ON COMPRESSIVE SENSING OF SPARSE COVARIANCE MATRICES USING DETERMINISTIC SENSING MATRICES

Alihan Kaplan, Volker Pohl

Lehrstuhl Theoretische Informationstechnik
Technische Universität München
80333 München, Germany

Dae Gwan Lee

Mathematisch-Geographische Fakultät
Katholische Universität Eichstätt-Ingolstadt
85071 Eichstätt, Germany

ABSTRACT

This paper considers the problem of determining the sparse covariance matrix $\mathbf{X}$ of an unknown data vector $\boldsymbol{x}$ by observing the covariance matrix $\mathbf{Y}$ of a compressive measurement vector $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$. We construct deterministic sensing matrices $\mathbf{A}$ for which the recovery of a k-sparse covariance matrix $\mathbf{X}$ from m values of $\mathbf{Y}$ is guaranteed with high probability. In particular, we show that the number of measurements m scales linearly with the sparsity k.

***Index Terms*—** Compressive sensing, covariance estimation, matrix sketching, statistical RIP

1. INTRODUCTION

Let $\boldsymbol{x} = (x_1, x_2, \dots, x_N)^{\mathrm{T}} \in \mathbb{C}^N$ be a vector of N independent, zero-mean random variables (r.v.) with covariance matrix $\mathbf{X} = \mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^*\right]$, and let $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$ be m linear measurements of $\boldsymbol{x}$ with the $m \times N$ measurement matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$. This paper considers the problem of determining $\mathbf{X}$ from the known covariance matrix

$$\mathbf{Y} = \mathbb{E}\left[\boldsymbol{y}\boldsymbol{y}^*\right] = \mathbf{A}\,\mathbb{E}\left[\boldsymbol{x}\boldsymbol{x}^*\right]\mathbf{A}^* = \mathbf{A}\mathbf{X}\mathbf{A}^* \tag{1}$$

of the observed measurements $\boldsymbol{y}$. This problem, also known as matrix sketching, appears in several problems of signal processing, like in array signal processing or communications [1].

In many applications, the covariance matrix $\mathbf{X}$ can assumed to be sparse in some sense. For example, if two r.v. x_i and x_j are known to be uncorrelated then the corresponding entries in the covariance matrix, namely $[\mathbf{X}]_{i,j} = \mathbb{E}[x_i\overline{x_j}]$ and $[\mathbf{X}]_{j,i} = \mathbb{E}[x_j\overline{x_i}]$, are equal to zero. So in cases where only a few entries of $\boldsymbol{x}$ are correlated, the matrix $\mathbf{X}$ will be sparse. Therefore, ideas from compressive sensing (CS) may be applied to find efficient sampling schemes which only need a few measurements to determine $\mathbf{X}$ [2]. In particular, it is natural to ask whether it is possible to find sensing matrices $\mathbf{A}$ with $m < N$ rows and such that $\mathbf{X}$ can uniquely recovered from $\mathbf{Y}$.

One common approach [3] is based on rewriting (1) as a standard linear compressive sensing (CS) problem by stacking the columns of $\mathbf{X}$ and $\mathbf{Y}$ into vectors $\widetilde{\boldsymbol{x}} = \mathrm{vec}(\mathbf{X}) \in \mathbb{C}^{N^2}$ and $\widetilde{\boldsymbol{y}} = \mathrm{vec}(\mathbf{Y}) \in \mathbb{C}^{m^2}$ respectively. This yields

$$\widetilde{\boldsymbol{y}} = \mathbf{C}\,\widetilde{\boldsymbol{x}} \qquad \text{with} \qquad \mathbf{C} = \overline{\mathbf{A}} \otimes \mathbf{A} \tag{2}$$

and wherein $\otimes$ stands for the usual Kronecker product of matrices. The problem is now to find a measurement matrix $\mathbf{A}$ such that the corresponding matrix $\mathbf{C} = \overline{\mathbf{A}} \otimes \mathbf{A}$ in (2) is a good measurement matrix for CS. Then the problem (2) can be solved uniquely by standard CS algorithms [4, 5].

In order to decide whether a given $\mathbf{A}$ is a good measurement matrix for CS, the so called *restricted isometric property (RIP)* is often applied [4, 6]:

Definition 1: *The k-th* restricted isometry constant (RIC) $\delta_k(\mathbf{A})$ *of a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ is the smallest $\delta \geq 0$ such that*

$$(1-\delta)\,\|\boldsymbol{x}\|_2^2 \leq \|\mathbf{A}\boldsymbol{x}\|_2^2 \leq (1+\delta)\,\|\boldsymbol{x}\|_2^2 \qquad \textit{for all } \boldsymbol{x} \in \Sigma_k^N\ ,$$

wherein Σ_k^N denotes the set of all k-sparse vectors $\boldsymbol{x} \in \mathbb{C}^N$.

The importance of the RIC stem from the fact that it provides guarantees for unique CS recovery. The following theorem gives just an example of such recovery guarantees (see, e.g., [4]).

Theorem 1: *Assume that the RIC for a matrix $\mathbf{A} \in \mathbb{C}^{m \times N}$ satisfies $\delta_{2k}(\mathbf{A}) < 1/3$, then every $\boldsymbol{x} \in \Sigma_k^N$ is the unique solution of*

$$\min\nolimits_{\boldsymbol{z} \in \mathbb{C}^N} \|\boldsymbol{z}\|_1 \quad \textit{subject to } \mathbf{A}\boldsymbol{z} = \mathbf{A}\boldsymbol{x}\ . \tag{3}$$

So, if $\delta_{2k}(\mathbf{A})$ is sufficiently small any k-sparse vector $\boldsymbol{x} \in \Sigma_k^N$ can be uniquely recovered from the m measurements $\boldsymbol{y} = \mathbf{A}\boldsymbol{x}$ by the optimization problem (3), known as *basis pursuit* [7]. There are two major problems with the RIP of matrices.

First, for a given matrix $\mathbf{A}$ the calculation of $\delta_k(\mathbf{A})$ requires a combinatorial search which is computationally infeasible for N large. Therefore, it is practically impossible to decide whether a given $\mathbf{A}$ satisfies the condition of Theorem 1. Mainly for this reason, probabilistic constructions where the entries of $\mathbf{A}$ are generated by independent identical distributed (i.i.d.) random variables are very common. Such probabilistic matrices are known to satisfy the k-RIP (with high probability) and the number of necessary measurements m is in the order of $k\log(N/k)$ [4, 6].

Secondly, in view of measurement matrices $\mathbf{C} = \overline{\mathbf{A}} \otimes \mathbf{A}$ with Kronecker structure, it is known [8] that $\delta_k(\overline{\mathbf{A}} \otimes \mathbf{A}) \geq \delta_k(\mathbf{A})$. So if $\mathbf{A} \in \mathbb{C}^{M \times N}$ satisfies the condition of Theorem 1, then every k-sparse vector in $\mathbb{C}^N$ can be recovered from the m measurements taken with $\mathbf{A}$, and in the best case the number of necessary measurements $m \sim k$ scales linearly with k. However, since $\delta_k(\overline{\mathbf{A}} \otimes \mathbf{A}) \geq \delta_k(\mathbf{A})$ Theorem 1 guarantees also only the recovery of k-sparse vectors in $\mathbb{C}^{N^2}$ from m^2 measurements taken with $\overline{\mathbf{A}} \otimes \mathbf{A} \in \mathbb{C}^{m^2 \times N^2}$, even though the signal space has a much larger dimension and although one has much more measurements. So this way, one only obtains recovery guarantees for matrices with Kronecker structure for measurement numbers $\widetilde{m} = m^2 \sim k^2$ which scale at least quadratically in the sparsity k, instead of the linear scaling for measurement matrices without Kronecker structure.

This work was supported by the German Research Foundation (DFG) within the priority program *Compressed Sensing in Information Processing (CoSIP)* under Grants PF 450/9–1 and PO 1347/3–1.

To overcome the first problem, an alternative approach to characterize good CS matrices was proposed in [9]. This approach is based on the so called *statistical RIP (StRIP)*, and we will recapture the main definitions and ideas shortly in Sec. 2. Then Sec. 3 applies this framework to study StRIP for sensing matrices with Kronecker structure. In particular, we will construct matrices $\mathbf{A} \in \mathbb{C}^{m\times N}$ such that $\overline{\mathbf{A}} \otimes \mathbf{A} \in \mathbb{C}^{m^2 \times N^2}$ satisfies a statistical recovery guaranty with $\widetilde{m} = m^2 \sim k \log(N)$ measurements.

2. NOTATIONS AND STRIP

General notations Let $\mathbf{\Phi} \in \mathbb{C}^{m\times N}$ be a matrix with m rows and N columns. The j-th column of $\mathbf{\Phi}$ will be denoted by $\boldsymbol{\varphi}_j$, and $\boldsymbol{\varphi}_j[k]$ stands for the k-th entry of $\boldsymbol{\varphi}_j \in \mathbb{C}^m$. Assume that the columns of $\mathbf{\Phi}$ are normalized by $\|\boldsymbol{\varphi}_j\| = 1$ for all $j = 1, \dots, N$. Then the *coherence* of $\mathbf{\Phi}$ is defined to be $\mu(\mathbf{\Phi}) = \max_{i\neq j} |\langle \boldsymbol{\varphi}_i, \boldsymbol{\varphi}_j\rangle|$. Moreover, if $\mathcal{K}$ is a subset of $\{1, 2, \dots, N\}$ then $\mathbf{\Phi}_{\mathcal{K}}$ stands for the matrix consisting of the columns of $\mathbf{\Phi}$ indexed by $\mathcal{K}$.

Statistical RIP In [9] a statistical version of the RIP was introduced to investigate deterministic measurement matrices for CS. Since the matrices are deterministic, the probability enters in the signal model. In the following we briefly discuss the main definitions and concepts.

Definition 2: *A matrix* $\mathbf{A} = \frac{1}{\sqrt{m}}\mathbf{\Phi} \in \mathbb{C}^{m\times N}$ *with* ℓ_2*-normalized columns is said to have* (k, δ, ϵ)-StRIP *if*

$$(1-\delta)\,\|\boldsymbol{x}\|_2^2 \leq \|\mathbf{A}\boldsymbol{x}\|_2^2 \leq (1+\delta)\,\|\boldsymbol{x}\|_2^2$$

holds with probability exceeding $1-\epsilon$ *for a random vector* $\boldsymbol{x} \in \Sigma_k^N$ *drawn from a uniform distribution over all* $\{z \in \Sigma_k^N \,:\, \|z\|_2 = 1\}$. *Further, we say that* A *has* (k, δ, ϵ)*-uniqueness-guaranteed StRIP (abbr.* (k, δ, ϵ)-UStRIP*) if*

$$\left\{\boldsymbol{z} \in \Sigma_k^N \,:\, \mathbf{A}\boldsymbol{z} = \mathbf{A}\boldsymbol{x}\right\} = \{\boldsymbol{x}\}$$

is also satisfied with probability exceeding $1-\epsilon$.

So a sensing matrix having (k, δ, ϵ)-*StRIP* satisfies the standard restricted isometry property (RIP) with high probability. Nevertheless, the StRIP property does not guarantee unique recovery in general, not even with high probability. The unique recovery is guaranteed for the class of UStRIP-matrices (by definition).

To identify matrices which have UStRIP, the following conditions were introduced in [9].

Definition 3: *A matrix* $\mathbf{A} = \frac{1}{\sqrt{m}}\mathbf{\Phi} \in \mathbb{C}^{m\times N}$ *with all entries of* $\mathbf{\Phi}$ *having absolute value* 1*, is said to be* η*-StRIP-able if the following three conditions are satisfied.*

(St1) *The rows of* $\mathbf{\Phi}$ *are mutually orthogonal, and the sum of all entries in each row is zero, i.e.,*

$$\sum_{j=1}^N \phi_j[k]\,\overline{\phi_j[\ell]} = 0 \qquad \text{if } k \neq \ell,$$
$$\sum_{j=1}^N \phi_j[k] = 0 \qquad \text{for all } k = 1, \dots, m.$$

(St2) *The columns of* $\mathbf{\Phi}$ *form a group under pointwise multiplication, i.e., for all* $j, j' \in \{1, \dots, N\}$ *there exists* $j'' \in 1, \dots, N$ *such that*

$$\phi_j[k]\,\phi_{j'}[k] = \phi_{j''}[k] \qquad \text{for all } k = 1, \dots, m\ .$$

In particular, there is one column of $\mathbf{\Phi}$ *(the identity element of this group) with all its entries equal to* 1*. Without loss of generality, we assume that* $\boldsymbol{\varphi}_1$ *is this identity vector.*

(St3) *There exists* $\eta > 0$ *such that*

$$\left|\sum_{k=1}^m \boldsymbol{\varphi}_j[k]\right|^2 \leq m^{2-\eta} \quad \text{for all } j = 2, 3, \dots, N\ .$$

Remark 1: Condition (St2) implies that $|\boldsymbol{\varphi}_j[k]| = 1$ for all j and k, and that the set of column vectors $\{\boldsymbol{\varphi}_j\}_{j=1}^N$ is closed under complex conjugation, i.e., for any j there is j' so that $\boldsymbol{\varphi}_{j'} = \overline{\boldsymbol{\varphi}_j}$.

Remark 2: The parameter $\eta > 0$ is closely related to the coherence of $\mathbf{A}$. Using (St2) and (St3), it follows that $\mu(\mathbf{A}) \leq m^{-\eta/2}$. So a large η implies small coherence of $\mathbf{A}$.

Remark 3: (St1)–(St3) imply that $\{\boldsymbol{\varphi}_j\}_{j=1}^N$ is a tight frame for $\mathbb{C}^m$ with frame bound m, i.e., $\sum_{j=1}^N |\langle \boldsymbol{x}, \boldsymbol{\varphi}_j\rangle|^2 = m\|\boldsymbol{x}\|^2, \forall\, \boldsymbol{x} \in \mathbb{C}^m$.

The next theorem shows that any η-StRIP-able matrix has (k, δ, ϵ)-UStRIP if $\eta > 1/2$ and if the sparsity k satisfies some conditions.

Theorem 2 (Theorem 8 in [9]): *Let* $\mathbf{A} = \frac{1}{\sqrt{m}}\mathbf{\Phi} \in \mathbb{C}^{m\times N}$ *be an* η*-StRIP-able matrix with* $\eta > 1/2$*. If* $k < 1 + (N-1)\delta$ *and* $m \geq c\,(k \log N)/\delta^2$ *for some constant* $c > 0$ *then* $\mathbf{A}$ *has* $(k, \delta, 2\epsilon)$-UStRIP *with* $\epsilon = 2\exp\left(-\left(\delta - \frac{k-1}{N-1}\right)^2 \frac{m^\eta}{8k}\right)$.

This theorem reduces the search of good deterministic sensing matrices to finding matrices that satisfy the conditions (St1)–(St3) with an $\eta > 1/2$. Whereas it is basically impossible to calculate the RIC for a given matrix $\mathbf{A}$, it is fairly easy to check whether $\mathbf{A}$ is η-StRIP-able. On the other side, conditions (St1)–(St3) are fairly strong restrictions on the structure of $\mathbf{A}$. Nevertheless, [9] derived numerous StRIP-able matrices.

The next section will apply this framework to characterize matrices of Kronecker structure which are *UStRIP*.

3. STRIP–ABLE KRONECKER MATRICES

We consider the following problem. Let $\mathbf{A}$ be a matrix which is known to be η-StRIP-able for some $\eta > 0$. Is the Kronecker product $\overline{\mathbf{A}} \otimes \mathbf{A}$ again η'-StRIP-able for some $\eta' > 0$? If so, what would be the value of η'? The next theorem will give a complete answer to these questions.

Theorem 3: *Assume* $\boldsymbol{A} \in \mathbb{C}^{n\times N}$ *is* η_A*-StRIP-able and* $\boldsymbol{B} \in \mathbb{C}^{m\times M}$ *is* η_B*-StRIP-able, then the following holds.*

(a) $\overline{\boldsymbol{A}}$ *is* $\eta_{\overline{A}}$*-StRIP-able with* $\eta_{\overline{A}} = \eta_A$.

(b) The matrix $\boldsymbol{C} = \boldsymbol{A} \otimes \boldsymbol{B} \in \mathbb{C}^{nm\times NM}$ *is* η_C*-StRIP-able with*

$$\eta_C = \begin{cases} \eta_A \frac{\ln(n)}{\ln(nm)} & \text{if } n^{\eta_A} \leq m^{\eta_B} \\ \eta_B \frac{\ln(m)}{\ln(nm)} & \text{if } n^{\eta_A} > m^{\eta_B}. \end{cases} \tag{4}$$

Proof: Let us write $\mathbf{A} = \frac{1}{\sqrt{n}}\mathbf{\Phi}$, $\mathbf{B} = \frac{1}{\sqrt{m}}\mathbf{\Psi}$, and $\mathbf{C} = \frac{1}{\sqrt{nm}}\mathbf{\Gamma}$. Clearly, the matrices $\mathbf{\Phi}, \mathbf{\Psi}, \mathbf{\Gamma}$ are related by $\mathbf{\Gamma} = \mathbf{\Phi} \otimes \mathbf{\Psi}$ and the entry of $\mathbf{\Gamma}$ in the (k, ℓ)-th column and (x, y)-th row is given by $\boldsymbol{\gamma}_{(k,\ell)}[x, y] = \boldsymbol{\varphi}_k[x]\boldsymbol{\psi}_\ell[y]$.

Part (a) is immediate by observing that each conditions (St1), (St2), (St3) for $\mathbf{\Phi}$ implies the respective condition for $\overline{\mathbf{\Phi}}$. Therefore $\overline{\mathbf{A}}$ is $\eta_{\overline{A}}$-StRIP-able with $\eta_{\overline{A}} = \eta_A$ (the same constant as $\mathbf{A}$).

To prove (b), we check conditions (St1), (St2) and (St3) for $\mathbf{\Gamma}$. First, observe that for $(x, y) \neq (x', y')$,

$$\begin{aligned} &\sum_{k=1}^N \sum_{\ell=1}^M \boldsymbol{\gamma}_{(k,\ell)}[x, y]\,\overline{\boldsymbol{\gamma}_{(k,\ell)}[x', y']} \\ &= \sum_{k=1}^N \sum_{\ell=1}^M \boldsymbol{\varphi}_k[x]\,\boldsymbol{\psi}_\ell[y]\,\overline{\boldsymbol{\varphi}_k[x']\,\boldsymbol{\psi}_\ell[y']} \\ &= \sum_{k=1}^N \boldsymbol{\varphi}_k[x]\,\overline{\phi_k[x']} \cdot \sum_{\ell=1}^M \psi_\ell[y]\,\overline{\psi_\ell[y']} = 0, \end{aligned}$$

using the fact that $\sum_{k=1}^{N} \boldsymbol{\varphi}_k[x]\overline{\boldsymbol{\varphi}_k[x']} = 0$ if $x \neq x'$, and that $\sum_{\ell=1}^{M} \boldsymbol{\psi}_\ell[y]\overline{\boldsymbol{\psi}_\ell[y']} = 0$ if $y \neq y'$. Moreover, for any (x, y),

$$\begin{aligned} \textstyle\sum_{k=1}^{N} \sum_{\ell=1}^{M} \boldsymbol{\gamma}_{(k,\ell)}[x,y] &= \textstyle\sum_{k=1}^{N} \sum_{\ell=1}^{M} \boldsymbol{\varphi}_k[x]\boldsymbol{\psi}_\ell[y] \\ &= \textstyle\sum_{k=1}^{N} \boldsymbol{\varphi}_k[x] \cdot \sum_{\ell=1}^{M} \boldsymbol{\psi}_\ell[y] = 0, \end{aligned}$$

where we have used that $\sum_{k=1}^{N} \boldsymbol{\varphi}_k[x] = \sum_{\ell=1}^{M} \boldsymbol{\psi}_\ell[y] = 0$. Therefore, $\boldsymbol{\Gamma}$ satisfies the condition (St1).

To verify the (St2) for $\boldsymbol{\Gamma}$, fix any (k, ℓ) and (k', ℓ'), where $1 \leq k, k' \leq N$, $1 \leq \ell, \ell' \leq M$. Since $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ satisfy (St2) there exist $1 \leq k'' \leq N$ and $1 \leq \ell'' \leq M$ such that $\boldsymbol{\varphi}_k[x]\boldsymbol{\varphi}_{k'}[x] = \boldsymbol{\varphi}_{k''}[x]$ for all x and $\boldsymbol{\psi}_\ell[y]\boldsymbol{\psi}_{\ell'}[y] = \boldsymbol{\psi}_{\ell''}[y]$ for all y. Then

$$\begin{aligned} \boldsymbol{\gamma}_{(k,\ell)}[x,y] \cdot \boldsymbol{\gamma}_{(k',\ell')}[x,y] &= \boldsymbol{\varphi}_k[x]\boldsymbol{\psi}_\ell[y] \cdot \boldsymbol{\varphi}_{k'}[x]\boldsymbol{\psi}_{\ell'}[y] \\ &= \boldsymbol{\varphi}_k[x]\boldsymbol{\varphi}_{k'}[x] \cdot \boldsymbol{\psi}_\ell[y]\boldsymbol{\psi}_{\ell'}[y] \\ &= \boldsymbol{\varphi}_{k''}[x] \cdot \boldsymbol{\psi}_{\ell''}[y] = \boldsymbol{\gamma}_{(k'',\ell'')}[x,y] \end{aligned}$$

which proves (St2).

Finally, we verify (St3) for $\boldsymbol{\Gamma}$. For $(k, \ell) \neq (1, 1)$ and with $1 \leq k \leq N$ and $1 \leq \ell \leq M$, we get

$$\begin{aligned} \Big|\textstyle\sum_{x=1}^{n} \sum_{y=1}^{m} \boldsymbol{\gamma}_{(k,\ell)}[x,y]\Big|^2 &= \Big|\textstyle\sum_{x=1}^{n} \sum_{y=1}^{m} \boldsymbol{\varphi}_k[x]\,\boldsymbol{\psi}_\ell[y]\Big|^2 \\ &= \big|\textstyle\sum_{x=1}^{n} \boldsymbol{\varphi}_k[x]\big|^2 \cdot \big|\textstyle\sum_{y=1}^{m} \boldsymbol{\psi}_\ell[y]\big|^2 \\ &= \begin{cases} n^2 \cdot m^{2-\eta_B} & \text{if } k = 1,\ \ell \neq 1, \\ n^{2-\eta_A} \cdot m^2 & \text{if } k = 1,\ \ell \neq 1, \\ n^{2-\eta_A} \cdot m^{2-\eta_B} & \text{if } k \neq 1,\ \ell \neq 1, \end{cases} \end{aligned}$$

and since $\eta_A, \eta_B > 0$, we have

$$\max_{(k,\ell)\neq(1,1)} \left|\sum_{x=1}^{n}\sum_{y=1}^{m} \boldsymbol{\gamma}_{(k,\ell)}[x,y]\right|^2 = \max\left\{n^2 m^{2-\eta_B},\ n^{2-\eta_A} m^2\right\}.$$

Setting $(nm)^{2-\eta_C} = \max\{n^2 m^{2-\eta_B}, n^{2-\eta_A} m^2\}$ gives

$$\eta_C = \begin{cases} \eta_A \frac{\ln(n)}{\ln(nm)} & \text{if } n^{\eta_A} \leq m^{\eta_B} \\ \eta_B \frac{\ln(m)}{\ln(nm)} & \text{if } n^{\eta_A} > m^{\eta_B} \end{cases}$$

which finishes the proof. ∎

Theorem 3 shows that the Kronecker product $\mathbf{C} = \mathbf{A} \otimes \mathbf{B}$ of two StRIP-able matrices $\mathbf{A}$ and $\mathbf{B}$ is again StRIP-able. However, the constant η_C given in (4) is always strictly smaller than both η_A and η_B, i.e., $\eta_C < \min\{\eta_A, \eta_B\}$. In view of Remark 2, this means that the coherence of a Kronecker product matrix is always worse (i.e., larger) than the coherence of the original matrices.

Motivated by the applications described in the introduction, we are interested in sensing matrices of the form $\overline{\mathbf{A}} \otimes \mathbf{A}$ (cf. (2)). For such matrices, Theorem 3 immediately yields the following statement.

Corollary 4: *If $\mathbf{A} \in \mathbb{C}^{m\times N}$ is an η-StRIP-able matrix then the Kronecker structured matrix $\overline{\mathbf{A}} \otimes \mathbf{A} \in \mathbb{C}^{m^2\times N^2}$ is $(\eta/2)$-StRIP-able.*

Combining Theorem 2 and Corollary 4, one obtains a sufficient condition under which a Kronecker product matrix $\overline{\mathbf{A}} \otimes \mathbf{A} \in \mathbb{C}^{m^2\times N^2}$ satisfies UStRIP.

Corollary 5: *Let $\mathbf{A} = \frac{1}{\sqrt{m}}\boldsymbol{\Phi} \in \mathbb{C}^{m\times N}$ be an η-StRIP-able matrix with $\eta > 1$. If $k < 1 + (N^2 - 1)\delta$ and $m^2 \geq c\,(2k \log N)/\delta^2$ for some constant $c > 0$, then $\overline{\mathbf{A}} \otimes \mathbf{A}$ has $(k, \delta, 2\epsilon)$-*UStRIP *with*

$$\epsilon = 2\exp\left(-\Big(\delta - \frac{k-1}{N^2-1}\Big)^2 \frac{m^{2\eta}}{8k}\right). \tag{5}$$

So if a matrix $\mathbf{A}$ satisfies the conditions of Corollary 5 then one needs in the order of $\widetilde{m} := m^2 \geq c\,k \log N$ measurements of the form $\boldsymbol{y} = \big(\overline{\mathbf{A}}\otimes\mathbf{A}\big)\boldsymbol{x}$ to recover k-sparse vectors $\boldsymbol{x} \in \mathbb{C}^{N^2}$ with high probability. Equivalently, every covariance matrix $\mathbf{X} \in \mathbb{C}^{N\times N}$ can be recovered from $\widetilde{m}$ values of the covariance matrix $\mathbf{Y} = \mathbf{A}\mathbf{X}\mathbf{A}^*$.

4. KRONECKER MATRICES WITH RECOVERY GUARANTEE

Corollary 5 requires that the StRIP constant η of an StRIP-able matrix $\mathbf{A} \in \mathbb{C}^{m\times N}$ has to be larger than 1 for $\overline{\mathbf{A}} \otimes \mathbf{A}$ to have UStRIP. To get a first idea which matrices might satisfy this condition, we recall from Remark 2 that the coherence of an η-StRIP-able matrix is upper bounded by $\mu(\mathbf{A}) \leq m^{-\eta/2}$. On the other side, $\mu(\mathbf{A})$ is known to be lower bounded by the Welsh bound [10]. So $\mu(\mathbf{A})$ always satisfies

$$\sqrt{\frac{N-m}{m(N-1)}} \leq \mu(\mathbf{A}) \leq \frac{1}{\sqrt{m^\eta}}.$$

From these inequalities, one easily derives an upper bound on η:

$$\eta \leq 1 + \ln\left(\tfrac{N-1}{N-m}\right)\tfrac{1}{\ln(m)}.$$

For $m > 1$, this upper bound is strictly larger than 1 but it gets very close to 1 for $m \ll N$ (as usually desired in CS). Since we are looking for matrices with $\eta > 1$, this means that we are searching for matrices $\mathbf{A}$ whose coherence is very close to the Welch bound, which means that the columns of $\mathbf{A}$ have to be close to an *equiangular tight frame* (ETF). In particular, we observe that if the coherence of $\mathbf{A}$ would achieve the Welch bound with equality then the Kronecker product $\overline{\mathbf{A}} \otimes \mathbf{A}$ would have UStRIP. Additionally, such an equal norm ETF needs to fulfill (St1) – (St3). A class of ETFs fulfilling these conditions are *equiangular harmonic frames (EHF)* [11, Chap. 5]. These frames are constructed by selecting certain rows from the DFT matrix. The selected rows are indexed by a so-called *difference sets* [12]. Note that by selecting arbitrary rows (apart from the first all ones row) of the DFT matrix, the partial DFT matrix fulfills (St1) and (St2), but not necessarily (St3). To be self-contained we give a short description of the construction of EHFs.

Definition 4: *An (N, m, ρ)-difference set is a set $\mathcal{K} \subset \mathbb{Z}_N$ of size m such that every nonzero element of the N-element cyclic group $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$ can be expressed as a difference of two elements from $\mathcal{K}$ in exactly ρ ways.*

Example 1: The set $\mathcal{K} = \{1, 2, 4\}$ is a $(7, 3, 1)$-difference set in the group $\mathbb{Z}_7$. Indeed, we have

$$\begin{array}{lll} 1 - 2 = 6, & 2 - 1 = 1, & 4 - 1 = 3, \\ 1 - 4 = 4, & 2 - 4 = 5, & 4 - 2 = 2, \end{array}$$

which shows that every nonzero element of $\mathbb{Z}_7$ is expressed as a difference of two elements from $\mathcal{K}$ in exactly one way.

Many types of difference sets are known. One particular example is the $(q^2 + q + 1, q + 1, 1)$-difference set (due to Singer [13]).

Proposition 1: *Let $\mathcal{K} \subset \mathbb{Z}_N$ be an (N, m, ρ)-difference set. Then the partial Fourier matrix $\mathbf{F}_{\mathcal{K}} \in \mathbb{C}^{m\times N}$ is η-StRIP-able with*

$$\eta = 2 - \frac{\ln(m-\rho)}{\ln(m)} > 1,$$

where $\mathbf{F}_{\mathcal{K}} = [\mathrm{e}^{\mathrm{i}2\pi jk/N}]_{k\in\mathcal{K},j=0,\ldots,N-1}$ is the partial Fourier matrix formed with the rows indexed by $\mathcal{K}$.

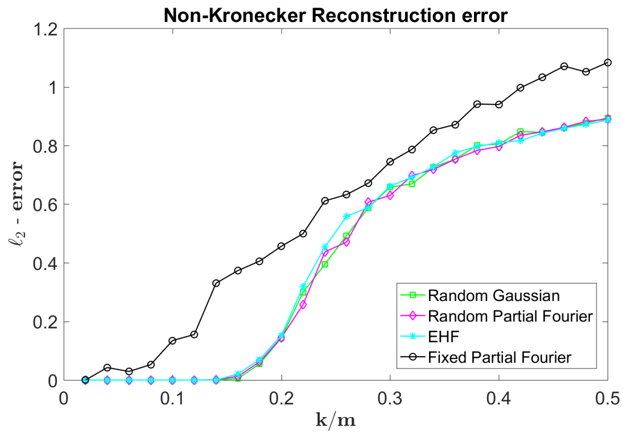

Fig. 1. Quadratic error of the solutions of (7) for non-Kronecker structured matrices $\mathbf{C}$. Horizontal axis: k/m (sparsity over number of measurements). Vertical axis: ℓ_2-error ($\|\widehat{\boldsymbol{x}}-\boldsymbol{x}\|^2 \,/\, \|\boldsymbol{x}\|^2$).

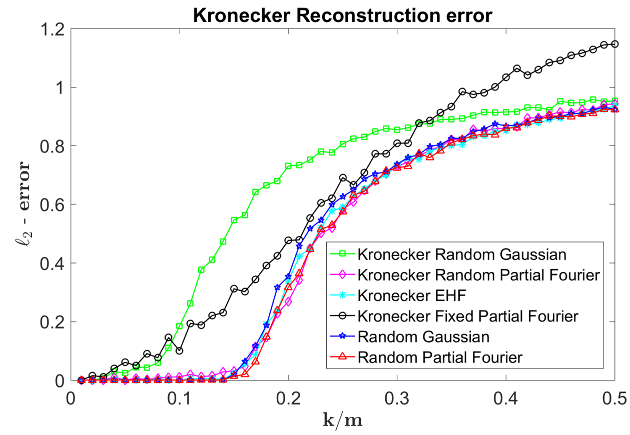

Fig. 2. Quadratic error of the optimal solutions of (7) for Kronecker structured matrices $\mathbf{C}$, and comparison with random Gaussian and random partial Fourier matrices. Axis as in Fig. 1.

Proof: Conditions (St1) and (St2) follow from the properties of the $N \times N$ DFT matrix. To verify (St3), let $\mathcal{K} = \{a_1, a_2, \ldots, a_m\} \subset \mathbb{Z}_N$ be given in increasing order. The entry of $\mathbf{F}_{\mathcal{K}}$ in the j-th column and k-th row is given by $\mathbf{f}_j[k] = \omega^{ja_k}$, where $\omega := e^{2\pi i/N}$. Then for any $j \in \{2, 3, ..., N\}$,

$$\begin{aligned} &\left|\textstyle\sum_{k=1}^m \mathbf{f}_j[k]\right|^2 = \left|\textstyle\sum_{k=1}^m \omega^{ja_k}\right|^2 = \textstyle\sum_{k,\ell=1}^m \omega^{j(a_k-a_\ell)} \\ &= m + \textstyle\sum_{k\neq\ell} \omega^{j(a_k-a_\ell)} = m + \rho(\omega+\omega^2+\ldots+\omega^{M-1}) \\ &= m - \rho, \end{aligned} \tag{6}$$

where we used the that $\mathcal{K}$ is an (N, m, ρ)-difference set and that $\sum_{\ell=0}^{m-1} \omega^\ell = 0$. Setting (6) equal to $m^{2-\eta}$ yields the desired expression for η. ∎

Since the constructed matrix $\mathbf{F}_{\mathcal{K}}$ is η-StRIP-able with $\eta > 1$, the corresponding Kronecker product $\overline{\mathbf{F}_{\mathcal{K}}} \otimes \mathbf{F}_{\mathcal{K}}$ is η-StRIP-able with $\eta > 1/2$ (cf. Corollary 4). Consequently $\overline{\mathbf{F}_{\mathcal{K}}} \otimes \mathbf{F}_{\mathcal{K}}$ has UStRIP according to Theorem 2, i.e. we have the following statement.

Corollary 6: *Let $\mathbf{F}_{\mathcal{K}} \in \mathbb{C}^{m\times N}$ be a matrix constructed as in Proposition 1 and let $\mathbf{C}_{\mathcal{K}} = \overline{\mathbf{F}_{\mathcal{K}}} \otimes \mathbf{F}_{\mathcal{K}}$. If $k < 1 + (N^2 - 1)\,\delta$ and if $m^2 \geq c\,(2k\log N)/\delta^2$ for some constant $c > 0$ then $\mathbf{C}_{\mathcal{K}}$ has $(k, \delta, 2\epsilon)$-UStRIP with ϵ given by* (5).

We remark again, that Corollary 6 implies in particular a statistical recovery guarantee (in the sense of Def. 2) for a Kronecker structured measurement matrix and where the number of measurements $\widetilde{m} = m^2$ scales *linearly* with the sparsity k.

5. NUMERICAL EXPERIMENTS

Finally, we present numerical experiments showing the effectiveness of the proposed measurement matrices. Before comparing the recovery performance of Kronecker product matrices $\mathbf{C} = (\overline{\mathbf{A}} \otimes \mathbf{A})$, we first check the performance of matrices $\mathbf{C}$ without Kronecker structure. To this end, we consider the following matrices $\mathbf{C}$, all of them having $m = 50$ rows and $N = 2451$ columns:

(i) *EHF*: $\mathbf{C} = \mathbf{F}_{\mathcal{K}}$ is the matrix constructed according to Prop. 1.

(ii) *deterministic partial Fourier*: the columns of $\mathbf{C}$ coincide with the first m columns of the $N \times N$ DFT matrix.

(iii) *random partial Fourier*: the rows of $\mathbf{C}$ are randomly chosen from the $N \times N$ DFT matrix.

(iv) *random Gaussian*: the entries of $\mathbf{C}$ are i.i.d normal distributed random variables.

Fig. 1 shows the corresponding simulation result for recovering a k-sparse vector $\boldsymbol{x} \in \mathbb{C}^N$ from linear measurements $\boldsymbol{y} = \mathbf{C}\,\boldsymbol{x}$, using basis pursuit (3), i.e. by solving

$$\widehat{\boldsymbol{x}} = \arg\min \|\boldsymbol{z}\|_1 \quad \text{subject to} \quad \mathbf{C}\boldsymbol{z} = \boldsymbol{y}\,, \boldsymbol{z} \in \mathbb{C}^N\,. \tag{7}$$

For these simulations, we varied the sparsity k of the data vector $\boldsymbol{x}$. On the horizontal axis we plot the normalized ℓ_2 reconstruction error $\|\widehat{\boldsymbol{x}}-\boldsymbol{x}\|^2 \,/\, \|\boldsymbol{x}\|^2$. For each k we generated 100 random k-sparse vectors $\boldsymbol{x}$ and averaged the reconstruction error over these 100 experiments. The simulation result for the deterministic partial Fourier matrix in Fig. 1 shows that not every choice of rows from the DFT matrix yields a good CS matrix. However, for a choice of rows that corresponds to an EHF the resulting measurement matrix is essentially as good as random Gaussian and random partial Fourier matrices, which are known to be good CS matrices.

In Fig 2, we compare the recovery performance of Kronecker product matrices $\mathbf{C} = (\overline{\mathbf{A}} \otimes \mathbf{A})$ for matrices $\mathbf{A} \in \mathbb{C}^{m\times N}$ as under (i)-(iv), denoted respectively by (i')-(iv'). Additionally, we consider random matrices $\mathbf{C}$ of size $m^2 \times N^2$ (without Kronecker product):

(v') *random partial Fourier*: the rows of $\mathbf{C}$ are randomly chosen from the $N^2 \times N^2$ DFT matrix.

(vi') *random Gaussian*: the entries of $\mathbf{C}$ are i.i.d normal distributed random variables.

For these simulations, we fixed $m = 10$ and $N = 91$, and the results where averaged over 100 random vectors $\boldsymbol{x}$. We observe that the Kronecker structure destroys the good behavior of the random Gaussian matrix which now performs worse. On the other side, we see that the Kronecker structured EHF matrix performs almost as good as the non-Kronecker-structured random partial Fourier and random Gaussian matrices. So for our *deterministic* EHF matrix the Kronecker structure does not harm its good CS properties. A similar behavior is observed for the random partial Fourier matrix.

6. REFERENCES

[1] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digit. Signal Process.*, vol. 8, no. 3, pp. 185–210, Jul. 1998.

[2] D. Romero, D. D. Ariananda, Z. Tian, and G. Leus, "Compressive covariance sensing: Structure-based compressive sensing beyond sparsity," *IEEE Signal Process. Mag.*, vol. 33, no. 1, pp. 78–93, Jan. 2016.

[3] G. Dasarathy, P. Shah, B. N. Bhaskar, and R. D. Nowak, "Sketching sparse matrices, covariances, and graphs via tensor products," *IEEE Trans. Inf. Theory*, vol. 61, no. 3, pp. 1373–1388, Mar. 2015.

[4] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Basel: Birkhäuser, 2013.

[5] M. Elad, *Sparse and Redundant Representations*. New York, USA: Springer, 2010.

[6] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[8] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 494–504, Feb. 2012.

[9] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 358–374, Apr. 2010.

[10] L. Welch, "Lower bounds on the maximum cross correlation of signals," *IEEE Trans. Inf. Theory*, vol. 20, no. 3, pp. 397–399, May 1974.

[11] P. G. Casazza and G. Kutyniok, *Finite Frames: Theory and Applications*. Birkhuser Basel, 2012.

[12] D. R. Stinson, *Combinatorial Designs: Construction and Analysis*. Springer, 2004.

[13] J. Singer, "A theorem in finite projective geometry and some applications to number theory," *Trans. Amer. Math. Soc.*, vol. 43, pp. 377–385, 1938.