PIECEWISE TOEPLITZ MATRICES-BASED SENSING FOR RANK MINIMIZATION

Kezhi Li *Cristian R. Rojas* *Saikat Chatterjee* *Håkan Hjalmarsson*

ACCESS Linnaeus Centre, Royal Institute of Technology (KTH)
{kezhi, crro, sach, hjalmars}@kth.se

ABSTRACT

This paper proposes a set of piecewise Toeplitz matrices as the linear mapping/sensing operator $\mathcal{A} : \mathbf{R}^{n_1 \times n_2} \to \mathbf{R}^M$ for recovering low rank matrices from few measurements. We prove that such operators efficiently encode the information so there exists a unique reconstruction matrix under mild assumptions. This work provides a significant extension of the compressed sensing and rank minimization theory, and it achieves a tradeoff between reducing the memory required for storing the sampling operator from $\mathcal{O}(n_1 n_2 M)$ to $\mathcal{O}(\max(n_1, n_2)M)$ but at the expense of increasing the number of measurements by r. Simulation results show that the proposed operator can recover low rank matrices efficiently with a reconstruction performance close to the cases of using random unstructured operators.

***Index Terms*—Rank minimization, Toeplitz matrix, compressed sensing, coherence**

I. INTRODUCTION

As a dual of the compressed sensing (CS) problem [1], the matrix rank minimization problem has been extensively studied in recent years [2]–[4]. This problem arises in many fields such as system identification [4], computer vision [3] and quantum state tomography [5], where notions of order, dimensionality or complexity can be expressed in terms of the rank of a matrix. Let $\mathbf{X}$ be an $n_1 \times n_2$ matrix of low rank rank r, the sampling operation can be expressed as

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) = [\langle \mathbf{A}_1, \mathbf{X} \rangle, \cdots, \langle \mathbf{A}_M, \mathbf{X} \rangle]^T = \mathbf{A}\text{vec}(\mathbf{X}), \tag{1}$$

where $\mathbf{y}$ is a vector of measurements, $\mathcal{A} : \mathbf{R}^{n_1 \times n_2} \to \mathbf{R}^M$ is a linear transformation, $\mathbf{A}$ denotes the transform in matrix format in which $\mathbf{A} \in \mathbf{R}^{M \times (n_1 \cdot n_2)}$ (without loss of generality we assume that $n_1 < n_2$); and $\langle \mathbf{A}_j, \mathbf{X} \rangle = \text{Tr}(\mathbf{A}_j^T \mathbf{X}), j \in \{1, \cdots M\}$, (vec) is the operator that vectorizes the matrix $\mathbf{X}$ by concatenating the columns as a long vertical vector $\mathbf{x} \in \mathbf{R}^N, N = n_1 n_2$.

In compressed sensing, besides conventional Gaussian/Bernoulli random sensing approaches, many structured/deterministic sensing matrices have been proven suitable for recovery of compressible signals. The interest in using structured sensing matrices in CS stems from the application needs, mainly due to their low complexity in computation and memory, as well as hardware implementation [6]–[8].

While we note several attempts in CS to use structured sensing matrices, contributions to low rank matrix reconstruction are more scarce. There are some efforts such as in [2], [9], [10]. An existing key condition of $\mathcal{A}$ is the so called rank restricted isometry property (r-RIP) [2]. However, examining the r-RIP of a given operator $\mathcal{A}$ is NP-hard. In this paper, we show that if the $\mathbf{A}$ consists of independent identically distributed (i.i.d.) random variables forming a piecewise Toeplitz structure, it is feasible to recover the objective low rank matrix from its measurements uniquely. Instead of studying its r-RIP, we convert the uniqueness problem of rank minimization into a compressed sensing problem, and analyze the performance bounds of proposed matrices by using the tools from structured sensing matrix analysis in CS. The extension is not trivial, since the vectorized low rank matrix is no longer sparse. By utilizing such technique we may reduce the memory required to store the sampling operator from $\mathcal{O}(n_1 n_2 M)$ to $\mathcal{O}(M \max(n_1, n_2))$ at the expense of a few measurements under mild assumptions.

The rest of the paper is organized as follows. In Section II we formulate the problem and introduce the proposed piecewise Toeplitz matrices. The main result and the corresponding proof are presented in Section III. Simulations are given in Section IV and finally Section V addresses the conclusion.

I-A. Relations to Previous Works

There are existing random or structured matrices for the rank minimization problem, such as [2], [9], [10]. The idea of this paper different from the previous work is to use Toeplitz structured operators inspired from structured sensing matrices in CS [6], [11]. After decomposing the low rank matrix, we analyze the uniqueness of recovering a block sparse vector (as defined in [12]). In [11] the authors also adopted the Gershgorin circle theorem to bound the eigenvalues of Toeplitz matrices, while in this paper the different coherence expression makes the analysis much more complicated. The idea of decomposing low rank matrix by its columns also relates to CUR matrix decomposition and the Nystrom method [13]. In contrast, to these previous approaches, the core problem here is the unique reconstruction rather than the decomposition, so the coherence and RIP analyses are adopted.

II. PIECEWISE TOEPLITZ MATRIX-BASED SENSING

II-A. Problem Formulation

To exploit the low rank property, we suppose that the low rank matrix $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_2}]$ has rank r so that r of its columns $\mathbf{x}_i$ can represent the remaining $(n_2 - r)$ columns explicitly by their linear combination. Denote the selected r columns as $\mathbf{x}^{\diamond}$, $\mathbf{x}^{\diamond} \subset \{\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_{n_2}\}$, $\diamond \subset \{1, 2, \cdots, n\}$, card$(\diamond) = r$; and the remaining $(n_2 - r)$ columns as $\mathbf{x}^{*}$, $* \subset \{1, \cdots, n_2\}$, card$(*) = n_2 - r$. We call them primary columns and secondary columns, respectively. Then the $(n_2 - r)$ secondary columns can be represented as

$$\mathbf{x}_i = \sum_{j=1}^{r} \alpha_{ij} \mathbf{x}_{\diamond_j}, \tag{2}$$

$i \in \{*_1, *_2, \cdots, *_{n_2-r}\}$. Its matrix multiplication form is

$$\begin{bmatrix} \mathbf{x}_{*_1} \\ \mathbf{x}_{*_2} \\ \vdots \\ \mathbf{x}_{*_{n_2-r}} \end{bmatrix} = \Bigg\{ \underbrace{\begin{bmatrix} \alpha_{*_1\diamond_1} & \cdots & \alpha_{*_1\diamond_r} \\ \alpha_{*_2\diamond_1} & \cdots & \alpha_{*_2\diamond_r} \\ \vdots & \ddots & \vdots \\ \alpha_{*_{n_2-r}\diamond_1} & \cdots & \alpha_{*_{n_2-r}\diamond_r} \end{bmatrix}}_{\boldsymbol{\alpha}} \otimes \mathbf{I}_{n_1} \Bigg\} \begin{bmatrix} \mathbf{x}_{\diamond_1} \\ \mathbf{x}_{\diamond_2} \\ \vdots \\ \mathbf{x}_{\diamond_r} \end{bmatrix} \tag{3}$$

where $\otimes$ denotes the Kronecker product. Normally $r \ll n_2$, i.e. $\boldsymbol{\alpha}$ is a tall matrix. Thus $\mathbf{x} = \text{vec}(\mathbf{X})$ can also be decomposed in a block sparse manner as

$$\mathbf{x} = \boldsymbol{\Psi} \cdot \mathbf{f} \tag{4}$$

where $\boldsymbol{\Psi}$ is a sparse matrix with block diagonal matrices $\alpha_{ij}\mathbf{I}_{n_1}$ or $\mathbf{I}_{n_1}$ of size $n_1 \times n_1$, $\mathbf{f}$ is a block r-sparse vector (adopting the definition in [12]) $\mathbf{f} = \left[\mathbf{0}, \cdots, \mathbf{x}_{\diamond_1}^T, \mathbf{0}, \cdots, \mathbf{x}_{\diamond_r}^T, \cdots\right]^T$. Then (1) can be written as:

$$\mathbf{y} = \mathcal{A}(\mathbf{X}) = \mathbf{A}\mathbf{x} = \mathbf{A}\boldsymbol{\Psi}\mathbf{f} = \boldsymbol{\Theta}\mathbf{f}, \tag{5}$$

where $\boldsymbol{\Theta} = \mathbf{A}\boldsymbol{\Psi}$. This block sparse problem has been studied by analyzing the block-restricted isometry constant [12]. However, here the matrix $\boldsymbol{\Theta}$ is formed by the multiplication of $\mathbf{A}$ and an unknown structured sparsifying matrix $\boldsymbol{\Psi}$ in (4). And obviously this $\boldsymbol{\Psi}$ is neither unitary, nor can be constructed delicately from complete bases. It is an unknown block diagonal sparse matrix determined by the low rank matrix $\mathbf{X}$. The same low rank matrix may even lead to multiple $\boldsymbol{\Psi}$ and $\mathbf{f}$. Thus the conventional theory of block CS may encounter difficulties for such a problem.

II-B. Piecewise Toeplitz Matrices

*Definition 1 (***Piecewise Toeplitz):** A set of matrices $\{\mathbf{A}_1, \mathbf{A}_2, \cdots, \mathbf{A}_M\}$ of size $n_1 \times n_2$ are defined as piecewise Toeplitz matrices if

$$\begin{aligned} \mathbf{a}_1 = \text{vec}(\mathbf{A}_1)^T &= [\mathbf{a}_{11}^T, \mathbf{a}_{12}^T, \cdots, \mathbf{a}_{1n_2}^T], \\ &\vdots \\ \mathbf{a}_M = \text{vec}(\mathbf{A}_M)^T &= [\mathbf{a}_{M1}^T, \mathbf{a}_{M2}^T, \cdots, \mathbf{a}_{Mn_2}^T], \end{aligned} \tag{6}$$

where $\mathbf{a}_{ij}$ denotes the jth column of the matrix $\mathbf{A}_i$, and their piecewise concatenation matrices

$$\mathbf{A}[1] = \begin{bmatrix} \mathbf{a}_{11}^T \\ \vdots \\ \mathbf{a}_{M1}^T \end{bmatrix} \quad \cdots \quad \mathbf{A}[n_2] = \begin{bmatrix} \mathbf{a}_{1n_2}^T \\ \vdots \\ \mathbf{a}_{Mn_2}^T \end{bmatrix} \tag{7}$$

are all Toeplitz.

Proposition 1: For the measurement process $\mathbf{y}(j) = \langle \mathbf{A}_j, \mathbf{X} \rangle, j \in \{1, \cdots, M\}$, and a matrix $\mathbf{X}$ of size $n_1 \times n_2$ and rank r, (1) is equivalent to $\mathbf{y} = \boldsymbol{\Theta}\mathbf{f}$, where $\mathbf{f}$ is a block r-sparse vector, and $\boldsymbol{\Theta}$ of size $M \times (n_1 n_2)$ has the structure $\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}[1] & \boldsymbol{\Theta}[2] & \cdots & \boldsymbol{\Theta}[i] & \cdots & \boldsymbol{\Theta}[n] \end{bmatrix}$ in which

$$\boldsymbol{\Theta}[i] = \begin{cases} \mathbf{A}[i] + \sum_{*} \alpha_{*i} \mathbf{A}[*] & \text{if} \quad i \in \{\diamond\} \\ \mathbf{0} & \text{if} \quad i \notin \{\diamond\} \end{cases} \tag{8}$$

for all $i \in \{1, \cdots, n_2\}$. $\{\mathbf{A}[i]\}$ are matrices derived from $\{\mathbf{A}_j\}$ by concatenating their columns piecewisely as in (7). $\{\diamond\}, \{*\}$ represent the sets of primary columns and secondary columns indexes, respectively. card$(\diamond) = r$, card$(*) = n_2 - r$. $\boldsymbol{\Theta}[i]$ are also Toeplitz when $i \in \{\diamond\}$.

Proof: Because $\mathbf{X}$ has rank r,

$$\mathbf{y} = [\mathbf{A}[1]\ \mathbf{A}[2]\ \cdots\ \mathbf{A}[n]]\, \boldsymbol{\Psi}\mathbf{f} = \boldsymbol{\Theta}\mathbf{f}. \tag{9}$$

Then it is straightforward to verify the expression in *Prop.* 1. Because $\mathbf{A}[i]$ are Toeplitz, $\boldsymbol{\Theta}[i]$ must be Toeplitz as well when $i \in \{\diamond\}$. ■

Remark: Since the decomposition $\mathbf{x} = \boldsymbol{\Psi}\mathbf{f}$ is not unique, $\boldsymbol{\Theta}$ have different expressions corresponding to various $\mathbf{f}$, which distinguishes (9) from CS with multiple solutions $\mathbf{f}$. Fortunately what we need to recover is not $\mathbf{f}$ but $\mathbf{x}$. Multiple $\mathbf{f}$ may lead to a unique solution $\mathbf{x}$.

III. MAIN RESULT AND PROOF

*Proposition 2 (***Unique Recovery):** The reconstruction of matrix $\mathbf{X}$ with size $n_1 \times n_2$ and rank r in (5) has a unique solution $\hat{\mathbf{X}}$ if $\boldsymbol{\Theta}\mathbf{f} \neq 0$ holds for every $\mathbf{f} \neq 0$ in (5) which is block $2r$-sparse, where $\boldsymbol{\Theta}$ has the structure in (8).

Proof: Assume that there is a new solution $\mathbf{X}^*$ to (5) with rank$(\mathbf{X}^*) \leq r$, $\mathbf{X}^* \neq \hat{\mathbf{X}}$, which means that

$$\hat{\mathbf{x}} = \hat{\boldsymbol{\Psi}}\hat{\mathbf{f}}, \mathbf{x}^* = \boldsymbol{\Psi}^*\mathbf{f}^*, \hat{\mathbf{x}} \neq \mathbf{x}^*. \tag{10}$$

Let $\mathbf{X}' = \mathbf{X}^* - \hat{\mathbf{X}}$, where $\mathbf{X}'$ is a nonzero matrix of rank at most $2r$. Then there must exist an $\mathbf{f}'$ that is block $2r$-sparse such that $\mathbf{A}\boldsymbol{\Psi}'\mathbf{f}' = \mathbf{0}$, where $\boldsymbol{\Psi}'\mathbf{f}' = \mathbf{X}'$, which contradicts the assumption. Please note that $\mathbf{f}' \neq \mathbf{f}^* - \hat{\mathbf{f}}$ if $\boldsymbol{\Psi}^* \neq \hat{\boldsymbol{\Psi}}$. The result is arrived from the above analyses. ■

Remark: *Prop.* 2 is an extension of the uniqueness guarantee from a sparse vector in CS to a low rank matrix. $\boldsymbol{\Theta}$ depends on the unknown matrix $\mathbf{X}$, $\mathbf{f}$ is not unique. However, because $\mathbf{X}'$ has rank at most $2r$, it must be decomposed into $\boldsymbol{\Psi}'$ and a $2r$-sparse vector $\mathbf{f}'$, which is in contradiction to the assumption.

Proposition 3 (ϵ **Bound):** Consider $\mathbf{\Theta} = \mathbf{A}\mathbf{\Psi}$ with the structure in (8). Denote by $\mathbf{\Theta}[\diamond]$ the submatrix formed by retaining the column blocks of $\mathbf{\Theta}$ indexed by $\diamond$. If the normalized Gram matrix $\mathbf{G}$ of every $\mathbf{\Theta}[\diamond]$ has bound

$$\begin{aligned} |g_{ii} - 1| &\leq \epsilon_1, \\ \mathcal{R}_i = \sum_{\substack{j=1 \\ j\neq i}} |g_{ij}| &\leq \epsilon_2, \end{aligned} \tag{11}$$

with some positive values $0 < \epsilon_1, \epsilon_2 < 1$, then the eigenvalues of $\mathbf{G}(\mathbf{\Theta}[\diamond])$ are bounded by $(1-\epsilon, 1+\epsilon)$, $\epsilon = \epsilon_1 + \epsilon_2$, and (5) has the unique block $2r$-sparse solution $\hat{\mathbf{f}}$ when ϵ is bounded by the RIP constant.

Proof: The proof is based on the Gershgorin circle theorem [14], and can be derived from the RIP and *Prop.* 2. ∎

Now we give the main result of this paper. The details of *Condition 1* and *Assumption 1* are provided in the Appendix.

*Theorem 1 (***Main Result):** Consider the measurements $\mathbf{y}(j) = \langle \mathbf{A}_j, \mathbf{X} \rangle, j \in \{1, \cdots, M\}$. Let $\mathbf{A}_j$ be piecewise random Toeplitz matrices whose entries satisfy *Condition 1*. $\mathbf{X}$ is a rank r matrix satisfying *Assumption 1*, when $r^2 < \mathcal{O}(n_1)$, $M \geq \mathcal{O}\left(r^2(n_1+n_2)\log(n_1 n_2)\right)$, then there exists a constant $c > 0$ such that for any fixed $\mathbf{X}$ has a unique solution $\hat{\mathbf{X}} = \mathbf{X}$ with probability exceeding $1 - \exp\left(-\frac{cM}{r^2 n_1^2}\right)$ as $n_2 = n_1^2$.

Proof: The proof exploits inner products of any two columns of $\mathbf{\Theta}$ in order to bound the eigenvalues. Denote the qth column in matrix $\mathbf{A}[p]$ as $\mathbf{A}[p,q], p \in \{1, \cdots, n_2\}, q \in \{1, \cdots, n_1\}$. For the entries in one row of the Gram matrix $\mathbf{G}$, there are four circumstances of non-zero $\theta[p_1, q_1]^T \theta[p_2, q_2]$: (1) $p_1 = p_2, q_1 = q_2$; (2) $p_1 \neq p_2, q_1 = q_2$; (3)$p_1 = p_2, q_1 \neq q_2$; (4) $p_1 \neq p_2, q_1 \neq q_2$. We will analyze them case by case.

(1) When $p_1 = p_2, q_1 = q_2$, they are the diagonal entries of the Gram matrix. Without loss of generality, we calculate $\theta_{\diamond_i q}^T \theta_{\diamond_i q}, i \in \{1, \cdots, r\}, q \in \{1, \cdots, n_1\}$, which implies $p_1 = \diamond_i, q_1 = q_2 = q$. Following the expressions in *Prop.* 1, the $((\diamond_i - 1)\cdot n_1 + q)$th column of $\mathbf{\Theta}$ can be calculated is

$$\theta[\diamond_i, q] = \mathbf{A}[\diamond_i, q] + \sum_{*} \alpha_{*\diamond_i} \mathbf{A}[*, q]. \tag{12}$$

Suppose $|\theta[\diamond_i, q](k)| \leq a$ where a is a positive bound. There is some positive value γ_i that $\mathbb{E}(\theta[\diamond_i, q]^2(k)) = \gamma_i^2 \sigma^2$ for every $k \in \{1, \cdots, M\}$, $\sigma^2 = 1/M$. By exploiting Hoeffding's inequality we have

$$\Pr\left\{ \left| \sum_{k=1}^{M} \theta[\diamond_i, q]^2(k) - \gamma_i^2 \sigma^2 M \right| \geq t_0 \right\} \leq 2\exp\left(-\frac{2t_0^2}{Ma^4}\right). \tag{13}$$

(2) When $p_1 \neq p_2, q_1 = q_2$, they represent two columns in different block $\mathbf{\Theta}[p_1], \mathbf{\Theta}[p_2]$ but have relatively the same internal positions. Let $p_1 = \diamond_i, p_2 = \diamond_j, q_1 = q_2 = q$, the two columns $\theta[\diamond_i, q], \theta[\diamond_j, q] = \mathbf{A}[\diamond_j, q]$ can be expressed as

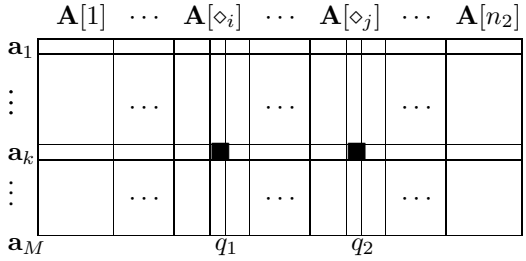


Fig. 1. The schematic diagram of matrix $\mathbf{A}$. $\mathbf{a}_k, \mathbf{A}[i]$ are defined in *Deft. 1*. q_1 and q_2 are internal column indexes. The two entries with black squares are denoted as $\mathbf{A}[\diamond_i, q_1](k), \mathbf{A}[\diamond_j, q_2](k)$, respectively.

$$\begin{aligned} \theta[\diamond_i, q] &= \mathbf{A}[\diamond_i, q] + \sum_{*} \alpha_{*\diamond_i} \mathbf{A}[*, q] \\ \theta[\diamond_j, q] &= \mathbf{A}[\diamond_j, q] + \sum_{*} \alpha_{*\diamond_j} \mathbf{A}[*, q]. \end{aligned} \tag{14}$$

Because all entries in $\mathbf{A}$ are i.i.d. in different blocks,

$$\begin{aligned} &\mathbb{E}(\theta[\diamond_i, q](k) \cdot \theta[\diamond_j, q](k)) \\ &= \mathbb{E}\left(\sum_{*} \alpha_{*\diamond_i} \mathbf{A}[*, q](k) \sum_{*} \alpha_{*\diamond_j} \mathbf{A}[*, q](k) \right) \\ &= \sum_{*} \alpha_{*\diamond_i} \alpha_{*\diamond_j} \cdot \sigma^2 \end{aligned} \tag{15}$$

Although $\theta[\diamond_i, q](k), \theta[\diamond_j, q](k)$ are dependent due to the mutual combination terms, we still can use Hoeffding's inequality to bound the summation, because for different k, $\theta[\diamond_i, q](k) \cdot \theta[\diamond_j, q](k)$ are i.i.d.. Let $\kappa_{ij} = \sum_{*} \alpha_{*\diamond_i} \alpha_{*\diamond_j}$, then there exists some positive t_1 such that

$$\Pr\left\{ \left| \theta[\diamond_i, q]^T \theta[\diamond_j, q] - \kappa_{ij} \sigma^2 M \right| \geq t_1 \right\} \leq 2\exp\left(-\frac{t_1^2}{2Ma^4}\right). \tag{16}$$

(3) When $p_1 = p_2, q_1 \neq q_2$, the two columns are in the same block $p_1 = p_2 = \diamond_i, i \in \{1, \cdots r\}$ with different internal index q_1, q_2:

$$\begin{aligned} \theta[\diamond_i, q_1] &= \mathbf{A}[\diamond_i, q_1] + \sum_{*} \alpha_{*\diamond_i} \mathbf{A}[*, q_1] \\ \theta[\diamond_i, q_2] &= \mathbf{A}[\diamond_i, q_2] + \sum_{*} \alpha_{*\diamond_i} \mathbf{A}[*, q_2]. \end{aligned} \tag{17}$$

Here a natural problem comes out: for these two columns, they are not independent any more due to the Toeplitz structure. Here we use "divide and conquer" technique that separates the sum into two groups that have no mutual terms. For instance, if $q_2 > q_1, q_2 - q_1 = d$, the sum can

be divided as

$$\theta[\diamond_i,q_1]^T\theta[\diamond_i,q_2] = \underbrace{\sum_{k=1}^{d}\theta[\diamond_i,q_1](k)\theta[\diamond_i,q_2](k+d)+\cdots}_{\text{first group}} + \underbrace{\sum_{k=d+1}^{2d}\theta[\diamond_i,q_1](k)\theta[\diamond_i,q_2](k+d)+\cdots}_{\text{second group}}, \tag{18}$$

and it is always possible to find a partition that divides $\theta[\diamond_i,q_1]^T\theta[\diamond_i,q_2]$ into two parts as sums with size $\{\frac{M}{2},\frac{M}{2}\}$ for even M and $\{\frac{M-1}{2},\frac{M+1}{2}\}$ for odd M. Thus for some positive value t_2,

$$\Pr\left\{\left|\theta[\diamond_i,q_1]^T\theta[\diamond_i,q_2]\right|\geq t_2\right\}\leq 4\exp\left(-\frac{t_2^2}{8Ma^4}\right). \tag{19}$$

(4) When $p_1\neq p_2, q_1\neq q_2$, similar to case (3), we divide $\theta^T_{\diamond_i q_1}\theta_{\diamond_j q_2}$ into 2 parts without mutual terms, giving

$$\Pr\left\{\left|\theta[\diamond_i,q_1]^T\theta[\diamond_j,q_2]\right|\geq t_3\right\}\leq 4\exp\left(-\frac{t_3^2}{8Ma^4}\right). \tag{20}$$

Finally, we normalize the Gram matrix and summarize the diagonal elements and off-diagonal elements. Assume that $\mathbf{X}$ satisfies the *statistical low rank property* defined in the *Def. 2* (see Appendix). Suppose $a=\sqrt{c_0/M}\gamma_m$, $\gamma_m=\max(\gamma_i)$. Let $\epsilon_{21}=\epsilon'_{21}+\frac{\kappa_{ij}}{\gamma^2},\epsilon'_{21}=\frac{t_1}{\gamma^2}$, $\gamma=\mathbb{E}(\gamma_i)$, and let $t_2=\gamma^2\epsilon_{22}, t_3=\gamma^2\epsilon_{23}, \epsilon_1=\epsilon_{2m}=\frac{1}{4}\epsilon, \epsilon_2=\sum_m\epsilon_{2m}, m=\{1,2,3\}$. After tedious calculation of bounding bias factor γ_i^2 and κ_{ij}, we obtain the normalized result by summarizing the four cases above and adopting the Gershgorin theorem

$$\Pr\left\{\bigcup_{i=1}^{n_1n_2}\{\sum_{j\neq i}^{rn_1}|g_{ij}|\geq\epsilon_2\}\right\}\leq 4n_1^2n_2^2\exp\left(-\frac{M\epsilon^2}{128c_0^2r^2n_1^2}\right). \tag{21}$$

Hence there must exist a constant $0<c<\frac{\epsilon^2}{128c_0^2}$ that

$$\Pr\{\text{Not Unique Recovery}\}\leq\exp\left(-\frac{cM}{r^2n_1^2}\right) \tag{22}$$

whenever

$$M\geq\left(\frac{384c_0^2}{\epsilon^2-128cc_0^2}\right)r^2(n_1+n_2)\log(n_1n_2), \tag{23}$$

as $r^2<\mathcal{O}(n_1), n_2=n_1^2, n_2\to\infty$ then $\Pr\to 0$. Use *Prop.* 2,3 to derive the last step, which completes the proof. ■

Remark: 1) For Gaussian matrices with i.i.d. entries, the measurements for the recovery of a low rank matrix should be at least $M\sim\mathcal{O}\left(r(n_1+n_2)\log(n_1n_2)\right)$ [2]. In our case, the measurement price is the extra factor r.

2) For random Gaussian matrices, one needs $\mathcal{O}(n_1n_2M)$ memory to store the operator. By using the piecewise Toeplitz structure, we are able to reduce the memory requirement to $\mathcal{O}\left((M+n_1)n_2\right)$ as $n_1<n_2<M$.

3) The proof holds when $r^2<\mathcal{O}(n_1), n_2=n_1^2$ in order to avoid the situation that $M\geq n_1n_2$. In practice it is feasible to apply $\mathbf{A}$ to the case when n_1, n_2 are close. Our simulations verify this conjecture numerically.

IV. SIMULATIONS

Extensive simulations have been carried out to compare the reconstruction performances of random and proposed operators. Here we present some results.

We utilize 3 different algorithms, including a) cvx toolbox to minimize the nuclear norm [15]; b) Alternating Least-Squares (ALS) algorithm [16] c) Directional-ALS algorithm [17], to compare the reconstruction results, respectively. For each algorithm, we recover a 50×50 random low rank matrix using different $\mathbf{A}$ such as Gaussian/Bernoulli operators, 3-valued operators [2] and finally random piecewise Toeplitz operators with truncated Gaussian entries. Fig. 1 depicts a comparison of reconstruction errors with increasing rank r at sampling rate $\rho=M/(n_1n_2)=0.3$. Each point is recorded as an average of 200 trials. From these curves one can observe that the performance of the proposed operator is close to that of random matrices, which are typically considered as the optimal universal operators. In addition, the proposed operators may be equipped with fast reconstruction algorithms potentially by exploiting the Toeplitz structure like in CS [7].

V. CONCLUSION

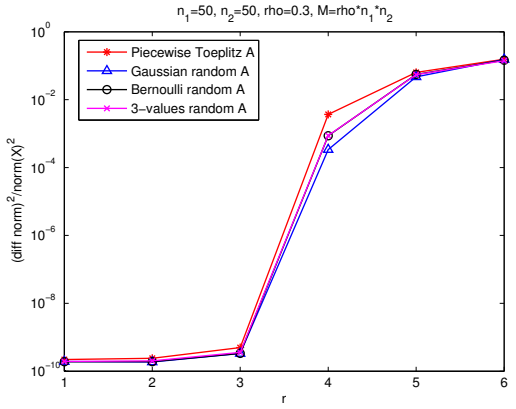
This paper proposes piecewise Toeplitz matrices as structured linear operators in the matrix minimization problem, and proves that it is feasible to recover the low rank matrix uniquely when the number of measurements exceeds $\mathcal{O}(r^2(n_1+n_2)\log(n_1n_2))$ under mild assumptions. Experimental results show that the proposed operators compare favorably with existing random operators.

VI. APPENDIX

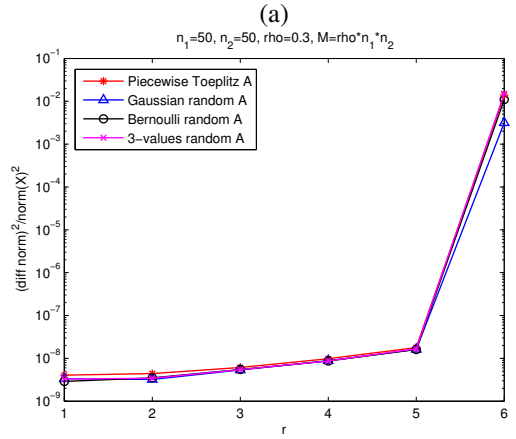
Condition 1 ($\mathbf{A}_j$)
$\mathbf{A}_j, j\in\{1,\ldots,M\}$ is an $n_1\times n_2$ matrix whose entries are bounded i.i.d. 0-mean $1/M$-variance random variables satisfying $|\mathbf{A}_j(p,q)|\leq\sqrt{c_0/M}$ for some $c_0>1$, $p\in\{1,\cdots,n_1\}, q\in\{1,\cdots,n_2\}$; in addition, $\{\mathbf{A}_1,\cdots,\mathbf{A}_M\}$ are a set of piecewise Toeplitz matrices defined in Sec. II-B.

Assumption 1 ($\boldsymbol{\alpha}$)
$\boldsymbol{\alpha}$ is the parameter matrix of r-rank $\mathbf{X}$ defined in (3). We assume that $\alpha_{*_p\diamond_q}, p\in\{1,\cdots,n-r\}, q\in\{1,\cdots,r\}$ satisfy the *statistical low rank property* defined below.

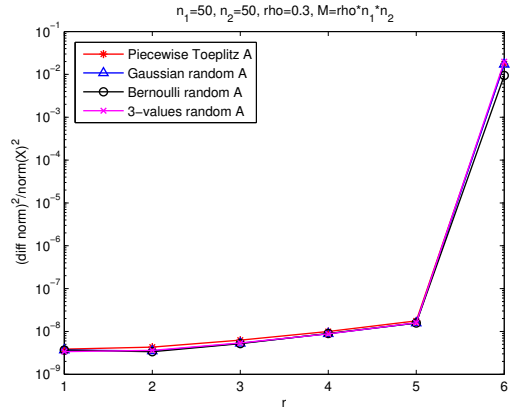
Definition 2 (**Statistical Low Rank Property):** A r-rank matrix $\mathbf{X}$ has the statistical low rank property if the α_{ij} in (3) are i.i.d. random variables with zero mean and variance σ_α^2.
In the proof of the main result we set $\sigma_\alpha^2=1/r$ to keep the variance of $\mathbf{x}_i$ identical due to (3).



(a)



(b)



(c)

Fig. 2. Simulation results (a) using cvx toolbox to minimize the nuclear norm of $\mathbf{X}$ (b) using Alternating Least-Squares (ALS) algorithm (c) using Directional-ALS algorithm.

VII. REFERENCES

[1] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inform. Theory*, vol. 52, pp. 1289–1306, Jul. 2006.

[2] B. Recht, M. Fazel, and P. Parillo, "Guaranteed minimum rank solution of matrix equations via nuclear norm minimization," *SIAM Rev.*, vol. 52, pp. 471–?01, 2007.

[3] E. J. Candés, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, pp. 1–37, Jun. 2011.

[4] I. Markovsky, "Structured low-rank approximation and its applications," *Automatica*, vol. 44, pp. 891–909, 2007.

[5] D. Gross, Y. Liu, S. T. Flammia, S. Becker, and J. Eisert, "Quantum state tomography via compressed sensing," *Phys. Rev. Lett.*, vol. 105, no. 15, pp. 150 401–150 404, Oct. 2010.

[6] M. Duarte and Y. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Processing*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[7] W. Yin, S. P. Morgan, J. Yang, and Y. Zhang, "Practical compressive sensing with Toeplitz and circulant matrices," Rice University CAAM Technical Report TR10-01, 2010.

[8] K. Li, L. Gan, and C. Ling, "Convolutional compressed sensing using deterministic sequences," *Signal Processing, IEEE Transactions on*, vol. 61, no. 3, pp. 740–752, 2013.

[9] M. A. Forbes and A. Shpilka, "On identity testing of tensors, low-rank recovery and compressed sensing," *Electronic Colloquium on Computational Comlexity Report*, no. 147, 2011.

[10] D. Gross, "Recovering low-rank matrices from few coefficients in any basis," *IEEE Trans. Inform. Theory*, vol. 57, no. 3, pp. 1548–1566, 2011.

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

[12] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[13] M. W. Mahoney and P. Drineas, "CUR matrix decompositions for improved data analysis," *Proceedings of the National Academy of Sciences*, vol. 106, no. 3, pp. 697–702, 2009.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1990.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[16] D. Zachariah, M. Sundin, M. Jansson, and S. Chatterjee, "Alternating least-squares for low-rank matrix reconstruction," *IEEE Signal Processing Letters*, vol. 19, no. 4, pp. 231–234, 2012.

[17] K. Li, M. Sundin, C. R. Rojas, S. Chatterjee, and M. Jansson, "Alternating strategies are good for low rank matrix reconstruction," 2014, preprint, available online.