FAST LOW RANK COLUMN-WISE COMPRESSIVE SENSING FOR ACCELERATED DYNAMIC MRI

Silpa Babu, Seyedehsara (Sara) Nayer*, Sajan Goud Lingala**, Namrata Vaswani**

*ECE dept, Iowa State University, USA
**BME dept, University of Iowa, USA

ABSTRACT

In recent work we developed a fast and sample-efficient gradient descent (GD) solution to the following "Low Rank column-wise Compressive Sensing (LRcCS)": recover an $n \times q$, rank-r matrix $\boldsymbol{X}^*$ from measurements $\boldsymbol{y}_k = \boldsymbol{A}_k \boldsymbol{x}_k^*$, $k = 1, 2, \dots, q$ when each $\boldsymbol{y}_k$ is an m-length vector with $m < n$, and the rank $r \ll \min(n, q)$. Accelerated dynamic MRI is a key application where this problem occurs. In this work, we show the power of our approach (and of its modification for the MRI setting) for four very different highly undersampled dynamic MRI applications. Without any application-specific parameter tuning, in most settings, our approach outperforms the state-of-the-art MRI methods, while also being significantly faster in all settings.

Index Terms— low-rank, compressed sensing, MRI

1. INTRODUCTION

Dynamic Magnetic Resonance Imaging (MRI) is a powerful imaging modality to non-invasively image time evolving phenomena in the human body. Some example applications include capturing the motion of the beating heart, motion of vocal tract during speaking, dynamics of contrast uptake in brain or cardiac perfusion MRI. However, a long standing challenge in MRI is its slow imaging speed which restricts its full potential in the achievable spatial or temporal resolution or slice coverage. From a signal processing standpoint, in MRI, one measures the 2D discrete Fourier transform (FT) of the unknown image (to be reconstructed), one FT coefficient (or one row of coefficients) at a time. This is what makes the imaging slow. Accelerated dynamic imaging involves the design of algorithms to accurately reconstruct the image sequence from undersampled sampled k-space (frequency domain) data.

1.1. Existing work and our contribution

MRI literature: LR or sparse model approaches. Since the work on compressive sensing in the early 2000s there has been extensive work on exploiting sparsity of the image or of the sequence in different dictionaries and bases in order to enable accelerated MR imaging in various applications, e.g., see [1, 2] and follow-up work. For settings where joint reconstruction of a set of similar images is needed, a low rank (LR) assumption on the matrix formed by arranging the images as its columns is a much more flexible model since it does not require knowledge of the sparsifying basis or dictionary. MR images change slowly over time and hence are well-modeled as being approximately LR. There has been prior work in the MRI literature on accelerating dynamic or joint MRI using the LR assumption [3, 4, 5, 6, 7]. *In this paper, we refer to this problem as LR column-wise Compressive Sensing (LRcCS).* The LRcCS solutions can be classified into two categories: (a) methods which enforce the LR constraint in an explicit manner (e.g. via explicit estimation of the temporal subspace from low spatial but high temporal resolution training data (e.g.,[4, 5]), or (b) methods that enforce the LR constraint in an implicit manner (e.g., via the nuclear norm or Schatten-p norm regularization with $p < 1$ as in k-t-SLR [6]). The implicit methods like k-t-SLR are useful because they offer more flexibility in handling arbitrary sampling patterns [6]. A challenge with all these methods is the need of tuning the regularization parameters, and other hyper-parameters associated with the iterative optimization algorithm. Moreover, the iterative optimization algorithms used in k-t-SLR and similar methods are known to be very slow when applied for large matrix dimensions, e.g., it takes several minutes for reconstructing typical datasets with 50-100 temporal frames. There is another line of work inspired by the robust PCA literature [8] that models the matrix formed by the MRI sequence as being LR plus sparse (L+S), e.g., see [9] (L+S-Otazo), and follow-up work, e.g., [10, 11]. Similar to k-t-SLR, two challenges with this line of work are (i) the need of tuning the parameters for different MRI applications, and (ii) the long computation times needed for the iterative solvers for the nuclear norm min based (or similar) optimization programs. Besides the above works, there has been extensive work on motion estimation and compensation before imposing the structural assumptions [12, 13, 14, 2, 15, 16]. While these handle motion better, these schemes also share both of the above challenges: each MRI application needs parameter tuning, and these methods are even slower.

Reconstruction algorithm speed is an important concern in applications needing low latency such as real-time interactive MRI, interventional MRI, or biofeedback imaging.

MRI literature: Deep Learning (DL) methods. The ap-

This work was partially supported by NSF grant CIF-2115200.

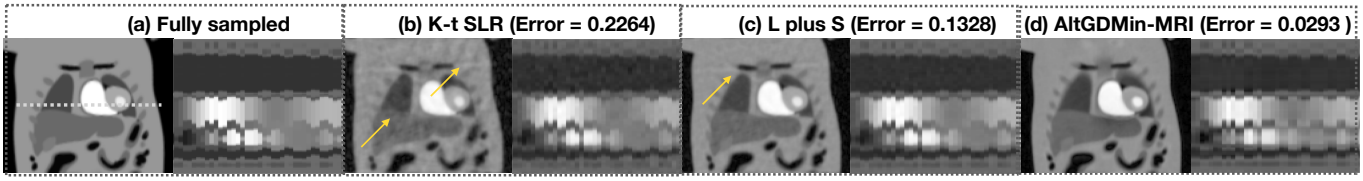

Fig. 1. *Example image recon from various reconstruction methods on the PINCAT perfusion dataset from retrospective undersampled data having 4 radial lines per frame. Each inset shows a spatial frame and an image time series along a horizontal cross section through the heart. The error numbers are also shown. Both k-t-SLR and L+S-Otazo schemes show residual alias artifacts (see yellow arrows). In contrast, altGDmin-MRI depict good image quality with minimal artifacts and blurring.*

proaches described above are drastically different from DL based reconstruction schemes, e.g., [17, 18]. DL models need a large number of fully sampled training data points. Such data are easily available for static imaging applications. However, it is not straightforward to acquire sufficient number of fully sampled image sequences for dynamic imaging applications, and definitely not for high time resolution ones. For this reason, a majority of DL models have been used to perform reconstruction frame by frame; this approach does not fully exploit redundancies along the temporal dimension. On the other hand, our approach, and also all the approaches mentioned above, exploit both spatial and temporal redundancies, do not need any training data, and consequently also do not need the expensive training phase that DL models need. DL model learning/training can be very computationally, and hence energy-wise, expensive.

Provable solutions with only simulation experiments. There are three existing provable solutions to LRcCS. The first is an Alternating Minimization solution that solves the harder magnitude-only generalization of LRcCS (LR Phase Retrieval) [19, 20]. The second (parallel work) studies a convex relaxation called mixed norm min [21]. The third [22] is a gradient descent (GD) based provable solution to LRcCS, that we called altGDmin. The convex solution is very slow, has bad experimental performance even for data simulated using the authors' model, and has a worse sample complexity than altGDmin [22, Table 1]. The AltMin solution is faster than the convex one, but still significantly slower than altGDmin [22]. Also, since it is designed for a harder problem, its sample complexity guarantee for LRcCS is sub-optimal compared to that of altGDmin. All guarantees are for Gaussian measurements but, as in case of (sparse) compressive sensing, the qualitative implications remain true also for MRI (random Fourier measurements).

Contributions. In this work, we explore the utility of the altGDmin approach for fast and highly-undersampled dynamic MR imaging. We develop an improved version of the basic altGDmin algorithm to address the specific issues that occur when trying to recover real dynamic MRI sequences. We show via extensive experiments on four very different dynamic MRI applications that, without any application-specific parameter tuning, altGDmin and altGDmin-MRI outperform both k-t-SLR and L+S-Otazo in most settings, while also being significantly faster than both; see Table 1, Fig. 1.

1.2. Problem setting and notation

LRcCS. We would like to recover an $n \times q$ rank-r matrix $\boldsymbol{X}^* = [\boldsymbol{x}_1^*, \boldsymbol{x}_2^*, \dots, \boldsymbol{x}_q^*]$, with $r \ll \min(q, n)$, from

$$\boldsymbol{y}_k = \boldsymbol{A}_k \boldsymbol{x}_k^*, \; k = 1, 2, \dots, q,$$

when $\boldsymbol{y}_k$ is an m-length vector. In accelerated dynamic MRI, $\boldsymbol{x}_k^*$ is the k-th vectorized image; thus the matrix $\boldsymbol{X}^*$ corresponds to the entire (unknown) image sequence.The $m \times n$ measurement matrices $\boldsymbol{A}_k$ are random Fourier with $m < n$ for accelerated imaging. The exact form of $\boldsymbol{A}_k$ is decided by the specific random sampling trajectory (specified in Sec. 3).

Everywhere, $\|.\|_F$ denotes the Frobenius norm, $\|.\|$ without a subscript denotes the (induced) l_2 norm, and $^\top$ denotes (conjugate) transpose. For two $n \times r$ matrices $\boldsymbol{U}_1, \boldsymbol{U}_2$ with orthonormal columns, we use $\mathrm{SD}(\boldsymbol{U}_1, \boldsymbol{U}_2) := \|(\boldsymbol{I} - \boldsymbol{U}_1 \boldsymbol{U}_1{}^\top)\boldsymbol{U}_2\|_F$ as the Subspace Distance (SD) between the subspaces spanned by their columns. This takes values between 0 and $\sqrt{r}$. Also, we let $\boldsymbol{X}^* \overset{\mathrm{SVD}}{=} \boldsymbol{U}^* \boldsymbol{\Sigma}^* \boldsymbol{B}^*$ denote its reduced (rank r) SVD, and $\kappa := \sigma_{\max}^*/\sigma_{\min}^*$ the condition number of $\boldsymbol{\Sigma}^*$. Notice here the non-standard form of writing the SVD: here $\boldsymbol{U}^*$ and $\boldsymbol{B}^{*\top}$ are tall matrices with orthonormal columns and rows respectively, $\boldsymbol{U}^*$ is $n \times r$ and $\boldsymbol{B}^*$ is $r \times q$. We let $\tilde{\boldsymbol{B}}^* := \boldsymbol{\Sigma}^* \boldsymbol{B}^*$.

Approx-LRcCS for dynamic MRI. MR image sequences are not exactly LR. Also, all images in the sequence have a certain common component that we can denote by $\bar{\boldsymbol{x}}^*$. Thus, the model $\boldsymbol{x}_k^* = \bar{\boldsymbol{x}}^* + \boldsymbol{z}_k^* + \boldsymbol{e}_k^*$, with $\boldsymbol{Z}^* = [\boldsymbol{z}_1^*, \boldsymbol{z}_2^*, \dots, \boldsymbol{z}_q^*]$ being LR, and $\boldsymbol{e}_k^*$ being small residual error in this model, is more practically valid. We explain this point in Sec. 2.2.

2. SOLVING LRCCS AND APPROX-LRCCS

Another way to understand our problem is as follows: each scalar measurement $\boldsymbol{y}_{ki}$ (i-th entry of $\boldsymbol{y}_k$) satisfies $\boldsymbol{y}_{ki} := \langle \boldsymbol{a}_{ki}, \boldsymbol{x}_k^* \rangle$, $i \in [m]$, $k \in [q]$ with $\boldsymbol{a}_{ki}{}^\top$ being the i-th row of $\boldsymbol{A}_k$. Observe that the measurements are not global, i.e., no $\boldsymbol{y}_{ki}$ is a function of the entire matrix $\boldsymbol{X}^*$. The measurements are global for each column but not across the different columns. We thus need to the following incoherence assumption to enable correct interpolation across the different columns [19]. This was introduced in [23] for LR matrix completion (LRMC) which is another LR problem with non-global measurements, but its model is symmetric across rows and columns.

Dataset (# radial lines)	k-t-SLR	L+S-Otazo	altGDmin	altGDmin-MRI
Vocal tract (4)	0.6328 (267.4)	0.4037 (6.2)	0.2270 (**1.2**)	**0.2006** (**5.2**)
Vocal tract (8)	0.4272 (277.7)	0.2388 (5.7)	0.2099 (**0.9**)	**0.1310** (**2.9**)
Vocal tract (16)	0.2423 (271.0)	0.1162 (3.7)	0.1570 (**0.9**)	**0.1023** (**2.1**)
PINCAT (4)	0.2264 (441.1)	0.1328 (3.8)	0.1035 (**1.3**)	**0.0293** (**2.3**)
PINCAT (8)	0.0924 (688.3)	0.0475 (2.6)	0.0467 (**1.4**)	**0.0197** (**2.0**)
PINCAT (16)	0.0477 (393.8)	0.0223 (2.6)	0.0334 (**1.5**)	**0.0145** (**1.8**)
Cardiac (4)	0.38694 (1318.9)	0.30685 (25.6)	0.4274(**5.1**)	**0.23814** (**12.8**)
Cardiac (8)	0.30164 (1185.3)	0.16623 (25.7)	0.3159 (**5.0**)	**0.15602** (**10.6**)
Cardiac (16)	0.23116 (1162.0)	**0.10561** (**26.4**)	0.2220 (**5.4**)	0.13705 (**8.1**)
Brain (4)	0.03887 (216.0)	0.02441 (3.5)	0.0445 (**0.5**)	**0.02214** (**0.6**)
Brain (8)	0.01754 (193.7)	**0.01055** (1.8)	0.0277 (**0.5**)	**0.01139** (**0.7**)
Brain (16)	0.00574 (169.2)	**0.00549** (1.3)	0.0191 (**0.5**)	0.00704 (**0.7**)

Table 1. *Comparing k-t-SLR, L+S-Otazo, altGDmin, altGDmin-MRI for 4 dynamic MRI sequences: Brain ($n = 16384, q = 24$), Cardiac ($n = 31104, q = 80$), PINCAT ($n = 16384, q = 50$), Vocal tract ($n = 10000, q = 59$). Each is recovered from golden-angle sampled measurements with 4, 8, 16 radial lines. c radial lines means $(128/c)$-times acceleration roughly. We report* **Error (Time)** *with Error defined in Sec. 3. In each row, the smallest error and the 2 smallest times are shown in* **bold**.

Assumption 2.1 (Right singular vectors' incoherence) *We assume that $\max_k \|\boldsymbol{b}_k^*\| \le \mu\sqrt{r/q}$. Treating κ as a constant, up to constants, this is equivalent to requiring that $\max_k \|\boldsymbol{x}_k^*\|^2 \le \tilde{\mu} \sum_{k=1}^q \|\boldsymbol{x}_k^*\|^2/q$.*

This assumption assumes that the "energy" (squared 2-norm) of the different columns $\boldsymbol{x}_k^*$ is similar so that so that the maximum energy is within a constant factor of its average value. This is valid for medical image sequences for which large energy changes across the sequence cannot happen.

2.1. The basic altGDmin algorithm (altGDmin-basic)

We would like to design a GD based solution to find the matrix $\boldsymbol{X}$ that minimizes $f(\boldsymbol{X}) := \sum_{k=1}^q \|\boldsymbol{y}_k - \boldsymbol{A}_k \boldsymbol{x}_k^*\|^2$ subject to the constraint that its rank is r or less. There are two commonly used GD approaches in LR recovery literature. The first is to use projected GD on $\boldsymbol{X}$: at each iteration, perform one step of GD w.r.t. $\boldsymbol{X}$, followed by projecting the resulting matrix onto the space of rank r matrices (by SVD) [24]. The second is to write $\boldsymbol{X} = \boldsymbol{U}\boldsymbol{B}$ where $\boldsymbol{U}$ is $n \times r$ and $\boldsymbol{B}$ is $r \times q$ and do alternating GD on $\boldsymbol{U}$ and $\boldsymbol{B}$ for a cost function that contains the data term, $f(\boldsymbol{U}\boldsymbol{B})$, plus a term that helps ensure that norms of $\boldsymbol{U}$ and $\boldsymbol{B}$ remain similar [25, 26]. Because of the specific asymmetric nature of the LRcCS measurement model, we need to be able to show that the column-wise error, $\max_k(\|\hat{\boldsymbol{x}}_k - \boldsymbol{x}_k^*\|/\|\boldsymbol{x}_k^*\|)$, decreases sufficiently with each algorithm iteration. As explained in detail in [22], this is not possible to show for either of the above approaches and hence neither will work for LRcCS. Moreover, even for LRMC where these ideas do work, the first approach is memory-intensive; while the second needs a GD step size that is $1/r$ or smaller [25, 26], making it r-times slower than GD with a constant step size.

The following modification of alternating GD, that we dub *altGDmin*, removes all the above limitations: it is not memory intensive, it works with a constant step size, and it helps guarantee the desired column-wise error bound decay. Consider $f(\boldsymbol{U}\boldsymbol{B}) := \sum_k \|\boldsymbol{y}_k - \boldsymbol{A}_k\boldsymbol{U}\boldsymbol{b}_k\|^2$. We use projected GD for updating $\boldsymbol{U}$ (one GD step w.r.t. $\boldsymbol{U}$ followed by projecting onto the space of orthonormal matrices by QR). For each new estimate of $\boldsymbol{U}$, we solve for $\boldsymbol{B}$ by minimizing over it while keeping $\boldsymbol{U}$ fixed at its current value. Because of the specific asymmetric nature of our model, the minimization is decoupled for the different columns of $\boldsymbol{B}$; consequently, it only involves solving q r-dimensional Least Squares (LS) problems, in addition to also first computing the matrices, $\boldsymbol{A}_k\boldsymbol{U}$, for use in the LS step. Thus the time needed is only $O(qmr^2 + qmnr) = O(mqnr)$. This is equal to the time needed to compute the gradient w.r.t. $\boldsymbol{U}$; and thus, the per-iteration cost of altGDmin is only $O(mqnr)$. We summarize the algorithm in Algorithm 2.1.

Since for $m < n$, our problem is not strongly convex, we need a carefully designed initialization. We obtain an initial estimate of $\boldsymbol{U}$ by computing the top r left singular vectors of the matrix $\boldsymbol{X}_0$ defined in Algorithm 2.1.

2.2. Approx-LRcCS and altGDmin-MRI

MR image sequences are not exactly LR. Also, since MR images are very slow changing, there is a certain "common" component in all the image frames in the sequence. In other words, the average of all the MR images in the sequence has a large magnitude compared to the largest singular value of the matrix formed by subtracting this common component out. Consequently κ for the latter matrix is much lower than if this common component were not subtracted out. Thus the following is a more appropriate model for dynamic MRI image sequences:

$$\boldsymbol{x}_k^* = \bar{\boldsymbol{x}}^* + \boldsymbol{z}_k^* + \boldsymbol{e}_k^*, \text{ for all } k \in [q],$$

where $\bar{\boldsymbol{x}}^*$ is the vectorized average image, the $\boldsymbol{z}_k^*$'s form a rank r matrix $\boldsymbol{Z}^* := [\boldsymbol{z}_1^*, \boldsymbol{z}_2^*, \dots, \boldsymbol{z}_q^*]$, $\boldsymbol{e}_k^*$ is the unstructured residual signal component and we assume that $\|\boldsymbol{e}_k^*\| \ll \|\boldsymbol{z}_k^*\| \ll \|\bar{\boldsymbol{x}}^*\|$.

Algorithm 1 *altGDmin-basic*

Input: $\boldsymbol{y}_k, \boldsymbol{A}_k, k \in [q]$

Parameters: rank r, GD step size η, Number of iterations, T, threshold multiplier $\tilde{C}$ in initialization

Initialization. $\boldsymbol{U}_0 \leftarrow$ top-r-singular-vectors of $\hat{\boldsymbol{X}}_0 := \frac{1}{m}\sum_{ki} \boldsymbol{a}_{ki}\boldsymbol{y}_{ki}\boldsymbol{e}_k{}^\top \mathbb{1}\left\{\boldsymbol{y}_{ki}^2 \le \tilde{C}\frac{\sum_{ki}\boldsymbol{y}_{ki}^2}{mq}\right\}$

for $t = 1$ **to** T **do**

  Let $\boldsymbol{U} \leftarrow \boldsymbol{U}_{t-1}$.

  for $k = 1$ **to** q **do**

    $(\boldsymbol{b}_k)_t \leftarrow (\boldsymbol{A}_k\boldsymbol{U})^\dagger \boldsymbol{y}_k$. Here $\boldsymbol{M}^\dagger := (\boldsymbol{M}^\top\boldsymbol{M})^{-1}\boldsymbol{M}^\top$.

    $(\hat{\boldsymbol{x}}_k)_t \leftarrow \boldsymbol{U}(\boldsymbol{b}_k)_t$

  end for

  Gradient: $\nabla_U f(\boldsymbol{U}\boldsymbol{B}_t) \leftarrow \sum_k \boldsymbol{A}_k^\top(\boldsymbol{A}_k\boldsymbol{U}(\boldsymbol{b}_k)_t - \boldsymbol{y}_k)(\boldsymbol{b}_k)_t^\top$

  GD step: $\hat{\boldsymbol{U}}^+ \leftarrow \boldsymbol{U} - \eta\nabla_U f(\boldsymbol{U}\boldsymbol{B}_t)$

  Projection (using QR): $\hat{\boldsymbol{U}}^+ \overset{\text{QR}}{=} \boldsymbol{U}^+\boldsymbol{R}^+$. Set $\boldsymbol{U}_t \leftarrow \boldsymbol{U}^+$.

end for

Output: $\hat{\boldsymbol{X}}_T := [(\hat{\boldsymbol{x}}_1)_T, (\hat{\boldsymbol{x}}_2)_T, \ldots, (\hat{\boldsymbol{x}}_q)_T]$.

We used ground truth data to verify the validity of the above model. For most of our sequences, the condition number of the first r singular values of $\boldsymbol{X}^*$ is much larger than that for $\boldsymbol{Z}^*$. For example, for the PINCAT sequence, it decreases from 17.1 to 3.4 after subtracting the average image, while for the vocal tract sequence it decreases from 8.8 to 2.5. Lower condition number means a smaller value of m suffices to get an accurate reconstruction (the sample complexity for random Gaussian measurements depends on κ^4).

Under the above assumption, we can estimate $\bar{\boldsymbol{x}}^*$ by solving the following LS problem: $\min_{\bar{\boldsymbol{x}}} \sum_{k=1}^q \|\boldsymbol{y}_k - \boldsymbol{A}_k\bar{\boldsymbol{x}}\|^2$. Denote the solution by $\hat{\bar{\boldsymbol{x}}}$. Next, we estimate $\boldsymbol{Z}^*$ by using the measurement residuals $\tilde{\boldsymbol{y}}_k := \boldsymbol{y}_k - \boldsymbol{A}_k\hat{\bar{\boldsymbol{x}}}$ as the input to altGDmin-basic. Denote its output by $\hat{\boldsymbol{Z}}$. The last step is to estimate the unstructured component $\boldsymbol{e}_k^*$ by using the new measurement residuals $\tilde{\tilde{\boldsymbol{y}}}_k := \boldsymbol{y}_k - \boldsymbol{A}_k\hat{\bar{\boldsymbol{x}}} - \boldsymbol{A}_k\hat{\boldsymbol{z}}_k$ and solving $\min_{\boldsymbol{e}} \|\tilde{\tilde{\boldsymbol{y}}}_k - \boldsymbol{A}_k\boldsymbol{e}\|^2$ for each k, while imposing the assumption that $\|\boldsymbol{e}\|^2$ is small. An indirect way to enforce this while also getting a fast algorithm is to run only a few iterations of GD to solve this minimization problem. The complete algorithm is summarized in Algorithm 2.

Algorithm 2 *altGDmin-MRI*

1. Solve $\min_{\bar{\boldsymbol{x}}} \sum_{k=1}^q \|\boldsymbol{y}_k - \boldsymbol{A}_k\bar{\boldsymbol{x}}\|^2$; denote solution by $\hat{\bar{\boldsymbol{x}}}$.
2. Compute $\tilde{\boldsymbol{y}}_k := \boldsymbol{y}_k - \boldsymbol{A}_k\hat{\bar{\boldsymbol{x}}}$, $k \in [q]$ and use these as inputs for Algorithm 2.1 (altGDmin-basic). Denote its output by $\hat{\boldsymbol{Z}}$.
3. For each $k \in q$, compute $\tilde{\tilde{\boldsymbol{y}}}_k := \boldsymbol{y}_k - \boldsymbol{A}_k\hat{\bar{\boldsymbol{x}}} - \boldsymbol{A}_k\hat{\boldsymbol{z}}_k$ and run 3 iterations of GD to solve $\min_{\boldsymbol{e}} \|\tilde{\tilde{\boldsymbol{y}}}_k - \boldsymbol{A}_k\boldsymbol{e}\|^2$. Denote the output by $\hat{\boldsymbol{e}}_k$.

Output $\hat{\boldsymbol{X}} := [\hat{\boldsymbol{x}}_1, \hat{\boldsymbol{x}}_2, \ldots, \hat{\boldsymbol{x}}_q]$ with $\hat{\boldsymbol{x}}_k = \hat{\bar{\boldsymbol{x}}} + \hat{\boldsymbol{z}}_k + \hat{\boldsymbol{e}}_k$.

2.3. Setting parameters automatically

Since the matrices $\boldsymbol{X}^*$ are only approximating LR, there is no one correct choice of rank r to use. One approach used often is to use the "$b\%$ energy threshold" on singular values of the initialization matrix $\hat{\boldsymbol{X}}_0$ (which is an approximation of $\boldsymbol{X}^*$). In addition, one always needs $\hat{r}$ to be sufficiently smaller than $\min(n, q)$ for the algorithm to take advantage of the LR assumption. A good heuristic is to set $r = \min(\hat{r}_{b\%}, \min(n, q)/10)$. Since the number of frames, q, is pretty small in all our datasets (ranges from 24 to 80), $q/10$ becomes the limiting factor in deciding the rank. Computing $\hat{r}_{b\%}$ is expensive because it requires computation of all the q nonzero singular values. Thus, in all our experiments in this paper, we use $r = \lfloor \min(n, q)/10 \rfloor = \lfloor q/10 \rfloor$. We set the GD step size $\eta = 0.14/\|\nabla_U f(\boldsymbol{U}^0\boldsymbol{B}^0)\|$ where $\boldsymbol{U}^0, \boldsymbol{B}^0$ are the initial estimates and we use $\tilde{C} = 6$ in the initialization step. To decide T (maximum number of iterations), we stop the GD loop when $\text{SD}(\boldsymbol{U}_{t-1}, \boldsymbol{U}_t) < 0.01\sqrt{r}$ while setting $T_{max} = 70$ so that no more than 70 iterations are run. For altGDmin-MRI, we use the CGLS code `https://web.stanford.edu/group/SOL/software/cgls/` (with at most 10 iterations or tolerance of 10^{-36}) to compute $\hat{\bar{\boldsymbol{x}}}$.

For datasets for which both n and q are large, one can set r using the "$b\%$ energy threshold" on the first $\min(n/10, q/10)$ singular values. We are using this approach in ongoing work when dealing with longer sequences.

3. EXPERIMENTS

Experiments were performed by retrospectively under-sampling fully sampled datasets in four dynamic MRI applications: (i) Brain perfusion MRI [27], (ii) Ungated free breathing cardiac perfusion MRI [12], (iii) PINCAT perfusion phantom [6], and (iv) Vocal tract MRI during speech production. The random sampling scheme is borrowed from [28]: it is a pseudo-radial trajectory where the polar coordinates were interpolated onto a Cartesian grid. The angular increment between successive radial spokes were determined by the golden angle (111.25 degrees). All experiments were conducted in MATLAB on the same PC. AltGDmin-MRI was compared with altGDmin-basic, k-t-SLR, L+S-Otazo. For the latter two, we used author provided code and parameters (the authors had tuned the parameters for cardiac imaging). The error value that we report is as follows [6]. Let $\text{dist}^2(\boldsymbol{x}^*, \hat{\boldsymbol{x}}) = \|\boldsymbol{x}^* - \hat{\boldsymbol{x}}\frac{\hat{\boldsymbol{x}}^\top\boldsymbol{x}^*}{\|\boldsymbol{x}^*\|^2}\|^2$ denote the scale invariant distance between two vectorized images with "scale" being a complex number (the reconstructed images can be complex-valued). We report the Monte Carlo (MC) average of $(\sum_{k=1}^q \text{dist}^2(\boldsymbol{x}_k^*, \hat{\boldsymbol{x}}_k))/\|\boldsymbol{X}^*\|_F^2$ over 5 MC iterations. The MC is over different realizations of $\boldsymbol{A}_k$s. In Table 1, *we report results as Error (Time); Time is the time taken by the algorithm in seconds*. As can be seen, in most cases, altGDmin-MRI has lower error than the other methods, and *it is much faster*, than both k-t-SLR and L+S-Otazo. We show a visual comparison in Fig. 1.

4. REFERENCES

[1] M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid mr imaging," *Magnetic Resonance in Medicine*, vol. 58(6), pp. 1182–1195, December 2007.

[2] L. Feng, L. Axel, H. Chandarana, K. T. Block, D. K. Sodickson, and R. Otazo, "Xd-grasp: golden-angle radial mri with reconstruction of extra motion-state dimensions using compressed sensing," *Magnetic resonance in medicine*, vol. 75, no. 2, pp. 775–788, 2016.

[3] A. S. Gupta and Z. Liang, "Dynamic imaging by temporal modeling with principal component analysis," 2001, p. 10.

[4] Z.-P. Liang, "Spatiotemporal imagingwith partially separable functions," in *2007 4th IEEE international symposium on biomedical imaging: from nano to macro*. IEEE, 2007, pp. 988–991.

[5] H. Pedersen, S. Kozerke, S. Ringgaard, K. Nehrke, and W. Y. Kim, "k-t pca: Temporally constrained k-t blast reconstruction using principal component analysis," *Magnetic resonance in medicine*, vol. 62, no. 3, pp. 706–716, 2009.

[6] S. G. Lingala, Y. Hu, E. DiBella, and M. Jacob, "Accelerated dynamic mri exploiting sparsity and low-rank structure: kt slr," *Medical Imaging, IEEE Transactions on*, vol. 30, no. 5, pp. 1042–1054, 2011.

[7] B. Zhao, J. P. Haldar, A. G. Christodoulou, and Z.-P. Liang, "Image reconstruction from highly undersampled-space data with joint partial separability and sparsity constraints," *Medical Imaging, IEEE Transactions on*, vol. 31, no. 9, pp. 1809–1820, 2012.

[8] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, 2011.

[9] R. Otazo, E. Candes, and D. K. Sodickson, "Low-rank plus sparse matrix decomposition for accelerated dynamic mri with separation of background and dynamic components," *Magnetic resonance in medicine*, vol. 73, no. 3, pp. 1125–1136, 2015.

[10] C. Y. Lin and J. A. Fessler, "Accelerated methods for low-rank plus sparse image reconstruction," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. IEEE, 2018, pp. 48–51.

[11] Y. Liu, T. Liu, J. Liu, and C. Zhu, "Smooth robust tensor principal component analysis for compressed sensing of dynamic mri," *Pattern Recognition*, vol. 102, p. 107252, 2020.

[12] S. G. Lingala, E. DiBella, and M. Jacob, "Deformation corrected compressed sensing (dc-cs): a novel framework for accelerated dynamic mri," *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 72–85, 2014.

[13] X. Chen, M. Salerno, Y. Yang, and F. H. Epstein, "Motion-compensated compressed sensing for dynamic contrast-enhanced mri using regional spatiotemporal sparsity and region tracking: Block low-rank sparsity with motion-guidance (blosm)," *Magnetic resonance in medicine*, vol. 72, no. 4, pp. 1028–1038, 2014.

[14] A. Tolouee, J. Alirezaie, and P. Babyn, "Nonrigid motion compensation in compressed sensing reconstruction of cardiac cine mri," *Magnetic resonance imaging*, vol. 46, pp. 114–120, 2018.

[15] A. G. Christodoulou, J. L. Shaw, C. Nguyen, Q. Yang, Y. Xie, N. Wang, and D. Li, "Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging," *Nature biomedical engineering*, vol. 2, no. 4, pp. 215–226, 2018.

[16] J. Y. Cheng, T. Zhang, M. T. Alley, M. Uecker, M. Lustig, J. M. Pauly, and S. S. Vasanawala, "Comprehensive multi-dimensional mri for the simultaneous assessment of cardiopulmonary anatomy and physiology," *Scientific reports*, vol. 7, no. 1, pp. 1–15, 2017.

[17] Z. Ke, W. Huang, Z.-X. Cui, J. Cheng, S. Jia, H. Wang, X. Liu, H. Zheng, L. Ying, Y. Zhu, and D. Liang, "Learned low-rank priors in dynamic mr imaging," *IEEE Transactions on Medical Imaging*, vol. 40, no. 12, pp. 3698–3710, 2021.

[18] A. H. Ahmed, H. Aggarwal, P. Nagpal, and M. Jacob, "Dynamic mri using deep manifold self-learning," in *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 1052–1055.

[19] S. Nayer, P. Narayanamurthy, and N. Vaswani, "Provable low rank phase retrieval," *IEEE Trans. Info. Th.*, March 2020.

[20] S. Nayer and N. Vaswani, "Sample-efficient low rank phase retrieval," *IEEE Trans. Info. Th.*, 2021.

[21] R. S. Srinivasa, K. Lee, M. Junge, and J. Romberg, "Decentralized sketching of low rank matrices," in *Neur. Info. Proc. Sys. (NeurIPS)*, 2019, pp. 10 101–10 110.

[22] S. Nayer and N. Vaswani, "Fast low rank column-wise compressive sensing," *arXiv:2102.10217*.

[23] E. J. Candes and B. Recht, "Exact matrix completion via convex optimization," *Found. of Comput. Math*, no. 9, pp. 717–772, 2008.

[24] Y. Cherapanamjeri, K. Gupta, and P. Jain, "Nearly-optimal robust matrix completion," *ICML*, 2016.

[25] X. Yi, D. Park, Y. Chen, and C. Caramanis, "Fast algorithms for robust pca via gradient descent," in *Neur. Info. Proc. Sys. (NeurIPS)*, 2016.

[26] Q. Zheng and J. Lafferty, "Convergence analysis for rectangular matrix completion using burer-monteiro factorization and gradient descent," *arXiv preprint arXiv:1605.07051*, 2016.

[27] S. G. Lingala and M. Jacob, "Blind compressive sensing dynamic mri," *Medical Imaging, IEEE Transactions on*, vol. 32, no. 6, pp. 1132–1145, 2013.

[28] S. Winkelmann, T. Schaeffter, T. Koehler, H. Eggers, and O. Doessel, "An optimal radial profile order based on the golden ratio for time-resolved mri," *IEEE transactions on medical imaging*, vol. 26, no. 1, pp. 68–76, 2006.