Math. Stat. Learn. (submitted)

Completion of Matrices with Low Description Complexity

Erwin Riegler, Günther Koliander, David Stotz, and Helmut Bölcskei

Abstract. We propose a theory for matrix completion that goes beyond the low-rank structure commonly considered in the literature and applies to general matrices of low description complexity. Specifically, complexity of the sets of matrices encompassed by the theory is measured in terms of Hausdorff and upper Minkowski dimensions. Our goal is the characterization of the number of linear measurements, with an emphasis on rank-1 measurements, needed for the existence of an algorithm that yields reconstruction, either perfect, with probability 1, or with arbitrarily small probability of error, depending on the setup. Concretely, we show that matrices taken from a set $\mathcal{U}$ such that $\mathcal{U} - \mathcal{U}$ has Hausdorff dimension s can be recovered from $k > s$ measurements, and random matrices supported on a set $\mathcal{U}$ of Hausdorff dimension s can be recovered with probability 1 from $k > s$ measurements. What is more, we establish the existence of recovery mappings that are robust against additive perturbations or noise in the measurements. Concretely, we show that there are β-Hölder continuous mappings recovering matrices taken from a set of upper Minkowski dimension s from $k > 2s/(1-\beta)$ measurements and, with arbitrarily small probability of error, random matrices supported on a set of upper Minkowski dimension s from $k > s/(1-\beta)$ measurements. The numerous concrete examples we consider include low-rank matrices, sparse matrices, QR decompositions with sparse R-components, and matrices of fractal nature.

Funding. The work of G. Koliander was supported by the Vienna Science and Technology Fund (WWTF): MA16-053.

1. Introduction

Matrix completion refers to the recovery of a low-rank matrix from a (small) subset of its entries or a (small) number of linear combinations thereof. This problem arises in a wide range of applications including quantum state tomography, face recognition, recommender systems, and sensor localization (see, e.g., [8,12] and references therein).

2020 Mathematics Subject Classification. Primary 68P30; Secondary 15A83.
Keywords. Matrix completion, compressed sensing, geometric measure theory, Hausdorff dimension, Minkowski dimension, rank-1 measurements, recurrent iterated function system.

The formal problem statement is as follows. Suppose we have k linear measurements of the matrix $\boldsymbol{X} \in \mathbb{R}^{m\times n}$ with $rank(\boldsymbol{X}) \leq r$ in the form of

$$\boldsymbol{y} = (\langle \boldsymbol{A}_1, \boldsymbol{X}\rangle \ \ldots \ \langle \boldsymbol{A}_k, \boldsymbol{X}\rangle)^{\mathsf{T}}, \tag{1.1}$$

where $\langle \boldsymbol{A}_i, \boldsymbol{X}\rangle = tr(\boldsymbol{A}_i^{\mathsf{T}}\boldsymbol{X})$ is the standard trace inner product between matrices; the $\boldsymbol{A}_i \in \mathbb{R}^{m\times n}$ are referred to as measurement matrices. The number of measurements, k, is typically much smaller than the total number of entries, mn, of $\boldsymbol{X}$. Depending on the $\boldsymbol{A}_i$, the measurements can simply be individual entries of $\boldsymbol{X}$ or general linear combinations thereof. Can we recover $\boldsymbol{X}$ from $\boldsymbol{y}$?

The vast literature on matrix completion (for a highly incomplete list see [5–10, 12, 18, 20–22, 28, 29, 34]) provides thresholds on the number of measurements needed for successful recovery of the unknown low-rank matrix $\boldsymbol{X}$, under various assumptions on the measurement matrices $\boldsymbol{A}_i$ and the low-rank models generating $\boldsymbol{X}$. For instance, in [18] the $\boldsymbol{A}_i$ are chosen randomly from a fixed orthonormal (w.r.t. the trace inner product) basis for $\mathbb{R}^{n\times n}$ and it is shown that an unknown $n \times n$ matrix $\boldsymbol{X}$ of rank no more than r can be recovered with high probability if $k \geq O(nr\nu \ln^2 n)$. Here, ν denotes the coherence [18, Def. 1] between the unknown matrix $\boldsymbol{X}$ and the orthonormal basis the $\boldsymbol{A}_i$ are drawn from.

The setting in [9] assumes random[1] measurement matrices A_i with the position of the only nonzero entry, which is equal to one, chosen uniformly at random. It is shown that almost all (a.a.) matrices $\boldsymbol{X}$ (where a.a. is with respect to the random orthogonal model [9, Def. 2.1]) of rank no more than r can be recovered with high probability (with respect to the measurement matrices) provided that the number of measurements satisfies $k \geq Cn^{1.25} r \ln n$, where C is a constant.

In [8] it is shown that for random measurement matrices A_i satisfying a certain concentration property, matrices $\boldsymbol{X}$ of rank no more than r can be recovered with high probability from $k \geq C(m + n)r$ measurements, where C is a constant.

The results on recovery thresholds reviewed so far as well as those in [5,6,10,28,29] all pertain to recovery through nuclear norm minimization.

In [14] measurement matrices A_i containing i.i.d. entries drawn from an absolutely continuous (with respect to Lebesgue measure) distribution are considered. It is shown that rank minimization (which is NP-hard, in general) recovers $n \times n$ matrices $\boldsymbol{X}$ of rank no more than r with probability 1 if $k > (2n - r)r$. Furthermore, it is established that all matrices $\boldsymbol{X}$ of rank no more than $n/2$ can be recovered, again through rank minimization and with probability 1, provided that $k \geq 4nr - 4r^2$.

The recovery thresholds in [14],[8] do not exhibit a $(\log n)$-factor, but assume significant richness in the random measurement matrices A_i. Storing and applying the

[1] We indicate random quantities by roman sans-serif letters such as A.

realizations of such measurement matrices is costly in terms of memory and computation time, respectively. To alleviate this problem, [6] considers rank-1 measurement matrices of the form $\mathbf{A}_i = \mathbf{a}_i \mathbf{b}_i^\top$, where the random vectors $\mathbf{a}_i \in \mathbb{R}^m$ and $\mathbf{b}_i \in \mathbb{R}^n$ are independent with i.i.d. Gaussian or sub-Gaussian entries; it is shown that nuclear norm minimization succeeds under the same recovery threshold as in [8], namely $k \geq C(m+n)r$ for some constant C.

The recovery of matrices that, along with the measurement matrices, belong to algebraic varieties was studied in [32, 35, 38]. As a byproduct, it is shown that almost all rank-r matrices in $\mathbb{R}^{m \times n}$ can be recovered from $k > (m+n-r)r$ measurements taken with measurement matrices of arbitrary rank.

Finally, the application of recent results on analog signal compression [1,2,33,37] to matrix completion yields recovery thresholds for a.a. measurement matrices $\boldsymbol{A}_i$ and random matrices $\mathbf{X}$ that have low description complexity in the sense of [1]. Specifically, the results in [1] can be transferred to matrix completion by writing the trace inner product $\langle \boldsymbol{A}_i, \boldsymbol{X} \rangle$ as the standard inner product between the vectorized matrices $\boldsymbol{A}_i$ and $\boldsymbol{X}$ (obtained by stacking the columns). The definition of "low description complexity" as put forward in [1] goes beyond the usual assumption of $\mathbf{X}$ having low rank. It essentially says that the matrix takes value in some low-dimensional[2] set $\mathcal{U} \subseteq \mathbb{R}^{m \times n}$ with probability 1. This set $\mathcal{U}$ can, for example, be the set of all matrices with rank no more than r, but much more general structures are possible.

Contributions. The purpose of this paper is to establish fundamental recovery thresholds (i.e., thresholds not restricted to a certain recovery scheme) for rank-1 measurement matrices $\boldsymbol{A}_i = \boldsymbol{a}_i \boldsymbol{b}_i^\top$ applied to data matrices $\boldsymbol{X}$ taking value in low-dimensional sets $\mathcal{U} \subseteq \mathbb{R}^{m \times n}$. Rank-1 measurement matrices are practically relevant due to reduced storage requirements and lower computational complexity in the evaluation of the trace inner product $\langle \boldsymbol{a}_i \boldsymbol{b}_i^\top, \boldsymbol{X} \rangle = \boldsymbol{a}_i^\top \boldsymbol{X} \boldsymbol{b}_i$. We consider both deterministic data matrices $\boldsymbol{X}$ with associated recovery guarantees for all $\boldsymbol{X} \in \mathcal{U}$ and random $\mathbf{X}$ accompanied by recovery guarantees either with probability 1 or with arbitrarily small probability of error. The recovery thresholds we obtain are in terms of the Hausdorff dimension of the support set $\mathcal{U}$ of the data matrices. Furthermore, we establish bounds–in terms of the upper Minkowski dimension of $\mathcal{U}$–on the number of measurements needed to guarantee Hölder continuous recovery, and hence robustness against additive perturbations or noise. Hausdorff and upper Minkowski dimension are particularly easy to characterize for countably rectifiable and rectifiable sets [16], respectively. These concepts comprise most practically relevant data structures such as low rank and sparsity in terms

[2] The precise dimension measures used in the definition of low description complexity in [1, 2, 33, 37] are, depending on the context, lower modified Minkowski dimension, upper Minkowski dimension, or Hausdorff dimension.

of the number of nonzero entries as well as the Kronecker product, matrix product, or sum of any such matrices. As an example of sets that do not fall into the rich class of rectifiable sets, but our theory still applies to, we consider sets of fractal nature. Specifically, we investigate attractor sets of recurrent iterated function systems [3].

Recovery thresholds for general (as opposed to rank-1) measurement matrices follow in a relatively straightforward manner through vectorization from the theory of lossless analog compression as developed in [1]. For the reader's convenience, we shall describe these extensions in brief wherever appropriate. We finally note that a preliminary version of part of the work reported in the present paper, specifically weaker results for more restrictive sets $\mathcal{U}$ of bounded Minkowski dimension and without statements on Hölder-continuous recovery, was presented in [30] by a subset of the authors.

Organization of the paper. In Section 2, we present recovery thresholds for matrices $\boldsymbol{X}$ taking value in a general set $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$. These results are formulated in terms of Hausdorff and upper Minkowski dimension of $\mathcal{U}$. In Section 3, we introduce the concept of rectifiable and countably rectifiable sets from geometric measure theory [16] and we characterize the upper Minkowski and Hausdorff dimensions of such sets. Furthermore, it is shown that many practically relevant sets of structured matrices are (countably) rectifiable. The particularization of our general recovery thresholds to the rectifiable case concludes this section. In Section 4, we particularize our general recovery thresholds to attractor sets of recurrent iterated function systems. The proofs of our main results, Theorems 2.1 and 2.2 and Propositions 2.1 and 2.2 are contained in Sections 5–8.

Notation. Capitalized boldface letters $\boldsymbol{A}, \boldsymbol{B}, \dots$ designate deterministic matrices and lowercase boldface letters $\boldsymbol{a}, \boldsymbol{b}, \dots$ stand for deterministic vectors. We use roman sans-serif letters for random quantities (e.g., x for a random vector and A for a random matrix). Random quantities are assumed to be defined on the Borel σ-algebra of the underlying space. $\mathrm{P}[\mathsf{X} \in \mathcal{U}]$ denotes the probability of X being in the Borel set $\mathcal{U}$. We write λ for Lebesgue measure. The superscript $^\mathsf{T}$ stands for transposition. The ordered singular values of a matrix $\boldsymbol{A}$ are denoted by $\sigma_1(\boldsymbol{A}) \geq \dots \geq \sigma_n(\boldsymbol{A})$. We write $\boldsymbol{A} \otimes \boldsymbol{B}$ for the Kronecker product of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, denote the trace of $\boldsymbol{A}$ by $tr(\boldsymbol{A})$, and let $\langle \boldsymbol{A}, \boldsymbol{B} \rangle = tr(\boldsymbol{A}^\mathsf{T}\boldsymbol{B})$ be the trace inner product of $\boldsymbol{A}$ and $\boldsymbol{B}$. Further, $\|\boldsymbol{A}\|_2 = \sqrt{\langle \boldsymbol{A}, \boldsymbol{A} \rangle}$ and $\|\boldsymbol{A}\|_0$ refers to the number of nonzero entries of $\boldsymbol{A}$. For the Euclidean space $(\mathbb{R}^k, \|\cdot\|_2)$, we denote the open ball of radius s centered at $\boldsymbol{u} \in \mathbb{R}^k$ by $\mathcal{B}_k(\boldsymbol{u}, s)$, $V(k, s)$ stands for its volume. Similarly, forthe Euclidean space $(\mathbb{R}^{m\times n}, \|\cdot\|_2)$, the open ball of radius s centered at $\boldsymbol{U} \in \mathbb{R}^{m\times n}$ is $\mathcal{B}_{m\times n}(\boldsymbol{U}, s)$. We set $\mathcal{M}_r^{m\times n} = \{\boldsymbol{X} \in \mathbb{R}^{m\times n} : rank(\boldsymbol{X}) \leq r\}$ and let $\mathcal{A}_s^{m\times n} = \{\boldsymbol{X} \in \mathbb{R}^{m\times n} : \|\boldsymbol{X}\|_0 \leq s\}$. The closure of the set $\mathcal{U}$ is denoted by $\overline{\mathcal{U}}$. The Cartesian product of the sets $\mathcal{A}$ and $\mathcal{B}$ in Euclidean space is written as $\mathcal{A} \times \mathcal{B}$ and their Minkowski difference is designated by $\mathcal{A} - \mathcal{B}$. The indicator function of a set $\mathcal{U}$ is designated by $\chi_{\mathcal{U}}$. For a bounded set $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ and $\delta > 0$, we denote the covering number of $\mathcal{U}$ by

$$N_\delta(\mathcal{U}) = \min\left\{N \in \mathbb{N} : \exists \boldsymbol{Y}_1, \dots, \boldsymbol{Y}_N \in \mathcal{U} \text{ s.t. } \mathcal{U} \subseteq \bigcup_{i=1}^{N} \mathcal{B}_{m\times n}(\boldsymbol{Y}_i, \delta)\right\}. \tag{1.2}$$

For $\mathcal{U}, \mathcal{V} \subseteq \mathbb{R}^{m\times n}$, we set $\mathcal{U} - \mathcal{V} = \{\boldsymbol{U} - \boldsymbol{V} : \boldsymbol{U} \in \mathcal{U}, \boldsymbol{V} \in \mathcal{V}\}$. We let $\dim_{\mathrm{H}}(\cdot)$ denote the Hausdorff dimension [15, Equation (3.10)], defined by

$$\dim_{\mathrm{H}}(\mathcal{U}) = \inf\{s \geq 0 : \mathscr{H}^s(\mathcal{U}) = 0\}, \tag{1.3}$$

where $\mathscr{H}^s(\mathcal{U}) = \lim_{\delta\to 0} \inf\left\{\sum_{i=1}^{\infty} \operatorname{diam}^s(\mathcal{U}_i) : \{\mathcal{U}_i\} \text{ is a } \delta\text{-cover of } \mathcal{U}\right\}$ is the s-dimensional Hausdorff measure of $\mathcal{U}$ [15, Equation (3.2)].

Furthermore, $\overline{\dim}_{\mathrm{B}}(\cdot)$ and $\underline{\dim}_{\mathrm{B}}(\cdot)$ refer to the upper and lower Minkowski dimension [15, Definition 2.1] defined as

$$\overline{\dim}_{\mathrm{B}}(\mathcal{U}) = \limsup_{\delta\to 0} \frac{\log N_\delta(\mathcal{U})}{\log(1/\delta)} \tag{1.4}$$

and

$$\underline{\dim}_{\mathrm{B}}(\mathcal{U}) = \liminf_{\delta\to 0} \frac{\log N_\delta(\mathcal{U})}{\log(1/\delta)}, \tag{1.5}$$

respectively. Finally, we note that [15, Proposition 3.4]

$$\dim_{\mathrm{H}}(\mathcal{U}) \leq \underline{\dim}_{\mathrm{B}}(\mathcal{U}) \leq \overline{\dim}_{\mathrm{B}}(\mathcal{U}) \tag{1.6}$$

for all nonempty subsets in Euclidean spaces.

2. Main Results

Our first main result provides a threshold for recovery of matrices $\boldsymbol{X}$ from rank-1 measurements in a very general setting. Specifically, the matrices $\boldsymbol{X}$ are assumed to take value in some set $\mathcal{U}$ and the recovery threshold is in terms of either $\dim_{\mathrm{H}}(\mathcal{U})$ or $\dim_{\mathrm{H}}(\mathcal{U} - \mathcal{U})$.

Theorem 2.1. *For every nonempty set* $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$*, the following holds:*

i) *The mapping*

$$\mathcal{U} \to \mathbb{R}^k \tag{2.1}$$

$$\boldsymbol{X} \mapsto (\boldsymbol{a}_1^\top \boldsymbol{X} \boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^\top \boldsymbol{X} \boldsymbol{b}_k)^\top \tag{2.2}$$

is one-to-one for Lebesgue a.a. $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$ *provided that* $\dim_{\mathrm{H}}(\mathcal{U} - \mathcal{U}) < k$.

ii) *Suppose that* $\mathcal{U}$ *is Borel and consider an* $m \times n$ *random matrix* X *satisfying* $\mathrm{P}[\mathsf{X} \in \mathcal{U}] = 1$. *Then, for Lebesgue a.a.* $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$, *there exists a Borel-measurable mapping* $g \colon \mathbb{R}^k \to \mathbb{R}^{m\times n}$ *satisfying*

$$\mathrm{P}\left[g\Big((\boldsymbol{a}_1^{\mathsf{T}}\mathsf{X}\boldsymbol{b}_1 \;\dots\; \boldsymbol{a}_k^{\mathsf{T}}\mathsf{X}\boldsymbol{b}_k)^{\mathsf{T}}\Big) \neq \mathsf{X}\right] = 0 \tag{2.3}$$

provided that $\dim_{\mathrm{H}}(\mathcal{U}) < k$.

Proof. See Section 5. ■

The first part of the theorem states that in the deterministic case, $k > \dim_{\mathrm{H}}(\mathcal{U} - \mathcal{U})$ rank-1 measurements suffice for unique recovery of $\boldsymbol{X} \in \mathcal{U}$ (except for measurement vectors $\boldsymbol{a}_i, \boldsymbol{b}_i$ supported on a Lebesgue null-set). In the probabilistic setting of the second part, $k > \dim_{\mathrm{H}}(\mathcal{U})$ measurements suffice for the existence of a Borel-measurable recovery mapping achieving zero error. While these are the most general versions of our recovery results, it can be difficult to evaluate $\dim_{\mathrm{H}}(\mathcal{U})$ and $\dim_{\mathrm{H}}(\mathcal{U} - \mathcal{U})$ for sets $\mathcal{U}$ with interesting structural properties. In Section 3, we shall see, however, that these dimensions are easily computed for rectifiable sets, which, in turn, encompass many structures of practical relevance such as low rank or sparsity.

A version of Theorem 2.1 in terms of lower Minkowski dimension instead of Hausdorff dimension was found by a subset of the authors of the present paper in [30, Theorem 2] and was subsequently extended by Li et al. to the complex-valued case in [25, Theorem 8]. As lower Minkowski dimension is always greater than or equal to Hausdorff dimension (see (1.6)), Theorem 2.1 strengthens [30, Theorem 2]. In addition, lower Minkowski dimension is defined for bounded sets $\mathcal{U}$ only, a restriction not shared by Hausdorff dimension.

A vectorization argument, concretely, stacking the columns of the data matrix $\boldsymbol{X}$ and the measurement matrices $\boldsymbol{A}_i$, shows that [2, Theorem 3.7] implies Item i) in Theorem 2.1 and [2, Corollary 3.4] implies Item ii), in both cases, however, for the rank-1 measurement matrices $\boldsymbol{a}_i\boldsymbol{b}_i^{\mathsf{T}}$ replaced by generic measurement matrices $\boldsymbol{A}_i \in \mathbb{R}^{m\times n}$. As the set of rank-1 matrices is a null-set when viewed as a subset of $\mathbb{R}^{m\times n}$, these results do not imply our Theorem 2.1. Furthermore, the technical challenges in establishing Theorem 2.1 are quite different from those encountered in [2], which, in turn, builds on [1]. In particular, here we need a stronger concentration of measure inequality (see Lemma 6.2 in Section 6).

The proof of Theorem 2.1 detailed in Section 5 is based on the following result, which is similar in spirit to the null-space property in compressed sensing theory [17, Theorem 2.13]:

Proposition 2.1. *Consider a nonempty set* $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ *with* $\dim_{\mathrm{H}}(\mathcal{U}) < k$. *Then,*

$$\{\boldsymbol{X} \in \mathcal{U}\backslash\{\boldsymbol{0}\} : (\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_1 \;\dots\; \boldsymbol{a}_k^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_k)^{\mathsf{T}} = \boldsymbol{0}\} = \emptyset, \tag{2.4}$$

for Lebesgue a.a. $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$.

Proof. See Section 6. ∎

Our second main result establishes thresholds for Hölder-continuous recovery, that is, recovery which exhibits robustness against additive perturbations or noise. Here, we have to impose the stricter technical condition of bounded upper Minkowski dimension $\overline{\dim}_{\mathrm{B}}(\mathcal{U})$ and, in turn, can only consider bounded sets $\mathcal{U}$.

Theorem 2.2. *For every nonempty and bounded set* $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ *and* $\beta \in (0, 1)$*, the following holds:*

i) *Suppose that*

$$\frac{\overline{\dim}_{\mathrm{B}}(\mathcal{U} - \mathcal{U})}{1-\beta} < k. \tag{2.5}$$

Then, for Lebesgue a.a. $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$*, there exists a* β*-Hölder continuous mapping* $g\colon \mathbb{R}^k \to \mathbb{R}^{m\times n}$ *satisfying*

$$g\Big((\boldsymbol{a}_1^\top \boldsymbol{X} \boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^\top \boldsymbol{X} \boldsymbol{b}_k)^\top\Big) = \boldsymbol{X}, \quad \textit{for all } \boldsymbol{X} \in \mathcal{U}. \tag{2.6}$$

ii) *Suppose that* $\mathcal{U}$ *is Borel with*

$$\frac{\overline{\dim}_{\mathrm{B}}(\mathcal{U})}{1-\beta} < k. \tag{2.7}$$

Fix $\varepsilon > 0$ *arbitrarily and consider an* $m \times n$ *random matrix* X *with* $\mathrm{P}[\mathsf{X} \in \mathcal{U}] = 1$*. Then, for Lebesgue a.a.* $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$*, there exists a* β*-Hölder continuous mapping* $g\colon \mathbb{R}^k \to \mathbb{R}^{m\times n}$ *satisfying*

$$\mathrm{P}\Big[g\Big((\boldsymbol{a}_1^\top \mathsf{X} \boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^\top \mathsf{X} \boldsymbol{b}_k)^\top\Big) \neq \mathsf{X}\Big] \le \varepsilon. \tag{2.8}$$

Proof. See Section 7. ∎

Again, the first part of the theorem concerns deterministic data matrices $\boldsymbol{X}$ for which $k > \overline{\dim}_{\mathrm{B}}(\mathcal{U} - \mathcal{U})/(1-\beta)$ rank-1 measurements (except for measurement vectors supported on a Lebesgue null-set) guarantee β-Hölder continuous recovery. The higher the desired Hölder exponent β, the larger the number of measurements has to be. In the probabilistic setting of the second part of the theorem, $k > \overline{\dim}_{\mathrm{B}}(\mathcal{U})/(1-\beta)$ measurements suffice for β-Hölder continuous recovery. We hasten to add that recovery is only with probability $1-\varepsilon$, where, however, ε can be arbitrarily small. Also note that the number of measurements, k, is independent of ε. We shall evaluate $\overline{\dim}_{\mathrm{B}}(\mathcal{U})$ for several rectifiable sets with interesting structural properties in Section 3 and for attractor sets of recurrent iterated function systems in Section 4.

A version of Theorem 2.2 for the rank-1 measurements replaced by measurements taken with general matrices can be obtained from results available in the literature. Specifically, the equivalent of Item i) in Theorem 2.2 follows from [31, Theorem 4.3], that of Item ii) is obtained from [33, Theorem 2], in both cases by vectorization.

The proof of Theorem 2.2 is again based on a variant of the null-space property as used in compressed sensing theory, concretely on the following result:

Proposition 2.2. *Consider a nonempty and bounded set $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$, and suppose that there exists a $\beta \in (0,1)$ such that*

$$\frac{\overline{\dim}_{\mathrm{B}}(\mathcal{U})}{1-\beta} < k. \tag{2.9}$$

Then,

$$\inf\left\{\frac{\|(\boldsymbol{a}_1^\mathsf{T}\boldsymbol{X}\boldsymbol{b}_1\ \ldots\ \boldsymbol{a}_k^\mathsf{T}\boldsymbol{X}\boldsymbol{b}_k)^\mathsf{T}\|_2}{\|\boldsymbol{X}\|_2^{1/\beta}} : \boldsymbol{X} \in \mathcal{U}\backslash\{\boldsymbol{0}\}\right\} > 0, \tag{2.10}$$

for Lebesgue a.a. $((\boldsymbol{a}_1 \ldots \boldsymbol{a}_k), (\boldsymbol{b}_1 \ldots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$.

Proof. See Section 6. ■

Regarding converse statements, i.e., the question of whether too few rank-1 measurements of a given random matrix X necessarily render unique reconstruction impossible, we note that [1, Corollary IV.2] allows a partial answer. Specifically, the simple characterization of the support set $\mathcal{U}$ of X through its dimension $\overline{\dim}_{\mathrm{B}}(\mathcal{U})$ does not enable a general impossibility result. If one assumes, however, that the vectorized version of X is k-analytic according to [1, Definition IV.2], then we can conclude that fewer than k measurements necessarily lead to reconstruction of X being impossible, with probability 1. This statement holds for arbitrary measurement matrices, so in particular also for rank-1 matrices.

3. Rectifiable Sets

To illustrate the practical applicability of the general recovery thresholds obtained in Theorems 2.1 and 2.2 and expressed in terms of Hausdorff and upper Minkowski dimension, we first introduce the concept of rectifiable sets, a central element of geometric measure theory [16]. The relevance of rectifiability derives itself from the fact that a broad class of structured data matrix support sets we are interested in turns out to be rectifiable. In addition, Hausdorff and upper Minkowski dimensions of rectifiable sets have been characterized in significant detail in the literature.

We start with the formal definition of rectifiable sets.

Definition 3.1. [16, Definition 3.2.14] *For $s \in \mathbb{N}$, the set $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ is*

i) *s-rectifiable if there exist a nonempty and compact set $\mathcal{A} \subseteq \mathbb{R}^s$ and a Lipschitz mapping $\varphi\colon \mathcal{A} \to \mathbb{R}^{m\times n}$ such that $\mathcal{U} = \varphi(\mathcal{A})$;*

ii) *countably s-rectifiable if it is the countable union of s-rectifiable sets;*

iii) *countably $\mathscr{H}^s$-rectifiable if it is $\mathscr{H}^s$-measurable and there exists a countably s-rectifiable set $\mathcal{V} \subseteq \mathbb{R}^{m\times n}$ such that $\mathscr{H}^s(\mathcal{U} \setminus \mathcal{V}) = 0$.*

We have the following obvious chain of implications:

$$s\text{-rectifiable} \Rightarrow \text{countably } s\text{-rectifiable} \Rightarrow \text{countably } \mathscr{H}^s\text{-rectifiable}.$$

Countably $\mathscr{H}^s$-rectifiable sets thus constitute the most general class.

We proceed to state preparatory results, which will be used later to establish that many practically relevant sets of structured matrices are (countably) s-rectifiable and to quantify the associated rectifiability parameter s.

Lemma 3.1. *(Properties of s-rectifiable sets)*

i) *If $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ is s-rectifiable, then it is t-rectifiable for all $t \in \mathbb{N}$ with $t > s$.*

ii) *For $\mathcal{U}_i \subseteq \mathbb{R}^{m\times n}$ s_i-rectifiable with $s_i \leq s$, $i = 1, \ldots, N$, the set*

$$\mathcal{U} = \bigcup_{i=1}^{N} \mathcal{U}_i \tag{3.1}$$

is s-rectifiable. In particular, the finite union of s-rectifiable sets is s-rectifiable.

iii) *If $\mathcal{U} \subseteq \mathbb{R}^{m_1\times n_1}$ is s-rectifiable and $\mathcal{V} \subseteq \mathbb{R}^{m_2\times n_2}$ is t-rectifiable, then $\mathcal{U} \times \mathcal{V}$ is $(s+t)$-rectifiable.*

iv) *Every compact subset of an s-dimensional C^1-submanifold* [23, Definition 5.3.1] *of $\mathbb{R}^{m\times n}$ is s-rectifiable.*

Proof. See Appendix A. ■

Lemma 3.2. *(Properties of countably s-rectifiable sets)*

i) *If $\mathcal{U} \subseteq \mathbb{R}^{m_1\times n_1}$ is countably s-rectifiable and $\mathcal{V} \subseteq \mathbb{R}^{m_2\times n_2}$ is countably t-rectifiable, then $\mathcal{U} \times \mathcal{V}$ is countably $(s+t)$-rectifiable.*

ii) *For $\mathcal{U}_i \subseteq \mathbb{R}^{m\times n}$ countably s_i-rectifiable with $s_i \leq s$, $i \in \mathbb{N}$, the set*

$$\mathcal{U} = \bigcup_{i\in\mathbb{N}} \mathcal{U}_i \tag{3.2}$$

is countably s-rectifiable.

iii) *Every s-dimensional C^1-submanifold* [23, Definition 5.3.1] *of $\mathbb{R}^{m\times n}$ is countably s-rectifiable. In particular, every s-dimensional affine subspace of $\mathbb{R}^{m\times n}$ is countably s-rectifiable.*

Proof. Follows from [1, Lemma III.1]. ∎

To establish the rectifiability of structured matrices obtained as products or sums of structured matrices, we need to understand the impact of continuous mappings on rectifiability. Specifically, we shall need the following result from [1, Lemma III. 3] for locally-Lipschitz mappings, i.e., functions that are Lipschitz continuous on all compact subsets:

Lemma 3.3. *Let $\mathcal{U} \subseteq \mathbb{R}^{m_1\times n_1}$ and let $f\colon \mathbb{R}^{m_1\times n_1} \to \mathbb{R}^{m_2\times n_2}$ be a locally-Lipschitz mapping.*

i) *If $\mathcal{U}$ is s-rectifiable, then $f(\mathcal{U})$ is s-rectifiable.*

ii) *If $\mathcal{U}$ is countably s-rectifiable, then $f(\mathcal{U})$ is countably s-rectifiable.*

We will mainly use the following generalization of Lemma 3.3:

Lemma 3.4. *Consider a locally-Lipschitz mapping $f\colon \bigtimes_{i=1}^{N} \mathbb{R}^{m_i\times n_i} \to \mathbb{R}^{m\times n}$, and suppose that $\mathcal{U}_i \subseteq \mathbb{R}^{m_i\times n_i}$, for $i = 1, \dots, N$.*

i) *If $\mathcal{U}_i$ is s_i-rectifiable, for $i = 1, \dots, N$, then $f(\mathcal{U}_1 \times \cdots \times \mathcal{U}_N)$ is s-rectifiable with $s = \sum_{i=1}^{N} s_i$.*

ii) *If $\mathcal{U}_i$ is countably s_i-rectifiable, for $i = 1, \dots, N$, then $f(\mathcal{U}_1 \times \cdots \times \mathcal{U}_N)$ is countably s-rectifiable with $s = \sum_{i=1}^{N} s_i$.*

Proof. Item i) follows from Item iii) of Lemma 3.1 and Item i) of Lemma 3.3, and Item ii) follows from Item i) of Lemma 3.2 and Item ii) of Lemma 3.3. ∎

Before we can particularize our results in Theorems 2.1 and 2.2, it remains to characterize the Hausdorff dimension and the upper Minkowski dimension of rectifiable sets in terms of their rectifiability parameters.

Lemma 3.5. *Let $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ be nonempty. Then, the following properties hold:*

i) *If $\mathcal{U}$ is countably $\mathscr{H}^s$-rectifiable, then*

$$\dim_{\mathrm{H}}(\mathcal{U}) \le s. \tag{3.3}$$

ii) *If $\mathcal{U} \subseteq \mathcal{V}$ with $\mathcal{V} \subseteq \mathbb{R}^{m\times n}$ s-rectifiable, then*

$$\overline{\dim}_{\mathrm{B}}(\mathcal{U}) \le s. \tag{3.4}$$

Proof. We first prove Item i). Since $\mathcal{U}$ is countably $\mathscr{H}^s$-rectifiable, by Definition 3.1, there exists a countably s-rectifiable set $\mathcal{V} \subseteq \mathbb{R}^{m\times n}$ with $\mathscr{H}^s(\mathcal{U} \setminus \mathcal{V}) = 0$. By [1, Lemma III.2], the upper modified Minkowski dimension of a countably s-rectifiable set $\mathcal{V}$ is upper-bounded by s. Combined with [15, Equation (3.27)], which states that the Hausdorff dimension of $\mathcal{V}$ is upper-bounded by the upper modified Minkowski dimension, this yields

$$\dim_{\mathrm{H}}(\mathcal{V}) \leq s. \tag{3.5}$$

Since $\mathscr{H}^s(\mathcal{U} \setminus \mathcal{V}) = 0$, the definition of Hausdorff dimension implies

$$\dim_{\mathrm{H}}(\mathcal{U} \setminus \mathcal{V}) \leq s \tag{3.6}$$

so that

$$\begin{align} \dim_{\mathrm{H}}(\mathcal{U}) &= \max\{\dim_{\mathrm{H}}(\mathcal{U} \cap \mathcal{V}), \dim_{\mathrm{H}}(\mathcal{U} \setminus \mathcal{V})\} \tag{3.7}\\ &\leq \max\{\dim_{\mathrm{H}}(\mathcal{V}), \dim_{\mathrm{H}}(\mathcal{U} \setminus \mathcal{V})\} \tag{3.8}\\ &\leq s, \tag{3.9} \end{align}$$

where (3.7) follows from countable stability of Hausdorff dimension [15, Section 3.2], in (3.8) we used monotonicity of Hausdorff dimension [15, Section 3.2], and (3.9) is by (3.5) and (3.6).

To establish Item ii), we note that, by Definition 3.1, a nonempty s-rectifiable set $\mathcal{V}$ can be written as $\mathcal{V} = \varphi(\mathcal{A})$ for a Lipschitz mapping $\varphi\colon \mathcal{A} \to \mathbb{R}^{m\times n}$ and a nonempty compact set $\mathcal{A} \subseteq \mathbb{R}^s$. We thus have

$$\begin{align} \overline{\dim}_{\mathrm{B}}(\mathcal{U}) &\leq \overline{\dim}_{\mathrm{B}}(\mathcal{V}) \tag{3.10}\\ &\leq \overline{\dim}_{\mathrm{B}}(\mathcal{A}) \tag{3.11}\\ &\leq s, \tag{3.12} \end{align}$$

where (3.10) and (3.12) follow from the monotonicity of upper Minkowski dimension [15, Section 2.2] upon noting that the compact set $\mathcal{A}$ is a subset of an open ball in $\mathbb{R}^s$ of sufficiently large radius, which has upper Minkowski dimension s, and in (3.11) we applied [15, Proposition 2.5, Item (a)]. ∎

The following result will be useful in particularizing our deterministic recovery thresholds in Item i) of Theorem 2.1 and Item i) of Theorem 2.2 to rectifiable sets.

Lemma 3.6. *Let $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ be nonempty. Then,*

$$\dim_H(\mathcal{U} - \mathcal{U}) \leq \overline{\dim}_{\mathrm{B}}(\mathcal{U} - \mathcal{U}) \leq 2\,\overline{\dim}_{\mathrm{B}}(\mathcal{U}). \tag{3.13}$$

If, in addition, $\mathcal{U}$ is (countably) s-rectifiable, then $\mathcal{U}-\mathcal{U}$ is (countably) $2s$-rectifiable with

$$\dim_{\mathrm{H}}(\mathcal{U}-\mathcal{U}) \leq 2s. \tag{3.14}$$

Proof. The first inequality in (3.13) is by (1.6) and the second inequality in (3.13) follows from [15, Proposition 2.5, Item (a)] with $f(\boldsymbol{A}_1,\boldsymbol{A}_2)=\boldsymbol{A}_1-\boldsymbol{A}_2$ and the product formula [15, Equation (7.9)]. The set $\mathcal{U}-\mathcal{U}$ is (countably) $2s$-rectifiable owing to Item i) (Item ii)) in Lemma 3.4 with $f(\boldsymbol{A}_1,\boldsymbol{A}_2)=\boldsymbol{A}_1-\boldsymbol{A}_2$. Together with Item i) in Lemma 3.5 this yields (3.14). ∎

We are now in a position to particularize the results in Theorems 2.1 and 2.2 to rectifiable sets.

Theorem 3.1. *(Recovery for rectifiable sets)*

i) *Let $\mathcal{U}\subseteq\mathbb{R}^{m\times n}$ be nonempty with $\mathcal{U}-\mathcal{U}$ countably $\mathscr{H}^s$-rectifiable. Then, for $k>s$ and Lebesgue a.a. $((\boldsymbol{a}_1\dots\boldsymbol{a}_k),(\boldsymbol{b}_1\dots\boldsymbol{b}_k))\in\mathbb{R}^{m\times k}\times\mathbb{R}^{n\times k}$, every $\boldsymbol{X}\in\mathcal{U}$ can be recovered uniquely from the rank-1 measurements*

$$(\boldsymbol{a}_1^\top\boldsymbol{X}\boldsymbol{b}_1 \ \dots\ \boldsymbol{a}_k^\top\boldsymbol{X}\boldsymbol{b}_k)^\top. \tag{3.15}$$

ii) *Let $\mathcal{U},\mathcal{V}\subseteq\mathbb{R}^{m\times n}$ be nonempty with $\mathcal{V}$ s-rectifiable and $\mathcal{U}-\mathcal{U}\subseteq\mathcal{V}$. Fix $\beta\in(0,1-s/k)$ with $k>s$. Then, for Lebesgue a.a. $((\boldsymbol{a}_1\dots\boldsymbol{a}_k),(\boldsymbol{b}_1\dots\boldsymbol{b}_k))\in\mathbb{R}^{m\times k}\times\mathbb{R}^{n\times k}$, every $\boldsymbol{X}\in\mathcal{U}$ can be recovered uniquely from the rank-1 measurements*

$$(\boldsymbol{a}_1^\top\boldsymbol{X}\boldsymbol{b}_1 \ \dots\ \boldsymbol{a}_k^\top\boldsymbol{X}\boldsymbol{b}_k)^\top \tag{3.16}$$

by a β-Hölder continuous mapping g.

iii) *Let $\mathcal{U}\subseteq\mathbb{R}^{m\times n}$ be nonempty, Borel, and countably $\mathscr{H}^s$-rectifiable. Suppose that the random matrix X satisfies $\mathrm{P}[\mathsf{X}\in\mathcal{U}]=1$. Then, for Lebesgue a.a. $((\boldsymbol{a}_1\dots\boldsymbol{a}_k),(\boldsymbol{b}_1\dots\boldsymbol{b}_k))\in\mathbb{R}^{m\times k}\times\mathbb{R}^{n\times k}$, there exists a Borel-measurable mapping $g\colon\mathbb{R}^k\to\mathbb{R}^{m\times n}$, satisfying*

$$\mathrm{P}\left[g\Big((\boldsymbol{a}_1^\top\mathsf{X}\boldsymbol{b}_1\ \dots\ \boldsymbol{a}_k^\top\mathsf{X}\boldsymbol{b}_k)^\top\Big)\neq\mathsf{X}\right]=0 \tag{3.17}$$

provided that $k>s$.

iv) *Let $\mathcal{U},\mathcal{V}\subseteq\mathbb{R}^{m\times n}$ be nonempty with $\mathcal{U}$ Borel, $\mathcal{V}$ s-rectifiable, and $\mathcal{U}\subseteq\mathcal{V}$. Suppose that the random matrix X satisfies $\mathrm{P}[\mathsf{X}\in\mathcal{U}]=1$. Fix $\varepsilon>0$ and let $\beta\in(0,1-s/k)$ with $k>s$. Then, for Lebesgue a.a. $((\boldsymbol{a}_1\dots\boldsymbol{a}_k),(\boldsymbol{b}_1\dots\boldsymbol{b}_k))\in\mathbb{R}^{m\times k}\times\mathbb{R}^{n\times k}$, there exists a β-Hölder continuous mapping $g\colon\mathbb{R}^k\to\mathbb{R}^{m\times n}$ satisfying*

$$\mathrm{P}\left[g\Big((\boldsymbol{a}_1^\top\mathsf{X}\boldsymbol{b}_1\ \dots\ \boldsymbol{a}_k^\top\mathsf{X}\boldsymbol{b}_k)^\top\Big)\neq\mathsf{X}\right]\leq\varepsilon. \tag{3.18}$$

Proof. The proof is a straightforward combination of results already established:

- Item i) is by Item i) in Lemma 3.5 combined with Item i) in Theorem 2.1.
- Item ii) is by Item ii) in Lemma 3.5 combined with Item i) in Theorem 2.2.
- Item iii) is by Item i) in Lemma 3.5 combined with Item ii) in Theorem 2.1.
- Item iv) is by Item ii) in Lemma 3.5 combined with Item ii) in Theorem 2.2.

■

We now apply Theorem 3.1 to various interesting structured sets and start with sparse matrices.

Example 3.1. *Let $\mathcal{A}_s^{m\times n}$ be the set of s-sparse matrices in $\mathbb{R}^{m\times n}$, i.e., the set of matrices with at most s nonzero entries. Further, let $\mathcal{A}_{\mathcal{I}}^{m\times n}$ denote the set of matrices that have their nonzero entries indexed by $\mathcal{I} \subseteq \{1,\dots,m\}\times\{1,\dots,n\}$. Obviously, $\mathcal{A}_{\mathcal{I}}^{m\times n}$ is a linear subspace of $\mathbb{R}^{m\times n}$ of dimension $|\mathcal{I}|$. By Item iii) of Lemma 3.2, the set $\mathcal{A}_{\mathcal{I}}^{m\times n}$ is hence countably $|\mathcal{I}|$-rectifiable. As $\mathcal{A}_s^{m\times n} = \bigcup_{\mathcal{I}:|\mathcal{I}|=s}\mathcal{A}_{\mathcal{I}}^{m\times n}$, it follows from Item ii) in Lemma 3.2 that $\mathcal{A}_s^{m\times n}$ is countably s-rectifiable. Also note that $\mathcal{A}_s^{m\times n} - \mathcal{A}_s^{m\times n} = \mathcal{A}_{2s}^{m\times n}$ is countably $2s$-rectifiable.*

Similarly, for every bounded subset $\mathcal{U} \subseteq \mathcal{A}_s^{m\times n}$, $\overline{\mathcal{U}}$ is s-rectifiable. This follows by first noting that $\overline{\mathcal{U}}$ is compact in $\mathbb{R}^{m\times n}$ and, therefore, for given $\mathcal{I}$, $\mathcal{A}_{\mathcal{I}}^{m\times n} \cap \overline{\mathcal{U}}$ is a compact subset of the linear subspace $\mathcal{A}_{\mathcal{I}}^{m\times n}$. Hence, by Items ii) and iv) of Lemma 3.1, $\overline{\mathcal{U}} = \bigcup_{\mathcal{I}:|\mathcal{I}|=s}(\mathcal{A}_{\mathcal{I}}^{m\times n} \cap \overline{\mathcal{U}})$ is s-rectifiable.

We can therefore apply the corresponding items of Theorem 3.1 to obtain recovery thresholds for Lebesgue a.a. $((\boldsymbol{a}_1\dots\boldsymbol{a}_k),(\boldsymbol{b}_1\dots\boldsymbol{b}_k)) \in \mathbb{R}^{m\times k}\times\mathbb{R}^{n\times k}$ for the following sets:

i) *If $\mathcal{U} \subseteq \mathcal{A}_s^{m\times n}$ is nonempty, then every $\boldsymbol{X} \in \mathcal{U}$ can be recovered uniquely from $k > 2s$ measurements since $\mathcal{U} - \mathcal{U} \subseteq \mathcal{A}_{2s}^{m\times n}$ is countably $\mathscr{H}^{2s}$-rectifiable.*

ii) *If $\mathcal{U} \subseteq \mathcal{A}_s^{m\times n}$ is nonempty and bounded, then every $\boldsymbol{X} \in \mathcal{U}$ can be recovered uniquely from $k > 2s$ measurements by a β-Hölder continuous mapping with $\beta \in (0, 1-2s/k)$ since $\mathcal{V} = \overline{\mathcal{U}-\mathcal{U}} \subseteq \mathcal{A}_{2s}^{m\times n}$ is $(2s)$-rectifiable.*

iii) *If $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ is nonempty, Borel, and satisfies $\mathscr{H}^s(\mathcal{U}\setminus\mathcal{A}_s^{m\times n}) = 0$, then every $\boldsymbol{X}$ with $\mathrm{P}[\boldsymbol{X}\in\mathcal{U}] = 1$ can be recovered from $k > s$ measurements with zero error probability since $\mathcal{U}$ is countably $\mathscr{H}^s$-rectifiable.*

iv) *If $\mathcal{U} \subseteq \mathcal{A}_s^{m\times n}$ is nonempty, Borel, and bounded, then every $\boldsymbol{X}$ with $\mathrm{P}[\boldsymbol{X} \in \mathcal{U}] = 1$ can be recovered from $k > s$ measurements with arbitrarily small error probability by a β-Hölder continuous mapping with $\beta \in (0, 1 - s/k)$ since $\mathcal{V} = \overline{\mathcal{U}} \subseteq \mathcal{A}_s^{m\times n}$ is s-rectifiable.*

We proceed to particularizing our recovery thresholds for low-rank matrices.

Example 3.2. *The set $\mathcal{M}_r^{m\times n}$ of matrices in $\mathbb{R}^{m\times n}$ that have rank no more than r is a finite union of $\{\mathbf{0}\}$ and C^1-submanifolds of $\mathbb{R}^{m\times n}$ of dimensions no more than $(m+n-r)r$. This follows by noting that the set of matrices in $\mathbb{R}^{m\times n}$ of fixed rank k is a C^1-submanifold of $\mathbb{R}^{m\times n}$ of dimension $(m+n-k)k$* [24, Ex. 5.30], [36, Ex. 1.7]. *Application of Items ii) and iii) in Lemma 3.2 therefore yields that $\mathcal{M}_r^{m\times n}$ is countably $(m+n-r)r$-rectifiable. Also note that $\mathcal{M}_r^{m\times n}-\mathcal{M}_r^{m\times n}=\mathcal{M}_{2r}^{m\times n}$.*

Similarly, for every bounded subset $\mathcal{U}\subseteq\mathcal{M}_r^{m\times n}$, $\overline{\mathcal{U}}$ is $(m+n-r)r$-rectifiable. This follows by first noting that $\overline{\mathcal{U}}$ is compact in $\mathbb{R}^{m\times n}$ and, therefore, the intersection of $\overline{\mathcal{U}}$ with any of the finitely many C^1-submanifolds participating in $\mathcal{M}_r^{m\times n}$ is a compact subset of a C^1-submanifold. Hence, by Items ii) and iv) of Lemma 3.1, $\overline{\mathcal{U}}$ is $(m+n-r)r$-rectifiable.

We can therefore apply the corresponding items of Theorem 3.1 to obtain recovery thresholds for Lebesgue a.a. $((\boldsymbol{a}_1\dots\boldsymbol{a}_k),(\boldsymbol{b}_1\dots\boldsymbol{b}_k))\in\mathbb{R}^{m\times k}\times\mathbb{R}^{n\times k}$ for the following sets:

i) *If $\mathcal{U}\subseteq\mathcal{M}_r^{m\times n}$ is nonempty, then every $\boldsymbol{X}\in\mathcal{U}$ can be recovered uniquely from $k>2(m+n-2r)r$ measurements since $\mathcal{U}-\mathcal{U}\subseteq\mathcal{M}_{2r}^{m\times n}$ is countably $\mathscr{H}^{2(m+n-2r)r}$-rectifiable.*

ii) *If $\mathcal{U}\subseteq\mathcal{M}_r^{m\times n}$ is nonempty and bounded, then every $\boldsymbol{X}\in\mathcal{U}$ can be recovered uniquely from $k>2(m+n-2r)r$ measurements by a β-Hölder continuous mapping with $\beta\in(0,1-2(m+n-2r)r/k)$ since $\mathcal{V}=\overline{\mathcal{U}-\mathcal{U}}\subseteq\mathcal{M}_{2r}^{m\times n}$ is $(2(m+n-2r)r)$-rectifiable.*

iii) *If $\mathcal{U}\subseteq\mathbb{R}^{m\times n}$ is nonempty, Borel, and satisfies $\mathscr{H}^s(\mathcal{U}\setminus\mathcal{M}_r^{m\times n})=0$, then every $\boldsymbol{X}$ with $\mathrm{P}[\boldsymbol{X}\in\mathcal{U}]=1$ can be recovered from $k>(m+n-r)r$ measurements with zero error probability since $\mathcal{U}$ is countably $\mathscr{H}^{(m+n-r)r}$-rectifiable.*

iv) *If $\mathcal{U}\subseteq\mathcal{M}_r^{m\times n}$ is nonempty, Borel, and bounded, then every $\boldsymbol{X}$ with $\mathrm{P}[\boldsymbol{X}\in\mathcal{U}]=1$ can be recovered from $k>(m+n-r)r$ measurements with arbitrarily small error probability by a β-Hölder continuous mapping with $\beta\in(0,1-(m+n-r)r/k)$ since $\mathcal{V}=\overline{\mathcal{U}}\subseteq\mathcal{M}_r^{m\times n}$ is $((m+n-r)r)$-rectifiable.*

We proceed with the development of our general theory by demonstrating that simple, albeit relevant algebraic manipulations preserve rectifiability and hence allow the direct statement of recovery thresholds in the spirit of Theorem 3.1 through application of the approach just described.

Lemma 3.7. *Let $\mathcal{U}_i\subseteq\mathbb{R}^{m\times n}$, for $i=1,2$, and define*

i) $\mathcal{A}=\{\boldsymbol{X}\boldsymbol{X}^\mathsf{T}:\boldsymbol{X}\in\mathcal{U}_1\}$,

ii) $\mathcal{A}_\times=\{\boldsymbol{X}_1\boldsymbol{X}_2^\mathsf{T}:\boldsymbol{X}_1\in\mathcal{U}_1,\boldsymbol{X}_2\in\mathcal{U}_2\}$,

iii) $\mathcal{A}_+=\{\boldsymbol{X}_1+\boldsymbol{X}_2:\boldsymbol{X}_1\in\mathcal{U}_1,\boldsymbol{X}_2\in\mathcal{U}_2\}$,

iv) $\mathcal{A}_{\otimes} = \{\boldsymbol{X}_1 \otimes \boldsymbol{X}_2 : \boldsymbol{X}_1 \in \mathcal{U}_1, \boldsymbol{X}_2 \in \mathcal{U}_2\}$.

If the sets $\mathcal{U}_i$ are (countably) s_i-rectifiable, for $i = 1, 2$, then $\mathcal{A}$ is (countably) s_1-rectifiable and $\mathcal{A}_{\times}, \mathcal{A}_{+}$, and $\mathcal{A}_{\otimes}$ are (countably) $(s_1 + s_2)$-rectifiable.

Proof. The mapping $\boldsymbol{X} \mapsto \boldsymbol{X}\boldsymbol{X}^{\mathsf{T}}$ is continuously differentiable, and hence locally Lipschitz. Thus, by Item ii) in Lemma 3.4, the set $\mathcal{A}$ is countably s_1-rectifiable for $\mathcal{U}_1$ countably s_1-rectifiable, and, by Item i) in Lemma 3.4, the set $\mathcal{A}$ is s_1-rectifiable for s_1-rectifiable $\mathcal{U}_1$. Similarly, all of the mappings $f_{\times}(\boldsymbol{X}_1, \boldsymbol{X}_2) \mapsto \boldsymbol{X}_1\boldsymbol{X}_2^{\mathsf{T}}$, $f_{+}(\boldsymbol{X}_1, \boldsymbol{X}_2) \mapsto \boldsymbol{X}_1 + \boldsymbol{X}_2$, and $f_{\otimes}(\boldsymbol{X}_1, \boldsymbol{X}_2) \mapsto \boldsymbol{X}_1 \otimes \boldsymbol{X}_2$ are continuously differentiable, and thus locally Lipschitz. Hence, by Item ii) in Lemma 3.4, the sets $\mathcal{A}_{\times}$, $\mathcal{A}_{+}$, and $\mathcal{A}_{\otimes}$ are countably $(s_1 + s_2)$-rectifiable when the sets $\mathcal{U}_i$ are countably s_i-rectifiable, for $i = 1, 2$. Likewise, by Item i) in Lemma 3.4, the sets $\mathcal{A}_{\times}$, $\mathcal{A}_{+}$, and $\mathcal{A}_{\otimes}$ are $(s_1 + s_2)$-rectifiable when the sets $\mathcal{U}_i$ are s_i-rectifiable, for $i = 1, 2$. ∎

Lemma 3.7 in combination with Examples 3.1 and 3.2 immediately yields recovery thresholds for sums, products, and Kronecker products of sparse and low-rank matrices and covers, e.g., the structured matrices discussed in [22]. A concrete example making use of Item ii) in Lemma 3.7 is the QR-decomposition of matrices with sparse R-components.

Example 3.3. *Let $m, n \in \mathbb{N}$ with $m \geq n$ and denote by $\mathcal{C}_s^{m\times n}$ the set of all matrices in $\mathbb{R}^{m\times n}$ with s-sparse upper triangular matrix in their QR-decomposition, i.e,*

$$\mathcal{C}_s^{m\times n} = \{\boldsymbol{Q}\boldsymbol{R} : \boldsymbol{Q} \in \mathcal{Q}^{m\times m}, \boldsymbol{R} \in \mathcal{R}_s^{m\times n}\}, \tag{3.19}$$

where $\mathcal{R}_s^{m\times n} \subseteq \mathcal{A}_s^{m\times n}$ designates the set of all s-sparse upper triangular matrices and $\mathcal{Q}^{m\times m}$ stands for the set of orthogonal matrices in $\mathbb{R}^{m\times m}$. Employing the same reasoning as in Example 3.1, it follows that $\mathcal{R}_s^{m\times n}$ is countably s-rectifiable. Further, $\mathcal{Q}^{m\times m}$ is a compact $m(m-1/2)$-dimensional C^1-submanifold of $\mathbb{R}^{m\times m}$ [11, Section 1.3.1] *and thus, by Item iv) in Lemma 3.1, $(m(m-1)/2)$-rectifiable. We therefore conclude that, by Item ii) in Lemma 3.7, $\mathcal{C}_s^{m\times n}$ is countably $(m(m-1)/2 + s)$-rectifiable. Further, thanks to Lemma 3.6, $\mathcal{C}_s^{m\times n} - \mathcal{C}_s^{m\times n}$ is countably $(m(m-1) + 2s)$-rectifiable.*

Now, consider a bounded subset $\mathcal{U} \subseteq \mathcal{C}_s^{m\times n}$. Then,

$$\mathcal{U}_2 = \left\{\boldsymbol{R} \in \mathcal{R}_s^{m\times n} : \exists \boldsymbol{Q} \in \mathcal{Q}^{m\times m} \text{ with } \boldsymbol{Q}\boldsymbol{R} \in \mathcal{U}\right\} \tag{3.20}$$

is bounded because multiplication by $\boldsymbol{Q}$ does not change the 2-norm and $\mathcal{U} \subseteq \tilde{\mathcal{U}} := \{\boldsymbol{Q}\boldsymbol{R} : \boldsymbol{Q} \in \mathcal{Q}^{m\times m}, \boldsymbol{R} \in \overline{\mathcal{U}_2}\}$. Now, $\mathcal{Q}^{m\times m}$ is $m(m-1)/2$-rectifiable, and using the same argumentation as in Example 3.1 with $\mathcal{R}_s^{m\times n}$ in place of $\mathcal{A}_s^{m\times n}$, it follows that $\overline{\mathcal{U}_2}$ is s-rectifiable. Thus, $\tilde{\mathcal{U}}$ is $(m(m-1)/2 + s)$-rectifiable owing to Item ii) in Lemma 3.7. Further, thanks to Lemma 3.6, $\tilde{\mathcal{U}} - \tilde{\mathcal{U}}$ is $(m(m-1) + 2s)$-rectifiable.

We can therefore apply the corresponding items of Theorem 3.1 to obtain recovery thresholds for Lebesgue a.a. $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$ *for the following sets:*

i) *If* $\mathcal{U} \subseteq C_s^{m\times n}$ *is nonempty, then every* $\boldsymbol{X} \in \mathcal{U}$ *can be recovered uniquely from* $k > m(m-1) + 2s$ *measurements since* $\mathcal{U} - \mathcal{U} \subseteq C_s^{m\times n} - C_s^{m\times n}$ *is countably* $\mathscr{H}^{m(m-1)+2s}$*-rectifiable.*

ii) *If* $\mathcal{U} \subseteq C_s^{m\times n}$ *is nonempty and bounded, then every* $\boldsymbol{X} \in \mathcal{U}$ *can be recovered uniquely from* $k > m(m-1) + 2s$ *measurements by a* β*-Hölder continuous mapping with* $\beta \in (0, 1 - (m(m-1) + 2s)/k)$ *since* $\mathcal{V} = \tilde{\mathcal{U}} - \tilde{\mathcal{U}} \subseteq C_s^{m\times n} - C_s^{m\times n}$ *is* $(m(m-1) + 2s)$*-rectifiable.*

iii) *If* $\mathcal{U} \subseteq \mathbb{R}^{m\times n}$ *is nonempty, Borel, and satisfies* $\mathscr{H}^s(\mathcal{U} \setminus C_s^{m\times n}) = 0$*, then every* $\boldsymbol{X}$ *with* $\mathrm{P}[\boldsymbol{X} \in \mathcal{U}] = 1$ *can be recovered from* $k > m(m-1)/2 + s$ *measurements with zero error probability since* $\mathcal{U}$ *is countably* $\mathscr{H}^{m(m-1)/2+s}$*-rectifiable.*

iv) *If* $\mathcal{U} \subseteq \mathcal{A}_s^{m\times n}$ *is nonempty, Borel, and bounded, then every* $\boldsymbol{X}$ *with* $\mathrm{P}[\boldsymbol{X} \in \mathcal{U}] = 1$ *can be recovered from* $k > m(m-1)/2 + s$ *measurements with arbitrarily small error probability by a* β*-Hölder continuous mapping with* $\beta \in (0, 1 - (m(m-1)/2 + s)/k)$ *since* $\mathcal{V} = \tilde{\mathcal{U}} \subseteq C_s^{m\times n}$ *is* $(m(m-1)/2 + s)$*-rectifiable.*

We finally note that since Lemma 3.4 holds for general $N \in \mathbb{N}$, Lemma 3.7 is readily extended to sums, products, and Kronecker products of more than two matrices. This extension allows to deal, inter alia, with singular value decompositions and eigendecompositions in a manner akin to Example 3.3. Another interesting example, which can be worked out using the same arguments as in Example 3.3 with Item iii) in Lemma 3.7 in place of Item ii) in Lemma 3.7, is the recovery of matrices that are sums of low-rank and sparse matrices.

4. Recurrent Iterated Function Systems

We now demonstrate how our theory can be applied to sets of fractal nature, which do not fall into the rich class of rectifiable sets. Specifically, we investigate attractor sets of recurrent iterated function systems defined as follows [3]. Let $\mathcal{K}$ be a compact subset of $(\mathbb{R}^m, \|\cdot\|_2)$ and fix $n \in \mathbb{N}$. For $i = 1, \dots, n$, let $w_i : \mathcal{K} \to \mathcal{K}$ be similitudes of contractivity $s_i \in [0, 1)$, i.e.,

$$\|w_i(x) - w_i(y)\|_2 = s_i\|x - y\|_2, \quad \text{for all } x, y \in \mathcal{K} \text{ and } i = 1, \dots, n, \tag{4.1}$$

and designate $\boldsymbol{w} = (w_1, \dots, w_n)^{\mathsf{T}}$. Finally, let $\boldsymbol{P} \in [0,1]^{n\times n}$ with entries $p_{i,j}$ in the i-th row and j-th column. The triple $(\mathcal{K}, \boldsymbol{w}, \boldsymbol{P})$ is referred to as a recurrent iterated function system. In what follows, we assume that $\boldsymbol{P}$ is

i) row-stochastic, i.e.,

$$\sum_{j=1}^{n} p_{i,j} = 1, \quad \text{for } i \in \{1, \dots, n\}, \text{ and} \tag{4.2}$$

ii) irreducible, i.e., for every $i, j \in \{1, \dots, n\}$, there exist $i_1, \dots, i_m \in \{1, \dots, n\}$ such that $i_1 = i$, $i_m = j$, and

$$p_{i_1,i_2} p_{i_2,i_3} \dots p_{i_{m-1},i_m} > 0. \tag{4.3}$$

Further, define the connectivity matrix $\boldsymbol{C} \in \{0,1\}^{n\times n}$ with entries $c_{i,j}$ in the i-th row and j-th column according to

$$c_{i,j} = \begin{cases} 1, & \text{if } p_{i,j} > 0 \\ 0, & \text{if } p_{i,j} = 0, \end{cases} \quad \text{for } i, j \in \{1, \dots, n\} \tag{4.4}$$

and set $I(i) = \{j \in \{1, \dots, n\} : c_{i,j} = 1\}$ for $i \in \{1, \dots, n\}$. Note that $\boldsymbol{P}$ is irreducible if and only if $\boldsymbol{C}$ is irreducible. For every recurrent iterated function system $(\mathcal{K}, \boldsymbol{w}, \boldsymbol{P})$, there exist unique nonempty compact sets $\mathcal{A}_1, \dots, \mathcal{A}_n \subseteq \mathcal{K}$ satisfying [3, Corollary 3.5]

$$\mathcal{A}_i = \bigcup_{j\in I(i)} w_i(\mathcal{A}_j), \quad \text{for } i = 1, \dots, n. \tag{4.5}$$

The set

$$\mathcal{U} = (\mathcal{A}_1, \dots, \mathcal{A}_n) \subseteq \mathbb{R}^{m\times n} \tag{4.6}$$

is called the attractor set of the recurrent iterated function system $(\mathcal{K}, \boldsymbol{w}, \boldsymbol{P})$. We say that the sets $\mathcal{A}_i$ in (4.5) are nonoverlapping if, for every $i \in \{1, \dots, n\}$,

$$\mathcal{A}_j \cap \mathcal{A}_k = \emptyset, \quad \text{for all } j, k \in I(i) \text{ with } j \neq k. \tag{4.7}$$

To apply the recovery thresholds from Theorems 2.1 and 2.2 to attractor sets $\mathcal{U}$ according to (4.6), we need the following dimension result:

Theorem 4.1. [3, Theorem 4.1] *Let $(\mathcal{K}, \boldsymbol{w}, \boldsymbol{P})$ be a recurrent iterated function system with $\boldsymbol{P}$ satisfying (4.2) and (4.3). For every $t \in (0, \infty)$, define the diagonal matrix $\boldsymbol{S}(t) = diag(s_1^t, s_2^t, \dots, s_n^t)$, where s_i is the contractivity of the similitude w_i, for $i = 1, \dots, n$. Let $\mathcal{U} = (\mathcal{A}_1, \dots, \mathcal{A}_n)$ be the attractor set of $(\mathcal{K}, \boldsymbol{w}, \boldsymbol{P})$ and suppose that the sets $\mathcal{A}_1, \dots, \mathcal{A}_n$ are nonoverlapping. Finally, let d be the unique positive number* *such that* 1 *is an eigenvalue of* $\boldsymbol{S}(d)\boldsymbol{C}$ *of maximum modulus (cf.* [3, Perron-Frobenius Theorem]*). Then, it holds that*

$$\max\left\{\overline{\dim}_{\mathrm{B}}(\mathcal{A}_i) : i = 1, \ldots, n\right\} = d. \tag{4.8}$$

One obtains the following immediate consequence:

Corollary 4.1. *Under the assumptions of Theorem 4.1, the attractor set* $\boldsymbol{\mathcal{U}} = (\mathcal{A}_1, \ldots, \mathcal{A}_n)$ *satisfies*

$$\overline{\dim}_{\mathrm{B}}(\boldsymbol{\mathcal{U}}) \le nd. \tag{4.9}$$

Proof. Follows from Theorem 4.1 and the product formula [15, Equation (7.9)]. ■

We next present a simple example application of Theorem 4.1 and Corollary 4.1, which can easily be extended to higher dimensions.

Example 4.1. *Let* $s \in (0, 1/2)$ *and consider the similitudes* $w_i\colon [0,1]^2 \to [0,1]^2$ *of contractivity* s *defined as* $w_i(\boldsymbol{x}) = s\,\boldsymbol{x} + \boldsymbol{b}_i$, *for* $i = 1, \ldots, 4$, *where* $\boldsymbol{b}_1 = (0,0)^{\mathsf{T}}$, $\boldsymbol{b}_2 = (1-s, 0)^{\mathsf{T}}$, $\boldsymbol{b}_3 = (0, 1-s)^{\mathsf{T}}$, *and* $\boldsymbol{b}_4 = (1-s, 1-s)^{\mathsf{T}}$. *Now, let* $\boldsymbol{P} \in [0,1]^{4\times 4}$ *be a row-stochastic matrix and suppose that* $p_{i,i} = 0$, *for* $i = 1, \ldots, 4$, *and* $p_{i,j} > 0$, *for* $i, j \in \{1, \ldots, 4\}$ *with* $i \neq j$. *Further, let* $\mathcal{A}_1, \ldots \mathcal{A}_4$ *be as in* (4.5) *and* $\boldsymbol{\mathcal{U}}$ *as in* (4.6). *By construction, the sets* $\mathcal{A}_i$ *are nonoverlapping as* (4.5) *implies*

$$\mathcal{A}_i \subseteq w_i([0,1]^2) \quad \text{for } i = 1, \ldots, 4 \tag{4.10}$$

and the w_i'*s have pairwise disjoint codomains. Next, note that, owing to* [19, Theorem 1.3.22], *the characteristic polynomial of the all-ones matrix*

$$\boldsymbol{J} = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \tag{4.11}$$

is given by

$$p_{\boldsymbol{J}}(x) = \det(\boldsymbol{J} - x\boldsymbol{I}) = (x-4)x^3, \tag{4.12}$$

where $\boldsymbol{I} = diag(1,\ 1,\ 1,\ 1)$. *We conclude that the characteristic polynomial of the matrix* $\boldsymbol{S}(t)\boldsymbol{C}$ *equals*

$$\begin{aligned} p_{\boldsymbol{S}(t)\boldsymbol{C}}(x) &= p_{s^t\boldsymbol{C}}(x) && (4.13)\\ &= s^{4t}\det(\boldsymbol{C} - s^{-t}x\boldsymbol{I}) && (4.14)\\ &= s^{4t}\det(\boldsymbol{J} - (s^{-t}x+1)\boldsymbol{I}) && (4.15)\\ &= s^{4t}p_{\boldsymbol{J}}(s^{-t}x+1) && (4.16)\\ &= s^{4t}(s^{-t}x-3)(s^{-t}x+1)^3, && (4.17) \end{aligned}$$

where (4.15) *follows from* $\boldsymbol{C} = \boldsymbol{J} - \boldsymbol{I}$ *and in* (4.17) *we applied* (4.12). *Hence, the eigenvalue of maximum modulus of* $\boldsymbol{S}(t)\boldsymbol{C}$ *is* $\lambda_{\max} = 3s^t$. *Setting* $t = \log(1/3)/\log(s)$ *therefore yields* $\lambda_{\max} = 1$ *so that*

$$\max\left\{\overline{\dim}_{\mathrm{B}}(\mathcal{A}_i) : i = 1, \dots, 4\right\} = \frac{\log(1/3)}{\log(s)} \tag{4.18}$$

owing to Theorem 4.1 and hence

$$\overline{\dim}_{\mathrm{B}}(\mathcal{U}) \leq \frac{4\log(1/3)}{\log(s)} \tag{4.19}$$

thanks to Corollary 4.1.

The upper bound in (4.9) now leads to the following recovery thresholds:

Theorem 4.2. *(Recovery of matrices taking values in attractor sets) Let* $\mathcal{U}$ *be the attractor set of a recurrent iterated function system satisfying the assumptions of Theorem 4.1. Then, the following statements hold.*

i) *For* $k > 2nd$ *and Lebesgue a.a.* $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$, *every* $\boldsymbol{X} \in \mathcal{U}$ *can be recovered uniquely from the rank-1 measurements*

$$(\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_k)^{\mathsf{T}}. \tag{4.20}$$

ii) *Let* $\beta \in \left(0, \left(1 - \frac{2nd}{k}\right)\right)$ *with* $k > 2nd$. *Then, recovery in Item i) can be accomplished by a* β*-Hölder continuous mapping* g.

iii) *Let* X *be a random matrix satisfying* $\mathrm{P}[\mathsf{X} \in \mathcal{U}] = 1$. *Then, for Lebesgue a.a.* $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$, *there exists a Borel-measurable mapping* $g\colon \mathbb{R}^k \to \mathbb{R}^{m\times n}$ *satisfying*

$$\mathrm{P}\left[g\left((\boldsymbol{a}_1^{\mathsf{T}}\mathsf{X}\boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^{\mathsf{T}}\mathsf{X}\boldsymbol{b}_k)^{\mathsf{T}}\right) \neq \mathsf{X}\right] = 0 \tag{4.21}$$

provided that $k > nd$.

iv) *Let* X *be a random matrix satisfying* $\mathrm{P}[\mathsf{X} \in \mathcal{U}] = 1$, *fix* $\varepsilon > 0$, *and let* $\beta \in \left(0, \left(1 - \frac{nd}{k}\right)\right)$ *with* $k > nd$. *Then, for Lebesgue a.a.* $((\boldsymbol{a}_1 \dots \boldsymbol{a}_k), (\boldsymbol{b}_1 \dots \boldsymbol{b}_k)) \in \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$, *there exists a* β*-Hölder continuous mapping* $g\colon \mathbb{R}^k \to \mathbb{R}^{m\times n}$ *satisfying*

$$\mathrm{P}\left[g\left((\boldsymbol{a}_1^{\mathsf{T}}\mathsf{X}\boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^{\mathsf{T}}\mathsf{X}\boldsymbol{b}_k)^{\mathsf{T}}\right) \neq \mathsf{X}\right] \leq \varepsilon. \tag{4.22}$$

Proof. With $\overline{\dim}_{\mathrm{B}}(\mathcal{U}) \leq nd$ from Corollary 4.1, we have

$$\dim_{\mathrm{H}}(\mathcal{U} - \mathcal{U}) \leq \overline{\dim}_{\mathrm{B}}(\mathcal{U} - \mathcal{U}) \leq 2\overline{\dim}_{\mathrm{B}}(\mathcal{U}) \leq 2nd \tag{4.23}$$

thanks to Lemma 3.6. The statements in Item i)–Item iv) now follow readily from the corresponding parts of Theorems 2.1 and 2.2. ∎

5. Proof of Theorem 2.1

Item i) is by linearity of the mapping defined in (2.1)–(2.2) combined with Proposition 2.1 applied to the set $\mathcal{U} - \mathcal{U}$.

The proof of Item ii) follows along the same lines as the proof of [1, Theorem II.1]. We therefore present a proof sketch only. First, note that by [2, Lemma 2.3], there exist compact sets $\mathcal{U}_i \subseteq \mathcal{U}$, $i \in \mathbb{N}$, such that $\mathrm{P}[\mathsf{X} \in \mathcal{V}] = 1$, where

$$\mathcal{V} = \bigcup_{i\in\mathbb{N}} \mathcal{U}_i. \tag{5.1}$$

Next, consider the "encoder" mapping[3]

$$e\colon \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times n} \to \mathbb{R}^k \tag{5.2}$$

$$\big(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{V}\big) \mapsto (\boldsymbol{a}_1^\mathsf{T} \boldsymbol{V} \boldsymbol{b}_1 \ \ldots \ \boldsymbol{a}_k^\mathsf{T} \boldsymbol{V} \boldsymbol{b}_k)^\mathsf{T}. \tag{5.3}$$

With the decomposition of $\mathcal{V}$ in (5.1), argumentation as in [1, Section V.A] (with the mapping $\|\boldsymbol{y} - \boldsymbol{A}\boldsymbol{v}\|_2$ in [1, (139)–(140)] replaced by $\|\boldsymbol{y} - e(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{V})\|_2$) implies the existence of a measurable mapping

$$\hat{g}\colon \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{k} \to \mathbb{R}^{m\times n} \tag{5.4}$$

$$\boldsymbol{A} \times \boldsymbol{B} \times \boldsymbol{y} \mapsto \boldsymbol{X} \tag{5.5}$$

such that

$$e\big(\boldsymbol{A}, \boldsymbol{B}, \hat{g}(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{y})\big) = \boldsymbol{y}, \text{ for all } \boldsymbol{A} \in \mathbb{R}^{m\times k}, \boldsymbol{B} \in \mathbb{R}^{n\times k} \text{ and } \boldsymbol{y} \in e(\{\boldsymbol{A}\} \times \{\boldsymbol{B}\} \times \mathcal{V}). \tag{5.6}$$

Moreover, the mapping $\hat{g}$ is guaranteed to deliver an $\boldsymbol{X} \in \mathcal{V}$ that is consistent if at least one such consistent $\boldsymbol{X} \in \mathcal{V}$ exists, otherwise an error is declared by delivering an error symbol not contained in $\mathcal{V}$. Next, for every $\boldsymbol{A} \in \mathbb{R}^{m\times k}$ and $\boldsymbol{B} \in \mathbb{R}^{n\times k}$, let $p_\mathrm{e}(\boldsymbol{A}, \boldsymbol{B})$ denote the probability of error defined as

$$p_\mathrm{e}(\boldsymbol{A}, \boldsymbol{B}) = \mathrm{P}[\hat{g}(\boldsymbol{A}, \boldsymbol{B}, e(\boldsymbol{A}, \boldsymbol{B}, \mathsf{X})) \neq \mathsf{X}]. \tag{5.7}$$

We now show that $p_\mathrm{e}(\boldsymbol{A}, \boldsymbol{B}) = 0$ for Lebesgue a.a. $(\boldsymbol{A}, \boldsymbol{B})$. We have

$$\int p_\mathrm{e}(\boldsymbol{A}, \boldsymbol{B})\, \mathrm{d}\lambda(\boldsymbol{A}, \boldsymbol{B}) \tag{5.8}$$

$$= \mathrm{E}\left[\lambda\big(\{(\boldsymbol{A}, \boldsymbol{B}) : g\big(\boldsymbol{A}, \boldsymbol{B}, e(\boldsymbol{A}, \boldsymbol{B}, \mathsf{X})\big) \neq \mathsf{X}\}\big)\chi_{\mathcal{V}}(\mathsf{X})\right] \tag{5.9}$$

$$\leq \mathrm{E}\left[\lambda\big(\{(\boldsymbol{A}, \boldsymbol{B}) : \{\widetilde{\boldsymbol{V}} \in \mathcal{V}_{\mathsf{X}} : e(\boldsymbol{A}, \boldsymbol{B}, \widetilde{\boldsymbol{V}}) = \boldsymbol{0}\} \neq \{\boldsymbol{0}\}\}\big)\right], \tag{5.10}$$

[3] $\boldsymbol{A}$ and $\boldsymbol{B}$ denote the matrices with $\boldsymbol{a}_i$ and $\boldsymbol{b}_i$, respectively, in their i-th column.

where (5.9) follows from Fubini's theorem [26, Theorem 1.14] together with $\mathrm{P}[\mathbf{X} \in \mathcal{V}] = 1$, and in (5.10), we set $\mathcal{V}_{\boldsymbol{X}} = \{\boldsymbol{V} - \boldsymbol{X} : \boldsymbol{V} \in \mathcal{V}\}$ and used the fact that, by (5.6), $\boldsymbol{V} := \hat{g}\big(\boldsymbol{A}, \boldsymbol{B}, e(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X})\big) \neq \boldsymbol{X}$ with $\boldsymbol{X} \in \mathcal{V}$ implies that $\boldsymbol{V} \in \mathcal{V} \backslash \{\boldsymbol{X}\}$ with

$$e(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X}) = e(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{V}), \tag{5.11}$$

i.e., $e(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{V} - \boldsymbol{X}) = \mathbf{0}$. Finally, since $\mathscr{H}^k(\mathcal{V}_{\boldsymbol{X}}) = \mathscr{H}^k(\mathcal{V})$ by the translation invariance of $\mathscr{H}^k$ and $\mathscr{H}^k(\mathcal{V}) = 0$ as a consequence of $\mathcal{V} \subseteq \mathcal{U}$ and $\dim_{\mathrm{H}}(\mathcal{U}) < k$, the expectation in (5.10) is equal to zero owing to Proposition 2.1. Finally, for fixed $\boldsymbol{A}, \boldsymbol{B}$, set $g = \hat{g}(\boldsymbol{A}, \boldsymbol{B}, \cdot)$. ∎

6. Proof of Proposition 2.1

For every $j \in \mathbb{N}$, set

$$\mathcal{A}(j) = \underbrace{\mathcal{B}_m(\mathbf{0}, j) \times \cdots \times \mathcal{B}_m(\mathbf{0}, j)}_{k \text{ times}} \quad \text{and} \tag{6.1}$$

$$\mathcal{B}(j) = \underbrace{\mathcal{B}_n(\mathbf{0}, j) \times \cdots \times \mathcal{B}_n(\mathbf{0}, j)}_{k \text{ times}}. \tag{6.2}$$

By countable subadditivity of Lebesgue measure, it suffices to show that

$$\big\{\boldsymbol{X} \in \mathcal{U} \backslash \{\mathbf{0}\} : (\boldsymbol{a}_1^\top \boldsymbol{X} \boldsymbol{b}_1 \ \ldots \ \boldsymbol{a}_k^\top \boldsymbol{X} \boldsymbol{b}_k)^\top = \mathbf{0}\big\} = \emptyset, \tag{6.3}$$

for Lebesgue a.a. $((\boldsymbol{a}_1 \ldots \boldsymbol{a}_k), (\boldsymbol{b}_1 \ldots \boldsymbol{b}_k)) \in \mathcal{A}(j) \times \mathcal{B}(j)$ and all $j \in \mathbb{N}$. By Lemma 6.1 below, (6.3) then holds, for all $j \in \mathbb{N}$, with probability one if the deterministic matrices $\big((\boldsymbol{a}_1 \ldots \boldsymbol{a}_k), (\boldsymbol{b}_1 \ldots \boldsymbol{b}_k)\big) \in \mathcal{A}(j) \times \mathcal{B}(j)$ are replaced by independent random matrices with columns $\mathbf{a}_i$, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_m(\mathbf{0}, j)$, and columns $\mathbf{b}_i$, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_n(\mathbf{0}, j)$. By countable subadditivity of Lebesgue measure, this finally implies that (2.4) can be violated only on a set of Lebesgue measure zero, which concludes the proof.

Lemma 6.1. *Let $s > 0$ and take $\mathbf{A} = (\mathbf{a}_1 \ldots \mathbf{a}_k)$ and $\mathbf{B} = (\mathbf{b}_1 \ldots \mathbf{b}_k)$ to be independent random matrices with columns $\mathbf{a}_i$, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_m(\mathbf{0}, s)$, and columns $\mathbf{b}_i$, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_n(\mathbf{0}, s)$. Consider $\mathcal{U} \subseteq \mathbb{R}^{m \times n}$ with $\dim_{\mathrm{H}}(\mathcal{U}) < k$. Then,*

$$P := \mathrm{P}\big[\exists \boldsymbol{X} \in \mathcal{U} \backslash \{\mathbf{0}\} : (\mathbf{a}_1^\top \boldsymbol{X} \mathbf{b}_1 \ \ldots \ \mathbf{a}_k^\top \boldsymbol{X} \mathbf{b}_k)^\top = \mathbf{0}\big] = 0. \tag{6.4}$$

Proof. For every $L \in \mathbb{N}$, let

$$\mathcal{U}_L = \Big\{\boldsymbol{X} \in \mathcal{U} : \tfrac{1}{L} < \sigma_1(\boldsymbol{X}) < L\Big\} \tag{6.5}$$

and set

$$P_L = \mathrm{P}\big[\exists X \in \mathcal{U}_L : (\mathbf{a}_1^\mathsf{T} X \mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T} X \mathbf{b}_k)^\mathsf{T} = \mathbf{0}\big]. \tag{6.6}$$

By the union bound, we have

$$P \leq \sum_{L\in\mathbb{N}} P_L. \tag{6.7}$$

We now fix $L \in \mathbb{N}$ arbitrarily and prove that $P_L = 0$. Let $\kappa = (k + \dim_\mathrm{H}(\mathcal{U}))/2$. As $k > \dim_\mathrm{H}(\mathcal{U})$ by assumption, it follows that $\dim_\mathrm{H}(\mathcal{U}) < \kappa < k$. In particular, $\kappa > \dim_\mathrm{H}(\mathcal{U})$ implies, by [15, Equation (3.11)], that $\mathscr{H}^\kappa(\mathcal{U}) = 0$ and in turn $\mathscr{H}^\kappa(\mathcal{U}_L) = 0$ by monotonicity of $\mathscr{H}^\kappa$. Thus, $\mathscr{M}^\kappa(\mathcal{U}_L) = 0$ by [15, Section 3.4], where the measure $\mathscr{M}^\kappa$ is defined according to

$$\mathscr{M}^\kappa(\mathcal{V}) = \lim_{d\to 0} \mathscr{M}_d^\kappa(\mathcal{V}) \tag{6.8}$$

with

$$\mathscr{M}_d^\kappa(\mathcal{V}) = \inf\left\{ \sum_{i\in\mathbb{N}} \varepsilon_i^\kappa : \mathcal{V} \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{B}_{m\times n}\big(X_i, \tfrac{\varepsilon_i}{2}\big) \right\}, \quad \text{for all } d > 0, \tag{6.9}$$

where the infimum is taken over all possible ball centers $X_i \in \mathbb{R}^{m\times n}$ and radii $\varepsilon_i \in (0, d)$, $i \in \mathbb{N}$. Since $\mathscr{M}_d^\kappa(\mathcal{U}_L)$ is nonnegative and monotonically nondecreasing as $d \to 0$, $\mathscr{M}^\kappa(\mathcal{U}_L) = 0$ implies $\mathscr{M}_d^\kappa(\mathcal{U}_L) = 0$, for all $d > 0$. Now, fix $d > 0$ and $\varepsilon \in (0, (\sqrt{k}L)^{-\kappa})$ arbitrarily. As $\mathscr{M}_d^\kappa(\mathcal{U}_L) = 0$, there must exist ball centers $X_i \in \mathbb{R}^{m\times n}$, $i \in \mathbb{N}$, and radii ε_i, $i \in \mathbb{N}$, such that

$$\mathcal{U}_L \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{B}_{m\times n}\big(X_i, \tfrac{\varepsilon_i}{2}\big) \tag{6.10}$$

and

$$\sum_{i\in\mathbb{N}} \varepsilon_i^\kappa < \varepsilon. \tag{6.11}$$

As (6.10) and (6.11) continue to hold upon removal of all i that satisfy

$$\mathcal{U}_L \cap \mathcal{B}_{m\times n}\big(X_i, \tfrac{\varepsilon_i}{2}\big) = \emptyset, \tag{6.12}$$

we can assume, w.l.o.g., that

$$\mathcal{U}_L \cap \mathcal{B}_{m\times n}\big(X_i, \tfrac{\varepsilon_i}{2}\big) \neq \emptyset, \quad \text{for all } i \in \mathbb{N}. \tag{6.13}$$

By doubling the radius, we can further construct a covering that has all its ball centers in $\mathcal{U}$. Concretely, by (6.13), for every $i \in \mathbb{N}$, there exists $\boldsymbol{Y}_i \in \mathcal{U}_L \cap \mathcal{B}_{m\times n}(\boldsymbol{X}_i, \varepsilon_i/2)$, and we have $\mathcal{B}_{m\times n}(\boldsymbol{X}_i, \varepsilon_i/2) \subseteq \mathcal{B}_{m\times n}(\boldsymbol{Y}_i, \varepsilon_i)$. Thus, by (6.10),

$$\mathcal{U}_L \subseteq \bigcup_{i\in\mathbb{N}} \mathcal{B}_{m\times n}(\boldsymbol{Y}_i, \varepsilon_i). \tag{6.14}$$

With the definition of $\mathcal{U}_L$ in (6.5), we now obtain for the shifted ball centers

$$\tfrac{1}{L} < \sigma_1(\boldsymbol{Y}_i) < L, \quad \text{for all } i \in \mathbb{N}. \tag{6.15}$$

A union bound argument applied to (6.6) in combination with (6.14) yields

$$P_L \le \sum_{i\in\mathbb{N}} \mathrm{P}\left[\exists \boldsymbol{X} \in \mathcal{B}_{m\times n}(\boldsymbol{Y}_i, \varepsilon_i) : (\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T} = \mathbf{0}\right]. \tag{6.16}$$

To bound the individual probabilities on the right-hand side of (6.16), we proceed as follows. Suppose that $\boldsymbol{X} \in \mathcal{B}_{m\times n}(\boldsymbol{Y}_i, \varepsilon_i)$ for some $i \in \mathbb{N}$. Then, we have

$$\begin{aligned}
&\|(\mathbf{a}_1^\mathsf{T}\boldsymbol{Y}_i\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{Y}_i\mathbf{b}_k)^\mathsf{T}\|_2 && (6.17)\\
&\le \|(\mathbf{a}_1^\mathsf{T}(\boldsymbol{X}-\boldsymbol{Y}_i)\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}(\boldsymbol{X}-\boldsymbol{Y}_i)\mathbf{b}_k)^\mathsf{T}\|_2 + \|(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T}\|_2 && (6.18)\\
&\le \sqrt{\sum_{j=1}^{k}\|\mathbf{a}_j\|_2^2\|\boldsymbol{X}-\boldsymbol{Y}_i\|_2^2\|\mathbf{b}_j\|_2^2} + \|(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T}\|_2 && (6.19)\\
&\le s^2\sqrt{k}\varepsilon_i + \|(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T}\|_2, && (6.20)
\end{aligned}$$

where in (6.20) we used that $\mathbf{a}_j$ and $\mathbf{b}_j$ are uniformly distributed on $\mathcal{B}_m(\mathbf{0}, s)$ and $\mathcal{B}_n(\mathbf{0}, s)$, respectively, and $\boldsymbol{X} \in \mathcal{B}_{m\times n}(\boldsymbol{Y}_i, \varepsilon_i)$ by assumption. Thus, the event that there exists $\boldsymbol{X} \in \mathcal{B}_{m\times n}(\boldsymbol{Y}_i, \varepsilon_i)$ satisfying

$$(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T} = \mathbf{0} \tag{6.21}$$

implies that $\|(\mathbf{a}_1^\mathsf{T}\boldsymbol{Y}_i\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{Y}_i\mathbf{b}_k)^\mathsf{T}\|_2 \le s^2\sqrt{k}\varepsilon_i$. Hence, we can further upper-bound P_L according to

$$\begin{aligned}
P_L &\le \sum_{i\in\mathbb{N}} \mathrm{P}\left[\|(\mathbf{a}_1^\mathsf{T}\boldsymbol{Y}_i\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{Y}_i\mathbf{b}_k)^\mathsf{T}\|_2 \le s^2\sqrt{k}\varepsilon_i\right] && (6.22)\\
&\le k^{\frac{k}{2}} \sum_{i\in\mathbb{N}} \varepsilon_i^k \frac{2^{\frac{k(m+n)}{2}}}{\sigma_1(\boldsymbol{Y}_i)^k}\left(1 + \log\left(\frac{\sigma_1(\boldsymbol{Y}_i)}{\sqrt{k}\varepsilon_i}\right)\right)^k && (6.23)\\
&\le C \sum_{i\in\mathbb{N}} \varepsilon_i^k \left(1 + \log\left(\frac{L}{\sqrt{k}\varepsilon_i}\right)\right)^k && (6.24)\\
&= C \sum_{i\in\mathbb{N}} \varepsilon_i^\kappa \varepsilon_i^{k-\kappa} \left(1 + \log\left(\frac{L}{\sqrt{k}\varepsilon_i}\right)\right)^k && (6.25)
\end{aligned}$$

with

$$C = 2^{\frac{k(m+n)}{2}} \left(L\sqrt{k}\right)^k, \tag{6.26}$$

where (6.23) is by Lemma 6.2 below for $\delta = s^2\sqrt{k}\varepsilon_i$ and $\boldsymbol{X} = \boldsymbol{Y}_i$ upon noting that

$$\begin{aligned} s^2\sqrt{k}\varepsilon_i &< s^2\sqrt{k}\varepsilon^{1/\kappa} && (6.27)\\ &< \frac{s^2}{L} && (6.28)\\ &< \sigma_1(\boldsymbol{Y}_i)s^2, \quad \text{for all } i \in \mathbb{N}. && (6.29) \end{aligned}$$

Here, (6.27) is by (6.11), in (6.28) we used $\varepsilon < (\sqrt{k}L)^{-\kappa}$ which holds by assumption, and (6.29) follows from (6.15). As the log-term in (6.25) is dominated by $\varepsilon_i^{k-\kappa}$ for $\varepsilon_i \to 0$ thanks to $k > \kappa$, (6.25) tends to zero for $\varepsilon \to 0$ by (6.11). We can therefore conclude that $P_L = 0$, which, as L was arbitrary, by (6.7), implies $P = 0$. ∎

Lemma 6.2. *Let* $\mathsf{A} = (\mathsf{a}_1 \dots \mathsf{a}_k)$ *and* $\mathsf{B} = (\mathsf{b}_1 \dots \mathsf{b}_k)$ *be independent random matrices, with columns* a_i, $i = 1, \dots, k$, *independent and uniformly distributed on* $\mathcal{B}_m(\mathbf{0}, s)$ *and columns* b_i, $i = 1, \dots, k$, *independent and uniformly distributed on* $\mathcal{B}_n(\mathbf{0}, s)$. *Suppose that* $\boldsymbol{X} \in \mathbb{R}^{m\times n}\backslash\{\mathbf{0}\}$. *Then, we have*

$$\mathrm{P}\left[\left\|(\mathsf{a}_1^\mathsf{T}\boldsymbol{X}\mathsf{b}_1 \ \dots \ \mathsf{a}_k^\mathsf{T}\boldsymbol{X}\mathsf{b}_k)^\mathsf{T}\right\|_2 \le \delta\right] \le \delta^k \frac{2^{\frac{k(m+n)}{2}}}{\sigma_1(\boldsymbol{X})^k s^{2k}}\left(1 + \log\left(\frac{s^2\sigma_1(\boldsymbol{X})}{\delta}\right)\right)^k, \tag{6.30}$$

for all $\delta \le \sigma_1(\boldsymbol{X})s^2$.

Proof. We have

$$\begin{aligned} &\mathrm{P}\left[\left\|(\mathsf{a}_1^\mathsf{T}\boldsymbol{X}\mathsf{b}_1 \dots \mathsf{a}_k^\mathsf{T}\boldsymbol{X}\mathsf{b}_k)^\mathsf{T}\right\|_2 \le \delta\right] && (6.31)\\ &= \mathrm{P}\left[\sum_{i=1}^k (\mathsf{a}_i^\mathsf{T}\boldsymbol{X}\mathsf{b}_i)^2 \le \delta^2\right] && (6.32)\\ &\le \mathrm{P}\left[|\mathsf{a}_i^\mathsf{T}\boldsymbol{X}\mathsf{b}_i| \le \delta, \ \text{for } i = 1, \dots, k\right] && (6.33)\\ &= \mathrm{P}\left[|\mathsf{a}^\mathsf{T}\boldsymbol{X}\mathsf{b}| \le \delta\right]^k && (6.34)\\ &\le \delta^k \frac{2^{\frac{k(m+n)}{2}}}{\sigma_1(\boldsymbol{X})^k s^{2k}}\left(1 + \log\left(\frac{s^2\sigma_1(\boldsymbol{X})}{\delta}\right)\right)^k, && (6.35) \end{aligned}$$

where in (6.34) a and b are independent with a uniformly distributed on $\mathcal{B}_m(\mathbf{0}, s)$ and b uniformly distributed on $\mathcal{B}_n(\mathbf{0}, s)$ and, therefore, we can apply Lemma 6.3 below to obtain (6.35). ∎

Lemma 6.3. [25, Lemma 17][4] *Let* a *and* b *be independent random vectors, with* a *uniformly distributed on* $\mathcal{B}_m(\mathbf{0}, s)$ *and* b *uniformly distributed on* $\mathcal{B}_n(\mathbf{0}, s)$*, and suppose that* $\boldsymbol{X} \in \mathbb{R}^{m\times n}\backslash\{\mathbf{0}\}$*. Then, we have*

$$\mathrm{P}[|\mathsf{a}^{\mathsf{T}}\boldsymbol{X}\mathsf{b}| \leq \delta] \leq \delta \frac{D_{m,n}}{\sigma_1(\boldsymbol{X})s^2}\left(1 + \log\left(\frac{s^2\sigma_1(\boldsymbol{X})}{\delta}\right)\right), \quad \textit{for all } \delta \leq \sigma_1(\boldsymbol{X})s^2, \tag{6.36}$$

where[5]

$$D_{m,n} = \frac{4V(n-1,1)V(m-1,1)}{V(m,1)V(n,1)} \tag{6.37}$$

$$\leq 2^{\frac{m+n}{2}}. \tag{6.38}$$

Proof. We start by applying Fubini's Theorem [26, Theorem 1.14] and rewriting

$$\mathrm{P}[|\mathsf{a}^{\mathsf{T}}\boldsymbol{X}\mathsf{b}| \leq \delta] = \frac{1}{V(m,s)V(n,s)} \int_{\mathcal{B}_m(\mathbf{0},s)} h(\boldsymbol{a})\,\mathrm{d}\lambda(\boldsymbol{a}) \tag{6.39}$$

with

$$h(\boldsymbol{a}) = \int_{\mathcal{B}_n(\mathbf{0},s)} \chi_{\{\boldsymbol{b}\in\mathbb{R}^n : |\boldsymbol{a}^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}| \leq \delta\}}(\boldsymbol{b})\,\mathrm{d}\lambda(\boldsymbol{b}). \tag{6.40}$$

Let $\boldsymbol{X} = \boldsymbol{U}\Sigma\boldsymbol{V}$ be a singular value decomposition of $\boldsymbol{X}$, where $\boldsymbol{U} \in \mathbb{R}^{m\times m}$ and $\boldsymbol{V} \in \mathbb{R}^{n\times n}$ are orthogonal matrices, and

$$\Sigma = \begin{pmatrix} \boldsymbol{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \in \mathbb{R}^{m\times n} \tag{6.41}$$

with $\boldsymbol{D} = diag(\sigma_1(\boldsymbol{X}) \ldots \sigma_r(\boldsymbol{X}))$ and $r = rank(\boldsymbol{X})$. Using the fact that Lebesgue measure on $\mathcal{B}_m(\mathbf{0}, s)$ and $\mathcal{B}_n(\mathbf{0}, s)$ is invariant under rotations, we can write

$$\mathrm{P}[|\mathsf{a}^{\mathsf{T}}\boldsymbol{X}\mathsf{b}| \leq \delta] = \frac{1}{V(m,s)V(n,s)} \int_{\mathcal{B}_m(\mathbf{0},s)} h(\boldsymbol{U}\boldsymbol{a})\,\mathrm{d}\lambda(\boldsymbol{a}) \tag{6.42}$$

and

$$h(\boldsymbol{U}\boldsymbol{a}) = \int_{\mathcal{B}_n(\mathbf{0},s)} \chi_{\{\boldsymbol{b}\in\mathbb{R}^n : |\boldsymbol{a}^{\mathsf{T}}\Sigma\boldsymbol{b}| \leq \delta\}}(\boldsymbol{b})\,\mathrm{d}\lambda(\boldsymbol{b}). \tag{6.43}$$

[4] Since the assumption $\delta \leq \sigma_1(\boldsymbol{X})s^2$ is missing in [25, Lemma 17] we present the proof of the lemma for completeness. A slightly weaker form of this result was first presented in [30, Lemma 5].

[5] We use the convention $V(0, s) = 1$, for all $s \in \mathbb{R}_+$.

We now make the dependence on the largest eigenvalue $\sigma_1(\boldsymbol{X})$ explicit according to

$$h(\boldsymbol{U}\boldsymbol{a}) = \int_{\mathcal{B}_{n-1}(\boldsymbol{0},s)} \int_{-s}^{s} \chi_{\{b_1\in\mathbb{R}:|\boldsymbol{a}^{\mathsf{T}}\Sigma\boldsymbol{b}|\le\delta,\|\boldsymbol{b}\|\le s\}}(b_1)\,\mathrm{d}\lambda(b_1)\,\mathrm{d}\lambda\big((b_2\dots b_n)^{\mathsf{T}}\big) \tag{6.44}$$

$$\le \int_{\mathcal{B}_{n-1}(\boldsymbol{0},s)} g\big((b_2\dots b_n)^{\mathsf{T}}\big)\,\mathrm{d}\lambda\big((b_2\dots b_n)^{\mathsf{T}}\big) \tag{6.45}$$

with

$$g\big((b_2\dots b_n)^{\mathsf{T}}\big) = \min\left\{2s, \int_{-\infty}^{\infty} \chi_{\{b_1\in\mathbb{R}:|\sum_{i=1}^{r}\sigma_i(\boldsymbol{X})a_ib_i|\le\delta\}}(b_1)\,\mathrm{d}\lambda(b_1)\right\} \tag{6.46}$$

$$= \min\left\{2s, \int_{-\infty}^{\infty} \chi_{\{b_1\in\mathbb{R}:|\sigma_1(\boldsymbol{X})a_1b_1|\le\delta\}}(b_1)\,\mathrm{d}\lambda(b_1)\right\} \tag{6.47}$$

$$= 2\min\left\{s, \frac{\delta}{\sigma_1(\boldsymbol{X})|a_1|}\right\}. \tag{6.48}$$

Using (6.45) and (6.46)–(6.48) in (6.42), we obtain

$$\mathrm{P}[|\mathbf{a}^{\mathsf{T}}\boldsymbol{X}\mathbf{b}|\le\delta] \le \frac{D_{m,n}}{s^2}\int_0^s \min\left\{s, \frac{\delta}{\sigma_1(\boldsymbol{X})a_1}\right\}\mathrm{d}\lambda(a_1) \tag{6.49}$$

$$= \frac{\delta D_{m,n}}{\sigma_1(\boldsymbol{X})s^2}\left(1+\log\left(\frac{s^2\sigma_1(\boldsymbol{X})}{\delta}\right)\right), \tag{6.50}$$

for all $\delta \le \sigma_1(\boldsymbol{X})s^2$. The upper bound on $D_{m,n}$ follows from $2^{k/2} < V(k,1) < 2^k$, for all $k \in \mathbb{N}$. ∎

7. Proof of Theorem 2.2

We first prove Item i). Consider the mapping

$$h\colon \mathcal{U} \to h(\mathcal{U}) \subseteq \mathbb{R}^k \tag{7.1}$$

$$\boldsymbol{X} \mapsto (\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_1\ \dots\ \boldsymbol{a}_k^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_k)^{\mathsf{T}}. \tag{7.2}$$

Application of Proposition 2.2 to $\mathcal{U}-\mathcal{U}$ establishes the existence of a $c>0$ such that

$$\|h(\boldsymbol{U}-\boldsymbol{V})\|_2 \ge c\|\boldsymbol{U}-\boldsymbol{V}\|_2^{1/\beta}, \quad \text{for all } \boldsymbol{U},\boldsymbol{V}\in\mathcal{U}, \tag{7.3}$$

for Lebesgue a.a. $((\boldsymbol{a}_1\dots\boldsymbol{a}_k),(\boldsymbol{b}_1\dots\boldsymbol{b}_k)) \in \mathbb{R}^{m\times k}\times\mathbb{R}^{n\times k}$. Hence, by [33, Lemma 2], h admits a β-Hölder continuous inverse $h^{-1}\colon h(\mathcal{U}) \to \mathcal{U}$, which can be extended to the desired β-Hölder continuous mapping g on $\mathbb{R}^k$ owing to [27, Theorem 1, Item ii)].

The proof of Item ii) follows along the same lines as that of [33, Theorem 2]. We therefore present a proof sketch only. By [15, Proposition 2.6], we can assume, w.l.o.g., that $\mathcal{U}$ is compact. Consider the sets $\mathcal{A}, \mathcal{A}_j \subseteq \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k} \times \mathbb{R}^{m\times n}$ defined according to

$$\mathcal{A} = \left\{ (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X}) : \inf \left\{ \frac{\|(\boldsymbol{a}_1^\top \boldsymbol{U} \boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^\top \boldsymbol{U} \boldsymbol{b}_k)^\top\|_2}{\|\boldsymbol{U}\|_2^{1/\beta}} : \boldsymbol{U} \in \mathcal{U}_{\boldsymbol{X}} \setminus \{\boldsymbol{0}\} \right\} = 0 \right\} \tag{7.4}$$

and

$$\mathcal{A}_j = \left\{ (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X}) : \inf \left\{ \frac{\|(\boldsymbol{a}_1^\top \boldsymbol{U} \boldsymbol{b}_1 \ \dots \ \boldsymbol{a}_k^\top \boldsymbol{U} \boldsymbol{b}_k)^\top\|_2}{\|\boldsymbol{U}\|_2^{1/\beta}} : \boldsymbol{U} \in \mathcal{U}_{\boldsymbol{X}} \setminus \{\boldsymbol{0}\} \right\} > \frac{1}{j} \right\}, \tag{7.5}$$

for all $j \in \mathbb{N}$, where

$$\mathcal{U}_{\boldsymbol{X}} = \{\boldsymbol{U} - \boldsymbol{X} : \boldsymbol{U} \in \mathcal{U}\}, \quad \text{for all } \boldsymbol{X} \in \mathbb{R}^{m\times n}. \tag{7.6}$$

By the same arguments as used in [33, Section VI], one can show that $\mathcal{A}$ is a measurable set. Application of Fubini's Theorem [26, Theorem 1.14] therefore yields

$$\int \mathrm{P}[(\boldsymbol{A}, \boldsymbol{B}, \mathbf{X}) \in \mathcal{A}] \, \mathrm{d}\lambda(\boldsymbol{A}, \boldsymbol{B}) = \mathrm{E}[\lambda\{(\boldsymbol{A}, \boldsymbol{B}) : (\boldsymbol{A}, \boldsymbol{B}, \mathbf{X}) \in \mathcal{A}\}]. \tag{7.7}$$

As the right-hand side of (7.7) equals zero owing to Proposition 2.2, it follows that

$$\mathrm{P}[(\boldsymbol{A}, \boldsymbol{B}, \mathbf{X}) \in \mathcal{A}] = 0, \quad \text{for Lebesgue a.a. } (\boldsymbol{A}, \boldsymbol{B}). \tag{7.8}$$

Since the complement of $\mathcal{A}$, denoted by $\mathcal{A}^{\mathrm{c}}$, can be written as

$$\mathcal{A}^{\mathrm{c}} = \bigcup_{j\in\mathbb{N}} \mathcal{A}_j, \tag{7.9}$$

application of [4, Lemma 3.4, Item (a)] together with (7.8) yields

$$\lim_{j\to\infty} \mathrm{P}[(\boldsymbol{A}, \boldsymbol{B}, \mathbf{X}) \in \mathcal{A}_j] = 1, \quad \text{for Lebesgue a.a. } (\boldsymbol{A}, \boldsymbol{B}). \tag{7.10}$$

Let $C \subseteq \mathbb{R}^{m\times k} \times \mathbb{R}^{n\times k}$ denote the set of matrices $(\boldsymbol{A}, \boldsymbol{B})$ for which (7.10) holds, and fix $\varepsilon > 0$ arbitrarily. Then, for every $(\boldsymbol{A}, \boldsymbol{B}) \in C$, there must exist a $J(\boldsymbol{A}, \boldsymbol{B}) \in \mathbb{N}$ such that

$$\mathrm{P}[(\boldsymbol{A}, \boldsymbol{B}, \mathbf{X}) \in \mathcal{A}_{J(\boldsymbol{A},\boldsymbol{B})}] \geq 1 - \varepsilon. \tag{7.11}$$

Next, for every $(\boldsymbol{A}, \boldsymbol{B}) \in C$, let

$$\mathcal{U}_{\boldsymbol{A},\boldsymbol{B}} = \{\boldsymbol{X} \in \mathcal{U} : (\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{X}) \in \mathcal{A}_{J(\boldsymbol{A},\boldsymbol{B})}\}. \tag{7.12}$$

Since $\mathrm{P}[\mathsf{X} \in \mathcal{U}] = 1$ by assumption, (7.11) yields

$$\mathrm{P}[\mathsf{X} \in \mathcal{U}_{\boldsymbol{A},\boldsymbol{B}}] \geq 1 - \varepsilon, \quad \text{for all } (\boldsymbol{A}, \boldsymbol{B}) \in \mathcal{C}. \tag{7.13}$$

Now, consider $(\boldsymbol{A}, \boldsymbol{B}) \in \mathcal{C}$ and fix $\boldsymbol{U}, \boldsymbol{V} \in \mathcal{U}_{\boldsymbol{A},\boldsymbol{B}}$ with $\boldsymbol{U} \neq \boldsymbol{V}$ but arbitrary otherwise. It follows that $\boldsymbol{U} - \boldsymbol{V} \in \mathcal{U}_{\mathbf{V}} \setminus \{\mathbf{0}\}$ and $(\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{V}) \in \mathcal{A}_{J(\boldsymbol{A},\boldsymbol{B})}$, and the definition of $\mathcal{A}_{J(\boldsymbol{A},\boldsymbol{B})}$ (see (7.5)) yields

$$\|\boldsymbol{U} - \boldsymbol{V}\|_2^{\frac{1}{\beta}} \leq J(\boldsymbol{A}, \boldsymbol{B}) \|(\boldsymbol{a}_1^{\mathsf{T}}(\boldsymbol{U} - \boldsymbol{V})\boldsymbol{b}_1 \ \ldots \ \boldsymbol{a}_k^{\mathsf{T}}(\boldsymbol{U} - \boldsymbol{V})\boldsymbol{b}_k)^{\mathsf{T}}\|_2. \tag{7.14}$$

By [33, Lemma 2], we can therefore conclude that, for every $(\boldsymbol{A}, \boldsymbol{B}) \in \mathcal{C}$, the mapping

$$f_{\boldsymbol{A},\boldsymbol{B}} \colon \mathcal{U}_{\boldsymbol{A},\boldsymbol{B}} \to \{(\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_1 \ \ldots \ \boldsymbol{a}_k^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_k)^{\mathsf{T}} : \boldsymbol{X} \in \mathcal{U}_{\boldsymbol{A},\boldsymbol{B}}\} \tag{7.15}$$
$$\boldsymbol{X} \mapsto (\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_1 \ \ldots \ \boldsymbol{a}_k^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_k)^{\mathsf{T}} \tag{7.16}$$

is injective with β-Hölder continuous inverse $f_{\boldsymbol{A},\boldsymbol{B}}^{-1}$. Finally, for every $(\boldsymbol{A}, \boldsymbol{B}) \in \mathcal{C}$, the mapping $f_{\boldsymbol{A},\boldsymbol{B}}^{-1}$ can be extended to the desired β-Hölder continuous mapping g on $\mathbb{R}^k$ by [27, Theorem 1, Item ii)]. ∎

8. Proof of Proposition 2.2

For every $j \in \mathbb{N}$, let $\mathcal{A}(j)$ and $\mathcal{B}(j)$ be as in (6.1) and (6.2), respectively. By countable subadditivity of Lebesgue measure, it suffices to show that

$$\inf \left\{ \frac{\|(\boldsymbol{a}_1^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_1 \ \ldots \ \boldsymbol{a}_k^{\mathsf{T}}\boldsymbol{X}\boldsymbol{b}_k)^{\mathsf{T}}\|_2}{\|\boldsymbol{X}\|_2^{1/\beta}} : \boldsymbol{X} \in \mathcal{U} \setminus \{\mathbf{0}\} \right\} > 0, \tag{8.1}$$

for Lebesgue a.a. $\big((\boldsymbol{a}_1 \ldots \boldsymbol{a}_k), (\boldsymbol{b}_1 \ldots \boldsymbol{b}_k)\big) \in \mathcal{A}(j) \times \mathcal{B}(j)$ and all $j \in \mathbb{N}$. Owing to Lemma 8.1 below, (8.1) then holds, for all $j \in \mathbb{N}$, with probability 1 if the deterministic matrices $\big((\boldsymbol{a}_1 \ldots \boldsymbol{a}_k), (\boldsymbol{b}_1 \ldots \boldsymbol{b}_k)\big) \in \mathcal{A}(j) \times \mathcal{B}(j)$ are replaced by independent random matrices with columns a_i, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_m(\mathbf{0}, j)$, and columns b_i, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_n(\mathbf{0}, j)$. By countable subadditivity of Lebesgue measure, this finally implies that (2.10) can be violated only on a set of Lebesgue measure zero, which finalizes the proof.

Lemma 8.1. *Let $s > 0$ and take $\mathsf{A} = (\mathsf{a}_1 \ \ldots \ \mathsf{a}_k)$ and $\mathsf{B} = (\mathsf{b}_1 \ \ldots \ \mathsf{b}_k)$ to be independent random matrices with columns a_i, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_m(\mathbf{0}, s)$, and columns b_i, $i = 1, \ldots, k$, independent and uniformly distributed on $\mathcal{B}_n(\mathbf{0}, s)$. Consider a nonempty and bounded set $\mathcal{U} \subseteq \mathbb{R}^{m \times n}$, and suppose that there* *exists a* $\beta \in (0,1)$ *such that*

$$\frac{\overline{\dim}_{\mathrm{B}}(\mathcal{U})}{k} < 1-\beta. \tag{8.2}$$

Then,

$$\mathrm{P}\left[\inf\left\{\frac{\|(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T}\|_2}{\|\boldsymbol{X}\|_2^{1/\beta}} : \boldsymbol{X} \in \mathcal{U}\backslash\{\mathbf{0}\}\right\} > 0\right] = 1. \tag{8.3}$$

Proof. Since $\mathcal{U}$ is bounded by assumption, there exists a $K > 0$ such that

$$\sigma_1(\boldsymbol{X}) \le K, \quad \text{for all } \boldsymbol{X} \in \mathcal{U}. \tag{8.4}$$

For every $j \in \mathbb{N}$, set

$$\mathcal{U}_j = \mathcal{U}\backslash\mathcal{B}_{m\times n}\big(\mathbf{0}, 2^{-\beta j}\big). \tag{8.5}$$

Then, (8.4) together with (8.5), upon using $\sigma_1(\boldsymbol{X}) \ge \|\boldsymbol{X}\|_2/rank(\boldsymbol{X})$, yields

$$\frac{2^{-\beta j}}{\sqrt{m}} \le \sigma_1(\boldsymbol{X}) \le K, \quad \text{for all } \boldsymbol{X} \in \mathcal{U}_j \text{ and } j \in \mathbb{N}. \tag{8.6}$$

By Lemma 8.2 below, it is sufficient to show that

$$\mathrm{P}\left[\exists J : \|(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T}\|_2 \ge 2^{-j}, \text{ for all } \boldsymbol{X} \in \mathcal{U}_j, j \ge J\right] = 1. \tag{8.7}$$

This will be established by arguing as follows. Suppose we can prove that there exists a $J \in \mathbb{N}$ such that

$$\sum_{j=J}^{\infty} \mathrm{P}\left[\exists \boldsymbol{X} \in \mathcal{U}_j : \|(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T}\|_2 < 2^{-j}\right] < \infty. \tag{8.8}$$

Then, the Borel-Cantelli Lemma [13, Theorem 2.3.1] implies

$$\mathrm{P}\left[\exists \boldsymbol{X} \in \mathcal{U}_j : \|(\mathbf{a}_1^\mathsf{T}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^\mathsf{T}\boldsymbol{X}\mathbf{b}_k)^\mathsf{T}\|_2 < 2^{-j}, \text{ for infinitely many } j \in \mathbb{N}\right] = 0, \tag{8.9}$$

which, in turn, implies (8.7).

It remains to establish (8.8), which will be effected through a covering argument. For every $j \in \mathbb{N}$, consider the covering ball center $\boldsymbol{Y}_i^{(j)} \in \mathcal{U}_j$ such that

$$\mathcal{U}_j \subseteq \bigcup_{i=1}^{N_{\mathcal{U}_j}(2^{-j})} \mathcal{B}_{m\times n}\Big(\boldsymbol{Y}_i^{(j)}, 2^{-j}\Big). \tag{8.10}$$

A union bound argument then yields

$$\mathrm{P}\left[\exists \boldsymbol{X} \in \mathcal{U}_j : \|(\mathbf{a}_1^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_k)^{\mathsf{T}}\|_2 < 2^{-j}\right] \tag{8.11}$$

$$\leq \sum_{i=1}^{N_{\mathcal{U}_j}(2^{-j})} \mathrm{P}\left[\exists \boldsymbol{X} \in \mathcal{B}_{m\times n}(\boldsymbol{Y}_i^{(j)}, 2^{-j}) : \|(\mathbf{a}_1^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_k)^{\mathsf{T}}\|_2 < 2^{-j}\right]. \tag{8.12}$$

Next, choose $J_1 \in \mathbb{N}$ such that

$$2^{-J_1(1-\beta)} \leq \frac{s^2}{(1+s^2\sqrt{k})\sqrt{m}}. \tag{8.13}$$

This implies $(1+s^2\sqrt{k})2^{-j} \leq \frac{2^{-\beta j}}{\sqrt{m}} s^2$, for all $j \geq J_1$, and thus, by (8.6), $(1+s^2\sqrt{k})2^{-j} \leq \sigma_1(\boldsymbol{X})s^2$, for all $\boldsymbol{X} \in \mathcal{U}_j$. Hence, for all $j \geq J_1$, we can bound each summand in (8.12) according to

$$\mathrm{P}\left[\exists \boldsymbol{X} \in \mathcal{B}_{m\times n}(\boldsymbol{Y}_i^{(j)}, 2^{-j}) : \|(\mathbf{a}_1^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_k)^{\mathsf{T}}\|_2 < 2^{-j}\right] \tag{8.14}$$

$$\leq \mathrm{P}\left[\left\|\left(\mathbf{a}_1^{\mathsf{T}}\boldsymbol{Y}_i^{(j)}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^{\mathsf{T}}\boldsymbol{Y}_i^{(j)}\mathbf{b}_k\right)^{\mathsf{T}}\right\|_2 < (1+s^2\sqrt{k})2^{-j}\right] \tag{8.15}$$

$$\leq (1+s^2\sqrt{k})^k 2^{-jk} \frac{2^{\frac{k(m+n)}{2}}}{\sigma_1(\boldsymbol{Y}_i^{(j)})^k s^{2k}} \left(1+\log\left(\frac{s^2\sigma_1(\boldsymbol{Y}_i^{(j)})}{(1+s^2\sqrt{k})2^{-j}}\right)\right)^k \tag{8.16}$$

$$\leq (s^{-2}+\sqrt{k})^k m^{\frac{k}{2}} 2^{-jk(1-\beta)} 2^{\frac{k(m+n)}{2}} \left(1+\log\left(\frac{s^2K}{1+s^2\sqrt{k}}\right)+j\log 2\right)^k, \tag{8.17}$$

where (8.15) is by (6.17)–(6.20) for $\varepsilon_i = 2^{-j}$, in (8.16) we applied Lemma 6.2 with $\delta = (1+s^2\sqrt{k})2^{-j}$ and $\boldsymbol{X} = \boldsymbol{Y}_i^{(j)}$, and in (8.17) we used (8.6). Inserting (8.14)–(8.17) into (8.11)–(8.12) results in

$$\mathrm{P}\left[\exists \boldsymbol{X} \in \mathcal{U}_j : \|(\mathbf{a}_1^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_1 \ \dots \ \mathbf{a}_k^{\mathsf{T}}\boldsymbol{X}\mathbf{b}_k)^{\mathsf{T}}\|_2 < 2^{-j}\right] \tag{8.18}$$

$$\leq C N_{\mathcal{U}}(2^{-j}) 2^{-jk(1-\beta)} (D + j\log 2)^k, \quad \text{for all } j \geq J_1, \tag{8.19}$$

with

$$C = (s^{-2}+\sqrt{k})^k m^{\frac{k}{2}} 2^{\frac{k(m+n)}{2}} \tag{8.20}$$

and

$$D = 1 + \log\left(\frac{s^2K}{1+s^2\sqrt{k}}\right). \tag{8.21}$$

Next, let

$$d = \frac{\overline{\dim}_{\mathrm{B}}(\mathcal{U}) + k(1-\beta)}{2}, \tag{8.22}$$

which implies $\overline{\dim}_{\mathrm{B}}(\mathcal{U}) < d < k(1-\beta)$ (see (2.9)). By (1.4) we have

$$\overline{\dim}_{\mathrm{B}}(\mathcal{U}) = \inf_{\ell \in \mathbb{N}} \sup_{j \geq \ell} \frac{\log N_{\mathcal{U}}(2^{-j})}{\log\left(2^{j}\right)}. \tag{8.23}$$

Thus, as a consequence of $d > \overline{\dim}_{\mathrm{B}}(\mathcal{U})$, there exists a $J_2 \in \mathbb{N}$ such that

$$N_{\mathcal{U}}\left(2^{-j}\right) \leq 2^{jd}, \quad \text{for all } j \geq J_2. \tag{8.24}$$

Now set $J = \max(J_1, J_2)$. Then, we have

$$\sum_{j=J}^{\infty} \mathrm{P}\left[\exists X \in \mathcal{U}_j : \|(\mathbf{a}_1^{\mathsf{T}} X \mathbf{b}_1 \ \ldots \ \mathbf{a}_k^{\mathsf{T}} X \mathbf{b}_k)^{\mathsf{T}}\|_2 < 2^{-j}\right] \tag{8.25}$$

$$\leq C \sum_{j=J}^{\infty} N_{\mathcal{U}}(2^{-j}) 2^{-jk(1-\beta)} (D + j \log 2)^k \tag{8.26}$$

$$\leq C \sum_{j=J}^{\infty} 2^{-j(k(1-\beta)-d)} (D + j \log 2)^k \tag{8.27}$$

$$< \infty, \tag{8.28}$$

where in (8.26) we used (8.18)–(8.19), (8.27) is by (8.24), and (8.28) follows from $d < k(1-\beta)$. ■

Lemma 8.2. *Consider a nonempty and bounded set $\mathcal{U} \subseteq \mathbb{R}^{m \times n} \setminus \{\mathbf{0}\}$ and let $f : \mathcal{U} \to \mathbb{R}^k$. Fix $\beta \in (0,1)$, and suppose that there exists a $J \in \mathbb{N}$ such that*

$$\|f(X)\|_2 \geq 2^{-j}, \quad \textit{for all } X \in \mathcal{U} \setminus \mathcal{B}_{m \times n}\left(\mathbf{0}, 2^{-\beta j}\right) \textit{ and } j \geq J. \tag{8.29}$$

Then, we have

$$\inf \left\{ \frac{\|f(X)\|_2}{\|X\|_2^{1/\beta}} : X \in \mathcal{U} \right\} > 0. \tag{8.30}$$

Proof. Follows from [33, Lemma 3] through vectorization. ■

A. Proof of Lemma 3.1

Item i) follows from [1, Lemma III.1, Item i)] through vectorization.

In order to prove Item ii), we first note that the sets $\mathcal{U}_i$ participating in $\mathcal{U}$ are all s-rectifiable by Item i). To see that a finite union of s-rectifiable sets is s-rectifiable, we first prove the statement for two sets and then note that the generalization to finitely many sets follows by induction. Let $\mathcal{A}$ and $\mathcal{B}$ be s-rectifiable. By the definition of rectifiability, there exist compact sets $\mathcal{C}, \mathcal{D} \subseteq \mathbb{R}^s$ and Lipschitz mappings $\varphi\colon \mathcal{C} \to \mathbb{R}^{m\times n}$ and $\psi\colon \mathcal{D} \to \mathbb{R}^{m\times n}$ such that $\mathcal{A} = \varphi(\mathcal{C})$ and $\mathcal{B} = \psi(\mathcal{D})$. As the sets $\mathcal{C}$ and $\mathcal{D}$ are compact, there exists a constant $R > 0$ such that $\mathcal{C} \cup \mathcal{D} \subseteq \mathcal{B}_s(\mathbf{0}, R)$. The set $\mathcal{D} + \{3R\}$ is thus disjoint from $\mathcal{C}$. We now define the function

$$\tilde{\varphi}\colon \mathcal{C} \cup (\mathcal{D} + \{3R\}) \to \mathbb{R}^{m\times n}$$
$$\boldsymbol{x} \mapsto \begin{cases} \varphi(\boldsymbol{x}), & \boldsymbol{x} \in \mathcal{C} \\ \psi(\boldsymbol{x} - 3R), & \boldsymbol{x} \in \mathcal{D} + \{3R\} \end{cases}.$$

The set $\mathcal{C} \cup (\mathcal{D} + \{3R\}) \subseteq \mathbb{R}^s$ is compact as the union of compact sets and $\tilde{\varphi}\big(\mathcal{C} \cup (\mathcal{D} + \{3R\})\big) = \mathcal{A} \cup \mathcal{B}$. It remains to establish that $\tilde{\varphi}$ is Lipschitz. Indeed, for vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{C}$ the Lipschitz property follows from the Lipschitz property of φ. Analogously, for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{D} + \{3R\}$ the Lipschitz property is inherited from that of ψ. For $\boldsymbol{x} \in \mathcal{C}$ and $\boldsymbol{y} \in \mathcal{D} + \{3R\}$, we have that $\|\boldsymbol{x} - \boldsymbol{y}\| \geq R$ and $\|\tilde{\varphi}(\boldsymbol{x}) - \tilde{\varphi}(\boldsymbol{y})\| \leq 2\max_{\boldsymbol{z} \in \mathcal{C}\cup(\mathcal{D}+\{3R\})} \|\tilde{\varphi}(\boldsymbol{z})\| =: M$. Thus, $\|\tilde{\varphi}(\boldsymbol{x}) - \tilde{\varphi}(\boldsymbol{y})\| \leq \frac{M}{R}\|\boldsymbol{x} - \boldsymbol{y}\|$ and we obtain Lipschitz continuity of $\tilde{\varphi}$ with Lipschitz constant given by the maximum of $\frac{M}{R}$ and the Lipschitz constants of φ and ψ.

To prove Item iii), let $\mathcal{U} \in \mathbb{R}^{m_1\times n_1}$ be s-rectifiable and $\mathcal{V} \in \mathbb{R}^{m_2\times n_2}$ t-rectifiable. By the definition of rectifiability, there exist compact sets $\mathcal{C} \subseteq \mathbb{R}^s$ and $\mathcal{D} \subseteq \mathbb{R}^t$ and Lipschitz mappings $\varphi\colon \mathcal{C} \to \mathbb{R}^{m_1\times n_1}$ and $\psi\colon \mathcal{D} \to \mathbb{R}^{m_2\times n_2}$ such that $\mathcal{U} = \varphi(\mathcal{C})$ and $\mathcal{V} = \psi(\mathcal{D})$. We can therefore write $\mathcal{U} \times \mathcal{V} = (\varphi \times \psi)(\mathcal{C} \times \mathcal{D})$ with $\mathcal{C} \times \mathcal{D} \subseteq \mathbb{R}^{s+t}$ compact and $\varphi \times \psi\colon \mathcal{C} \times \mathcal{D} \to \mathbb{R}^{m_1\times n_1} \times \mathbb{R}^{m_2\times n_2}$ Lipschitz.

It remains to establish Item iv). Let $\mathcal{K}$ be a compact subset of an s-dimensional C^1-submanifold $\mathcal{M} \subseteq \mathbb{R}^{m\times n}$. The statement is trivial if $\mathcal{K} = \emptyset$. We hence assume that $\mathcal{K}$ is nonempty. By [23, Definition 5.3.1], we can write

$$\mathcal{M} = \bigcup_{\boldsymbol{X} \in \mathcal{M}} \varphi_{\boldsymbol{X}}(\mathcal{U}_{\boldsymbol{X}}), \tag{A.1}$$

where, for every $\boldsymbol{X} \in \mathcal{M}$, $\mathcal{U}_{\boldsymbol{X}} \subseteq \mathbb{R}^s$ is open, and $\varphi_{\boldsymbol{X}}\colon \mathcal{U}_{\boldsymbol{X}} \to \mathbb{R}^{m\times n}$ is a one-to-one C^1-map satisfying $\boldsymbol{X} \in \varphi_{\boldsymbol{X}}(\mathcal{U}_{\boldsymbol{X}})$ and $\varphi_{\boldsymbol{X}}(\mathcal{U}_{\boldsymbol{X}}) = \mathcal{V}_{\boldsymbol{X}} \cap \mathcal{M}$ with $\mathcal{V}_{\boldsymbol{X}} \subseteq \mathbb{R}^{m\times n}$ open. Since there exists a real analytic diffeomorphism between $\mathbb{R}^s$ and $\mathcal{B}_s(\mathbf{0}, 1)$ [1, Lemma K.10], we can assume, w.l.o.g., that the sets $\mathcal{U}_{\boldsymbol{X}}$ are all bounded. As $\mathcal{K} \subseteq \mathcal{M}$ is compact by assumption, there must exist a finite set $\{X_i : i = 1, \ldots, N\} \subseteq \mathcal{M}$ such that

$$\mathcal{K} \subseteq \bigcup_{i=1}^{N} \varphi_{X_i}(\mathcal{U}_{X_i}) \tag{A.2}$$

and $\mathcal{V}_{X_i} \cap \mathcal{K} \neq \emptyset$, for $i = 1, \ldots, N$. With the set $\{X_i : i = 1, \ldots, N\} \subseteq \mathcal{M}$, we can now write

$$\mathcal{K} = \bigcup_{i=1}^{N} (\varphi_{X_i}(\mathcal{U}_{X_i}) \cap \mathcal{K}) \tag{A.3}$$

$$= \bigcup_{i=1}^{N} \varphi_i(\mathcal{U}_i) \tag{A.4}$$

$$= \bigcup_{i=1}^{N} \varphi_i(\overline{\mathcal{U}}_i), \tag{A.5}$$

where in (A.4) we set $\varphi_i = \varphi_{X_i}$ and $\mathcal{U}_i = \mathcal{U}_{X_i} \cap \varphi_i^{-1}(\mathcal{K})$, and (A.5) is by $\mathcal{K} = \overline{\mathcal{K}}$ and the continuity of φ_i. The claim now follows from Item ii) applied to A.5.

References

[1] G. Alberti, H. Bölcskei, C. De Lellis, G. Koliander, and E. Riegler, Lossless analog compression. *IEEE Trans. Inf. Theory* **65** (2019), no. 11, 7480–7513

[2] K. Barański, Y. Gutman, and A. Śpiewak, A probabilistic Takens theorem. *Nonlinearity* **33** (2020), no. 9, 4940–4966

[3] M. F. Barnsley, J. H. Elton, and D. P. Hardin, *Recurrent iterated function systems*, pp. 3–31. Springer, Boston, MA, 1989

[4] R. G. Bartle, *The elements of integration and Lebesgue measure*. Wiley, New York, NY, 1995

[5] T. T. Cai and A. Zhang, Sparse representation of a polytope and recovery of sparse signals and low-rank matrices. *IEEE Trans. Inf. Theory* **60** (2014), no. 1, 122–132

[6] T. T. Cai and A. Zhang, ROP: Matrix recovery via rank-one projections. *Ann. Stat.* **43** (2015), no. 1, 102–138

[7] E. J. Candès and Y. Plan, Matrix completion with noise. *Proc. IEEE* **98** (2010), no. 6, 925–936

[8] E. J. Candès and Y. Plan, Tight oracle inequalities for low-rank matrix recovery from a minimal number of noisy random measurements. *IEEE Trans. Inf. Theory* **4** (2011), no. 57, 2342–2359

[9] E. J. Candès and B. Recht, Exact matrix completion via convex optimization. *Found. Comput. Math.* **9** (2009), no. 6, 717–772

[10] E. J. Candès and T. Tao, The power of convex relaxation: Near-optimal matrix completion. *IEEE Trans. Inf. Theory* **56** (2010), no. 5, 2053–2080

[11] Y. Chikuse, *Statistics on special manifolds*. Springer, New York, NY, 2003

[12] D. L. Donoho, M. Gavish, and A. Montanari, The phase transition of matrix recovery from gaussian measurements matches the minimax MSE of matrix denoising. *Proc. Natl. Acad. Sci. USA* **110** (2013), no. 21, 8405–8410

[13] R. Durett, *Probability: Theory and examples*. 4th edn., Cambridge Univ. Press, New York, NY, 2010

[14] Y. C. Eldar, D. Needell, and Y. Plan, Uniqueness conditions for low-rank matrix recovery. *Appl. Comput. Harmon. Anal.* **33** (2012), no. 2, 309–314

[15] K. Falconer, *Fractal geometry*. 3rd edn., Wiley, New York, NY, 2014

[16] H. Federer, *Geometric measure theory*. Springer, New York, NY, 1969

[17] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing*. Birkhäuser, Basel, Switzerland, 2013

[18] D. Gross, Recovering low-rank matrices from few coefficients in any basis. *IEEE Trans. Inf. Theory* **57** (2011), no. 3, 1548–1566

[19] R. A. Horn and C. R. Johnson, *Matrix analysis*. 2nd edn., Cambridge Univ. Press, Cambridge, U.K., 2013

[20] R. H. Keshavan, A. Montanari, and S. Oh, Matrix completion from a few entries. *IEEE Trans. Inf. Theory* **56** (2010), no. 6, 2980–2998

[21] R. H. Keshavan, A. Montanari, and S. Oh, Matrix completion from noisy entries. *J. Mach. Learn. Res.* **11** (2010), 2057–2078

[22] O. Klopp, Y. Lu, A. B. Tsybakov, and H. H. Zhou, Structured matrix estimation and completion. *Bernoulli* **25** (2019), no. 4B, 3883–3911

[23] S. G. Krantz and H. R. Parks, *Geometric integration theory*. Birkhäuser, Princeton, NJ, 2008

[24] J. M. Lee, *Introduction to smooth manifolds*. 1st edn., Springer, New York, NY, 2000

[25] Y. Li, K. Lee, and Y. Bresler, Identifiability and stability in blind deconvolution under minimal assumptions. *IEEE Trans. Inf. Theory* **63** (2017), no. 7, 4619–4633

[26] P. Mattila, *Geometry of sets and measures in Euclidean space: Fractals and rectifiability*. Cambridge Univ. Press, Cambridge, UK, 1995

[27] G. J. Minty, On the extension of Lipschitz, Lipschitz-Hölder continuous, and monotone functions. *Bull. Am. Math. Soc.* **76** (1970), no. 2, 334–339

[28] B. Recht, A simpler approach to matrix completion. *J. Mach. Learn. Res.* **12** (2011), 341–3430

[29] B. Recht, M. Fazel, and P. A. Parrilo, Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization. *SIAM Rev.* **52** (2010), no. 3, 471–501

[30] E. Riegler, D. Stotz, and H. Bölcskei, Information-theoretic limits of matrix completion. In *Proc. IEEE Int. Symp. Inf. Theory*, pp. 1836–1840, 2015

[31] J. C. Robinson, *Dimensions, embeddings, and attractors*. Cambridge University Press, Cambridge, NY, 2011

[32] Y. Rong, Y. Wang, and Z. Xu, Almost everywhere injectivity conditions for the matrix recovery problem. *Appl. Comput. Harmon. Anal.* **50** (2021), 386–400

[33] D. Stotz, E. Riegler, E. Agustsson, and H. Bölcskei, Almost lossless analog signal separation and probabilistic uncertainty relations. *IEEE Trans. Inf. Theory* **63** (2017), no. 9, 5445–5460

[34] V. Y. F. Tan, L. Balzano, and S. C. Draper, Rank minimization over finite fields: Fundamental limits and coding-theoretic interpretations. *IEEE Trans. Inf. Theory* **58** (2012), no. 4, 2018–2039

[35] Y. Wang and Z. Xu, Generalized phase retrieval: Measurement number, matrix recovery and beyond. *Appl. Comput. Harmon. Anal.* **47** (2019), no. 2, 423–446

[36] R. O. Wells Jr., *Differential analysis on complex manifolds*. 3rd edn., Springer, 2008

[37] Y. Wu and S. Verdú, Rényi information dimension: Fundamental limits of almost lossless analog compression. *IEEE Trans. Inf. Theory* **56** (2010), no. 8, 3721–3748

[38] Z. Xu, The minimal measurement number for low-rank matrix recovery. *Appl. Comput. Harmon. Anal.* **44** (2018), no. 2, 497–508

Erwin Riegler
Chair for Mathematical Information Science, ETH Zurich, Sternwartstrasse 7, 8092 Zurich, Switzerland; eriegler@mins.ee.ethz.ch

Günther Koliander
ARI, Austrian Academy of Sciences, Wohllebengasse 12-14, 1040 Vienna, Austria; guenther.koliander@oeaw.ac.at

David Stotz
Kantonsschule Schaffhausen, Pestalozzistrasse 20, 8200 Schaffhausen, Switzerland; david-stotz@web.de

Helmut Bölcskei
Chair for Mathematical Information Science, ETH Zurich, Sternwartstrasse 7, 8092 Zurich, Switzerland; hboelcskei@ethz.ch