cases}. \tag{10.48}$$

Note that the Gaussian choice in (10.47) will be equivalent to hard-clustering if the clusters are well-separated. In this case, the **L** matrix will have a block structure with no interactions between clusters. When the points are well-distributed in the manifold, the Gaussian choice will promote the interaction between the points, thus facilitating data-sharing between the corresponding images.

Alternating minimization schemes: In many cases (e.g., signals are patches in an image), the signals $\mathbf{x}_i$ are either noisy or jointly measured using a single-rank deficient linear operator **E**. In this case, the estimation of the weights from aliased data using (10.47) often results in poor results. An approach to overcome this challenge is to pose the recovery as the minimization of the cost function [67,68]

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \|\mathbf{E}(\mathbf{X}) - \mathbf{S}\|^2 + \lambda \sum_{i=1}^{N} \sum_{i=1}^{N} \eta\left(\|\mathbf{x}_i - \mathbf{x}_j\|\right). \tag{10.49}$$

Here, $\eta(\cdot)$ is a nonconvex function of its argument (e.g., $\ell_p; p < 1$ norm). The nonconvexity of the regularization term will encourage each i to be influenced by signals in the immediate proximity, while being minimally impacted by far away points. This criterion can be minimized by alternating between the estimation of the weights (10.47) and the recovery of the signals (10.41). Continuation schemes that start with a quadratic or convex η, and gradually change it to the desired nonconvex cost function during iterations have been introduced to encourage the convergence to the global minimum with improved results in compressed sensing applications.

Sparse optimization: The work in [46,64] proposes to estimate the Laplacian by assuming the weight matrix (and equivalently the Laplacian entries) to be sparse. Specifically, it aims to express each signal

i as a sparse linear combination of the other signals. The intuition is that each signal on the manifold can be expressed as a sparse linear combination of its neighbors

$$\mathbf{W}_{i,j}^* = \arg \min_{\mathbf{W}_{i,j}; \sum_j \mathbf{W}_{i,j}=1} \|\mathbf{s}_i - \sum_j \mathbf{W}_{i,j}\mathbf{s}_j\|^2 + \lambda \sum_j \|\mathbf{W}_{i,j}\|_{\ell_1}. \tag{10.50}$$

This approach is reported to yield improved results over proximity-based methods [7,12,3,54] in (10.47). This approach has similarities to local linear embedding [62], where each signal is expressed as a weighted linear combination of its neighbors.

Kernel based projection: Note that the approach in (10.46) approximates the signals using the eigen vectors corresponding to the lowest eigenvalues of the Laplacian matrix to approximate/denoise them. This is equivalent to approximating the signals using eigenvectors corresponding to the highest eigenvalues of the normalized kernel matrix. Kernel PCA is widely used in machine learning to approximate signals living on manifolds. The eigenvalues of kernel matrix is observed to decay rapidly when the signals are living on smooth manifolds or clustered [6], which is used to denoise the signals.

Early manifold approaches in MRI relied on explicit polynomial features [66,47]. Specifically, polynomial features of the signals were computed, followed by performing PCA in the feature space. Once the features are projected onto a lower dimensional subspace, these methods use the explicit inversion formula to obtain the corresponding signals in the original space; the corresponding signals are termed as pre-images of the projections derived by PCA. This approach demonstrated improved performance over PCA. A challenge for this direct approach is the difficulty applying it to large images, where the explicit lifting is not possible.

Kernel PCA regularization: The low-dimensional structure of the weight matrix is explicitly used for the joint estimation of the Laplacian matrix and the signals from undersampled measurements in [56]. We now show a simple example to illustrate that the kernel matrix can be exactly low-rank if the original points lie on simple constructs such as surfaces or manifolds. For example, the points on a unit circle in two dimensions live on the zero set of $x^2 + y^2 - 1$. Hence, one can consider a nonlinear lifting of the points

$$\begin{bmatrix} x \\ y \end{bmatrix} \rightarrow \begin{bmatrix} 1 \\ x^2 \\ y^2 \end{bmatrix} = \phi\left(\begin{bmatrix} x \\ y \end{bmatrix}\right). \tag{10.51}$$

Because the high-dimensional features of every point on the circle satisfies

$$\begin{bmatrix} -1 & 1 & 1 \end{bmatrix} \underbrace{\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1^2 & x_2^2 & \dots & x_N^2 \\ y_1^2 & y_2^2 & \dots & y_N^2 \end{bmatrix}}_{\Phi(\mathbf{X})} = 0, \tag{10.52}$$

we can conclude that $\Phi(\mathbf{X})$ is rank deficient by one. If the points lie on the intersection of two polynomials, one would have more null-space vectors, further reducing the rank of $\Phi(\mathbf{X})$.

Generalizing this simple illustration, we can show that the kernel matrix $\mathbf{W}$ exactly low-rank when the manifold $\mathcal{M} \in \mathbb{R}^N <$ can be expressed as the zero-level set of a finite linear combination of basis

functions $\psi(\mathbf{x}) = \sum_{k=1}^{B} c_k \, \varphi_k(\mathbf{x})$ as follows:

$$\mathcal{M} = \{x \mid \sum c_k \, \varphi_k(\mathbf{x}) = 0\}. \tag{10.53}$$

Here, $\varphi_k(\mathbf{r})$ are basis functions[1] (e.g., polynomials, exponentials) that span the high-dimensional space. In this case, the feature matrix

$$\mathbf{\Phi}(\mathbf{X}) = \begin{bmatrix} \underbrace{\begin{bmatrix} \varphi_1(\mathbf{x}_1) \\ \vdots \\ \varphi_S(\mathbf{x}_1) \end{bmatrix}}_{\phi(\mathbf{r}_1)} & \cdots & \underbrace{\begin{bmatrix} \varphi_1(\mathbf{x}_N) \\ \vdots \\ \varphi_S(\mathbf{x}_N) \end{bmatrix}}_{\phi(\mathbf{x}_N)} \end{bmatrix} \tag{10.54}$$

is low-rank.

The mapping from the original points to the feature vectors $\phi(x)$ can be viewed as a nonlinear lifting. The low-rank structure of this matrix implies that the lifted points lie in a low-dimensional subspace. The lifted points can hence be viewed as low-dimensional latent vectors that compactly represent the signals. The algorithms that use the low-rank property of the feature vectors may be viewed as structured low-rank algorithms with the nonlinear mapping ϕ. Following the approach in Section 10.4, we recover the signals on the manifold from its linear measurements as

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \|\mathbf{E}(\mathbf{X}) - \mathbf{S}\|^2 + \lambda \|\mathbf{\Phi}(\mathbf{X})\|_*. \tag{10.55}$$

This problem cannot be solved in practical applications since the feature vectors $\phi(x)$ are high-dimensional. Hence, one can use the *kernel-trick* that allows the direct computation of the inner-product of the high-dimensional feature maps

$$\langle \phi(\mathbf{r}_i), \phi(\mathbf{r}_j) \rangle = \kappa(\mathbf{x}_i - \mathbf{x}_j) \tag{10.56}$$

as nonlinear functions of the image differences $\mathbf{x}_i - \mathbf{x}_j$ without the direct computation of the feature maps, which is computationally challenging. Here, κ is a function that is dependent on the specific feature maps. This approach is often referred to as the *kernel-trick*, which is widely used in machine learning to avoid explicit lifting to higher-dimensional space. We note that the IRLS approach described in Section 10.4.2 only depends on the inner products of the features, unlike many of the nuclear norm-minimization algorithms. This property can be made use of to solve (10.55), which alternates between

$$\mathbf{X}^* = \arg\min_{\mathbf{X}} \|\mathbf{E}(\mathbf{X}) - \mathbf{S}\|^2 + \lambda \|\mathbf{\Phi}(\mathbf{X}) \, \kappa(\mathbf{X})^{-1}\|_F^2 \tag{10.57}$$

and the evaluation of the kernel matrix $\kappa(\mathbf{X})$

$$(\kappa(\mathbf{X}))_{i,j} = \langle \phi(\mathbf{r}_i), \phi(\mathbf{r}_j) \rangle = \kappa(\mathbf{r}_i, \mathbf{r}_j). \tag{10.58}$$

---

[1] The basis functions may also be chosen as a subset of the Mercer decomposition of the kernel function.

This approach eliminates the need for the explicit evaluation of the high-dimensional feature maps $\phi(f_i)$. Since the regularization functional in (10.57) is nonquadratic, this term is solved using steepest descent; this approach is shown to be similar to the alternating strategy in (10.5.1.2) (i.e., alternation between (10.47) and the recovery of the signals (10.41)), where (10.47) is replaced by

$$\mathbf{L} = \mathcal{K}(\mathbf{X}) \odot \mathcal{K}(\mathbf{X})^{-\frac{1}{2}}, \tag{10.59}$$

where $\odot$ denotes point-wise multiplication of the matrices.

10.5.1.3 *Image recovery assuming smooth patch manifold*

Nonlocal means is an early and powerful algorithm for patch-based image denoising [12,3]. It estimates the Laplacian matrix from noisy data as in (10.47), followed by (10.41) to recover the denoised signals. Note that the evaluation of the $p^2 \times p^2$ patch matrix, as well as its use in (10.41), is computationally expensive. Several assumptions on the structure of the weight matrix (e.g., block diagonal assuming that the similar patches are in the spatial neighborhood) have been introduced to speed up the computations. The BM3D approach is also related, when the structure is determined by the Euclidean proximity of the patches.

Recently, some researchers have proposed to use the decomposition in (10.45) to further improve the denoising performance [69]. Each of the columns of $\mathbf{V}$ are termed as nonlocal basis functions, which shows the similarity between the regions. The corresponding $\mathbf{U}$ basis functions are termed as local basis functions. The spatial variation of the nonlocal basis functions (coefficients of the expansion) allows the subspace to be adapted to each patch, depending on the local neighborhood on the manifold.

When the recovery of image from undersampled Fourier measurements are considered, an alternating minimization scheme that minimizes (10.49) is adopted [68,67]. By making use of the redundancy between the patches, this approach is observed to offer good image quality during reconstruction. A similar strategy, where similarity between image patches in dynamic MRI reconstruction is used in PRICE [42], which offers implicit motion compensation. This approach has conceptual similarities to [13] and the kernel PCA approach in the patch setting [63]. Considering the improved performance offered by kernel PCA methods used in the dynamic MRI setting, we expect better performance with this scheme in the patch setting.

10.5.2 Application to dynamic MRI

The manifold structure of images in a dynamic time series is used to recover them from undersampled data [55,56,54,43]. In particular, the images in a free-breathing cardiac dataset can be viewed as nonlinear functions of the cardiac and respiratory phases; the images can be assumed as points on a low-dimensional surface.

A navigated strategy was used in [54]. In particular, each image is sampled by a sampling pattern that includes a common set of k-space locations, termed as navigators. Specifically, one would obtain $\mathbf{Z} = \mathbf{BX}$, where $\mathbf{B}$ is the sampling operator corresponding to the k-space navigators. In this case, one can approximate the $\mathbf{W}$ matrix as

$$\mathbf{W}_{i,j} \approx \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|^2}{\sigma_z^2}\right). \tag{10.60}$$

(a) Similarity to XD-GRASP

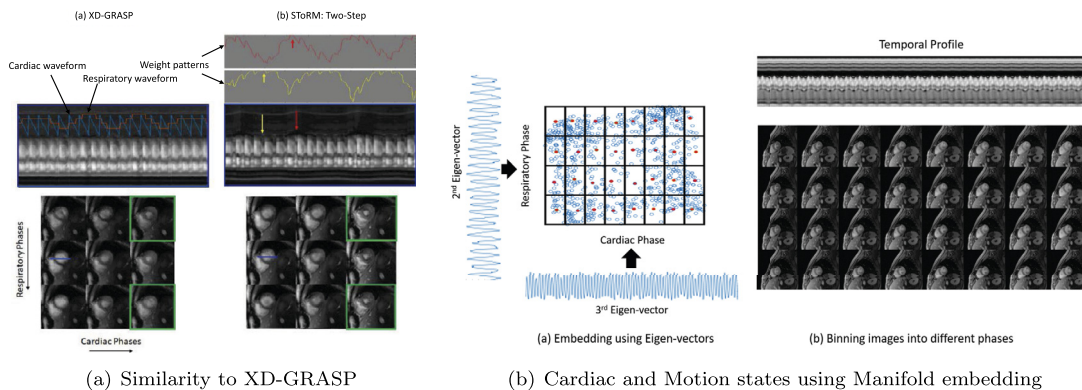(b) Cardiac and Motion states using Manifold embedding

FIGURE 10.8

Illustration of Smoothness Regularization on Manifolds (STORM) framework and its similarity to XD GRASP for the recovery of free-breathing and ungated MRI data. XD-GRASP bins the data to different cardiac/respiratory phases, followed by the joint recovery of the images. By contrast, STORM estimates a Laplacian matrix that has conceptual similarities to the XD-GRASP approach. The manifold Laplacian is estimated from the k-space navigators using (10.57) and (10.59). The reconstructed data is compared against self-gated XD-GRASP reconstruction of the same data. (**b**) uses the eigenvectors of the Laplacian matrix to bin the reconstructed data into cardiac and respiratory phases. This figure is copied from [56] with permission from IEEE.

Specifically, the navigators are expected to indicate the structure of the manifold.

A challenge with the direct implementation of the manifold-aware recovery in (10.43) in the multidimensional setting is its high computational complexity. The factorization approach is considered in (10.46) enables significant reduction in the computational complexity and memory demand of the algorithm [56]. In practice, 20–30 basis functions were observed to recover the dataset, which offers a 20–30 fold reduction in the memory demand. This approach also estimated the Laplacian matrix from the navigators using the kernel low-rank algorithm was used to estimate the Laplacian in [56], which offered improved performance over [54].

When navigators are not available, the formulation in (10.49) makes it possible to jointly estimate the Laplacian and the signals from the data itself [43]. A key benefit of this approach is that one can customize the Laplacian to different spatial regions. In particular, [43] split the images to patches, each with a different Laplacian. The kernel low-rank algorithm (10.55) is used instead of (10.49) to estimate the Laplacian in [1]. The iterative strategies including (10.49) and (10.55) are reported to yield far superior results compared to the proximity based methods. More studies are needed for the comparison of the iterative methods for the Laplacian matrix [1]. See Fig. 10.8 for an illustration of the manifold recovery. The results show that the generative STORM approach facilitates a significant reduction in acquisition time compared to the analysis counterparts.

10.6 Software

The MATLAB® software associated with this paper is available at https://github.com/sajanglingala/data_adaptive_recon_MRI.

10.7 Summary

This chapter reviewed several learning-based models that are used in MR image reconstruction, building upon compressed sensing methods in Chapter 6 and low-rank methods in Chapter 9. All the approaches reviewed in this chapter make use of the manifold structure of substructures (e.g., patches, pixel time series, images in the time series) of the dataset. The models differ in the representation of the data manifold, resulting in algorithms that rely on matrix factorization. In particular, the subregions are used to populate the columns of a structured matrix, which is factorized into two submatrices. Even though these approaches resemble low-rank methods reviewed extensively in Chapter 9, the main distinction is the nature of the priors used on the factors. The priors encourage the learning of basis functions and coefficients with specific properties, which often provide improved approximations of the data manifold. The matrix structure and basis functions promote the sharing of information between subregions of the dataset, thus facilitating the recovery of the dataset from highly undersampled data.

References

[1] Ahmed AH, Zhou R, Yang Y, Nagpal P, Salerno M, Jacob M. Free-breathing and ungated dynamic MRI using navigator-less spiral SToRM. IEEE Trans Med Imaging 2020:1. https://doi.org/10.1109/TMI.2020.3008329.

[2] Akçakaya M, Basha TA, Goddu B, Goepfert LA, Kissinger KV, Tarokh V, et al. Low-dimensional-structure self-learning and thresholding: regularization beyond compressed sensing for MRI Reconstruction. Magn Reson Med 2011;66. https://doi.org/10.1002/mrm.22841.

[3] Awate SP, Whitaker RT. Unsupervised, information-theoretic, adaptive image filtering for image restoration. IEEE Trans Pattern Anal Mach Intell 2006;28:364–76. https://doi.org/10.1109/TPAMI.2006.64. Available from: http://ieeexplore.ieee.org/document/1580482/.

[4] Balachandrasekaran A, Magnotta V, Jacob M. Recovery of damped exponentials using structured low rank matrix completion. IEEE Trans Med Imaging 2017. https://doi.org/10.1109/TMI.2017.2726995.

[5] Balachandrasekaran A, Mani M, Jacob M. Calibration-free B0 correction of EPI data using structured low rank matrix recovery. IEEE Trans Med Imaging 2019;38. https://doi.org/10.1109/TMI.2018.2876423.

[6] Belkin M, Niyogi P. Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput 2003. https://doi.org/10.1162/089976603321780317.

[7] Belkin M, Niyogi P, Sindhwani V. Manifold regularization: a geometric framework for learning from labeled and unlabeled examples. J Mach Learn Res 2006;7:2399–434. https://doi.org/10.1016/j.neuropsychologia.2009.02.028.

[8] Bhave S, Lingala SG, Johnson CP, Magnotta VA, Jacob M. Accelerated whole-brain multi-parameter mapping using blind compressed sensing. Magn Reson Med 2016. https://doi.org/10.1002/mrm.25722.

[9] Bhave S, Lingala SG, Newell JD, Nagle SK, Jacob M. Blind compressed sensing enables 3-dimensional dynamic free breathing magnetic resonance imaging of lung volumes and diaphragm motion. Invest Radiol 2016;51:387–99. https://doi.org/10.1097/RLI.0000000000000253.

[10] Biswas S, Aggarwal HK, Jacob M. Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM. Magn Reson Med 2019. https://doi.org/10.1002/mrm.27706.

[11] Brinegar C, Schmitter SS, Mistry NN, Johnson GA, Liang ZP. Improving temporal resolution of pulmonary perfusion imaging in rats using the partially separable functions model. Magn Reson Med 2010;64. https://doi.org/10.1002/mrm.22500.

[12] Buades A, Coll B, Morel JM. A non-local algorithm for image denoising. In: Proceedings - 2005 IEEE computer society conference on computer vision and pattern recognition, CVPR 2005. IEEE Computer Society; 2005. p. 60–5.

[13] Bustin A, Lima da Cruz G, Jaubert O, Lopez K, Botnar RM, Prieto C. High-dimensionality undersampled patch-based reconstruction (HD-PROST) for accelerated multi-contrast MRI. Magn Reson Med 2019;81(6). https://doi.org/10.1002/mrm.27694.

[14] Candès EJ, Fernandez-Granda C. Towards a mathematical theory of super-resolution. Commun Pure Appl Math 2014;67:906–56.

[15] Candes EJ, Wakin MB. An introduction to compressive sampling: a sensing/sampling paradigm that goes against the common knowledge in data acquisition. IEEE Signal Process Mag 2008;25:21–30. https://doi.org/10.1109/MSP.2007.914731.

[16] Dabov K, Foi A, Katkovnik V, Egiazarian K. Image denoising by sparse 3-D transform-domain collaborative filtering. IEEE Trans Image Process 2007;16:2080–95. https://doi.org/10.1109/TIP.2007.901238.

[17] Danielyan A, Katkovnik V, Egiazarian K. BM3D frames and variational image deblurring. IEEE Trans Image Process 2012. https://doi.org/10.1109/TIP.2011.2176954.

[18] Feng L, Axel L, Chandarana H, Block KT, Sodickson DK, Otazo R. Xd-grasp: golden-angle radial mri with reconstruction of extra motion-state dimensions using compressed sensing. Magn Reson Med 2016;75:775–88.

[19] Goud S, Hu Y, Jacob M. Real-time cardiac MRI using low-rank and sparsity penalties. In: 2010 7th IEEE international symposium on biomedical imaging: from nano to macro, ISBI 2010 - proceedings; 2010.

[20] Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, et al. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 2002;47:1202–10.

[21] Haldar JP. Low-rank modeling of local -space neighborhoods (LORAKS) for constrained MRI. IEEE Trans Med Imaging 2014;33:668–81. https://doi.org/10.1109/TMI.2013.2293974.

[22] Haldar JP. Low-rank modeling of local k-space neighborhoods (LORAKS) for constrained MRI. IEEE Trans Med Imaging 2014. https://doi.org/10.1109/TMI.2013.2293974.

[23] Haldar JP, Zhuo J. P-LORAKS: low-rank modeling of local k-space neighborhoods with parallel imaging data. Magn Reson Med 2016. https://doi.org/10.1002/mrm.25717.

[24] Jacob M, Mani MP, Ye JC. Structured low-rank algorithms: theory, magnetic resonance applications, and links to machine learning. IEEE Signal Process Mag 2020;37:54–68. https://doi.org/10.1109/MSP.2019.2950432.

[25] Jin KH, Lee D, Ye JC. A novel k-space annihilating filter method for unification between compressed sensing and parallel MRI. In: Proceedings - international symposium on biomedical imaging; 2015.

[26] Jin KH, Lee D, Ye JC. A general framework for compressed sensing and parallel MRI using annihilating filter based low-rank Hankel matrix. IEEE Trans Comput Imaging 2016;2. https://doi.org/10.1109/tci.2016.2601296.

[27] Kay SM, Marple SL. Spectrum analysis—a modern perspective. Proc IEEE 1981. https://doi.org/10.1109/PROC.1981.12184.

[28] Lai P, Lustig M, Vasanawala SS, Brau AC. ESPIRiT (efficient eigenvector-based L1SPIRiT) for compressed sensing parallel imaging - theoretical interpretation and improved robustness for overlapped FOV prescription. Electr Eng 2011;19.

[29] Lee D, Jin KH, Kim EY, Park SH, Ye JC. Acceleration of MR parameter mapping using annihilating filter-based low rank Hankel matrix (ALOHA). Magn Reson Med 2016. https://doi.org/10.1002/mrm.26081.

[30] Liang ZP. Spatiotemporal imaging with partially separable functions. In: 2007 4th IEEE international symposium on biomedical imaging: from nano to macro - proceedings; 2007. p. 988–91.

[31] Liang ZP, Lauterbur PC. A generalized series approach to MR spectroscopic imaging. IEEE Trans Med Imaging 1991;10. https://doi.org/10.1109/42.79470.

[32] Lingala SG, Dibella E, Adluru G, McGann C, Jacob M. Accelerating free breathing myocardial perfusion MRI using multi coil radial k - T SLR. Phys Med Biol 2013. https://doi.org/10.1088/0031-9155/58/20/7309.

[33] Lingala SG, Hu Y, Dibella E, Jacob M. Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR. IEEE Trans Med Imaging 2011;30:1042–54. https://doi.org/10.1109/TMI.2010.2100850. Available from: https://pubmed.ncbi.nlm.nih.gov/21292593/.

[34] Lingala SG, Jacob M. Blind compressed sensing with sparse dictionaries for accelerated dynamic MRI. In: Proceedings - international symposium on biomedical imaging; 2013.

[35] Lingala SG, Jacob M. Blind compressive sensing dynamic MRI. IEEE Trans Med Imaging 2013;32:1132–45. https://doi.org/10.1109/TMI.2013.2255133.

[36] Lobos RA, Kim TH, Hoge WS, Haldar JP. Navigator-free EPI ghost correction using low-rank matrix modeling: theoretical insights and practical improvements. In: Proceedings of the 25th annual meeting of ISMRM; 2017.

[37] Mani M, Aggarwal HK, Magnotta V, Jacob M. Improved MUSSELS reconstruction for high-resolution multi-shot diffusion weighted imaging. Magn Reson Med 2020;83. https://doi.org/10.1002/mrm.28090.

[38] Mani M, Jacob M. Fast iterative algorithm for the reconstruction of multishot non-Cartesian diffusion data. Magn Reson Med 2014. Available from: http://onlinelibrary.wiley.com/doi/10.1002/mrm.25486/full.

[39] Mani M, Jacob M, Kelley D, Magnotta V. Multi-shot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS). Magn Reson Med 2017;78. https://doi.org/10.1002/mrm.26382.

[40] Mani M, Magnotta V, Jacob M. A general algorithm for compensation of trajectory errors: application to radial imaging; 2018. Available from: http://doi.wiley.com/10.1002/mrm.27148.

[41] Mani M, Magnotta V, Kelley D, Jacob M. Comprehensive reconstruction of multi-shot multi-channel diffusion data using mussels. In: Proceedings of the annual international conference of the IEEE engineering in medicine and biology society, EMBS; 2016.

[42] Mohsin YQ, Lingala SG, DiBella E, Jacob M. Accelerated dynamic MRI using patch regularization for implicit motion compensation. Magn Reson Med 2017;77:1238–48. https://doi.org/10.1002/mrm.26215. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27091812. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5300957. http://doi.wiley.com/10.1002/mrm.26215.

[43] Mohsin YQ, Poddar S, Jacob M. Free-breathing ungated cardiac MRI using iterative SToRM (i-SToRM). IEEE Trans Med Imaging 2019;38. https://doi.org/10.1109/TMI.2019.2908140.

[44] Morrison RL, Jacob M, Do MN. Multichannel estimation of coil sensitivities in parallel MRI. In: 2007 4th IEEE international symposium on biomedical imaging: from nano to macro - proceedings; 2007.

[45] Nakarmi U, Slavakis K, Lyu J, Ying L. M-MRI: a manifold-based framework to highly accelerated dynamic magnetic resonance imaging. In: Proceedings - international symposium on biomedical imaging; 2017.

[46] Nakarmi U, Slavakis K, Ying L. MLS: joint manifold-learning and sparsity-aware framework for highly accelerated dynamic magnetic resonance imaging. In: Proceedings - international symposium on biomedical imaging; 2018.

[47] Nakarmi U, Wang Y, Lyu J, Liang D, Ying L. A kernel-based low-rank (KLR) model for low-dimensional manifold recovery in highly accelerated dynamic MRI. IEEE Trans Med Imaging 2017. https://doi.org/10.1109/TMI.2017.2723871.

[48] Ongie G, Biswas S, Jacob M. Convex recovery of continuous domain piecewise constant images from nonuniform Fourier samples. IEEE Trans Signal Process 2018;66:236–50. https://doi.org/10.1109/TSP.2017.2750111.

[49] Ongie G, Jacob M. Super-resolution MRI using finite rate of innovation curves. In: Proceedings - international symposium on biomedical imaging. IEEE Computer Society; 2015. p. 1248–51.

[50] Ongie G, Jacob M. Off-the-grid recovery of piecewise constant images from few Fourier samples. SIAM J Imaging Sci 2016;9:1004–41. https://doi.org/10.1137/15M1042280.

[51] Ongie G, Jacob M. A fast algorithm for convolutional structured low-rank matrix recovery. IEEE Trans Comput Imaging 2017;3:535–50. https://doi.org/10.1109/tci.2017.2721819.

[52] Ortega A, Frossard P, Kovacevic J, Moura JM, Vandergheynst P. Graph signal processing: overview, challenges, and applications. Proc IEEE 2018. https://doi.org/10.1109/JPROC.2018.2820126.

[53] Peyré G. Manifold models for signals and images. Comput Vis Image Underst 2009. https://doi.org/10.1016/j.cviu.2008.09.003.

[54] Poddar S, Jacob M. Dynamic MRI using SmooThness Regularization on Manifolds (SToRM). IEEE Trans Med Imaging 2016. https://doi.org/10.1109/TMI.2015.2509245.

[55] Poddar S, Mohsin Y, Ansah D, Thattaliyath B, Ashwath R, Jacob M. Free-breathing cardiac MRI using bandlimited manifold modelling. IEEE Trans Comput Imaging 2015;35:1106–15.

[56] Poddar S, Mohsin YQ, Ansah D, Thattaliyath B, Ashwath R, Jacob M. Manifold recovery using kernel low-rank regularization: application to dynamic imaging. IEEE Trans Comput Imaging 2019;3:478–91. https://doi.org/10.1109/tci.2019.2893598. Available from: https://ieeexplore.ieee.org/document/8625515/.

[57] Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. IEEE Trans Med Imaging 2011;30:1028–41. https://doi.org/10.1109/TMI.2010.2090538.

[58] Ravishankar S, Bresler Y. Learning sparsifying transforms. IEEE Trans Signal Process 2013;61:1072–86. https://doi.org/10.1109/TSP.2012.2226449.

[59] Ravishankar S, Bresler Y. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging. SIAM J Imaging Sci 2015;8:2519–57. https://doi.org/10.1137/141002293.

[60] Ravishankar S, Ye JC, Fessler JA. Image reconstruction: from sparsity to data-adaptive methods and machine learning. Proc IEEE 2020;108. https://doi.org/10.1109/JPROC.2019.2936204.

[61] Recht B, Fazel M, Parrilo PA. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. SIAM Rev 2010;52:471–501.

[62] Roweis ST, Saul LK. Nonlinear dimensionality reduction by locally linear embedding. Science 2000. https://doi.org/10.1126/science.290.5500.2323.

[63] Schmidt JF, Santelli C, Kozerke S. Mr image reconstruction using block matching and adaptive kernel methods. PLoS ONE 2016;11:e0153736.

[64] Shetty GN, Slavakis K, Bose A, Nakarmi U, Scutari G, Ying L. Bi-linear modeling of data manifolds for dynamic-MRI recovery. IEEE Trans Med Imaging 2020. https://doi.org/10.1109/TMI.2019.2934125.

[65] Shin PJ, Larson PE, Ohliger MA, Elad M, Pauly JM, Vigneron DB, et al. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion. Magn Reson Med 2014. https://doi.org/10.1002/mrm.24997.

[66] Wang Y, Ying L. Undersampled dynamic magnetic resonance imaging using kernel principal component analysis. In: 2014 36th annual international conference of the IEEE engineering in medicine and biology society, EMBC 2014; 2014.

[67] Yang Z, Jacob M. Robust non-local regularization framework for motion compensated dynamic imaging without explicit motion estimation. In: Proceedings - international symposium on biomedical imaging. NIH Public Access; 2012. p. 1056–9. Available from: /pmc/articles/PMC3956771/?report=abstract, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3956771/.

[68] Yang Z, Jacob M. Nonlocal regularization of inverse problems: a unified variational framework. IEEE Trans Image Process 2013;22:3192–203. https://doi.org/10.1109/TIP.2012.2216278.

[69] Yin R, Gao T, Lu YM, Daubechies I. A tale of two bases: local-nonlocal regularization on image patches with convolution framelets. SIAM J Imaging Sci 2017. https://doi.org/10.1137/16M1091447.

[70] Zhao B, Haldar JP, Liang ZP. PSF model-based reconstruction with sparsity constraint algorithm and application to real-time cardiac MRI. In: 2010 annual international conference of the IEEE engineering in

medicine and biology society, EMBC'10. NIH Public Access; 2010. p. 3390–3. Available from: /pmc/articles/PMC3121182/?report=abstract, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3121182/.

[71] Zou Q, Ahmed AH, Nagpal P, Kruger S, Jacob M. Dynamic imaging using a deep generative SToRM (Gen-SToRM) model. IEEE Trans Med Imaging 2021 Nov;40(11):3102–12. https://doi.org/10.1109/TMI.2021.3065948. Epub 2021 Oct 27. PMID: 33720831; PMCID: PMC8590205.

[72] Zou Q, Ahmed AH, Nagpal P, Priya S, Schulte R, Jacob M. Generative storm: a novel approach for joint alignment and recovery of multi-slice dynamic mri. Available from: https://arxiv.org/abs/2101.08196.