CONVOLUTIONAL APPROXIMATIONS TO LINEAR DIMENSIONALITY REDUCTION OPERATORS

Swayambhoo Jain and Jarvis Haupt

Department of Electrical and Computer Engineering
University of Minnesota – Twin Cities, Minneapolis, MN - 55455
{jainx174,jdhaupt}@umn.edu

ABSTRACT

This paper examines the existence of efficiently implementable approximations of a general real linear dimensionality reduction (LDR) operator. The specific focus is on approximating a given LDR operator with a partial circulant structured matrix (a matrix whose rows are related by circular shifts) as these constructions allow for low-memory footprint and computationally efficient implementations. Our main contributions are theoretical: we quantify how well general matrices may be approximated (in a Frobenius sense) by partial circulant structured matrices, and also consider a variation of this problem where the aim is only to accurately approximate the action of a given LDR operator on a restricted set of inputs. For the latter setting, we also propose a sparsity-regularized alternating minimization based algorithm for learning partial circulant approximations from data, and provide experimental evidence demonstrating the potential efficacy of this approach on real-world data.

Index Terms— Circulant matrices, subspace learning, matrix factorization, sparse regularization, big data

1. INTRODUCTION

Linear dimensionality reduction (LDR) is at the core of many applications in signal processing, statistics, and machine learning. LDR methods yield low-dimensional linear mappings which aid in visualizing, efficient processing and extracting features from noisy high-dimensional data. Principal component analysis, perhaps one of the most classical and widely used techniques for high-dimensional data analysis, is essentially an LDR technique designed to preserve the variance of the original data in the lower-dimensional space. It has been widely used for data compression, de-noising, and as a dimensionality reducing pre-processing step for other analyses (e.g., clustering, classification, etc.) [1]. Similarly, the design of LDR operators for other specific purposes has resulted in techniques such as linear discriminant analysis, canonical correlations analysis, and many more. See the survey paper [2] for many additional examples.

Recently random LDR operators have emerged as a useful tool for universal "pre-compression" in high-dimensional inference tasks. For example, fully random LDR operators are at the core of the initial investigations into *compressed sensing* (CS) (see, e.g., [3]); other, more structured, LDR operators – both non-adaptive (see, e.g., [4–10]) and adaptive (see, e.g., [11–20]) – have also been examined recently in the context of CS and sparse inference.

This work was supported in part by DARPA/ONR Grant No. N66001-11-1-4090 and the DARPA Young Faculty Award, Grant No. N66001-14-1-4047.

With the ever increasing dimensionality and volume of data, the implementation of LDR operators is itself becoming resource intensive. A (here, real) LDR operator can be represented as a matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ (with $m < n$) and it generally requires $\mathcal{O}(mn)$ operations for applying it on an arbitrary vector $\mathbf{x} \in \mathbb{R}^n$. The number of operations can be superlinear in n even for modest values of m (e.g., when $m = n^\beta$ for $\beta \in (0, 1]$ the complexity is $\mathcal{O}(n^{1+\beta})$). Given these considerations, structured LDR operators are particularly attractive as they offer efficient implementations. The specific focus of this paper is on employing *partial circulant approximations* to general LDR operators. There are two main motivations for considering such approximations: (1) from a system perspective, partial circulant matrices arise naturally when describing (sampled outputs of) certain linear time invariant (LTI) systems, and (2) from a computational perspective, partial circulant matrices admit fast memory efficient implementations via the fast Fourier transform (FFT) and sub-sampling operations. In this paper, we try to answer the fundamental question – *How well can a partial circulant structured matrix approximate a given LDR matrix?* We investigate the quality of approximation for several partial circulant approximations, and provide some initial experimental results to demonstrate potential efficacy of these methods for large-scale data processing tasks.

1.1. Connections to Existing Work

Circulant approximations to square matrices are classical in linear algebra; for example, circulant preconditioners for linear systems were examined in [21–23]. In works motivated by "optical information processing," several efforts have examined fundamental aspects of approximating square matrices by products of circulant and diagonal matrices [24–28]. Here, our focus is on LDR matrices (not square matrices), so results from these works are not directly applicable here.

Random partial circulant matrices have also been studied recently in compressed sensing tasks [29–33], and random partial circulant matrices with diagonal pre-processors have been proposed as computationally efficient methods for Johnson-Lindenstrauss (JL) embeddings in [34–37]. The work [38] established the viability of using random partial circulant matrices for embedding manifold-structured data. In contrast to these works, here our aim is to approximate the action of a given LDR matrix, not necessarily to perform JL embeddings.

1.2. Our Contributions

Our first main result pertains to approximating $\mathbf{A}$ by a partial circulant matrix $\tilde{\mathbf{A}} = \mathbf{SC}$, obtained by choosing m distinct rows from an $n \times n$ circulant matrix $\mathbf{C}$ via a row-subsampling matrix $\mathbf{S}$ that

is a permuted subset of m rows of $\mathbf{I}_n$ (the $n \times n$ identity matrix). Such matrices enjoy a low computational complexity of $\mathcal{O}(n \log n)$ as well as a concise $\mathcal{O}(m+n)$ storage cost. Despite these potential benefits, our first result here is negative – we establish that "most" LDR matrices are not well-approximated (in a Frobenius sense) by partial circulant matrices.

We then propose a generalization that uses approximations of the form $\tilde{\mathbf{A}} = \mathbf{PSC}$, where $\mathbf{C}$ and $\mathbf{S}$ are as above, except that $\mathbf{S}$ has some $m' \geq m$ rows, and $\mathbf{P}$ is an $m \times m'$ "post-processing" matrix. Operating with this type of matrix approximation requires $\mathcal{O}(mm' + n \log n)$ operations in general and an $\mathcal{O}(mm' + m' + n)$ storage cost, both of which can be as low as $\mathcal{O}(n \log n)$, e.g., when $m' = \mathcal{O}(n^{1/2})$. We present a new result that connects the approximation error to the so-called *Gaussian width* of the set of "post-processed" partial circulant matrices of the form above.

Next, we examine a modified paradigm wherein we exploit the fact that signals of interest often reside on a *restricted* input domain (e.g., a union of subspaces, manifold, etc.). In such settings, we may restrict our approximation to mimicking the action of $\mathbf{A}$ on these inputs. We provide a concise argument establishing the efficacy of this more general approach for certain restricted inputs, describe a *data-driven* approach to learning the factors of the approximating matrix, and provide empirical evidence to demonstrate the viability of this approach.

All results are stated here without proof; detailed proofs appear in a working manuscript in preparation [39].

1.3. Outline

After introducing few preliminaries, in Section 2 we present a fundamental approximation result regarding partial circulant approximations to general LDR matrices. In Section 3 we analyze the partial circulant approximations with "post-processing". In Section 4 we present a generalized "data-driven" approach. Numerical experiments on a real-world dataset are discussed in Section 5. The conclusion and future directions of research are provided in Section 6.

1.4. Preliminaries

We introduce some preliminary concepts and notation that will be used throughout the paper. First, we define

$$\mathbf{R} = \begin{bmatrix} \mathbf{0}_{(n-1)\times 1} & \mathbf{I}_{n-1} \\ 1 & \mathbf{0}_{1\times(n-1)} \end{bmatrix},$$

to be the "right rotation" matrix, in that post-multiplication of a row vector by $\mathbf{R}$ implements a circular shift to the right by one position. Analogously, post-multiplying a row vector by $\mathbf{L} = \mathbf{R}^T$ implements a circular shift to the left by one position; note that $\mathbf{LR} = \mathbf{I}_n$. We represent an $n \times n$ (real) circulant matrix by

$$\mathbf{C} = \begin{bmatrix} c_1 & c_2 & \cdots & c_n \\ c_n & c_1 & \cdots & c_{n-1} \\ & & \ddots & \\ c_2 & c_3 & \cdots & c_1 \end{bmatrix} = \begin{bmatrix} \mathbf{c}^T \\ \mathbf{c}^T\mathbf{R} \\ \vdots \\ \mathbf{c}^T\mathbf{R}^{n-1} \end{bmatrix}, \tag{1}$$

where $\mathbf{c} = [c_1 \ \cdots \ c_n]^T \in \mathbb{R}^n$. We let $\mathcal{C}_n$ denote the set of all (real) n-dimensional circulant matrices. The set $\mathcal{S}_m$ denotes the set of $m \times n$ row sub-sampling matrices whose rows comprise m different rows of $\mathbf{I}_n$. For $m < n$, the set of all $m \times n$ real *partial* circulant matrices is

$$\mathcal{PC}_{m,n} = \left\{ \mathbf{SC} \in \mathbb{R}^{m\times n} \mid \mathbf{S} \in \mathcal{S}_m, \mathbf{C} \in \mathcal{C}_n \right\}. \tag{2}$$

The set of partial circulant matrices with m'-rows and a post-processing matrix defined as follows

$$\mathcal{PC}_{m,m',n} = \left\{ \mathbf{PSC} \;\middle|\; \mathbf{P} \in \mathbb{R}^{m\times m'}, \mathbf{S} \in \mathcal{S}_{m'}, \mathbf{C} \in \mathcal{C}_n \right\}. \tag{3}$$

For $\mathbf{A} \in \mathbb{R}^{m\times n}$, we denote its m individual rows by $\mathbf{A}_{i,:}$ for $i = 1, \ldots, m$ and its n columns by $\mathbf{A}_{:,j} \in \mathbb{R}^m$ for $j = 1, \ldots, n$. The squared Frobenius norm of $\mathbf{A}$ is $\|\mathbf{A}\|_F^2 = \sum_{i,j} |A_{i,j}|^2$, and $\|\mathbf{A}\|_{2,1} = \sum_{j=1}^n \|\mathbf{A}_{:,j}\|_2$, where $\|\mathbf{A}_{:,j}\|_2$ is the Euclidean norm of $\mathbf{A}_{:,j}$.

2. APPROXIMATION WITH PARTIAL CIRCULANT MATRICES

We first consider approximating the given LDR matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ with a partial circulant $\mathbf{Z} \in \mathcal{PC}_{m,n}$ by minimizing the Frobenius norm distance of $\mathbf{A}$ from $\mathcal{PC}_{m,n}$. In this setting, the minimum approximation error is

$$\mathcal{E}_{\mathcal{PC}_{m,n}}(\mathbf{A}) = \min_{\mathbf{Z}\in\mathcal{PC}_{m,n}} \|\mathbf{A} - \mathbf{Z}\|_F^2. \tag{4}$$

Evaluating (4) is a non-convex optimization problem. In the following lemma, we provide a precise characterization of the minimum achievable approximation error for a given $\mathbf{A}$.

Lemma 2.1. *For $\mathbf{A} \in \mathbb{R}^{m\times n}$, we have*

$$\mathcal{E}_{\mathcal{PC}_{m,n}}(\mathbf{A}) = \|\mathbf{A}\|_F^2 - R^2(\mathbf{A}), \tag{5}$$

where $R(\mathbf{A})$ is defined as

$$R(\mathbf{A}) = \max_{\mathbf{f}\in\mathcal{F}} \frac{\|\sum_{i=1}^m \mathbf{A}_{i,:}\mathbf{L}^{f_i}\|_2}{\sqrt{m}}, \tag{6}$$

and $\mathcal{F}$ is the set of m-dimensional vectors with distinct components taking integer values in $[0, n-1]$ defined as

$$\mathcal{F} = \left\{ [f_1, \cdots, f_m] \in \{0, ..., n-1\}^m \middle| f_i \neq f_j \ \forall i \neq j \right\}.$$

We call the term $R(\mathbf{A})$ in (6) the *Rubik's Score* of the matrix $\mathbf{A}$, inspired by the fact that $R(\mathbf{A})$ is maximized when the circular shifts of the rows $\{\mathbf{A}_{i,:}\}$ are "maximally aligned". A large value of the *Rubik's Score* implies better approximation. In can easily be shown that $0 \leq R(\mathbf{A}) \leq \|\mathbf{A}\|_F$, with $R(\mathbf{A}) = \|\mathbf{A}\|_F$ for all $\mathbf{A} \in \mathcal{PC}_{m,n}$. Calculation of $R(\mathbf{A})$ is difficult as it involves maximization over the discrete set $\mathcal{F}$. We can gain additional insight using a probabilistic technique – instead of quantifying the approximation error for a fixed $\mathbf{A}$, we consider drawing matrices $\mathbf{A}$ randomly, so that their unit-normed row-wise vectorized representations are uniformly distributed on the unit sphere in $\mathbb{R}^{mn}$. (These matrices can be generated as matrices with iid zero-mean Gaussian elements.) We then quantify the *proportion* of matrices whose (optimal) partial circulant approximation error is at most a fixed fraction of their squared Frobenius norm. With this, we establish the following theorem

Theorem 2.1. *For $2 \leq m \leq n$, let $\mathbf{A} \in \mathbb{R}^{m\times n}$ have iid $\mathcal{N}(0,1)$ entries. Then for $\delta \in [0, 0.125)$, and n is sufficiently large, there exists a positive constant $c(\delta)$ such that*

$$\Pr(\mathcal{E}_{\mathcal{PC}_{m,n}}(\mathbf{A}) \leq \delta \|\mathbf{A}\|_F^2) = \mathcal{O}(e^{-c(\delta)\cdot mn}).$$

Theorem 2.1 implies that the proportion of large matrices that can be approximated to high accuracy by partial circulant matrices decays exponentially with the product of matrix dimensions mn. Stated another way, most large matrices are not well-approximated by partial circulant matrices.

3. APPROXIMATION USING PARTIAL CIRCULANT MATRICES WITH POST-PROCESSING

We now consider a more general framework in which we approximate $\mathbf{A}$ by a partial circulant matrix with post-processing in the set $\mathcal{PC}_{m,m',n}$. In this setting, the minimum approximation error is

$$\mathcal{E}_{\mathcal{PC}_{m,m',n}}(\mathbf{A}) = \min_{\mathbf{Z}\in\mathcal{PC}_{m,m',n}} \|\mathbf{A}-\mathbf{Z}\|_F^2. \tag{7}$$

Again evaluating $\mathcal{E}_{\mathcal{PC}_{m,m',n}}(\mathbf{A})$ is a non-convex problem. However, quantifying the improvements offered by this expanded approximation model is possible when we make use of the following result.

Lemma 3.1. *For $\mathbf{A}\in\mathbb{R}^{m\times n}$, we have*

$$\mathcal{E}_{\mathcal{PC}_{m,m',n}}(\mathbf{A}) = \|\mathbf{A}\|_F^2 - \left[\max_{\mathbf{Z}\in\widetilde{\mathcal{PC}}_{m,m',n}} Tr\left(\mathbf{A}^T\mathbf{Z}\right)\right]^2,$$

where $\widetilde{\mathcal{PC}}_{m,m',n} = \left\{\frac{\mathbf{Z}}{\|\mathbf{Z}\|_F}\middle|\mathbf{Z}\neq\mathbf{0},\mathbf{Z}\in\mathcal{PC}_{m,m',n}\right\}$.

The quantity $\max_{\tilde{\mathbf{Z}}\in\widetilde{\mathcal{PC}}_{m,m',n}} \mathrm{Tr}\left(\mathbf{A}^T\tilde{\mathbf{Z}}\right)$ in Lemma 3.1 is analogous to the Rubik's score in the previous section. For a given $\mathbf{A}$, it determines quality of approximation with the partial circulant matrix with post-processing. Also, since $\mathcal{PC}_{m,n}\subset\widetilde{\mathcal{PC}}_{m,m',n}$ it is easy to see that the Rubik's score satisfies $R(\mathbf{A})\leq\max_{\tilde{\mathbf{Z}}\in\widetilde{\mathcal{PC}}_{m,m',n}}\mathrm{Tr}\left(\mathbf{A}^T\tilde{\mathbf{Z}}\right)$. This implies that the approximations with partial circulant matrices that employ post-processing are guaranteed to be no worse than those without post-processing.

We codify the approximation error for general matrices via the following result.

Theorem 3.1. *Assuming that $\mathbf{A}\in\mathbb{R}^{m\times n}$ have iid $\mathcal{N}(0,1)$ entries, we have*

$$\mathbb{E}_{\mathbf{A}}\left(\mathcal{E}_{\mathcal{PC}_{m,m',n}}(\mathbf{A})\right)\leq mn - \left[\omega(\widetilde{\mathcal{PC}}_{m,m',n})\right]^2, \tag{8}$$

where $\omega(\tilde{\mathcal{PC}}_{m,m',n})$ is the Gaussian width of $\widetilde{\mathcal{PC}}_{m,m',n}$, defined as

$$\omega(\widetilde{\mathcal{PC}}_{m,m',n}) = \mathbb{E}_{\mathbf{A}}\left(\max_{\mathbf{Z}\in\widetilde{\mathcal{PC}}_{m,m',n}}\mathrm{tr}\left(\mathbf{A}^{\mathrm{T}}\mathbf{Z}\right)\right).$$

Theorem 3.1 provides a fundamental insight that the average approximation quality of partial circulant matrices with post-processing matrix is related to a notion of the size of the set $\widetilde{\mathcal{PC}}_{m,m',n}$, quantified via the Gaussian width. Therefore, understanding the dependence between the Gaussian width and m' is crucial. The following lemma sheds some light on this dependence.

Lemma 3.2. *The Gaussian width of $\widetilde{\mathcal{PC}}_{m,m',n}$ is bounded as*

$$\omega(\widetilde{\mathcal{PC}}_{m,m',n})\geq\frac{mm'}{\sqrt{1+mm'}}.$$

Using the lower bound on the Gaussian width in Lemma 3.2 along with the result in Theorem 3.1 we arrive at the following upper bound on the expected error

$$\mathbb{E}\left(\mathcal{E}_{\mathcal{PC}_{m,m',n}}(\mathbf{A})\right)\leq mn\left(1-\frac{m'}{n}\frac{mm'}{1+mm'}\right) \tag{9}$$

The above inequality gives insight in how the expected approximation error decreases with increasing m': since for modest values of m,m' the term $\frac{mm'}{1+mm'}\approx 1$, we see that the (average) approximation error decreases roughly linearly in m'.

4. THE DATA DRIVEN APPROACH

The underlying theme of the approximation techniques considered so far is that they approximate the given LDR matrix globally. However, in many practical applications where LDR methods are employed (e.g. PCA, LDA, Compressive Sensing etc.) the data to be processed is not arbitrary, but lie in some *restricted input domain* (e.g., in a low-dimensional subspace, a union of low-dimensional subspaces, distinct clusters, etc.). In these settings, the approximation can be much better, provided that the inner dimension m' exceed a measure of complexity of the restricted input domain. This is made concrete here by the following existence result.

Theorem 4.1. *Let $\mathbf{A}\in\mathbb{R}^{m\times n}$ be any fixed matrix, and let $\mathcal{X}$ be any finite set of n-dimensional unit-norm vectors. For any $\epsilon\in(0,1)$, there exists a post-processing $\mathbf{P}\in\mathbb{R}^{m\times m'}$, sampling matrix $\mathbf{S}\in\mathbb{R}^{m'\times n}$ comprised of rows of identity matrix $\mathbf{I}_n$, and circulant matrix $\mathbf{C}\in\mathbb{C}^{n\times n}$ for which*

$$\sup_{\mathbf{x}\in\mathcal{X}}\|\mathbf{A}\mathbf{x}-\mathbf{PSC}\mathbf{x}\|_2\leq\epsilon\|\mathbf{A}\|_F,$$

provided that $m'>c_1\epsilon^{-2}\log(c_2m|\mathcal{X}|)\log^4(n)$ where c_1 and c_2 are universal positive constants.

We note that above result is for finite sized sets but extensions to general sets are also possible. This line of investigation leads naturally to a *data-driven* approach in which we are given a data matrix $\mathbf{X}\in\mathbb{R}^{n\times p}$ whose columns are "representative" of the restricted domain for the problem of interest and we want to control the worst-case approximation error of the action of $\mathbf{A}$ on $\mathbf{X}$. Here we investigate a related (more tractable) variant of this approach, where we assume we are given a "representative" data matrix $\mathbf{X}\in\mathbb{R}^{n\times p}$ and aim to minimize the aggregate error $\|\mathbf{AX}-\mathbf{PSCX}\|_F^2$. This approach ensures the approximation be accurate in an *average* sense.

Naively, this minimization would need to be solved for various values of m'. Here we propose an alternative approach: since for a fixed value of m' the resulting matrix $\mathbf{PS}$ obtained by this approach has exactly m' non-zero columns, we propose to effectively combine the actions of the sampling and post processing matrices. We let $\mathbf{M}\triangleq\mathbf{PS}$, and seek *column sparsity* in $\mathbf{M}$ using the $\|\mathbf{M}\|_{2,1}$ regularization term; this results in an optimization of the form

$$\min_{\mathbf{M}\in\mathbb{R}^{m\times n},\mathbf{C}\in\mathcal{C}_n}\|\mathbf{AX}-\mathbf{MCX}\|_F^2+\lambda\|\mathbf{M}\|_{2,1}+\mu\|\mathbf{C}\|_F^2,$$

where $\lambda>0,\mu>0$ are the regularization parameters. In this formulation the non-zero columns m' vary with λ. The regularization term $\|\mathbf{C}\|_F^2$ is needed to fix scaling ambiguity introduced due to the matrix product term $\mathbf{MC}$.

This problem is jointly non-convex in $\mathbf{M}$ and $\mathbf{C}$ due to the matrix multiplication term $\mathbf{MC}$, so we propose an alternating minimization based approach shown in Algorithm 1. Note that both $\mathbf{C}$ as well as $\mathbf{M}$ update steps are convex problems, and can be solved using existing software (e.g., SLEP [40]). For large scale datasets computationally faster alternatives leveraging the circulant structure are also possible. Due to space limits we do not discuss these here.

5. EXPERIMENTAL EVALUATION

We evaluate these approaches using the processed COIL-20 image database [41]. This database contains 128×128 images which were vectorized to obtain a data matrix $\mathbf{X}$ whose columns represent 1440 vectorized images from the dataset. We took rows of $\mathbf{A}$ as the

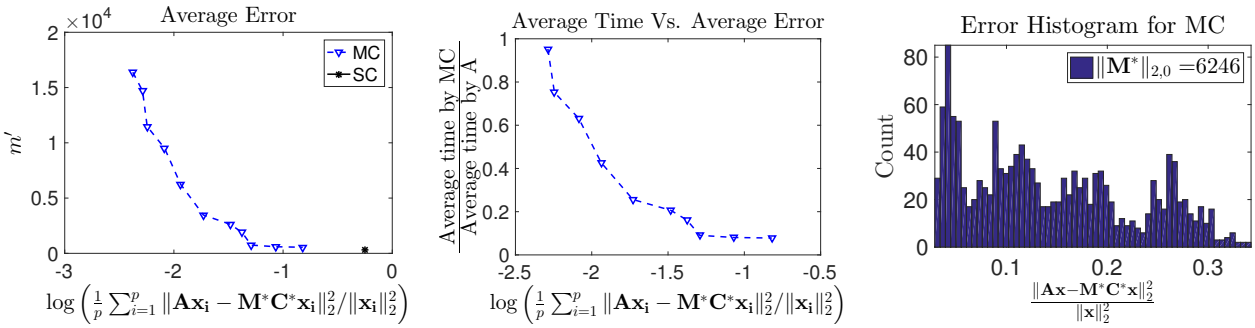

Fig. 1: Results for data-driven approximation of LDR matrix obtained by top 300 principal components of training data obtained by resized images from COIL-20 database. The first panel (left to right) contains log average relative approximation error vs. m'. The second panel shows the ratio of average time taken by the matrix approximation and $\mathbf{A}$ vs. the log average relative approximation error. The third panel shows the histogram of errors with $m' = 6246$.

Algorithm 1 The "Data-Driven" approach algorithm.

Inputs: LDR matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, parameters $\lambda, \mu, \epsilon > 0$,
Matrix of "representative" data $\mathbf{X} \in \mathbb{R}^{n\times p}$,
Initialize: $\mathbf{M}^{(0)} = \mathbf{U}\Sigma$ (from the SVD $\mathbf{AX} = \mathbf{U}\Sigma\mathbf{V}^T$)
$\text{obj}^{(0)} = \|\mathbf{AX}\|_F^2$
repeat
$\mathbf{C}^{(t)} = \arg\min_{\mathbf{C}\in\mathcal{C}_n} \|\mathbf{AX} - \mathbf{M}^{(t-1)}\mathbf{CX}\|_F^2 + \mu\|\mathbf{C}\|_F^2$
$\mathbf{M}^{(t)} = \arg\min_{\mathbf{M}\in\mathbb{R}^{m\times n}} \|\mathbf{AX} - \mathbf{MC}^{(t)}\mathbf{X}\|_F^2 + \lambda\|\mathbf{M}\|_{2,1}$
$\text{obj}^{(t)} = \|\mathbf{AX} - \mathbf{M}^{(t)}\mathbf{C}^{(t)}\mathbf{X}\|_F^2 + \mu\|\mathbf{C}^{(t)}\|_F^2 + \lambda\|\mathbf{M}^{(t)}\|_{2,1}$
until $\text{obj}^{(t)} - \text{obj}^{(t-1)} \leq \epsilon \cdot \text{obj}^{(t-1)}$
Output: $\mathbf{M}^* = \mathbf{M}^{(t)}, \mathbf{C}^* = \mathbf{C}^{(t)}$

top 300 principal component vectors of the training data. In Algorithm 1, we use $\mu = 0.1$ and vary λ to obtain $\mathbf{M}^*$ and $\mathbf{C}^*$ for each value of λ, and quantified the normalized error on the training set vs. column sparsity of $\mathbf{M}^*$. We also plot the results for the circulant approximation to the matrix $\mathbf{A}$ without the post-processing.

The first panel (left to right) in Figure 1 plots m' vs. log average relative error for "data-driven" approximations with and without the post-processing matrix. We can see from the plot that approximating with the post-processing matrix (shown in blue dotted line with triangular marker) incurs far less error as compared to approximation without the post-processing matrix (shown in black star marker). This plot demonstrates the superiority of approximations with post-processing relative to those without post-processing.

The second panel (left to right) in Figure 1 gives insight into the relative time taken by matrix approximations shown as blue markers in the plot in first panel as compared to the original LDR matrix $\mathbf{A}$; it shows the ratio of average time taken by $\mathbf{A}$ and by its "data-driven" approximation with post-processing matrix obtained from Algorithm 1 versus the log average relative error. A C programming language based implementation was used for matrix-vector multiplication by $\mathbf{A}$ as well as for its circulant structured approximation. The average time was obtained by averaging over 10 trials. We can see that the multiplication with approximations obtained by Algorithm 1 is faster than the given LDR matrix $\mathbf{A}$. The speed is due to the FFT based implementation of matrix vector multiplication. We can see that as m' increases the speed of matrix vector multiplication decreases but the log average relative error also decreases. For a representative λ corresponding to the sparsity of $m' = 6246$, we computed a histogram of the normalized approximation errors $\|\mathbf{Ax} - \mathbf{MCx}\|_2^2/\|\mathbf{Ax}\|_2^2$ for each point in the training dataset. The histogram is plotted in third panel of the Figure 1. Most of the relative errors are relatively small, demonstrating that our approach provides fairly accurate approximation.

6. CONCLUSION AND FUTURE WORK

We investigated the problem of approximating an arbitrary LDR matrix via various partial circulant structured matrices, presented several fundamental results, and evaluated numerically a "data-driven" partial circulant approximation approach. Future directions of research include extension of the basic analytical framework developed here to other structured matrix approximations with low implementation complexities (e.g., sparse matrices and fast JL embeddings [42, 43]).

7. REFERENCES

[1] I. Jolliffe, *Principal component analysis*, Wiley Online Library, 2002.

[2] J. P. Cunningham and Z. Ghahramani, "Unifying linear dimensionality reduction," *arXiv preprint arXiv:1406.0873*, 2014.

[3] Y. C. Eldar and G. Kutyniok, *Compressed sensing: Theory and applications*, Cambridge University Press, 2012.

[4] M. Elad, "Optimized projections for compressed sensing," *IEEE Trans. Signal Processing*, vol. 55, no. 12, pp. 5695–5702, 2007.

[5] M. Rudelson and R. Vershynin, "On sparse reconstruction from Fourier and Gaussian measurements," *Communications on Pure and Applied Mathematics*, vol. 61, no. 8, pp. 1025–1045, 2008.

[6] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE Journal of Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 358–374, 2010.

[7] J. Haupt, L. Applebaum, and R. Nowak, "On the restricted isometry of deterministically subsampled Fourier matrices," in *IEEE Conf. on Info. Sciences and Systems*, 2010, pp. 1–6.

[8] A. Ashok and M. A. Neifeld, "Compressive imaging: hybrid measurement basis design," *J. Optical Society of America A*, vol. 28, no. 6, pp. 1041–1050, 2011.

[9] W. R. Carson, M. Chen, M. R. D. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM Journal on Imaging Sciences*, vol. 5, no. 4, pp. 1185–1212, 2012.

[10] S. Jain, A. Soni, and J. Haupt, "Compressive measurement designs for estimating structured signals in structured clutter: A Bayesian experimental design approach," in *Asilomar Conf. on Signals, Systems and Computers*. IEEE, 2013, pp. 163–167.

[11] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. on Signal Processing*, vol. 56, no. 6, pp. 2346–2356, 2008.

[12] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Compressive distilled sensing: Sparse recovery using adaptivity in compressive measurements," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, 2009, pp. 1551–1555.

[13] J. Haupt, R. Castro, and R. Nowak, "Adaptive sensing for sparse signal recovery," in *Proc. IEEE DSP Workshop and Workshop on Sig. Processing Education*, 2009, pp. 702–707.

[14] P. Indyk, E. Price, and D. P. Woodruff, "On the power of adaptivity in sparse recovery," in *Proc. IEEE Foundations of Computer Science*, 2011, pp. 285–294.

[15] M. Iwen and A. Tewfik, "Adaptive group testing strategies for target detection and localization in noisy environments," *IEEE Trans. on Sig. Processing*, vol. 60, no. 5, pp. 2344–2353, 2012.

[16] J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in *Proc. IEEE Statistical Sig. Processing Workshop*, 2012, pp. 401–404.

[17] J. M. Duarte-Carvajalino, G. Yu, L. Carin, and G. Sapiro, "Task-driven adaptive statistical compressive sensing of gaussian mixture models," *IEEE Trans. Signal Processing*, vol. 61, no. 3, pp. 585–600, 2013.

[18] A. Soni and J. Haupt, "On the fundamental limits of recovering tree sparse vectors from noisy linear measurements," *IEEE Trans. on Inform. Theory*, vol. 60, no. 1, pp. 133–149, 2014.

[19] M. Malloy and R. Nowak, "Near-optimal adaptive compressive sensing," *IEEE Trans. on Inform. Theory*, vol. 60, no. 4, pp. 4001–4012, 2014.

[20] R. M. Castro and E. Tánczos, "Adaptive compressed sensing for estimation of structured sparse sets," *arXiv preprint arXiv:1410.4593*, 2014.

[21] G. Strang, "A proposal for Toeplitz matrix calculations," *Studies in Applied Mathematics*, vol. 74, no. 2, pp. 171–176, 1986.

[22] T. F. Chan, "An optimal circulant preconditioner for Toeplitz systems," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, no. 4, pp. 766–771, 1988.

[23] E. E. Tyrtyshnikov, "Optimal and superoptimal circulant preconditioners," *SIAM Journal on Matrix Analysis and Applications*, vol. 13, no. 2, pp. 459–473, 1992.

[24] J. Müller-Quade, H. Aagedal, T. Beth, and M. Schmid, "Algorithmic design of diffractive optical systems for information processing," *Physica D: Nonlinear Phenomena*, vol. 120, no. 1, pp. 196–205, 1998.

[25] M. Schmid, R. Steinwandt, J. Müller-Quade, M. Rötteler, and T. Beth, "Decomposing a matrix into circulant and diagonal factors," *Linear Algebra and its Applications*, vol. 306, no. 1, pp. 131–143, 2000.

[26] M. Huhtanen, "How real is your matrix?," *Linear algebra and its applications*, vol. 424, no. 1, pp. 304–319, 2007.

[27] M. Huhtanen, "Factoring matrices into the product of two matrices," *BIT Numerical Mathematics*, vol. 47, no. 4, pp. 793–808, 2007.

[28] M. Huhtanen, "Approximating ideal diffractive optical systems," *Journal of Math. Analysis and Applications*, vol. 345, no. 1, pp. 53–62, 2008.

[29] J. Romberg, "Compressive sensing by random convolution," *SIAM Journal on Imaging Sciences*, vol. 2, no. 4, pp. 1098–1128, 2009.

[30] J. Haupt, W. U. Bajwa, G. Raz, and R. Nowak, "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *IEEE Trans. Information Theory*, vol. 56, no. 11, pp. 5862–5875, 2010.

[31] H. Rauhut, "Compressive sensing and structured random matrices," in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, M. Fornasier, Ed., vol. 9, pp. 1–92. deGruyter, 2010.

[32] H. Rauhut, J. Romberg, and J. A. Tropp, "Restricted isometries for partial random circulant matrices," *Applied and Computational Harmonic Analysis*, vol. 32, no. 2, pp. 242–254, 2012.

[33] W. Yin, S. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with Toeplitz and circulant matrices," in *Visual Communications and Image Processing 2010*. International Society for Optics and Photonics, 2010, p. 77440K.

[34] A. Hinrichs and J. Vybíral, "Johnson-Lindenstrauss lemma for circulant matrices," *Random Structures & Alg.*, vol. 39, no. 3, pp. 391–398, 2011.

[35] J. Vybíral, "A variant of the Johnson-Lindenstrauss lemma for circulant matrices," *Journal of Functional Analysis*, vol. 260, no. 4, pp. 1096–1105, 2011.

[36] F. Krahmer and R. Ward, "New and improved Johnson-Lindenstrauss embeddings via the restricted isometry property," *SIAM Journal on Mathematical Analysis*, vol. 43, no. 3, pp. 1269–1281, 2011.

[37] H. Zhang and L. Cheng, "New bounds for circulant Johnson-Lindenstrauss embeddings," *arXiv preprint arXiv:1308.6339*, 2013.

[38] H. L. Yap, M. B. Wakin, and C. J. Rozell, "Stable manifold embeddings with structured random matrices," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 4, pp. 720–730, 2013.

[39] S. Jain and J. Haupt, "On convolutional approximations to linear dimensionality reduction operators for large scale data processing," *In Preparation*.

[40] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, 2009.

[41] Sameer A Nene, Shree K Nayar, Hiroshi Murase, et al., "Columbia object image library (COIL-20)," Tech. Rep., technical report CUCS-005-96, 1996.

[42] N. Ailon and B. Chazelle, "Approximate nearest neighbors and the fast Johnson-Llindenstrauss transform," in *Proc. ACM Symp. on Theory of Computing*, 2006, pp. 557–563.

[43] A. Dasgupta, R. Kumar, and T. Sarlós, "A sparse Johnson-Lindenstrauss transform," in *Proc. ACM Symp. on Theory of Computing*, 2010, pp. 341–350.