

# Dictionary, Structured Low-rank, and Manifold learning based reconstruction

Mathews Jacob, Sajan Goud Lingala, and Merry Mani

---

## Abstract

This chapter reviews machine learned image representations for accelerated MRI. Unlike prior approaches that use fixed models for image recovery, recent methods relies on advances in machine learning to adapt the representation to the data. The improved representation translates to quite significant gains in acceleration. We review global low-rank representations, local low-rank methods that approximate groups of signals by subspaces, dictionary learning methods, and smooth manifold based approaches. All of the schemes learn the representation either from exemplar data, calibration data, or jointly learn the representation and images from undersampled data. We also review applications of these frameworks to static and dynamic imaging applications.

---

## 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, which were often carefully engineered to the data. In contrast, learning the low dimensional representations from data itself offers improved efficiency. For instance, 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*, which 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 be broadly classified as

1. approaches that capitalize the complex *non-linear redundancies* with the datasets, which low-rank methods are not capable of capitalizing on. Sparse dictionary learning and smooth manifold representations fall in 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 non-linear kernel priors, allowing us to account for non-linear redundancies in the data resulting from motion and contrast changes. Please see Fig. 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 allows one to capitalize 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

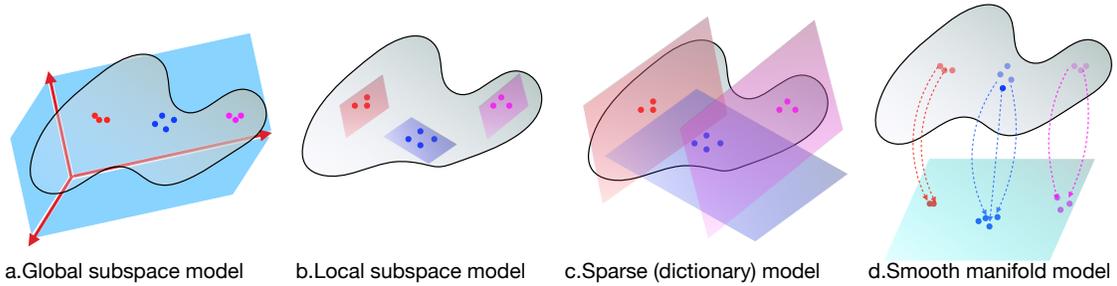


Figure 1: Overview of signals extracted from the image volumes, whose redundancy is capitalized in this chapter: The signals (images/patches/timeprofiles) in MRI datasets have extensive redundancy as described in Fig. 2; they can be viewed as points lying on a manifold in high dimensional spaces. The 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 non-linear manifold to a low-dimensional subspace denoted by the plane; the structure of the low-dimensional subspace is used to recover the signals.

compressed sensing algorithms. Please see Fig. 2 for an illustration of the signals extracted from the images, whose redundancies are leveraged to recover the image.

The chapter is organized as follows. Following a brief overview of the background in Section 2, we review dictionary based methods in Section 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 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 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 4. We review smooth manifold models, which are efficient in capturing non-linear redundancies in the dataset in Section 5. An overview and broad classification of the methods reviewed in this chapter is given in Table 1. In this chapter, the different topics are not presented in the chronological order in which they were introduced in the MRI setting. Rather, our focus is on grouping the different methods into broader themes to facilitate easy comprehension of the links and the generalizations of the various approaches.

## 2. Background

### 2.1. Acquisition scheme

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

$$s_{i,j} = \int x(\mathbf{r})c_j(\mathbf{r}) \exp(-j\mathbf{k}_i^T \mathbf{r})d\mathbf{r} + \eta_{i,j} \quad (1)$$

Here,  $c_j(\mathbf{r})$  are the sensitivity profiles of the  $j^{\text{th}}$  coil and  $\mathbf{k}_i$  is the location in k-space, while  $\eta$  denotes noise. The measurement process described in (1) can be compactly represented as

$$\mathbf{s} = \mathbf{E}(\mathbf{x}), \quad (2)$$

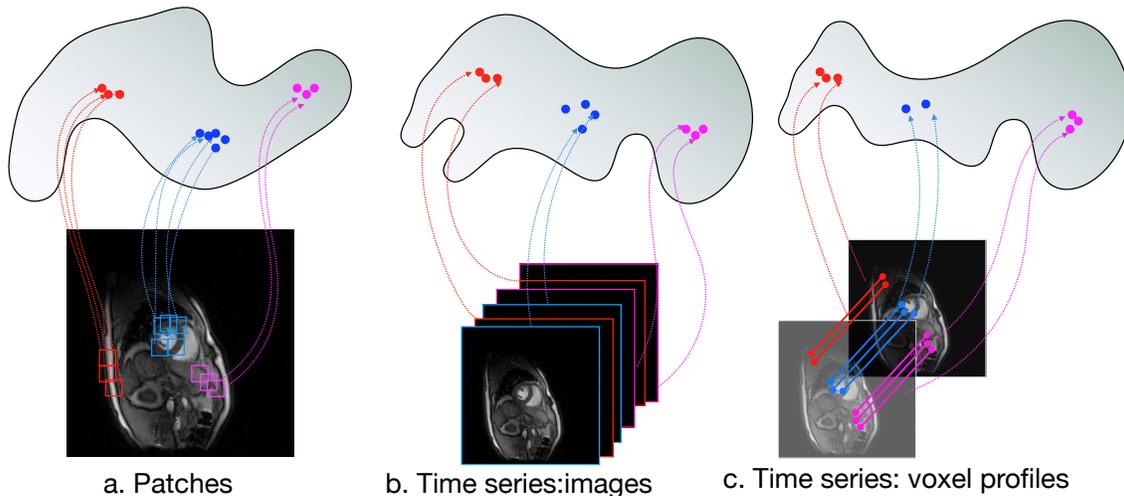


Figure 2: Types of signals used by different algorithms: the MR methods reviewed in this chapter account for the redundancy within the image volumes, which often manifest as correlations between image sub-parts that we broadly refer to as patches. Depending upon the shape of the patches, they could be 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 non-local 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 accounted 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.

where  $\mathbf{E}$  is the 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 (eg: dynamic imaging applications), the acquisition model correspond to

$$s_{i,j}(t) = \int x(\mathbf{r}, t) c_j(\mathbf{r}) \exp(-j\mathbf{k}_i^T \mathbf{r}) d\mathbf{r}, \quad (3)$$

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  $\hat{\mathbf{x}}$  to denote the discrete Fourier coefficients of the signal on a Cartesian grid. Note that the multichannel measurements in (1) need not be in the Cartesian domain.

## 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 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 linearly increasing with the same slope, all the three pixel images will fall on a straight line passing through the origin. When modeling larger 1-D images, one may extract patches consisting of three consecutive pixels from the image, each of which can be viewed as a point on the above line. This approach can be viewed as a global subspace model for the image. As discussed in the next section, the above patch signals can be collected into a  $3 \times m$  matrix, whose columns are the patches. The above global linear model is restrictive and may not approximate real world 1-D 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 live 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 global subspace model considered above.

The above 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 more complex inter-dependencies between the pixels, beyond the piecewise models discussed above. 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, 56].

The above idea can be extended to patches in 2D images, the pixels within the patches may have non-linear relations 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. 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 the unique redundancies in each application.

### 2.3. Capitalization of structure 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  $\mathcal{T}(\mathbf{x})$ :

$$\mathcal{T}(\mathbf{x}) = [P_1(\mathbf{x}) \quad P_2(\mathbf{x}) \quad \dots \quad P_n(\mathbf{x})] \quad (4)$$

Here,  $P_1, \dots, P_n$  are patch extraction operators, which extract a patch from image and convert it to 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  $i^{\text{th}}$  patch extraction operator is chosen as the  $i^{\text{th}}$  image in a time series  $P_i(x(\mathbf{r}, t)) = x(\mathbf{r}, i); i = 1, \dots, T$ , we obtain

$$\mathcal{T}(\mathbf{x}) = [x(\mathbf{r}, 1) \quad \dots \quad x(\mathbf{r}, T)], \quad (5)$$

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. 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. 6. Hence, the methods that use the low-rank property of these structured matrices are called structured low-rank (SLR) methods [68, 21, 40, 52, 23, 30, 26, 54, 43]. 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  $\mathbf{x} \rightarrow \mathcal{T}(\mathbf{x})$  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.

#### 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

$$\mathcal{T}(\mathbf{x}) = \mathbf{\Phi}\mathbf{W}^T, \quad (6)$$

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  $r$ . 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.  $\ell_2$  or  $\ell_1$  norms) on the factors discussed below 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 above 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:

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \lambda \|\mathcal{T}(\mathbf{x})\|_* \quad (7)$$

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 [64]:

$$\|\mathcal{T}(\mathbf{x})\|_* = \|\mathbf{\Phi}\|_F^2 + \|\mathbf{W}\|_F^2, \text{ where } \mathcal{T}(\mathbf{x}) = \mathbf{\Phi}\mathbf{W}^T, \quad (8)$$

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:

$$\{\mathbf{\Phi}^*, \mathbf{W}^*\} = \arg \min_{\mathbf{\Phi}, \mathbf{W}} \|\mathbf{E}(\mathbf{\Phi}\mathbf{W}^T) - \mathbf{s}\|^2 + \lambda (\|\mathbf{\Phi}\|_F^2 + \|\mathbf{W}\|_F^2), \quad (9)$$

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 above low-rank methods.

### 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. 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. State-of-the art patch based methods such as BM3D [17, 16] hence 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. 1.(b). A challenge with 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, 70, 71, 72] extend this approach to image reconstruction applications. These methods either estimate the clusters from zero-filled MRI data [2, 70], or alternate between clustering and recovery of images [71, 72]. More information on the alternating scheme is provided in Section 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.

### 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. 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), by only allowing only a few coefficients of the representation to be non-zero. For instance, the sparse model [15] represents the signal as

$$x(\mathbf{r}) \approx \sum_{i=1}^K w_i \varphi_i(\mathbf{r}) = \underbrace{[\varphi_1(\mathbf{r}) \quad \varphi_2(\mathbf{r}) \quad \dots \quad \varphi_K(\mathbf{r})]}_{\Phi} \underbrace{\begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_K \end{bmatrix}}_{\mathbf{w}^T}, \text{ where } \|\mathbf{w}\|_{\ell_0} \leq k \quad (10)$$

Here,  $\|\mathbf{w}\|_{\ell_0}$  denotes the number of non-zero 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)  $\Phi$ . Rather than using fixed dictionaries, several authors have proposed to adapt or learn the dictionaries or transforms from the data [60, 35, 36, 61, 62, 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 [60, 61, 8, 63] or from a single undersampled dataset [35, 36, 62, 8] in a joint manner. The first approach is termed pre-learning, while the joint learning of the dictionary and coefficients from the data is termed blind compressed sensing. Since the  $\ell_0$  norm is not convex, a common approach is to approximate it with the  $\ell_1$  norm that is convex [15]. Rather than employing dictionaries, the use of analysis operators  $\Psi$  is also a common approach.

### 3.2. Dictionary pre-learning

In pre-learning, the learning of the dictionary from a family of fully sampled signals  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N]$  is posed as an optimization problem [60, 61]:

$$\{\Phi, \mathbf{W}\} = \arg \min_{\Phi, \mathbf{W}} \|\mathbf{X} - \Phi \mathbf{W}^T\|^2 + \lambda_1 \|\mathbf{W}\|_{\ell_0} + \lambda_2 \mathcal{R}(\Phi), \quad (11)$$

where  $\mathcal{R}(\Phi)$  is a regularization penalty on the dictionary atoms. A simple choice for  $\mathcal{R}(\Phi) = \|\Phi\|_F^2$  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 low condition number, or be a combination of orthogonal transforms were introduced by several authors [60, 35, 36, 61, 62, 8].

Note that the above problem in (11) is convex in  $\mathbf{W}$  if  $\Phi$  is known. Likewise, for many of the common choices of  $\mathcal{R}$ , the optimization of  $\Phi$  assuming  $\mathbf{W}$  to be known, is also a simple problem. For

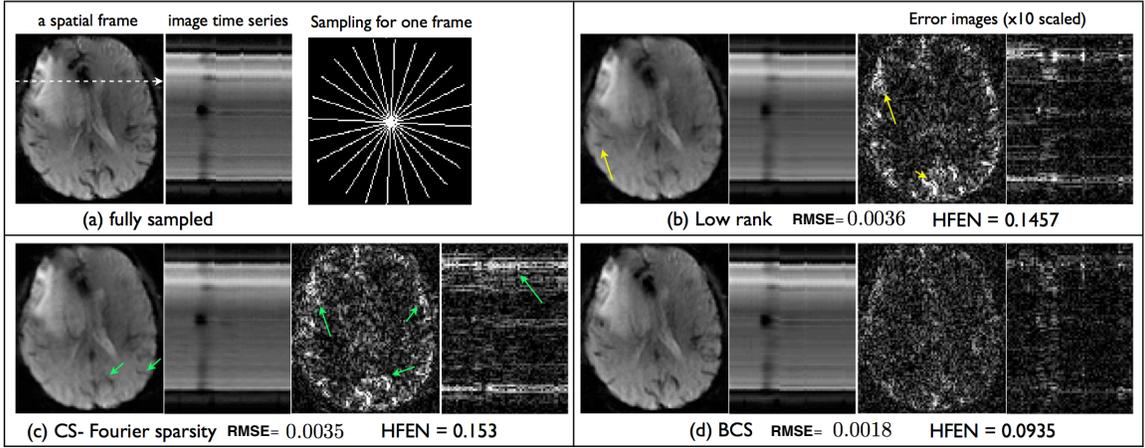


Figure 3: Comparison of low rank, Fourier sparsity, and Blind compressed sensing (BCS) reconstructions on a brain perfusion MRI dataset retrospectively under sampled at an under sampling factor of 10 fold. BCS is shown to provide superior spatial and temporal fidelity in characterizing the contrast agent temporal dynamics compared to low rank and CS reconstructions.

instance, when  $\mathcal{R}(\Phi) = \|\Phi\|_F^2$  is the Frobenius norm, the solution of  $\Phi$  is a quadratic problem that has an analytical solution. However, the joint optimization of  $\Phi$  and  $\mathbf{W}$  is a non-convex problem. Nevertheless, this problem has been well studied by several researchers, especially when  $\Phi$  is a transform [60, 35, 36, 61, 62, 8]. The learning of the transform to the class of signals results in improved performance over the use of standard transforms such as wavelet transform.

### 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 be sparse in a dictionary. Here the signal matrix in (11) is the patch matrix  $\mathbf{X} = \mathcal{T}(\mathbf{x})$  is of dimension  $p^2 \times n$ . The ability to choose the non-zero coefficients for each patch facilitates the use of the same dictionary for the entire image. The above formulation can be extended to learning from multiple images by horizontally stacking the patch matrices as  $[\mathcal{T}(x_1), \dots, \mathcal{T}(x_N)]$ . 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 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  $\Phi$  is learned from multiple images, it can be used to recover images from undersampled measurements [60] as

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \alpha \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \|\mathcal{T}(\mathbf{x}) - \Phi \mathbf{W}^T\|^2 + \lambda \|\mathbf{W}\|_{\ell_p}, \quad (12)$$

which is the extension of compressive sensing (CS) to image patches. The second and third term encourages the patches in the solution to be sparse linear combination of atoms in  $\Phi$ , 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  $\alpha$ .

### 3.3. Blind compressive sensing (BCS)

BCS schemes estimate the dictionary and the coefficients directly from the measured undersampled 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.

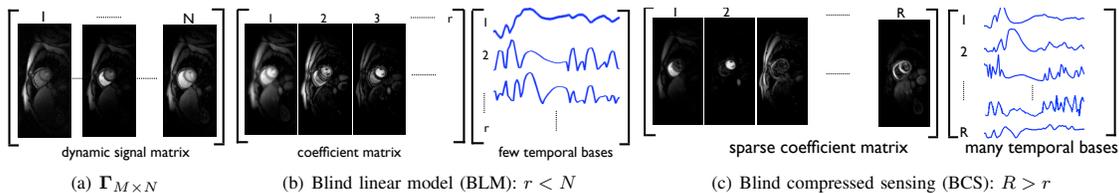


Figure 4: Comparison of blind compressed sensing (BCS) and low-rank (blind linear model) representations of dynamic imaging data: The Casorati form of the dynamic signal is shown in (a). The BLM and BCS decompositions are respectively shown in (b) and (c). BCS uses a large over-complete dictionary, unlike the orthogonal dictionary with few basis functions in BLM; ( $R > r$ ). Note that the coefficients/spatial weights in BCS are sparse 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 1st, 2nd 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 [36] with permission from IEEE.

### 3.3.1. Application of BCS to dynamic MRI

In the context of dynamic MRI, blind compressed sensing methods learn the bilinear model in (11) directly from undersampled data [35, 36, 62, 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  $\mathbf{X} = \Phi \mathbf{W}^T$  amounts to expressing the temporal profiles of each pixel as a linear combination of the columns of the dictionary  $\Phi$ . 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. 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 temporal basis functions chosen 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  $\Phi$  and its coefficients  $\mathbf{W}$  are hence learned directly from the undersampled dataset itself as

$$\{\Phi, \mathbf{W}\} = \arg \min_{\Phi, \mathbf{W}} \|\mathbf{E}(\Phi \mathbf{W}^T) - \mathbf{s}\|^2 + \lambda_1 \|\mathbf{W}\|_{\ell_0} + \lambda_2 \mathcal{R}(\Phi), \quad (13)$$

The use of blind compressed sensing scheme in (13) further improves the quality of dynamic MRI reconstructions [35, 36, 8] compared to linear (low-rank) models. Specifically, low-rank models uses the same basis functions for the voxel profiles of each pixel. The projection of the time series to the signal subspace results in non-local 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 non-zero coefficients) in sparse models can potentially differ from 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 (eg. heart, lung). An illustration of the blind compressed sensing (BCS) approach is shown in Fig. 3, where it is compared against global low-rank methods and approaches that use sparse models with fixed (Fourier) dictionary.

### 3.3.2. Application of BCS to static imaging

The blind compressed sensing formulation in (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 [62] as:

$$\{\mathbf{x}, \mathbf{W}, \Phi\} = \arg \min_{\mathbf{x}, \mathbf{W}, \Phi} \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \alpha \|\mathcal{T}(\mathbf{x}) - \Phi \mathbf{W}^T\|^2 + \lambda_1 \|\mathbf{W}\|_{\ell_p} + \lambda_2 \mathcal{R}(\Phi) \quad (14)$$

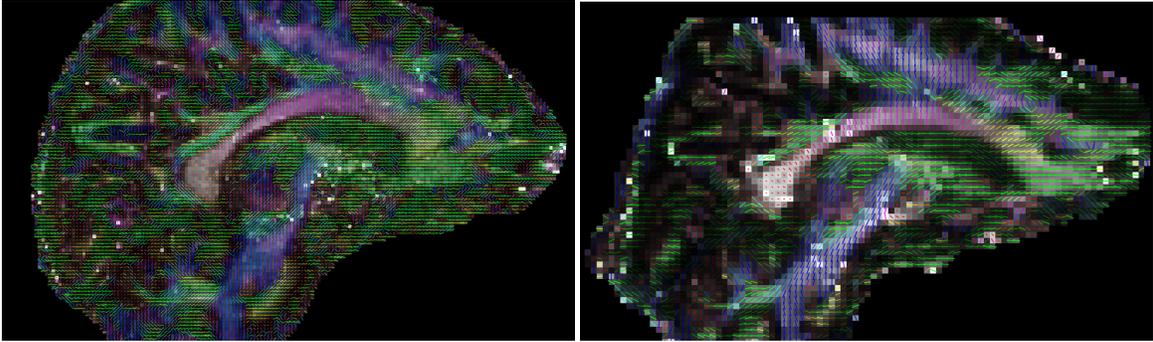


Figure 5: High resolution diffusion MRI enabled by the patch low-rank methods in k-space. The image on the left side is acquired using a 4-shot diffusion weighted EPI scan at 1.1mm isotropic resolution on 7 Tesla MRI with a standard clinical strength gradient of 40 mTesla/m. For comparison, a typical 2mm isotropic resolution single-shot diffusion MRI scan slice-matched from the same study, is provided on the right. The multi-shot data on the left is reconstructed using the MUSSELS method which 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.

Here, the first term is the data-consistency term, which 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  $\mathcal{T}(\mathbf{x}) = \Phi \mathbf{W}^T$ , where the coefficient matrix  $\mathbf{W}$  is sparse and the dictionary is compact under a prior  $\mathcal{R}(\Phi)$  such as  $\|\Phi\|_{\ell_2}$  or  $\|\Phi\|_{\ell_1}$ . Rather than introducing the factorization as a constraint, the formulation in (14) relies on a penalty term; when  $\alpha \rightarrow \infty$ , the solution will satisfy  $\mathcal{T}(\mathbf{x}) = \Phi \mathbf{W}^T$ .

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

##### 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 [29], simultaneous autocalibrating and k-space estimation (SAKE) [68], 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 [30, 26, 27, 25], structured low-rank (SLR) [52, 54, 50] and Multi-shot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS) [40, 43]. We now briefly review some of the low-rank relations resulting from specific signal properties.

##### 4.1.1. Low rank relations in multi-channel MRI

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

$$x_i(\mathbf{r}) = x(\mathbf{r}) c_i(\mathbf{r}), \quad i = 1, \dots, N_{\text{channels}} \quad (15)$$

where  $c_i(\mathbf{r})$  is the sensitivity weighting of the  $i^{\text{th}}$  receiver coil. The multichannel relations specified by (15) results in image domain annihilation relations [46]

$$\underbrace{x(\mathbf{r})c_1(\mathbf{r})}_{x_1(\mathbf{r})}c_2(\mathbf{r}) - \underbrace{x(\mathbf{r})c_2(\mathbf{r})}_{x_2(\mathbf{r})}c_1(\mathbf{r}) = 0. \quad (16)$$

One can take the Fourier transforms of both sides of the above equation to obtain [29, 68]

$$\widehat{x}_1 * \widehat{c}_2 - \widehat{x}_2 * \widehat{c}_1 = 0 \quad (17)$$

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

$$\mathbf{c} * x = \mathbf{c}^T \mathcal{T}(\mathbf{x}), \quad (18)$$

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 (17) as

$$\mathbf{c}_2^T \mathcal{T}(\widehat{\mathbf{x}}_1) - \mathbf{c}_1^T \mathcal{T}(\widehat{\mathbf{x}}_2) = 0 \quad (19)$$

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

$$\underbrace{\begin{bmatrix} \widehat{\mathbf{c}}_2^T & -\widehat{\mathbf{c}}_1^T & 0 & \dots \\ \widehat{\mathbf{c}}_3^T & 0 & -\widehat{\mathbf{c}}_1^T & \dots \\ \vdots & \vdots & \ddots & \dots \\ \widehat{\mathbf{c}}_{N_c}^T & 0 & 0 & \dots \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \mathcal{T}(\widehat{\mathbf{x}}_1) \\ \mathcal{T}(\widehat{\mathbf{x}}_2) \\ \vdots \\ \mathcal{T}(\widehat{\mathbf{x}}_{N_c}) \end{bmatrix}}_{\mathcal{M}(\widehat{\mathbf{X}})} = \mathbf{0}. \quad (20)$$

Note that each of the rows of  $\mathbf{P}$  are null-vectors of  $\mathcal{M}(\widehat{\mathbf{X}})$ , which are linearly independent. Hence, the matrix  $\mathcal{M}(\widehat{\mathbf{X}})$  is low-rank [29, 68]. Here  $\widehat{\mathbf{X}} = [\widehat{\mathbf{x}}_1, \dots, \widehat{\mathbf{x}}_{N_c}]$  is the multichannel data in the Fourier domain. The above multichannel convolution relations can also be rewritten as

$$\widehat{\mathbf{X}} = \underbrace{(\mathcal{I} - \mathbf{P})}_{\mathbf{G}} \mathcal{M}(\widehat{\mathbf{X}}). \quad (21)$$

where the operator  $\mathcal{I}$  in (21) extracts  $\widehat{\mathbf{X}}$  from  $\mathcal{M}(\widehat{\mathbf{X}})$  (*i.e.*,  $\mathcal{I}\{\mathcal{M}(\widehat{\mathbf{X}})\} = \widehat{\mathbf{X}}$ ). The above relation forms the basis of the auto-calibrating parallel MRI reconstruction method, SAKE [68], which interpolates the missing k-space samples of the accelerated acquisition, based on the structured low-rank property. The SAKE relation in (21) can be viewed as a generalization of GRAPPA reconstruction method for multi-channel MRI. [20].

The formulations in (16)-(20) is general enough to be applied to settings beyond multi-channel MRI. Researchers have used these relations in several contexts such as the calibration-less compensation of phase errors in multichannel diffusion MRI [41, 39], correction of Nyquist ghost artifacts in echo-planar imaging [43, 37], and correction of trajectory errors in radial MRI [42]. In all of these cases, different segments of k-space experience different phase errors [41, 43, 37]. These phase errors can be modeled as image domain weighting (similar to the coil sensitivity weighting studied above), allowing the use of the above relations (16)-(20), to derive the structured data matrix with the low-rank property. Figure 5 shows the application of the patch low-rank idea for the reconstruction of high resolution diffusion MRI data from multi-shot acquisitions, where the phase compensation of the multi-shot data were achieved in a calibration-less manner using patch low-rank.

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 relations 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 relations of the form

$$x(\mathbf{r}) \cdot f_i(\mathbf{r}) = 0 \quad (22)$$

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

$$\widehat{x}(\mathbf{r}) * \widehat{f}_i(\mathbf{r}) = 0 \quad (23)$$

resulting in annihilation relations 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 relation in (22). The above relation implies that

$$\mathcal{T}(\widehat{\mathbf{x}}) \mathbf{F} = 0. \quad (24)$$

or equivalently  $\mathcal{T}(\mathbf{x})$  is a low-rank matrix. Haldar et. al has empirically showed 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 under-sampled recovery.

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

$$\widehat{x}[k] - \widehat{x}[-k]^* = 0. \quad (25)$$

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 relations and hence is low-rank. These results translate to structured low-rank algorithms that account for the above patch low-rank structure.

4.1.3. *Low-rank structure resulting from continuous domain sparsity*

Here, we discuss the low-rank relations 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

$$x(r) = \sum_{i=1}^R \gamma_i \delta(r - r_i), \quad (26)$$

for sparse images, where  $\gamma_i$  are the weights and  $r_i$  are the location of the impulses, which are not necessarily on a uniform grid. The SLR [53] and ALOHA settings [26] extends the Fourier domain annihilation relations discussed in the previous sections for the recovery of continuous domain sparse signals.

To see how the above sparse image model can harness the Fourier domain annihilation relations, we first discuss a simple 1-D case. The seminal work by Prony dating back to 1885 showed that 1-D exponential signal of the form  $\widehat{x}[k] = \alpha^k$  can be annihilated by convolution as follows [28] :

$$\widehat{x}[k] * h[k] = \underbrace{\alpha^k}_{\widehat{x}[k]} - \alpha \underbrace{\alpha^{k-1}}_{\widehat{x}[k-1]} = 0, \quad (27)$$

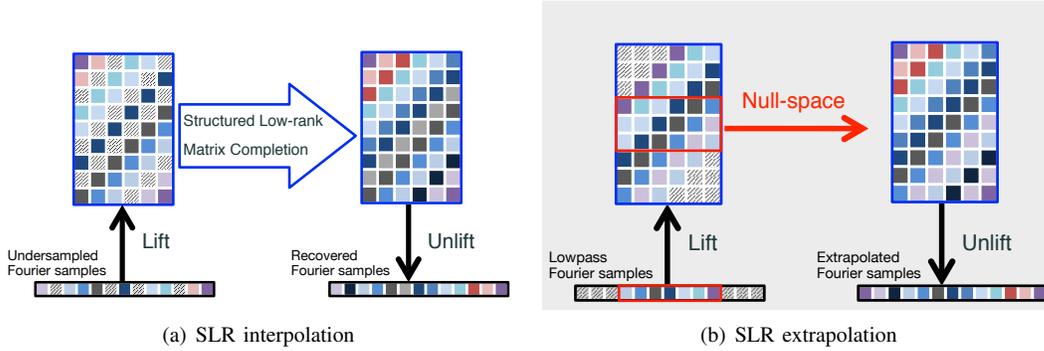


Figure 6: Illustration of SLR-based interpolation and extrapolation methods in the context of 1-D FRI. (a) In SLR interpolation, the data is acquired on a non-uniformly sub-sampled Fourier grid. The SLR interpolation scheme relies on a lifting of the signal samples to a Hankel matrix, which has missing entries indicated by the hashed boxes. The one-to-one relation 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.

where  $h[k]$  is a two tap filter given by  $[1, -\alpha]$ . This theory is relevant to MR images because when  $\alpha = \exp(jr_0)$ ,  $\hat{x}[k] = \exp(jr_0 k)$  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  $r_0, \dots, r_k$ , 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 above 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 relation in (27) for  $k$  impulses for the super-resolution recovery can thus be compactly expressed as [28, 21, 52, 26]

$$\mathbf{h}^T \underbrace{\begin{bmatrix} \hat{x}[0] & \hat{x}[1] & \dots & \hat{x}[N-k] \\ \hat{x}[1] & \hat{x}[2] & \dots & \hat{x}[N-k+1] \\ \vdots & \vdots & \ddots & \vdots \\ \hat{x}[k] & \hat{x}[1] & \dots & \hat{x}[N] \end{bmatrix}}_{\mathcal{T}(\hat{\mathbf{x}})} = \mathbf{0} \quad (28)$$

Note that  $\mathcal{T}(\hat{\mathbf{x}})$  is a patch matrix obtained by lifting the 1-D signal  $\hat{x}[n]$ , which are the Fourier coefficients of  $\mathbf{x}$ . The columns of  $\mathcal{T}(\hat{\mathbf{x}})$  correspond to  $(k+1) \times 1$  patches in  $\mathbf{x}$ . (28) implies that the matrix  $\mathcal{T}(\hat{\mathbf{x}})$  has a null-space vector. If the number of impulses is  $K' < k$ , the above matrix will have  $k - K' + 1$  null-space vectors [28, 26]. In other words, the rank of the matrix  $\mathcal{T}(\hat{\mathbf{x}})$  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 (27), we obtain

$$\hat{h}(r) \cdot x(r) = 0, \quad (29)$$

where  $\hat{h}(r)$  and  $x(r)$  the inverse Fourier transforms of  $h[k]$  and  $\hat{x}[k]$ , respectively. Note that the convolution in (27) is translated to the point-by-point multiplication [52, 53]. For the example

considered in (27), we get

$$\mu(r) = 1 - \underbrace{\exp(jr_0)}_{\alpha} \exp(-jr) = 1 - \exp(-j(r - r_0)) \quad (30)$$

Note that this exponential is a first order polynomial that is non-zero 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 non-zero locations of  $x(r)$ . This interpretation will be useful in the next section for the discussion of piecewise smooth images.

#### 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 more sparse [52] 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 1-D signal  $\widehat{\partial_{r_1} x} = jk_{r_1} \widehat{x}$ . Thus, replacing  $\widehat{x}$  by  $\widehat{\partial_{r_1} f}$  will result in a matrix with a significantly smaller rank.

The generalization of the above idea to multiple dimensions is not straightforward from a theoretical perspective. Specifically, the gradient of a piecewise constant image is non-zero 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 2-D bandlimited function  $\mu(\mathbf{r})$  [52, 53]. This model amounts to stating that the piecewise constant image has smooth edge contours. In this case, we have

$$\mu(\mathbf{r}) \cdot \underbrace{[\partial_{r_1} x(\mathbf{r}) \quad \partial_{r_2} x(\mathbf{r})]}_{\nabla x(\mathbf{r})} = 0 \quad (31)$$

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

$$\mathcal{T}_2(\hat{\mathbf{x}}) = \begin{bmatrix} \mathcal{T}(\widehat{\partial_{r_1} x}) & \mathcal{T}(\widehat{\partial_{r_2} x}) \end{bmatrix}, \quad (32)$$

where  $\mathcal{T}(\widehat{\partial_{r_1} x})$  is the 2-D patch matrix of the Fourier coefficients of the partial derivative of  $\mathbf{x}$ . The same approach can be extended to piecewise polynomials by replacing (32) with a matrix with more partial derivatives. The number of rows is 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. Theoretical results show that the best performance is obtained when the matrix  $\mathcal{T}_2(\widehat{\partial_{r_1} x})$  is square shaped, which roughly correspond 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.

#### 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 pixels change in an exponential fashion (e.g.  $\rho[\mathbf{r}, \mathbf{n}] = \alpha[\mathbf{r}]^n + c$ , where  $c$  and  $\alpha$  are arbitrary constants). Such a signal can be annihilated by a finite difference filter, whose parameters depend on  $\alpha$ . 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 (27), we see that such exponential signals satisfy an annihilation relation. 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 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].

#### 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 relations, which translate to low-rank relations 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 relations is posed as an optimization problem [22, 52, 53, 26]. We will first discuss the multi-channel parallel MRI case where the recovery of the images from the under-sampled measurements can be written as an unconstrained optimization

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \lambda \|\mathcal{T}(\hat{\mathbf{x}})\|_*, \quad (33)$$

Here,  $\mathcal{T}(\hat{\mathbf{x}})$  is the lifted structured matrix formed from the multi-channel convolution relations in (20). The above 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. Figure 6 shows a schematic of this approach where the above 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 (33), as described in [22, 52, 53, 26].

#### 4.3. Iterative reweighted least square (IRLS) algorithm

The IRLS scheme relies on the approximation of the nuclear norm [54]:

$$\|\mathcal{T}_2(\hat{\mathbf{x}})\|_* \approx \|\mathbf{Q} \mathcal{T}(\hat{\mathbf{x}})\|^2, \quad (34)$$

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

$$\mathbf{Q} = (\mathcal{T}(\hat{\mathbf{x}})\mathcal{T}(\hat{\mathbf{x}})^T)^{-1/4} \quad (35)$$

The main benefits of the IRLS algorithm [54] is that it significantly reduces the computational complexity and the memory demand. In addition, we use this algorithm some of the existing SLR methods in Section 4.4 and to to connect kernel PCA with manifold methods in Section 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 (34) can be viewed as the energy of the projection of  $\mathcal{T}(\mathbf{x})$  on to the null-space.

Using the above approximation and the structure of the matrix  $T(\mathbf{x})$ , one can solve (28) by alternating between (35) and

$$f^* = \arg \min_f \|\mathbf{E}(\mathbf{x}) - \mathbf{s}\|^2 + \lambda \sum_{i=1}^{p^2} \|\hat{\mathbf{x}} * \mathbf{q}_i\|^2, \quad (36)$$

where  $\mathbf{q}_i$  are the columns of  $\mathbf{Q}$  [54]. The main benefit of (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]

$$\|\mathcal{T}_2(\hat{\mathbf{x}})\|_* \approx \arg \min_{\mathbf{U}, \mathbf{V}}, \|\mathbf{U}\|_F^2 + \|\mathbf{V}\|_F^2 \quad (37)$$

This method do not involve SVD computation, thus speeding up the minimization.

#### 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 (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 [52, 29]

$$\mathbf{Q} = \arg \min \|\mathbf{Q} \mathcal{T}(\mathbf{k}_{\text{central}})\|^2 \quad \text{such that} \quad \|\mathbf{Q}\|_F = 1 \quad (38)$$

using eigen value 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 (36) with the knowledge of the filters, resulting in computationally efficient solution. Figure 6 shows a schematic of this approach where the above 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 discussed above. The popular ESPIRiT reconstruction uses the null space property to estimate the coil sensitivities using an eigen decomposition [29]. In particular, once  $\mathbf{Q}$  or equivalently the signal subspace of  $\mathcal{T}_2$  is obtained, it performs a pixel by pixel eigen 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.

## 5. Smooth manifold models

Smooth manifold models use non-linear representations, which are more powerful in capturing the non-linear relations 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. 1). Methods relying on smooth manifold models include non-local means [12], non-local regularization [44], kernel methods [47, 69, 49, 48, 67], STORM [1, 45, 57, 10, 59], and recent extensions of STORM using deep generative models [74, 75]. 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 using the regularization term  $\|\nabla \mathbf{x}\|^2$ . This approach capitalizes on the fact that the intensities of the adjacent pixels are similar. Most algorithms for Tikhonov regularization rely on the Laplacian of the image  $\Delta \mathbf{x}$ . For 1-D images with  $n$  pixels, the Laplacian is often approximated by the  $x \times n$  finite difference

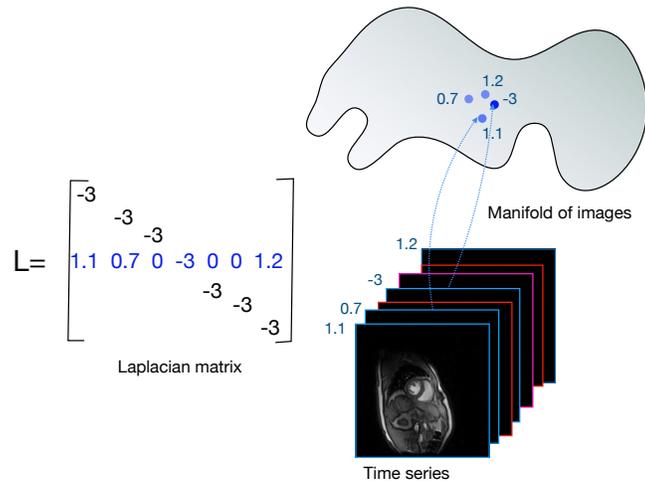


Figure 7: Illustration of the Laplacian matrix used in manifold methods. Each row of the Laplacian matrix may be thought of a second order derivative operator in the manifold domain. In this example, the fourth row of the matrix correspond to the finite difference operator for the fourth image in the 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.

matrix

$$\mathbf{L} = \begin{bmatrix} -2 & 1 & 0 & \dots & 1 \\ 1 & -2 & 1 & \dots & 0 \\ 0 & 1 & -2 & 1 & \dots \\ 1 & 0 & \dots & 1 & -2 \end{bmatrix}. \quad (39)$$

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. 7). This approach has strong ties to kernel methods [7] used in machine learning. In particular, one performs the smoothing in a non-linearly transformed feature space, defined by the non-linear mapping function  $\varphi(\mathbf{x})$ . The neighborhood structure or Laplacian is determined based on the distances between the non-linearly 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 non-linear mapping  $\varphi$  that is expensive to compute.

Recently, some authors have considered the use of deep generative models to account for the manifold structure of images in a dynamic dataset [74, 75]. Rather than considering a non-linear lifting of the data to a feature space, these approaches assume that the images are non-linear mappings of low-dimensional latent vectors that capture the variability in the data. The latent vectors capture the variability in the dynamic dataset (e.g. cardiac and respiratory variations). The latent vectors as well as the non-linear mapping are learned from the undersampled measurements using back-propagation. Since these models is to synthesize the images as non-linear mappings of low-dimensional latent vectors, these approaches are called as synthesis or generative manifold methods (see Fig. 8). By contrast, the earlier kernel methods that perform non-linear lifting of

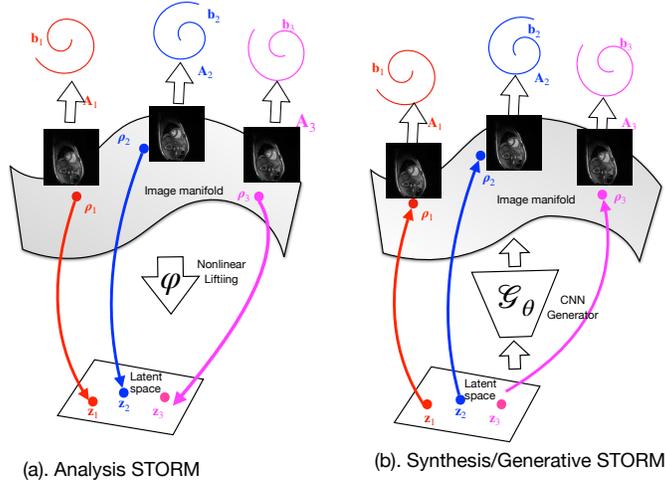


Figure 8: Illustration of the two classes of manifold methods. The early analysis methods [57, 47, 69, 49, 59, 45, 48, 10, 67, 1] rely on learning the Laplacian matrix, which captures the neighborhood structure of images in a non-linearly transformed domain. The non-linear mapping  $\varphi$  determines the lifting, and hence is an integral part of the algorithms. By contrast, the recent deep generative models [74, 75] represent each image in the time series as non-linear mappings of low-dimensional latent variables, which capture the natural variability in the data. Since these approaches synthesize the images, they are termed as synthesis models. The parameters of the deep generator  $\mathcal{G}_\theta$  as well as the latent vectors that represent each image  $\mathbf{z}_i$  are learned from the undersampled k-t space measurements.

images to a feature domain, where the smoothness is imposed, are termed as analysis methods.

### 5.1. Analysis manifold methods

The recovery of a smooth multidimensional function  $\mathbf{f} : \mathbb{R}^m \rightarrow \mathbb{R}^n$  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  $\mathbf{r}_k$ :  $\mathbf{x}_k = \mathbf{f}(\mathbf{r}_k)$ . Here,  $\mathbf{f}$  is a non-linear function, and  $\mathbf{r}_k$  are as the co-ordinates that are unknown. We note that as the latent co-ordinates  $\mathbf{r}$  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]:

$$\mathbf{f}^* = \arg \min_{\mathbf{f}, \{\mathbf{r}_k\}} \|\mathbf{f}(\mathbf{r}_k) - \mathbf{s}_k\|^2 + \eta \int_{\mathcal{M}} \|\nabla \mathbf{f}\|^2 d\mathbf{r}, \quad (40)$$

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]

$$\mathbf{f}^* = \arg \min_{\mathbf{f}, \{\mathbf{r}_j\}} \|\mathbf{f}(\mathbf{r}_k) - \mathbf{s}_k\|^2 + \eta \sum_{i=1}^N \sum_{j=1}^N \mathbf{W}_{i,j} \|\mathbf{f}(\mathbf{r}_i) - \mathbf{f}(\mathbf{r}_j)\|^2, \quad (41)$$

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

$$\mathbf{W}_{i,j} = \exp\left(-\frac{\|\mathbf{f}(\mathbf{r}_i) - \mathbf{f}(\mathbf{r}_j)\|^2}{\sigma^2}\right). \quad (42)$$

The weights capture the geometry of the manifold. Specifically, closer point pairs on  $\mathcal{M}$  will have higher weights, while distant point pairs will have smaller weights. We will discuss more

sophisticated approaches for the estimation of weights in Section 5.1.2. We note that the weighted sum can be expressed in a compact form as

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

Here,  $\mathbf{X} = [\mathbf{x}_1 \ \dots \ \mathbf{x}_N] = [\mathbf{f}(\mathbf{r}_1) \ \dots \ \mathbf{f}(\mathbf{r}_N)]$  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  $\mathbf{D} = \text{diag}(\sum_j \mathbf{W}_{i,j})$ . See Fig. 7 for an illustration of the structure of this matrix. Thus, the optimization scheme in (41) can also be written as

$$\mathbf{X} = \arg \min_{\mathbf{X}} \|\mathbf{X} - \mathbf{S}\|^2 + \eta \text{trace}(\mathbf{X}\mathbf{L}\mathbf{X}^T) \quad (43)$$

and  $\mathbf{S} = [\mathbf{s}_1 \ \dots \ \mathbf{s}_N]$ . The above 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 [55].

#### 5.1.1. Relation to factor models and binning based approaches

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

$$\mathbf{L} = \mathbf{\Phi}\mathbf{\Lambda}\mathbf{\Phi}^T \quad (44)$$

It is well known that the eigen functions (columns of  $\mathbf{\Phi}$ ) are basis functions of functions on  $\mathcal{M}$ , analogous to Fourier exponentials being eigen functions of Laplacian operator in Euclidean space [55]. In particular, eigen functions corresponding to smaller eigen values of  $\mathbf{L}$  correspond to smooth functions on  $\mathcal{M}$ ; the eigen values 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

$$\mathbf{X} = \mathbf{\Phi}\mathbf{W}^T, \quad (45)$$

where  $\mathbf{U}$  can be viewed as the coefficients. Using this property, one can rewrite (43) as

$$\mathbf{W}^* = \arg \min_{\mathbf{W}} \|\mathbf{\Phi}\mathbf{W}^T - \mathbf{Z}\|^2 + \eta \underbrace{\text{trace}(\mathbf{W}\mathbf{\Lambda}\mathbf{W}^T)}_{\sum_{i=1}^N \lambda_i \|\mathbf{w}_i\|^2} \quad (46)$$

Note that (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  $k^{\text{th}}$  row of  $\mathbf{W}$ . The weights are expected to be similar for points closer on the manifold; the active  $\mathbf{\Phi}$  basis vectors in each manifold neighborhood provides a local linear representation (similar to local PCA) on the manifold. Since the eigen values  $\lambda_i$  can be viewed as the *frequency* or the measure of roughness, one can attenuate the high frequency components on the manifold by increasing  $\eta$ , thus obtaining smoother signals on the manifold. For computational efficiency, one may also truncate the representation by ignoring the basis functions corresponding to higher eigen values.

When the images may 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 eigen values. The eigen vectors  $\varphi_i$  corresponding to the zero eigen values will be the indicator vectors of the clusters. If the remaining eigen values are much higher, one can approximate (46) as the independent recovery of each cluster from the measured data. We note that the hard binning approach pursued in GRASP or X-D 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 non-zero eigen values are not ignored in the reconstruction,

the additional eigen functions (corresponding to non-zero eigen values) 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.

### 5.1.2. Estimation of manifold Laplacian

As discussed previously,  $\mathbf{L}$  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, 57]. For instance, they are chosen as

$$\mathbf{W}_{i,j} \approx \exp\left(-\frac{\|\mathbf{s}_i - \mathbf{s}_j\|^2}{\sigma_s^2}\right). \quad (47)$$

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

$$\mathbf{W}_{i,j} = \begin{cases} 1 & \text{if } \|\mathbf{s}_i - \mathbf{s}_j\|^2 \leq \sigma_s^2 \\ 0 & \text{else} \end{cases} \quad (48)$$

Note that the Gaussian choice in (47) will be equivalent to hard-clustering if the clusters are well-separated. In this case, the  $\mathbf{L}$  matrix will have a block structure with no interactions between clusters. In contrast to the hard-clustering approach, the Gaussian choice is appropriate when the points are well-distributed on the manifold.

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  $\mathbf{E}$ . In this case, the estimation of the weights from aliased data using (47) often results in poor results. An approach to overcome this challenge is to post the recovery as the minimization of the cost function [70, 71]:

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} \|\mathbf{E}(\mathbf{X}) - \mathbf{S}\|^2 + \lambda \sum_{i=1}^N \sum_{j=1}^N \eta(\|\mathbf{x}_i - \mathbf{x}_j\|) \quad (49)$$

Here,  $\eta(\cdot)$  is a non-convex function of its argument (eg.,  $\ell_p$ ;  $p < 1$  norm). The non-convexity 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. The above criterion can be minimized by alternating between the estimation of the weights (47) and the recovery of the signals (41). Homotopy continuation schemes that start with a quadratic or convex  $\eta$ , and gradually change it to the desired one 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 [48, 67] 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  $\mathbf{i}$  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.

$$\mathbf{W}_{i,j}^* = \arg \min_{\mathbf{W}_{i,j}; \sum_j \mathbf{W}_{i,j} = 1} \|\mathbf{s}_i - \sum_j \mathbf{W}_{i,j} \mathbf{s}_j\|^2 + \lambda \sum_j \|\mathbf{W}_{i,j}\|_{\ell_1} \quad (50)$$

This approach is reported to yield improved results over proximity based methods [7, 12, 3, 57] in (47). This approach has similarities to local linear embedding [65], where each signal is expressed as a weighted linear combination of its neighbors.

Kernel based projection: Note that the approach in (46) approximates the signals using the eigen vectors corresponding to the lowest eigen values of the Laplacian matrix to approximate/denise

them. This is equivalent to approximating the signals using eigen vectors corresponding to the highest eigen values of the normalized kernel matrix. Kernel PCA is widely used in machine learning to approximate signals living on manifolds. The eigen values 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 [69, 49]. Specifically, polynomial features of the signals were computed, followed by performing PCA. Once the features are projected to a lower dimensional subspace, these methods use the explicit inversion formula to obtain the pre-image available for polynomial kernels. This approach demonstrated improved performance over PCA. A challenge with this direct approach is the difficulty to apply 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 [59]. In particular, the kernel matrix is exactly low-rank when the manifold  $\mathcal{M} \in \mathbb{R}^N <$  is the zero level set of a finite linear combination of basis functions  $\psi(\mathbf{x}) = \sum_{k=1}^B c_k \varphi_k(\mathbf{x})$ .

$$\mathcal{M} = \{x | \sum c_k \varphi_k(\mathbf{x}) = 0\}. \quad (51)$$

Here  $\varphi_k(\mathbf{r})$  are basis functions<sup>1</sup> (e.g. polynomials, exponentials) that span the high dimensional space.

$$\Phi(\mathbf{X}) = \begin{bmatrix} \underbrace{\begin{bmatrix} \varphi_1(\mathbf{x}_1) \\ \vdots \\ \varphi_S(\mathbf{x}_1) \end{bmatrix}}_{\phi(\mathbf{r}_1)} & \cdots & \underbrace{\begin{bmatrix} \varphi_1(\mathbf{x}_N) \\ \vdots \\ \varphi_S(\mathbf{x}_N) \end{bmatrix}}_{\phi(\mathbf{x}_N)} \end{bmatrix} \quad (52)$$

is low rank. The mapping from the original points to the feature vectors can be viewed as a non-linear 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 non-linear mapping  $\phi$ . Following the approach in Section 4, we recover the signals on the manifold from its linear measurements as

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} \|\mathbf{E}(\mathbf{X}) - \mathbf{S}\|^2 + \lambda \|\Phi(\mathbf{X})\|_*. \quad (53)$$

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

$$\langle \phi(\mathbf{r}_i), \phi(\mathbf{r}_j) \rangle = \kappa(\mathbf{x}_i - \mathbf{x}_j) \quad (54)$$

as non-linear functions of the image differences  $\mathbf{x}_i - \mathbf{x}_j$  without the direct computation of the feature maps, which is computationally challenging. Here,  $\kappa$  is a function that is dependent on the specific feature maps.

The *kernel-trick* is widely used in machine learning to translate algorithms that depend on inner-products to feature spaces. In particular, by simply changing the inner-products by kernel functions. We note that the IRLS approach described in Section 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 this property to solve (53), which alternates between

$$\mathbf{X}^* = \arg \min_{\mathbf{X}} \|\mathbf{E}(\mathbf{X}) - \mathbf{S}\|^2 + \lambda \|\Phi(\mathbf{X}) \kappa(\mathbf{X})^{-1}\|_F^2. \quad (55)$$

---

<sup>1</sup>The basis functions may also be chosen as a subset of the Mercer decomposition of the kernel function.

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

$$(\kappa(\mathbf{X}))_{i,j} = \langle \phi(\mathbf{r}_i), \phi(\mathbf{r}_j) \rangle = \kappa(\mathbf{r}_i, \mathbf{r}_j) \quad (56)$$

This approach eliminates the need for the explicit evaluation of the high dimensional feature maps  $\phi(f_i)$ . Since the regularization functional in (55) is non-quadratic, this term is solved using steepest descent; this approach is shown to be similar to the alternating strategy in (5.1.2) (i.e, alternation between (47) and the recovery of the signals (41)), where (47) is replaced by

$$\mathbf{L} = \mathcal{K}(\mathbf{X}) \odot \mathcal{K}(\mathbf{X})^{-\frac{1}{2}}, \quad (57)$$

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

### 5.1.3. Image recovery assuming smooth patch manifold

Non-local means is an early and powerful algorithm for patch based image denoising [12, 3]. It estimates the Laplacian matrix from noisy data as in (47), followed by (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 (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 (45) to further improve the denoising performance [72]. Each of the columns of  $\mathbf{V}$  are termed as non-local 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 non-local 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 (49) is adopted [71, 70]. 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 [44], which offers implicit motion compensation. This approach has conceptual similarities to [13] and the kernel PCA approach in the patch setting [66]. 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.

## 5.2. Synthesis Manifold Recovery

The main focus of synthesis manifold approaches us to capitalize on the power of deep convolutional neural networks (CNN) to introduce a memory efficient generative or synthesis formulation. These methods assume the image volumes in the dataset are smooth non-linear functions of a few latent variables, i.e.,  $x(t) = \mathcal{G}_\theta(z_t)$ , where  $z_t$  are the latent vectors in a low-dimensional space.  $x(t)$  is the  $t^{\text{th}}$  generated image frame in the time series. This explicit formulation implies that the image volumes lie on a smooth non-linear manifold in a high-dimensional ambient space (see Fig 8.(b)). A subject-specific deep convolutional neural network (CNN) to represent the non-linear transformation  $\mathcal{G}_\theta$  in Fig. 8. The parameters of the generator  $\theta$  as well as the low-dimensional latent vectors  $\mathbf{z}_i$  are jointly estimated only from the undersampled measurements. This approach is different from traditional CNN approaches that require extensive fully sampled training data. We penalize the norm of the gradients of the non-linear mapping to constrain the manifold to be smooth, while temporal gradients of the latent vectors are penalized to obtain a smoothly varying time-series. The proposed scheme brings in the spatial regularization provided by the convolutional network and offers a significantly compressed representation of the data. Specifically, the number of parameters required by the model (CNN weights and latent vectors) are several orders of magnitude smaller than

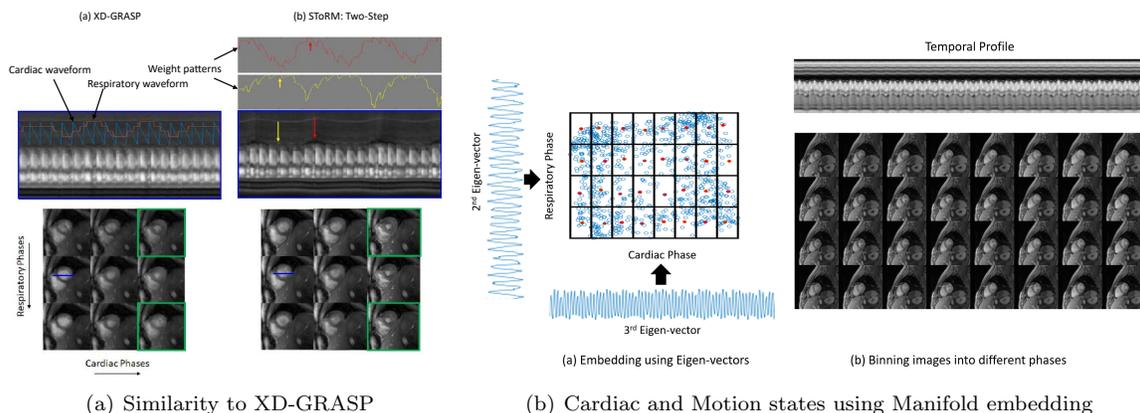


Figure 9: Illustration of Smoothness Regularization on Manifolds (STORM) framework and its similarity to XD-GRASP for the recovery of free-breathing and ungated MRI data. XD-GRASP bins the data to different cardiac/respiratory phases, followed by the joint recovery of the images. By contrast, STORM estimates a Laplacian matrix that has conceptual similarities to the XD-GRASP approach. The manifold Laplacian is estimated from the k-space navigators using (55) and (57). The reconstructed data is compared against self-gated XD-GRASP reconstruction of the same data. (b) uses the eigen vectors of the Laplacian matrix to bin the reconstructed data into cardiac and respiratory phases. This figure is copied from [59] with permission from IEEE.

required for the direct representation of the images. The compact model proportionately reduces the number of measurements needed to recover the images. In addition, the compression also enables algorithms with much smaller memory footprint and computational complexity. The main benefit of the proposed scheme is the improvement in image quality and the orders-of-magnitude reduction in memory demand compared to traditional manifold models. The recovery from the measured data is posed as an optimization problem, where the latent variables  $\mathbf{z}_i$  and the CNN parameters  $\theta$  are obtained as the minimum of the cost function:

$$\mathcal{C}(\mathbf{z}, \theta) = \underbrace{\sum_{i=1}^N \|\mathcal{A}_i(\mathcal{G}_\theta[\mathbf{z}_i]) - \mathbf{b}\|^2}_{\text{data term}} + \lambda_1 \underbrace{\|J_{\mathbf{z}}\mathcal{G}_\theta(\mathbf{z})\|^2}_{\text{distance regularization}} + \lambda_2 \underbrace{\|\nabla_t \mathbf{z}_t\|^2}_{\text{latent regularization}}$$

The first term is the data consistency term, which compares the measurements of the generated images with the k-t space data. The second term is a regularization penalty on the CNN parameters, which minimizes overfitting, while the last term is a smoothness penalty of the latent vectors. In particular, the last term encourages smooth variation of the latent vectors, capitalizing on the smooth nature of cardiac and respiratory motion.

### 5.3. Application to Dynamic MRI

The manifold structure of images in a dynamic time series is used to recover them from under-sampled data [58, 59, 57, 45]. In particular, the images in a free breathing cardiac dataset can be viewed as non-linear 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 [57]. 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  $\mathbf{Z} = \mathbf{B}\mathbf{X}$ , where  $\mathbf{B}$  is the sampling operator corresponding to the k-space navigators. In this case, one can approximate the  $\mathbf{W}$  matrix as

$$\mathbf{W}_{i,j} \approx \exp\left(-\frac{\|\mathbf{z}_i - \mathbf{z}_j\|^2}{\sigma_z^2}\right). \quad (58)$$

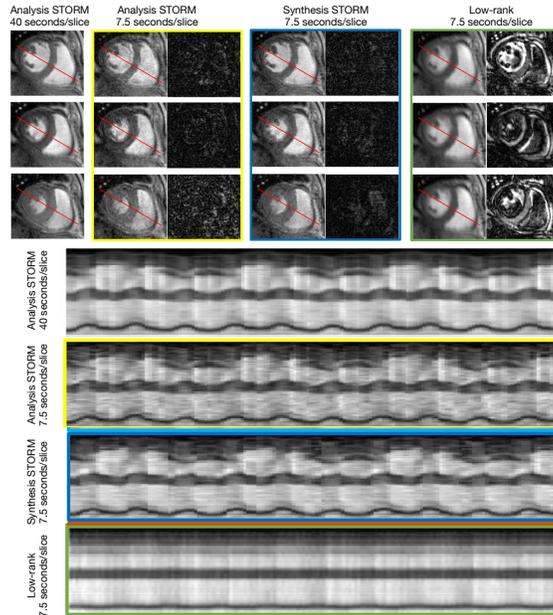


Figure 10: Comparison of Analysis and Deep-Synthesis STORM approaches, applied to free breathing and ungated cardiac MRI. The improved regularization offered by synthesis manifold approach enables a significant reduction in data demand. The synthesis STORM approach from 7.5 seconds of data is able to offer reconstructions that are similar to analysis STORM reconstructions from 40 seconds of acquisition. The results are significantly better than low-rank methods.

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

A challenge with the direct implementation of the manifold-aware recovery in (43) in the multi-dimensional setting is its high computational complexity. The factorization approach is considered in (46) allows to significantly reduce the computational complexity and memory demand of the algorithm [59]. 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 [59], which offered improved performance over [57].

When navigators are not available, the formulation in (49) allows to jointly estimate the Laplacian and the signals from the data itself [45]. A key benefit of this approach is that one can customize the Laplacian to different spatial regions. In particular, [45] splitted the images to patches, each with a different Laplacian. The kernel low-rank algorithm (53) is used instead of (49) to estimate the Laplacian in [1]. The iterative strategies including (49) and (53) 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. 9 for an illustration of the manifold recovery. The comparison of analysis and synthesis manifold methods are shown in Fig. 5.3. The results show that the generative STORM approach facilitate a significant reduction in acquisition time compared to the analysis counterparts.

## 6. Software

The MATLAB software associated with this paper is available at [https://github.com/sajanglingala/data\\_adaptive\\_recon\\_MRI](https://github.com/sajanglingala/data_adaptive_recon_MRI)

## 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 sub-structures (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 sub-regions are used to populate the columns of a structured matrix, which is factorized into two sub-matrices. 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 sub-regions of the dataset, thus facilitating the recovery of the dataset from highly undersampled data.

## References

- [1] Ahmed, A.H., Zhou, R., Yang, Y., Nagpal, P., Salerno, M., Jacob, M., 2020. Free-breathing and ungated dynamic MRI using navigator-less spiral STORM. *IEEE Transactions on Medical Imaging* , 1–1doi:10.1109/TMI.2020.3008329.
- [2] Akçakaya, M., Basha, T.A., Goddu, B., Goepfert, L.A., Kissinger, K.V., Tarokh, V., Manning, W.J., Nezafat, R., 2011. Low-dimensional-structure self-learning and thresholding: Regularization beyond compressed sensing for MRI Reconstruction. *Magnetic Resonance in Medicine* 66. doi:10.1002/mrm.22841.
- [3] Awate, S.P., Whitaker, R.T., 2006. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28, 364–376. URL: <http://ieeexplore.ieee.org/document/1580482/>, doi:10.1109/TPAMI.2006.64.
- [4] Balachandrasekaran, A., Magnotta, V., Jacob, M., 2017. Recovery of Damped Exponentials Using Structured Low Rank Matrix Completion. *IEEE Transactions on Medical Imaging* doi:10.1109/TMI.2017.2726995.
- [5] Balachandrasekaran, A., Mani, M., Jacob, M., 2019. Calibration-Free B0 Correction of EPI Data Using Structured Low Rank Matrix Recovery. *IEEE Transactions on Medical Imaging* 38. doi:10.1109/TMI.2018.2876423.
- [6] Belkin, M., Niyogi, P., 2003. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation* doi:10.1162/089976603321780317.
- [7] Belkin, M., Niyogi, P., Sindhwani, V., 2006. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of Machine Learning Research* 7, 2399–2434. doi:10.1016/j.neuropsychologia.2009.02.028.
- [8] Bhave, S., Lingala, S.G., Johnson, C.P., Magnotta, V.A., Jacob, M., 2016a. Accelerated whole-brain multi-parameter mapping using blind compressed sensing. *Magnetic Resonance in Medicine* doi:10.1002/mrm.25722.
- [9] Bhave, S., Lingala, S.G., Newell, J.D., Nagle, S.K., Jacob, M., 2016b. Blind compressed sensing enables 3-dimensional dynamic free breathing magnetic resonance imaging of lung volumes and diaphragm motion. *Investigative Radiology* 51, 387–99. doi:10.1097/RLL.0000000000000253.

- [10] Biswas, S., Aggarwal, H.K., Jacob, M., 2019. Dynamic MRI using model-based deep learning and SToRM priors: MoDL-SToRM. *Magnetic Resonance in Medicine* doi:10.1002/mrm.27706.
- [11] Brinegar, C., Schmitter, S.S., Mistry, N.N., Johnson, G.A., Liang, Z.P., 2010. Improving temporal resolution of pulmonary perfusion imaging in rats using the partially separable functions model. *Magnetic Resonance in Medicine* 64. doi:10.1002/mrm.22500.
- [12] Buades, A., Coll, B., Morel, J.M., 2005. 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*. pp. 60–65. doi:10.1109/CVPR.2005.38.
- [13] Bustin, A., Lima da Cruz, G., Jaubert, O., Lopez, K., Botnar, R.M., Prieto, C., 2019. High-dimensionality undersampled patch-based reconstruction (HD-PROST) for accelerated multi-contrast MRI. *Magnetic Resonance in Medicine* 81. doi:10.1002/mrm.27694.
- [14] Candès, E.J., Fernandez-Granda, C., 2014. Towards a mathematical theory of super-resolution. *Communications on pure and applied Mathematics* 67, 906–956.
- [15] Candes, E.J., Wakin, M.B., 2008. An introduction to compressive sampling: A sensing/sampling paradigm that goes against the common knowledge in data acquisition. *IEEE Signal Processing Magazine* 25, 21–30. doi:10.1109/MSP.2007.914731.
- [16] Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K., 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16, 2080–2095. doi:10.1109/TIP.2007.901238.
- [17] Danielyan, A., Katkovnik, V., Egiazarian, K., 2012. BM3D frames and variational image deblurring. *IEEE Transactions on Image Processing* doi:10.1109/TIP.2011.2176954.
- [18] Feng, L., Axel, L., Chandarana, H., Block, K.T., Sodickson, D.K., Otazo, R., 2016. Xd-grasp: golden-angle radial mri with reconstruction of extra motion-state dimensions using compressed sensing. *Magnetic resonance in medicine* 75, 775–788.
- [19] Goud, S., Hu, Y., Jacob, M., 2010. 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*. doi:10.1109/ISBI.2010.5490154.
- [20] Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magn. Reson. Med* 47, 1202–1210.
- [21] Haldar, J.P., 2014a. Low-Rank Modeling of Local  $k$ -Space Neighborhoods (LORAKS) for Constrained MRI. *Medical Imaging, IEEE Transactions on* 33, 668–681. doi:10.1109/TMI.2013.2293974.
- [22] Haldar, J.P., 2014b. Low-Rank Modeling of Local  $k$ -Space Neighborhoods (LORAKS) for Constrained MRI. *IEEE Transactions on Medical Imaging* doi:10.1109/TMI.2013.2293974.
- [23] Haldar, J.P., Zhuo, J., 2016. P-LORAKS: Low-rank modeling of local  $k$ -space neighborhoods with parallel imaging data. *Magnetic Resonance in Medicine* doi:10.1002/mrm.25717.
- [24] Jacob, M., Mani, M.P., Ye, J.C., 2020. Structured Low-Rank Algorithms: Theory, Magnetic Resonance Applications, and Links to Machine Learning. *IEEE Signal Processing Magazine* 37, 54–68. doi:10.1109/MSP.2019.2950432.

- [25] Jin, K.H., Lee, D., Ye, J.C., 2015. A novel k-space annihilating filter method for unification between compressed sensing and parallel MRI, in: Proceedings - International Symposium on Biomedical Imaging. doi:10.1109/ISBI.2015.7163879.
- [26] Jin, K.H., Lee, D., Ye, J.C., 2016a. A General Framework for Compressed Sensing and Parallel MRI Using Annihilating Filter Based Low-Rank Hankel Matrix. IEEE Transactions on Computational Imaging 2. doi:10.1109/tci.2016.2601296.
- [27] Jin, K.H., Lee, D., Ye, J.C., 2016b. A General Framework for Compressed Sensing and Parallel MRI Using Annihilating Filter Based Low-Rank Hankel Matrix. IEEE Transactions on Computational Imaging doi:10.1109/tci.2016.2601296.
- [28] Kay, S.M., Marple, S.L., 1981. Spectrum Analysis—A Modern Perspective. Proceedings of the IEEE doi:10.1109/PROC.1981.12184.
- [29] Lai, P., Lustig, M., Vasanawala, S.S., Brau, a.C., 2011. ESPIRiT ( Efficient Eigenvector-Based L1SPIRiT ) for Compressed Sensing Parallel Imaging - Theoretical Interpretation and Improved Robustness for Overlapped FOV Prescription. Electrical Engineering 19.
- [30] Lee, D., Jin, K.H., Kim, E.Y., Park, S.H., Ye, J.C., 2016. Acceleration of MR parameter mapping using annihilating filter-based low rank hankel matrix (ALOHA). Magnetic Resonance in Medicine doi:10.1002/mrm.26081.
- [31] Liang, Z.P., 2007. Spatiotemporal imaging with partially separable functions, in: 2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings, pp. 988–991. doi:10.1109/ISBI.2007.357020.
- [32] Liang, Z.P., Lauterbur, P.C., 1991. A Generalized Series Approach to MR Spectroscopic Imaging. IEEE Transactions on Medical Imaging 10. doi:10.1109/42.79470.
- [33] Lingala, S.G., Dibella, E., Adluru, G., McGann, C., Jacob, M., 2013. Accelerating free breathing myocardial perfusion MRI using multi coil radial k - T SLR. Physics in Medicine and Biology doi:10.1088/0031-9155/58/20/7309.
- [34] Lingala, S.G., Hu, Y., Dibella, E., Jacob, M., 2011. Accelerated dynamic MRI exploiting sparsity and low-rank structure: K-t SLR. IEEE Transactions on Medical Imaging 30, 1042–1054. URL: <https://pubmed.ncbi.nlm.nih.gov/21292593/>, doi:10.1109/TMI.2010.2100850.
- [35] Lingala, S.G., Jacob, M., 2013a. Blind compressed sensing with sparse dictionaries for accelerated dynamic MRI, in: Proceedings - International Symposium on Biomedical Imaging. doi:10.1109/ISBI.2013.6556398.
- [36] Lingala, S.G., Jacob, M., 2013b. Blind compressive sensing dynamic MRI. IEEE Transactions on Medical Imaging 32, 1132–1145. doi:10.1109/TMI.2013.2255133.
- [37] Lobos, R.A., Kim, T.H., Hoge, W.S., Haldar, J.P., 2017. Navigator-free EPI ghost correction using low-rank matrix modeling: Theoretical insights and practical improvements, in: In Proceedings of the 25th Annual Meeting of ISMRM, Honolulu, Hawaii, USA.
- [38] Mani, M., Aggarwal, H.K., Magnotta, V., Jacob, M., 2020a. Improved MUSSELS reconstruction for high-resolution multi-shot diffusion weighted imaging. Magnetic Resonance in Medicine 83. doi:10.1002/mrm.28090.
- [39] Mani, M., Aggarwal, H.K., Magnotta, V., Jacob, M., 2020b. Improved MUSSELS reconstruction for high-resolution multi-shot diffusion weighted imaging. Magnetic Resonance in Medicine 83. doi:10.1002/mrm.28090.

- [40] Mani, M., Jacob, M., 2014. Fast iterative algorithm for the reconstruction of multishot non-cartesian diffusion data. *Magnetic Resonance in ...* URL: <http://onlinelibrary.wiley.com/doi/10.1002/mrm.25486/full>.
- [41] Mani, M., Jacob, M., Kelley, D., Magnotta, V., 2017. Multi-shot sensitivity-encoded diffusion data recovery using structured low-rank matrix completion (MUSSELS). *Magnetic Resonance in Medicine* 78. doi:10.1002/mrm.26382.
- [42] Mani, M., Magnotta, V., Jacob, M., 2018. A general algorithm for compensation of trajectory errors: Application to radial imaging. URL: <http://doi.wiley.com/10.1002/mrm.27148>, doi:10.1002/mrm.27148.
- [43] Mani, M., Magnotta, V., Kelley, D., Jacob, M., 2016. 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*. doi:10.1109/EMBC.2016.7590897.
- [44] Mohsin, Y.Q., Lingala, S.G., DiBella, E., Jacob, M., 2017. Accelerated dynamic MRI using patch regularization for implicit motion compensation. *Magnetic Resonance in Medicine* 77, 1238–1248. URL: <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>, doi:10.1002/mrm.26215.
- [45] Mohsin, Y.Q., Poddar, S., Jacob, M., 2019. Free-Breathing Ungated Cardiac MRI Using Iterative SToRM (i-SToRM). *IEEE Transactions on Medical Imaging* 38. doi:10.1109/TMI.2019.2908140.
- [46] Morrison, R.L., Jacob, M., Do, M.N., 2007. Multichannel estimation of coil sensitivities in parallel MRI, in: *2007 4th IEEE International Symposium on Biomedical Imaging: From Nano to Macro - Proceedings*. doi:10.1109/ISBI.2007.356802.
- [47] Nakarmi, U., Slavakis, K., Lyu, J., Ying, L., 2017a. M-MRI: A manifold-based framework to highly accelerated dynamic magnetic resonance imaging, in: *Proceedings - International Symposium on Biomedical Imaging*. doi:10.1109/ISBI.2017.7950458.
- [48] Nakarmi, U., Slavakis, K., Ying, L., 2018. MLS: Joint manifold-learning and sparsity-aware framework for highly accelerated dynamic magnetic resonance imaging, in: *Proceedings - International Symposium on Biomedical Imaging*. doi:10.1109/ISBI.2018.8363789.
- [49] Nakarmi, U., Wang, Y., Lyu, J., Liang, D., Ying, L., 2017b. A Kernel-Based Low-Rank (KLR) Model for Low-Dimensional Manifold Recovery in Highly Accelerated Dynamic MRI. *IEEE Transactions on Medical Imaging* doi:10.1109/TMI.2017.2723871.
- [50] Ongie, G., Biswas, S., Jacob, M., . Convex recovery of continuous domain piecewise constant images from nonuniform Fourier samples. *IEEE Transactions on Signal Processing* 66, 236–250.
- [51] Ongie, G., Biswas, S., Jacob, M., 2018. Convex Recovery of Continuous Domain Piecewise Constant Images From Nonuniform Fourier Samples. *IEEE Transactions on Signal Processing* 66, 236–250. doi:10.1109/TSP.2017.2750111.
- [52] Ongie, G., Jacob, M., 2015. Super-resolution MRI using finite rate of innovation curves, in: *Proceedings - International Symposium on Biomedical Imaging, IEEE Computer Society*. pp. 1248–1251. doi:10.1109/ISBI.2015.7164100.

- [53] Ongie, G., Jacob, M., 2016. Off-the-grid recovery of piecewise constant images from few fourier samples. *SIAM Journal on Imaging Sciences* 9, 1004–1041. doi:10.1137/15M1042280.
- [54] Ongie, G., Jacob, M., 2017. A Fast Algorithm for Convolutional Structured Low-Rank Matrix Recovery. *IEEE Transactions on Computational Imaging* 3, 535–550. doi:10.1109/tci.2017.2721819.
- [55] Ortega, A., Frossard, P., Kovacevic, J., Moura, J.M., Vandergheynst, P., 2018. Graph Signal Processing: Overview, Challenges, and Applications. *Proceedings of the IEEE* doi:10.1109/JPROC.2018.2820126.
- [56] Peyré, G., 2009. Manifold models for signals and images. *Computer Vision and Image Understanding* doi:10.1016/j.cviu.2008.09.003.
- [57] Poddar, S., Jacob, M., 2016. Dynamic MRI Using Smoothness Regularization on Manifolds (SToRM). *IEEE Transactions on Medical Imaging* doi:10.1109/TMI.2015.2509245.
- [58] Poddar, S., Mohsin, Y., Ansah, D., Thattaliyath, B., Ashwath, R., Jacob, M., 2015. Free-breathing cardiac MRI using bandlimited manifold modelling. *IEEE Trans. on Computational Imaging* 35, 1106–1115.
- [59] Poddar, S., Mohsin, Y.Q., Ansah, D., Thattaliyath, B., Ashwath, R., Jacob, M., 2019. Manifold Recovery Using Kernel Low-Rank Regularization: Application to Dynamic Imaging. *IEEE Transactions on Computational Imaging* 3, 478 – 491. URL: <https://ieeexplore.ieee.org/document/8625515/>, doi:10.1109/tci.2019.2893598.
- [60] Ravishankar, S., Bresler, Y., 2011. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Transactions on Medical Imaging* 30, 1028–1041. doi:10.1109/TMI.2010.2090538.
- [61] Ravishankar, S., Bresler, Y., 2013. Learning sparsifying transforms. *IEEE Transactions on Signal Processing* 61, 1072–1086. doi:10.1109/TSP.2012.2226449.
- [62] Ravishankar, S., Bresler, Y., 2015. Efficient blind compressed sensing using sparsifying transforms with convergence guarantees and application to magnetic resonance imaging. *SIAM Journal on Imaging Sciences* 8, 2519–2557. doi:10.1137/141002293.
- [63] Ravishankar, S., Ye, J.C., Fessler, J.A., 2020. Image Reconstruction: From Sparsity to Data-Adaptive Methods and Machine Learning. *Proceedings of the IEEE* 108. doi:10.1109/JPROC.2019.2936204.
- [64] Recht, B., Fazel, M., Parrilo, P.A., 2010. Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM review* 52, 471–501.
- [65] Roweis, S.T., Saul, L.K., 2000. Nonlinear dimensionality reduction by locally linear embedding. *Science* doi:10.1126/science.290.5500.2323.
- [66] Schmidt, J.F., Santelli, C., Kozerke, S., 2016. Mr image reconstruction using block matching and adaptive kernel methods. *PloS one* 11, e0153736.
- [67] Shetty, G.N., Slavakis, K., Bose, A., Nakarmi, U., Scutari, G., Ying, L., 2020. Bi-Linear Modeling of Data Manifolds for Dynamic-MRI Recovery. *IEEE Transactions on Medical Imaging* doi:10.1109/TMI.2019.2934125.
- [68] Shin, P.J., Larson, P.E., Ohliger, M.A., Elad, M., Pauly, J.M., Vigneron, D.B., Lustig, M., 2014. Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion, in: *Magnetic Resonance in Medicine*. doi:10.1002/mrm.24997.

- [69] Wang, Y., Ying, L., 2014. 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. doi:10.1109/EMBC.2014.6943894.
- [70] Yang, Z., Jacob, M., 2012. Robust non-local regularization framework for motion compensated dynamic imaging without explicit motion estimation, in: Proceedings - International Symposium on Biomedical Imaging, NIH Public Access. pp. 1056–1059. URL: /pmc/articles/PMC3956771/?report=abstract <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3956771/>, doi:10.1109/ISBI.2012.6235740.
- [71] Yang, Z., Jacob, M., 2013. Nonlocal regularization of inverse problems: A unified variational framework. IEEE Transactions on Image Processing 22, 3192–3203. doi:10.1109/TIP.2012.2216278.
- [72] Yin, R., Gao, T., Lu, Y.M., Daubechies, I., 2017. A tale of two bases: Local-nonlocal regularization on image patches with convolution framelets. SIAM Journal on Imaging Sciences doi:10.1137/16M1091447.
- [73] Zhao, B., Haldar, J.P., Liang, Z.P., 2010. 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. pp. 3390–3393. URL: /pmc/articles/PMC3121182/?report=abstract <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3121182/>, doi:10.1109/IEMBS.2010.5627934.
- [74] Zou, Q., Ahmed, A., Nagpal, P., Kruger, S., Jacob, M., in press. Dynamic imaging using a deep generative storm (gen-storm) model. IEEE Transactions on Medical Imaging .
- [75] Zou, Q., Ahmed, A.H., Nagpal, P., Priya, S., Schulte, R., Jacob, M., . Generative storm: A novel approach for joint alignment and recovery of multi-slice dynamic mri generative storm: A novel approach for joint alignment and recovery of multi-slice dynamic mri. <https://arxiv.org/abs/2101.08196> .

Method	Signal Type		Patch		Images	
	Two-step	Joint	Two-step	Joint	Two-step	Joint
Low-rank image domain	BM3D [16, 17]	CLEAR [19], LOST [2], HD-PROST [?] ] PRICE [44]	-	-	-	-
		ESPIRIT[44] GRAPPA [20] SLR[52]	LORAKS [21, 23, 22] SLR [54, 53, 51, 52, 50] MUSSELS [43, 41, 42, 38]	PSF [11, 73, 32, 31]	k-t SLR [19, 34, 33]	
Low-rank k-space	ESPIRIT[44] GRAPPA [20] SLR[52]	LORAKS [21, 23, 22] SLR [54, 53, 51, 52, 50] MUSSELS [43, 41, 42, 38]	-	-	ALOHA [30, 26, 27, 25] SLR [4, 5]	-
		Transform Learning [61, 60, 63]	BCS [62]	-	-	BCS [9, 8, 35, 36]
Dictionary learning	NLM [12] UINTA [3]	Nonlocal Regularization [71, 71, 70] Patch-STORM [45]	Kernel PCA [49, 69] STORM [57] MLS [67, 48] KLR-STORM [59]	-	-	-
		Manifold smoothness manifold	Spiral-STORM [1]	-	-	-

Table 1: Broad classification of methods reviewed in this chapter.