

On the Restricted Isometry of the Columnwise Khatri-Rao Product

Saurabh Khanna, *Student Member, IEEE* and Chandra R. Murthy, *Senior Member, IEEE*

Dept. of ECE, Indian Institute of Science, Bangalore, India
 {saurabh, cmurthy}@iisc.ac.in

Abstract—The columnwise Khatri-Rao product of two matrices is an important matrix type, reprising its role as a structured sensing matrix in many fundamental linear inverse problems. Robust signal recovery in such inverse problems is often contingent on proving the restricted isometry property (RIP) of a certain system matrix expressible as a Khatri-Rao product of two matrices. In this work, we analyze the RIP of a generic columnwise Khatri-Rao product by deriving two upper bounds for its k^{th} order Restricted Isometry Constant (k -RIC) for different values of k . The first RIC bound is computed in terms of the individual RICs of the input matrices participating in the Khatri-Rao product. The second RIC bound is probabilistic in nature, and is specified in terms of the input matrix dimensions. We show that the Khatri-Rao product of a pair of $m \times n$ sized random matrices comprising independent and identically distributed subgaussian entries satisfies k -RIP with arbitrarily high probability, provided m exceeds $\mathcal{O}(\sqrt{k} \log^{3/2} n)$. This is a substantially milder condition compared to $\mathcal{O}(k \log n)$ rows needed to guarantee k -RIP of the input subgaussian random matrices participating in the Khatri-Rao product. Our results confirm that the Khatri-Rao product exhibits stronger restricted isometry compared to its constituent matrices for the same RIP order. The proposed RIC bounds are potentially useful in obtaining improved performance guarantees in several sparse signal recovery and tensor decomposition problems.

Index Terms—Khatri-Rao product, Kronecker product, compressive sensing, Restricted isometry property, covariance matrix estimation, multiple measurement vectors, PARAFAC, CANDECOMP, tensor decomposition, direction of arrival estimation.

I. INTRODUCTION

The Khatri-Rao product, denoted by the symbol \odot , is a columnwise Kronecker product, which was originally introduced by Khatri and Rao in [1]. For any two matrices $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2 \dots, \mathbf{a}_p]$ and $\mathbf{B} = [\mathbf{b}_1, \mathbf{b}_2 \dots, \mathbf{b}_p]$ of sizes $m \times p$ and $n \times p$, respectively, the columnwise Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ is a matrix of dimension $mn \times p$ defined as

$$\mathbf{A} \odot \mathbf{B} = [\mathbf{a}_1 \otimes \mathbf{b}_1 \quad \mathbf{a}_2 \otimes \mathbf{b}_2 \quad \dots \quad \mathbf{a}_p \otimes \mathbf{b}_p], \quad (1)$$

where $\mathbf{a} \otimes \mathbf{b}$ denotes the Kronecker product [2] between vectors \mathbf{a} and \mathbf{b} . That is, each column of $\mathbf{A} \odot \mathbf{B}$ is the Kronecker product between the respective columns of the two input matrices \mathbf{A} and \mathbf{B} . In this article, we shall refer to the columnwise Khatri-Rao product as simply the Khatri-Rao product or the KR product. Since the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ comprises all pairwise Kronecker product combinations of the columns of the input matrices, it can be shown that $\mathbf{A} \odot \mathbf{B} = (\mathbf{A} \otimes \mathbf{B})\mathbf{J}$, where \mathbf{J} is a $p^2 \times p$ selection matrix with columns as a subset of the standard basis in \mathbb{R}^{p^2} [3].

Khatri-Rao product matrices are encountered in several linear inverse problems of fundamental importance. Recent

examples include compressive sensing [4], [5], covariance matrix estimation [6], [7], direction of arrival estimation [8] and tensor decomposition [9]. In each of these examples, the KR product $\mathbf{A} \odot \mathbf{B}$, for certain $m \times n$ sized system matrices \mathbf{A} and \mathbf{B} , plays the role of the sensing matrix used to generate linear measurements \mathbf{y} of an unknown signal vector \mathbf{x} according to

$$\mathbf{y} = (\mathbf{A} \odot \mathbf{B}) \mathbf{x} + \mathbf{w}, \quad (2)$$

where \mathbf{w} represents the additive measurement noise. It is now well established in the sparse signal recovery literature [10]–[12] that, if the signal of interest, \mathbf{x} , is a k -sparse¹ vector in \mathbb{R}^n , it can be stably recovered from its noisy underdetermined linear observations $\mathbf{y} \in \mathbb{R}^{m^2}$ ($m^2 < n$) in a computationally efficient manner provided that the sensing matrix (here, $\mathbf{A} \odot \mathbf{B}$) satisfies the restricted isometry property defined next.

A matrix $\Phi \in \mathbb{R}^{m^2 \times n}$ is said to satisfy the *Restricted Isometry Property* (RIP) [13] of order k , if there exists a constant $\delta_k(\Phi) \in (0, 1)$, such that for all k -sparse vectors $\mathbf{z} \in \mathbb{R}^n$,

$$(1 - \delta_k(\Phi)) \|\mathbf{z}\|_2^2 \leq \|\Phi \mathbf{z}\|_2^2 \leq (1 + \delta_k(\Phi)) \|\mathbf{z}\|_2^2. \quad (3)$$

The smallest constant $\delta_k(\Phi)$ for which (3) holds for all k -sparse \mathbf{z} is called the k^{th} order restricted isometry constant or the k -RIC of Φ . Matrices with small k -RICs are good encoders for storing/sketching high dimensional vectors with k or fewer nonzero entries [14]. For example, $\delta_k(\mathbf{A} \odot \mathbf{B}) < 0.307$ is a sufficient condition for a unique k -sparse solution to (2) in the noiseless case, and its perfect recovery via the ℓ_1 minimization technique [15]. As pointed out earlier, in many structured signal recovery problems, the main sensing matrix can be expressed as a columnwise Khatri-Rao product between two matrices. Thus, from a practitioner’s viewpoint, it is pertinent to study the restricted isometry property of a columnwise Khatri-Rao product matrix, which is the focus of this work.

A. Applications involving Khatri-Rao matrices

We briefly describe some examples where it is required to show the restricted isometry property of a KR product matrix.

1) **Support recovery of joint sparse vectors from underdetermined linear measurements:** Suppose $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_L$ are unknown joint sparse signals in \mathbb{R}^n with a common k -sized support denoted by an index set \mathcal{S} . A canonical problem in multi-sensor signal processing is concerned with the recovery of the common support \mathcal{S} of the unknown signals from their

¹A vector is said to be k -sparse if at most k of its entries are nonzero.

noisy underdetermined linear measurements $\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_L \in \mathbb{R}^m$ generated according to

$$\mathbf{y}_j = \mathbf{A}\mathbf{x}_j + \mathbf{w}_j, \quad 1 \leq j \leq L, \quad (4)$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ ($m < n$) is a known measurement matrix, and $\mathbf{w}_j \in \mathbb{R}^m$ models the noise in the measurements. This problem arises in many practical applications such as MIMO channel estimation, cooperative wideband spectrum sensing in cognitive radio networks, target localization, and direction of arrival estimation. In [16], the support set \mathcal{S} is recovered as the support of $\hat{\gamma}$, the solution to the Co-LASSO problem:

$$\text{Co-LASSO: } \min_{\gamma \geq 0} \left\| \text{vec}(\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}}) - (\mathbf{A} \odot \mathbf{A})\gamma \right\|_2^2 + \lambda \|\gamma\|_1, \quad (5)$$

where $\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}} \triangleq \frac{1}{L} \sum_{j=1}^L \mathbf{y}_j \mathbf{y}_j^T$. From compressive sensing theory [11], the RIP of $\mathbf{A} \odot \mathbf{A}$ (also called the self Khatri-Rao product of \mathbf{A}) determines the stability of the sparse solution in the Co-LASSO problem. In M-SBL [17], a different support recovery algorithm, $\mathbf{A} \odot \mathbf{A}$ satisfying $2k$ -RIP can guarantee exact recovery of \mathcal{S} from multiple measurements [18].

2) *Vandermonde decomposition of Toeplitz matrices:*

According to a classical result by Carathéodory and Fejér [19], any $n \times n$ positive semidefinite Toeplitz matrix \mathbf{T} of rank $r < n$ admits the following decomposition

$$\mathbf{T} = \mathbf{A}\mathbf{P}\mathbf{A}^T, \quad (6)$$

where \mathbf{P} is an $n \times n$ positive semidefinite diagonal matrix with an r -sparse diagonal, and \mathbf{A} is an $n \times n$ Vandermonde matrix with uniformly sampled complex sinusoids of different frequencies as its columns. This Toeplitz decomposition underpins subspace based spectrum estimation methods such as Multiple Signal Classification (MuSiC) and Estimation of Parameters by Rotationally Invariant Techniques (ESPRIT) [20], [21]. By replacing \mathbf{T} with a data covariance matrix, the r -sparse support of $\text{diag}(\mathbf{P})$ in (6) corresponds to the r -dimensional signal subspace of the data. Estimation of $\mathbf{p} \triangleq \text{diag}(\mathbf{P})$ is tantamount to finding an r -sparse solution to the vectorized form of (6), i.e., $\text{vec}(\mathbf{T}) = (\mathbf{A} \odot \mathbf{A})\mathbf{p}$. Here again, the recovery of a unique r -sparse solution for \mathbf{p} can be guaranteed if $\mathbf{A} \odot \mathbf{A}$ satisfies the RIP of order $2r$.

3) *PARAFAC model for low-rank three-way arrays:*

Consider an $I \times J \times K$ tensor $\underline{\mathbf{X}}$ of rank r . We can express $\underline{\mathbf{X}}$ as the sum of r rank-one three way arrays as $\underline{\mathbf{X}} = \sum_{i=1}^r \mathbf{a}_i \circ \mathbf{b}_i \circ \mathbf{c}_i$, where $\mathbf{a}_i, \mathbf{b}_i, \mathbf{c}_i$ are loading vectors of dimension I, J, K , respectively, and \circ denotes the vector outer product. The tensor $\underline{\mathbf{X}}$ itself can be arranged into a matrix as $\mathbf{X} = [\text{vec}(\mathbf{X}_1), \text{vec}(\mathbf{X}_2), \dots, \text{vec}(\mathbf{X}_K)]$. In the parallel factor analysis (PARAFAC) model [22], the matrix \mathbf{X} can be approximated as

$$\mathbf{X} \approx (\mathbf{A} \odot \mathbf{B})\mathbf{C}^T, \quad (7)$$

where \mathbf{A}, \mathbf{B} and \mathbf{C} are the loading matrices with columns as the loading vectors $\mathbf{a}_i, \mathbf{b}_i$ and \mathbf{c}_i , respectively. In many problems such as direction of arrival estimation using a 2D-antenna array, the loading matrix \mathbf{C} turns out to be row-sparse matrix [23]. In such cases, the uniqueness of the PARAFAC model shown in (7) depends on the restricted isometry property of the Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$.

Finding the exact k th order RIC of any matrix entails searching for the smallest and largest eigenvalues among all possible k -column submatrices of the matrix, which is, in general, an NP hard task [24]. In this work, we follow an alternative approach to analyzing the RIP of a KR product matrix. We seek to derive tight upper bounds for its RICs.

B. Related Work

Perhaps the most straightforward way to analyze the RICs of the KR product matrix is to use the eigenvalue interlacing theorem [25], which relates the singular values of any k -column submatrix of the KR product to the singular values of the Kronecker product. This is possible because any k columns of the KR product can together be interpreted as a submatrix of the Kronecker product. However, barring the maximum and minimum singular values of the Kronecker product, there is no explicit characterization of its non-extremal singular values available, that can be used to obtain tight bounds the k -RIC of the KR product. Bounding the RIC using the extreme singular values of the Kronecker product matrix turns out to be too loose to be useful. In this context, we note that an upper bound for the k -RIC of the Kronecker product matrix is derived in terms of the k -RICs of the input matrices in [4], [26]. However, the k -RIC of the KR product matrix is yet to be analyzed.

Recently, [27], [28] gave probabilistic lower bounds for the minimum singular value of the columnwise KR product between two or more matrices. These bounds are limited to randomly constructed input matrices, and are polynomial in the matrix size. In [29], it is shown that for any two matrices \mathbf{A} and \mathbf{B} , the Kruskal-rank² of $\mathbf{A} \odot \mathbf{B}$ has a lower bound in terms of $\text{K-rank}(\mathbf{A})$ and $\text{K-rank}(\mathbf{B})$. In fact, $\text{Krank}(\mathbf{A} \odot \mathbf{B})$ is at least as high as $\max(\text{K-rank}(\mathbf{A}), \text{K-rank}(\mathbf{B}))$, thereby suggesting that $\mathbf{A} \odot \mathbf{B}$ satisfies a stronger restricted isometry property than both \mathbf{A} and \mathbf{B} . The RIC bounds presented in this work ratify this fact.

A closely related yet weaker notion of restricted isometry constant is the τ -robust K-rank, denoted by K-rank_τ . For a given matrix Φ , the $\text{K-rank}_\tau(\Phi)$ is defined as the largest k for which every $n \times k$ submatrix of Φ has its smallest singular value larger than $1/\tau$. In [27], it is shown that the τ -robust K-rank is super-additive, implying that the K-rank_τ of the Khatri-Rao product is strictly larger than individual K-rank_τ s of the input matrices. We show a similar result for the restricted isometry constants of the KR product matrix.

Our work is perhaps closest to [7], which provides a polynomially tight probabilistic upper bound for the k -RIC (defined using ℓ_1 -norm in (3)) of the Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$, when the input matrices \mathbf{A} and \mathbf{B} are the adjacency matrices of two independent uniformly random δ -left regular bipartite graphs. This work instead assumes \mathbf{A} and \mathbf{B} to be random matrices with independent subgaussian elements.

C. Our Contributions

We derive two upper bounds on the k -RIC of the columnwise KR product of two $m \times n$ sized matrices \mathbf{A} and \mathbf{B} . The

²The Kruskal rank of any matrix \mathbf{A} is the largest integer r such that any r columns of \mathbf{A} are linearly independent.

bounds are listed below.

- 1) A deterministic upper bound for the k -RIC of $\mathbf{A} \odot \mathbf{B}$ in terms of the k -RICs of the input matrices \mathbf{A} and \mathbf{B} . The bound is valid for $k \leq m$, and for input matrices with unit ℓ_2 -norm columns.
- 2a) A probabilistic upper bound for the k -RIC of $\mathbf{A} \odot \mathbf{B}$ in terms of k and the input matrix dimensions (m, n) , for \mathbf{A}, \mathbf{B} as random matrices with i.i.d. subgaussian elements. The probabilistic bound is polynomially tight with respect to the input matrix dimension n . The bound is valid for $k \leq n$.
- 2b) A probabilistic upper bound for the k -RIC of the self KR product $\mathbf{A} \odot \mathbf{A}$ in terms of m, n , and k , for \mathbf{A} as a random matrix with i.i.d. subgaussian elements. Although the RIC bounds for the self KR product and the general KR product with distinct input matrices are of similar form, the derivation of former RIC bound is slightly more intricate as it involves showing sharp concentration for functions of dependent random variables.

A key idea in the RIC analysis is to use the fact (stated formally as Proposition 12) that for any two matrices \mathbf{A} and \mathbf{B} , the Gram matrix of their KR product $(\mathbf{A} \odot \mathbf{B})^T (\mathbf{A} \odot \mathbf{B})$ can be interpreted as the Hadamard product (element wise multiplication) between $\mathbf{A}^T \mathbf{A}$ and $\mathbf{B}^T \mathbf{B}$. The Hadamard product form is more analytically tractable than columnwise Kronecker product form of the KR matrix.

Lately, in several machine learning problems, the necessary and sufficient conditions for successful signal recovery have been reported in terms of the RICs of a certain Khatri-Rao product matrix serving as a *pseudo* sensing matrix [8], [16]. In light of this, our proposed RIC bounds are quite timely, and pave the way towards obtaining order-wise tight sample complexity bounds in several fundamental learning problems.

The rest of this article is organized as follows. In Secs. II and III, we present our main results: deterministic and probabilistic RIC bounds, respectively, for a generic columnwise KR product matrix. Sec. III also discusses about the RIP of the *self* Khatri-Rao product of a matrix with itself, an important matrix type encountered in the sparse diagonal covariance matrix estimation problem. In Secs. IV and VI, we provide some background concepts needed in proving the proposed RIC bounds. Secs. V and VII provide the detailed proofs of the deterministic and probabilistic RIC bounds, respectively. Final conclusions are presented in Sec. VIII.

Notation: In this work, bold lowercase letters are used for representing both scalar random variables as well as vectors. Bold uppercase letters are reserved for matrices. The ℓ_2 -norm of vector \mathbf{x} is denoted by $\|\mathbf{x}\|_2$. For an $m \times n$ matrix \mathbf{A} , $\|\mathbf{A}\|$ denotes its operator norm, $\|\mathbf{A}\| = \sup_{\mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq 0} \frac{\|\mathbf{A}\mathbf{x}\|_2}{\|\mathbf{x}\|_2}$. The Hilbert-Schmidt (or Frobenius) norm of \mathbf{A} is defined as $\|\mathbf{A}\|_{HS} = \sum_{i=1}^m \sum_{j=1}^n |A_{i,j}|^2$. The symbol $[n]$ denotes the index set $\{1, 2, \dots, n\}$. For any index set $\mathcal{S} \subseteq [n]$, $\mathbf{A}_{\mathcal{S}}$ denotes the submatrix comprising the columns of \mathbf{A} indexed by \mathcal{S} . The matrices $\mathbf{A} \otimes \mathbf{B}$, $\mathbf{A} \circ \mathbf{B}$ and $\mathbf{A} \odot \mathbf{B}$ denote the Kronecker product, Hadamard product and columnwise Khatri-Rao product of \mathbf{A} and \mathbf{B} , respectively. $\mathbf{A} \leq \mathbf{B}$ implies that $\mathbf{B} - \mathbf{A}$ is a positive semidefinite matrix. \mathbf{A}^T , \mathbf{A}^H , \mathbf{A}^{-1} ,

and \mathbf{A}^\dagger denote the transpose, conjugate-transpose, inverse and generalized matrix inverse operations, respectively.

II. DETERMINISTIC k -RIC BOUND

In this section, we present our first upper bound on the k -RIC of a generic columnwise KR product $\mathbf{A} \odot \mathbf{B}$, for any two similar sized matrices \mathbf{A} and \mathbf{B} with normalized columns. The bound is given in terms of the k -RICs of \mathbf{A} and \mathbf{B} .

Theorem 1. *Let \mathbf{A} and \mathbf{B} be $m \times n$ sized real-valued matrices with unit ℓ_2 -norm columns and satisfying the k^{th} order restricted isometry property with constants $\delta_k^{\mathbf{A}}$ and $\delta_k^{\mathbf{B}}$, respectively. Then, their columnwise Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ satisfies the restricted isometry property with k -RIC at most δ^2 , where $\delta \triangleq \max(\delta_k^{\mathbf{A}}, \delta_k^{\mathbf{B}})$, i.e.,*

$$(1 - \delta^2) \|\mathbf{z}\|_2^2 \leq \|(\mathbf{A} \odot \mathbf{B})\mathbf{z}\|_2^2 \leq (1 + \delta^2) \|\mathbf{z}\|_2^2 \quad (8)$$

holds for all k -sparse vectors $\mathbf{z} \in \mathbb{R}^n$.

Proof. The proof is provided in Section V. \square

Remark 1: The RIC bound for $\mathbf{A} \odot \mathbf{B}$ in Theorem 1 is relevant only when $\delta_k(\mathbf{A})$ and $\delta_k(\mathbf{B})$ lie in $(0, 1)$, which is true only for $k \leq m$. In other words, the above k -RIC characterization for $\mathbf{A} \odot \mathbf{B}$ requires the input matrices \mathbf{A} and \mathbf{B} to be k -RIP compliant.

Remark 2: Since the input matrices \mathbf{A} and \mathbf{B} satisfy k -RIP with $\delta_k(\mathbf{A}), \delta_k(\mathbf{B}) \in (0, 1)$, it follows from Theorem 1 that $\delta_k(\mathbf{A} \odot \mathbf{B})$ is strictly smaller than $\max(\delta_k(\mathbf{A}), \delta_k(\mathbf{B}))$. If $\mathbf{B} = \mathbf{A}$, the special case of self Khatri-Rao product $\mathbf{A} \odot \mathbf{A}$ arises, for which

$$\delta_k(\mathbf{A} \odot \mathbf{A}) < \delta_k^2(\mathbf{A}). \quad (9)$$

Above implies that the self Khatri-Rao product $\mathbf{A} \odot \mathbf{A}$ is a better restricted isometry compared to \mathbf{A} itself. This observation is in alignment with the expanding Kruskal rank and shrinking mutual coherence of the self Khatri-Rao product reported in [16]. In fact, for $k = 2$, the 2-RIC bound (9) exactly matches the mutual coherence bound shown in [16].

For $k \in (m, m^2]$, using ([30], Theorem 1), one can show that $\delta_k(\mathbf{A} \odot \mathbf{B}) \leq (\sqrt{k} + 1) \delta_{\sqrt{k}}$, where $\delta_{\sqrt{k}} = \max(\delta_{\sqrt{k}}(\mathbf{A}), \delta_{\sqrt{k}}(\mathbf{B}))$. This bound, however, loses its tightness and quickly becomes unattractive for larger k values of k . Finding a tighter k -RIC upper bound for the $k > m$ case remains an open problem.

To gauge the tightness of the proposed k -RIC bound for $\mathbf{A} \odot \mathbf{B}$, we present its simulation-based quantification for the case when the input matrices \mathbf{A} and \mathbf{B} are random Gaussian matrices with i.i.d. $\mathcal{N}(0, 1/m)$ entries. Fig. 1 plots $\delta_k(\mathbf{A}), \delta_k(\mathbf{B}), \delta_k(\mathbf{A} \odot \mathbf{B})$ and the upper bound $\overline{\delta_k(\mathbf{A} \odot \mathbf{B})} = (\max(\delta_k(\mathbf{A}), \delta_k(\mathbf{B})))^2$ for a range of input matrix dimension m . The aspect ratio m/n of the input matrices is fixed to 0.5.³ For computational tractability, we restrict our analysis to the cases $k = 2$ and 3. The RICs: $\delta_k(\mathbf{A}), \delta_k(\mathbf{B})$ and

³ While the $m \times n$ matrices \mathbf{A} and \mathbf{B} may represent highly underdetermined linear systems (when $m \ll n$), their $m^2 \times n$ sized Khatri-Rao product $\mathbf{A} \odot \mathbf{B}$ can become an overdetermined system. In fact, many covariance matching based sparse support recovery algorithms [16], [17], [31] exploit this fact to offer significantly better support reconstruction performance.

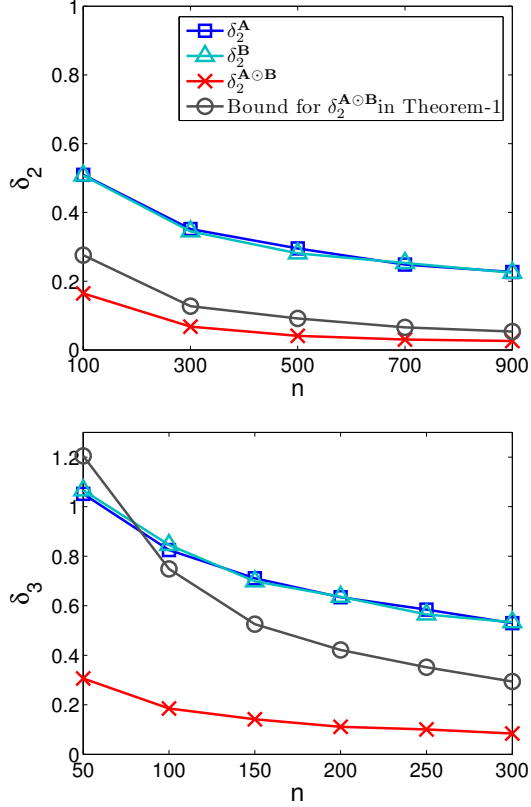


Fig. 1: Variation of k -RICs of \mathbf{A} , \mathbf{B} , $\mathbf{A} \odot \mathbf{B}$ and the proposed upper bound with increasing input matrix dimensions. The top and the bottom plots are for $k = 2$ and 3, respectively. Each data point is averaged over 10 trials.

$\delta_k(\mathbf{A} \odot \mathbf{B})$ are computed by exhaustively searching for the worst conditioned submatrix comprising k columns of \mathbf{A} , \mathbf{B} and $\mathbf{A} \odot \mathbf{B}$, respectively. From Fig. 1, we observe that the proposed k -RIC upper bound becomes tighter as the input matrices grow in size.

III. PROBABILISTIC k -RIC BOUND

The deterministic RIC bound for the columnwise KR product discussed in the previous section is only applicable when $k \leq m$. It also fails to explain the empirical observation that the KR product $\mathbf{A} \odot \mathbf{B}$ often satisfies k -RIP in spite of the input matrices \mathbf{A} and \mathbf{B} failing to do so. These concerns are addressed by our second RIC bound for the columnwise KR product. This second RIC bound is probabilistic in nature and is applicable to the KR product of random input matrices with i.i.d. subgaussian entries. Below, we define a subgaussian random variable and state some of its properties.

Definition 1. (*Subgaussian Random Variable*): A random variable \mathbf{x} is called subgaussian, if its tail probability is dominated by that of a Gaussian random variable. In other words, there exists a constant $K > 0$ such that $\mathbb{P}(|\mathbf{x}| \geq t) \leq e^{-t^2/K^2}$.

Gaussian, Bernoulli and all bounded random variables are subgaussian random variables. For a subgaussian random variable, its p^{th} order moment grows only as fast as $\mathcal{O}(p^{p/2})$ [32]. In other words, there exists $K_1 > 0$ such that

$$(\mathbb{E} |\mathbf{x}|^p)^{\frac{1}{p}} \leq K_1 \sqrt{p}, \quad p \geq 1. \quad (10)$$

The minimum such K_1 is called the *subgaussian* or ψ_2 norm of the random variable \mathbf{x} , i.e.,

$$\|\mathbf{x}\|_{\psi_2} = \sup_{p \geq 1} p^{-1/2} (\mathbb{E} |\mathbf{x}|^p)^{\frac{1}{p}}. \quad (11)$$

Given a pair of random input matrices with i.i.d. subgaussian entries, Theorem 2 presents a new upper bound on the k -RIC of their columnwise KR product.

Theorem 2. Suppose \mathbf{A} and \mathbf{B} are $m \times n$ random matrices with real i.i.d. subgaussian entries, such that $\mathbb{E} \mathbf{A}_{ij} = 0$ and $\mathbb{E} \mathbf{A}_{ij}^2 = 1$, and $\|\mathbf{A}_{ij}\|_{\psi_2} \leq K$, and similarly for \mathbf{B} . Let $K_o \triangleq \max(K, 1)$. Then, the k^{th} order restricted isometry constant of $\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}$, denoted by δ_k , satisfies $\delta_k \leq \delta$ with probability at least $1 - 8en^{-2(c_2\gamma-1)}$ for all $\gamma > 1/c_2$, and c_2 a universal positive constant, provided that

$$m \geq \left(\frac{16\sqrt{2\pi}\gamma^{3/2}K_o^2}{\delta} \right) \sqrt{k} (\log n)^{3/2}.$$

Proof. The proof is provided in Section VII. \square

The normalization constant \sqrt{m} used while computing the KR product $\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}$ ensures that the columns of the input matrices $\frac{\mathbf{A}}{\sqrt{m}}$, $\frac{\mathbf{B}}{\sqrt{m}}$ have unit average energy, i.e. $\mathbb{E} \|\mathbf{a}_i/\sqrt{m}\|_2 = \mathbb{E} \|\mathbf{b}_i/\sqrt{m}\|_2 = 1$ for $1 \leq i \leq n$. Column normalization is a key assumption towards correct modelling of the isotropic, norm-preserving nature of the effective sensing matrix $\frac{1}{m}(\mathbf{A} \odot \mathbf{B})$, an attribute found in most sensing matrices employed in practice.

Theorem 2 implies that

$$\delta_k \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}} \right) \leq 16\sqrt{2\pi}\gamma^{3/2}K_o^2 \left(\frac{\sqrt{k} (\log n)^{3/2}}{m} \right) \quad (12)$$

with probability exceeding $1 - 8en^{-2(c_2\gamma-1)}$. Thus, the above k -RIC bound decreases as m increases, which is intuitively appealing. Interestingly, for fixed k and n , the above k -RIC upper bound for $\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}$ decays as $\mathcal{O}(\frac{1}{m})$. This is a significant improvement over the $\mathcal{O}(\frac{1}{\sqrt{m}})$ decay rate [11] already known for the individual k -RICs of the input subgaussian matrices $\frac{\mathbf{A}}{\sqrt{m}}$ and $\frac{\mathbf{B}}{\sqrt{m}}$. Thus, for any m , the Khatri-Rao product $\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}}$ exhibits stronger restricted isometry property, with significantly smaller k -RICs in comparison to the k -RICs for the input matrices.

In some applications, the effective sensing matrix can be expressed as the self-Khatri Rao product $\frac{1}{m}(\mathbf{A} \odot \mathbf{A})$ of a certain column normalized system matrix $\frac{\mathbf{A}}{\sqrt{m}}$ with itself [16]. In Theorem 3 below, we present the k -RIC bound for the special case of self-Khatri-Rao product matrices.

Theorem 3. Let \mathbf{A} be an $m \times n$ random matrix with real i.i.d. subgaussian entries, such that $\mathbb{E} \mathbf{A}_{ij} = 0$ and $\mathbb{E} \mathbf{A}_{ij}^2 = 1$, and $\|\mathbf{A}_{ij}\|_{\psi_2} \leq K$. Then, the k^{th} order restricted isometry constant of the column normalized self Khatri-Rao product $\frac{1}{m}(\mathbf{A} \odot \mathbf{A})$ satisfies $\delta_k \leq \delta$ with probability at least $1 - 4en^{-2(c_2\gamma-1)}$ for any $\gamma \geq 1/c_2$, provided

$$m \geq \left(\frac{16\sqrt{2\pi}\gamma^{3/2}K_o^2}{\delta} \right) \sqrt{k} (\log n)^{3/2}.$$

Here, $K_o \triangleq \max(K, 1)$ and $c_2 > 0$ is a universal constant.

Proof. A proof sketch is provided in Appendix F. \square

Theorem 3 implies that

$$\delta_k \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}} \right) \leq 16\sqrt{2\pi}\gamma^{3/2}K_o^2 \left(\frac{\sqrt{k}(\log n)^{3/2}}{m} \right) \quad (13)$$

with probability exceeding $1 - 4en^{-2(c_2\gamma-1)}$. The above k -RIC bound for the self Khatri-Rao product scales with m, n , and k in a similar fashion as the asymmetric Khatri-Rao product.

Remark 3: From [11], for an $m \times n$ matrix \mathbf{A} with i.i.d. subgaussian entries, $\delta_k(\frac{\mathbf{A}}{\sqrt{m}}) \leq \sqrt{\frac{2Ck \log(eN/k)}{m}}$ with high probability. Now, even though the columns of \mathbf{A} do not exactly have unit norm, directly using this in Theorem 1 for the sake of comparison results in $\delta_k(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}}) \leq \frac{2Ck \log(eN/k)}{m}$, when $k \leq m$. Our probabilistic bound in (13) is therefore tighter than the deterministic bound in Theorem 1 by a multiplicative factor of $\mathcal{O}(\sqrt{k})$. Note that, the deterministic bound is valid for any pair of input matrices with normalized columns, while the probabilistic bound holds when the input matrices have i.i.d. subgaussian entries.

Remark 4: In covariance matching based signal support recovery algorithms like Co-LASSO [16] and M-SBL [17], given a system matrix $\frac{\mathbf{A}}{\sqrt{m}}$, $\delta_{2k}(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}}) \leq 1$ is one of the sufficient conditions for exact recovery of the true k -sparse support of the unknown signal of interest [18]. This sufficiency condition is met with arbitrarily high probability according to Theorem 3, when the subgaussian matrix \mathbf{A} has at least $m = \mathcal{O}(\sqrt{k} \log^{3/2} n)$ rows. This is a significantly milder condition when compared with $\mathcal{O}(k \log n)$ rows required for exact k -sparse support recovery by conventional compressive sensing algorithms such as SOMP [33]. In fact, to the best of our knowledge, our k -RIC bound for the self Khatri-Rao product provides the first ever theoretical confirmation for the empirically observed performance of covariance matching based sparse support recovery algorithms, i.e., the number of measurements m need to scale as only $\mathcal{O}(\sqrt{k})$ rather than $\mathcal{O}(k)$ in the case of conventional support recovery algorithms.

IV. PRELIMINARIES FOR DETERMINISTIC k -RIC BOUND

In this section, we present some preliminary concepts and results which are necessary for the derivation of the deterministic k -RIC bound in Theorem 1. For the sake of brevity, we provide proofs only for claims that have not been explicitly shown in their cited sources.

A. Properties of the Kronecker and Hadamard product

For any two matrices \mathbf{A} and \mathbf{B} of dimensions $m \times n$ and $p \times q$, the kronecker product $\mathbf{A} \otimes \mathbf{B}$ is the $mp \times nq$ matrix

$$\mathbf{A} \otimes \mathbf{B} = \begin{pmatrix} a_{11}\mathbf{B} & a_{12}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ a_{21}\mathbf{B} & a_{22}\mathbf{B} & \cdots & a_{2n}\mathbf{B} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & a_{m2}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}. \quad (14)$$

The following Proposition relates the spectral properties of the Kronecker product and its constituent matrices.

Proposition 4 (7.1.10 in [2]). *Let $\mathbf{A} \in \mathbb{R}^{n \times n}$ and $\mathbf{B} \in \mathbb{R}^{p \times p}$ admit eigenvalue decompositions $\mathbf{U}_A \Lambda_A \mathbf{U}_A^T$ and $\mathbf{U}_B \Lambda_B \mathbf{U}_B^T$, respectively. Then,*

$$(\mathbf{U}_A \otimes \mathbf{U}_B)(\Lambda_A \otimes \Lambda_B)(\mathbf{U}_A \otimes \mathbf{U}_B)^T$$

yields the eigenvalue decomposition for $\mathbf{A} \otimes \mathbf{B}$.

For any two matrices of matching dimensions, say $m \times n$, their Hadamard product $\mathbf{A} \circ \mathbf{B}$ is obtained by elementwise multiplication of the entries of the input matrices, i.e.,

$$[\mathbf{A} \circ \mathbf{B}]_{i,j} = a_{ij}b_{ij} \quad \text{for } i \in [m], j \in [n]. \quad (15)$$

The Hadamard product $\mathbf{A} \circ \mathbf{B}$ is a principal submatrix of the Kronecker product $\mathbf{A} \otimes \mathbf{B}$ [3], [34]. For $n \times n$ sized square matrices \mathbf{A} and \mathbf{B} , one can write,

$$\mathbf{A} \circ \mathbf{B} = \mathbf{J}^T (\mathbf{A} \otimes \mathbf{B}) \mathbf{J}, \quad (16)$$

where \mathbf{J} is an $n^2 \times n^2$ sized selection matrix constructed entirely from 0's and 1's which satisfies $\mathbf{J}^T \mathbf{J} = \mathbf{I}_n$.

In Proposition 5, we present an upper bound on the spectral radius of a generic Hadamard product.

Proposition 5. *For every $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$, we have*

$$\sigma_{\max}(\mathbf{A} \circ \mathbf{B}) \leq r_{\max}(\mathbf{A})c_{\max}(\mathbf{B}) \quad (17)$$

where $\sigma_{\max}(\cdot)$, $r_{\max}(\cdot)$ and $c_{\max}(\cdot)$ are the largest singular value, the largest row norm and the largest column norm of the input matrix, respectively.

Proof. See Theorem 5.5.3 in [35]. \square

We now state an important result about the Hadamard product of two positive semidefinite matrices.

Proposition 6 (Mond and Pečarić [36]). *Let \mathbf{A} and \mathbf{B} be positive semidefinite $n \times n$ Hermitian matrices and let r and s be two nonzero integers such that $s > r$. Then,*

$$(\mathbf{A}^s \circ \mathbf{B}^s)^{1/s} \geq (\mathbf{A}^r \circ \mathbf{B}^r)^{1/r}. \quad (18)$$

In Propositions 7 and 8, we state some spectral properties of correlation matrices and their Hadamard products. Correlation matrices are symmetric positive semidefinite matrices with diagonal entries equal to one. Later on, we will exploit the fact that the singular values of the columnwise KR product are related to the singular values of the Hadamard product of certain correlation matrices.

Proposition 7. *If \mathbf{A} is an $n \times n$ correlation matrix, then $\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}$ is a doubly stochastic matrix.*

Proof. See Appendix A. \square

Proposition 8 (Werner [37]). *For any correlation matrices \mathbf{A} and \mathbf{B} of the same size, we have $\mathbf{A}^{1/2} \circ \mathbf{B}^{1/2} \leq \mathbf{I}$, where $\mathbf{A}^{1/2}$ and $\mathbf{B}^{1/2}$ are the positive square roots of \mathbf{A} and \mathbf{B} , respectively.*

Proof. Since \mathbf{A} is a correlation matrix, from Proposition 7, it follows that $\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}$ is doubly stochastic. Since the rows and columns of $\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}$ sum to unity, we have $r_{\max}(\mathbf{A}^{1/2}) = c_{\max}(\mathbf{A}^{1/2}) = 1$. Similarly, $r_{\max}(\mathbf{B}^{1/2}) = c_{\max}(\mathbf{B}^{1/2}) = 1$. Then, from Proposition 5, it follows that the largest eigenvalue of $\mathbf{A}^{1/2} \circ \mathbf{B}^{1/2}$ is at most unity. \square

B. Matrix Kantorovich Inequalities

Matrix Kantorovich inequalities relate positive definite matrices by inequalities in the sense of the Löwner partial order.⁴ These inequalities can be used to extend the Löwner partial order to the Hadamard product of positive definite matrices. Our proposed RIC bound relies on the tightness of these Kantorovich inequalities and their extensions.

A matrix version of the Kantorovich inequality was first proposed by Marshall and Olkin in [38]. It is stated below as Proposition 9.

Proposition 9 (Marshall and Olkin [38]). *Let \mathbf{A} be an $n \times n$ positive definite matrix. Let \mathbf{A} admit the Schur decomposition $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$ with unitary \mathbf{U} and $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$ such that $\lambda_i \in [m, M]$. Then, we have*

$$\mathbf{A}^2 \leq (M + m)\mathbf{A} - mM\mathbf{I}_n. \quad (19)$$

The above inequality (19) is the starting point for obtaining a variety of forward and reverse Kantorovich-type matrix inequalities for positive definite matrices. In Propositions 10 and 11, we state specific forward and reverse inequalities, respectively, which are relevant to us.

Proposition 10 (Liu and Neudecker [39]). *Let \mathbf{A} be an $n \times n$ positive definite Hermitian matrix, with eigenvalues in $[m, M]$. Let \mathbf{V} be an $n \times n$ matrix such that $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. Then,*

$$\mathbf{V}^T\mathbf{A}^2\mathbf{V} - (\mathbf{V}^T\mathbf{A}\mathbf{V})^2 \leq \frac{1}{4}(M - m)^2\mathbf{I}. \quad (20)$$

Proposition 11 (Liu and Neudecker [40]). *Let \mathbf{A} and \mathbf{B} be $n \times n$ positive definite matrices. Let m and M be the minimum and maximum eigenvalues of $\mathbf{B}^{1/2}\mathbf{A}^{-1}\mathbf{B}^{1/2}$. Let \mathbf{X} be an $n \times p$ matrix with rank q ($n > p \geq q$). Then, we have*

$$(\mathbf{X}^T\mathbf{B}\mathbf{X})(\mathbf{X}^T\mathbf{A}\mathbf{X})^\dagger(\mathbf{X}^T\mathbf{B}\mathbf{X}) \geq \frac{4mM}{(M + m)^2}\mathbf{X}^T\mathbf{B}\mathbf{A}^{-1}\mathbf{B}\mathbf{X}. \quad (21)$$

Proposition 10 can be proved using (19) by pre- and post-multiplying by \mathbf{V}^T and \mathbf{V} , respectively, followed by completion of squares for the right hand side terms. The proof of Proposition 11 is given in [40].

C. Kantorovich Matrix Inequalities for the Hadamard Products of Positive Definite Matrices

Lemmas 1 and 2 stated below extend the Kantorovich inequalities from the previous subsection to Hadamard products.

Lemma 1 (Liu and Neudecker [39]). *Let \mathbf{A} and \mathbf{B} be $n \times n$ positive definite matrices, with m and M denoting the minimum and maximum eigenvalues of $\mathbf{A} \otimes \mathbf{B}$. Then, we have*

$$\mathbf{A}^2 \circ \mathbf{B}^2 \leq (\mathbf{A} \circ \mathbf{B})^2 + \frac{1}{4}(M - m)^2\mathbf{I}_n. \quad (22)$$

Proof. Let \mathbf{J} be the selection matrix such that $\mathbf{J}^T\mathbf{J} = \mathbf{I}$ and $\mathbf{A} \circ \mathbf{B} = \mathbf{J}^T(\mathbf{A} \otimes \mathbf{B})\mathbf{J}$. Then, by applying Proposition 10 with \mathbf{A} replaced with $\mathbf{A} \otimes \mathbf{B}$, and \mathbf{V} replaced with \mathbf{J} , we obtain

$$\mathbf{J}^T(\mathbf{A} \otimes \mathbf{B})^2\mathbf{J} - (\mathbf{J}^T(\mathbf{A} \otimes \mathbf{B})\mathbf{J})^2 \leq \frac{1}{4}(M - m)^2\mathbf{I}.$$

⁴The Löwner partial order here refers to the relation “ \leq ”. For positive definite matrices \mathbf{A} and \mathbf{B} , $\mathbf{A} \leq \mathbf{B}$ if and only if $\mathbf{B} - \mathbf{A}$ is a positive semi-definite matrix.

Using Fact 8.21.29 in [2], i.e., $(\mathbf{A} \otimes \mathbf{B})^2 = \mathbf{A}^2 \otimes \mathbf{B}^2$, we get

$$\mathbf{J}^T(\mathbf{A}^2 \otimes \mathbf{B}^2)\mathbf{J} - (\mathbf{A} \circ \mathbf{B})^2 \leq \frac{1}{4}(M - m)^2\mathbf{I},$$

Finally, by observing that $\mathbf{J}^T(\mathbf{A}^2 \otimes \mathbf{B}^2)\mathbf{J} = \mathbf{A}^2 \circ \mathbf{B}^2$, we obtain the desired result. \square

Lemma 2 (Liu [41]). *Let \mathbf{A} , \mathbf{B} be $n \times n$ positive definite correlation matrices. Then,*

$$\mathbf{A}^{1/2} \circ \mathbf{B}^{1/2} \geq \frac{2cd}{c^2 + d^2}\mathbf{I}_n \quad (23)$$

where the eigenvalues of $\mathbf{A}^{1/2}$ and $\mathbf{B}^{1/2}$ lie inside $[c, d]$.

Lemma 2 follows from Proposition 11, by replacing \mathbf{A} with $\mathbf{I}_n \otimes \mathbf{B}$, \mathbf{B} with $\mathbf{A}^{1/2} \otimes \mathbf{B}^{1/2}$, and \mathbf{X} with \mathbf{J} , where \mathbf{J} is the $n^2 \times n$ binary selection matrix such that $\mathbf{J}^T\mathbf{J} = \mathbf{I}$ and $\mathbf{A} \circ \mathbf{B} = \mathbf{J}^T(\mathbf{A} \otimes \mathbf{B})\mathbf{J}$.

V. PROOF OF THE DETERMINISTIC k -RIC BOUND (THEOREM 1)

The key idea used in bounding the k -RIC of the columnwise KR product $\mathbf{A} \odot \mathbf{B}$ is the observation that the Gram matrix of $\mathbf{A} \odot \mathbf{B}$ can be interpreted as a Hadamard product between the two correlation matrices $\mathbf{A}^T\mathbf{A}$ and $\mathbf{B}^T\mathbf{B}$, as mentioned in the following Proposition.

Proposition 12 (Rao and Rao [42]). *For $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{m \times n}$,*

$$(\mathbf{A} \odot \mathbf{B})^T(\mathbf{A} \odot \mathbf{B}) = (\mathbf{A}^T\mathbf{A}) \circ (\mathbf{B}^T\mathbf{B}) \quad (24)$$

Proof. See Proposition 6.4.2 in [42]. \square

Then, by using the forward and reverse Kantorovich matrix inequalities, we obtain the proposed upper bound for k -RIC of $\mathbf{A} \odot \mathbf{B}$ as explained in the following arguments.

Without loss of generality, let $S \subset [n]$ be an arbitrary index set representing the nonzero support of \mathbf{z} in (8), with $|S| \leq k$. Let \mathbf{A}_S denote the $m \times |S|$ submatrix of \mathbf{A} , constituting $|S|$ columns of \mathbf{A} indexed by the set S . Let \mathbf{B}_S be constructed similarly. Since $\delta_k(\mathbf{A}), \delta_k(\mathbf{B}) < 1$, both $\mathbf{A}_S, \mathbf{B}_S$ have full rank, and consequently the associated Gram matrices $\mathbf{A}_S^T\mathbf{A}_S, \mathbf{B}_S^T\mathbf{B}_S$ are positive definite. Further, since \mathbf{A} and \mathbf{B} have unit norm columns, both $\mathbf{A}_S^T\mathbf{A}_S$ and $\mathbf{B}_S^T\mathbf{B}_S$ are correlation matrices with unit diagonal entries. Using Proposition 12, we can write

$$(\mathbf{A}_S \odot \mathbf{B}_S)^T(\mathbf{A}_S \odot \mathbf{B}_S) = \mathbf{A}_S^T\mathbf{A}_S \circ \mathbf{B}_S^T\mathbf{B}_S. \quad (25)$$

Next, for $k \leq m$, by applying Lemma 1 to the positive definite matrices $(\mathbf{A}_S^T\mathbf{A}_S)^{1/2}$ and $(\mathbf{B}_S^T\mathbf{B}_S)^{1/2}$, we get,

$$\begin{aligned} \mathbf{A}_S^T\mathbf{A}_S \circ \mathbf{B}_S^T\mathbf{B}_S &\leq \left((\mathbf{A}_S^T\mathbf{A}_S)^{\frac{1}{2}} \circ (\mathbf{B}_S^T\mathbf{B}_S)^{\frac{1}{2}} \right)^2 + \frac{1}{4}(M - m)^2\mathbf{I}_k \\ &\leq \mathbf{I}_k + \frac{1}{4}(M - m)^2\mathbf{I}_k, \end{aligned} \quad (26)$$

where the second inequality is a consequence of the unity bound on the spectral radius of the Hadamard product between correlation matrices, shown in Proposition 8. In (26), M and m are upper and lower bounds for the maximum and minimum eigenvalues of $(\mathbf{A}_S^T\mathbf{A}_S)^{1/2} \otimes (\mathbf{B}_S^T\mathbf{B}_S)^{1/2}$, respectively. From the restricted isometry of \mathbf{A} and \mathbf{B} , and by application of

Proposition 4, the minimum and maximum eigenvalues of $(\mathbf{A}_S^T \mathbf{A}_S)^{1/2} \otimes (\mathbf{B}_S^T \mathbf{B}_S)^{1/2}$ are lower and upper bounded by $\sqrt{(1 - \delta_k^{\mathbf{A}})(1 - \delta_k^{\mathbf{B}})}$ and $\sqrt{(1 + \delta_k^{\mathbf{A}})(1 + \delta_k^{\mathbf{B}})}$, respectively. By introducing $\delta \triangleq \max(\delta_k^{\mathbf{A}}, \delta_k^{\mathbf{B}})$, it is easy to check that the eigenvalues of $(\mathbf{A}_S^T \mathbf{A}_S)^{1/2} \otimes (\mathbf{B}_S^T \mathbf{B}_S)^{1/2}$ also lie inside the interval $[1 - \delta, 1 + \delta]$. Plugging $m = 1 - \delta$ and $M = 1 + \delta$ in (26), and by using (25), we get

$$(\mathbf{A}_S \odot \mathbf{B}_S)^T (\mathbf{A}_S \odot \mathbf{B}_S) \leq (1 + \delta^2) \mathbf{I}_k. \quad (27)$$

Similarly, by applying Lemma 2 to $\mathbf{A}_S^T \mathbf{A}_S$ and $\mathbf{B}_S^T \mathbf{B}_S$ with $c = \sqrt{1 - \delta}$ and $d = \sqrt{1 + \delta}$, we obtain

$$(\mathbf{A}_S^T \mathbf{A}_S)^{1/2} \circ (\mathbf{B}_S^T \mathbf{B}_S)^{1/2} \geq (\sqrt{1 - \delta^2}) \mathbf{I}_k.$$

From Proposition 6, we have $\mathbf{A}_S^T \mathbf{A}_S \circ \mathbf{B}_S^T \mathbf{B}_S \geq ((\mathbf{A}_S^T \mathbf{A}_S)^{1/2} \circ (\mathbf{B}_S^T \mathbf{B}_S)^{1/2})^2$. Therefore, we can write

$$\mathbf{A}_S^T \mathbf{A}_S \circ \mathbf{B}_S^T \mathbf{B}_S \geq (1 - \delta^2) \mathbf{I}_k.$$

Further, using (25), we get

$$(\mathbf{A}_S \odot \mathbf{B}_S)^T (\mathbf{A}_S \odot \mathbf{B}_S) \geq (1 - \delta^2) \mathbf{I}_k. \quad (28)$$

Finally, Theorem 1's statement follows from (27) and (28).

VI. PRELIMINARIES FOR PROBABILISTIC k -RIC BOUND

In this section, we briefly discuss some concentration results on certain functions of subgaussian random variables which will appear in the proofs of Theorems 2 and 3.

A. A Tail Probability for Subgaussian Vectors

The Theorem below presents the *Hanson-Wright* inequality [43], [44], a tail probability for a quadratic form constructed using independent subgaussian random variables.

Theorem 13 (Rudelson and Vershynin [44]). *Let $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n) \in \mathbb{R}^n$ be a random vector with independent components \mathbf{x}_i satisfying $\mathbb{E}\mathbf{x}_i = 0$ and $\|\mathbf{x}_i\|_{\psi_2} \leq K$. Let \mathbf{A} be an $n \times n$ matrix. Then, for every $t \geq 0$,*

$$\begin{aligned} & \mathbb{P} \left\{ |\mathbf{x}^T \mathbf{A} \mathbf{x} - \mathbb{E} \mathbf{x}^T \mathbf{A} \mathbf{x}| > t \right\} \\ & \leq 2 \exp \left[-c \min \left(\frac{t^2}{K^4 \|\mathbf{A}\|_{HS}^2}, \frac{t}{K^2 \|\mathbf{A}\|} \right) \right] \end{aligned}$$

where c is a universal positive constant.

B. Concentration of Supremum of Linear Combination of Bounded, Nonnegative Independent Random Variables

The following theorem bounds the subgaussian tail of a function of several independent random variables satisfying a relaxed bounded difference property.

Theorem 14. *Consider a general real-valued function of n independent random variables $\mathbf{z} = f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n)$ and \mathbf{z}_i denotes an \mathbf{x}_i -measurable random variable defined by $\mathbf{z}_i = \inf_{\mathbf{x}'_i} f(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}'_i, \dots, \mathbf{x}_n)$. Assume that the random variable \mathbf{z} is such that there exists a constant $\nu > 0$, for which,*

$$\sum_{i=1}^n (\mathbf{z} - \mathbf{z}_i)^2 \leq \nu$$

almost surely. Then, for all $t > 0$,

$$\mathbb{P}(\mathbf{z} - \mathbb{E}\mathbf{z} \geq t) \leq e^{-t^2/2\nu}.$$

Proof. See Theorem 6.7 in [45]. \square

Theorem 14 can be used to show a subgaussian tail for the maximal linear combination of bounded nonnegative random variables, stated below.

Lemma 3. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n$ be i.i.d. nonnegative, bounded random variables satisfying $\mathbf{x}_i \leq b$. Then, $\mathbf{y} \triangleq \sup_{\mathbf{z} \in \mathbb{R}^n, \|\mathbf{z}\|_2 \leq 1} \sum_{i=1}^n z_i \mathbf{x}_i$ is a nonnegative random variable which satisfies the following tail inequality:*

$$\mathbb{P}(\mathbf{y} - \mathbb{E}\mathbf{y} \geq t) \leq e^{-t^2/2b^2}.$$

Proof. See Appendix B. \square

The following proposition bounds the expectation of a nonnegative subgaussian random variable.

Proposition 15. *Let \mathbf{x} be a nonnegative random variable with a subgaussian tail, i.e., for $t \geq 0$,*

$$\mathbb{P}(\mathbf{z} - \mathbb{E}\mathbf{z} \geq t) \leq e^{-t^2/2\nu}$$

for some $\nu > 0$. Then, $\mathbb{E}\mathbf{z} \leq \sqrt{2\pi\nu}$.

Proof. See Appendix C. \square

VII. PROOF OF THE PROBABILISTIC k -RIC BOUND (THEOREM 2)

The proof of Theorem 2 starts with a variational definition of the k -RIC, $\delta_k \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}} \right)$ given below.

$$\delta_k \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}} \right) = \sup_{\substack{\mathbf{z} \in \mathbb{R}^n, \\ \|\mathbf{z}\|_2 = 1, \|\mathbf{z}\|_0 \leq k}} \left| \left\| \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}} \right) \mathbf{z} \right\|_2^2 - 1 \right|. \quad (29)$$

In order to find a probabilistic upper bound for δ_k , we intend to find a constant $\delta \in (0, 1)$ such that $\mathbb{P}(\delta_k \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}} \right) \geq \delta)$ is arbitrarily close to zero. We therefore consider the tail event

$$\mathcal{E} \triangleq \left\{ \sup_{\substack{\mathbf{z} \in \mathbb{R}^n, \\ \|\mathbf{z}\|_2 = 1, \|\mathbf{z}\|_0 \leq k}} \left| \left\| \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}} \right) \mathbf{z} \right\|_2^2 - 1 \right| \geq \delta \right\}, \quad (30)$$

and show that for m sufficiently large, $\mathbb{P}(\mathcal{E})$ can be driven arbitrarily close to zero. In other words, the constant δ serves as a probabilistic upper bound for $\delta_k \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{B}}{\sqrt{m}} \right)$.

Let \mathcal{U}_k be the set of all k or less sparse unit norm vectors in \mathbb{R}^n . Then, using Proposition 12, the tail event in (30) can be rewritten as

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &= \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \mathbf{z}^T (\mathbf{A} \odot \mathbf{B})^T (\mathbf{A} \odot \mathbf{B}) \mathbf{z} - m^2 \right| \geq \delta m^2 \right) \\ &= \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \mathbf{z}^T (\mathbf{A}^T \mathbf{A} \circ \mathbf{B}^T \mathbf{B}) \mathbf{z} - m^2 \right| \geq \delta m^2 \right) \\ &= \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1}^n z_i z_j (\mathbf{a}_i^T \mathbf{a}_j) (\mathbf{b}_i^T \mathbf{b}_j) - m^2 \right| \geq \delta m^2 \right) \end{aligned}$$

where \mathbf{a}_i and \mathbf{b}_i denote the i th column of \mathbf{A} and \mathbf{B} , respectively. Further, by applying the triangle inequality and the union bound, we get

$$\begin{aligned} \mathbb{P}(\mathcal{E}) &\leq \mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n z_i^2 \|\mathbf{a}_i\|_2^2 \|\mathbf{b}_i\|_2^2 - m^2 \right| \geq \alpha \delta m^2\right) + \\ &\mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j \mathbf{a}_i^T \mathbf{a}_j \mathbf{b}_i^T \mathbf{b}_j \right| \geq (1 - \alpha) \delta m^2\right). \end{aligned} \quad (31)$$

In the above, $\alpha \in (0, 1)$ is a variational union bound parameter which can be optimized at a later stage. We now proceed to find separate upper bounds for each of the two probability terms in (31).

The first probability term in (31) admits the following sequence of relaxations.

$$\begin{aligned} &\mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n z_i^2 \|\mathbf{a}_i\|_2^2 \|\mathbf{b}_i\|_2^2 - m^2 \right| \geq \alpha \delta m^2\right) \\ &\stackrel{(a)}{\leq} \mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n z_i^2 \left| \|\mathbf{a}_i\|_2^2 \|\mathbf{b}_i\|_2^2 - m^2 \right| \geq \alpha \delta m^2\right) \\ &\stackrel{(b)}{\leq} \mathbb{P}\left(\max_{1 \leq i \leq n} \left| \|\mathbf{a}_i\|_2^2 \|\mathbf{b}_i\|_2^2 - m^2 \right| \geq \alpha \delta m^2\right) \\ &\stackrel{(c)}{=} \mathbb{P}\left(\bigcup_{1 \leq i \leq n} \left\{ \left| \|\mathbf{a}_i\|_2^2 \|\mathbf{b}_i\|_2^2 - m^2 \right| \geq \alpha \delta m^2 \right\}\right) \\ &\stackrel{(d)}{\leq} \sum_{i=1}^n \mathbb{P}\left(\left| \|\mathbf{a}_i\|_2^2 \|\mathbf{b}_i\|_2^2 - m^2 \right| \geq \alpha \delta m^2\right) \\ &\stackrel{(e)}{=} n \mathbb{P}\left(\left| \|\mathbf{a}_1\|_2^2 \|\mathbf{b}_1\|_2^2 - m^2 \right| \geq \alpha \delta m^2\right). \\ &\stackrel{(f)}{\leq} n \mathbb{P}\left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \left| \|\mathbf{b}_1\|_2^2 - m \right| \geq \alpha \beta \delta m^2\right) \\ &\quad + 2n \mathbb{P}\left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \frac{\alpha(1 - \beta)\delta m}{2}\right) \\ &\stackrel{(g)}{\leq} 2n \mathbb{P}\left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \sqrt{\alpha \beta \delta m}\right) \\ &\quad + 2n \mathbb{P}\left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \frac{\alpha(1 - \beta)\delta m}{2}\right). \\ &\stackrel{(h)}{\leq} 4n \mathbb{P}\left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \frac{\alpha \delta m}{2} \left(1 - \frac{\alpha \delta}{4}\right)\right) \\ &\stackrel{(i)}{\leq} 4ne^{-cm \frac{\alpha^2 \delta^2}{4K^4} (1 - \alpha \delta/4)^2}. \end{aligned} \quad (32)$$

In the above, step (a) follows from the triangle inequality combined with the fact that z_i^2 's sum to one. The inequality in step (b) is a consequence of the fact that any nonnegative convex combination of n arbitrary numbers is at most the maximum among the n numbers. Step (c) is obtained by simply rewriting the tail event for the maximum of n random variables as the union of tail events for the individual random variables. Step (d) is obtained by using the union bound over values of index $i \in [n]$. Step (e) follows because the columns of \mathbf{A} and \mathbf{B} are i.i.d.. Step (f) is the union bound combined with the fact that for any two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^m$, the following triangle inequality holds:

$$\left| \|\mathbf{a}\|_2^2 \|\mathbf{b}\|_2^2 - m^2 \right| \leq \left(\left| \|\mathbf{a}\|_2^2 - m \right| \right) \left(\left| \|\mathbf{b}\|_2^2 - m \right| \right)$$

$$+ m \left| \left| \|\mathbf{a}\|_2^2 - m \right| + m \left| \left| \|\mathbf{b}\|_2^2 - m \right| \right|.$$

In step (d), $\beta \in (0, 1)$ is a variational union bound parameter. Step (g) is once again the union bound and exploits the i.i.d. nature of the columns of \mathbf{A} and \mathbf{B} . Step (h) is obtained by setting $\beta = \alpha \delta/4$. Lastly, step (i) is the Hanson-Wright inequality (Theorem 13) applied to the subgaussian vector \mathbf{a}_1 .

Next, we turn our attention to the second probability term in (31). It can be upper bounded as follows:

$$\begin{aligned} &\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j \mathbf{a}_i^T \mathbf{a}_j \mathbf{b}_i^T \mathbf{b}_j \right| \\ &\leq \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 \left(\sum_{j=1, j \neq i}^n |z_j| |u_{ij} v_{ij}| \right) \\ &\leq \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 \right) \\ &\quad \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z}' \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z'_j| |u_{ij} v_{ij}| \right), \end{aligned} \quad (33)$$

where $u_{ij} \triangleq \langle \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2}, \mathbf{a}_j \rangle$ and $v_{ij} \triangleq \langle \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|_2}, \mathbf{b}_j \rangle$. From the approximate rotational invariance of subgaussian random variables [32], it can be shown that the inner products u_{ij} and v_{ij} are also subgaussian with $\|u_{ij}\|_{\psi_2} = \|v_{ij}\|_{\psi_2} \leq c_3 K$, with $c_3 > 0$ being a universal numerical constant.

Using (33), and by applying the union bound, the second probability term in (31) can be written as the sum of following two tail probabilities.

$$\begin{aligned} &\mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j \mathbf{a}_i^T \mathbf{a}_j \mathbf{b}_i^T \mathbf{b}_j \right| \geq (1 - \alpha) \delta m^2\right) \\ &\leq \mathbb{P}\left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{j=1, j \neq i}^n z_j u_{ij} v_{ij} \right| \geq 4\sqrt{2\pi} K_o^2 (\gamma \log n)^{3/2}\right) \\ &\quad + \mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 \geq \frac{(1 - \alpha) \delta m^2}{4\sqrt{2\pi} K_o^2 (\gamma \log n)^{3/2}}\right) \end{aligned} \quad (34)$$

The following Lemmas provide upper bounds for the two tail probability terms in (34).

Lemma 4. *Let \mathbf{A} and \mathbf{B} be $m \times n$ random matrices as specified in Theorem 2. Let $u_{ij} \triangleq \langle \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2}, \mathbf{a}_j \rangle$ and $v_{ij} \triangleq \langle \frac{\mathbf{b}_i}{\|\mathbf{b}_i\|_2}, \mathbf{b}_j \rangle$, where \mathbf{a}_i and \mathbf{b}_i denote the i^{th} column of \mathbf{A} and \mathbf{B} , respectively. Then,*

$$\begin{aligned} &\mathbb{P}\left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| |u_{ij} v_{ij}| \geq 4\sqrt{2\pi} K_o^2 (\gamma \log n)^{3/2}\right) \\ &\leq \frac{2e}{n^{2(c_2 \gamma - 1)}}, \end{aligned} \quad (35)$$

where $K_o \triangleq \max(K, 1)$ and $c_2 > 0$ is a universal constant.

Proof. See Appendix D. \square

Lemma 5. Let \mathbf{A} and \mathbf{B} be $m \times n$ random matrices with independent zero mean, unit variance, i.i.d. subgaussian entries satisfying $\|\mathbf{A}_{ij}\|_{\psi_2} = \|\mathbf{B}_{ij}\|_{\psi_2} \leq K$. Then,

$$\mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 \geq \frac{(1-\alpha)\delta m^2}{4\sqrt{2\pi}K_o^2(\gamma \log n)^{3/2}}\right) \leq 2ne^{-\frac{cm}{K_o^4}}, \quad (36)$$

provided $m \geq \frac{8\sqrt{2\pi}K_o^2\gamma^{3/2}}{(1-\alpha)\delta} \sqrt{k} \log^{3/2} n$. Here, $K_o = \max(K, 1)$ and $c > 0$ is a universal constant.

Proof. See Appendix E. \square

By using Lemmas 4 and 5 together in (34), we get

$$\mathbb{P}\left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j \mathbf{a}_i^T \mathbf{a}_j \mathbf{b}_i^T \mathbf{b}_j \right| \geq (1-\alpha)\delta m^2\right) \leq \frac{2e}{n^{2(c_2\gamma-1)}} + 2ne^{-\frac{cm}{K_o^4}} \leq \frac{4e}{n^{2(c_2\gamma-1)}}, \quad (37)$$

provided $m \geq \max\left(\frac{8\sqrt{2\pi}K_o^2\gamma^{3/2}}{(1-\alpha)\delta} \sqrt{k} \log^{3/2} n, \frac{(2c_2\gamma-1)K_o^4 \log n}{c}\right)$. Using (32) and (37) together in (31), we get

$$\mathbb{P}(\mathcal{E}) \leq \frac{4}{n^{\left(\frac{c\alpha^2\delta^2(1-\alpha\delta/4)^2 m}{4K_o^4 \log n} - 1\right)}} + \frac{4e}{n^{2(c_2\gamma-1)}} \leq \frac{8e}{n^{2(c_2\gamma-1)}}, \quad (38)$$

provided $m \geq \max\left(\frac{8\sqrt{2\pi}K_o^2\gamma^{3/2}}{(1-\alpha)\delta} \sqrt{k} \log^{3/2} n, \frac{(2c_2\gamma-1)\xi K_o^4 \log n}{c}\right)$, where $\xi = \max\left(1, \frac{4}{\alpha^2\delta^2(1-\alpha\delta/4)^2}\right)$.

Lastly, by choosing $\alpha = 1/2$, we get $\mathbb{P}(\delta_k \geq \delta) = \mathbb{P}(\mathcal{E}) \leq 8en^{-2(c_2\gamma-1)}$, provided that

$$m \geq \max\left(\frac{16\sqrt{2\pi}K_o^2\gamma^{3/2}}{\delta} \sqrt{k} \log^{3/2} n, \frac{(2c_2\gamma-1)\xi K_o^4 \log n}{c}\right),$$

where $\xi = \max\left(1, \frac{16}{\gamma\delta^2(1-\delta/8)^2}\right)$. Note that, in terms of k and n , the first term in the inequality for m scales as $\sqrt{k} \log^{3/2} n$; it dominates the second term, which scales as $\log n$. This ends our proof.

VIII. CONCLUSIONS

In this work, we have analyzed the restricted isometry property of the columnwise Khatri-Rao product matrix in terms of its restricted isometry constants. We gave two upper bounds for the k -RIC of a generic columnwise Khatri-Rao product matrix. The first k -RIC bound, a deterministic bound, is valid for the Khatri-Rao product of an arbitrary pair of input matrices of the same size with normalized columns. It is conveniently computed in terms of the k -RICs of the input matrices. We also gave a probabilistic RIC bound for the columnwise KR product of a pair of random matrices with i.i.d. subgaussian entries. The probabilistic RIC bound is one of the key components needed for computing tight sample complexity bounds for several machine learning algorithms.

The analysis of the RIP of Khatri-Rao product matrices in this article can be extended in multiple ways. The current RIC bounds can be extended to the Khatri-Rao product of three or more matrices. More importantly, in order to relate the RICs

to the dimensions of the input matrices, we had to resort to the randomness in their entries. Removing this randomness aspect of our results could be an interesting direction for future work.

APPENDIX

A. Proof of Proposition 7

Proof. Since \mathbf{A} is a correlation matrix, it admits the Schur decomposition, $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, with unitary \mathbf{U} and eigenvalue matrix $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n)$. Since \mathbf{A} is positive semi-definite, its nonnegative square-root exists and is given by $\mathbf{A}^{1/2} = \mathbf{U}\mathbf{\Lambda}^{1/2}\mathbf{U}^T$. Consider

$$\begin{aligned} \mathbf{A}^{1/2} \circ \mathbf{A}^{1/2} &= \sum_{i=1}^n \lambda_i^{1/2} \mathbf{u}_i \mathbf{u}_i^T \circ \sum_{j=1}^n \lambda_j^{1/2} \mathbf{u}_j \mathbf{u}_j^T \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{1/2} \lambda_j^{1/2} (\mathbf{u}_i \mathbf{u}_i^T) \circ (\mathbf{u}_j \mathbf{u}_j^T) \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{1/2} \lambda_j^{1/2} (\mathbf{u}_i \circ \mathbf{u}_j) (\mathbf{u}_i \circ \mathbf{u}_j)^T. \end{aligned} \quad (39)$$

Above, the second equality follows from the distributive property of the Hadamard product and the last step follows from Fact 7.6.2 in [2]. Using (39), we can show that the rows and columns of $\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}$ sum to one, as follows:

$$\begin{aligned} (\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}) \mathbf{1} &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{1/2} \lambda_j^{1/2} (\mathbf{u}_i \circ \mathbf{u}_j) (\mathbf{u}_i \circ \mathbf{u}_j)^T \mathbf{1} \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{1/2} \lambda_j^{1/2} (\mathbf{u}_i \circ \mathbf{u}_j) (\mathbf{u}_i \circ \mathbf{1})^T \mathbf{u}_j \\ &= \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{1/2} \lambda_j^{1/2} (\mathbf{u}_i \circ \mathbf{u}_j) \mathbf{u}_i^T \mathbf{u}_j \\ &= \sum_{i=1}^n \lambda_i (\mathbf{u}_i \circ \mathbf{u}_i) = \mathbf{d} \quad (\text{say}). \end{aligned}$$

The above arguments follow from the orthonormality of the columns of \mathbf{U} , and repeated application of Fact 7.6.1 in [2]. Note that for $k \in [n]$, $\mathbf{d}(k) = \sum_{i=1}^n \lambda(i) (\mathbf{u}_i(k))^2 = [\mathbf{U}\mathbf{\Lambda}\mathbf{U}^T]_{kk} = \mathbf{A}_{kk} = 1$. Thus, we have shown that $(\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}) \mathbf{1} = \mathbf{1}$. Likewise, it can be shown that $\mathbf{1}^T (\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}) = \mathbf{1}^T$. Thus, $\mathbf{A}^{1/2} \circ \mathbf{A}^{1/2}$ is doubly stochastic. \square

B. Proof of Lemma 3

Define $\mathbf{y}^{(i)} \triangleq \inf_{\mathbf{x}'_i} \sup_{\mathbf{z} \in \mathbb{R}^n, \|\mathbf{z}\|_2 \leq 1} (z_1 \mathbf{x}_1 + \dots + z_i \mathbf{x}'_i + \dots + z_n \mathbf{x}_n)$.

Further, let $\mathbf{z}^* \triangleq \arg \sup_{\mathbf{z} \in \mathbb{R}^n, \|\mathbf{z}\|_2 \leq 1} \sum_{j=1}^n z_j \mathbf{x}_j$. Since \mathbf{x}_i are non-negative, clearly \mathbf{z}^* is a nonnegative, unit norm vector. Then, for $i \in [n]$,

$$\begin{aligned} \mathbf{y} - \mathbf{y}^{(i)} &= \sup_{\mathbf{z} \in \mathbb{R}^n, \|\mathbf{z}\|_2 \leq 1} \sum_{j=1}^n z_j \mathbf{x}_j - \inf_{\mathbf{x}'_i} \sup_{\mathbf{z}' \in \mathbb{R}^n, \|\mathbf{z}'\|_2 \leq 1} \left(\sum_{\substack{j=1, \\ j \neq i}}^n z'_j \mathbf{x}_j + z'_i \mathbf{x}'_i \right) \\ &\leq \sum_{j=1}^n z_j^* \mathbf{x}_j - \inf_{\mathbf{x}'_i} \left(\sum_{j=1, j \neq i}^n z_j^* \mathbf{x}_j + z_i^* \mathbf{x}'_i \right) \end{aligned}$$

$$= z_i^* \left(\mathbf{x}_i - \inf_{\mathbf{x}'_i} \mathbf{x}'_i \right) = z_i^* \mathbf{x}_i. \quad (40)$$

Since $\mathbf{x}_i \leq b$ and $\|\mathbf{z}^*\|_2 \leq 1$, we have $\sum_{i=1}^n (\mathbf{y} - \mathbf{y}^{(i)})^2 \leq \sum_{i=1}^n \mathbf{z}_i^{*2} b^2 \leq b^2$. Finally, by invoking Theorem 14 with $\nu = b^2$, we obtain the desired tail bound for \mathbf{y} . \square

C. Proof of Proposition 15

The said upper bound for $\mathbb{E}\mathbf{z}$ can be derived as shown below.

$$\begin{aligned} \mathbb{E}\mathbf{z} &= \int_0^\infty \mathbb{P}(\mathbf{z} \geq t) dt = \int_{-\mathbb{E}\mathbf{z}}^\infty \mathbb{P}(\mathbf{z} \geq r + \mathbb{E}\mathbf{z}) dr \\ &\leq \int_{-\mathbb{E}\mathbf{z}}^\infty e^{-r^2/2\nu} dr \leq \int_{-\infty}^\infty e^{-\frac{r^2}{2\nu}} dr \leq \sqrt{2\pi\nu}. \end{aligned}$$

\square

D. Proof of Lemma 4

Let $\mathcal{Q} \subseteq (\mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n})$ be the collection of matrix tuples (\mathbf{X}, \mathbf{Y}) satisfying $|u_{ij}v_{ij}| < 2K^2\gamma \log n$ for all $1 \leq i, j \leq n, j \neq i$, where u_{ij} and v_{ij} are evaluated as $\langle \frac{\mathbf{x}_i^T \mathbf{x}_j}{\|\mathbf{x}_i\|_2} \rangle$ and $\langle \frac{\mathbf{y}_i^T \mathbf{y}_j}{\|\mathbf{y}_i\|_2} \rangle$, respectively. Then, for subgaussian random matrices \mathbf{A}, \mathbf{B} as defined in Lemma 4,

$$\begin{aligned} \mathbb{P}((\mathbf{A}, \mathbf{B}) \notin \mathcal{Q}) &= \mathbb{P} \left(\bigcup_{\substack{1 \leq i, j \leq n \\ j \neq i}} \{|u_{ij}v_{ij}| \geq 2K^2\gamma \log n\} \right) \\ &\stackrel{(a)}{\leq} \sum_{\substack{1 \leq i, j \leq n \\ j \neq i}} \mathbb{P}(|u_{ij}v_{ij}| \geq 2K^2\gamma \log n) \\ &\stackrel{(b)}{\leq} \frac{n^2}{2} \mathbb{P}(|u_{12}v_{12}| \geq 2K^2\gamma \log n) \\ &\stackrel{(c)}{\leq} n^2 \mathbb{P}(|u_{12}| \geq \sqrt{2}K\sqrt{\gamma \log n}) \\ &\stackrel{(d)}{\leq} en^2 e^{-2c_1\gamma \log n} = \frac{e}{n^{2(c_1\gamma-1)}}. \end{aligned} \quad (41)$$

In the above, steps (a), (b), and (c) are obtain by applying the union bound over $\binom{n}{2}$ combinations of distinct (i, j) pairs and exploiting the identical subgaussian tails of the random variables u_{ij} and v_{ij} , $i \neq j$. Step (d) is simply the standard subgaussian tail bound, with $c_1 > 0$ a universal constant. Thus, for $\gamma > 1/c_1$, the random variable $|u_{ij}v_{ij}|$, $i \neq j$, is uniformly bounded by $2K^2\gamma \log n$ with arbitrarily high probability, for sufficiently large n .

Let \mathcal{E}_2 denote the event that the matrix tuple $(\mathbf{A}, \mathbf{B}) \in \mathcal{Q}$. Then, by the law of total probability, we have

$$\begin{aligned} &\mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| |u_{ij}v_{ij}| \geq t \right) \\ &\leq \mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| |u_{ij}v_{ij}| \geq t \mid \mathcal{E}_2 \right) + \mathbb{P}(\mathcal{E}_2^c) \\ &\leq n \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}v_{1j}| \geq t \mid \mathcal{E}_2 \right) + \mathbb{P}(\mathcal{E}_2^c). \end{aligned} \quad (42)$$

Let $\mathcal{E}_2^{\mathbf{a}, \mathbf{b}} \triangleq \mathcal{E}_2 \cap \{(\mathbf{X}, \mathbf{Y}) \in \mathbb{R}^{m \times n} \times \mathbb{R}^{m \times n} : \mathbf{x}_1 = \mathbf{a}, \mathbf{y}_1 = \mathbf{b}\}$ denote the set of matrix tuples (\mathbf{X}, \mathbf{Y}) with their first columns fixed as \mathbf{a} and $\mathbf{b} \in \mathbb{R}^m$, respectively. Then, conditioned on $\mathcal{E}_2^{\mathbf{a}, \mathbf{b}}$, the random variables $|u_{1j}v_{1j}|$, $2 \leq j \leq n$ are i.i.d., bounded and nonnegative random variables. By applying Lemma 3, it follows that

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}v_{1j}| - \mathbb{E} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}v_{1j}| \geq t \mid \mathcal{E}_2^{\mathbf{a}, \mathbf{b}} \right) \leq e^{-t^2/8K^4\gamma^2 \log^2 n}. \quad (43)$$

It is to be noted here that the above tail bound in the RHS does not depend on \mathbf{a} and \mathbf{b} . This leads to an interesting consequence explained in the following discourse.

Let \mathcal{T} be an event such that $\mathbb{P}(\mathcal{T} \mid \mathcal{E}_2^{\mathbf{a}, \mathbf{b}}) \leq \eta$, where η is independent of \mathbf{a}, \mathbf{b} . Since $\mathcal{E}_2 = \bigcup_{\mathbf{a}, \mathbf{b} \in \mathbb{R}^n} \mathcal{E}_2^{\mathbf{a}, \mathbf{b}}$, we have

$$\begin{aligned} \mathbb{P}(\mathcal{T} \mid \mathcal{E}_2) &= \frac{\mathbb{P}(\mathcal{T} \cap \mathcal{E}_2)}{\mathbb{P}(\mathcal{E}_2)} = \frac{\mathbb{P} \left(\bigcup_{\mathbf{a}, \mathbf{b} \in \mathbb{R}^n} (\mathcal{T} \cap \mathcal{E}_2^{\mathbf{a}, \mathbf{b}}) \right)}{\mathbb{P}(\mathcal{E}_2)} \\ &= \frac{\int_{\mathbf{a}, \mathbf{b}} \mathbb{P}(\mathcal{T} \cap \mathcal{E}_2^{\mathbf{a}, \mathbf{b}})}{\mathbb{P}(\mathcal{E}_2)} = \frac{\int_{\mathbf{a}, \mathbf{b}} \mathbb{P}(\mathcal{T} \mid \mathcal{E}_2^{\mathbf{a}, \mathbf{b}}) \mathbb{P}(\mathcal{E}_2^{\mathbf{a}, \mathbf{b}})}{\mathbb{P}(\mathcal{E}_2)} \\ &\leq \eta \frac{\int_{\mathbf{a}, \mathbf{b}} \mathbb{P}(\mathcal{E}_2^{\mathbf{a}, \mathbf{b}})}{\mathbb{P}(\mathcal{E}_2)} = \eta \frac{\mathbb{P}(\mathcal{E}_2)}{\mathbb{P}(\mathcal{E}_2)} = \eta. \end{aligned} \quad (44)$$

Coming back to our proof, using similar arguments as above⁵ and on account of the tail bound in the RHS in (43) being independent of \mathbf{a} and \mathbf{b} , it follows that

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{ij}v_{ij}| - \mathbb{E} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}v_{1j}| \geq t \mid \mathcal{E}_2 \right) \leq e^{-t^2/8K^4\gamma^2 \log^2 n}. \quad (45)$$

Further, using Proposition 15, the probability term in the LHS of the inequality in (45) can be replaced by a smaller probability term as shown below.

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}v_{1j}| \geq t + 2\sqrt{2\pi}K^2\gamma \log n \mid \mathcal{E}_2 \right) \leq e^{-t^2/8K^2\gamma^2 \log^2 n}.$$

Setting $t = 4K(\gamma \log n)^{3/2}$, we get

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}v_{1j}| \geq 4\sqrt{2\pi}K_o^2(\gamma \log n)^{\frac{3}{2}} \mid \mathcal{E}_2 \right) \leq \frac{1}{n^{2\gamma}}. \quad (46)$$

Finally, by combining (41), (42) and (46), we obtain the desired tail bound:

$$\begin{aligned} &\mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| |u_{ij}v_{ij}| \geq 4\sqrt{2\pi}K_o^2(\gamma \log n)^{3/2} \right) \\ &\leq \frac{1}{n^{2\gamma-1}} + \frac{e}{n^{2(c_1\gamma-1)}} \leq \frac{2e}{n^{2(c_2\gamma-1)}}, \end{aligned} \quad (47)$$

where $c_2 = \min(c_1, 1)$ is a universal positive constant.

⁵The tail bound in (45) can also be obtained by using the law of total probability extended to conditional probabilities.

E. Proof of Lemma 5

Consider the tail event

$$\mathcal{E}_3 = \left\{ \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 \geq \frac{(1-\alpha)\delta m^2}{4\sqrt{2\pi}K_o^2(\gamma \log n)^{3/2}} \right\}. \quad (48)$$

While bounding the tail probability $\mathbb{P}(\mathcal{E}_3)$, the supremum with respect to \mathbf{z} can be circumvented as shown below.

$$\begin{aligned} \mathbb{P}(\mathcal{E}_3) &= \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n \frac{|z_i|}{\sqrt{k}} \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 \geq \frac{(1-\alpha)\delta m^2}{4\sqrt{2\pi}kK_o^2(\gamma \log n)^{3/2}} \right) \\ &\leq \mathbb{P} \left(\max_{1 \leq i \leq n} \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2 \geq \frac{(1-\alpha)\delta m^2}{4\sqrt{2\pi}kK_o^2(\gamma \log n)^{3/2}} \right). \quad (49) \end{aligned}$$

The above inequality is obtained by observing that since \mathbf{z} has unit ℓ_2 -norm, by Cauchy-Schwarz inequality, we have $\frac{1}{\sqrt{k}} \sum_{i=1}^n |z_i| \leq 1$. Consequently, the linear combination $\sum_{i=1}^n \frac{|z_i|}{\sqrt{k}} \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2$ can be at most $\max_{1 \leq i \leq n} \|\mathbf{a}_i\|_2 \|\mathbf{b}_i\|_2$, a quantity independent of \mathbf{z} .

Since \mathbf{a}_i and \mathbf{b}_i , the column vectors of \mathbf{A} and \mathbf{B} , respectively, are i.i.d., the following series of union bounds apply.

$$\begin{aligned} \mathbb{P}(\mathcal{E}_3) &\leq n\mathbb{P} \left(\|\mathbf{a}_1\|_2 \|\mathbf{b}_1\|_2 \geq \frac{(1-\alpha)\delta m^2}{4\sqrt{2\pi}kK_o^2(\gamma \log n)^{3/2}} \right) \\ &\leq 2n\mathbb{P} \left(\|\mathbf{a}_1\|_2 \geq \sqrt{\frac{(1-\alpha)\delta m^2}{4\sqrt{2\pi}kK_o^2(\gamma \log n)^{3/2}}} \right) \\ &\leq 2n\mathbb{P} \left(\|\mathbf{a}_1\|_2^2 - m \geq m \left(\frac{(1-\alpha)\delta m}{4\sqrt{2\pi}kK_o^2(\gamma \log n)^{3/2}} - 1 \right) \right). \quad (50) \end{aligned}$$

Since \mathbf{a}_1 is a subgaussian vector with independent entries, by the Hanson-Wright concentration inequality in Theorem 13, the desired tail probability can be obtained as follows:

$$\mathbb{P}(\mathcal{E}_3) \leq 2ne^{-\frac{cm}{K_o^4} \left(\frac{(1-\alpha)\delta m}{4\sqrt{2\pi}kK_o^2\gamma^{3/2}\sqrt{k}(\log n)^{3/2}} - 1 \right)} \leq 2ne^{-cm/K_o^4}. \quad (51)$$

provided $m \geq \left(\frac{8\sqrt{2\pi}K_o^2\gamma^{3/2}}{(1-\alpha)\delta} \right) \sqrt{k}(\log n)^{3/2}$, and with $c > 0$ being a universal constant. \square

F. Proof Sketch for Theorem 3

The proof of Theorem 3 is along similar lines as that of Theorem 2; the sketch here focuses on the differences in the steps. We consider the tail event

$$\mathcal{E}_1 \triangleq \left\{ \sup_{\mathbf{z} \in \mathcal{U}_k} \left| \left\| \left(\frac{\mathbf{A}}{\sqrt{m}} \odot \frac{\mathbf{A}}{\sqrt{m}} \right) \mathbf{z} \right\|_2^2 - 1 \right| \geq \delta \right\}, \quad (52)$$

and show that for sufficiently large m , $\mathbb{P}(\mathcal{E}_1)$ can be driven arbitrarily close to zero, thereby implying that δ is a probabilistic upper bound for $\delta_k((\mathbf{A}/\sqrt{m} \odot \mathbf{A}/\sqrt{m}))$. Now, $\mathbb{P}(\mathcal{E}_1)$ admits the following union bound:

$$\begin{aligned} \mathbb{P}(\mathcal{E}_1) &= \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \mathbf{z}^T (\mathbf{A} \odot \mathbf{A})^T (\mathbf{A} \odot \mathbf{A}) \mathbf{z} - m^2 \right| \geq \delta m^2 \right) \\ &= \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \mathbf{z}^T (\mathbf{A}^T \mathbf{A} \circ \mathbf{A}^T \mathbf{A}) \mathbf{z} - m^2 \right| \geq \delta m^2 \right) \\ &= \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1}^n z_i z_j (\mathbf{a}_i^T \mathbf{a}_j)^2 - m^2 \right| \geq \delta m^2 \right) \end{aligned}$$

$$\begin{aligned} &\leq \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n z_i^2 \|\mathbf{a}_i\|_4^2 - m^2 \right| \geq \alpha \delta m^2 \right) \\ &+ \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j (\mathbf{a}_i^T \mathbf{a}_j)^2 \right| \geq (1-\alpha)\delta m^2 \right). \quad (53) \end{aligned}$$

In the above, the second identity follows from Proposition 12. The last inequality uses the triangle inequality followed by the union bound, with $\alpha \in (0, 1)$ being a variational parameter to be optimized later. Similar to the proof of Theorem 2, we now derive separate upper bounds for each of the two probability terms in (53).

The first term in (53) admits the following series of relaxations.

$$\begin{aligned} &\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n z_i^2 \|\mathbf{a}_i\|_2^4 - m^2 \right| \geq \alpha \delta m^2 \right) \\ &\stackrel{(a)}{\leq} \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n z_i^2 \left| \|\mathbf{a}_i\|_2^4 - m^2 \right| \geq \alpha \delta m^2 \right) \\ &\stackrel{(b)}{\leq} \mathbb{P} \left(\max_{1 \leq i \leq n} \left| \|\mathbf{a}_i\|_2^4 - m^2 \right| \geq \alpha \delta m^2 \right) \\ &\stackrel{(c)}{\leq} n\mathbb{P} \left(\left| \|\mathbf{a}_1\|_2^4 - m^2 \right| \geq \alpha \delta m^2 \right). \\ &\stackrel{(d)}{\leq} n\mathbb{P} \left(\left| \|\mathbf{a}_1\|_2^2 - m \right|^2 \geq \alpha \beta \delta m^2 \right) \\ &\quad + 2n\mathbb{P} \left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \frac{\alpha(1-\beta)\delta m}{2} \right) \\ &\stackrel{(e)}{\leq} n\mathbb{P} \left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \sqrt{\alpha \beta \delta m} \right) \\ &\quad + 2n\mathbb{P} \left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \frac{\alpha(1-\beta)\delta m}{2} \right). \\ &\stackrel{(f)}{\leq} 3n\mathbb{P} \left(\left| \|\mathbf{a}_1\|_2^2 - m \right| \geq \frac{\alpha \delta m}{2} \left(1 - \frac{\alpha \delta}{4} \right) \right) \\ &\stackrel{(g)}{\leq} 3ne^{-cm \frac{\alpha^2 \delta^2}{4K_o^4} (1-\alpha \delta/4)^2}. \quad (54) \end{aligned}$$

In the above, step (a) is the triangle inequality. The inequality in step (b) follows from the fact that nonnegative convex combination of n arbitrary numbers is at most the maximum among the n numbers. Step (c) is a union bound. Step (d) is also a union bounding argument with $\beta \in (0, 1)$ as a variational parameter, combined with the fact that for any vector \mathbf{a} , the triangle inequality $\left| \|\mathbf{a}\|_2^4 - m^2 \right| \leq \left| \|\mathbf{a}\|_2^2 - m \right|^2 + 2m \left| \|\mathbf{a}\|_2^2 - m \right|$ is always true. Step (f) is obtained by choosing the union bound parameter $\beta = \alpha \delta/4$. Finally, step (g) is the Hanson-Wright inequality (Theorem 13) applied to the subgaussian vector \mathbf{a}_1 .

We now derive an upper bound for the second probability term in (53). Note that

$$\begin{aligned} &\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j (\mathbf{a}_i^T \mathbf{a}_j)^2 \right| \\ &\leq \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2^2 \left(\sum_{j=1, j \neq i}^n |z_j| u_{ij}^2 \right) \end{aligned}$$

$$\leq \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \right) \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z}' \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z'_j| u_{ij}^2 \right), \quad (55)$$

where $u_{ij} \triangleq \langle \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2}, \mathbf{a}_j \rangle$. The inner product u_{ij} is a subgaussian random variable satisfying $\|u_{ij}\|_{\psi_2} \leq c_4 K$, where c_4 is a universal numerical constant.

Using (55) in (53) yields the following union bound.

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j (\mathbf{a}_i^T \mathbf{a}_j)^2 \right| \geq (1-\alpha) \delta m^2 \right) \\ & \leq \mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z}' \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z'_j| u_{ij}^2 \geq 4\sqrt{2\pi} K_o^2 (\gamma \log n)^{3/2} \right) \\ & \quad + \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \geq \frac{(1-\alpha) \delta m^2}{4\sqrt{2\pi} K_o^2 (\gamma \log n)^{3/2}} \right). \quad (56) \end{aligned}$$

The following Lemmas 6 and 7 provide upper bounds for the two tail probability terms occurring in the RHS of (56).

Lemma 6. *Let \mathbf{A} be an $m \times n$ random matrix as specified in Theorem 3. Let $u_{ij} \triangleq \langle \frac{\mathbf{a}_i}{\|\mathbf{a}_i\|_2}, \mathbf{a}_j \rangle$. Then,*

$$\mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| u_{ij}^2 \geq 4\sqrt{2\pi} K_o^2 (\gamma \log n)^{3/2} \right) \leq \frac{e}{n^{2(c_2\gamma-1)}},$$

where $K_o \triangleq \max(K, 1)$ and $c_2 > 0$ is a universal constant.

Proof. See Appendix G. \square

Lemma 7. *Let \mathbf{A} be an $m \times n$ random matrix with zero mean, unit variance, i.i.d. subgaussian entries satisfying $\|\mathbf{A}_{ij}\|_{\psi_2} \leq K$. Then,*

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{i=1}^n |z_i| \|\mathbf{a}_i\|_2 \geq \frac{(1-\alpha) \delta m^2}{4\sqrt{2\pi} K_o^2 (\gamma \log n)^{3/2}} \right) \leq n e^{-\frac{cm}{K_o^4}}, \quad (57)$$

provided $m \geq \frac{8\sqrt{2\pi} K_o^2 \gamma^{3/2}}{(1-\alpha)\delta} \sqrt{k} \log^{3/2} n$. Here, $c > 0$ is a universal constant.

Proof. The proof is similar to that of Lemma 5, which is given in Appendix E. \square

By using the tail probability bounds in Lemmas 6 and 7 together in (56), we obtain the following upper bound.

$$\begin{aligned} & \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \left| \sum_{i=1}^n \sum_{j=1, j \neq i}^n z_i z_j (\mathbf{a}_i^T \mathbf{a}_j)^2 \right| \geq (1-\alpha) \delta m^2 \right) \\ & \leq \frac{e}{n^{2(c_2\gamma-1)}} + n e^{-cm/K_o^4} = \frac{e}{n^{2(c_2\gamma-1)}} + \frac{1}{n^{\left(\frac{cm}{K_o^4 \log n} - 1\right)}} \\ & \leq \frac{2e}{n^{2(c_2\gamma-1)}}. \quad (58) \end{aligned}$$

The above inequality is true for $m \geq \max \left(\frac{8\sqrt{2\pi} K_o^2 \gamma^{3/2}}{(1-\alpha)\delta} \sqrt{k} \log^{3/2} n, \frac{(2c_2\gamma-1) K_o^4 \log n}{c} \right)$. By combining (54), (56) and (58), we obtain the following tail bound.

$$\mathbb{P}(\mathcal{E}) \leq \frac{3}{n^{\left(\frac{c\alpha^2\delta^2(1-\alpha\delta/4)^2 m}{4K_o^4 \log n} - 1\right)}} + \frac{2e}{n^{2(c_2\gamma-1)}} \leq \frac{4e}{n^{2(c_2\gamma-1)}}, \quad (59)$$

provided $m \geq \max \left(\frac{8\sqrt{2\pi} K_o^2 \gamma^{3/2}}{(1-\alpha)\delta} \sqrt{k} \log^{3/2} n, \frac{(2c_2\gamma-1)\xi K_o^4 \log n}{c} \right)$, where $\xi = \max \left(1, \frac{4}{\alpha^2 \delta^2 (1-\alpha\delta/4)^2} \right)$. Finally, by choosing $\alpha = 1/2$, we have $\mathbb{P}(\mathcal{E}) \leq 4en^{-2(c_2\gamma-1)}$, provided that $m \geq \max \left(\frac{16\sqrt{2\pi} K_o^2 \gamma^{3/2}}{\delta} \sqrt{k} \log^{3/2} n, \frac{(2c_2\gamma-1)\xi K_o^4 \log n}{c} \right)$, with $\xi = \max \left(1, \frac{16}{\delta^2 (1-\delta/8)^2} \right)$. Note that, in terms of k and n , the first term in the inequality for m scales as $\sqrt{k} \log^{3/2} n$; it dominates the second term, which scales as $\log n$.

G. Proof of Lemma 6

Let $\mathcal{W} \subseteq \mathbb{R}^{m \times n}$ be the collection of all $m \times n$ sized matrices \mathbf{X} with columns $\mathbf{x}_i, i \in [n]$, and satisfying $|u_{ij}| \triangleq \frac{|\mathbf{x}_i^T \mathbf{x}_j|}{\|\mathbf{x}_i\|_2} < K\sqrt{2\gamma \log n}$ for all $1 \leq i, j \leq n, j \neq i$. Then, for a subgaussian random matrix \mathbf{A} ,

$$\begin{aligned} \mathbb{P}(\mathbf{A} \notin \mathcal{W}) &= \mathbb{P} \left(\bigcup_{1 \leq i, j \leq n, j \neq i} \left\{ |u_{ij}| \geq K\sqrt{2\gamma \log n} \right\} \right) \\ &\leq \sum_{1 \leq i, j \leq n, j \neq i} \mathbb{P} \left(|u_{ij}| \geq K\sqrt{2\gamma \log n} \right) \\ &\leq \frac{n^2}{2} \mathbb{P} \left(|u_{12}| \geq K\sqrt{2\gamma \log n} \right) \\ &\leq \frac{en^2}{2} e^{-2c_1\gamma \log n} = \frac{e}{2n^{2(c_1\gamma-1)}}, \quad (60) \end{aligned}$$

where $c_1 > 0$ is a universal constant. In the above, the first and second inequalities are due to the union bound and the identically distributed nature of $u_{ij}, j \neq i$. The last inequality is simply the subgaussian tail bound. Thus, for $\gamma \geq 1/c_1$, and for sufficiently large n , $K\sqrt{2\gamma \log n}$ acts as a uniform upper bound for the random variables $|u_{ij}|, j \neq i$ with arbitrarily high probability.

Let \mathcal{E}_3 denote the event that the matrix $\mathbf{A} \in \mathcal{W}$. Then, by law of total probability,

$$\begin{aligned} & \mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| |u_{ij}|^2 \geq t \right) \\ & \leq \mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| |u_{ij}|^2 \geq t \mid \mathcal{E}_3 \right) + \mathbb{P}(\mathcal{E}_3^c) \\ & \leq n \mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}|^2 \geq t \mid \mathcal{E}_3 \right) + \mathbb{P}(\mathcal{E}_3^c). \quad (61) \end{aligned}$$

Conditioned on event \mathcal{E}_3 and the column \mathbf{a}_1 assuming some fixed value, the random variables $|u_{1j}|^2$ are bounded, nonnegative and i.i.d. random variables. By Lemma 3, it follows that

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}|^2 - \mathbb{E} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}|^2 \geq t \mid \mathcal{E}_3, \mathbf{a}_1 \right) \leq e^{-t^2/8K^4\gamma^2 \log^2 n}. \quad (62)$$

Since the above tail probability bound is independent of \mathbf{a}_1 , it remains unchanged after marginalizing the conditional

probability in the LHS over all values of \mathbf{a}_1 , i.e.,

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}|^2 - \mathbb{E} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}|^2 \geq t \mid \mathcal{E}_3 \right) \leq e^{-t^2/8K^4\gamma^2 \log^2 n}. \quad (63)$$

Further, using Proposition 15, the probability term in the LHS of the inequality in (63) can be replaced by a smaller tail probability as shown below.

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{ij}|^2 \geq t + 2\sqrt{2\pi}K^2\gamma \log n \mid \mathcal{E}_3 \right) \leq e^{-t^2/8K^2\gamma^2 \log^2 n}.$$

Setting $t = 4K(\gamma \log n)^{3/2}$, we get

$$\mathbb{P} \left(\sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=2}^n |z_j| |u_{1j}|^2 \geq 4\sqrt{2\pi}K_o^2(\gamma \log n)^{3/2} \mid \mathcal{E}_3 \right) \leq \frac{1}{n^{2\gamma}}. \quad (64)$$

Finally, by combining (60), (61) and (64), we obtain the desired tail bound.

$$\mathbb{P} \left(\max_{1 \leq i \leq n} \sup_{\mathbf{z} \in \mathcal{U}_k} \sum_{j=1, j \neq i}^n |z_j| |u_{ij}|^2 \geq 4\sqrt{2\pi}K_o^2(\gamma \log n)^{3/2} \right) \leq \frac{1}{n^{2\gamma-1}} + \frac{e}{2n^{2(c_1\gamma-1)}} \leq \frac{e}{n^{2(c_2\gamma-1)}},$$

where $c_2 = \min(1, c_1)$ is a universal positive constant.

REFERENCES

- [1] C. G. Khatri and C. R. Rao, "Solutions to some functional equations and their applications to characterization of probability distributions," *Sankhya: The Indian Journal of Statistics, Series A (1961-2002)*, vol. 30, no. 2, pp. 167–180, 1968.
- [2] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas (Second Edition)*. Princeton University Press, 2009.
- [3] S. Liu and G. Trenkler, "Hadamard, Khatri-Rao, Kronecker and other matrix products," *International Journal of Information and System Sciences*, vol. 4, no. 1, pp. 160–177, 2007.
- [4] M. F. Duarte and R. G. Baraniuk, "Kronecker compressive sensing," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 494–504, Feb. 2012.
- [5] A. Koochakzadeh and P. Pal, "Sparse source localization in presence of co-array perturbations," in *International Conference on Sampling Theory and Applications (SampTA)*, May 2015, pp. 563–567.
- [6] 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.
- [7] 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.
- [8] W. K. Ma, T. H. Hsieh, and C. Y. Chi, "DOA estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: A Khatri-Rao subspace approach," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2168–2180, Apr. 2010.
- [9] N. D. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *IEEE Signal Process. Lett.*, vol. 19, no. 11, pp. 757–760, Nov. 2012.
- [10] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Jan. 2006.
- [11] S. Foucart and H. Rauhut, *A Mathematical Introduction to Compressive Sensing*. Birkhäuser Basel, 2013.
- [12] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [13] E. J. Candès and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec 2005.
- [14] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.
- [15] T. T. Cai, L. Wang, and G. Xu, "New bounds for restricted isometry constants," *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4388–4394, Sept 2010.
- [16] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 711–726, Feb. 2015.
- [17] D. P. Wipf and B. D. Rao, "An empirical Bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Transactions on Signal Processing*, vol. 55, no. 7, pp. 3704–3716, Jul. 2007.
- [18] S. Khanna and C. R. Murthy, "On the support recovery of jointly sparse Gaussian sources using sparse Bayesian learning," *CoRR*, vol. abs/1703.04930. [Online]. Available: <http://arxiv.org/abs/1703.04930>
- [19] C. Carathéodory and L. Fejér, "Über den zusammenhang der extremen von harmonischen funktionen mit ihren koefizienten un über den picard-landauschen satz," *Rendiconti del Circolo Matematico di Palermo*, vol. 32, no. 1, pp. 218–239, 1911.
- [20] P. Stoica and R. Moses, *Spectral Analysis of Signals*. Pearson Prentice Hall, 2005.
- [21] Z. Yang, L. Xie, and P. Stoica, "Vandermonde decomposition of multilevel Toeplitz matrices with application to multidimensional super-resolution," *IEEE Trans. Inf. Theory*, vol. 62, no. 6, pp. 3685–3701, Jun. 2016.
- [22] N. D. Sidiropoulos and A. Kyrillidis, "Multi-way compressed sensing for sparse low-rank tensors," *IEEE Signal Processing Letters*, vol. 19, no. 11, pp. 757–760, Nov. 2012.
- [23] D. Nion and N. D. Sidiropoulos, "A PARAFAC-based technique for detection and localization of multiple targets in a MIMO radar system," in *Proc. ICASSP*, 2009, pp. 2077–2080.
- [24] A. M. Tillmann and M. E. Pfetsch, "The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1248–1259, Feb. 2014.
- [25] R. A. Horn and C. R. Johnson, Eds., *Matrix Analysis*. Cambridge University Press, 1986.
- [26] S. Jokar, "Sparse recovery and Kronecker products," in *44th Annual Conference on Information Sciences and Systems*, Mar. 2010, pp. 1–4.
- [27] A. Bhaskara, M. Charikar, A. Moitra, and A. Vijayaraghavan, "Smoothed analysis of tensor decompositions," in *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, 2014, pp. 594–603.
- [28] J. Anderson, M. Belkin, N. Goyal, L. Rademacher, and J. R. Voss, "The more, the merrier: the blessing of dimensionality for learning large Gaussian mixtures," in *27th Annual Conference on Learning Theory*, Jun. 2014, pp. 1135–1164.
- [29] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N-way arrays," *Journal of Chemometrics*, vol. 14, no. 3, pp. 229–239, 2000.
- [30] P. Koiran and A. Zouzias, "Hidden cliques and the certification of the restricted isometry property," *IEEE Trans. Inf. Theory*, vol. 60, no. 8, pp. 4999–5006, Aug. 2014.
- [31] S. Khanna and C. R. Murthy, "Rényi divergence based covariance matching pursuit of joint sparse support," in *Sig. Process. Advances in Wireless Commun. (SPAWC)*, Jul. 2017, pp. 1 – 5.
- [32] R. Vershynin, "Introduction to the non-asymptotic analysis of random matrices," 2010.
- [33] J. A. Tropp, "Greed is good: algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [34] G. Visick, "A quantitative version of the observation that the Hadamard product is a principal submatrix of the Kronecker product," *Linear Algebra and its Applications*, vol. 304, no. 13, pp. 45 – 68, 2000.
- [35] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge; New York: Cambridge University Press, 1994.
- [36] B. Mond and J. E. Pečarić, "Inequalities for the Hadamard product of matrices," *SIAM J. Matrix Anal. Appl.*, vol. 19, no. 1, pp. 66–70, Jan. 1998.
- [37] H. J. Werner, "Hadamard product of square roots of correlation matrices," *Image*, vol. 26, pp. 1–32, Apr. 2001.
- [38] A. W. Marshall and I. Olkin, "Reversal of the Lyapunov, Hölder, and Minkowski inequalities and other extensions of the Kantorovich inequality," *Journal of Mathematical Analysis and Applications*, vol. 8, no. 3, pp. 503 – 514, 1964.

- [39] S. Liu and H. Neudecker, "Several matrix Kantorovich-type inequalities," *Journal of Mathematical Analysis and Applications*, vol. 197, no. 1, pp. 23 – 26, 1996.
- [40] —, "Kantorovich inequalities and efficiency comparisons for several classes of estimators in linear models," *Statistica Neerlandica*, vol. 51, no. 3, pp. 345–355, 1997.
- [41] S. Liu, "On the Hadamard product of square roots of correlation matrices," *Econometric Theory*, vol. 18, no. 4, p. 1007, Aug. 2002.
- [42] C. R. Rao and M. B. Rao, *Matrix Algebra and Its Applications to Statistics and Econometrics*. Singapore: World Scientific, 1998.
- [43] R. Adamczak, "A note on the Hanson-Wright inequality for random vectors with dependencies," *Electron. Commun. Probab.*, vol. 20, pp. 72–84, 2015.
- [44] M. Rudelson and R. Vershynin, "Hanson-Wright inequality and sub-gaussian concentration," *Electron. Commun. Probab.*, vol. 18, pp. 82–91, 2013.
- [45] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.