Information-Theoretic Limits of Matrix Completion

Erwin Riegler, David Stotz, and Helmut Bölcskei

Dept. IT & EE, ETH Zurich, Switzerland
Email: {eriegler, dstotz, boelcskei}@nari.ee.ethz.ch

***Abstract*—We propose an information-theoretic framework for matrix completion. The theory goes beyond the low-rank structure and applies to general matrices of "low description complexity". Specifically, we consider random matrices $\mathbf{X} \in \mathbb{R}^{m\times n}$ of arbitrary distribution (continuous, discrete, discrete-continuous mixture, or even singular). With $\mathcal{S} \subseteq \mathbb{R}^{m\times n}$ an ε-support set of X, i.e., $\mathrm{P}[\mathbf{X} \in \mathcal{S}] \geq 1-\varepsilon$, and $\underline{\dim}_{\mathrm{B}}(\mathcal{S})$ denoting the lower Minkowski dimension of $\mathcal{S}$, we show that $k > \underline{\dim}_{\mathrm{B}}(\mathcal{S})$ measurements of the form $\langle \mathrm{A}_i, \mathrm{X}\rangle$, with A_i denoting the measurement matrices, suffice to recover X with probability of error at most ε. The result holds for Lebesgue a.a. A_i and does not need incoherence between the A_i and the unknown matrix X. We furthermore show that $k > \underline{\dim}_{\mathrm{B}}(\mathcal{S})$ measurements also suffice to recover the unknown matrix X from measurements taken with rank-one A_i, again this applies to a.a. rank-one A_i. Rank-one measurement matrices are attractive as they require less storage space than general measurement matrices and can be applied faster. Particularizing our results to the recovery of low-rank matrices, we find that $k > (m+n-r)r$ measurements are sufficient to recover matrices of rank at most r. Finally, we construct a class of rank-r matrices that can be recovered with arbitrarily small probability of error from $k < (m+n-r)r$ measurements.**

I. 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 of its entries. This problem arises in a wide range of applications, including quantum state tomography, face recognition, recommender systems, and sensor localization (see, e.g., [1] and references therein).

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

$$\mathrm{y} = (\langle \mathrm{A}_1, \mathrm{X}\rangle, ..., \langle \mathrm{A}_k, \mathrm{X}\rangle)^{\mathsf{T}} \in \mathbb{R}^k \tag{1}$$

where $\mathrm{A}_i \in \mathbb{R}^{m\times n}$ denotes the measurement matrices and $\langle \cdot, \cdot \rangle$ stands for the standard trace inner product between matrices in $\mathbb{R}^{m\times n}$. The number of measurements k is typically much smaller than the total number of entries, mn, of X. Depending on the A_i, the measurements can simply be individual entries of X or general linear combinations thereof.

The vast literature on matrix completion, for a highly incomplete list see [1]–[8], provides guarantees for the recovery of the unknown low-rank matrix X from the measurements y, under various assumptions on the measurement matrices A_i and the low-rank models generating X. For example, in [2] the A_i are assumed to be chosen randomly from an orthonormal basis for $\mathbb{R}^{n\times n}$ and it is shown that an unknown $n \times n$ matrix X of rank at most r can be recovered with high probability if $k = \mathcal{O}(nr\nu \ln^2 n)$. Here, ν quantifies the incoherence between the unknown matrix X and the orthonormal basis for $\mathbb{R}^{n\times n}$ the A_i are drawn from.

The setting in [3] assumes random measurement matrices $\mathbf{A}_i$ with the position of the only nonzero entry chosen uniformly at random. It is shown that almost all (a.a.) matrices (with respect to the random orthogonal model [3, Def. 2.1]) of rank at most 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 numerical constant.

In [1] it is shown that for measurement matrices $\mathbf{A}_i$ containing i.i.d. entries (that are, e.g., Gaussian), a matrix X of rank at most r can be recovered with high probability from $k \geq C(m+n)r$ measurements, where C is a constant. The recovery guarantees in [1]–[8] all pertain to recovery through nuclear norm minimization. In [9] measurement matrices $\mathbf{A}_i$ 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 an $n \times n$ matrix X of rank at most r with probability one if $k > (2n - r)r$. It is furthermore shown in [9] that all matrices X of rank at most $n/2$ can be recovered, again with probability one, provided that $k \geq 4nr - 4r^2$. The recovery thresholds in [1], [9] do not exhibit a $\log n$ term, but assume significant richness in the random measurement matrices $\mathbf{A}_i$. Storing and applying such measurement matrices is costly in terms of memory and computation time. To overcome this problem [8] considers rank-one measurement matrices of the form $\mathbf{A}_i = \mathbf{a}_i \mathbf{b}_i^{\mathsf{T}}$, where $\mathbf{a}_i \in \mathbb{R}^m$ and $\mathbf{b}_i \in \mathbb{R}^n$ are independent with i.i.d. Gaussian or sub-Gaussian entries, and shows that nuclear norm minimization succeeds under the same recovery threshold as in [1], namely $k \geq C(m+n)r$.

Contributions: Inspired by the work of Wu and Verdú on analog signal compression [10], we formulate an information-theoretic framework for almost lossless matrix completion. The theory is general in the sense of going beyond the low-rank structure and applying to general matrices of "low description complexity". Specifically, we consider random matrices $\mathbf{X} \in \mathbb{R}^{m\times n}$ of arbitrary distribution (continuous, discrete, discrete-continuous mixture, or even singular). With $\mathcal{S} \subseteq \mathbb{R}^{m\times n}$ an ε-support set of $\mathbf{X}$, i.e., $\mathrm{P}[\mathbf{X} \in \mathcal{S}] \geq 1-\varepsilon$, and $\underline{\dim}_{\mathrm{B}}(\mathcal{S})$ denoting the lower Minkowski dimension (see Definition 3) of $\mathcal{S}$, we show that $k > \underline{\dim}_{\mathrm{B}}(\mathcal{S})$ measurements suffice to recover $\mathbf{X}$ with probability of error no more than ε. The result holds for Lebesgue a.a. measurement matrices A_i and does not need any incoherence between the A_i and the unknown matrix $\mathbf{X}$. What is more, we show that $k > \underline{\dim}_{\mathrm{B}}(\mathcal{S})$ measurements also suffice for recovery from measurements taken with rank-one A_i, again this applies to a.a. rank-one A_i.

Particularizing our results to low-rank matrices $\mathbf{X}$, we show that $\mathbf{X}$ of rank at most r can be recovered from $k > (m+n-r)r$ measurements taken with either general A_i or with rank-one A_i. Perhaps surprisingly, it turns out that, depending on the specific distribution of the low-rank matrix $\mathbf{X}$, even fewer than $(m+n-r)r$ measurements can suffice. We construct a class of examples that illuminates this phenomenon.

Notation: Roman letters $\mathrm{A}, \mathrm{B}, \ldots$ designate deterministic matrices and $\mathrm{a}, \mathrm{b}, \ldots$ stands for deterministic vectors. Boldface letters $\mathbf{A}, \mathbf{B}, \ldots$ and $\mathbf{a}, \mathbf{b}, \ldots$ denote random matrices and vectors, respectively. For the distribution of a random matrix $\mathbf{A}$ we write $\mu_{\mathbf{A}}$ and we use $\mu_{\mathbf{a}}$ to designate the distribution of a random vector $\mathbf{a}$. The superscript $^{\mathsf{T}}$ stands for transposition. For $\mathrm{A} = (\mathrm{a}_1, \ldots, \mathrm{a}_n) \in \mathbb{R}^{m\times n}$ we let $\mathrm{vec}(\mathrm{A}) = (\mathrm{a}_1^{\mathsf{T}}, \ldots, \mathrm{a}_n^{\mathsf{T}})^{\mathsf{T}}$. For a rank-$r$ matrix $\mathrm{A} \in \mathbb{R}^{m\times n}$ with ordered singular values $\sigma_1(\mathrm{A}) \geq \ldots \geq \sigma_r(\mathrm{A})$, we set $\Delta(\mathrm{A}) = \prod_{i=1}^{r} \sigma_i(\mathrm{A})$. For a matrix A, $\mathrm{tr}(\mathrm{A})$ denotes its trace. For matrices A, B of the same dimensions, $\langle \mathrm{A}, \mathrm{B}\rangle = \mathrm{tr}(\mathrm{A}^{\mathsf{T}}\mathrm{B})$ is the trace inner product between A and B. We write $\|\mathrm{A}\|_2 = \sqrt{\langle \mathrm{A}, \mathrm{A}\rangle}$ for the Euclidean norm of the matrix A. For the Euclidean space $(\mathbb{R}^k, \|\cdot\|_2)$, we denote the open ball of radius s centered at $\mathrm{u} \in \mathbb{R}^k$ by $\mathcal{B}_k(\mathrm{u}, s)$, $V(k, s)$ and $A(k-1, s)$ stand for its volume and the area of its closure, respectively. Similarly, for the Euclidean space $(\mathbb{R}^{m\times n}, \|\cdot\|_2)$, we denote the open ball of radius s centered at $\mathrm{A} \in \mathbb{R}^{m\times n}$ by $\mathcal{B}_{m\times n}(\mathrm{A}, s)$. We write $\mathcal{M}_r^{m\times n}$ and $\mathcal{N}_r^{m\times n}$ for the set of matrices $\mathrm{A} \in \mathbb{R}^{m\times n}$ with $\mathrm{rank}(\mathrm{A}) \leq r$ and $\mathrm{rank}(\mathrm{A}) = r$, respectively.

II. Almost lossless matrix completion

We start by formulating the almost lossless matrix completion framework.

Definition 1. For a random matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ of arbitrary distribution $\mu_{\mathbf{X}}$ with Lebesgue decompositon $\mu_{\mathbf{X}} = \mu_{\mathbf{X}}^{\mathrm{c}} + \mu_{\mathbf{X}}^{\mathrm{d}} + \mu_{\mathbf{X}}^{\mathrm{s}}$ (continuous, discrete, and singular components, respectively), an $(m \times n, k)$ code consists of

(i) linear measurements $(\langle \mathrm{A}_1, \cdot\rangle, \ldots, \langle \mathrm{A}_k, \cdot\rangle)^{\mathsf{T}} : \mathbb{R}^{m\times n} \to \mathbb{R}^k$;

(ii) a measurable decoder $g : \mathbb{R}^k \to \mathbb{R}^{m\times n}$.

For given measurement matrices A_i, we say that a decoder g achieves error probability ε if

$$\mathrm{P}\big[g\big((\langle \mathrm{A}_1, \mathbf{X}\rangle, \ldots, \langle \mathrm{A}_k \mathbf{X}\rangle)^{\mathsf{T}}\big) \neq \mathbf{X}\big] \leq \varepsilon.$$

Definition 2. For $\varepsilon \geq 0$, we call a nonempty bounded set $\mathcal{S} \subseteq \mathbb{R}^{m\times n}$ an ε-support set of the random matrix $\mathbf{X} \in \mathbb{R}^{m\times n}$ if $\mathrm{P}[\mathbf{X} \in \mathcal{S}] \geq 1 - \varepsilon$.

Definition 3. (Minkowski dimension[1]) Let $\mathcal{S}$ be a nonempty bounded set in $\mathbb{R}^{m\times n}$. The lower Minkowski dimension of $\mathcal{S}$ is defined as

$$\underline{\dim}_{\mathrm{B}}(\mathcal{S}) = \liminf_{\rho\to 0} \frac{\log N_{\mathcal{S}}(\rho)}{\log \frac{1}{\rho}}$$

and the upper Minkowski dimension is

$$\overline{\dim}_{\mathrm{B}}(\mathcal{S}) = \limsup_{\rho\to 0} \frac{\log N_{\mathcal{S}}(\rho)}{\log \frac{1}{\rho}}$$

where $N_{\mathcal{S}}$ denotes the covering number of $\mathcal{S}$ given by

$$N_{\mathcal{S}}(\rho) = \min\Big\{k \in \mathbb{N} \mid \mathcal{S} \subseteq \bigcup_{i\in\{1,\ldots,k\}} \mathcal{B}_{m\times n}(\mathrm{M}_i, \rho),\ \mathrm{M}_i \in \mathbb{R}^{m\times n}\Big\}.$$

If $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) = \overline{\dim}_{\mathrm{B}}(\mathcal{S}) =: \dim_{\mathrm{B}}(\mathcal{S})$, we simply say that $\dim_{\mathrm{B}}(\mathcal{S})$ is the Minkowski dimension of $\mathcal{S}$.

[1]This quantity is sometimes also referred to as box-counting dimension, which is the origin for the subscript B in the notation $\dim_{\mathrm{B}}(\cdot)$ used below.

III. Main results

The following result formalizes the statement on the operationally relevant description complexity being given by the lower Minkowski dimensions of ε-support sets of $\mathbf{X}$.

Theorem 1. *Let $\mathcal{S} \subseteq \mathbb{R}^{m\times n}$ be an ε-support set of $\mathbf{X} \in \mathbb{R}^{m\times n}$. Then, for Lebesgue a.a. measurement matrices A_i, $i = 1, \ldots, k$, there exists a decoder achieving error probability ε, provided that $k > \underline{\dim}_{\mathrm{B}}(\mathcal{S})$.*

Proof. See Section V. ☐

Remark 1. The central conceptual element in the proof of Theorem 1 is the following probabilistic null space property, first reported in [11] in the context of almost lossless analog signal separation. For a.a. measurement matrices A_i, $i = 1, \ldots, k$, the dimension of the kernel of the mapping $\mathrm{X} \mapsto (\langle \mathrm{A}_1, \mathrm{X}\rangle, \ldots, \langle \mathrm{A}_k, \mathrm{X}\rangle)^{\mathsf{T}}$ is $mn - k$. If the lower Minkowski dimension of a set $\mathcal{S}$ is smaller than k, the set $\mathcal{S}$ will intersect the kernel of this mapping at most trivially. What is remarkable here is that the notions of Euclidean dimension (for the kernel of the mapping) and of lower Minkowski dimension (for $\mathcal{S}$) are compatible.

We next particularize Theorem 1 for low-rank matrices. To this end, we first establish an upper bound on $\overline{\dim}_{\mathrm{B}}(\mathcal{S})$ for nonempty and bounded subsets of $\mathcal{M}_r^{m\times n}$.

Lemma 1. *Let $\mathcal{S} \subseteq \mathcal{M}_r^{m\times n}$ be a nonempty bounded set. Then*

$$\overline{\dim}_{\mathrm{B}}(\mathcal{S}) \leq (m+n-r)r.$$

Proof. We can decompose $\mathcal{M}_r^{m\times n}$ according to

$$\mathcal{M}_r^{m\times n} = \bigcup_{i=0}^{r} \mathcal{N}_i^{m\times n}.$$

By [12, Ex. 5.30], $\mathcal{N}_i^{m\times n}$ is an embedded submanifold of $\mathbb{R}^{m\times n}$ of dimension $(m+n-i)i$, $i = 1, \ldots, r$. Let $\mathcal{I} = \big\{i \in \{1, \ldots, r\} \mid \mathcal{S} \cap \mathcal{N}_i^{m\times n} \neq \emptyset\big\}$. Then, for each $i \in \mathcal{I}$, $\mathcal{S} \cap \mathcal{N}_i^{m\times n}$ is a nonempty bounded set and, therefore, $\overline{\dim}_{\mathrm{B}}(\mathcal{S} \cap \mathcal{N}_i^{m\times n})$ is well-defined. By [13, Sec. 3.2, Properties (i) and (ii)], $\overline{\dim}_{\mathrm{B}}(\mathcal{S} \cap \mathcal{N}_i^{m\times n}) \leq (m+n-i)i$, $i \in \mathcal{I}$. Since the upper Minkowski dimension is finitely stable [13, Sec. 3.2, Property (iii)], we get

$$\begin{aligned}\overline{\dim}_{\mathrm{B}}(\mathcal{S}) &= \max_{i\in\mathcal{I}} \overline{\dim}_{\mathrm{B}}(\mathcal{S} \cap \mathcal{N}_i^{m\times n}) \\ &\leq (m+n-r)r\end{aligned}$$

where in the last step we used the monotonicity of $f(s) = (m+n-s)s$ in the range $s \in [0, (m+n)/2]$ together with $r \leq (m+n)/2$, which in turn follows from $r \leq \min(m, n)$. ☐

We can now put the pieces together to get the desired statement on low-rank matrices.

Remark 2. Lemma 1 together with $\underline{\dim}_{\mathrm{B}}(\cdot) \leq \overline{\dim}_{\mathrm{B}}(\cdot)$, when used in Theorem 1, implies that for $\mathbf{X} \in \mathcal{M}_r^{m\times n}$ and every $\varepsilon > 0$, there exists a decoder that achieves error probability ε for Lebesgue a.a. measurement matrices A_i, $i = 1,\dots,k$, provided that $k > (m+n-r)r$.

While the sufficient condition $k > (m+n-r)r$ in Remark 2 is intuitively appealing as $(m+n-r)r$ is the dimension of the manifold $\mathcal{N}_r^{m\times n}$, it is actually the lower Minkowski dimensions of ε-support sets of $\mathbf{X} \in \mathcal{M}_r^{m\times n}$ that are of operational significance. Specifically, depending on the distribution of $\mathbf{X}$, a smaller (than $(m+n-r)r$) number of measurements may suffice for recovery of $\mathbf{X}$ with probability of error at most ε. The following example illuminates this phenomenon.

Example 1. Let $\mathbf{X} = \mathbf{X}_1^\mathsf{T}\mathbf{X}_2 \in \mathcal{M}_r^{m\times n}$, where $\mathbf{X}_1 \in \mathbb{R}^{r\times m}$ and $\mathbf{X}_2 \in \mathbb{R}^{r\times n}$ are independent. Suppose that $\mathbf{X}_1$ has l_1 columns at positions drawn uniformly at random and containing i.i.d. Gaussian entries with all other columns equal to zero and $\mathbf{X}_2$ has l_2 columns at positions drawn uniformly at random and containing i.i.d. Gaussian entries with all other columns equal to zero. Suppose further that $r \leq l_1 < m/2$ and $r \leq l_2 \leq n/2 - 1/r$. The assumptions $l_i \geq r$, $i = 1,2$, guarantee that $\mathrm{P}[\operatorname{rank}(\mathbf{X}) = r] = 1$. Next, we construct an ε-support set $\mathcal{T}$ for $\mathbf{X}$ with $\dim_{\mathrm{B}}(\mathcal{T}) \leq (l_1 + l_2)r$, which by Theorem 1 together with $(l_1+l_2)r < (m+n)r/2 - 1 \leq (m+n-r)r - 1$ proves that we can recover the rank-r matrix $\mathbf{X}$ with probability of error at most ε from strictly less than $(m+n-r)r$ measurements.

Let $\mathcal{A}_l^{r\times m} \subseteq \mathcal{M}_r^{r\times m}$ be the set of $r \times m$ matrices with no more than l nonzero columns. Choose $L \in \mathbb{N}$ sufficiently large for (i) $\mathcal{S}_1 = \mathcal{A}_{l_1}^{r\times m} \cap \mathcal{B}_{r\times m}(0, L)$ to be an $\varepsilon/2$-support set of $\mathbf{X}_1$ and (ii) $\mathcal{S}_2 = \mathcal{A}_{l_2}^{r\times n} \cap \mathcal{B}_{r\times n}(0, L)$ to be an $\varepsilon/2$-support set of $\mathbf{X}_2$. By [13, Sec. 3.2, Properties (i) and (iii)], we have

$$\dim_{\mathrm{B}}(\mathcal{S}_i) = l_i r \tag{2}$$

which is simply the maximum number of nonzero entries of $\mathrm{X}_i \in \mathcal{S}_i$, $i = 1, 2$. Set $\mathcal{T} = \{\mathrm{X}_1^\mathsf{T}\mathrm{X}_2 \mid \mathrm{X}_i \in \mathcal{S}_i, i = 1,2\}$. Then,

$$\begin{aligned}\mathrm{P}[\mathbf{X} \in \mathcal{T}] &= \mathrm{P}[\mathbf{X}_1 \in \mathcal{S}_1, \mathbf{X}_2 \in \mathcal{S}_2] \\ &= \mathrm{P}[\mathbf{X}_1 \in \mathcal{S}_1]\,\mathrm{P}[\mathbf{X}_2 \in \mathcal{S}_2] \\ &\geq 1 - \varepsilon.\end{aligned}$$

The triangle inequality implies that for all $\mathrm{X}_i, \bar{\mathrm{X}}_i \in \mathcal{S}_i$, $i = 1, 2$, we have

$$\begin{aligned}&\|\mathrm{X}_1^\mathsf{T}\mathrm{X}_2 - \bar{\mathrm{X}}_1^\mathsf{T}\bar{\mathrm{X}}_2\|_2 \\ &\leq \|\mathrm{X}_1^\mathsf{T}\mathrm{X}_2 - \bar{\mathrm{X}}_1^\mathsf{T}\mathrm{X}_2\|_2 + \|\bar{\mathrm{X}}_1^\mathsf{T}\mathrm{X}_2 - \bar{\mathrm{X}}_1^\mathsf{T}\bar{\mathrm{X}}_2\|_2 \\ &\leq L(\|\mathrm{X}_1 - \bar{\mathrm{X}}_1\|_2 + \|\mathrm{X}_2 - \bar{\mathrm{X}}_2\|_2)\end{aligned} \tag{3}$$

where we used $\mathcal{S}_1 \subseteq \mathcal{B}_{r\times m}(0, L)$ and $\mathcal{S}_2 \subseteq \mathcal{B}_{r\times n}(0, L)$. Let $N_{\mathcal{S}_i}(\rho)$ be the covering number of $\mathcal{S}_i$, $i = 1, 2$. We can cover $\mathcal{S}_i$ by $N_{\mathcal{S}_i}(\rho)$ balls of radius ρ with centers $\bar{\mathrm{X}}_{j_i}$, $j_i = 1,\dots,N_{\mathcal{S}_i}(\rho)$, $i = 1, 2$. Therefore, (3) implies that $\mathcal{T}$ can be covered by $N_{\mathcal{S}_1}(\rho)N_{\mathcal{S}_2}(\rho)$ balls of radius $2L\rho$ centered at $\bar{\mathrm{X}}_{j_1}^\mathsf{T}\bar{\mathrm{X}}_{j_2}$, $j_i = 1,\dots,N_{\mathcal{S}_i}(\rho)$, $i = 1, 2$. This yields $N_{\mathcal{T}}(2L\rho) \leq N_{\mathcal{S}_1}(\rho)N_{\mathcal{S}_2}(\rho)$ and we finally get

$$\begin{aligned}\dim_{\mathrm{B}}(\mathcal{T}) &= \lim_{\rho\to 0} \frac{\log N_{\mathcal{T}}(2L\rho)}{\log \frac{1}{2L\rho}} \\ &\leq \lim_{\rho\to 0} \frac{\log\big(N_{\mathcal{S}_1}(\rho)N_{\mathcal{S}_2}(\rho)\big)}{\log \frac{1}{2L\rho}} \\ &= \lim_{\rho\to 0} \frac{\log N_{\mathcal{S}_1}(\rho)}{\log \frac{1}{2L\rho}} + \lim_{\rho\to 0} \frac{\log N_{\mathcal{S}_2}(\rho)}{\log \frac{1}{2L\rho}} \\ &= l_1 r + l_2 r\end{aligned}$$

where we used (2) in the last step.

Remark 3. The derivation of the recovery thresholds in [9] is also based on a null space property similar to the one discussed in Remark 1. The relevant dimension in [9] is the dimension $(m+n-r)r$ of the manifold $\mathcal{N}_r^{m\times n}$. Example 1 above, however, shows that $k < (m+n-r)r$ measurements can suffice for recovery of rank-r matrices, thereby corroborating the operational significance of the lower Minkowski dimensions of ε-support sets of $\mathbf{X}$.

IV. Rank-one measurement matrices

Rank-one measurement matrices, i.e., matrices $\mathrm{A}_i = \mathrm{a}_i\mathrm{b}_i^\mathsf{T}$ with $\mathrm{a}_i \in \mathbb{R}^m$ and $\mathrm{b}_i \in \mathbb{R}^n$, $i = 1,\dots,k$, are attractive as they require less storage space than general measurement matrices and can also be applied faster. Interestingly, Theorem 1 continues to hold for rank-one measurement matrices although they exhibit much less richness than general measurement matrices. The technical challenges in establishing this result are quite different from those encountered in the case of general measurement matrices. In particular, we will need a stronger concentration of measure inequality (cf. Lemma 4).

Theorem 2. *Let $\mathcal{S} \subseteq \mathbb{R}^{m\times n}$ be an ε-support set of $\mathbf{X} \in \mathbb{R}^{m\times n}$. Then, for Lebesgue a.a. $\mathrm{a}_i \in \mathbb{R}^m$ and $\mathrm{b}_i \in \mathbb{R}^n$ and corresponding measurement matrices $\mathrm{A}_i = \mathrm{a}_i\mathrm{b}_i^\mathsf{T}$, $i = 1,\dots,k$, there exists a decoder achieving error probability ε, provided that $k > \underline{\dim}_{\mathrm{B}}(\mathcal{S})$.*

Proof. See Section V. □

Remark 4. Example 1 can be shown to carry over to rank-one measurement matrices A_i.

Remark 5. Theorem 2, when used in combination with Lemma 1, implies that for $\mathbf{X} \in \mathcal{M}_r^{m\times n}$ and every $\varepsilon > 0$, there exists a decoder achieving error probability ε for Lebesgue a.a. $\mathrm{a}_i \in \mathbb{R}^m$ and $\mathrm{b}_i \in \mathbb{R}^n$, provided that $k > (m+n-r)r$. In contrast, the threshold $k \geq C(m+n)r$ in [8] for rank-one measurements requires $\mathbf{a}_i$ and $\mathbf{b}_i$ to be independent random vectors containing i.i.d. Gaussian or sub-Gaussian entries. In addition, the constant C in [8] remains unspecified.

V. Proofs of Theorems 1 and 2

For both proofs, we first construct a measurable map $g : \mathbb{R}^k \to \mathbb{R}^{m\times n}$ such that

$$\begin{aligned}&\mathrm{P}\big[\big[g\big((\langle \mathrm{A}_1, \mathbf{X}\rangle,\dots,\langle \mathrm{A}_k, \mathbf{X}\rangle)^\mathsf{T}\big) \neq \mathbf{X}\big] \\ &\leq \mathrm{P}\big[\exists \mathrm{Z} \in \mathcal{S}_{\mathbf{X}}\backslash\{0\} \big| (\langle \mathrm{A}_1, \mathrm{Z}\rangle,\dots,\langle \mathrm{A}_k, \mathrm{Z}\rangle)^\mathsf{T} = 0, \mathbf{X} \in \mathcal{S}\big] + \varepsilon\end{aligned} \tag{4}$$

with $\mathcal{S}_{\mathrm{X}} = \{\mathrm{W} - \mathrm{X} \mid \mathrm{W} \in \mathcal{S}\}$ for $\mathrm{X} \in \mathcal{S}$. The proofs are then concluded by showing that

$$\mathrm{P}\big[\exists \mathrm{Z} \in \mathcal{S}_{\mathbf{X}} \backslash \{0\} \big| (\langle \mathrm{A}_1, \mathrm{Z}\rangle, ..., \langle \mathrm{A}_k, \mathrm{Z}\rangle)^{\mathsf{T}} = 0, \mathbf{X} \in \mathcal{S}\big] = 0 \tag{5}$$

for Lebesgue a.a. matrices $\mathrm{A}_i \in \mathbb{R}^{m \times n}$, $i = 1,..., k$, in the case of Theorem 1 and for Lebesgue a.a. vectors $\mathrm{a}_i \in \mathbb{R}^m$ and $\mathrm{b}_i \in \mathbb{R}^n$ with $\mathrm{A}_i = \mathrm{a}_i \mathrm{b}_i^{\mathsf{T}}$, $i = 1,..., k$, in the case of Theorem 2.

Proof of (4): Let $\mathcal{S} \subseteq \mathbb{R}^{m \times n}$ be an ε-support set of $\mathbf{X}$ with $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < k$. We define a measurable map g as follows:

$$g(\mathrm{y}) = \begin{cases} \mathrm{Z}, & \text{if } \big\{\mathrm{W} \in \mathcal{S} \mid (\langle \mathrm{A}_1, \mathrm{W}\rangle, ..., \langle \mathrm{A}_k, \mathrm{W}\rangle)^{\mathsf{T}} = \mathrm{y}\big\} = \{\mathrm{Z}\} \\ \mathrm{E}, & \text{else} \end{cases}$$

where E is an arbitrary, but fixed, matrix in $\mathbb{R}^{m \times n} \setminus \mathcal{S}$ (used to declare a decoding error). Then, we have

$$\begin{aligned} &\mathrm{P}\big[g\big((\langle \mathrm{A}_1, \mathbf{X}\rangle, ..., \langle \mathrm{A}_k, \mathbf{X}\rangle)^{\mathsf{T}}\big) \neq \mathbf{X}\big] \\ &\leq \mathrm{P}\big[g\big((\langle \mathrm{A}_1, \mathbf{X}\rangle, ..., \langle \mathrm{A}_k, \mathbf{X}\rangle)^{\mathsf{T}}\big) \neq \mathbf{X}, \mathbf{X} \in \mathcal{S}\big] + \mathrm{P}[\mathbf{X} \notin \mathcal{S}] \\ &\leq \mathrm{P}\big[g\big((\langle \mathrm{A}_1, \mathbf{X}\rangle, ..., \langle \mathrm{A}_k, \mathbf{X}\rangle)^{\mathsf{T}}\big) \neq \mathbf{X}, \mathbf{X} \in \mathcal{S}\big] + \varepsilon \qquad (6) \\ &= \mathrm{P}\big[g\big((\langle \mathrm{A}_1, \mathbf{X}\rangle, ..., \langle \mathrm{A}_k, \mathbf{X}\rangle)^{\mathsf{T}}\big) = \mathrm{E}, \mathbf{X} \in \mathcal{S}\big] + \varepsilon \qquad (7) \\ &= \mathrm{P}\big[\exists \mathrm{Z} \in \mathcal{S}_{\mathbf{X}} \backslash \{0\} \big| (\langle \mathrm{A}_1, \mathrm{Z}\rangle, ..., \langle \mathrm{A}_k, \mathrm{Z}\rangle)^{\mathsf{T}} = 0, \mathbf{X} \in \mathcal{S}\big] + \varepsilon \end{aligned}$$

where (6) is a consequence of S being an ε-support set and in (7) we used that the decoder declares an error if and only if $|\{\mathrm{W} \in \mathcal{S} \mid (\langle \mathrm{A}_1, \mathrm{W}\rangle, ..., \langle \mathrm{A}_k, \mathrm{W}\rangle)^{\mathsf{T}} = \mathrm{y}\}| > 1$ for $\mathrm{y} = (\langle \mathrm{A}_1, \mathrm{X}\rangle, ..., \langle \mathrm{A}_k, \mathrm{X}\rangle)^{\mathsf{T}}$ with $\mathrm{X} \in \mathcal{S}$.

Finishing the proof of Theorem 1: Let $s > 0$ and suppose that $\mathbf{A}_1, ..., \mathbf{A}_k$, $i = 1, ..., k$, are independent and uniformly distributed on $\mathcal{B}_{m \times n}(0, s)$. Then, we have

$$\begin{aligned} &\int\limits_{(\mathcal{B}_{m\times n}(0,s))^k} \mathrm{P}\big[\exists \mathrm{Z} \in \mathcal{S}_{\mathbf{X}} \backslash \{0\} \mid (\langle \mathrm{A}_1, \mathrm{Z}\rangle, ..., \langle \mathrm{A}_k, \mathrm{Z}\rangle)^{\mathsf{T}} = 0, \mathbf{X} \in \mathcal{S}\big] \\ &\qquad \mathrm{d}\mu_{\mathbf{A}_1} \times ... \times \mathrm{d}\mu_{\mathbf{A}_k} \\ &= \int\limits_{\mathcal{S}} \mathrm{P}\big[\exists \mathrm{Z} \in \mathcal{S}_{\mathrm{X}} \backslash \{0\} \mid (\langle \mathbf{A}_1, \mathrm{Z}\rangle, ..., \langle \mathbf{A}_k, \mathrm{Z}\rangle)^{\mathsf{T}} = 0\big] \, \mathrm{d}\mu_{\mathbf{X}} \qquad (8) \\ &= 0 \qquad (9) \end{aligned}$$

where (8) is a consequence of Fubini's theorem for non-negative measurable functions and (9) follows from Lemma 2 below. With $\mathbb{R}^{m \times n} = \bigcup_{s \in \mathbb{N}} \mathcal{B}_{m \times n}(0, s)$ and since s is arbitrary, (5) holds for Lebesgue a.a. measurement matrices A_i, which concludes the proof of Theorem 1.

Finishing the proof of Theorem 2: Let $s > 0$ and suppose that $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_k] \in \mathbb{R}^{m \times k}$ and $\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_k] \in \mathbb{R}^{n \times k}$ are independent random matrices with columns $\mathbf{a}_i$ independent and uniformly distributed on $\mathcal{B}_m(0, s)$ and columns $\mathbf{b}_i$ independent and uniformly distributed on $\mathcal{B}_n(0, s)$. Then, we have

$$\begin{aligned} &\int\limits_{(\mathcal{B}_m(0,s)\times\mathcal{B}_n(0,s))^k} \mathrm{P}\big[\exists \mathrm{Z} \in \mathcal{S}_{\mathbf{X}} \backslash \{0\} \mid (\mathrm{a}_1^{\mathsf{T}} \mathrm{Z} \mathrm{b}_1, ..., \mathrm{a}_k^{\mathsf{T}} \mathrm{Z} \mathrm{b}_k)^{\mathsf{T}} = 0, \mathbf{X} \in \mathcal{S}\big] \\ &\qquad \mathrm{d}\mu_{\mathbf{a}_1} \times \mathrm{d}\mu_{\mathbf{b}_1} \times ... \times \mathrm{d}\mu_{\mathbf{a}_k} \times \mathrm{d}\mu_{\mathbf{b}_k} \\ &= \int\limits_{\mathcal{S}} \mathrm{P}\big[\exists \mathrm{Z} \in \mathcal{S}_{\mathrm{X}} \backslash \{0\} \mid (\mathbf{a}_1^{\mathsf{T}} \mathrm{Z} \mathbf{b}_1, ..., \mathbf{a}_k^{\mathsf{T}} \mathrm{Z} \mathbf{b}_k)^{\mathsf{T}} = 0\big] \, \mathrm{d}\mu_{\mathbf{X}} \qquad (10) \\ &= 0 \qquad (11) \end{aligned}$$

where (10) is a consequence of Fubini's theorem for nonnegative measurable functions and (11) follows from Lemma 3 below. Again, with $\mathbb{R}^l = \bigcup_{s \in \mathbb{N}} \mathcal{B}_l(0, s)$ and since s is arbitrary, (5) holds for Lebesgue a.a. vectors $\mathrm{a}_i \in \mathbb{R}^m$ and $\mathrm{b}_i \in \mathbb{R}^n$, thereby finishing the proof of Theorem 2. □

Lemma 2. *Let $s > 0$ and $\mathbf{A}_1, ..., \mathbf{A}_k$, $i = 1, ..., k$, be independent and uniformly distributed on $\mathcal{B}_{m \times n}(0, s)$. Suppose that $\mathcal{U} \subseteq \mathbb{R}^{m \times n}$ is a nonempty bounded set with $\underline{\dim}_{\mathrm{B}}(\mathcal{U}) < k$. Then, we have*

$$\mathrm{P}\big[\exists \mathrm{X} \in \mathcal{U} \backslash \{0\} \mid (\langle \mathbf{A}_1, \mathrm{X}\rangle, ..., \langle \mathbf{A}_k, \mathrm{X}\rangle)^{\mathsf{T}} = 0\big] = 0. \tag{12}$$

Proof. Follows from rewriting the trace inner products $\langle \mathbf{A}_i, \mathrm{X}\rangle$, $i = 1, ..., k$ as inner products between vectors in $\mathbb{R}^{mn}$ and subsequent application of [11, Prop. 1]. □

Lemma 3. *Let $s > 0$ and take $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_k] \in \mathbb{R}^{m \times k}$ and $\mathbf{B} = [\mathbf{b}_1, ..., \mathbf{b}_k] \in \mathbb{R}^{n \times k}$ to be independent random matrices with columns $\mathbf{a}_i$, $i = 1, ..., k$, independent and uniformly distributed on $\mathcal{B}_m(0, s)$ and columns $\mathbf{b}_i$, $i = 1, ..., k$, independent and uniformly distributed on $\mathcal{B}_n(0, s)$. Suppose that $\mathcal{U} \subseteq \mathbb{R}^{m \times n}$ is a nonempty bounded set with $\underline{\dim}_{\mathrm{B}}(\mathcal{U}) < k$. Then, we have*

$$P := \mathrm{P}\big[\exists \mathrm{X} \in \mathcal{U} \backslash \{0\} \big| (\mathbf{a}_1^{\mathsf{T}} \mathrm{X} \mathbf{b}_1, ..., \mathbf{a}_k^{\mathsf{T}} \mathrm{X} \mathbf{b}_k)^{\mathsf{T}} = 0\big] = 0.$$

Proof. Let $R = \max_{\mathrm{X} \in \mathcal{U}} \operatorname{rank}(\mathrm{X})$ and set

$$\mathcal{U}_{L,r} = \Big\{\mathrm{X} \in \mathcal{U} \mid \Delta(\mathrm{X}) > \frac{1}{L}, \sigma_1(\mathrm{X}) < L, \operatorname{rank}(\mathrm{X}) = r\Big\}$$

for $L \in \mathbb{N}$ and $r = 1, ..., R$. By the union bound, we have

$$P \leq \sum_{L \in \mathbb{N}} \sum_{r=1}^{R} P_{L,r} \tag{13}$$

where

$$P_{L,r} = \mathrm{P}\big[\exists \mathrm{X} \in \mathcal{U}_{L,r} \mid (\mathbf{a}_1^{\mathsf{T}} \mathrm{X} \mathbf{b}_1, ..., \mathbf{a}_k^{\mathsf{T}} \mathrm{X} \mathbf{b}_k)^{\mathsf{T}} = 0\big].$$

We now prove by contradiction that $P_{L,r} = 0$ for all $L \in \mathbb{N}$ and all $r \in \{1, ..., R\}$. Suppose that there exists an $L \in \mathbb{N}$ and an $r \in \{1, ..., R\}$ such that $P_{L,r} > 0$ (by definition, $P_{L,r} \geq 0$). For this pair $\{L, r\}$, we would then have

$$\liminf_{\rho \to 0} \frac{\log P_{L,r}}{\log \frac{1}{\rho}} = 0. \tag{14}$$

For $\rho > 0$, let $N_{\mathcal{U}_{L,r}}(\rho)$ be the covering number of the set $\mathcal{U}_{L,r}$ and denote corresponding covering balls centered at $\mathrm{M}_i(\rho) \in \mathbb{R}^{m \times n}$ as $\mathcal{B}_{m \times n}(\mathrm{M}_i(\rho), \rho)$, $i = 1, ..., N_{\mathcal{U}_{L,r}}(\rho)$. We now fix $N_{\mathcal{U}_{L,r}}(\rho)$ matrices

$$\mathrm{X}_i(\rho) \in \mathcal{B}_{m \times n}(\mathrm{M}_i(\rho), \rho) \cap \mathcal{U}_{L,r}, \quad i = 1, ..., N_{\mathcal{U}_{L,r}}(\rho). \tag{15}$$

Since

$$\mathcal{B}_{m \times n}(\mathrm{M}_i(\rho), \rho) \subseteq \mathcal{B}_{m \times n}(\mathrm{X}_i(\rho), 2\rho), \quad i = 1, ..., N_{\mathcal{U}_{L,r}}(\rho)$$

by the triangle inequality, we get

$$P_{L,r} \leq \sum_{i=1}^{N_{\mathcal{U}_{L,r}}(\rho)} \mathrm{P}\big[\exists \mathrm{X} \in \mathcal{B}_{m\times n}(\mathrm{M}_i(\rho),\rho) \mid (\mathbf{a}_1^\mathsf{T}\mathrm{X}\mathbf{b}_1,\dots,\mathbf{a}_k^\mathsf{T}\mathrm{X}\mathbf{b}_k)^\mathsf{T} = 0\big]$$
$$\leq \sum_{i=1}^{N_{\mathcal{U}_{L,r}}(\rho)} \mathrm{P}\big[\exists \mathrm{X} \in \mathcal{B}_{m\times n}(\mathrm{X}_i(\rho),2\rho) \mid (\mathbf{a}_1^\mathsf{T}\mathrm{X}\mathbf{b}_1,\dots,\mathbf{a}_k^\mathsf{T}\mathrm{X}\mathbf{b}_k)^\mathsf{T} = 0\big]$$
$$\leq \sum_{i=1}^{N_{\mathcal{U}_{L,r}}(\rho)} \mathrm{P}\big[\exists \mathrm{X} \in \mathcal{B}_{m\times n}(\mathrm{X}_i(\rho),2\rho) \mid \|(\mathbf{a}_1^\mathsf{T}\mathrm{X}\mathbf{b}_1,\dots,\mathbf{a}_k^\mathsf{T}\mathrm{X}\mathbf{b}_k)^\mathsf{T}\|_2 \leq \rho\big], \quad \rho > 0. \tag{16}$$

Now, for $\mathrm{a}_i \in \mathcal{B}_m(0,s)$, $\mathrm{b}_i \in \mathcal{B}_n(0,s)$, and $\mathrm{X} \in \mathcal{B}_{m\times n}(\mathrm{X}_i(\rho),2\rho)$, we have

$$\begin{aligned}
&\|(\mathrm{a}_1^\mathsf{T}\mathrm{X}_i(\rho)\mathrm{b}_1,\dots,\mathrm{a}_k^\mathsf{T}\mathrm{X}_i(\rho)\mathrm{b}_k)^\mathsf{T}\|_2\\
&\leq \|(\mathrm{a}_1^\mathsf{T}(\mathrm{X}-\mathrm{X}_i(\rho))\mathrm{b}_1,\dots,\mathrm{a}_k^\mathsf{T}(\mathrm{X}-\mathrm{X}_i(\rho))\mathrm{b}_k)^\mathsf{T}\|_2\\
&\quad + \|(\mathrm{a}_1^\mathsf{T}\mathrm{X}\mathrm{b}_1,\dots,\mathrm{a}_k^\mathsf{T}\mathrm{X}\mathrm{b}_k)^\mathsf{T}\|_2\\
&\leq \sqrt{\sum_{j=1}^{k}\|\mathrm{a}_j\|_2^2\|\mathrm{X}-\mathrm{X}_i(\rho)\|_2^2\|\mathrm{b}_j\|_2^2}\\
&\quad + \|(\mathrm{a}_1^\mathsf{T}\mathrm{X}\mathrm{b}_1,\dots,\mathrm{a}_k^\mathsf{T}\mathrm{X}\mathrm{b}_k)^\mathsf{T}\|_2\\
&\leq 2s^2\sqrt{k}\rho + \|(\mathrm{a}_1^\mathsf{T}\mathrm{X}\mathrm{b}_1,\dots,\mathrm{a}_k^\mathsf{T}\mathrm{X}\mathrm{b}_k)^\mathsf{T}\|_2, \quad \rho > 0.
\end{aligned} \tag{17}$$

Inserting (17) into (16) allows us to further upper-bound $P_{L,r}$ according to

$$\begin{aligned}
P_{L,r} &\leq \sum_{i=1}^{N_{\mathcal{U}_{L,r}}(\rho)} \mathrm{P}\big[\|(\mathbf{a}_1^\mathsf{T}\mathrm{X}_i(\rho)\mathbf{b}_1,\dots,\mathbf{a}_k^\mathsf{T}\mathrm{X}_i(\rho)\mathbf{b}_k)^\mathsf{T}\|_2 \leq \rho(1+2s^2\sqrt{k})\big]\\
&\leq N_{\mathcal{U}_{L,r}}(\rho)\rho^k g(L,r,k,m,n,s,\rho)^k, \quad \rho > 0
\end{aligned} \tag{18}$$

where

$$g(L,r,k,m,n,s,\rho) = \frac{2(2s)^{m+n-r-1}(1+2s^2\sqrt{k})L}{V(m,s)V(n,s)} \times \begin{cases} 2 + 2\log\max\left(\frac{s^2L}{\rho},1\right), & \text{if } r = 1\\ 2^r\left(\frac{\rho(1+2s^2\sqrt{k})}{s}\right)^{r-1} + \frac{A(r-1,1)(sL)^{r-1}}{r-1}, & \text{if } r > 1.\end{cases}$$

Here, we applied the concentration of measure inequality in Lemma 4 below with $\delta = \rho(1+2s^2\sqrt{k})$ and used the fact that $1/\Delta(\mathrm{X}_i(\rho)) < L$, $\sigma_1(\mathrm{X}_i(\rho)) < L$, and $\mathrm{rank}(X_i(\rho)) = r$ (recall that by (15) all matrices $\mathrm{X}_i(\rho)$ are in the set $\mathcal{U}_{L,r}$). With the upper bound on $P_{L,r}$ in (18) we now get

$$\begin{aligned}
&\liminf_{\rho\to 0}\frac{\log P_{L,r}}{\log\frac{1}{\rho}}\\
&\leq \liminf_{\rho\to 0}\frac{\log(N_{\mathcal{U}_{L,r}}(\rho)) + k\log\rho + k\log g(L,r,k,m,n,s,\rho)}{\log\frac{1}{\rho}}\\
&= \liminf_{\rho\to 0}\frac{\log(N_{\mathcal{U}_{L,r}}(\rho))}{\log\frac{1}{\rho}} - k + \lim_{\rho\to 0}\frac{k\log g(L,r,k,m,n,s,\rho)}{\log\frac{1}{\rho}}\\
&= \liminf_{\rho\to 0}\frac{\log(N_{\mathcal{U}_{L,r}}(\rho))}{\log\frac{1}{\rho}} - k\\
&\leq \liminf_{\rho\to 0}\frac{\log(N_{\mathcal{U}}(\rho))}{\log\frac{1}{\rho}} - k && (19)\\
&= \underline{\dim}_{\mathrm{B}}(\mathcal{U}) - k\\
&< 0 && (20)
\end{aligned}$$

where (19) follows from $\mathcal{U}_{L,r} \subseteq \mathcal{U}$ and in the last step we used that $\underline{\dim}_{\mathrm{B}}(\mathcal{U}) < k$, by assumption. Since (20) contradicts (14), $P_{L,r} = 0$ for all $L \in \mathbb{N}$ and all $r \in \{1,\dots,R\}$. By (13), this establishes that $P = 0$. □

Lemma 4. *Let* $\mathbf{A} = [\mathbf{a}_1,\dots,\mathbf{a}_k] \in \mathbb{R}^{m\times k}$ *and* $\mathbf{B} = [\mathbf{b}_1,\dots,\mathbf{b}_k] \in \mathbb{R}^{n\times k}$ *be independent random matrices, with columns* $\mathbf{a}_i$, $i = 1,\dots,k$, *independent and uniformly distributed on* $\mathcal{B}_m(0,s)$ *and columns* $\mathbf{b}_i$, $i = 1,\dots,k$, *independent and uniformly distributed on* $\mathcal{B}_n(0,s)$. *Suppose that* $\mathrm{X} \in \mathbb{R}^{m\times n}$ *with* $r = \mathrm{rank}(\mathrm{X}) > 0$. *Then, we have*

$$\mathrm{P}\big[\big\|(\mathbf{a}_1^\mathsf{T}\mathrm{X}\mathbf{b}_1,\dots,\mathbf{a}_k^\mathsf{T}\mathrm{X}\mathbf{b}_k)^\mathsf{T}\big\|_2 \leq \delta\big] \leq \delta^k f(\mathrm{X},s,\delta)^k$$

with

$$f(\mathrm{X},s,\delta) = \frac{2(2s)^{m+n-r-1}}{\Delta(\mathrm{X})V(m,s)V(n,s)} \times \begin{cases} 2 + 2\log\max\left(\frac{s^2\sigma_1(\mathrm{X})}{\delta},1\right), & \text{if } r = 1\\ 2^r\left(\frac{\delta}{s}\right)^{r-1} + \frac{A(r-1,1)(s\sigma_1(\mathrm{X}))^{r-1}}{r-1}, & \text{if } r > 1.\end{cases} \tag{21}$$

Proof. Omitted due to space limitations. □

References

[1] 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*, vol. 4, no. 57, pp. 2342–2359, Apr. 2011.

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

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

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

[5] B. Recht, M. Fazel, and P. A. Parrilo, "Guaranteed minimum-rank solutions of linear matrix equations via nuclear norm minimization," *SIAM Review*, vol. 52, no. 3, pp. 471–501, Dec. 2010.

[6] B. Recht, "A simpler approach to matrix completion," *J. Mach. Learn. Res.*, vol. 12, pp. 341–3430, 2011.

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

[8] ——, "ROP: Matrix recovery via rank-one projections," *Ann. Stat.*, vol. 43, no. 1, pp. 102–138, 2015.

[9] Y. C. Eldar, D. Needell, and Y. Plan, "Uniqueness conditions for low-rank matrix recovery," *Appl. Comp. Harm. Anal.*, vol. 33, no. 2, pp. 309–314, 2012.

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

[11] D. Stotz, E. Riegler, and H. Bölcskei, "Almost lossless analog signal separation," in *Proc. IEEE Int. Symp. Inf. Theory*, Istanbul, Turkey, Jul. 2013, pp. 106–110.

[12] J. M. Lee, *Introduction to Smooth Manifolds*, 1st ed. New York, NY: Springer, 2000.

[13] K. Falconer, *Fractal Geometry*, 1st ed. New York, NY: Wiley, 1990.