

# Improved Computational Efficiency of Locally Low Rank MRI Reconstruction Using Iterative Random Patch Adjustments

Andres Saucedo, Stamatios Lefkimmiatis, Novena Rangwala, Kyunghyun Sung

**Abstract**—This paper presents and analyzes an alternative formulation of the locally low-rank (LLR) regularization framework for magnetic resonance image (MRI) reconstruction. Generally, LLR-based MRI reconstruction techniques operate by dividing the underlying image into a collection of matrices formed from image patches. Each of these matrices is assumed to have low rank due to the inherent correlations among the data, whether along the coil, temporal, or multi-contrast dimensions. LLR regularization has been successful for various MRI applications such as parallel imaging and accelerated quantitative parameter mapping. However, a major limitation of most conventional implementations of LLR regularization is the use of multiple sets of overlapping patches. Although the use of overlapping patches leads to effective shift-invariance, it also results in high computational load, which limits the practical utility of LLR regularization for MRI. To circumvent this problem, alternative LLR-based algorithms instead shift a single set of non-overlapping patches at each iteration, thereby achieving shift-invariance and avoiding block artifacts. A novel contribution of this paper is to provide a mathematical framework and justification of LLR regularization with iterative random patch adjustments (LLR-IRPA). This method is compared with a state-of-the-art LLR regularization algorithm based on overlapping patches, and it is shown experimentally that results are similar but with the advantage of much reduced computational load. We also present theoretical results demonstrating the effective shift invariance of the LLR-IRPA approach, and we show reconstruction examples and comparisons in both retrospectively and prospectively undersampled MRI acquisitions, and in T1 parameter mapping.

**Index Terms**—compressive sensing, parallel imaging, parameter mapping, locally low-rank regularization

## I. INTRODUCTION

CURRENT medical magnetic resonance imaging (MRI) largely depends on undersampled data acquisitions, combined with specialized reconstruction techniques, to reach the levels of spatio-temporal resolutions and volumetric coverage necessary for practical clinical purposes [1],[2],[3]. These accelerated imaging methods recover images from highly reduced k-space samples by implementing optimization schemes that incorporate *a priori* knowledge of the underlying image information. Parallel imaging takes advantage of the inherent data redundancy available from multiple coil measurements, while compressed sensing (CS) exploits the low-dimensional representation of spatiotemporal image characteristics with respect to suitable sparsifying transforms or matrix decompositions [4],[5]. One of the most recently developed ideas in CS includes the notion of low-rank constrained reconstruction, which is based on the fact that image data tends to be

highly correlated across, for example, the temporal and/or coil dimensions [6]. In the globally low-rank (GLR) model, a time series or multi-coil image set, when treated jointly in matrixized form, can be accurately represented by a matrix of much lower rank relative to the number of time points or coils [7]. The reconstruction of such an undersampled image set is generally posed as a low-rank constrained matrix completion optimization problem. Several researchers have demonstrated significant advantages in image quality and improved temporal resolution from utilizing this low-rank optimization framework in dynamic imaging [8],[9],[10], parallel imaging [11], functional imaging [12], real-time imaging [13], and accelerated quantitative parameter mapping [14].

Despite the success of the GLR approach, recent studies have shown that adopting a locally low-rank (LLR) model—which assumes correlations across multiple images only within a relatively small neighborhood of pixels—yields more favorable results and involves less computational load than the GLR approach. Studies by [15] have shown that the LLR framework may provide better performance in terms of the trade-off between imaging speed and data fidelity. This framework has been applied in parallel imaging, where Tzasko and Manduca introduced an image domain-based calibration-free method that is based on the observation that coil sensitivities vary smoothly in space, such that images are locally-correlated along the coil dimension [16]. In [17], Zhang et al. demonstrated accelerated T1 and T2 mapping using a similar concept, except that the local image correlations are assumed to exist across images taken with different pulse-sequence parameters. Locally low-rank constraints have facilitated the combination of both compressed sensing and parallel imaging reconstruction techniques, most notably in the case of dynamic cardiac and contrast-enhanced imaging [18]. Importantly, LLR-based reconstruction has the distinct advantage of requiring considerably less computational load and memory requirements than its GLR counterpart [16].

However, the LLR approach has certain drawbacks that result from the particular way of defining the set of patches into which the underlying image is decomposed. The collection of patches, or partition, delineates the local regions in the image series where low-rank submatrices may be formed. Like most patch-based reconstruction methods, LLR-based reconstruction can be implemented by using sets of either overlapping or non-overlapping patches. A partition consisting of overlapping patches can minimize the appearance of block artifacts as the transform becomes approximately shift-invariant, but it

comes with the disadvantage of high computational cost due to the large number of patches involved. In contrast, a partition composed of non-overlapping, covering patches can greatly reduce the computational load, but it inevitably leads to block artifacts<sup>1</sup>. Both of these strategies, therefore, limit the utility of LLR regularization for practical MRI applications.

Here, we present an implementation of LLR regularization with iterative random patch adjustments (LLR-IRPA) that utilizes an effectively shift-invariant, patch-based transform without high computational cost. This paper promotes the case of using partitions consisting of non-overlapping patches, since these types of partitions require the computation of a much smaller number of singular-value matrix decompositions (SVD) [19], in comparison to using partitions formed from overlapping patches. Inspired by the work of Xu and Yin [20], we propose to randomly shift the partition at each iteration of the iterative singular value soft-thresholding algorithm that is used for solving the associated optimization problem [21]. We compare LLR-IRPA with one of the state-of-the-art LLR regularization methods, CLEAR [16], which uses multiple sets of overlapping patches. We show that LLR-IRPA performs equivalently or even better than CLEAR but with the added advantage of substantially reduced computational load. This strategy is not limited strictly to parallel MRI, but can be extended within the context of MR quantitative parameter mapping. As also shown in the results, the LLR-IRPA approach applies equally well to accelerated T1 mapping from undersampled, calibrationless k-space data.

We provide theoretical support to justify LLR-IRPA, based on results related to cycle spinning in the wavelet-based iterative soft-thresholding algorithm (ISTA) [22]. A proof of convergence of the LLR-IRPA algorithm is given within the ISTA framework, although all reconstruction methods are implemented with FISTA, an accelerated version of ISTA. Similar to non-cycled wavelet-based  $\ell_1$ -regularization, locally low-rank regularization based on a fixed partition lacks the necessary shift invariance to prevent residual block artifacts. A key advantage of LLR-IRPA is the simultaneous suppression of block artifacts and reduction of computational load through iterative random shifts.

This paper is organized as follows: In section II we describe the theoretical framework for locally low-rank reconstruction with and without iterative random patch adjustments. Section III presents the experimental methods for undersampled parallel imaging and quantitative parameter mapping reconstructions. In section IV, numerical and imaging results are presented which show equivalent or comparable performance of LLR-IRPA in relation to CLEAR. In Section V, we discuss potential implications and extensions of the method, followed by concluding remarks in Section VI and further theoretical considerations in the Appendix.

## II. THEORY

### A. Locally Low-Rank Regularization (LLR) based on a fixed partition of non-overlapping patches

For simplicity, we consider the case of multi-coil two-dimensional (2D) imaging, although this framework can be adapted to more general 3D or multi-contrast imaging. In the case of 2D multi-coil images, the  $M \times N$  images from all  $C$  coils are vectorized into  $\mathbf{x} \in \mathbb{C}^{MNC}$ . The forward model of the data acquisition is then represented by

$$\mathbf{y} = \mathcal{F}\mathbf{x} + \mathbf{n} \quad (1)$$

where  $\mathcal{F} : \mathbb{C}^{MNC} \mapsto \mathbb{C}^{KC}$  is the Fourier undersampling operator,  $\mathbf{y} \in \mathbb{C}^{KC}$  is the undersampled k-space data, and  $\mathbf{n} \in \mathbb{C}^{KC}$  is a vector of i.i.d Gaussian noise, where  $K < MN$ . The reduction factor (RF) for the acquisition is defined as  $MN/K$ .

To form locally low-rank submatrices, the fixed-partition LLR framework divides the underlying image into a partition  $\Omega$  of non-overlapping, covering patches. The number of patches within  $\Omega$  is denoted by  $|\Omega|$ , and each is labeled as  $q \in \Omega$ , where  $q = 1, 2, \dots, |\Omega|$ . For a given set of patch dimensions, the image plane can be divided in a number of different ways by shifting the partition by different pixel amounts along each dimension. Therefore, we denote any particular shift of the partition  $\Omega$  by  $\Omega_k$ , where  $k \in \{1, 2, \dots, N_\Omega\}$ , and  $N_\Omega$  is the total number of distinct shifts. Each patch is assumed to have dimensions  $m \times n$ , where at the image boundaries either (i) periodic boundary conditions are imposed, or (ii) the patch is zero-padded in regions beyond the boundary. This paper adopts the latter approach.

Let  $\mathcal{P}_q : \mathbb{C}^{MNC} \mapsto \mathbb{C}^{mn \times C}$  be the linear operator that extracts from  $\mathbf{x}$  the image data corresponding to the  $q^{\text{th}}$  patch of the partition  $\Omega$  and forms a matrix  $\mathcal{P}_q(\mathbf{x})$  whose  $k^{\text{th}}$  column is the vectorized patch from the  $k^{\text{th}}$  coil image,  $k = 1, 2, \dots, C$ . (Fig. 1). The underlying image is modeled as locally correlated within a relatively small region across the coil dimension. The key assumption is that the coil sensitivities are locally smooth, so that the cumulative coil sensitivity within a sufficiently small region, i.e. patch, has a low-order representation with respect to some basis. Therefore, in cases where the image patches do not consist entirely of noise, the matrix  $\mathcal{P}_q(\mathbf{x})$  will be low rank [16]. This locally-low rank property is central to the LLR regularization scheme for reconstruction from undersampled data.

Using the inner product  $\langle A, B \rangle_{\mathbb{C}^{mn \times C}} = \text{Re}(\text{tr}(A^H B))$  over  $\mathbb{C}^{mn \times C}$ , the adjoint operator  $\mathcal{P}_q^*$  satisfies

$$\langle \mathcal{P}_q(\mathbf{x}), Y \rangle_{\mathbb{C}^{mn \times C}} = \langle \mathbf{x}, \mathcal{P}_q^*(Y) \rangle_2 \quad (2)$$

for any  $\mathbf{x} \in \mathbb{C}^{MNC}$  and  $Y \in \mathbb{C}^{mn \times C}$ . Specifically, it is defined as the linear operator  $\mathcal{P}_q^* : \mathbb{C}^{mn \times C} \mapsto \mathbb{C}^{MNC}$  that maps all vectorized patches in  $\mathcal{P}_q(\mathbf{x})$  back to their respective locations in a vector in  $\mathbb{C}^{MNC}$ , where all the entries in this vector are zero except those corresponding to the  $q^{\text{th}}$  patch.

With these operations, we can now define the linear operator  $\mathcal{T}_\Omega : \mathbb{C}^{MNC} \mapsto \mathcal{X}$ , where  $\mathcal{X} \equiv \mathbb{C}^{|\Omega| \times mn \times C}$ , and each

<sup>1</sup>Reconstruction example showing block artifacts is included in the supplementary material (Figure S1).

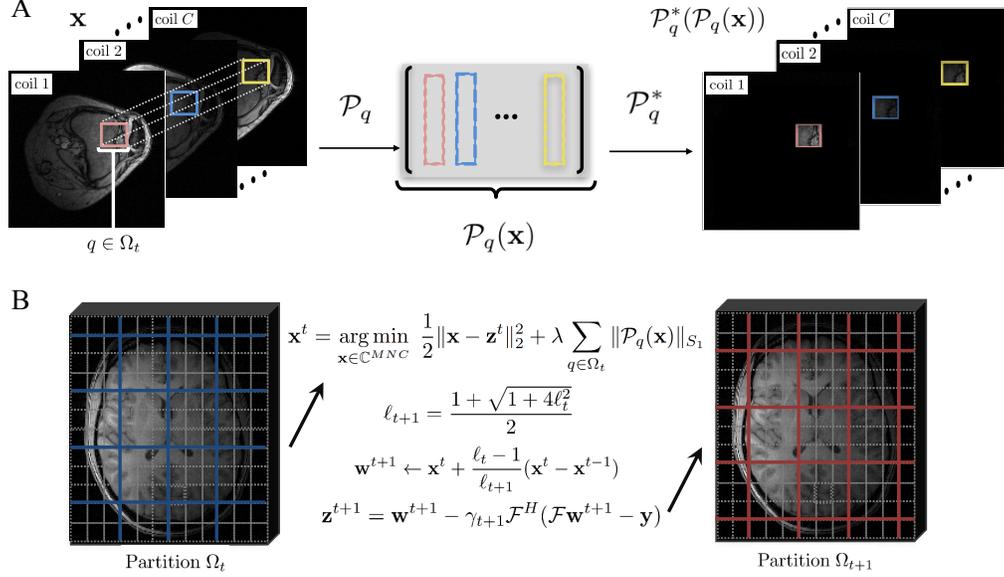


Fig. 1: (A) The patch location  $q \in \Omega_t$  from each image  $\mathbf{x}^{(j)}$  in the multi-coil set  $\mathbf{x}$  ( $j = 1, 2, \dots, C$ ) is extracted to form the columns of the locally-low rank matrix  $\mathcal{P}_q(\mathbf{x})$ . The adjoint operation is applied to reconfigure this matrix onto the original image space. (B) Illustration of the FISTA scheme with random shift of partition  $\Omega_t$  (blue) at iteration  $t$  to partition  $\Omega_{t+1}$  (red) at iteration  $t + 1$ . The patch is zero-padded in locations outside the image boundary.

component of  $\mathcal{T}_\Omega \mathbf{x}$  is given as

$$[\mathcal{T}_\Omega \mathbf{x}]_q = \mathcal{P}_q(\mathbf{x}) \quad (3)$$

for  $q = 1, 2, \dots, |\Omega|$ . For  $\mathbf{X}, \mathbf{Y} \in \mathcal{X}$ , the inner product is

$$\langle \mathbf{X}, \mathbf{Y} \rangle_{\mathcal{X}} = \sum_{q=1}^{|\Omega|} \text{Re}(\text{tr}(\mathbf{Y}_q^H \mathbf{X}_q)) \quad (4)$$

and norm  $\|\mathbf{X}\|_{\mathcal{X}} = \sqrt{\langle \mathbf{X}, \mathbf{X} \rangle_{\mathcal{X}}}$ , where the components  $\mathbf{X}_q, \mathbf{Y}_q \in \mathbb{C}^{mn \times C}$ . The adjoint  $\mathcal{T}_\Omega^* : \mathcal{X} \mapsto \mathbb{C}^{MNC}$  is the linear operator that satisfies

$$\langle \mathcal{T}_\Omega \mathbf{x}, \mathbf{Y} \rangle_{\mathcal{X}} = \langle \mathbf{x}, \mathcal{T}_\Omega^* \mathbf{Y} \rangle_2 \quad (5)$$

for any  $\mathbf{Y} \in \mathcal{X}$  and  $\mathbf{x} \in \mathbb{C}^{MNC}$ , and it is defined as

$$\mathcal{T}_\Omega^* \mathbf{Y} = \sum_{q=1}^{|\Omega|} \mathcal{P}_q^* \mathbf{Y}_q \quad (6)$$

Because all possible partitions consist only of non-overlapping, contiguous patches that completely cover the entire image [20], we also have that

$$\mathcal{T}_\Omega^*(\mathcal{T}_\Omega \mathbf{x}) = \mathbf{x} \quad (7)$$

$$\mathcal{T}_\Omega(\mathcal{T}_\Omega^* \mathbf{Y}) = \mathbf{Y} \quad (8)$$

Since the rank of a matrix is a non-convex function and minimization of rank is an NP hard problem, matrix rank is approximated by the Schatten 1-norm, also known as the nuclear norm. This norm is the closest convex relaxation of matrix rank. We recall the definition of the Schatten  $p$ -norm [23] of a matrix  $A \in \mathbb{C}^{n_1 \times n_2}$  as

$$\|A\|_{S_p} = \|\boldsymbol{\sigma}(A)\|_p \quad (9)$$

where  $\boldsymbol{\sigma}(A)$  is the vector of singular values of  $A$ ,  $\sigma_i(A)$  is the  $i^{\text{th}}$  singular value, and  $\|\cdot\|_p$  is the  $\ell_p$  norm. Based on the Schatten 1-norm, the patch-based locally-low rank regularization term can be defined in terms of the mixed  $\ell_1$ - $S_1$

norm [24] which, for an element  $\mathbf{X} \in \mathcal{X}$ , is defined as

$$\|\mathbf{X}\|_{1,1} = \sum_{q=1}^{|\Omega|} \|\mathbf{X}_q\|_{S_1} \quad (10)$$

Accordingly, the optimization problem is formulated as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{MNC}} \frac{1}{2} \|\mathbf{y} - \mathcal{F}\mathbf{x}\|_2^2 + \lambda \|\mathcal{T}_\Omega \mathbf{x}\|_{1,1} \quad (11)$$

where  $\lambda \geq 0$  is a regularization parameter that balances the trade-off between data fidelity and the locally low-rank representation of the image. Equation (11) represents the general formulation for recovering a locally-low rank image  $\mathbf{x}$  from its undersampled measurements, assuming a particular partition  $\Omega$  of non-overlapping, covering patches.

### B. Optimization

We use the ISTA formalism [25], [26] to solve (11), since the regularization functional is convex but non-smooth [24]. ISTA and its variants, such as FISTA, are majorization-minimization (MM) algorithms that successively minimize a sequence of surrogate functions that upper bound the original objective function. Using an initial estimate  $\mathbf{x}_0$ , a quadratic upper bound  $f$  of the objective function in (11) can be written as

$$f(\mathbf{x}, \mathbf{x}_0) = \frac{\alpha}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 + \lambda \|\mathcal{T}_\Omega \mathbf{x}\|_{1,1} \quad (12)$$

where  $\mathbf{z} = \mathbf{x}_0 + \frac{1}{\alpha} \mathcal{F}^H(\mathbf{y} - \mathcal{F}\mathbf{x}_0)$  and  $\alpha \geq \lambda_{\max}(\mathcal{F}^H \mathcal{F})$ . The algorithm proceeds by iteratively minimizing (12), setting  $\mathbf{x}_0$  to the solution of the previous iteration. To minimize the function in (12), we use the fact that the dual of the mixed  $\ell_1$ - $S_1$  norm is the mixed  $\ell_\infty$ - $S_\infty$  norm [24], [27]. Thus, for  $\mathbf{X} \in \mathcal{X}$ , the  $\ell_1$ - $S_1$  norm can be written equivalently as

$$\|\mathbf{X}\|_{1,1} = \max_{\boldsymbol{\Psi} \in B_{\infty, \infty}} \langle \boldsymbol{\Psi}, \mathbf{X} \rangle_{\mathcal{X}} \quad (13)$$

TABLE I: nRMSE results of restrospective undersampling along two phase-encoding directions, for the brain and knee data sets, at various reduction factors (RF) and patch sizes (PS). Results for CLEAR, LLR-IRPA, and CLEAR using iterative patch adjustmens (CLEAR-IRPA) are shown.

	PS	RF=3			RF=4			RF=5			RF=6			RF=7		
		CLEAR	CLEAR-IRPA	LLR-IRPA												
Brain	4	2.57E-2	2.54E-2	2.62E-2	3.2E-2	2.95E-2	3.01E-2	3.39E-2	3.31E-2	3.40E-2	3.71E-2	3.59E-2	3.66E-2	4.13E-2	3.93E-2	4.07E-2
	6	2.53E-2	2.52E-2	2.58E-2	2.94E-2	2.90E-2	2.96E-2	3.28E-2	3.22E-2	3.25E-2	3.56E-2	3.46E-2	3.52E-2	3.93E-2	3.82E-2	3.92E-2
	8	2.57E-2	2.54E-2	2.59E-2	2.95E-2	2.91E-2	2.96E-2	3.27E-2	3.21E-2	3.26E-2	3.53E-2	3.45E-2	3.50E-2	3.90E-2	3.80E-2	3.96E-2
	10	2.57E-2	2.58E-2	2.61E-2	2.93E-2	2.90E-2	2.95E-2	3.28E-2	3.21E-2	3.26E-2	3.54E-2	3.46E-2	3.52E-2	3.89E-2	3.79E-2	3.87E-2
	12	2.59E-2	2.57E-2	2.62E-2	2.95E-2	2.92E-2	2.99E-2	3.28E-2	3.20E-2	3.26E-2	3.53E-2	3.48E-2	3.59E-2	3.87E-2	3.79E-2	3.92E-2
	14	2.62E-2	2.60E-2	2.62E-2	2.99E-2	2.92E-2	3.00E-2	3.28E-2	3.21E-2	3.30E-2	3.55E-2	3.47E-2	3.56E-2	3.89E-2	3.81E-2	3.93E-2
	14	2.05E-2	2.05E-2	2.15E-2	2.51E-2	2.51E-2	2.61E-2	3.08E-2	3.06E-2	3.28E-2	3.82E-2	3.78E-2	3.97E-2	4.50E-2	4.32E-2	4.59E-2
Knee	6	2.13E-2	2.13E-2	2.15E-2	2.53E-2	2.53E-2	2.58E-2	2.96E-2	2.95E-2	3.03E-2	3.56E-2	3.51E-2	3.61E-2	4.06E-2	4.06E-2	4.26E-2
	8	2.18E-2	2.18E-2	2.21E-2	2.56E-2	2.56E-2	2.61E-2	2.96E-2	2.96E-2	2.98E-2	3.47E-2	3.49E-2	3.52E-2	3.92E-2	3.91E-2	4.04E-2
	10	2.22E-2	2.22E-2	2.23E-2	2.61E-2	2.59E-2	2.59E-2	2.99E-2	2.97E-2	3.01E-2	3.52E-2	3.46E-2	3.49E-2	3.89E-2	3.89E-2	3.92E-2
	12	2.26E-2	2.24E-2	2.26E-2	2.64E-2	2.63E-2	2.63E-2	3.01E-2	3.01E-2	3.02E-2	3.52E-2	3.48E-2	3.49E-2	3.90E-2	3.89E-2	3.93E-2
	14	2.29E-2	2.28E-2	2.29E-2	2.67E-2	2.69E-2	2.68E-2	3.07E-2	3.07E-2	3.09E-2	3.53E-2	3.54E-2	3.53E-2	3.94E-2	3.98E-2	4.02E-2
	14	2.05E-2	2.05E-2	2.15E-2	2.51E-2	2.51E-2	2.61E-2	3.08E-2	3.06E-2	3.28E-2	3.82E-2	3.78E-2	3.97E-2	4.50E-2	4.32E-2	4.59E-2
	14	2.13E-2	2.13E-2	2.15E-2	2.53E-2	2.53E-2	2.58E-2	2.96E-2	2.95E-2	3.03E-2	3.56E-2	3.51E-2	3.61E-2	4.06E-2	4.06E-2	4.26E-2

where the set  $B_{\infty, \infty}$  denotes the  $\ell_{\infty}$ - $S_{\infty}$  unit norm ball

$$B_{\infty, \infty} = \{\Psi \in \mathcal{X} : \|\Psi_q\|_{S_{\infty}} \leq 1, \forall q = 1, 2, \dots, |\Omega|\} \quad (14)$$

Using these definitions and the adjoint operator  $\mathcal{T}_{\Omega}^*$ , the minimization of (12) can be expressed equivalently as

$$\tilde{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{M \times N \times C}} \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 + \frac{\lambda}{\alpha} \max_{\Psi \in B_{\infty, \infty}} \langle \mathcal{T}_{\Omega}^* \Psi, \mathbf{x} \rangle_2 \quad (15)$$

Because the objective function in (15) is strictly convex in  $\mathbf{x}$  and concave in  $\Psi$ , an optimal saddle-point  $(\tilde{\mathbf{x}}, \tilde{\Psi})$  exists [24] at which the objective function attains a common value, and the order of minimization and maximization does not affect the solution. Thus, defining  $\mathcal{L}(\mathbf{x}, \Psi) = \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 + \frac{\lambda}{\alpha} \langle \mathcal{T}_{\Omega}^* \Psi, \mathbf{x} \rangle_2$ ,

$$\begin{aligned} \min_{\mathbf{x} \in \mathbb{C}^{M \times N \times C}} \max_{\Psi \in B_{\infty, \infty}} \mathcal{L}(\mathbf{x}, \Psi) &= \mathcal{L}(\tilde{\mathbf{x}}, \tilde{\Psi}) \\ &= \max_{\Psi \in B_{\infty, \infty}} \min_{\mathbf{x} \in \mathbb{C}^{M \times N \times C}} \mathcal{L}(\mathbf{x}, \Psi) \end{aligned} \quad (16)$$

From (16), one can identify the primal objective function  $\rho(\mathbf{x})$  and the dual objective function  $s(\Psi)$  as

$$\rho(\mathbf{x}) = \max_{\Psi \in B_{\infty, \infty}} \mathcal{L}(\mathbf{x}, \Psi) = \frac{1}{2} \|\mathbf{x} - \mathbf{z}\|_2^2 + \frac{\lambda}{\alpha} \|\mathcal{T}_{\Omega} \mathbf{x}\|_{1,1} \quad (17)$$

$$s(\Psi) = \min_{\mathbf{x} \in \mathbb{C}^{M \times N \times C}} \mathcal{L}(\mathbf{x}, \Psi) = \frac{1}{2} \left( \|\mathbf{z}\|_2^2 - \|\mathbf{z} - \frac{\lambda}{\alpha} \mathcal{T}_{\Omega}^* \Psi\|_2^2 \right) \quad (18)$$

Accordingly, one can find the minimizer  $\tilde{\mathbf{x}}$  of  $\rho(\mathbf{x})$  by finding the maximizer  $\tilde{\Psi}$  of  $s(\Psi)$ , using the relation

$$\tilde{\mathbf{x}} = \mathbf{z} - \frac{\lambda}{\alpha} \mathcal{T}_{\Omega}^* \tilde{\Psi} \quad (19)$$

From (18), one can see that by using (5), (7), and (8),

$$\begin{aligned} \max_{\Psi \in B_{\infty, \infty}} s(\Psi) &= \min_{\Psi \in B_{\infty, \infty}} \frac{1}{2} \|\mathbf{z} - \frac{\lambda}{\alpha} \mathcal{T}_{\Omega}^* \Psi\|_2^2 \\ &= \min_{\Psi \in B_{\infty, \infty}} \left\| \frac{\alpha}{\lambda} \mathcal{T}_{\Omega} \mathbf{z} - \Psi \right\|_{\mathcal{X}}^2 \end{aligned} \quad (20)$$

Therefore, the maximizer of (18) can be found by projecting  $\frac{\alpha}{\lambda} \mathcal{T}_{\Omega} \mathbf{z} \in \mathcal{X}$  onto the  $B_{\infty, \infty}$  unit norm ball. This projection onto  $B_{\infty, \infty}$  can be done by projecting each of the components of  $\frac{\alpha}{\lambda} \mathcal{T}_{\Omega} \mathbf{z}$  onto the unit norm ball  $B_{S_{\infty}}$ , the space of matrices with Schatten  $\infty$ -norm  $\leq 1$  [24]. If the singular value decomposition of the component matrix  $\frac{\alpha}{\lambda} [\mathcal{T}_{\Omega} \mathbf{z}]_q = \mathbf{U}_q \text{diag}(\boldsymbol{\sigma}(\frac{\alpha}{\lambda} [\mathcal{T}_{\Omega} \mathbf{z}]_q)) \mathbf{V}_q^H$ , then its projection onto  $B_{S_{\infty}}$  is

$$\begin{aligned} P_{B_{S_{\infty}}} \left( \frac{\alpha}{\lambda} [\mathcal{T}_{\Omega} \mathbf{z}]_q \right) &= \frac{\alpha}{\lambda} \mathbf{U}_q \text{diag}(\min(\boldsymbol{\sigma}([\mathcal{T}_{\Omega} \mathbf{z}]_q), \frac{\lambda}{\alpha} \mathbf{1})) \mathbf{V}_q^H \\ &= \tilde{\Psi}_q \end{aligned} \quad (21)$$

Based on (19) and (21), we conclude that

$$\tilde{\mathbf{x}} = \mathbf{z} - \frac{\lambda}{\alpha} \mathcal{T}_{\Omega}^* \tilde{\Psi} = \mathcal{T}_{\Omega}^* \left( \mathcal{T}_{\Omega} \mathbf{z} - \frac{\lambda}{\alpha} \tilde{\Psi} \right)$$

$$\Rightarrow \tilde{\mathbf{x}} = \sum_{q=1}^{|\Omega|} \mathcal{P}_q^* \left( \mathbf{U}_q \text{diag}(\mathcal{S}_{\beta}[\boldsymbol{\sigma}([\mathcal{T}_{\Omega} \mathbf{z}]_q)]) \mathbf{V}_q^H \right) \quad (22)$$

where  $\mathcal{S}_{\beta}[\boldsymbol{\sigma}(\mathbf{X}_q)] = \max(\boldsymbol{\sigma}(\mathbf{X}_q) - \beta, 0)$  is defined as the soft-thresholding operator, applied component-wise on the vector of singular values  $\boldsymbol{\sigma}(\mathbf{X}_q)$ . Importantly, we note that the minimization of  $\rho(\mathbf{x})$  is the proximal mapping of the patch-based regularizer  $\lambda \|\mathcal{T}_{\Omega} \mathbf{x}\|_{1,1}$ :

$$\tilde{\mathbf{x}} = \min_{\mathbf{x} \in \mathbb{C}^{M \times N \times C}} \rho(\mathbf{x}) \equiv \text{prox}_{\lambda \|\mathcal{T}_{\Omega} \cdot\|_{1,1}} \left( \mathbf{z}; \frac{1}{\alpha} \right) \quad (23)$$

Summarizing the above, the overall iterative soft-thresholding scheme based on FISTA becomes

$$\mathbf{z}^t \leftarrow \mathbf{w}^t - \gamma_t \mathcal{F}^H (\mathcal{F} \mathbf{w}^t - \mathbf{y}) \quad (24)$$

$$\mathbf{x}^t = \text{prox}_{\lambda \|\mathcal{T}_{\Omega} \cdot\|_{1,1}} (\mathbf{z}^t; \gamma_t) \quad (25)$$

$$\ell_{t+1} \leftarrow \frac{1 + \sqrt{1 + 4\ell_t^2}}{2} \quad (26)$$

$$\mathbf{w}^{t+1} \leftarrow \mathbf{x}^t + \frac{\ell_t - 1}{\ell_{t+1}} (\mathbf{x}^t - \mathbf{x}^{t-1}) \quad (27)$$

where  $\gamma_t \leq \frac{1}{\alpha}$  is the gradient descent step size and  $\mathbf{w}^1 = \mathcal{F}^H \mathbf{y}$ . In short, (22) represents reconstructing the image after singular-value thresholding of each matrix formed from each patch in the partition. The iterations (24) to (27) represent the FISTA technique for solving (11), using a fixed partition  $\Omega$ .

### C. Locally Low-Rank Regularization (LLR) based on overlapping patches

LLR regularization based on overlapping patches is exemplified by the state-of-the-art CLEAR algorithm [16]. In CLEAR, the entire image is covered by patches that are overlapped by a certain factor of the image patch dimensions. Suppose that patch dimensions of  $m \times n$  are used to cover a  $M \times N$  image matrix, where  $m|M$  and  $n|N$ . The amount of overlap along the first dimension is designated by a factor  $0 < r \leq 1$  of  $m$ , i.e., patches overlap each successive patch by  $rm$  pixels. Along the other dimension, patches are overlapped by  $sn$  pixels ( $0 < s \leq 1$ ). Then the total number of overlapping patches  $P$  that cover the image is

$$P = \left( \frac{N + n(1-s)}{sn} \right) \left( \frac{M + m(1-r)}{rm} \right) \quad (28)$$

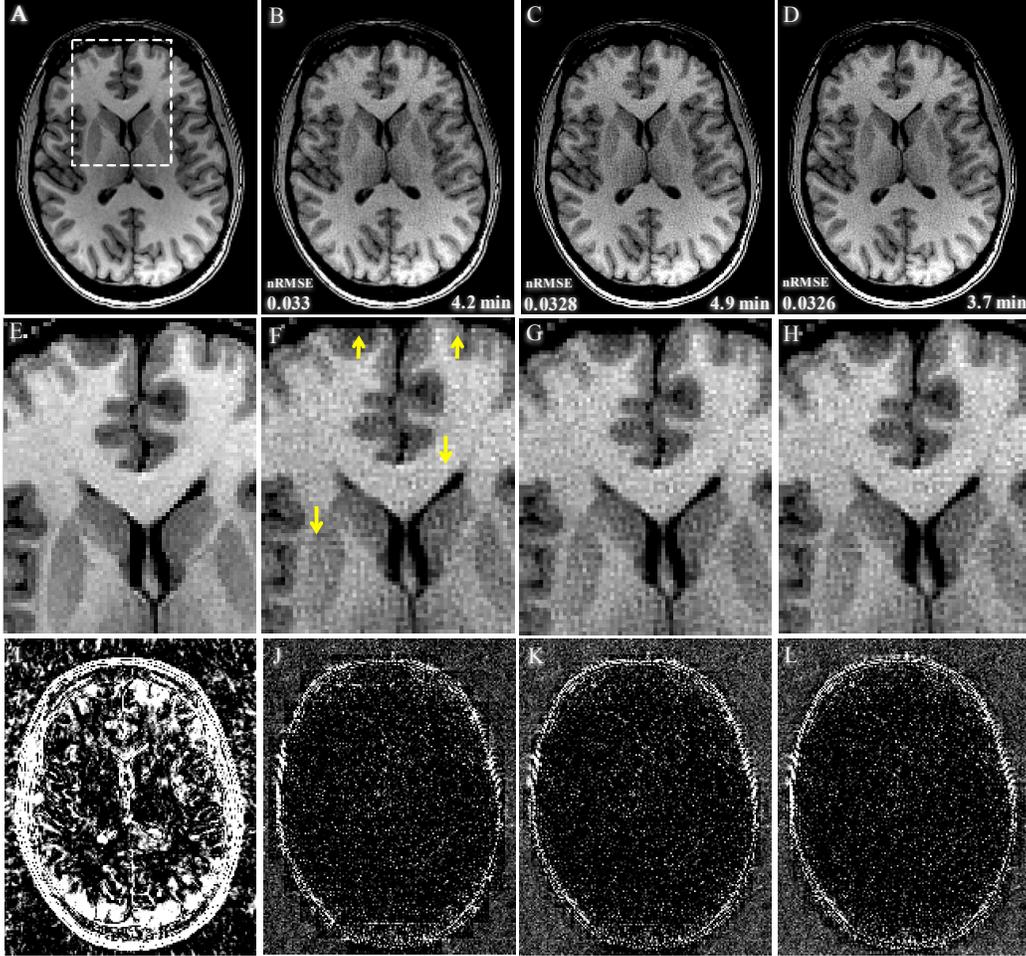


Fig. 2: Brain images reconstructed with  $RF = 5$  and  $PS = 10 \times 10$ . Reconstruction times and nRMSE values included. (A) Reference image, (B) result using non-overlapping patches, (C) CLEAR, (D) LLR-IRPA. (E)-(H) corresponding zoom-in images. (I) Absolute difference image between the reference and zero-filled reconstructed image. (J), (K), and (L) are absolute difference images for non-overlapping patch-based, CLEAR and LLR-IRPA reconstructions, respectively. Note the reduction in block artifacts (indicated by yellow arrows) using CLEAR and LLR-IRPA in (G) and (H). Window level: 2.4 - 5.6% of the maximum reference signal in (A).

Let  $\Gamma$  be the set of non-overlapping partitions that, when taken together, correspond to the entire collection of  $P$  overlapping patches. Then the optimization problem of CLEAR can be expressed within the framework described in the previous section as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \in \mathbb{C}^{M \times N \times C}} \frac{1}{2} \|\mathbf{y} - \mathcal{F}\mathbf{x}\|_2^2 + \lambda \cdot rs \sum_{\Omega \in \Gamma} \|\mathcal{T}_{\Omega}\mathbf{x}\|_{1,1} \quad (29)$$

Note that this problem entails solving for a significantly higher amount of SVD's, as compared to (11) in which at most only  $(\frac{M}{m} + 1)(\frac{N}{n} + 1)$  are computed. In addition, the regularization term must be multiplied by a factor  $rs$  in order to compensate for taking the SVD's of multiple overlapping patches. Due to overlapping patches, each image patch is actually not independent of the others, since any particular patch in the output image contains contributions from the image patches that surround it. Therefore, including this factor in the regularization term essentially represents a heuristic approach to reconstructing the final image from sets of overlapping patches [28].

#### D. Locally Low-Rank Regularization with Iterative Random Patch Adjustments (LLR-IRPA)

In the LLR regularization approach described in section B, the partition  $\Omega$  remains fixed throughout all iterations. As an alternative strategy, the proposed LLR-IRPA method updates the partition at each iteration. This modification leads to reduced appearance of block artifacts while exhibiting similar behavior to CLEAR in terms of reconstruction error and convergence rate. The partition is shifted by random amounts in each direction. The FISTA iterations are modified such that the partition updates as  $\Omega_{k_{t+1}} \leftarrow \Omega_{k_t}$  where  $k_t \in \{1, 2, \dots, N_{\Omega}\}$  is chosen at random for each iteration  $t$ , and  $\Omega_{k_t}$  runs through the  $N_{\Omega}$  possible partitions as the iterative process continues<sup>2</sup>. Note that in this case, each image patch is independent of the others throughout the reconstruction process, in contrast to CLEAR. A major feature is that LLR-IRPA achieves the property

<sup>2</sup>Algorithm pseudocode for both LLR-IRPA and CLEAR is included in the supplementary material (Figure S2).

of shift-invariance without the need of overlapping patches, leading to a much reduced computational load compared to CLEAR.

Although at each iteration we obtain a solution using a different patch arrangement, and thus a different decomposition of the image, we provide a proof which shows that the LLR-IRPA iterations converge to a solution that represents the outcome from averaging the singular value thresholdings from all unique patch-based SVD's of the entire image matrix, i.e., the regularizer is effectively shift-invariant. These results are inspired by the variational justification for cycle spinning using the wavelet transform [22], and a proof of convergence is provided in the Appendix. For simplicity, this proof is given in the context of ISTA, which is equally valid for the FISTA algorithm.

### III. METHODS

The effectiveness of the proposed algorithm was tested with retrospectively-undersampled data in parallel imaging and quantitative parameter mapping experiments<sup>3</sup>, as well as with prospectively undersampled contrast-enhanced magnetic resonance angiography (CE-MRA) data. Both the CLEAR and LLR-IRPA algorithms, including the reconstruction experiments, were implemented 'in-house' using MATLAB (The Mathworks, Natick, MA) and run on a Linux workstation with a 4.4 GHz CPU and 96 GB memory.

Two types of undersampling schemes were tested: (1) reducing the number of samples along a single phase-encoding direction (1D undersampling), and (2) reducing the number of samples along two phase-encoding directions (2D undersampling). For the 2D undersampling case, variable-density, Poisson-disk undersampling masks of various reduction factors (RF) were applied to the fully-sampled k-space to simulate accelerated acquisitions. The 2D probability density function that characterizes the sampling density was set as a normalized Gaussian with standard deviations  $\sigma_x$  and  $\sigma_y$  equal to one-fourth of the corresponding image dimensions  $N_x$  and  $N_y$ , in pixels. For the 1D undersampling case, the random sampling density was determined by a one-dimensional normalized Gaussian probability density function with standard deviation equal to one-fourth of the corresponding largest image dimension. For the cases involving quantitative parameter mapping, this standard deviation was set equal to one-fourth the size of the actual phase-encoding dimension. For the CLEAR algorithm, the extent of overlapping was set as one-half the dimensions of the image patch, so that  $r = s = \frac{1}{2}$  in (28). This choice of overlapping ratio provides a balanced trade-off between computational load and the extent of shift-invariance of CLEAR's patch-based regularizer.

We used the first-order, fast iterative soft-thresholding algorithm (FISTA) to solve the associated nuclear-norm regularized optimization problem. For simplicity, we take ISTA as the algorithmic framework with which to show theoretical results concerning the per-iteration random shifts of the image partitions. All reconstructions were performed with a maximum

<sup>3</sup>Methods and Results sections for quantitative parameter mapping experiments are included in Supplementary Section 1.

TABLE II: CLEAR and LLR-IRPA nRMSE results of retrospective undersampling along a single phase-encoding direction, for the brain and knee data sets, at various reduction factors (RF) and patch sizes (PS)

		RF=2		RF=2.5		RF=3	
		CLEAR	LLR-IRPA	CLEAR	LLR-IRPA	CLEAR	LLR-IRPA
Brain	PS						
	6	2.99E-2	3.03E-2	3.74E-2	3.69E-2	5.32E-2	5.38E-2
	8	3.02E-2	3.03E-2	3.69E-2	3.71E-2	5.36E-2	5.30E-2
	10	3.08E-2	3.07E-2	3.72E-2	3.76E-2	5.39E-2	5.37E-2
	12	3.13E-2	3.11E-2	3.80E-2	3.74E-2	5.40E-2	5.46E-2
Knee	6	2.57E-2	2.60E-2	3.10E-2	3.14E-2	3.55E-2	3.53E-2
	8	2.59E-2	2.61E-2	3.10E-2	3.13E-2	3.56E-2	3.57E-2
	10	2.68E-2	2.61E-2	3.17E-2	3.19E-2	3.63E-2	3.60E-2
	12	2.66E-2	2.66E-2	3.17E-2	3.17E-2	3.59E-2	3.60E-2

number of 100 iterations. Moreover, we set a stopping criterion dependent on the relative error between successive solutions, namely  $\|x^{t+1} - x^t\|/\|x^t\| < 10^{-5}$ . The regularization parameter  $\lambda$  for both CLEAR and LLR-IRPA was set as the estimated standard deviation of the collection of singular values from all matrices formed from all image patches, computed at each iteration. This estimate was obtained using the median absolute deviation of the collection of singular values.

The quality of the reconstruction was quantified using the normalized root mean square error (nRMSE), defined as

$$\text{nRMSE} = \frac{1}{\max(X_0)} \frac{\|X_0 - X\|_2}{\sqrt{N}} \quad (30)$$

where  $X_0$  is the true image and  $X$  is the reconstruction image, and  $N$  is the number of pixels. The algorithmic performances were compared on the basis of image quality and nRMSE values. Difference images were computed and the rate of convergence was characterized by plotting nRMSE as a function of iteration. The criteria to determine the image quality includes the extent of visible blocky artifacts and preservation of edge-features.

#### A. Parallel Imaging - Retrospective Undersampling

Two fully-sampled data sets, one of the knee and the brain, were acquired from a healthy volunteer after Institutional Review Board (IRB) approval. The data set of the knee was acquired with a 3T Skyra (Siemens Healthcare, Erlangen, Germany) MRI scanner, using a 3D GRE sequence with the following acquisition parameters: matrix size  $160 \times 160$ , isotropic resolution of  $1 \text{ mm}^2$ , TE = 3.78 ms, TR = 8.6 ms, flip angle  $15^\circ$ , and bandwidth of 810 Hz/px. The brain data set was acquired with the same scanner and sequence, using the following parameters: matrix size  $224 \times 224$ , isotropic resolution of  $1 \text{ mm}^2$ , TE = 3.5 ms, TR = 8.5 ms, flip angle  $12^\circ$ , and bandwidth of 400 Hz/px. The knee and brain data sets consisted of 15 and 20 channels, respectively.

These 3D data sets were retrospectively undersampled at reduction factors (RF) of 3, 4, 5, 6, and 7, and were reconstructed using square patch sizes (PS) of side lengths of 4, 6, 8, 10, 12, and 14 pixels. For 2D undersampling, a fully-sampled, square region with a 12-pixel side was retained in each undersampling mask. In the 1D undersampling case, 16 fully sampled central

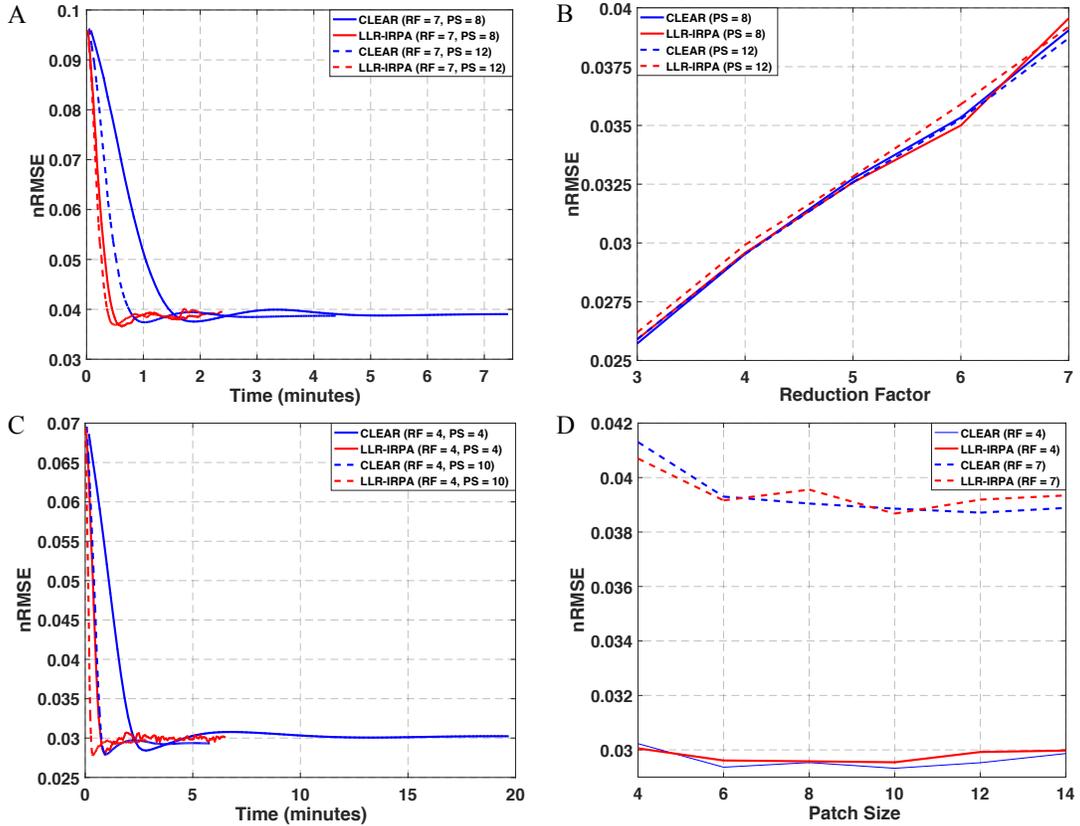


Fig. 3: (A) Algorithmic convergence in terms of nRMSE, at RF = 7 and with PS = 8 and 12. (B) nRMSE values as a function of reduction factor for PS = 8 and 12, using 2D undersampling (C) Algorithmic convergence in terms of nRMSE, at RF = 4 and with PS = 4 and 10. (D) comparing the difference in the effect of patch size on the resulting nRMSE value for RF = 4 and 7. Plots based on reconstruction results from retrospectively undersampled 2D brain image data.

phase-encoding lines were retained for undersampling the k-space of the brain data set, while 12 fully-sampled central lines were kept in the mask for undersampling the k-space of the knee data set. Each data set was reconstructed for each combination of reduction factor and patch size.

To emphasize the advantage of LLR-IRPA in terms of both computational efficiency and reconstruction accuracy, these data sets are also reconstructed with CLEAR using iterative random patch adjustments (CLEAR-IRPA). Reconstruction times and nRMSE values resulting from this additional reconstruction approach are compared with LLR-IRPA.

### B. Parallel Imaging - Prospective Undersampling

Multi-phase, contrast-enhanced cardiac- and ventilator-gated MR angiography data [29] was acquired with prospective variable density Poisson-disk undersampling pattern on a 6-month-old pediatric patient with congenital heart disease. The data was acquired with a 3T Trio (Siemens Healthcare, Erlangen, Germany) scanner. The acquisition matrix size was  $480 \times 266 \times 128$ , with a total of 12 channels, and a fully-sampled central  $24 \times 24$  square region. Further image acquisition details are as described in [29]. This data was prospectively underampled with RF = 6.5, and was subsequently

reconstructed using CLEAR and LLR-IRPA with PS = 4, 6, 8, 10, 12, and 14.

## IV. RESULTS

### A. Parallel Imaging - Retrospective Undersampling

As shown in Fig. 2, reconstructions from the retrospectively undersampled data sets<sup>4</sup> show that LLR-IRPA leads to similar reduction in block artifacts and comparable fidelity to the actual images, in relation to the results from CLEAR. This observation can be seen from the reconstructed and difference images, in which LLR-IRPA exhibits minimal block artifacts and recovers structural features just as well or better than CLEAR. We observe also that, in terms of nRMSE, the larger patch sizes (PS = 10,12,14) lead to better recovery of images undersampled at the higher reduction factors (RF = 5,6,7). For comparison, we also include in the figures images reconstructed with LLR regularization using non-overlapping patches.

Numerical results listed in Table I and Table II show that LLR-IRPA is as stable as CLEAR with respect to changes in patch size, in terms of nRMSE. The nRMSE values from LLR-IRPA regularization follow CLEAR's trend as the reduction

<sup>4</sup>Results for the knee data set are included in the supplementary material (Figure S3).

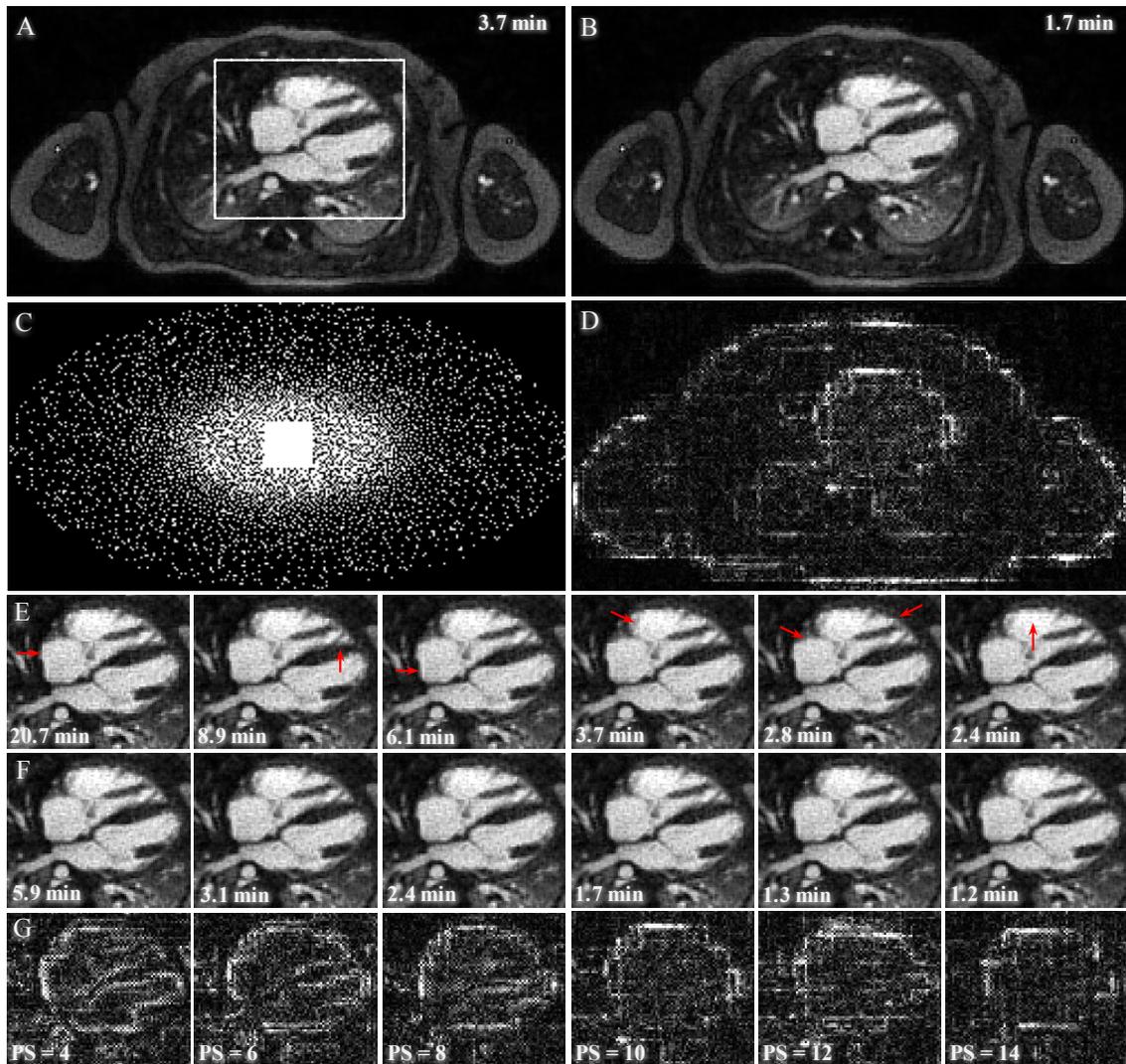


Fig. 4: Reconstruction results and times for the prospectively undersampled MRA data set, at a reduction factor of 6.5. (A) CLEAR result and (B) LLR-IRPA result using PS = 10. (C) undersampled variable-density k-space sampling. (D) difference between (A) and (B). Cropped image series (E) shows the progression of CLEAR imaging results as patch size increases from 4 to 14. Red arrows indicates block artifacts in the CLEAR reconstruction. (F) Cropped image series from LLR-IRPA, in which block artifacts are more suppressed in comparison to CLEAR. (G) Difference images between (E) and (F). Window level: 0 - 7.6% of the maximum signal in (A).

factor increases, for any given patch size (Fig. 3). Similarly, at a fixed reduction factor, the nRMSE values obtained from LLR-IRPA are as low or slightly lower than those produced from CLEAR. Finally, it is evident that the LLR-IRPA strategy leads to convergence of the algorithm to a final value that is very close or equal to that of images reconstructed by CLEAR, although at a much faster rate. These results are consistent with what is expected because, similar to the case of wavelet cycle spinning, the use of various shifted patch arrangements reduces the appearance of block artifacts and avoids the extra computational cost of overlapping patches. In terms of the time per iteration for 2D multi-coil reconstruction experiments, Table III demonstrates that LLR-IRPA is computationally more efficient than CLEAR, being approximately 3-4 times faster per iteration.

To show that LLR-IRPA is indeed shift-invariant at no extra computational cost, we also include in Table II and Table III nRMSE values and computation times from the CLEAR algorithm implemented with per-iterative random shifting. As seen from the tables, the use of per-iterative random shifting in addition to overlapping patches in the CLEAR algorithm results in almost the same performance compared to LLR-IRPA. However, the computational time per iterations is similar to that of conventional CLEAR. Thus, LLR-IRPA achieves a performance that would result from combining the benefits of overlapping patches and random shifting, but with no extra computational cost. Thus, LLR-IRPA retains a definite advantage in both computational efficiency and reconstruction accuracy.

### B. Parallel Imaging - Prospective Undersampling

Similar to the results from retrospective undersampling, the prospectively undersampled MRA data set shows comparable results in image quality, except that CLEAR displays more apparent block artifacts than LLR-IRPA within the myocardium and along its edges, as shown in Fig. 4. Due to the shift invariance induced from using different image partitions, the LLR-IRPA reconstructed images remain relatively stable and less prone to block artifacts as a function of patch size. We also see in the difference images between LLR-IRPA and CLEAR that various residual, block artifacts that remain due to the overlapping patches used by CLEAR. One can also observe that LLR-IRPA also avoids artifacts along edges of the anatomy. Importantly, note that the difference images in Fig. 4 indicate the contrast in the levels of residual blocks artifacts produced by CLEAR. This result shows that LLR-IRPA can avoid these types of artifacts without compromising image quality. Unlike the retrospectively undersampled experiments, this data retained a much larger central sampling region of size  $24 \times 24$ , and slightly more dense sampling in the central region of k-space. Even in such a case, LLR-IRPA still provides an advantage in terms of robustness to differences in patch size.

## V. DISCUSSION

Although the imaging results show the similarity of the LLR-IRPA and CLEAR reconstruction approaches, it is important to note that these results are affected by the choice of sampling pattern, the sampling density, and the extent of the fully sampled central region, both in the 2D and 1D undersampling experiments. Even though the shift invariance due the random shifting of the patch grid can be reasonably expected to provide consistent improvement in image quality, it may provide more advantageous results in other instances where the sampling density or fully sampled region may not be so favorable. This type of case may arise in a pure calibrationless setting in which no fully sampled region may be acquired, or in cases where the sampling scheme does not produce an incoherent sampling pattern. In these cases, LLR-IRPA may still have an advantage due to the added redundancy and shift invariance that results from iterative shifting of the image partition. It is important to note that the method of [17], which also shifts partitions at each iteration, depends on an auto-calibration acquisition, so that it may not be as applicable to general sampling schemes as the calibrationless LLR-IRPA technique, or CLEAR. The LLR-IRPA algorithm is applicable for arbitrary undersampled k-space trajectories without incurring the higher computational cost of CLEAR. Although CLEAR yields equivalent results, its algorithmic framework implicitly assumes that each patch is independent when applying singular-value thresholding and reconstructing each patch, which is not necessarily the case because this method uses overlapping patches. The step of dividing the resulting image by the number of times that each pixel is overlapped, as in (29), is at best a heuristic approach for reconstructing the final image from multiple overlapping patches [28]. This approach only renders the optimization problem of CLEAR more mathematically tractable. In contrast, the

LLR-IRPA algorithm treats each patch independently from the others throughout the iterative process, and the framework described in the Theory section provides a mathematically justifiable basis for solving the optimization scheme with shifting, non-overlapping patches.

As mentioned in the Theory section, each partition  $\Omega$  can be shifted in a number of  $N_\Omega$  different ways. In contrast to the wavelet cycle spinning strategy, each shift is performed randomly and the number of shifts is dependent on the prescribed patch size. The number of iterations needed to traverse all possible shifts is at least  $N_\Omega$ , and this number increases as the patch size becomes larger. In the reconstruction performed in this study, the limit of 100 iterations can be approximately sufficient for patch sizes  $\leq 10 \times 10$ . However, results for patch sizes of  $12 \times 12$  and  $14 \times 14$  may not reflect those that would be obtained using the corresponding greater number of iterations. Despite the lower number of iterations used, the reconstruction results for these patch sizes still demonstrate considerable improvement, as seen in the MRA images where block artifacts are clearly suppressed in reconstructions with LLR-IRPA. This result shows the robustness of the LLR-IRPA strategy, which maintains improved computational efficiency compared to CLEAR without introducing block artifacts.

In terms of computational complexity, LLR-IRPA has a significant advantage over CLEAR, without sacrificing algorithmic performance or image quality. In terms of computational cost, to reconstruct a 2D multi-coil  $M \times N \times C$  image set using  $m \times n$  patches (where  $m|N$  and  $n|C$ ), LLR-IRPA would require at most  $Q = (\frac{M}{m} + 1)(\frac{N}{n} + 1)$  SVD computations of  $mn \times C$  matrices per iteration, whereas CLEAR requires  $P = (\frac{N+n(1-s)}{sn})(\frac{M+m(1-r)}{rm})$ , where  $r$  and  $s$  are as given in (28). Suppose, as adopted in the experiments, that  $r = s = \frac{1}{2}$ , and that  $\frac{M}{m} = a$  and  $\frac{N}{n} = b$ . Then  $Q = ab + a + b + 1$  while  $P = 4ab + 3(a + b) + \frac{9}{4} \approx 4ab + 3(a + b) + 3$ . To quote the result given in [16], the number of floating point operations (FLOPS) required for one thin SVD calculation would be

$$\text{FLOPS} = \min(14mnC^2 + 8C^3, 6mnC^2 + 20C^3)$$

Thus, as seen from the comparison of  $P$  and  $Q$ , this means that CLEAR would require approximately 3 to 4 times more FLOPS than LLR-IRPA. This estimation also agrees with the results in Table III, which show that the average per-iteration time of LLR-IRPA can be roughly 3 to 4 times shorter than that of CLEAR. Evidently, this makes LLR-IRPA a much more efficient reconstruction scheme that produces comparable if not better imaging results to CLEAR. However, it is important to note that the use of graphical processing units (GPU) would highly accelerated these iteration times, yet the efficiency of LLR-IRPA would still outperform CLEAR.

In the context of local, patch-based methods for accelerated quantitative parameter mapping, we also note that alternative SVD-based techniques such as dictionary learning for blind compressive sensing also show promise in reducing the computational load of locally low-rank regularized reconstruction [30], [31]. In the absence of a learning step, however, LLR-IRPA represents a highly efficient class of algorithms for implementing effectively shift-invariant locally low-rank reconstruction in a calibrationless setting. Globally low rank

TABLE III: Average time (seconds) **per iteration** for CLEAR, LLR-IRPA, and CLEAR using iterative random patch adjustments (CLEAR-IRPA) at various reduction factors (RF) and patch sizes (PS), for retrospective undersampling experiments. The matrix sizes for the brain and knee images are  $224 \times 224 \times 20$  and  $160 \times 160 \times 15$ , respectively.

		RF=3			RF=4			RF=5			RF=6			RF=7		
PS		CLEAR	CLEAR-IRPA	LLR-IRPA												
Brain	4	12.7	19.1	3.4	11.8	18.8	3.9	10.5	18.7	4.2	9.7	17.7	3.3	9.6	19.7	3.3
	6	7.3	9.6	2.0	5.9	9.5	2.3	5.2	9.3	2.5	5.1	9.3	1.8	5.2	10.3	1.7
	8	8.2	8.2	1.5	6.0	8.7	2.4	4.1	6.6	3.8	4.4	6.3	1.9	4.5	8.7	1.4
	10	7.5	6.3	1.6	3.5	5.4	1.7	3.3	4.9	2.5	3.4	5.9	1.5	3.4	6.6	1.3
	12	4.3	4.5	1.9	2.3	3.9	2.0	2.6	3.8	1.9	2.6	4.9	1.1	2.6	4.8	1.2
14	2.1	3.6	1.6	2.5	3.8	1.6	2.2	3.1	1.1	2.1	4.5	1.0	2.2	4.0	1.4	
Knee	4	5.9	8.8	1.5	6.0	8.6	1.7	5.4	8.9	1.8	4.9	8.5	2.0	4.9	9.3	3.0
	6	4.7	4.9	0.9	5.7	6.0	1.5	3.4	4.8	1.5	3.9	5.1	1.7	3.2	5.2	1.7
	8	2.4	2.8	0.6	3.2	3.5	1.0	2.1	2.9	1.0	2.1	3.5	1.1	1.7	2.8	1.3
	10	1.7	2.0	0.5	2.1	2.4	0.8	1.4	2.0	0.7	1.4	2.5	1.1	1.2	2.0	0.8
	12	1.4	1.5	0.4	2.0	1.9	0.7	1.0	1.6	0.6	1.2	1.9	1.0	1.0	1.5	0.9
14	1.4	1.5	0.4	1.8	1.6	0.6	0.8	1.4	0.5	1.0	1.6	0.8	0.8	1.4	0.7	

regularization techniques, such as the k-space domain based SAKE method for parallel imaging [11], can also operate in a calibrationless setting, yet they may entail significantly more FLOPS per iteration than the locally low rank LLR-IRPA and CLEAR methods, as discussed in [16].

Further considerations include examining the performance of LLR-IPRA in more diverse acquisition and reconstruction settings, for example in dynamic reconstructions using more advanced algorithms based on constrained formulations and accelerated convergence schemes. However, while more advance algorithms may improve the convergence rates of LLR-IRPA and CLEAR, the relative computational efficiency between the two methods would not change because CLEAR would still require the computation of many more SVD's per iteration than LLR-IRPA. Therefore, the reconstruction speed of each method may increase, but the relative improvement of LLR-IRPA over CLEAR would remain valid.

## VI. CONCLUSION

In this paper, we have shown that LLR-IRPA retains the same level of image reconstruction and quantitative parameter mapping results compared to overlapping patch-based LLR regularization, in terms of image quality and nRMSE, but with the distinct advantage of substantially reduced computational load. We describe this patch adjustment strategy for LLR regularization and set a theoretical framework for formulating this novel development in the context of patch-based image reconstruction techniques. This technique is an improvement over conventional LLR-based algorithms such as CLEAR, since the computational load is substantially reduced without promoting block artifacts. In fact, the amount of acceleration from using random shifting has been shown to reach up to a factor of 3 to 4. The iterative random patch adjustment strategy is experimentally shown to suppress these artifacts while retaining the convergence rate of the more computationally expensive CLEAR algorithm. Experimental results and theoretical analysis of the proposed method support these findings. The implications of LLR-IRPA is to facilitate the application of LLR-based regularization for clinical MRI applications.

## ACKNOWLEDGMENT

This study was supported in part by the National Institute of Health (NIH) under Grant No. NIH U01-HD087221 and by the UCLA Eugene Cota-Robles Fellowship. The author thanks Dr. Joshua Trzasko for sharing his implementation of CLEAR, as well as Fei Han and Ziwu Zhou for providing CE-MRA data for our prospective undersampling experiments.

## APPENDIX A

### PROOF OF SHIFT INVARIANCE

In order to justify the shift invariance that results from random iterative patch adjustments, we prove the result for the deterministic case, in which the iterations traverse all possible partition shifts. Similar to the case of cycle spinning using the wavelet transform, in the context of patch-based LLR-IRPA regularization, the partition  $\Omega$  is changed by shifting it by some amount at each iteration. For simplicity, this proof of convergence is given in the context of ISTA and is equally valid for the case of FISTA which is an accelerated version of ISTA [26]. In the context of ISTA, the iterations are given by

$$\mathbf{z}^t = \mathbf{x}^{t-1} - \gamma_t \mathcal{F}^H (\mathcal{F} \mathbf{x}^{t-1} - \mathbf{y}) \quad (31)$$

$$\mathbf{x}^t = \text{prox}_{\lambda \|\cdot\|_{1,1}} (\mathbf{z}^t; \gamma_t) \quad (32)$$

Using the notation introduced in the Theory section, we write the cost function associated with the partition  $\Omega_k$  as

$$H_k(\mathbf{x}) = \underbrace{\frac{1}{2} \|\mathbf{y} - \mathcal{F} \mathbf{x}\|_2^2}_{D(\mathbf{x})} + \underbrace{\lambda \|\mathcal{T}_{\Omega_k} \mathbf{x}\|_{1,1}}_{G_{\Omega_k}(\mathbf{x})} \quad (34)$$

We write the cost function whose regularization term represents the average of the nuclear norms of all locally low-rank matrices formed from the  $N_\Omega$  unique shifts of  $\Omega$ :

$$\begin{aligned} h(\mathbf{x}) &= \frac{1}{N_\Omega} \sum_{k=1}^{N_\Omega} H_k(\mathbf{x}) \\ &= D(\mathbf{x}) + \frac{1}{N_\Omega} \sum_{k=1}^{N_\Omega} G_{\Omega_k}(\mathbf{x}) \end{aligned} \quad (35)$$

The following assumptions are made:

- The feasible set  $\mathcal{C} \subseteq \mathbb{C}^{MNC}$  is nonempty, convex, closed and bounded, i.e.,  $\exists d > 0$  such that for any  $\mathbf{x}, \mathbf{y} \in \mathcal{C}$ ,  $\|\mathbf{x} - \mathbf{y}\|_2 \leq d$ .
- The data fidelity term  $D$  is continuously differentiable with Lipschitz continuous gradient, i.e.,  $\exists L > 0$  such that  $\forall \mathbf{x}, \mathbf{y} \in \mathcal{C}$ ,  $\|\nabla D(\mathbf{x}) - \nabla D(\mathbf{y})\|_2 \leq L \|\mathbf{x} - \mathbf{y}\|_2$

- The gradient of  $D$  is bounded and the subgradients of  $G_{\Omega_k}$  are bounded, i.e.,  $\exists S > 0$  such that  $\forall \mathbf{x} \in \mathcal{C}$ ,  $\|\nabla D(\mathbf{x})\|_2 \leq S$  and  $\|\partial G_{\Omega_k}(\mathbf{x})\|_2 \leq S$ .

Referring to Theorem 1 of [22], if  $L > \lambda_{\max}(\mathcal{F}^H \mathcal{F})$  and the step size  $\gamma_t = 1/L\sqrt{t}$ , then we claim that the sequence  $\{\mathbf{x}^t\}$  generated according to Equations (31) and (32) satisfies

$$\lim_{t \rightarrow \infty} h(\mathbf{x}^t) = h^* \quad (36)$$

where  $h^* = \min_{\mathbf{x} \in \mathcal{C}} h(\mathbf{x})$ . The claim in (36) essentially says that shifting the partitions throughout the iteration process leads to the minimization of the cost function that simultaneously minimizes the nuclear norms of the submatrices formed from all partition shifts. Therefore, the underlying regularizer in (35) is effectively shift-invariant.

For simplicity, we consider the case of an image with square dimensions  $N \times N$  and square patch sizes ( $m = n$ ), such that  $n$  divides  $N$  (results for arbitrary dimensions can be obtained through similar arguments). In this case  $N_{\Omega} = n^2$  and the number of patches  $|\Omega_k|$  within any partition  $\Omega_k$  falls within  $[(\frac{N}{n})^2, (\frac{N}{n})^2 + 2\frac{N}{n} + 1]$ . We also note that in patch-based LLR regularization, it is generally assumed that each of the  $(N + n)^2$  possible (overlapping) patch locations lead to low-rank submatrices. Therefore, all submatrices  $\mathcal{P}_q(\mathbf{x})$  formed by  $q \in \Omega_k$  for any  $k$ , are assumed low-rank.

We first characterize the subdifferential of the function  $G_{\Omega_k}(\mathbf{x}) = \lambda \|\mathcal{T}_{\Omega_k} \mathbf{x}\|_{1,1}$  to show that the subgradients are indeed bounded [33]. To do so, we find the conjugate function of the mixed  $\ell_1$ - $S_1$  norm. Denote the function  $F(\mathbf{X}) = \|\mathbf{X}\|_{1,1}$  for  $\mathbf{X} \in \mathcal{X}$  and denote the function  $g(\mathbf{x}) = \mathcal{T}_{\Omega_k} \mathbf{x}$  for  $\mathbf{x} \in \mathbb{C}^{MNC}$ . Using the duality between the  $\ell_1$ - $S_1$  norm and the  $\ell_{\infty}$ - $S_{\infty}$  norm, it is straightforward to show the conjugate function  $F^*$  of  $F$  is

$$F^*(\mathbf{X}) = \begin{cases} 0 & \|\mathbf{X}\|_{\infty, \infty} \leq 1 \\ \infty & \text{otherwise,} \end{cases} \quad (37)$$

i.e., the conjugate is the indicator function of the  $\ell_{\infty}$ - $S_{\infty}$  unit norm ball. Recalling the well-known fact from convex analysis [33], [34] that if  $\mathbf{G} \in \partial F(\mathbf{X})$  then

$$F^*(\mathbf{G}) + F(\mathbf{X}) = \langle \mathbf{G}, \mathbf{X} \rangle_{\mathcal{X}} \quad (38)$$

and noting that  $G_{\Omega_k}(\mathbf{x}) = \lambda(F \circ g)(\mathbf{x}) = \lambda F(g(\mathbf{x}))$ , one can use the chain rule to characterize the subdifferential of  $G_{\Omega_k}$  at  $\mathbf{x}$  as the set

$$\begin{aligned} \partial G_{\Omega_k}(\mathbf{x}) = \\ \{ \lambda \mathcal{T}_{\Omega_k}^* \mathbf{Y} \in \mathbb{C}^{MNC} \mid \|\mathbf{Y}\|_{\infty, \infty} \leq 1, \\ \langle \mathcal{T}_{\Omega_k}^* \mathbf{Y}, \mathbf{x} \rangle_2 = \|\mathcal{T}_{\Omega_k} \mathbf{x}\|_{1,1} \} \end{aligned} \quad (39)$$

This characterization implies that the subgradients of  $G_{\Omega_k}(\mathbf{x}) = \lambda \|\mathcal{T}_{\Omega_k} \mathbf{x}\|_{1,1}$  are indeed bounded. We also recall the important fact from convex analysis that if

$$\mathbf{x} = \text{prox}_{\lambda \|\mathcal{T}_{\Omega_k} \cdot\|_{1,1}}(\mathbf{z}; \gamma_t) \quad (40)$$

then  $\mathbf{z} - \mathbf{x} \in \lambda \partial \|\mathcal{T}_{\Omega_k} \mathbf{x}\|_{1,1}$ .

Given these observations, we can follow the steps as in the proof of Lemma 1 [22] to conclude that  $\forall \mathbf{x}^* \in \mathcal{C}$ ,

$$\begin{aligned} H_{k_t}(\mathbf{x}^t) - H_{k_t}(\mathbf{x}^*) \leq \\ \frac{1}{2\gamma_t} (\|\mathbf{x}^{t-1} - \mathbf{x}^*\|_2 - \|\mathbf{x}^t - \mathbf{x}^*\|_2) + 6\gamma_t S^2 \end{aligned} \quad (41)$$

For convenience, we restate Lemma 2 of [22] in this context

as

$$\lim_{n \rightarrow \infty} \left\{ \frac{1}{nN_{\Omega}} \sum_{t=1}^{nN_{\Omega}} H_{k_t}(\mathbf{x}^t) \right\} = h(\bar{\mathbf{x}}) \quad (42)$$

where  $\{\mathbf{x}^t\}$  is a sequence in  $\mathcal{C}$  such that  $\mathbf{x}^t \rightarrow \bar{\mathbf{x}}$ .

In a similar manner to the proof of Theorem 1 in [22], if we let  $\mathbf{x}^*$  denote a minimizer of the function  $h$  and sum the bound in (41), then

$$\sum_{t=1}^{nN_{\Omega}} (H_{k_t}(\mathbf{x}^t) - H_{k_t}(\mathbf{x}^*)) \leq \frac{d^2}{2\gamma_{nN_{\Omega}}} + 6S^2 \sum_{t=1}^{nN_{\Omega}} \gamma_t \quad (43)$$

Choosing the step size  $\gamma_t = 1/L\sqrt{t}$  and dividing both sides of (43) by  $nN_{\Omega}$ , the above inequality can be simplified to

$$\frac{1}{nN_{\Omega}} \sum_{t=1}^{nN_{\Omega}} H_{k_t}(\mathbf{x}^t) - h(\mathbf{x}^*) \leq \frac{C}{\sqrt{n}} \quad (44)$$

for a constant  $C$  depending on the parameters  $d$ ,  $S$  and  $L$  defined above. Therefore, using (42) and (44),

$$\begin{aligned} 0 \leq h(\bar{\mathbf{x}}) - h(\mathbf{x}^*) \\ = \lim_{n \rightarrow \infty} \left\{ \frac{1}{nN_{\Omega}} \sum_{t=1}^{nN_{\Omega}} H_{k_t}(\mathbf{x}^t) \right\} - h^* \leq 0 \end{aligned} \quad (45)$$

where the first inequality is due to the optimality of  $\mathbf{x}^*$ . This result shows that the sequence generated by (31) and (32) converges to a minimizer  $\mathbf{x}^*$  of the function  $h$ , which is the cost-function that incorporates the effective shift-invariant regularizer that penalizes the nuclear norms of all submatrices that can be formed from all the  $N_{\Omega}$  shifts of the partition  $\Omega$ . As stated above, this argument considers the deterministic shifting strategy in which all possible  $N_{\Omega}$  partitions are traversed throughout the iteration process. However, as is the case for wavelet-based cycle spinning, the most practical method to implement this technique is to shift the partition randomly at each iteration. Although it considers the deterministic case, this proof serves as a justification for the more practical method of iterative random shifting.

## REFERENCES

- [1] M. Murphy, M. Alley, J. Demmel, K. Keutzer, S. Vasanawala, and M. Lustig, "Fast-SPIRiT compressed sensing parallel imaging MRI: Scalable parallel implementation and clinically feasible runtime," *IEEE Trans. Med. Imag.*, vol. 31, no. 6, pp. 1250–1262, 2012.
- [2] S. Vasanawala, M. Murphy, M. T. Alley, P. Lai, K. Keutzer, J. M. Pauly, and M. Lustig, "Practical parallel imaging compressed sensing MRI: Summary of two years of experience in accelerating body MRI of pediatric patients," in *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1039–1043.
- [3] S. S. Vasanawala, M. T. Alley, B. A. Hargreaves, R. A. Barth, J. M. Pauly, and M. Lustig, "Improved pediatric MR imaging with compressed sensing 1," *Radiology*, vol. 256, no. 2, pp. 607–616, 2010.
- [4] A. Majumdar and R. K. Ward, "Exploiting rank deficiency and transform domain sparsity for MR image reconstruction," *Magn. Reson. Imag.*, vol. 30, no. 1, pp. 9–18, 2012.
- [5] J. P. Haldar and Z.-P. Liang, "Spatiotemporal imaging with partially separable functions: A matrix recovery approach," in *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, 2010, pp. 716–719.
- [6] —, "Low-rank approximations for dynamic imaging," in *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, 2011, pp. 1052–1055.
- [7] J. Trzasko, "Exploiting local low-rank structure in higher-dimensional MRI applications," in *SPIE Optical Engineering and Applications*. Int. Society for Optics and Photonics, 2013, pp. 821–885.
- [8] A. Majumdar, "Improved dynamic MRI reconstruction by exploiting sparsity and rank-deficiency," *Magn. Reson. Imag.*, vol. 31, no. 5, pp. 789–795, 2013.

- [9] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic MRI exploiting sparsity and low-rank structure: k-t SLR," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1042–1054, 2011.
- [10] R. Otazo, E. Candès, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components," *Magn. Reson. Med.*, 2014.
- [11] P. J. Shin, P. E. Larson, M. A. Ohliger, M. Elad, J. M. Pauly, D. B. Vigneron, and M. Lustig, "Calibrationless parallel imaging reconstruction based on structured low-rank matrix completion," *Magn. Reson. Med.*, vol. 72, no. 4, pp. 959–970, 2014.
- [12] M. Chiew, S. Smith, P. Koopmans, N. Graedel, T. Blumensath, and K. Miller, "k-t FASTER: Acceleration of functional MRI data acquisition using low rank constraints," *Magn. Reson. Med.*, 2014.
- [13] B. Zhao, J. P. Haldar, C. Brinegar, and Z.-P. Liang, "Low rank matrix recovery for real-time cardiac MRI," in *IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*. IEEE, 2010, pp. 996–999.
- [14] B. Zhao, W. Lu, T. K. Hitchens, F. Lam, C. Ho, and Z.-P. Liang, "Accelerated MR parameter mapping with low-rank and sparsity constraints," *Magn. Reson. Med.*, 2014.
- [15] J. Trzasko, A. Manduca, and E. Borisch, "Local versus global low-rank promotion in dynamic MRI series reconstruction," in *Proc. Int. Symp. Magn. Reson. Med.*, 2011, p. 4371.
- [16] J. D. Trzasko and A. Manduca, "Calibrationless parallel MRI using CLEAR," in *Record of the 45th Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*. IEEE, 2011, pp. 75–79.
- [17] T. Zhang, J. M. Pauly, and I. R. Levesque, "Accelerating parameter mapping with a locally low rank constraint," *Magn. Reson. Med.*, 2014.
- [18] T. Zhang, J. Y. Cheng, A. G. Potnick, R. A. Barth, M. T. Alley, M. Uecker, M. Lustig, J. M. Pauly, and S. S. Vasanawala, "Fast pediatric 3D free-breathing abdominal dynamic contrast enhanced MRI with high spatiotemporal resolution," *Magn. Reson. Imag.*, vol. 41, no. 2, pp. 460–473, 2015.
- [19] R. Dony, "Karhounen-Loeve transform," *The Transform and Data Compression Handbook*, 2004.
- [20] Y. Xu and W. Yin, "A fast patch-dictionary method for whole image recovery," *UCLA CAM Report*, 2013.
- [21] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM Journal on Optimization*, vol. 20, no. 4, pp. 1956–1982, 2010.
- [22] U. S. Kamilov, E. Bostan, and M. Unser, "Variational justification of cycle spinning for wavelet-based solutions of inverse problems," *IEEE Signal Process. Lett.*, vol. 21, no. 11, pp. 1326–1330, 2014.
- [23] R. Bhatia, *Matrix Analysis*. New York: Springer-Verlag, 1997.
- [24] S. Lefkimmiatis, J. P. Ward, and M. Unser, "Hessian Schatten-norm regularization for linear inverse problems," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1873–1888, 2013.
- [25] M. Guerquin-Kern, M. Häberlin, K. P. Pruessmann, and M. Unser, "A fast wavelet-based reconstruction method for magnetic resonance imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 9, pp. 1649–1660, 2011.
- [26] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM Journal on Imaging Sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [27] J. Liu and J. Ye, "Efficient  $l1/lq$  norm regularization," *arXiv preprint arXiv:1009.4766*, 2010.
- [28] J. Mairal, F. Bach, J. Ponce, G. Sapiro, and A. Zisserman, "Non-local sparse models for image restoration," in *Computer Vision, 2009 IEEE 12th International Conference on*. IEEE, 2009, pp. 2272–2279.
- [29] F. Han, S. Rapacchi, S. Khan, I. Ayad, I. Salusky, S. Gabriel, A. Plotnik, J. Paul Finn, and P. Hu, "Four-dimensional, multiphase, steady-state imaging with contrast enhancement (MUSIC) in the heart: A feasibility study in children," *Magn. Reson. Med.*, 2014.
- [30] S. Bhave, S. G. Lingala, C. P. Johnson, V. A. Magnotta, and M. Jacob, "Accelerated whole-brain multi-parameter mapping using blind compressed sensing," *Magn Reson Med*, 2015.
- [31] Y. Zhu, Q. Zhang, Q. Liu, Y.-X. J. Wang, X. Liu, H. Zheng, D. Liang, and J. Yuan, "Panda-t1 $\rho$ : Integrating principal component analysis and dictionary learning for fast t1 $\rho$  mapping," *Magnetic resonance in medicine*, vol. 73, no. 1, pp. 263–272, 2015.
- [32] J. P. Wansapura, S. K. Holland, R. S. Dunn, and W. S. Ball, "NMR relaxation times in the human brain at 3.0 tesla," *Magn. Reson. Imag.*, vol. 9, no. 4, pp. 531–538, 1999.
- [33] A. S. Lewis, "The convex analysis of unitarily invariant matrix functions," *Journal of Convex Analysis*, vol. 2, no. 1, pp. 173–183, 1995.
- [34] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.