Almost Lossless Analog Signal Separation and Probabilistic Uncertainty Relations

David Stotz†, Erwin Riegler⋆, Eirikur Agustsson⋆, and Helmut Bölcskei⋆
†Kantonsschule Schaffhausen, 8200 Schaffhausen, Switzerland, Email: david-stotz@web.de
⋆Dept. of IT & EE, ETH Zurich, 8092 Zurich, Switzerland, Email: {eriegler, boelcskei}@nari.ee.ethz.ch,
aeirikur@vision.ee.ethz.ch

***Abstract*—We propose an information-theoretic framework for analog signal separation. Specifically, we consider the problem of recovering two analog signals, modeled as general random vectors, from the noiseless sum of linear measurements of the signals. Our framework is inspired by the groundbreaking work of Wu and Verdú (2010) on analog compression and encompasses, inter alia, inpainting, declipping, super-resolution, the recovery of signals corrupted by impulse noise, and the separation of (e.g., audio or video) signals into two distinct components. The main results we report are general achievability bounds for the compression rate, i.e., the number of measurements relative to the dimension of the ambient space the signals live in, under either measurability or Hölder continuity imposed on the separator. Furthermore, we find a matching converse for sources of mixed discrete-continuous distribution. For measurable separators our proofs are based on a new probabilistic uncertainty relation which shows that the intersection of generic subspaces with general sets of sufficiently small Minkowski dimension is empty. Hölder continuous separators are dealt with by introducing the concept of regularized probabilistic uncertainty relations. The probabilistic uncertainty relations we develop are inspired by embedding results in dynamical systems theory due to Sauer et al. (1991) and—conceptually—parallel classical Donoho-Stark and Elad-Bruckstein uncertainty principles at the heart of compressed sensing theory. Operationally, the new uncertainty relations take the theory of sparse signal separation beyond traditional sparsity—as measured in terms of the number of non-zero entries—to the more general notion of low description complexity as quantified by Minkowski dimension. Finally, our approach also allows to significantly strengthen key results in Wu and Verdú (2010).**

***Index Terms*—Signal separation, compressed sensing, uncertainty relations, Minkowski dimension, Shannon theory.**

I. Introduction

We consider the following signal separation problem: Recover the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ from the noiseless observation

$$\boldsymbol{w} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{B}\boldsymbol{z}, \tag{1}$$

where $\boldsymbol{A}$ and $\boldsymbol{B}$ are measurement matrices. Numerous signal processing problems can be cast in the form (1), e.g., inpainting, declipping, super-resolution, the recovery of signals corrupted by impulse noise, and the separation of (e.g., audio or video) signals into two distinct components. For a detailed exposition of the specifics of (1) for each of these applications and for corresponding references, we refer the interested reader to [3, Sec. 1].

The sparse signal recovery literature [3]–[12] provides separation guarantees under sparsity constraints on the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$. Specifically, the sparsity thresholds in, e.g., [3], [7], [11], [12] are functions of the coherence parameters [3] of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and hold for *all* $\boldsymbol{y}$ and $\boldsymbol{z}$, but suffer from the "square-root bottleneck" [8], which states that the number of measurements, i.e., the number of entries of $\boldsymbol{w}$, has to scale quadratically in the total number of non-zero entries in $\boldsymbol{y}$ and $\boldsymbol{z}$. For random signals $\boldsymbol{y}$ and $\boldsymbol{z}$, the probabilistic results in [4], [5], [10] overcome the square-root bottleneck, but hold "only" with overwhelming probability. For $\boldsymbol{B}$ equal to the identity matrix and $\boldsymbol{A}$ a random orthogonal matrix it is shown in [9] that the probability of failure of an ℓ_1-based separation algorithm decays exponentially in the dimension of the ambient space, provided that $\boldsymbol{y}$ and $\boldsymbol{z}$ satisfy certain convex cone conditions.

Contributions. The goal of this paper is to develop an information-theoretic framework for signal separation. Specifically, inspired by the groundbreaking work of Wu and Verdú on analog compression [13], we consider the problem of recovering y and z from $\mathsf{w} = \boldsymbol{A}\mathsf{y} + \boldsymbol{B}\mathsf{z}$, where y and z are random, possibly dependent, and of general distributions, i.e., mixtures of discrete, absolutely continuous, and singular distributions. Our results are asymptotic in the sense that the probability of error can be made arbitrarily small by increasing the dimensions of w, y, and z. In practical signal separation problems of this form one often encounters a specific structure for one of the matrices (assumed to be $\boldsymbol{B}$ here, without loss of generality (w.l.o.g.)); for example, the matrix could represent a certain dictionary under which a class of signals is sparse. We will therefore be interested in statements that hold for a given $\boldsymbol{B}$. Our separation guarantees will, indeed, be seen to apply to deterministic $\boldsymbol{B}$ and a.a.[1] $\boldsymbol{A}$ (with the set of exceptions for $\boldsymbol{A}$ depending on the specific choice of $\boldsymbol{B}$). Moreover, they do not provide worst-case guarantees like the coherence-based results in, e.g., [3], [7], [11], [12], but are rather in terms of probability of separation error with respect to (w.r.t.) the constituents y and z, and as such do not depend on coherence parameters. Specifically, we study the asymptotic setting $\ell, n \to \infty$ where the random vectors $\mathsf{y} \in \mathbb{R}^{n-\ell}$ and $\mathsf{z} \in \mathbb{R}^{\ell}$ are sections of random processes; for each n, we let $\ell = \lfloor \lambda n \rfloor$ and $k = \lfloor Rn \rfloor$

The material in this paper was presented in part at the IEEE Int. Symp. on Inf. Theory, Istanbul, Turkey, Jul. 2013 [1]. An Online Addendum [2] to this paper is available at http://www.nari.ee.ethz.ch/commth/research/downloads/sigsep_addendum.pdf.

[1]Throughout the paper a.a. stands for "Lebesgue almost all".

for parameters $\lambda, R \in [0, 1]$. We refer to R as the compression rate as it equals (approximately) the ratio between the number of measurements, k, and the total number of entries, n, in y and z. Our first main result, Theorem 1, shows that for each (deterministic) full-rank matrix $\boldsymbol{B} \in \mathbb{R}^{k\times\ell}$, with $k \geqslant \ell$, and a.a. matrices $\boldsymbol{A} \in \mathbb{R}^{k\times(n-\ell)}$, there exists a measurable[2] separator recovering y and z from w with arbitrarily small probability of error, provided that n is sufficiently large and the compression rate R is larger than the description complexity of the concatenated random source vector $\mathsf{x} := [\mathsf{y}^{\mathrm{T}} \ \ \mathsf{z}^{\mathrm{T}}]^{\mathrm{T}}$ as quantified by its Minkowski dimension compression rate $R_{\mathrm{B}}(\varepsilon)$ (see Definition 4). In practice, when recovery is to be performed from noisy, quantized, or otherwise perturbed versions of the measurement w, it is desirable to impose continuity/smoothness constraints on the separator. The second main result of this paper, reported in Theorem 2, shows that for each (deterministic) full-rank matrix $\boldsymbol{B} \in \mathbb{R}^{k\times\ell}$, with $k \geqslant \ell$, and a.a. matrices $\boldsymbol{A} \in \mathbb{R}^{k\times(n-\ell)}$, there exists a β-Hölder continuous separator achieving error probability ε provided that $R > R_{\mathrm{B}}(\varepsilon)$ and $\beta < 1 - \frac{R_{\mathrm{B}}(\varepsilon)}{R}$. We hasten to add that we do not specify explicit separators that achieve our thresholds, rather we prove existence results absent computational considerations. In contrast, many of the recovery thresholds available in the literature pertain to ℓ_1-norm-based recovery algorithms, see, e.g., [3]–[5], [7]–[12]. In the case of mixed discrete-continuous source distributions a converse matching the general—w.r.t. the nature of the source distributions—achievability statements in Theorems 1 and 2 can be obtained. This establishes the Minkowski dimension compression rate $R_{\mathrm{B}}(\varepsilon)$ as the critical rate for successful separation when the source distributions are mixed discrete-continuous.

In principle one could rewrite (1) in the form $\boldsymbol{w} = [\boldsymbol{A} \ \boldsymbol{B}][\boldsymbol{y}^{\mathrm{T}} \ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}$ and consider applying the results in [13] with $\boldsymbol{H} = [\boldsymbol{A} \ \boldsymbol{B}]$. However, the theory developed in [13] leads to statements that apply to a.a. matrices $\boldsymbol{H}$, whereas here, for reasons mentioned above, we seek statements that apply for a *given* matrix $\boldsymbol{B}$, and fixing $\boldsymbol{B}$ results in $\boldsymbol{H} = [\boldsymbol{A} \ \boldsymbol{B}]$-matrices supported on a set of Lebesgue measure zero. A direct application of the results in [13] to the signal separation problem is hence not possible; we therefore develop a new proof methodology and new mathematical tools. The foundation of our approach stems from dynamical systems theory [14]. Specifically, we establish a new technique for showing that the intersection of generic subspaces (of finite-dimensional Euclidean spaces) and arbitrary sets of sufficiently small Minkowski dimension is empty. This leads to statements that have the flavor of a probabilistic uncertainty relation akin to the classical (deterministic) Donoho-Stark [12] and Elad-Bruckstein [15] uncertainty relations underlying much of compressed sensing theory. Our result on Hölder continuous separators is based on a regularized probabilistic uncertainty relation, a concept which does not seem to have a counterpart in classical compressed sensing theory. Finally, we note that applying our mathematical machinery to the analog compression framework in [13] leads to a simplification of the proof of [13, Thm. 18, 1)] and to significant strengthening of [13, Thm. 18, 2)], as detailed in Section VIII.

[2]Throughout the paper, the term measurable refers to Borel measurability.

Notation. For a relation $\# \in \{<, >, \leqslant, \geqslant, =, \neq, \in, \notin\}$, we write $f(n) \overset{\centerdot}{\#} g(n)$ if there exists an $N \in \mathbb{N}$ such that $f(n) \# g(n)$ holds for all $n \geqslant N$. Leb^n stands for the n-dimensional Lebesgue measure and $\mathcal{B}^{\otimes n}$ refers to the Borel σ-algebra on $\mathbb{R}^n$. Matrices are denoted by capital boldface and vectors by lowercase boldface letters. We let $\|\cdot\|$ be the ℓ_2-norm on $\mathbb{R}^n$ and set $\|\boldsymbol{A}\| := \sup_{\|\boldsymbol{x}\|=1} \|\boldsymbol{A}\boldsymbol{x}\|$. The $n \times n$ identity matrix is $\boldsymbol{I}_n$ and $\boldsymbol{F}_n$ stands for the n-dimensional discrete Fourier transform (DFT) matrix. Sets are represented by calligraphic letters. For $\mathcal{A}, \mathcal{B} \subseteq \mathbb{R}^n$, we set $\mathcal{A} \ominus \mathcal{B} := \{\boldsymbol{a} - \boldsymbol{b} : \boldsymbol{a} \in \mathcal{A}, \boldsymbol{b} \in \mathcal{B}\}$. $B^n(\boldsymbol{x}, \delta)$ is the open ℓ_2-ball of radius δ centered at $\boldsymbol{x} \in \mathbb{R}^n$, and its volume is given by $\alpha(n, \delta) = \mathrm{Leb}^n(B^n(\boldsymbol{x}, \delta))$. We use sans-serif letters, e.g. x, for random quantities and roman letters, e.g. $\boldsymbol{x}$, for deterministic quantities. For a random variable X and a random vector x, μ_{X} and μ_{x} denote the respective distributions, integration w.r.t. these distributions is indicated by $\mu_{\mathsf{X}}(\mathrm{d}x)$ and $\mu_{\mathsf{x}}(\mathrm{d}\boldsymbol{x})$. For Borel sets $\mathcal{A}$, we let $\mathbb{1}_{\mathcal{A}}(x)$ be the indicator function on $\mathcal{A}$. Constants which depend exclusively on parameters $\alpha_1, ..., \alpha_n$ are written as $c(\alpha_1, ..., \alpha_n)$ or $C(\alpha_1, ..., \alpha_n)$, where the constants may take on different values in different appearances.

Outline of the paper. In Section II, we first introduce our information-theoretic framework for the signal separation problem and then state the achievability result for measurable separators, Theorem 1, followed by the achievability result for Hölder continuous separators, Theorem 2. Section III contains the probabilistic uncertainty relation the proof of Theorem 1 is based on. In Section IV, we present the proof of Theorem 1. Section V introduces the regularized probabilistic uncertainty relation underlying the proof of Theorem 2, which is provided in Section VI. In Section VII, we particularize our results for mixed discrete-continuous source distributions and we derive a converse matching the corresponding achievability results. Finally, in Section VIII we show how our mathematical techniques lead to a simplification of the proof of [13, Thm. 18, 1)] and to significant strengthening of the statement [13, Thm. 18, 2)].

II. Statement of the main results

We begin by introducing our information-theoretic framework for signal separation. The recovery of the vectors $\boldsymbol{y}$ and $\boldsymbol{z}$ from the noiseless observation $\boldsymbol{w}$ in (1) can be rephrased as the recovery of $[\boldsymbol{y}^{\mathrm{T}} \ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}$ from the linear measurements

$$\boldsymbol{w} = [\boldsymbol{A} \ \boldsymbol{B}] \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{z} \end{bmatrix}. \tag{2}$$

An information-theoretic framework for analog compression, i.e., for the problem of recovering x from the linear measurements $\mathsf{w} = \boldsymbol{H}\mathsf{x}$, was introduced in [13]. The main achievability result in [13] provides conditions on the compression rate R—in terms of the source vector's Minkowski dimension compression rate—for exact recovery to be possible at arbitrarily small probability of error as the blocklength n goes to infinity. While the information-theoretic framework for signal separation we develop here is inspired by the analog compression framework in [13], there are fundamental differences between these two problems. Specifically, the signal separation applications outlined in Section I (again, we refer to [3, Sec. 1] for specifics) mandate taking specific structural properties of $\boldsymbol{A}$ and $\boldsymbol{B}$ into account. For example, for the recovery of signals corrupted by impulse noise or narrowband interference one of the matrices $\boldsymbol{A}$, $\boldsymbol{B}$ equals the identity matrix or the DFT matrix. This will be accounted for by taking $\boldsymbol{B}$ to be deterministic and fixed throughout the paper. As noted in the introduction, addressing this problem requires new techniques, namely probabilistic uncertainty relations akin to the (deterministic) Donoho-Stark [12] and Elad-Bruckstein [15] uncertainty relations, extended to frames and undercomplete signal sets in [16], and underlying much of compressed sensing theory. These probabilistic uncertainty relations will allow us to make statements that apply to a.a. $\boldsymbol{A}$ for a fixed $\boldsymbol{B}$ (with the set of exceptions for $\boldsymbol{A}$ depending on the specific choice of $\boldsymbol{B}$).

We next define the specifics of our setup.

Definition 1: Suppose that $(\mathsf{Y}_i)_{i\in\mathbb{N}}$ and $(\mathsf{Z}_i)_{i\in\mathbb{N}}$ are stochastic processes on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$. Then, for $n \in \mathbb{N}$, we define the concatenated source vector $\mathbf{x}$ of dimension n as $\mathbf{x} = [\mathsf{X}_1 \dots \mathsf{X}_n]^{\mathrm{T}}$ according to

$$\begin{aligned} \mathsf{X}_i &= \mathsf{Y}_i, \quad \text{for } i \in \{1,\dots,n-\ell\} \\ \mathsf{X}_{n-\ell+i} &= \mathsf{Z}_i, \quad \text{for } i \in \{1,\dots,\ell\}, \end{aligned}$$

where $\ell = \lfloor \lambda n \rfloor$ with the parameter $\lambda \in [0,1]$ representing the asymptotic fraction of components in $\mathbf{x}$ corresponding to the Z_i's.

We emphasize that the distributions of the components Y_i and Z_i in the above definition are general in the sense that they can be a mixture of discrete, continuous, and singular distributions, i.e., $\mu = \mu_{\mathrm{d}} + \mu_{\mathrm{c}} + \mu_{\mathrm{s}}$.

The encoding–decoding part comprises

(i) a measurement matrix $\boldsymbol{H} = [\boldsymbol{A}\ \boldsymbol{B}]\colon \mathbb{R}^{n-\ell} \times \mathbb{R}^{\ell} \to \mathbb{R}^k$, where $\boldsymbol{A} \in \mathbb{R}^{k\times(n-\ell)}$ and $\boldsymbol{B} \in \mathbb{R}^{k\times\ell}$;

(ii) a separator $g : \mathbb{R}^k \to \mathbb{R}^{n-\ell} \times \mathbb{R}^{\ell}$.

We will deal with separators g that are measurable and with g that are, in addition, β-Hölder continuous, i.e., for a given $\beta > 0$ they satisfy

$$\|g(\boldsymbol{x}_1) - g(\boldsymbol{x}_2)\| \leqslant c\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{\beta}, \qquad \text{for all } \boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^k,$$

where $c > 0$ is a constant. Hölder continuous separators are relevant in the context of recovery from noisy, quantized, or otherwise perturbed measurements, but the class of Hölder continuous mappings is significantly smaller than that of measurable mappings.

Definition 2: For $\mathbf{x}$ as in Definition 1 and a given measurement matrix $\boldsymbol{H} = [\boldsymbol{A}\ \boldsymbol{B}]$, we say that there exists a (measurable or β-Hölder continuous) separator that achieves rate $R \in [0,1]$ with error probability $\varepsilon \in (0,1)$ if there exists a sequence (w.r.t. n) of (measurable or β-Hölder continuous) maps g such that $k = \lfloor Rn \rfloor$ and

$$\mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathbf{x}) \neq \mathbf{x}] \mathrel{\dot{\leqslant}} \varepsilon.$$

Next, we quantify the description complexity of $\mathbf{x}$ with general distribution (possibly containing a singular component) through the Minkowski dimension of approximate support sets for $\mathbf{x}$. The Minkowski dimension is sometimes also referred to as box-counting dimension, which explains the origin for the subscript B in the notation $\dim_{\mathrm{B}}(\cdot)$ used below. We start with the definition of Minkowski dimension for general sets.

Definition 3: (Minkowski dimension, [17]). Let $\mathcal{S}$ be a non-empty bounded set in $\mathbb{R}^n$. Define the lower and upper Minkowski dimension of $\mathcal{S}$ as

$$\underline{\dim}_{\mathrm{B}}(\mathcal{S}) = \liminf_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}} \tag{3}$$

$$\overline{\dim}_{\mathrm{B}}(\mathcal{S}) = \limsup_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}}, \tag{4}$$

where $N_{\mathcal{S}}(\delta)$ is the covering number of $\mathcal{S}$ given by

$$N_{\mathcal{S}}(\delta) = \min\Big\{ m \in \mathbb{N} : \mathcal{S} \subseteq \bigcup_{i\in\{1,\dots,m\}} B^n(\boldsymbol{x}_i, \delta),\ \boldsymbol{x}_i \in \mathbb{R}^n \Big\}. \tag{5}$$

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

Remark 1: In Lemma 5 in Appendix B we show that Minkowski dimension can be defined equivalently by replacing $N_{\mathcal{S}}(\delta)$ in (3), (4) by the modified covering number

$$M_{\mathcal{S}}(\delta) = \min\Big\{ m \in \mathbb{N} : \mathcal{S} \subseteq \bigcup_{i\in\{1,\dots,m\}} B^n(\boldsymbol{x}_i, \delta),\ \boldsymbol{x}_i \in \mathcal{S} \Big\}, \tag{6}$$

which is in terms of covering balls that have their centers in the set $\mathcal{S}$. This equivalent definition is often convenient as the covering ball centers inherit structural properties from the set $\mathcal{S}$.

As our framework involves statements that are asymptotic in the blocklength n, we will need a description complexity measure that applies to random processes. This leads to the notion of Minkowski dimension compression rate.

Definition 4: (Minkowski dimension compression rate, [13]). For $\mathbf{x}$ as in Definition 1 and $\varepsilon > 0$, we define the lower and upper Minkowski dimension compression rate as

$$\underline{R}_{\mathrm{B}}(\varepsilon) = \limsup_{n\to\infty} \underline{a}_n(\varepsilon), \quad \text{where}$$

$$\underline{a}_n(\varepsilon) = \inf\left\{ \frac{\underline{\dim}_{\mathrm{B}}(\mathcal{S})}{n} : \mathcal{S} \subseteq \mathbb{R}^n,\ \mathbb{P}[\mathbf{x} \in \mathcal{S}] \ \geqslant 1 - \varepsilon \right\},$$

and

$$\overline{R}_{\mathrm{B}}(\varepsilon) = \limsup_{n\to\infty} \overline{a}_n(\varepsilon), \quad \text{where}$$

$$\overline{a}_n(\varepsilon) = \inf\left\{ \frac{\overline{\dim}_{\mathrm{B}}(\mathcal{S})}{n} : \mathcal{S} \subseteq \mathbb{R}^n,\ \mathbb{P}[\mathbf{x} \in \mathcal{S}] \ \geqslant 1 - \varepsilon \right\}.$$

If $\underline{R}_{\mathrm{B}}(\varepsilon) = \overline{R}_{\mathrm{B}}(\varepsilon)$, we define the Minkowski dimension compression rate as $R_{\mathrm{B}}(\varepsilon) := \underline{R}_{\mathrm{B}}(\varepsilon) = \overline{R}_{\mathrm{B}}(\varepsilon)$.

The following theorem constitutes our first main result.

Theorem 1: Let $\mathbf{x}$ be as in Definition 1. Take $\varepsilon > 0$ and let $R > \underline{R}_{\mathrm{B}}(\varepsilon)$. Then, for every full-rank matrix $\boldsymbol{B} \in \mathbb{R}^{k\times\ell}$, with $k \geqslant \ell$, and for a.a. matrices $\boldsymbol{A} \in \mathbb{R}^{k\times(n-\ell)}$, where $k = \lfloor Rn \rfloor$, there exists a measurable separator g such that

$$\mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathbf{x}) \neq \mathbf{x}] \mathrel{\dot{\leqslant}} \varepsilon. \tag{7}$$

Proof: See Section IV. ■

Remark 2: The set of exceptions for $\boldsymbol{A}$ depends on the specific choice of the full-rank matrix $\boldsymbol{B}$. The proof of Theorem 1 further reveals that the minimum $N \in \mathbb{N}$ for (7) to hold for all $n \geqslant N$ depends on the distribution of $\mathbf{x}$ only and is independent of the matrices $\boldsymbol{A}$ and $\boldsymbol{B}$.

Remark 3: In [13, Thm. 18, 1)] it was shown—in the context of analog compression—that every rate R with $R > \overline{R}_{\mathrm{B}}(\varepsilon)$ is achievable for a.a. measurement matrices $\boldsymbol{H} \in \mathbb{R}^{k\times n}$. The proof of [13, Thm. 18, 1)] relies on intricate properties of invariant measures on Grassmannian manifolds under the action of the orthogonal group. The new proof technique we develop here is based on two key elements, a probabilistic uncertainty relation formalized in Proposition 1 and a concentration of measure result stated in Lemma 1. Specifically, the probabilistic uncertainty relation says that the $(n-k)$-dimensional null-space of $\boldsymbol{H} = [\boldsymbol{A}\ \boldsymbol{B}]$ and the approximate support set $\mathcal{S}$ of $\mathbf{x}$ do not intersect if the Minkowski dimension of $\mathcal{S}$ is smaller than k. Underlying this result is the basic idea that two objects—in general relative position—whose dimensions do not add up to at least the dimension of their ambient space do not intersect. What is surprising is that Euclidean dimension (for the null-space of $\boldsymbol{H}$) and Minkowski dimension (for the support set $\mathcal{S}$) are compatible dimensionality notions in this context.

Remark 4: As pointed out by an anonymous reviewer, it is possible to deduce a proof of Theorem 1 starting from the analog compression result [13, Thm. 18, 1)], which applies to a.a. matrices $\boldsymbol{H} = [\boldsymbol{A}\ \boldsymbol{B}]$, using a version of Fubini's theorem for complete measures [18, Thm. 2.39]. The resulting overall proof is, however, more technical than our proof and does not uncover the underlying probabilistic uncertainty relation. Moreover, the proof technique we develop also applies to the analog compression problem [13] and leads to a simplification of the proof of [13, Thm. 18, 1)] and to significant strengthening of the statement [13, Thm. 18, 2)], as detailed in Section VIII.

While Theorem 1 provides guarantees for the existence of a *measurable* separator, a natural follow-up question is whether we can make a similar statement under continuity/smoothness constraints imposed on the separator. This question is relevant when separation is to be performed from quantized, noisy, or otherwise perturbed observations. It turns out that it is, indeed, possible for fixed $\boldsymbol{B}$ and a.a. $\boldsymbol{A}$ to guarantee the existence of measurable separators that are, in addition, Hölder continuous, even though Hölder continuity is a much stronger property than measurability alone. It is therefore not surprising that the corresponding statement we obtain is weaker, but actually only slightly so, than that for measurable separators. Specifically, we establish the existence of a β-Hölder continuous separator with the threshold $R > \overline{R}_{\mathrm{B}}(\varepsilon)$ instead of $R > \underline{R}_{\mathrm{B}}(\varepsilon)$, provided that $\beta < 1 - \frac{\overline{R}_{\mathrm{B}}(\varepsilon)}{R}$.

Theorem 2: Let $\mathbf{x}$ be as in Definition 1, $R > \overline{R}_{\mathrm{B}}(\varepsilon)$, for $\varepsilon > 0$, and fix $\beta > 0$ such that

$$\beta < 1 - \frac{\overline{R}_{\mathrm{B}}(\varepsilon)}{R}.$$

Then, for every fixed full-rank matrix $\boldsymbol{B} \in \mathbb{R}^{k\times \ell}$, with $k \geqslant \ell$, and a.a. matrices $\boldsymbol{A} \in \mathbb{R}^{k\times(n-\ell)}$, where $k = \lfloor Rn \rfloor$, there exists a β-Hölder continuous separator g such that

$$\mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathbf{x}) \neq \mathbf{x}] \ \dot{\leqslant}\ \varepsilon + \kappa,$$

where $\kappa > 0$ is an arbitrarily small constant.

Remark 5: As in Theorem 1, the set of exceptions for $\boldsymbol{A}$ depends on the specific choice of the full-rank matrix $\boldsymbol{B}$. The constant κ honors the fact that we have to excise a small set of concatenated source vectors on which the separator may fail to be Hölder-continuous. The proof of Theorem 2 is based on a regularized probabilistic uncertainty relation reported in Section V. The regularization accounts for the Hölder-continuity of the separator.

III. Probabilistic Uncertainty Relation

The central conceptual element in the proof of Theorem 1 is a probabilistic uncertainty relation, which leads to uniqueness guarantees for recovery of $\boldsymbol{y}$ and $\boldsymbol{z}$ from $\boldsymbol{w}$ in (2). Formally, the question of uniqueness boils down to asking whether different concatenated source vectors $\boldsymbol{x} = [\boldsymbol{y}^{\mathrm{T}}\ \ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}$ and $\boldsymbol{x}' = [\boldsymbol{y}'^{\mathrm{T}}\ \ \boldsymbol{z}'^{\mathrm{T}}]^{\mathrm{T}}$ exist such that

$$[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{x} = [\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{x}',$$

or, equivalently,

$$\boldsymbol{A}\boldsymbol{p} = \boldsymbol{B}\boldsymbol{q}, \tag{8}$$

with difference vectors $\boldsymbol{p} := \boldsymbol{y} - \boldsymbol{y}'$ and $\boldsymbol{q} := \boldsymbol{z}' - \boldsymbol{z}$. In the context of compressed sensing where $\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{z}$, and $\boldsymbol{z}'$ are sparse signals, $\boldsymbol{p}$ and $\boldsymbol{q}$ are sparse as well so that (8) would imply the existence of a non-zero signal $\boldsymbol{s} := \boldsymbol{A}\boldsymbol{p} = \boldsymbol{B}\boldsymbol{q}$ that can be sparsely represented in both dictionaries $\boldsymbol{A}$ and $\boldsymbol{B}$. Uncertainty principles are at the heart of compressed sensing theory and state that no such $\boldsymbol{s}$ can exist if the signals $\boldsymbol{y}, \boldsymbol{y}', \boldsymbol{z}$, and $\boldsymbol{z}'$ and hence $\boldsymbol{p}$ and $\boldsymbol{q}$ are sufficiently sparse and the dictionaries $\boldsymbol{A}$ and $\boldsymbol{B}$ are sufficiently incoherent, thereby guaranteeing that, for a given $\boldsymbol{w}$, there is a unique pair $(\boldsymbol{y}, \boldsymbol{z})$ such that $\boldsymbol{w} = \boldsymbol{A}\boldsymbol{y} + \boldsymbol{B}\boldsymbol{z}$. Specifically, the Donoho-Stark uncertainty principle [12] applies to the square matrices $\boldsymbol{A} = \boldsymbol{I}_n$ and $\boldsymbol{B} = \boldsymbol{F}_n$, and states that there exists no pair of vectors $(\boldsymbol{p}, \boldsymbol{q}) \neq \boldsymbol{0}$ with $2n_{\boldsymbol{p}}n_{\boldsymbol{q}} < n$ satisfying (8), where $n_{\boldsymbol{p}}$ and $n_{\boldsymbol{q}}$ denote the number of non-zero entries in $\boldsymbol{p}$ and $\boldsymbol{q}$, respectively. Elad and Bruckstein [15] generalized the Donoho-Stark uncertainty principle to arbitrary orthonormal bases $\boldsymbol{A}$ and $\boldsymbol{B}$ and found that no pair of vectors $(\boldsymbol{p}, \boldsymbol{q}) \neq \boldsymbol{0}$ with $(n_{\boldsymbol{p}} + n_{\boldsymbol{q}})/2 < 1/\mu$ satisfying (8) exists. Here,

$$\mu := \sup_{1\leqslant i,j\leqslant n} |\langle \boldsymbol{a}_i, \boldsymbol{b}_j \rangle|$$

is the coherence of $\boldsymbol{A} = [\boldsymbol{a}_1 \dots \boldsymbol{a}_n]$ and $\boldsymbol{B} = [\boldsymbol{b}_1 \dots \boldsymbol{b}_n]$. This uncertainty principle was further extended to redundant and undercomplete dictionaries in [16]. The essence of uncertainty relations is that uniqueness in signal separation or signal recovery can be enforced by demanding that the signals to be separated or recovered, respectively, be sufficiently sparse, provided that the underlying dictionaries are incoherent enough.

The central tool in the proof of Theorem 1 is a probabilistic uncertainty relation obtained as follows. We first rewrite (8) as

$[\boldsymbol{A}\ \boldsymbol{B}][\boldsymbol{p}^{\mathrm{T}}\ -\boldsymbol{q}^{\mathrm{T}}]^{\mathrm{T}}=\mathbf{0}$ and then assume that $[\boldsymbol{p}^{\mathrm{T}}\ -\boldsymbol{q}^{\mathrm{T}}]^{\mathrm{T}}$ lies in a set $\mathcal{S}$ of (sufficiently) small Minkowski dimension. The probabilistic uncertainty relation, stated formally in Proposition 1, says that for fixed $\boldsymbol{B}$ and for a.a. $\boldsymbol{A}$, there is no $[\boldsymbol{p}^{\mathrm{T}}\ -\boldsymbol{q}^{\mathrm{T}}]^{\mathrm{T}}\in\mathcal{S}\backslash\{\mathbf{0}\}$ such that $[\boldsymbol{A}\ \boldsymbol{B}][\boldsymbol{p}^{\mathrm{T}}\ -\boldsymbol{q}^{\mathrm{T}}]^{\mathrm{T}}=\mathbf{0}$ or equivalently (8) holds. Minkowski dimension here replaces sparsity in terms of the number of non-zero entries as a measure of description complexity of the signals to be separated.

Proposition 1: Let $\mathcal{S}\subseteq\mathbb{R}^n$ be non-empty and bounded such that $\underline{\dim}_{\mathrm{B}}(\mathcal{S})<k$, and let $\boldsymbol{B}\in\mathbb{R}^{k\times\ell}$, with $k\geqslant\ell$, be a matrix with $\operatorname{rank}(\boldsymbol{B})=\ell$. Then,

$$\{\boldsymbol{x}\in\mathcal{S}\backslash\{\mathbf{0}\}\ :\ [\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{x}=\mathbf{0}\}=\emptyset, \tag{9}$$

for a.a. $\boldsymbol{A}\in\mathbb{R}^{k\times(n-\ell)}$.

Proof: We show that (9) holds with probability (w.p.) 1 for the random matrix $\mathbf{A}=[\mathbf{a}_1\ \dots\ \mathbf{a}_k]^{\mathrm{T}}$, where the $\mathbf{a}_i$ are i.i.d. uniform on $B^{n-\ell}(\mathbf{0},r)$ and $r>0$ is arbitrary. Since r can, in particular, be chosen arbitrarily large, this establishes that the Lebesgue measure of matrices $\boldsymbol{A}$ violating (9) is zero. We split $\boldsymbol{x}=[\boldsymbol{y}^{\mathrm{T}}\ \ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}$, where $\boldsymbol{y}\in\mathbb{R}^{n-\ell}$ and $\boldsymbol{z}\in\mathbb{R}^{\ell}$, and note that, thanks to the full-rank assumption on $\boldsymbol{B}$, it suffices to show that

$$\mathbb{P}\left[\exists\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}\backslash\{\mathbf{0}\}:[\mathbf{A}\ \boldsymbol{B}]\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\mathbf{0}\right]=0 \tag{10}$$

for sets $\mathcal{S}$ that have the norm of the $\boldsymbol{y}$-parts of their elements bounded away from zero. To see this, we first note that $\boldsymbol{B}$, by virtue of being full-rank, maps non-zero vectors to non-zero vectors. For $[\boldsymbol{y}^{\mathrm{T}}\ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}\in\mathcal{S}\backslash\{\mathbf{0}\}$, $\boldsymbol{A}\boldsymbol{y}+\boldsymbol{B}\boldsymbol{z}=\mathbf{0}$ is therefore possible only for $\boldsymbol{y}\neq\mathbf{0}$ as $\boldsymbol{y}=\mathbf{0}$ would lead to $\boldsymbol{B}\boldsymbol{z}=\mathbf{0}$ which in turn would result in $[\boldsymbol{y}^{\mathrm{T}}\ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}=\mathbf{0}$. We can hence rewrite (10) as

$$\begin{aligned}&\mathbb{P}\left[\exists\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}\backslash\{\mathbf{0}\}:[\mathbf{A}\ \boldsymbol{B}]\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\mathbf{0}\right]\\&\quad=\mathbb{P}\left[\exists\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}\backslash\{\mathbf{0}\},\boldsymbol{y}\neq\mathbf{0}:[\mathbf{A}\ \boldsymbol{B}]\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\mathbf{0}\right].\end{aligned} \tag{11}$$

A union bound argument applied to (11) then yields

$$\begin{aligned}&\mathbb{P}\left[\exists\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}\backslash\{\mathbf{0}\}:[\mathbf{A}\ \boldsymbol{B}]\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\mathbf{0}\right]\\&\leqslant\sum_{m=1}^{\infty}\mathbb{P}\left[\exists\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}\backslash\{\mathbf{0}\},\|\boldsymbol{y}\|\geqslant\frac{1}{m}:[\mathbf{A}\ \boldsymbol{B}]\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\mathbf{0}\right].\end{aligned} \tag{12}$$

This allows us to conclude that (10) is established by showing that

$$\mathbb{P}\left[\exists\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}':[\mathbf{A}\ \boldsymbol{B}]\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\mathbf{0}\right]=0, \tag{13}$$

for all non-empty bounded sets $\mathcal{S}'\subseteq\mathcal{S}$ with

$$\inf\left\{\|\boldsymbol{y}\|:\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}'\right\}>0, \tag{14}$$

as this implies that each term in the series in (12) equals zero. Note that we no longer need to excise $\mathbf{0}$ from $\mathcal{S}'$ in (13), as $\mathcal{S}'$ is guaranteed not to contain $\mathbf{0}$ by definition, cf. (14).

We next employ a covering argument, which reduces the question of the existence of $[\boldsymbol{y}^{\mathrm{T}}\ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}\in\mathcal{S}'$ such that

$$\mathbf{A}\boldsymbol{y}+\boldsymbol{B}\boldsymbol{z}=\mathbf{0} \tag{15}$$

to the question of the existence of covering ball centers satisfying (15). For reasons that will become clear towards the end of the proof, we employ the modified covering number $M_{\mathcal{S}'}(\delta)$ (defined in (6)), which requires the covering ball centers to lie in $\mathcal{S}'$. This implies that the covering ball centers $[\boldsymbol{y}_i^{\mathrm{T}}\ \boldsymbol{z}_i^{\mathrm{T}}]^{\mathrm{T}}$ satisfy

$$\min_i\|\boldsymbol{y}_i\|\geqslant\inf\left\{\|\boldsymbol{y}\|:\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}'\right\}>0. \tag{16}$$

By definition of $\underline{\dim}_{\mathrm{B}}(\cdot)$ in (3) there exists a sequence of covering ball radii δ_j tending to zero with corresponding covering ball centers $\boldsymbol{x}_1,\dots,\boldsymbol{x}_{M_{\mathcal{S}'}(\delta_j)}\in\mathcal{S}'$ such that

$$\lim_{j\to\infty}\frac{\log M_{\mathcal{S}'}(\delta_j)}{\log\frac{1}{\delta_j}}=\underline{\dim}_{\mathrm{B}}(\mathcal{S}'). \tag{17}$$

Next, we note that

$$\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\|\leqslant c(k,r,\|\boldsymbol{B}\|)\|\boldsymbol{x}\|,\quad\text{for all }\boldsymbol{x}\in\mathbb{R}^n, \tag{18}$$

since i) $\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\|\leqslant\|[\mathbf{A}\ \boldsymbol{B}]\|\cdot\|\boldsymbol{x}\|$, ii) $\|[\mathbf{A}\ \boldsymbol{B}]\|\leqslant\|\mathbf{A}\|+\|\boldsymbol{B}\|$, and iii)

$$\begin{aligned}\|\mathbf{A}\boldsymbol{y}\|&=\sqrt{\langle\mathbf{a}_1,\boldsymbol{y}\rangle^2+\ldots+\langle\mathbf{a}_k,\boldsymbol{y}\rangle^2}\\&\leqslant\sqrt{\|\mathbf{a}_1\|^2\|\boldsymbol{y}\|^2+\ldots+\|\mathbf{a}_k\|^2\|\boldsymbol{y}\|^2}\\&<r\sqrt{k}\|\boldsymbol{y}\|,\end{aligned} \tag{19}$$

for all $\boldsymbol{y}\in\mathbb{R}^{n-\ell}$, implying $\|\mathbf{A}\|<r\sqrt{k}$, where in (19) we used $\mathbf{a}_i\in B^{n-\ell}(\mathbf{0},r)$, for $i=1,...,k$. Putting things together, we find that

$$\begin{aligned}&\mathbb{P}\left[\exists\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}\in\mathcal{S}':[\mathbf{A}\ \boldsymbol{B}]\begin{bmatrix}\boldsymbol{y}\\\boldsymbol{z}\end{bmatrix}=\mathbf{0}\right]\\&\leqslant\sum_{i=1}^{M_{\mathcal{S}'}(\delta_j)}\mathbb{P}[\exists\boldsymbol{x}\in B^n(\boldsymbol{x}_i,\delta_j):[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}=\mathbf{0}] &(20)\\&\leqslant\sum_{i=1}^{M_{\mathcal{S}'}(\delta_j)}\mathbb{P}[\exists\boldsymbol{x}\in B^n(\boldsymbol{x}_i,\delta_j):\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\|<\delta_j]\\&\leqslant\sum_{i=1}^{M_{\mathcal{S}'}(\delta_j)}\mathbb{P}[\|\mathbf{A}\boldsymbol{y}_i+\boldsymbol{B}\boldsymbol{z}_i\|<(c(k,r,\|\boldsymbol{B}\|)+1)\delta_j] &(21)\\&\leqslant C(n,k,r,\|\boldsymbol{B}\|)\,M_{\mathcal{S}'}(\delta_j)\,\delta_j^k, &(22)\end{aligned}$$

where (20) follows from a union bound over the covering balls $B^n(\boldsymbol{x}_i,\delta_j)$ of $\mathcal{S}'$ and in (21) we set $\boldsymbol{x}_i=[\boldsymbol{y}_i^{\mathrm{T}}\ \boldsymbol{z}_i^{\mathrm{T}}]^{\mathrm{T}}$ and used

$$\begin{aligned}\|\mathbf{A}\boldsymbol{y}_i+\boldsymbol{B}\boldsymbol{z}_i\|&=\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_i\|\\&\leqslant\|[\mathbf{A}\ \boldsymbol{B}](\boldsymbol{x}_i-\boldsymbol{x})\|+\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\|\\&\leqslant c(k,r,\|\boldsymbol{B}\|)\delta_j+\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\|.\end{aligned}$$

Finally, (22) is by application of the concentration of measure result in Lemma 1 below (for fixed, albeit arbitrarily large,

r), where we used (16) to deduce that $\boldsymbol{y}_i \neq \mathbf{0}$, which, in turn, allows us to absorb the term $1/\|\boldsymbol{y}_i\|^k$ into the constant $C(n,k,r,\|\boldsymbol{B}\|)$. Now,

$$\lim_{j\to\infty} \frac{\log\big(M_{\mathcal{S}'}(\delta_j)\delta_j^k\big)}{\log\frac{1}{\delta_j}} = \lim_{j\to\infty} \frac{\log M_{\mathcal{S}'}(\delta_j)}{\log\frac{1}{\delta_j}} - k \quad (23)$$
$$= \underline{\dim}_{\mathrm{B}}(\mathcal{S}') - k$$
$$< 0,$$

where we used (17) together with $\underline{\dim}_{\mathrm{B}}(\mathcal{S}') \leqslant \underline{\dim}_{\mathrm{B}}(\mathcal{S})$ thanks to $\mathcal{S}' \subseteq \mathcal{S}$. Since $\lim_{j\to\infty}\log(1/\delta_j) = \infty$, the convergence of the left-hand side (LHS) of (23) to a finite negative number implies that $\lim_{j\to\infty}\log\big(M_{\mathcal{S}'}(\delta_j)\delta_j^k\big) = -\infty$ and hence $\lim_{j\to\infty} M_{\mathcal{S}'}(\delta_j)\delta_j^k = 0$. Taking the limit $j\to\infty$ in (20)–(22) implies that the LHS in (20) equals zero, which concludes the proof. ■

Proposition 1 shows that we can enforce uniqueness of the solution in the recovery of $\boldsymbol{y}$ and $\boldsymbol{z}$ from $\boldsymbol{w}$ in (2) by requiring that $\boldsymbol{x} = [\boldsymbol{y}^{\mathrm{T}}\ \boldsymbol{z}^{\mathrm{T}}]^{\mathrm{T}}$ lie in a set with small enough Minkowski dimension. We note that this condition is in terms of a general measure for the description complexity of x, namely Minkowski dimension, and includes the case of traditional sparsity as measured in terms of the number of non-zero entries. Section VII elaborates on this matter.

It remains to establish the concentration of measure result employed in the proof of Proposition 1.

Lemma 1: Let $\mathbf{A} = [\mathbf{a}_1 \dots \mathbf{a}_k]^{\mathrm{T}}$ be a random matrix in $\mathbb{R}^{k\times p}$ where the $\mathbf{a}_i$ are i.i.d. uniform on $B^p(\mathbf{0},r)$, for $r>0$. Then, for each $\boldsymbol{u}\in\mathbb{R}^p\backslash\{\mathbf{0}\}$, each $\boldsymbol{v}\in\mathbb{R}^k$, and all $\delta>0$, we have

$$\mathbb{P}[\|\mathbf{A}\boldsymbol{u}+\boldsymbol{v}\| < \delta] \leqslant C(p,k,r)\frac{\delta^k}{\|\boldsymbol{u}\|^k}.$$

Proof: We start by noting that, by assumption, the random matrix $\mathbf{A}$ is uniformly distributed in the k-fold product set $B^p(\mathbf{0},r)\times\dots\times B^p(\mathbf{0},r)$, which is of Lebesgue measure $\alpha(p,r)^k$. We therefore have

$$\mathbb{P}[\|\mathbf{A}\boldsymbol{u}+\boldsymbol{v}\| < \delta]$$
$$= \frac{1}{\alpha(p,r)^k}\,\mathrm{Leb}^{kp}\{\boldsymbol{A}\in B^p(\mathbf{0},r)\times\dots\times B^p(\mathbf{0},r):$$
$$\|\boldsymbol{A}\boldsymbol{u}+\boldsymbol{v}\|<\delta\}$$
$$\leqslant \frac{1}{\alpha(p,r)^k}\prod_{i=1}^{k}\mathrm{Leb}^p\{\boldsymbol{a}_i\in B^p(\mathbf{0},r): |\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{u}+v_i|<\delta\} \quad (24)$$
$$= \frac{1}{\alpha(p,r)^k}\prod_{i=1}^{k}\mathrm{Leb}^p\bigg\{\boldsymbol{U}\boldsymbol{a}_i\in B^p(\mathbf{0},r):$$
$$\Big|(\boldsymbol{U}\boldsymbol{a}_i)^{\mathrm{T}}\frac{\boldsymbol{u}}{\|\boldsymbol{u}\|}+\frac{v_i}{\|\boldsymbol{u}\|}\Big|<\frac{\delta}{\|\boldsymbol{u}\|}\bigg\} \quad (25)$$
$$= \frac{1}{\alpha(p,r)^k}\prod_{i=1}^{k}\mathrm{Leb}^p\bigg\{\boldsymbol{a}_i\in B^p(\mathbf{0},r):$$
$$\Big|\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{e}_1+\frac{v_i}{\|\boldsymbol{u}\|}\Big|<\frac{\delta}{\|\boldsymbol{u}\|}\bigg\} \quad (26)$$
$$\leqslant \frac{(2r)^{k(p-1)}}{\alpha(p,r)^k}\prod_{i=1}^{k}\mathrm{Leb}^1\bigg\{a_i\in\mathbb{R}: \Big|a_i+\frac{v_i}{\|\boldsymbol{u}\|}\Big|<\frac{\delta}{\|\boldsymbol{u}\|}\bigg\} \quad (27)$$
$$= \frac{(2r)^{k(p-1)}(2\delta)^k}{\alpha(p,r)^k\|\boldsymbol{u}\|^k}, \quad (28)$$

where (24) holds by the multiplicativity of Lebesgue measure and because $\|\boldsymbol{A}\boldsymbol{u}+\boldsymbol{v}\|<\delta$ implies $|\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{u}+v_i|<\delta$, for all i, (25) follows from $\boldsymbol{u}\neq\mathbf{0}$ and the fact that Leb^p is invariant under rotations, with the specific rotation $\boldsymbol{U}^{\mathrm{T}}$ considered here taking $\boldsymbol{u}/\|\boldsymbol{u}\|$ into $\boldsymbol{e}_1 = [1\ 0\ \dots\ 0]^{\mathrm{T}}\in\mathbb{R}^p$, in (26) we relabel $\boldsymbol{U}\boldsymbol{a}_i\in B^p(\mathbf{0},r)$ as $\boldsymbol{a}_i\in B^p(\mathbf{0},r)$, and in (27) we denote the first component of the vector $\boldsymbol{a}_i$ by a_i, we relax the condition on the magnitude of a_i to $a_i\in\mathbb{R}$, and we use the monotonicity of Lebesgue measure together with the fact that the magnitudes of the remaining components of $\boldsymbol{a}_i$ are less than or equal to r. Finally, (28) follows by noting that in (27) we take the product over the Lebesgue measures of intervals of length $2\delta/\|\boldsymbol{u}\|$. ■

The proofs of Proposition 1 and Lemma 1 are inspired by the proofs of [14, Lem. 4.2, Lem. 4.3], but use a new proof technique that is more direct. Finally, we note that the probabilistic uncertainty relation developed here is a quite general tool and has been applied to establish information-theoretic limits of matrix completion [19] and of phase retrieval [20].

IV. Proof of Theorem 1

Since $R > \underline{R}_{\mathrm{B}}(\varepsilon) = \limsup_{n\to\infty}\underline{a}_n(\varepsilon)$ and $k = \lfloor Rn\rfloor$, both by assumption, we have

$$\underline{a}_n(\varepsilon) \,\dot{<}\, \frac{k}{n},$$

which, together with the definition of $\underline{a}_n(\varepsilon)$, implies that there exists a sequence[3] of non-empty compact sets[4] $\mathcal{U} := \mathcal{U}_n \subseteq \mathbb{R}^n$ such that

$$\underline{\dim}_{\mathrm{B}}(\mathcal{U}) \,\dot{<}\, k \quad (29)$$
$$\text{and}\quad \mathbb{P}[\mathsf{x}\in\mathcal{U}] \geqslant 1-\varepsilon. \quad (30)$$

In the remainder of the proof we take n to be sufficiently large for (29) to hold in the $\#$-sense and we drop the dot-notation. Let $\boldsymbol{B}\in\mathbb{R}^{k\times\ell}$ be an arbitrary but fixed full-rank matrix with $k\geqslant\ell$. Now, consider the mapping $\mathbb{R}^{k\times(n-\ell)}\times\mathbb{R}^n\times\mathbb{R}^k\to\mathbb{R}$, $(\boldsymbol{A},\boldsymbol{u},\boldsymbol{v})\mapsto\|[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}-\boldsymbol{v}\|$. Since this mapping is continuous and $[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}=\boldsymbol{v}$ if and only if $\|[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}-\boldsymbol{v}\|=0$, [21, Prop. 14.33 and Cor. 14.6] and the compactness of $\mathcal{U}$ imply that there exists a measurable mapping $f\colon\mathbb{R}^{k\times(n-\ell)}\times\mathbb{R}^k\to\mathbb{R}^n$ satisfying[5]

$$f(\boldsymbol{A},\boldsymbol{v})\in\{\boldsymbol{u}\in\mathcal{U}:[\boldsymbol{A}\ \ \boldsymbol{B}]\boldsymbol{u}=\boldsymbol{v}\},$$

if $\{\boldsymbol{u}\in\mathcal{U}:[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}=\boldsymbol{v}\}\neq\emptyset$ and $f(\boldsymbol{A},\boldsymbol{v})=\boldsymbol{e}\in\mathbb{R}^n\backslash\mathcal{U}$ else. The mapping $g=f(\boldsymbol{A},\cdot)$ therefore constitutes a valid (i.e., measurable) separator. This separator is guaranteed to deliver a $\boldsymbol{u}\in\mathcal{U}$ that is consistent with the observation $\boldsymbol{v}$ (in the sense of $[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}=\boldsymbol{v}$) if at least one such consistent $\boldsymbol{u}$

[3]The symbol $\mathcal{U}$ actually denotes the sequence $\mathcal{U}_n$. We decided to drop the index n for simplicity of exposition.

[4]Since lower Minkowski dimension is invariant under set closure ([17, Prop. 3.4]), we can assume, w.l.o.g., that $\mathcal{U}$ is compact.

[5]For the detailed arguments leading to this statement, we refer to [22].

exists, otherwise an error is declared by delivering the "error symbol" $\boldsymbol{e}$. We can now define the set

$$\begin{aligned}\mathcal{A} &:= \{(\boldsymbol{A},\boldsymbol{x}) \in \mathbb{R}^{k\times(n-\ell)} \times \mathcal{U} : f(\boldsymbol{A},[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{x}) \neq \boldsymbol{x}\} \quad (31)\\ &= \{(\boldsymbol{A},\boldsymbol{x}) \in \mathbb{R}^{k\times(n-\ell)} \times \mathcal{U} : g([\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{x}) \neq \boldsymbol{x}\}\end{aligned}$$

and upper-bound the probability of decoding error according to

$$\begin{aligned}p_{\mathrm{e}}(\boldsymbol{A}) &:= \mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathbf{x}) \neq \mathbf{x}] && (32)\\ &= \mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathbf{x}) \neq \mathbf{x}, \mathbf{x} \in \mathcal{U}]\\ &\quad + \mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathbf{x}) \neq \mathbf{x}, \mathbf{x} \notin \mathcal{U}]\\ &\leqslant \mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathbf{x}) \neq \mathbf{x}, \mathbf{x} \in \mathcal{U}] + \varepsilon && (33)\\ &= \mathbb{P}[(\boldsymbol{A},\mathbf{x}) \in \mathcal{A}] + \varepsilon, \quad \boldsymbol{A} \in \mathbb{R}^{k\times(n-\ell)}, && (34)\end{aligned}$$

where (33) follows from (30). Since $\mathcal{A}$ is measurable[5], we can apply Fubini's theorem [23, Thm. 1.14] to the indicator function on $\mathcal{A}$ and get

$$\begin{aligned}&\int_{\mathbb{R}^{k\times(n-\ell)}} \mathbb{P}[(\boldsymbol{A},\mathbf{x}) \in \mathcal{A}]\mathrm{d}\boldsymbol{A} && (35)\\ &= \int_{\mathcal{U}} \mathrm{Leb}^{k(n-\ell)}\{\boldsymbol{A} : (\boldsymbol{A},\boldsymbol{x}) \in \mathcal{A}\}\mu_{\mathbf{x}}(\mathrm{d}\boldsymbol{x}).\end{aligned}$$

Note that for $\boldsymbol{v} = [\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{x}$ with $\boldsymbol{x} \in \mathcal{U}$, the separator g can make an error only if there exists a $\boldsymbol{u} \in \mathcal{U}\backslash\{\boldsymbol{x}\}$ that is consistent with $\boldsymbol{v}$, i.e., if $\boldsymbol{v} = [\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}$ for some $\boldsymbol{u} \in \mathcal{U}\backslash\{\boldsymbol{x}\}$. We therefore have

$$\mathcal{A} \subseteq \{(\boldsymbol{A},\boldsymbol{x}) \in \mathbb{R}^{k\times(n-\ell)} \times \mathcal{U} : \ker([\boldsymbol{A}\ \boldsymbol{B}]) \cap \mathcal{U}_{\boldsymbol{x}} \neq \{\boldsymbol{0}\}\},$$

where

$$\mathcal{U}_{\boldsymbol{x}} = \{\boldsymbol{u} - \boldsymbol{x} : \boldsymbol{u} \in \mathcal{U}\}, \quad \boldsymbol{x} \in \mathcal{U},$$

which implies

$$\{\boldsymbol{A} : (\boldsymbol{A},\boldsymbol{x}) \in \mathcal{A}\} \subseteq \{\boldsymbol{A} : \ker([\boldsymbol{A}\ \boldsymbol{B}]) \cap \mathcal{U}_{\boldsymbol{x}} \neq \{\boldsymbol{0}\}\},$$

for all $\boldsymbol{x} \in \mathcal{U}$. The monotonicity of Lebesgue measure therefore yields

$$\begin{aligned}&\mathrm{Leb}^{k(n-\ell)}\{\boldsymbol{A} : (\boldsymbol{A},\boldsymbol{x}) \in \mathcal{A}\}\\ &\leq \mathrm{Leb}^{k(n-\ell)}\{\boldsymbol{A} : \ker([\boldsymbol{A}\ \boldsymbol{B}]) \cap \mathcal{U}_{\boldsymbol{x}} \neq \{\boldsymbol{0}\}\}, \quad (36)\end{aligned}$$

for all $\boldsymbol{x} \in \mathcal{U}$. The probabilistic uncertainty relation, Proposition 1, with $\mathcal{S} = \mathcal{U}_{\boldsymbol{x}}$ and $\underline{\dim}_{\mathrm{B}}(\mathcal{U}_{\boldsymbol{x}}) < k$ ((lower) Minkowski dimension is invariant under translation, as seen by translating covering balls accordingly, and hence $\underline{\dim}_{\mathrm{B}}(\mathcal{U}_{\boldsymbol{x}}) = \underline{\dim}_{\mathrm{B}}(\mathcal{U})$) implies that (36) equals zero for all $\boldsymbol{x} \in \mathcal{U}$. Therefore, (35) equals zero as well, which, by (32)–(34), implies $p_{\mathrm{e}}(\boldsymbol{A}) \leq \varepsilon$ for a.a. $\boldsymbol{A}$ and thereby completes the proof. ∎

V. Regularized Probabilistic Uncertainty Relation

In this section, we develop the regularized probabilistic uncertainty relation the proof of Theorem 2 is based on. We start with results on the existence of Hölder continuous separators.

Definition 5: For $\mathcal{A} \subseteq \mathbb{R}^n$, $\mathcal{B} \subseteq \mathbb{R}^m$, and $\beta > 0$, a map $f\colon \mathcal{A} \to \mathcal{B}$ is β-Hölder continuous if there exists a constant $c > 0$ such that for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{A}$ we have

$$\|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\| \leqslant c\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{\beta}.$$

Lemma 2: For a map $f\colon \mathcal{A} \to \mathcal{B}$, where $\mathcal{A} \subseteq \mathbb{R}^n$ and $\mathcal{B} \subseteq \mathbb{R}^m$, there exist $c > 0$ and $\beta > 0$ such that

$$c\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{1/\beta} \leqslant \|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\|, \quad (37)$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{A}$, if and only if f is injective and $f^{-1}\colon f(\mathcal{A}) \to \mathcal{A}$ is β-Hölder continuous.

Proof: If (37) holds, then f is injective as for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{A}$ with $\boldsymbol{x}_1 \neq \boldsymbol{x}_2$, we have

$$\|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\| \geqslant c\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{1/\beta} > 0,$$

and hence $f(\boldsymbol{x}_1) \neq f(\boldsymbol{x}_2)$. Therefore, $f^{-1}\colon f(\mathcal{A}) \to \mathcal{A}$ is well-defined. Moreover, for all $\boldsymbol{y}_1, \boldsymbol{y}_2 \in f(\mathcal{A})$ we can find $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{A}$ such that $f(\boldsymbol{x}_i) = \boldsymbol{y}_i$ and hence β-Hölder continuity of f^{-1} follows from

$$\begin{aligned}\|f^{-1}(\boldsymbol{y}_1) - f^{-1}(\boldsymbol{y}_2)\| &= \|\boldsymbol{x}_1 - \boldsymbol{x}_2\|\\ &\leqslant \frac{1}{c^{\beta}}\|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\|^{\beta}\\ &= \frac{1}{c^{\beta}}\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|^{\beta},\end{aligned}$$

where the inequality is by (37).

Conversely, suppose that f is injective and $f^{-1}\colon f(\mathcal{A}) \to \mathcal{A}$ is β-Hölder continuous. Then, for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{A}$, by β-Hölder continuity of f^{-1} there exists a constant C such that

$$\|f^{-1}(f(\boldsymbol{x}_1)) - f^{-1}(f(\boldsymbol{x}_2))\| \leqslant C\|f(\boldsymbol{x}_1) - f(\boldsymbol{x}_2)\|^{\beta}.$$

Since $f^{-1}(f(\boldsymbol{x}_i)) = \boldsymbol{x}_i$ by injectivity of f, this implies (37) with $c := 1/C^{1/\beta}$. ∎

For a linear map f, e.g., the map induced by a realization of the random matrix $[\mathbf{A}\ \boldsymbol{B}]$, verifying (37) reduces to checking the condition

$$\inf_{\boldsymbol{x} \in (\mathcal{A}\ominus\mathcal{A})\backslash\{\boldsymbol{0}\}} \frac{\|f(\boldsymbol{x})\|}{\|\boldsymbol{x}\|^{1/\beta}} > 0. \quad (38)$$

We next provide a sufficient condition—that is convenient to check—for (38) to hold. For expositional simplicity, we formulate the condition for general sets $\mathcal{S}$ in place of $\mathcal{A} \ominus \mathcal{A}$. The condition we establish essentially consists of checking whether the elements in the set obtained upon excision of a ball of radius $2^{-j\beta}$ from $\mathcal{S}$ map to points outside a ball of radius 2^{-j}. A related approach was used in [13, p. 3736].

Lemma 3: Let $\mathcal{S}$ be a nonempty and bounded set in $\mathbb{R}^n$, $\mathcal{S} \neq \{\boldsymbol{0}\}$, $f\colon \mathcal{S} \to \mathbb{R}^k$, $\beta \in (0,1)$, and $\delta_j := 2^{-j}$. If there is a $J \in \mathbb{N}$ such that for all $j \geqslant J$ we have

$$\|f(\boldsymbol{x})\| \geqslant \delta_j, \quad \text{for all } \boldsymbol{x} \in \mathcal{S}\backslash B^n(\boldsymbol{0}, \delta_j^{\beta}), \quad (39)$$

then

$$\inf_{\boldsymbol{x} \in \mathcal{S}\backslash\{\boldsymbol{0}\}} \frac{\|f(\boldsymbol{x})\|}{\|\boldsymbol{x}\|^{1/\beta}} > 0. \quad (40)$$

Proof: We show that there exists a constant $c(\mathcal{S}) > 0$ such that

$$c(\mathcal{S})\|\boldsymbol{x}\|^{1/\beta} \leqslant \|f(\boldsymbol{x})\|, \quad \text{for all } \boldsymbol{x} \in \mathcal{S},$$

which is equivalent to (40). Let $\boldsymbol{x} \in \mathcal{S}$, $\mathcal{S}_j := \mathcal{S}\backslash B^n(\boldsymbol{0}, \delta_j^{\beta})$, and $i_{\boldsymbol{x}} := \min\{i \in \mathbb{N} : \boldsymbol{x} \in \mathcal{S}_i\}$ (see Figure 1 for an illustration).

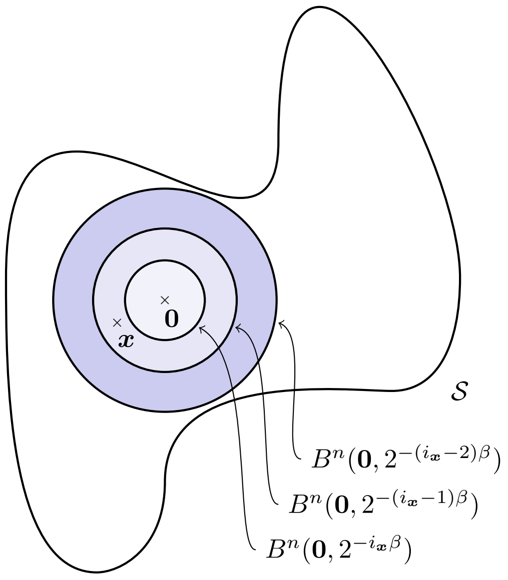

Fig. 1: Illustration of the set $\mathcal{S}_j$ for $j = i_{\boldsymbol{x}} - 2, i_{\boldsymbol{x}} - 1, i_{\boldsymbol{x}}$.

We then find that

$$\|f(\boldsymbol{x})\| \geqslant \begin{cases} \delta_{i_{\boldsymbol{x}}}, & \text{if } i_{\boldsymbol{x}} \geqslant J \\ \delta_J, & \text{if } i_{\boldsymbol{x}} < J \end{cases} \tag{41}$$

$$\geqslant \begin{cases} \|\boldsymbol{x}\|^{1/\beta}/2, & \text{if } i_{\boldsymbol{x}} \geqslant J \\ 2^{-J} \frac{\|\boldsymbol{x}\|^{1/\beta}}{\sup_{\boldsymbol{u}\in\mathcal{S}} \|\boldsymbol{u}\|^{1/\beta}}, & \text{if } i_{\boldsymbol{x}} < J \end{cases} \tag{42}$$

$$\geqslant c(\mathcal{S})\|\boldsymbol{x}\|^{1/\beta}, \tag{43}$$

where (41) follows, for $i_{\boldsymbol{x}} \geqslant J$, from $\boldsymbol{x} \in \mathcal{S}_{i_{\boldsymbol{x}}}$ together with (39); and for $i_{\boldsymbol{x}} < J$, from $\boldsymbol{x} \in \mathcal{S}_J$ and (39) with $j = J$. In (42) we used $\|\boldsymbol{x}\| < 2^{-(i_{\boldsymbol{x}}-1)\beta}$, for $i_{\boldsymbol{x}} \geqslant J$, and for $i_{\boldsymbol{x}} < J$ we apply the trivial bound $\|\boldsymbol{x}\| \leqslant \sup_{\boldsymbol{u}\in\mathcal{S}} \|\boldsymbol{u}\|$. Finally, in (43) we set $c(\mathcal{S}) = \min\left\{1/2, \frac{2^{-J}}{\sup_{\boldsymbol{u}\in\mathcal{S}} \|\boldsymbol{u}\|^{1/\beta}}\right\}$, and we note that $c(\mathcal{S}) > 0$ by virtue of $\mathcal{S}$ being bounded and $J < \infty$. ■

We are now ready to present the announced regularized probabilistic uncertainty relation. In the original probabilistic uncertainty relation, stated in Proposition 1, we showed that for fixed $\boldsymbol{B}$, for a.a. $\boldsymbol{A}$ there are no non-zero vectors in $\mathcal{S}$ that map to zero under $[\boldsymbol{A}\ \boldsymbol{B}]$ provided that the lower Minkowski dimension of $\mathcal{S}$ is sufficiently small. The regularized version of this result states that the norm of the image of a vector $\boldsymbol{x} \in \mathcal{S}$ under $[\boldsymbol{A}\ \boldsymbol{B}]$ does not become too small relative to $\|\boldsymbol{x}\|$. This will then allow us to deduce the existence of a Hölder continuous separator in Theorem 2 by applying Lemma 2.

Proposition 2: Let $\boldsymbol{B} \in \mathbb{R}^{k\times\ell}$, with $k \geqslant \ell$, have $\operatorname{rank}(\boldsymbol{B}) = \ell$, let $\mathcal{S} \subseteq \mathbb{R}^n$ be non-empty and bounded with $\mathcal{S} \neq \{\mathbf{0}\}$ and $\overline{\dim}_{\mathrm{B}}(\mathcal{S}) < k$, and fix $\beta \in \mathbb{R}$ such that

$$0 < \beta < 1 - \frac{\overline{\dim}_{\mathrm{B}}(\mathcal{S})}{k}. \tag{44}$$

Then, for a.a. $\boldsymbol{A} \in \mathbb{R}^{k\times(n-\ell)}$, we have

$$\inf_{\boldsymbol{x}\,\in\,\mathcal{S}\setminus\{\mathbf{0}\}} \frac{\|[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{x}\|}{\|\boldsymbol{x}\|^{1/\beta}} > 0. \tag{45}$$

Proof: As in the proof of Proposition 1, we show that for the random matrix $\mathbf{A} = [\mathbf{a}_1 \ldots \mathbf{a}_k]^{\mathrm{T}}$, with the $\mathbf{a}_i$ i.i.d. uniform on $B^{n-\ell}(\mathbf{0}, r)$ with arbitrary $r > 0$, (45) holds w.p. 1. Since r can be taken arbitrarily large, this establishes that the Lebesgue measure of matrices $\boldsymbol{A}$ for which (45) does not hold is zero. As (45) is an inequality of the form (40) we can apply Lemma 3 with $f(\boldsymbol{x}) = [\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}$ to conclude that showing

$$\mathbb{P}[\exists J \in \mathbb{N} : \|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| \geqslant \delta_j, \quad \text{for all } j \geqslant J \text{ and all } \boldsymbol{x} \in \mathcal{S}_j] = 1,$$

with $\delta_j := 2^{-j}$ and $\mathcal{S}_j := \mathcal{S}\setminus B^n(\mathbf{0}, \delta_j^{\beta})$, establishes the proof. Applying the Borel-Cantelli Lemma [24, Thm. 2.3.1] to the complementary events it follows that it suffices to show that

$$\sum_{j=0}^{\infty} \mathbb{P}[\exists \boldsymbol{x} \in \mathcal{S}_j : \|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| < \delta_j] < \infty. \tag{46}$$

The basic idea for establishing (46) is to cover $\mathcal{S}_j$ with balls of radius δ_j and to upper-bound the probabilities in (46) by probabilities that are in terms of the corresponding covering ball centers. Specifically, with the minimum number of balls of radius δ_j needed to cover $\mathcal{S}_j$ denoted by $M_j := M_{\mathcal{S}_j}(\delta_j)$ and the corresponding ball centers $\boldsymbol{x}_1^{(j)}, \ldots, \boldsymbol{x}_{M_j}^{(j)} \in \mathcal{S}_j$, we establish that

$$\mathbb{P}[\exists \boldsymbol{x} \in \mathcal{S}_j : \|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| < \delta_j] \tag{47}$$

$$\leqslant \sum_{i=1}^{M_j} \mathbb{P}[\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_i^{(j)}\| < (L+1)\delta_j]. \tag{48}$$

Here, $L := c(k, r, \|\boldsymbol{B}\|)$ is the constant in (18). To prove (47)–(48), first note that the existence of an $\boldsymbol{x} \in \mathcal{S}_j$ such that $\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| < \delta_j$ implies $\boldsymbol{x} \in B^n(\boldsymbol{x}_{i_0}^{(j)}, \delta_j)$ for some $i_0 \in \{1, ..., M_j\}$, since the balls $B^n(\boldsymbol{x}_i^{(j)}, \delta_j)$, $i = 1, ..., M_j$, cover $\mathcal{S}_j$. It then follows that

$$\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_{i_0}^{(j)}\| \leqslant \|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| + \|[\mathbf{A}\ \boldsymbol{B}](\boldsymbol{x}_{i_0}^{(j)} - \boldsymbol{x})\| \tag{49}$$

$$< \delta_j + L\delta_j = (L+1)\delta_j, \tag{50}$$

where we used $\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| < \delta_j$ and (18). From (49), (50), and a union bound argument we then get (47)–(48). We now turn to bounding the terms in the sum of (48) and will then use these bounds in (46) to establish the final result. Let us start by writing the covering ball centers as

$$\boldsymbol{x}_i^{(j)} = \begin{bmatrix} \boldsymbol{y}_i^{(j)} \\ \boldsymbol{z}_i^{(j)} \end{bmatrix},$$

with $\boldsymbol{y}_i^{(j)} \in \mathbb{R}^{n-\ell}$ and $\boldsymbol{z}_i^{(j)} \in \mathbb{R}^{\ell}$, and splitting them into two groups according to

$$\{\boldsymbol{x}_1^{(j)}, \ldots, \boldsymbol{x}_{M_j}^{(j)}\} = \mathcal{X}_1^{(j)} \cup \mathcal{X}_2^{(j)},$$

where

$$\mathcal{X}_1^{(j)} := \{\boldsymbol{x}_i^{(j)} : \|\boldsymbol{y}_i^{(j)}\| < c\|\boldsymbol{z}_i^{(j)}\|\} \tag{51}$$

$$\mathcal{X}_2^{(j)} := \{\boldsymbol{x}_i^{(j)} : \|\boldsymbol{y}_i^{(j)}\| \geqslant c\|\boldsymbol{z}_i^{(j)}\|\}, \tag{52}$$

with the constant $c > 0$ chosen below. The reasoning behind this splitting is as follows. For ball centers in $\mathcal{X}_1^{(j)}$, we establish that the corresponding probabilities of (48) equal zero for sufficiently large j, whereas for ball centers in $\mathcal{X}_2^{(j)}$, we use the concentration inequality in Lemma 1 to establish that the corresponding terms in the sum in (48) are sufficiently small to result in a finite upper bound as required in (46). We first note that for all ball centers

$$\|\boldsymbol{x}_i^{(j)}\|^2 = \|\boldsymbol{y}_i^{(j)}\|^2 + \|\boldsymbol{z}_i^{(j)}\|^2 \geqslant \delta_j^{2\beta},$$

by virtue of $\boldsymbol{x}_i^{(j)} \in \mathcal{S}_j$. We now turn to the set $\mathcal{X}_1^{(j)}$. From (51) we get

$$(c^2+1)\|\boldsymbol{z}_i^{(j)}\|^2 \geqslant \|\boldsymbol{y}_i^{(j)}\|^2 + \|\boldsymbol{z}_i^{(j)}\|^2 \geqslant \delta_j^{2\beta}. \tag{53}$$

This allows us to deduce that

$$\begin{aligned} \|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_i^{(j)}\| &\geqslant \|\boldsymbol{B}\boldsymbol{z}_i^{(j)}\| - \|\mathbf{A}\boldsymbol{y}_i^{(j)}\| && (54)\\ &\geqslant C_{\boldsymbol{B}}\|\boldsymbol{z}_i^{(j)}\| - \|\mathbf{A}\|\|\boldsymbol{y}_i^{(j)}\| && (55)\\ &\geqslant C_{\boldsymbol{B}}\|\boldsymbol{z}_i^{(j)}\| - c\|\mathbf{A}\|\|\boldsymbol{z}_i^{(j)}\| && (56)\\ &\geqslant \frac{C_{\boldsymbol{B}} - cL}{\sqrt{1+c^2}}\delta_j^{\beta}, && (57)\end{aligned}$$

where in (54) we applied the reverse triangle inequality, for (55) we note that there exists a constant $C_{\boldsymbol{B}} > 0$ such that $\|\boldsymbol{B}\boldsymbol{z}\| \geqslant C_{\boldsymbol{B}}\|\boldsymbol{z}\|$, for all $\boldsymbol{z} \in \mathbb{R}^{\ell}$, as a consequence of $\boldsymbol{B}$ being full-rank, in (56) we used $\boldsymbol{x}_i^{(j)} \in \mathcal{X}_1^{(j)}$, and in (57) we employed (53) and $\|\mathbf{A}\| \leqslant \|[\mathbf{A}\ \boldsymbol{B}]\| \leqslant L$, where L was defined right after (48). Since $\beta \in (0,1)$, δ_j^{β} can be made arbitrarily large relative to δ_j (i.e., $\delta_j^{\beta}/\delta_j = 2^{j(1-\beta)}$ can be made arbitrarily large) by taking j sufficiently large. Specifically, choosing[6] $c > 0$ to ensure $C_{\boldsymbol{B}} - cL > 0$, we can find a $J_1 \in \mathbb{N}$ such that

$$\frac{C_{\boldsymbol{B}} - cL}{\sqrt{1+c^2}}\delta_j^{\beta} \geqslant (L+1)\delta_j, \quad \text{for all } j \geqslant J_1. \tag{58}$$

By (57) this implies

$$\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_i^{(j)}\| \geqslant (L+1)\delta_j,$$

for all $j \geqslant J_1$, and hence establishes that

$$\mathbb{P}[\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_i^{(j)}\| < (L+1)\delta_j] = 0, \tag{59}$$

for $\boldsymbol{x}_i^{(j)} \in \mathcal{X}_1^{(j)}$ and $j \geqslant J_1$.

Next, consider $\boldsymbol{x}_i^{(j)} \in \mathcal{X}_2^{(j)}$. From (52) we get

$$\left(1 + \frac{1}{c^2}\right)\|\boldsymbol{y}_i^{(j)}\|^2 \geqslant \|\boldsymbol{y}_i^{(j)}\|^2 + \|\boldsymbol{z}_i^{(j)}\|^2 \geqslant \delta_j^{2\beta},$$

which, using the concentration inequality in Lemma 1, allows us to conclude that

$$\begin{aligned} &\mathbb{P}[\|\mathbf{A}\boldsymbol{y}_i^{(j)} + \boldsymbol{B}\boldsymbol{z}_i^{(j)}\| < (L+1)\delta_j] \\ &\leqslant C(n,k,r)\frac{(L+1)^k\delta_j^k}{\|\boldsymbol{y}_i^{(j)}\|^k} \\ &\leqslant C(n,k,r,L)\left(\sqrt{1+\frac{1}{c^2}}\right)^k \frac{2^{-jk}}{2^{-\beta jk}}. && (60)\end{aligned}$$

[6]This is possible since $C_{\boldsymbol{B}}/L > 0$.

Putting things together, we obtain

$$\begin{aligned} &\sum_{j=0}^{\infty}\mathbb{P}[\exists \boldsymbol{x} \in \mathcal{S}_j : \|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| < \delta_j] \\ &\leqslant J_1 + \sum_{j=J_1}^{\infty}\mathbb{P}[\exists \boldsymbol{x} \in \mathcal{S}_j : \|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}\| < \delta_j] && (61)\\ &= J_1 + \sum_{j=J_1}^{\infty}\sum_{i=1}^{M_j}\mathbb{P}[\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_i^{(j)}\| < (L+1)\delta_j] && (62)\\ &= J_1 + \sum_{j=J_1}^{\infty}\sum_{\boldsymbol{x}_i^{(j)} \in \mathcal{X}_2^{(j)}}\mathbb{P}[\|[\mathbf{A}\ \boldsymbol{B}]\boldsymbol{x}_i^{(j)}\| < (L+1)\delta_j] && (63)\\ &\leqslant J_1 + \sum_{j=J_1}^{\infty}\sum_{\boldsymbol{x}_i^{(j)} \in \mathcal{X}_2^{(j)}} C(n,k,r,L)\left(\sqrt{1+\frac{1}{c^2}}\right)^k \frac{2^{-jk}}{2^{-\beta jk}} && (64)\\ &\leqslant J_1 + C(n,k,r,L,c)\sum_{j=J_1}^{\infty} M_j 2^{-jk(1-\beta)} && (65)\\ &\leqslant J_1 + C(n,k,r,L,c,\mathcal{S})\sum_{j=J_1}^{\infty} 2^{jd'}2^{-jk(1-\beta)} && (66)\\ &= J_1 + C(n,k,r,L,c,\mathcal{S})\sum_{j=J_1}^{\infty} 2^{-jk\left(1-\frac{d'}{k}-\beta\right)} \\ &< \infty. && (67)\end{aligned}$$

Here, in (61) we upper-bounded the probability of the terms for $j < J_1$ by 1, where J_1 was defined in (58), (62) is by (47)–(48), (63) follows from (59), in (64) we invoked (60), and (65) holds since $|\mathcal{X}_2^{(j)}| \leqslant M_j$. For (66), we set $d' = \overline{\dim}_{\mathrm{B}}(\mathcal{S}) + \alpha$ with $\alpha > 0$ small enough so that $1 - \frac{d'}{k} > \beta$, which is possible by (44), and we used

$$\begin{aligned} M_j = M_{\mathcal{S}_j}(\delta_j) &\leqslant N_{\mathcal{S}_j}(\delta_j/2) && (68)\\ &\leqslant N_{\mathcal{S}}(\delta_j/2) && (69)\\ &\leqslant C(\mathcal{S})\delta_j^{-d'}, && (70)\end{aligned}$$

for all $j \in \mathbb{N}$, where (68) follows from a triangle inequality argument (cf. (97)), (69) holds as[7] $\mathcal{S}_j \subseteq \mathcal{S}$, for all j, and (70) is a consequence of

i) $\overline{\dim}_{\mathrm{B}}(\mathcal{S}) < d'$ and thus $N_{\mathcal{S}}(\delta_j/2) \leqslant (\delta_j/2)^{-d'} = 2^{d'}\delta_j^{-d'}$ for sufficiently large j by definition of $\limsup$, and
ii) $C(\mathcal{S})$ taken sufficiently large so that (70) also holds for the (finite number of) j's for which $N_{\mathcal{S}}(\delta_j/2) \leqslant (\delta_j/2)^{-d'}$ does not hold. Note that $C(\mathcal{S})$ is guaranteed to be finite as the set $\mathcal{S}$ is bounded and therefore the covering numbers $N_{\mathcal{S}}(\delta_j/2)$ are finite for all j.

Finally, (67) follows from

$$1 - \frac{d'}{k} > \beta,$$

which is by choice of d'. This completes the proof. ■

[7]Note that we resort to the original covering number $N_\delta(\mathcal{A})$ in this argument, as for the modified covering number $M_\delta(\mathcal{A})$ the relation $\mathcal{A} \subseteq \mathcal{B}$ does not imply, in general, that $M_\delta(\mathcal{A}) \leqslant M_\delta(\mathcal{B})$.

VI. Proof of Theorem 2

We start with preparatory material. Since, by assumption, $\beta > 0$ is fixed and satisfies $1 - \frac{\overline{R}_{\mathrm{B}}(\varepsilon)}{R} > \beta$, we can find an $\alpha > 0$ such that

$$1 - \frac{\overline{R}_{\mathrm{B}}(\varepsilon) + \alpha}{R} > \beta. \tag{71}$$

Let $k' := (\overline{R}_{\mathrm{B}}(\varepsilon) + \alpha)n$. By definition of $\overline{R}_{\mathrm{B}}(\varepsilon)$, we can find a sequence of non-empty compact[8] sets $\mathcal{U} \subseteq \mathbb{R}^n$ such that

$$\overline{\dim}_{\mathrm{B}}(\mathcal{U}) \mathrel{\dot{<}} k' \tag{72}$$
$$\text{and} \quad \mathbb{P}[\mathsf{x} \in \mathcal{U}] \geqslant 1 - \varepsilon.$$

Moreover, we have

$$1 - \frac{k'}{k} = \frac{\lfloor Rn \rfloor - (\overline{R}_{\mathrm{B}}(\varepsilon) + \alpha)n}{\lfloor Rn \rfloor} \tag{73}$$
$$\mathrel{\dot{\geqslant}} \frac{\lfloor Rn \rfloor - (\overline{R}_{\mathrm{B}}(\varepsilon) + \alpha)n}{Rn} \tag{74}$$
$$> \frac{R - \frac{1}{n} - \overline{R}_{\mathrm{B}}(\varepsilon) - \alpha}{R} \tag{75}$$
$$\mathrel{\dot{>}} \beta, \tag{76}$$

where (74) follows from $\lfloor Rn \rfloor \leqslant Rn$ and $\lfloor Rn \rfloor - (\overline{R}_{\mathrm{B}}(\varepsilon) + \alpha)n > (R - \frac{1}{n})n - (\overline{R}_{\mathrm{B}}(\varepsilon) + \alpha)n \mathrel{\dot{>}} 0$ (since $R > \overline{R}_{\mathrm{B}}(\varepsilon) + \alpha$, by choice of α), in (75) we used $Rn - 1 < \lfloor Rn \rfloor$, and in (76) we invoked (71). In the remainder of the proof, we take n sufficiently large for (72)–(76) to hold in the #-sense and drop the dot-notation. For $\boldsymbol{B}$ as in the statement of the theorem, $\boldsymbol{A} \in \mathbb{R}^{k \times (n-\ell)}$, and $\boldsymbol{x} \in \mathbb{R}^n$, we set[9]

$$\mathcal{A} := \left\{ (\boldsymbol{A}, \boldsymbol{x}) : \inf_{\boldsymbol{u} \in \mathcal{U}_{\boldsymbol{x}} \setminus \{\boldsymbol{0}\}} \frac{\|[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}\|}{\|\boldsymbol{u}\|^{1/\beta}} = 0 \right\},$$

with

$$\mathcal{U}_{\boldsymbol{x}} = \{\boldsymbol{u} - \boldsymbol{x} : \boldsymbol{u} \in \mathcal{U}\}.$$

Since (upper) Minkowski dimension is invariant under translation (as seen by translating covering balls accordingly), we have $\overline{\dim}_{\mathrm{B}}(\mathcal{U}_{\boldsymbol{x}}) = \overline{\dim}_{\mathrm{B}}(\mathcal{U})$ which, together with (72) and (73)–(76), implies $1 - \frac{\overline{\dim}_{\mathrm{B}}(\mathcal{U}_{\boldsymbol{x}})}{k} > 1 - \frac{k'}{k} > \beta$ for all $\boldsymbol{x} \in \mathbb{R}^n$. We can therefore apply the regularized probabilistic uncertainty relation, Proposition 2, to each $\mathcal{U}_{\boldsymbol{x}}$ with $\boldsymbol{x} \in \mathbb{R}^n$ and get

$$\mathrm{Leb}^{k(n-\ell)}\{\boldsymbol{A} : (\boldsymbol{A}, \boldsymbol{x}) \in \mathcal{A}\} = 0, \tag{77}$$

for all $\boldsymbol{x} \in \mathbb{R}^n$. Integrating (77) w.r.t. $\mu_{\mathsf{x}}(\mathrm{d}\boldsymbol{x})$ yields

$$\int_{\mathbb{R}^n} \mathrm{Leb}^{k(n-\ell)}\{\boldsymbol{A} : (\boldsymbol{A}, \boldsymbol{x}) \in \mathcal{A}\} \mu_{\mathsf{x}}(\mathrm{d}\boldsymbol{x}) = 0. \tag{78}$$

We next show that $\mathcal{A}$ is measurable, which will allow us to change the order of integration in (78) by applying Fubini's theorem [23, Thm. 1.14] to the indicator function on $\mathcal{A}$. This will be accomplished by showing that the sets

$$\mathcal{A}_j := \left\{ (\boldsymbol{A}, \boldsymbol{x}) : \inf_{\boldsymbol{u} \in \mathcal{U}_{\boldsymbol{x}} \setminus \{\boldsymbol{0}\}} \frac{\|[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}\|}{\|\boldsymbol{u}\|^{1/\beta}} > \frac{1}{j} \right\}$$

are measurable for all $j \in \mathbb{N}$ and using[10]

$$\mathcal{A}^{\mathrm{c}} = \bigcup_{j \in \mathbb{N}} \mathcal{A}_j, \tag{79}$$

where $\mathcal{A}^{\mathrm{c}}$ denotes the complement of $\mathcal{A}$ in $\mathbb{R}^{k \times (n-\ell)} \times \mathbb{R}^n$. Indeed, for each $j \in \mathbb{N}$, we can write

$$\begin{aligned}\mathcal{A}_j &= \left\{ (\boldsymbol{A}, \boldsymbol{x}) : j\|[\boldsymbol{A}\ \boldsymbol{B}]\boldsymbol{u}\| \geq \|\boldsymbol{u}\|^{1/\beta},\ \text{for all } \boldsymbol{u} \in \mathcal{U}_{\boldsymbol{x}} \right\} \\ &= \big\{ (\boldsymbol{A}, \boldsymbol{x}) : j\|[\boldsymbol{A}\ \boldsymbol{B}](\boldsymbol{u} - \boldsymbol{x})\| \geq \|\boldsymbol{u} - \boldsymbol{x}\|^{1/\beta}, \\ &\qquad \text{for all } \boldsymbol{u} \in \mathcal{U} \big\} \\ &= \big\{ (\boldsymbol{A}, \boldsymbol{x}) : \inf_{\boldsymbol{u} \in \mathcal{U}} \big(j\|[\boldsymbol{A}\ \boldsymbol{B}](\boldsymbol{u} - \boldsymbol{x})\| - \|\boldsymbol{u} - \boldsymbol{x}\|^{1/\beta} \big) \\ &\qquad \geq 0 \big\},\end{aligned} \tag{80}$$

where (80) is measurable as a consequence of [21, Prop. 14.40], upon noting that $\mathcal{U}$ is compact and

$$h(\boldsymbol{A}, \boldsymbol{x}, \boldsymbol{u}) := j\|[\boldsymbol{A}\ \boldsymbol{B}](\boldsymbol{u} - \boldsymbol{x})\| - \|\boldsymbol{u} - \boldsymbol{x}\|^{1/\beta}$$

is a continuous mapping. Fubini's theorem therefore yields

$$\int_{\mathbb{R}^n} \mathrm{Leb}^{k(n-\ell)}\{\boldsymbol{A} : (\boldsymbol{A}, \boldsymbol{x}) \in \mathcal{A}\} \mu_{\mathsf{x}}(\mathrm{d}\boldsymbol{x}) \tag{81}$$
$$= \int_{\mathbb{R}^{k \times (n-\ell)}} \mathbb{P}[(\boldsymbol{A}, \mathsf{x}) \in \mathcal{A}] \mathrm{d}\boldsymbol{A}. \tag{82}$$

Combining (78) with (81)–(82), we conclude that

$$\mathbb{P}[(\boldsymbol{A}, \mathsf{x}) \in \mathcal{A}] = 0, \quad \text{for a.a. } \boldsymbol{A},$$

which is equivalent to

$$\mathbb{P}[(\boldsymbol{A}, \mathsf{x}) \in \mathcal{A}^{\mathrm{c}}] = 1, \quad \text{for a.a. } \boldsymbol{A}. \tag{83}$$

Using (79), (83), and $\mathcal{A}_j \subseteq \mathcal{A}_{j+1}$, for all $j \in \mathbb{N}$, [25, Lem. 3.4, Part (a)] implies that

$$\lim_{j \to \infty} \mathbb{P}[(\boldsymbol{A}, \mathsf{x}) \in \mathcal{A}_j] = 1, \quad \text{for a.a. } \boldsymbol{A}.$$

Therefore, for every $\kappa > 0$, there exists a $J(\boldsymbol{A}) \in \mathbb{N}$ such that

$$\mathbb{P}\big[(\boldsymbol{A}, \mathsf{x}) \in \mathcal{A}_{J(\boldsymbol{A})}\big] \geqslant 1 - \kappa, \quad \text{for a.a. } \boldsymbol{A}.$$

Moreover, since $\mathbb{P}[\mathsf{x} \in \mathcal{U}] \geq 1 - \varepsilon$, a union bound argument yields

$$\mathbb{P}[\mathsf{x} \in \mathcal{U}_{\boldsymbol{A}}] \geqslant 1 - \kappa - \varepsilon, \quad \text{for a.a. } \boldsymbol{A},$$

where

$$\mathcal{U}_{\boldsymbol{A}} := \left\{ \boldsymbol{x} : (\boldsymbol{A}, \boldsymbol{x}) \in \mathcal{A}_{J(\boldsymbol{A})} \right\} \cap \mathcal{U}.$$

Since $\mathcal{U}_{\boldsymbol{A}} \subseteq \mathcal{A}_{J(\boldsymbol{A})}$, we can conclude that for a.a. $\boldsymbol{A}$ the following holds:

$$\|\boldsymbol{x}_1 - \boldsymbol{x}_2\|^{1/\beta} \leqslant J(\boldsymbol{A})\|[\boldsymbol{A}\ \boldsymbol{B}](\boldsymbol{x}_1 - \boldsymbol{x}_2)\|,$$

for all $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathcal{U}_{\boldsymbol{A}}$. By Lemma 2, $\boldsymbol{H} = [\boldsymbol{A}\ \boldsymbol{B}]$ is therefore injective on $\mathcal{U}_{\boldsymbol{A}}$ and its inverse $\boldsymbol{H}^{-1} \colon \boldsymbol{H}(\mathcal{U}_{\boldsymbol{A}}) \to \mathcal{U}_{\boldsymbol{A}}$ is β-Hölder continuous, and by [26, Thm. 1, ii)] (restated below for completeness) with $\mathcal{V} = \mathbb{R}^k$, $\mathcal{W} = \mathbb{R}^n$, and $\mathcal{B}(\boldsymbol{A}) = \boldsymbol{H}(\mathcal{U}_{\boldsymbol{A}})$ the inverse $\boldsymbol{H}^{-1}$ can be extended to a β-Hölder continuous mapping $g_{\boldsymbol{H}} \colon \mathbb{R}^k \to \mathbb{R}^n$. Again, this statement holds for a.a. $\boldsymbol{A}$. Finally, thanks to injectivity of $\boldsymbol{H}$ on $\mathcal{U}_{\boldsymbol{A}}$, for a.a. $\boldsymbol{A}$, we

[8] Since upper Minkowski dimension is invariant under set closure ([17, Prop. 3.4]), we can assume, w.l.o.g., that $\mathcal{U}$ is compact.

[9] We use the convention $\inf(\emptyset) = \infty$.

[10] Complements and countable unions of measurable sets are again measurable.

have $g_{\boldsymbol{H}}(\boldsymbol{H}\boldsymbol{x}) = \boldsymbol{x}$ for all $\boldsymbol{x} \in \mathcal{U}_{\boldsymbol{A}}$ and for a.a. $\boldsymbol{A}$, and therefore

$$\mathbb{P}[g_{\boldsymbol{H}}(\boldsymbol{H}\mathsf{x}) \neq \mathsf{x}] \leqslant \mathbb{P}[\mathsf{x} \notin \mathcal{U}_{\boldsymbol{A}}] \leq \varepsilon + \kappa,$$

for a.a. $\boldsymbol{A}$. This completes the proof. ■

Finally, for the reader's convenience, we provide the following (reformulated) version of the statement [26, Thm. 1, ii)].

Theorem 3: Let $\mathcal{V}, \mathcal{W}$ be Euclidean spaces and let $g\colon \mathcal{B} \to \mathcal{W}$ be β-Hölder continuous with $0 < \beta < 1$ and $\mathcal{B} \subseteq \mathcal{V}$. Then, g can be extended to a β-Hölder continuous mapping on all of $\mathcal{V}$.

VII. To sparse signal separation

Converses for the achievability statements in Theorems 1 and 2 seem difficult to obtain for general sources. We can, however, build on [13, Thm. 15], which establishes a converse for the analog compression problem for sources of mixed discrete-continuous distribution, and derive a converse to Theorems 1 and 2 for mixed discrete-continuous sources. Mixed discrete-continuous sources are of particular interest as their Minkowski dimension effectively quantifies the number of non-zero entries and hence reflects the traditional sparsity notion used, e.g., in [3], [7], [12], [27], [28]. Specifically, we consider concatenated source vectors x with independent entries of mixed discrete-continuous distribution and possibly different mixture parameters for the constituent processes $(\mathsf{Y}_i)_{i\in\mathbb{N}}$ and $(\mathsf{Z}_i)_{i\in\mathbb{N}}$.

Definition 6: We say that x in Definition 1 has a mixed discrete-continuous distribution if for each $n \in \mathbb{N}$ the random variables X_i for $i \in \{1, \dots, n\}$ are independent and distributed according to

$$\mu_{\mathsf{X}_i} = \begin{cases} (1-\rho_1)\mu_{\mathrm{d}_1} + \rho_1\mu_{\mathrm{c}_1}, & i \in \{1, \dots, n-\ell\} \\ (1-\rho_2)\mu_{\mathrm{d}_2} + \rho_2\mu_{\mathrm{c}_2}, & i \in \{n-\ell+1, \dots, n\}, \end{cases}$$

where $0 \leqslant \rho_i \leqslant 1$ are mixture parameters, $\ell = \lfloor \lambda n \rfloor$, the μ_{d_i} are discrete distributions, and the μ_{c_i} are absolutely continuous (w.r.t. Lebesgue measure) distributions.

Before stating the converse, we extend—to concatenated source vectors—[13, Thm. 6], which shows that, indeed, the Minkowski dimension compression rate of mixed discrete-continuous sources reflects the traditional notion of sparsity. Specifically, if the discrete parts μ_{d_1}, μ_{d_2} are Dirac measures at 0, i.e., $\mu_{\mathrm{d}_1} = \mu_{\mathrm{d}_2} = \delta_0$, then the non-zero entries of x can be generated only by the continuous parts μ_{c_1}, μ_{c_2}. With

$$\widetilde{\mathsf{Y}}_i := \mathbb{1}_{\mathbb{R}\setminus\{0\}}(\mathsf{Y}_i), \quad i = 1, ..., n-\ell,$$
$$\widetilde{\mathsf{Z}}_i := \mathbb{1}_{\mathbb{R}\setminus\{0\}}(\mathsf{Z}_i), \quad i = n-\ell+1, ..., n,$$

the fraction of non-zero entries in x is given by

$$\frac{1}{n}\left(\sum_{i=1}^{n-\ell}\widetilde{\mathsf{Y}}_i + \sum_{i=n-\ell+1}^{n}\widetilde{\mathsf{Z}}_i\right). \tag{84}$$

Letting $n \to \infty$ in (84), we obtain

$$\begin{aligned} &\frac{1}{n}\left(\sum_{i=1}^{n-\ell}\widetilde{\mathsf{Y}}_i + \sum_{i=n-\ell+1}^{n}\widetilde{\mathsf{Z}}_i\right) \\ &= \frac{n-\ell}{n}\frac{1}{n-\ell}\sum_{i=1}^{n-\ell}\widetilde{\mathsf{Y}}_i + \frac{\ell}{n}\frac{1}{\ell}\sum_{i=n-\ell+1}^{n}\widetilde{\mathsf{Z}}_i \\ &\xrightarrow{\mathbb{P}} (1-\lambda)\rho_1 + \lambda\rho_2, \end{aligned}$$

where we used

$$\begin{aligned} \lim_{n\to\infty}(n-\ell)/n &= \lim_{n\to\infty}(n-\lfloor\lambda n\rfloor)/n \\ &= (1-\lambda), \end{aligned}$$

as $(1-\lambda)n \leqslant n - \lfloor\lambda n\rfloor < (1-\lambda)n + 1$. Similarly,

$$\begin{aligned} \lim_{n\to\infty}\ell/n &= \lim_{n\to\infty}\lfloor\lambda n\rfloor/n \\ &= \lambda, \end{aligned}$$

as $\lambda n - 1 < \lfloor\lambda n\rfloor \leqslant \lambda n$, and

$$\frac{1}{n-\ell}\sum_{j=1}^{n-\ell}\widetilde{\mathsf{Y}}_i \xrightarrow{\mathbb{P}} \rho_1$$
$$\frac{1}{\ell}\sum_{j=n-\ell+1}^{n}\widetilde{\mathsf{Z}}_i \xrightarrow{\mathbb{P}} \rho_2,$$

by the weak law of large numbers and $\mathbb{E}[\widetilde{\mathsf{Y}}_i] = \mathbb{P}[\widetilde{\mathsf{Y}}_i = 1] = \rho_1$, $\mathbb{E}[\widetilde{\mathsf{Z}}_i] = \mathbb{P}[\widetilde{\mathsf{Z}}_i = 1] = \rho_2$. This shows that the fraction of non-zero entries in x converges—in probability—to $(1-\lambda)\rho_1 + \lambda\rho_2$. The next result establishes that the Minkowski dimension compression rate $R_{\mathrm{B}}(\varepsilon)$ of mixed discrete-continuous sources equals, for all $\varepsilon \in (0,1)$, the asymptotic fraction of non-zero entries given by $(1-\lambda)\rho_1 + \lambda\rho_2$.

Proposition 3: Suppose that x is distributed according to Definition 6. Then, we have

$$R_{\mathrm{B}}(\varepsilon) = (1-\lambda)\rho_1 + \lambda\rho_2,$$

for all $\varepsilon \in (0,1)$.

Proof: The proof follows closely [13, Thm. 15] and is therefore not detailed here. Interested readers can, however, consult the Online Addendum to this paper [2, Sec. II] for the proof of [13, Thm. 15] adapted to our setting. ■

We are now ready to state the converse for measurable separators.

Proposition 4: Suppose that x is distributed according to Definition 6 and let $\varepsilon \in (0,1)$. Then, the existence of a measurement matrix $\boldsymbol{H} = [\boldsymbol{A}\ \boldsymbol{B}] : \mathbb{R}^{n-\ell} \times \mathbb{R}^{\ell} \to \mathbb{R}^k$ and a corresponding measurable separator $g\colon \mathbb{R}^k \to \mathbb{R}^{n-\ell} \times \mathbb{R}^{\ell}$, with $k = \lfloor Rn \rfloor$, such that

$$\mathbb{P}[g([\boldsymbol{A}\ \boldsymbol{B}]\mathsf{x}) \neq \mathsf{x}] \;\dot{\leqslant}\; \varepsilon, \tag{85}$$

imply $R \geqslant R_{\mathrm{B}}(\varepsilon)$.

Proof: The proof does not have to account for the fact that $\boldsymbol{H} = [\boldsymbol{A}\ \boldsymbol{B}]$ contains a fixed block $\boldsymbol{B}$ and follows closely the converse part of [13, Thm. 6]. We therefore do not include the details here, but, again, refer the interested reader to the Online Addendum [2, Sec. III]. ■

Combining the achievability statements in Theorems 1 and 2, and Propositions 3 and 4, we can conclude that, for mixed discrete-continuous sources,

$$R_{\mathrm{B}}(\varepsilon) = (1-\lambda)\rho_1 + \lambda\rho_2,$$

is the critical rate in the following sense:

- For $R > R_{\mathrm{B}}(\varepsilon)$, for every fixed full-rank matrix $\boldsymbol{B} \in \mathbb{R}^{k\times\ell}$, with $k \geqslant \ell$, and for a.a. $\boldsymbol{A}$ (where the set of exceptions for $\boldsymbol{A}$ depends on the specific choice of $\boldsymbol{B}$), there exists a measurable separator g satisfying (85), as well as a β-Hölder continuous separator g for fixed β with $\beta < 1 - \frac{\overline{R}_{\mathrm{B}}(\varepsilon)}{R}$ satisfying (85) with ε replaced by $\varepsilon + \kappa$ for arbitrarily small $\kappa > 0$,
- for $R = R_{\mathrm{B}}(\varepsilon)$, we cannot make a general statement on the existence of a separator,
- and for $R < R_{\mathrm{B}}(\varepsilon)$ there does not exist a single pair $(g, [\boldsymbol{A}\ \boldsymbol{B}])$, with g measurable, satisfying (85).

As $R \approx k/n$, where n is the ambient dimension and k the number of measurements, the threshold $R_{\mathrm{B}}(\varepsilon) = (1-\lambda)\rho_1 + \lambda\rho_2$ identifies the critical number of measurements relative to the ambient dimension as the number of non-zero entries in $\mathbf{x}$.

Relation to classical uncertainty relations in compressed sensing. Comparing the threshold obtained from our probabilistic uncertainty relations to the thresholds available in the compressed sensing literature, we note the following. The Donoho-Stark [12] and Elad-Bruckstein [15] uncertainty principles hold for *all* $\boldsymbol{y}$ and $\boldsymbol{z}$, but suffer from the "square-root bottleneck" [8]. It is well known that the inequalities leading to the square-root bottleneck are saturated by very special combinations of signals and dictionaries, e.g., a Dirac comb for $\boldsymbol{A} = \boldsymbol{I}_n$ and $\boldsymbol{B} = \boldsymbol{F}_n$ [12]. Relaxing these deterministic thresholds by considering random models for the signals and dictionaries [10], [27]–[29] leads to thresholds that exhibit a "$\log n$-factor". The thresholds that follow from our probabilistic uncertainty relations exclude an arbitrarily small set of signals, a set of $\boldsymbol{A}$-matrices of Lebesgue measure zero, are asymptotic in n, and suffer neither from the square-root bottleneck nor from the $\log n$-factor. Moreover, they are best possible as the same threshold would be obtained if the support sets of $\boldsymbol{y}$ and $\boldsymbol{z}$ were known a priori and only the values of the non-zero entries in the concatenated source vector were to be recovered. The set of exceptions for $\boldsymbol{A}$ in the "a.a.-statement" in Theorems 1 and 2 depending on the specific choice of $\boldsymbol{B}$ can be interpreted as a mild incoherence condition between $\boldsymbol{A}$ and $\boldsymbol{B}$ akin to those in [12], [15]. In fact, we have a phase transition phenomenon, which states that above the critical rate $R_{\mathrm{B}}(\varepsilon)$ a.a. matrices $\boldsymbol{A}$ are "incoherent" to a given matrix $\boldsymbol{B}$, whereas below the critical rate there is not a single pair of matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ that admits separation via a measurable separator g. Finally, as already noted, our regularized probabilistic uncertainty relation does not seem to have a counterpart in classical compressed sensing theory.

Remark 6: The results above show that for mixed discrete-continuous sources $\mathbf{x}$ the Minkowski dimension compression rate is small if the asymptotic fraction $(1-\lambda)\rho_1 + \lambda\rho_2$ of non-zero entries in $\mathbf{x}$ is small, i.e., if the source vectors are sparse in the classical sense. Another factor that can lead to small Minkowski dimension compression rate is statistical dependence between the constituents $\mathbf{y}$ and $\mathbf{z}$. For example, consider the declipping problem [3] where a signal that is sparse in the dictionary $\boldsymbol{A}$ is to be recovered from its clipped version. Specifically, we observe $\boldsymbol{A}\mathbf{y} + \mathbf{z}$ with $\mathbf{z} = g_a(\boldsymbol{A}\mathbf{y}) - \boldsymbol{A}\mathbf{y}$, where g_a denotes entry-wise clipping to the values $\pm a$. If clipping is not too aggressive, the signal $\mathbf{z}$ will be sparse in the identity basis $\boldsymbol{B} = \boldsymbol{I}_\ell$ (see Fig. 2). Here $\mathbf{y}$ and $\mathbf{z}$ are of the same dimension, i.e., $\lambda = 1/2$. Moreover, $\mathbf{z}$ is completely determined by $\mathbf{y}$, which, as proved in Lemma 4 in Appendix A, implies that

$$R_{\mathrm{B}}^{\mathbf{x}}(\varepsilon) = \frac{1}{2} R_{\mathrm{B}}^{\mathbf{y}}(\varepsilon),$$

where $R_{\mathrm{B}}^{\mathbf{x}}(\varepsilon)$ is the Minkowski dimension compression rate of $\mathbf{x} = [\mathbf{y}^{\mathrm{T}}\ \mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}$ and $R_{\mathrm{B}}^{\mathbf{y}}(\varepsilon)$ is the Minkowski dimension compression rate of $\mathbf{y}$ only. If the components of $\mathbf{y}$ are i.i.d. and of discrete-continuous mixture $(1-\rho)\mu_{\mathrm{d}} + \rho\mu_{\mathrm{c}}$, it follows from Proposition 3 that $R_{\mathrm{B}}^{\mathbf{y}}(\varepsilon) = \rho$ and consequently

$$R_{\mathrm{B}}^{\mathbf{x}}(\varepsilon) = \frac{1}{2}\rho.$$

The description complexity of $\mathbf{x}$ is therefore determined by the fraction of components in $\mathbf{y}$ that are continuously distributed. As expected, we find that the critical rate here is half the critical rate for a mixed discrete-continuous source with independent $\mathbf{y}$ and $\mathbf{z}$ and $\rho_1 = \rho_2 = \rho$.

VIII. Strengthening of [13, Thm. 18] and simplifying its proof

In this section, we sketch how the probabilistic uncertainty relation, Proposition 1, and the regularized probabilistic uncertainty relation, Proposition 2, can be applied to devise a simplification of the proof of [13, Thm. 18, 1)] and a significant strengthening of the statement [13, Thm. 18, 2)]. We begin by restating [13, Thm. 18, 1)] in our notation and terminology.

Theorem 4 ([13, Thm. 18, 1)]): Let $\mathbf{x} = [\mathsf{X}_1 \dots \mathsf{X}_n]^{\mathrm{T}}$ be a source vector of dimension n with underlying stochastic source process $(\mathsf{X}_i)_{i\in\mathbb{N}}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$. Take $\varepsilon > 0$ and let $R > \overline{R}_{\mathrm{B}}(\varepsilon)$. Then, for a.a. $\boldsymbol{H} \in \mathbb{R}^{k\times n}$, there exists a measurable decoder g such that

$$\mathbb{P}[g(\boldsymbol{H}\mathbf{x}) \neq \mathbf{x}] \mathrel{\dot{\leqslant}} \varepsilon,$$

where $k = \lfloor Rn \rfloor$.

This statement can be recovered from Theorem 1 by setting $\lambda = 0$, which, in fact, yields a slight improvement upon [13, Thm. 18, 1)], namely the condition $R > \overline{R}_{\mathrm{B}}(\varepsilon)$ in [13, Thm. 18, 1)] is replaced by $R > \underline{R}_{\mathrm{B}}(\varepsilon)$. For our simplified proof, we start by particularizing the probabilistic uncertainty relation, Proposition 1, to $\ell = 0$.

Corollary 1: Let $\mathcal{S} \subseteq \mathbb{R}^n$ be non-empty and bounded such that $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) < k$. Then, we have

$$\{\boldsymbol{x} \in \mathcal{S}\backslash\{\boldsymbol{0}\} : \boldsymbol{H}\boldsymbol{x} = \boldsymbol{0}\} = \emptyset,$$

for a.a. $\boldsymbol{H} \in \mathbb{R}^{k\times n}$.

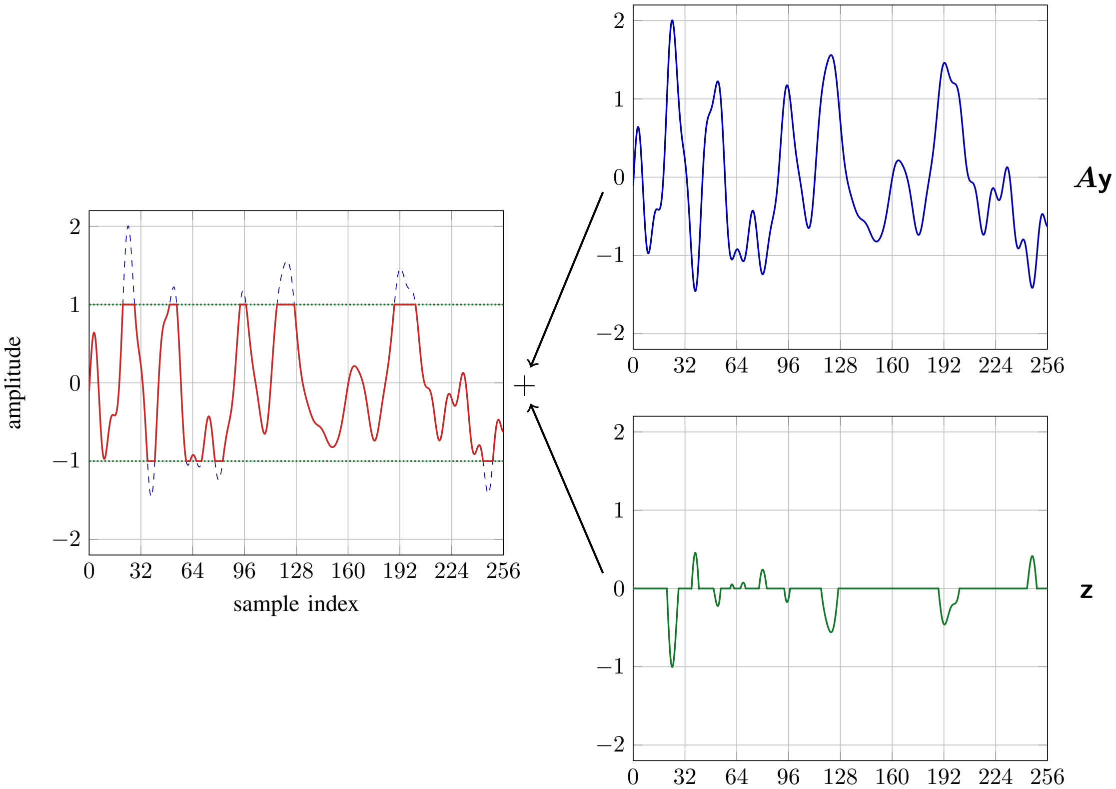

Fig. 2: Declipping of signals as a sparse signal separation problem.

We refer to this result as a probabilistic null-space property as it is a statement on the intersection of the null-space of $\boldsymbol{H}$ with the set $\mathcal{S}$. Our alternative, simplified proof of [13, Thm. 18, 1)] goes as follows. As in the proof of Theorem 1 we choose a sequence of compact sets $\mathcal{U} \subseteq \mathbb{R}^n$ satisfying (29) and (30). Let $\boldsymbol{e} \in \mathbb{R}^n \setminus \mathcal{U}$. Again, it follows from [21, Prop. 14.33 and Cor. 14.6] and the compactness of $\mathcal{U}$ that there exists a measurable mapping $f \colon \mathbb{R}^{k\times n} \times \mathbb{R}^k \to \mathbb{R}^n$ satisfying

$$f(\boldsymbol{H}, \boldsymbol{v}) \in \{\boldsymbol{u} \in \mathcal{U} : \boldsymbol{H}\boldsymbol{u} = \boldsymbol{v}\}, \quad \text{if } \{\boldsymbol{u} \in \mathcal{U} : \boldsymbol{H}\boldsymbol{u} = \boldsymbol{v}\} \neq \emptyset,$$

and $f(\boldsymbol{A}, \boldsymbol{v}) = \boldsymbol{e}$ else. The mapping $g = f(\boldsymbol{H}, \cdot)$ therefore constitutes a valid (i.e., measurable) decoder. Let

$$p_{\mathrm{e}}(\boldsymbol{H}) := \mathbb{P}[g(\boldsymbol{H}\mathbf{x}) \neq \mathbf{x}], \quad \boldsymbol{H} \in \mathbb{R}^{k\times n}.$$

Repeating the steps in (31)–(36) with $\boldsymbol{H}$ in place of $[\boldsymbol{A}\ \boldsymbol{B}]$, it follows that $p_{\mathrm{e}}(\boldsymbol{H}) \leq \varepsilon$ for a.a. $\boldsymbol{H} \in \mathbb{R}^{k\times n}$ provided we can show that

$$\mathrm{Leb}^{kn}\{\boldsymbol{H} : \ker(\boldsymbol{H}) \cap \mathcal{U}_{\boldsymbol{x}} \neq \{\mathbf{0}\}\} = 0, \tag{86}$$

for all $\boldsymbol{x} \in \mathcal{U}$, where

$$\mathcal{U}_{\boldsymbol{x}} = \{\boldsymbol{u} - \boldsymbol{x} : \boldsymbol{u} \in \mathcal{U}\}, \quad \boldsymbol{x} \in \mathcal{U}.$$

Applying Corollary 1 with $\mathcal{S} = \mathcal{U}_{\boldsymbol{x}}$, (86) holds as a consequence of $\underline{\dim}_{\mathrm{B}}(\mathcal{U}_{\boldsymbol{x}}) < k$, thereby completing the proof. The application of the probabilistic null-space property, Corollary 1, replaces the arguments in [13, Thm. 18, 1)] that are based on properties of invariant measures on Grassmannian manifolds. Finally, we note that, instead of particularizing Proposition 1 to $\ell = 0$, the probabilistic null-space property in Corollary 1 can also be proved directly with considerably less logistic effort, as done in [1].

Next, we restate [13, Thm. 18, 2)] in our notation and terminology.

Theorem ([13, Thm. 18, 2)]): Let $\mathbf{x} = [\mathsf{X}_1 \ldots \mathsf{X}_n]^{\mathrm{T}}$ be a source vector of dimension n with underlying stochastic process $(\mathsf{X}_i)_{i\in\mathbb{N}}$ on $(\mathbb{R}^{\mathbb{N}}, \mathcal{B}^{\otimes\mathbb{N}})$. Take $\varepsilon > 0$, and let $R > \overline{R}_{\mathrm{B}}(\varepsilon)$ and $\beta > 0$ be fixed such that

$$\beta < 1 - \frac{\overline{R}_{\mathrm{B}}(\varepsilon)}{R}.$$

Then, there exist $\boldsymbol{H} \in \mathbb{R}^{k\times n}$ and a corresponding β-Hölder continuous decoder g such that

$$\mathbb{P}[g(\boldsymbol{H}\mathbf{x}) \neq \mathbf{x}] \dot{\leqslant} \varepsilon + \kappa, \tag{87}$$

where $k = \lfloor Rn \rfloor$ and $\kappa > 0$ can be chosen arbitrarily small.

Particularizing Theorem 2 to $\lambda = 0$, we obtain a substantial strengthening of [13, Thm. 18, 2)], as [13, Thm. 18, 2)] states the existence of an $\boldsymbol{H}$ with a corresponding g satisfying (87), whereas our result says that for *a.a.* $\boldsymbol{H}$ there is a corresponding g satisfying (87). The crucial element in accomplishing this strengthening is the regularized probabilistic uncertainty relation in Proposition 2.

Appendix A
Minkowski Dimension Compression Rate for Declipping Example

We consider the declipping problem where $\lambda = 1/2$ and we observe $\mathbf{z} = g_a(\boldsymbol{A}\mathbf{y}) - \boldsymbol{A}\mathbf{y}$, with g_a denoting entry-wise clipping to the values $\pm a$ for some $a > 0$. We introduce the notation

$$\underline{R}_{\mathrm{B}}^{\mathbf{x}}(\varepsilon), \underline{a}_n^{\mathbf{x}}(\varepsilon), \overline{R}_{\mathrm{B}}^{\mathbf{x}}(\varepsilon), \overline{a}_n^{\mathbf{x}}(\varepsilon)$$
$$\underline{R}_{\mathrm{B}}^{\mathbf{y}}(\varepsilon), \underline{a}_n^{\mathbf{y}}(\varepsilon), \overline{R}_{\mathrm{B}}^{\mathbf{y}}(\varepsilon), \overline{a}_n^{\mathbf{y}}(\varepsilon)$$

for the quantities in Definition 4 corresponding to the processes $\mathbf{x} = [\mathbf{y}^{\mathrm{T}}\ \mathbf{z}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbf{y}$, respectively.

Lemma 4: For $\varepsilon > 0$, we have

$$\underline{R}_{\mathrm{B}}^{\mathbf{x}}(\varepsilon) = \frac{1}{2}\underline{R}_{\mathrm{B}}^{\mathbf{y}}(\varepsilon) \quad \text{and} \quad \overline{R}_{\mathrm{B}}^{\mathbf{x}}(\varepsilon) = \frac{1}{2}\overline{R}_{\mathrm{B}}^{\mathbf{y}}(\varepsilon).$$

Proof: We only prove the first identity and note that the second is obtained by simply replacing $\underline{\dim}_{\mathrm{B}}(\cdot)$ by $\overline{\dim}_{\mathrm{B}}(\cdot)$ in the arguments below. Let us begin by showing that

$$\underline{R}_{\mathrm{B}}^{\mathbf{x}}(\varepsilon) \leqslant \frac{1}{2}\underline{R}_{\mathrm{B}}^{\mathbf{y}}(\varepsilon). \tag{88}$$

Recall that $\ell = \lfloor \frac{n}{2} \rfloor$, and suppose that we are given a set $\mathcal{S} \subseteq \mathbb{R}^{n-\ell}$ such that $\mathbb{P}[\mathbf{y} \in \mathcal{S}] \geqslant 1-\varepsilon$. Set

$$\mathcal{T} := \{[\boldsymbol{y}^{\mathrm{T}}\ \ (g_a(\boldsymbol{A}\boldsymbol{y}) - \boldsymbol{A}\boldsymbol{y})^{\mathrm{T}}]^{\mathrm{T}} : \boldsymbol{y} \in \mathcal{S}\} \subseteq \mathbb{R}^n,$$

and note that for all $\boldsymbol{y}_1, \boldsymbol{y}_2 \in \mathbb{R}^{n-\ell}$ we have

$$\begin{aligned}\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| \leqslant \|[\boldsymbol{y}_1^{\mathrm{T}}\ \ (g_a(\boldsymbol{A}\boldsymbol{y}_1) - \boldsymbol{A}\boldsymbol{y}_1)^{\mathrm{T}}]^{\mathrm{T}} \\ - [\boldsymbol{y}_2^{\mathrm{T}}\ \ (g_a(\boldsymbol{A}\boldsymbol{y}_2) - \boldsymbol{A}\boldsymbol{y}_2)^{\mathrm{T}}]^{\mathrm{T}}\| \end{aligned} \tag{89}$$

and

$$\begin{aligned}&\|[\boldsymbol{y}_1^{\mathrm{T}}\ \ (g_a(\boldsymbol{A}\boldsymbol{y}_1) - \boldsymbol{A}\boldsymbol{y}_1)^{\mathrm{T}}]^{\mathrm{T}} - [\boldsymbol{y}_2^{\mathrm{T}}\ \ (g_a(\boldsymbol{A}\boldsymbol{y}_2) - \boldsymbol{A}\boldsymbol{y}_2)^{\mathrm{T}}]^{\mathrm{T}}\| \\ &\leqslant \|\boldsymbol{y}_1 - \boldsymbol{y}_2\| + \|g_a(\boldsymbol{A}\boldsymbol{y}_1) - \boldsymbol{A}\boldsymbol{y}_1 - (g_a(\boldsymbol{A}\boldsymbol{y}_2) - \boldsymbol{A}\boldsymbol{y}_2)\| \qquad (90) \\ &\leqslant (1 + \|\boldsymbol{A}\|)\|\boldsymbol{y}_1 - \boldsymbol{y}_2\| + \|g_a(\boldsymbol{A}\boldsymbol{y}_1) - g_a(\boldsymbol{A}\boldsymbol{y}_2)\| \qquad (91) \\ &\leqslant (1 + 2\|\boldsymbol{A}\|)\|\boldsymbol{y}_1 - \boldsymbol{y}_2\|, \qquad (92)\end{aligned}$$

where (89), (90), and (91) follow from the triangle inequality, and (92) holds as $|g_a(y_1) - g_a(y_2)| \leqslant |y_1 - y_2|$, for all $y_1, y_2 \in \mathbb{R}$. Combining (89) and (90)–(92), it follows that for $\delta > 0$

$$N_{\mathcal{S}}(\delta) \leqslant N_{\mathcal{T}}(\delta) \leqslant N_{\mathcal{S}}((1 + 2\|\boldsymbol{A}\|)\delta), \tag{93}$$

which implies $\underline{\dim}_{\mathrm{B}}(\mathcal{S}) = \underline{\dim}_{\mathrm{B}}(\mathcal{T})$. Since $2(n-\ell) = 2(n - \lfloor n/2 \rfloor) \leqslant n+2$, we obtain

$$\frac{1}{2}\frac{\underline{\dim}_{\mathrm{B}}(\mathcal{S})}{n-\ell} = \frac{1}{2}\frac{\underline{\dim}_{\mathrm{B}}(\mathcal{T})}{n-\ell} \geqslant \frac{\underline{\dim}_{\mathrm{B}}(\mathcal{T})}{n+2},$$

and therefore $\frac{1}{2}\underline{a}_{n-\ell}^{\mathbf{y}}(\varepsilon) \geqslant \frac{n}{n+2}\underline{a}_n^{\mathbf{x}}(\varepsilon)$. As $\lim_{n\to\infty} n/(n+2) = 1$, we get (88).

To prove

$$\underline{R}_{\mathrm{B}}^{\mathbf{x}}(\varepsilon) \geqslant \frac{1}{2}\underline{R}_{\mathrm{B}}^{\mathbf{y}}(\varepsilon), \tag{94}$$

we consider a set $\mathcal{U} \subseteq \mathbb{R}^n$ such that $\mathbb{P}[\mathbf{x} \in \mathcal{U}] \geqslant 1-\varepsilon$. Setting $\mathcal{V} := \{\boldsymbol{y} \in \mathbb{R}^{n-\ell} : [\boldsymbol{y}^{\mathrm{T}}\ \ (g_a(\boldsymbol{A}\boldsymbol{y}) - \boldsymbol{A}\boldsymbol{y})^{\mathrm{T}}]^{\mathrm{T}} \in \mathcal{U}\}$, we have $\mathbb{P}[\mathbf{y} \in \mathcal{V}] = \mathbb{P}[\mathbf{x} \in \mathcal{U}]$. For the set $\widetilde{\mathcal{U}} := \{[\boldsymbol{y}^{\mathrm{T}}\ \ (g_a(\boldsymbol{A}\boldsymbol{y}) - \boldsymbol{A}\boldsymbol{y})^{\mathrm{T}}]^{\mathrm{T}} : \boldsymbol{y} \in \mathcal{V}\}$ we have $\underline{\dim}_{\mathrm{B}}(\widetilde{\mathcal{U}}) = \underline{\dim}_{\mathrm{B}}(\mathcal{V})$ by the same arguments as in (89)–(93). Moreover, by definition of $\mathcal{V}$ we have $\widetilde{\mathcal{U}} \subseteq \mathcal{U}$ and therefore $\underline{\dim}_{\mathrm{B}}(\widetilde{\mathcal{U}}) \leqslant \underline{\dim}_{\mathrm{B}}(\mathcal{U})$, which implies

$$\frac{1}{2}\frac{\underline{\dim}_{\mathrm{B}}(\mathcal{V})}{n-\ell} = \frac{1}{2}\frac{\underline{\dim}_{\mathrm{B}}(\widetilde{\mathcal{U}})}{n-\ell} \leqslant \frac{1}{2}\frac{\underline{\dim}_{\mathrm{B}}(\mathcal{U})}{n-\ell} \leqslant \frac{\underline{\dim}_{\mathrm{B}}(\mathcal{U})}{n},$$

where in the last step we used $2(n-\ell) \geqslant n$. This shows that $\underline{a}_n^{\mathbf{x}}(\varepsilon) \geqslant \frac{1}{2}\underline{a}_{n-\ell}^{\mathbf{y}}(\varepsilon)$ which establishes (94) and thereby completes the proof. ∎

Appendix B
Alternative definition of Minkowski dimension

In this section, we prove that Minkowski dimension can equivalently be defined through the modified covering number (6). Similar arguments for different modifications of the covering number (5) can be found in [17, Equivalent Definitions 3.1].

Lemma 5: The Minkowski dimension of a non-empty bounded set $\mathcal{S} \subseteq \mathbb{R}^n$ does not change when the covering balls in the definitions (3), (4) are restricted to have their centers inside the set $\mathcal{S}$, that is, we have

$$\liminf_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}} = \liminf_{\delta\to 0} \frac{\log M_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}} \tag{95}$$

$$\limsup_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}} = \limsup_{\delta\to 0} \frac{\log M_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}}, \tag{96}$$

where $N_{\mathcal{S}}(\delta)$ is the covering number of $\mathcal{S}$ given by

$$N_{\mathcal{S}}(\delta) = \min\Big\{m \in \mathbb{N} : \mathcal{S} \subseteq \bigcup_{i\in\{1,\dots,m\}} B^n(\boldsymbol{x}_i, \delta),\ \boldsymbol{x}_i \in \mathbb{R}^n\Big\},$$

and $M_{\mathcal{S}}(\delta)$ is the covering number of $\mathcal{S}$ with the covering balls centered in $\mathcal{S}$, i.e.,

$$M_{\mathcal{S}}(\delta) = \min\Big\{m \in \mathbb{N} : \mathcal{S} \subseteq \bigcup_{i\in\{1,\dots,m\}} B^n(\boldsymbol{x}_i, \delta),\ \boldsymbol{x}_i \in \mathcal{S}\Big\}.$$

Proof: Since $N_{\mathcal{S}}(\delta) \leqslant M_{\mathcal{S}}(\delta)$, the "$\leqslant$"-part in (95) and (96) is immediate. To establish the "$\geqslant$"-part, we consider a set of covering balls of $\mathcal{S}$ of radius $\delta/2$ and corresponding centers $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{N_{\mathcal{S}}(\delta/2)} \in \mathbb{R}^n$. Note that these centers do not necessarily lie in $\mathcal{S}$. Since $N_{\mathcal{S}}(\delta/2)$ is the minimum number of balls with radius $\delta/2$ needed to cover $\mathcal{S}$, the intersection $B^n(\boldsymbol{x}_i, \delta/2) \cap \mathcal{S}$ must be non-empty for all $i = 1, \dots, N_{\mathcal{S}}(\delta/2)$. We now choose an arbitrary point $\boldsymbol{y}_i \in (B^n(\boldsymbol{x}_i, \delta/2) \cap \mathcal{S})$ and note that

$$\|\boldsymbol{u} - \boldsymbol{y}_i\| \leqslant \|\boldsymbol{u} - \boldsymbol{x}_i\| + \|\boldsymbol{x}_i - \boldsymbol{y}_i\| \leqslant \|\boldsymbol{u} - \boldsymbol{x}_i\| + \delta/2,$$

for all $\boldsymbol{u} \in \mathbb{R}^n$, which implies $B^n(\boldsymbol{x}_i, \delta/2) \subseteq B^n(\boldsymbol{y}_i, \delta)$, $i = 1, \dots, N_{\mathcal{S}}(\delta/2)$. It therefore follows that $B^n(\boldsymbol{y}_i, \delta)$, $i = 1, \dots, N_{\mathcal{S}}(\delta/2)$, is a covering of $\mathcal{S}$ with balls of radius δ all centered in $\mathcal{S}$. This implies

$$M_{\mathcal{S}}(\delta) \leqslant N_{\mathcal{S}}(\delta/2), \tag{97}$$

and hence

$$\frac{\log M_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}} \leqslant \frac{\log N_{\mathcal{S}}(\delta/2)}{\log \frac{1}{\delta}} = \frac{\log N_{\mathcal{S}}(\delta/2)}{\log \frac{2}{\delta}} \underbrace{\frac{\log \frac{2}{\delta}}{\log \frac{1}{\delta}}}_{\xrightarrow{\delta\to 0} 1}. \tag{98}$$

Taking $\liminf_{\delta\to 0}$ and $\limsup_{\delta\to 0}$ on both sides of (98) yields the "$\geqslant$"-part in (95) and (96), respectively, according to

$$\liminf_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta/2)}{\log \frac{2}{\delta}} = \liminf_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}},$$

$$\limsup_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta/2)}{\log \frac{2}{\delta}} = \limsup_{\delta\to 0} \frac{\log N_{\mathcal{S}}(\delta)}{\log \frac{1}{\delta}}.$$

∎

Acknowledgments

The authors are thankful to the anonymous reviewers for very insightful comments. In particular, we would like to acknowledge one of the reviewers for pointing out an alternative proof of Theorem 1 (cf. Remark 4).

References

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

[2] D. Stotz, E. Riegler, E. Agustsson, and H. Bölcskei, "Addendum to "Almost lossless analog signal separation and probabilistic uncertainty relations"," *available online: http://www.nari.ee.ethz.ch/commth/research/downloads/sigsep_addendum.pdf*, May 2017.

[3] C. Studer, P. Kuppinger, G. Pope, and H. Bölcskei, "Recovery of sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 58, no. 5, pp. 3115–3130, May 2012.

[4] X. Li, "Compressed sensing and matrix completion with constant proportion of corruptions," *Constructive Approximation*, vol. 37, no. 1, pp. 79–99, Feb. 2013.

[5] J. Wright and Y. Ma, "Dense error correction via ℓ^1-minimization," *IEEE Trans. Inf. Theory*, vol. 56, no. 7, pp. 3540–3560, Jul. 2010.

[6] D. L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Natl. Acad. Sci.*, vol. 106, no. 45, pp. 18 914–18 919, 2009.

[7] D. L. Donoho and G. Kutyniok, "Microlocal analysis of the geometric separation problem," *Comm. Pure and Appl. Math.*, vol. 66, no. 1, pp. 1–47, Jan. 2013.

[8] J. A. Tropp, "On the conditioning of random subdictionaries," *Applied and Computational Harmonic Analysis*, vol. 25, pp. 1–24, 2008.

[9] M. B. McCoy and J. A. Tropp, "Sharp recovery bounds for convex demixing, with applications," *Found. Comp. Math.*, vol. 14, no. 3, pp. 503–567, Jun. 2014.

[10] G. Pope, A. Bracher, and C. Studer, "Probabilistic recovery guarantees for sparsely corrupted signals," *IEEE Trans. Inf. Theory*, vol. 59, no. 5, pp. 3104–3116, May 2013.

[11] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

[12] D. L. Donoho and P. B. Stark, "Uncertainty principles and signal recovery," *SIAM Journal on Applied Mathematics*, vol. 49, no. 3, pp. 906–931, Jun. 1989.

[13] 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.

[14] T. D. Sauer, J. A. Yorke, and M. Casdagli, "Embedology," *Journal of Statistical Physics*, vol. 65, no. 3-4, pp. 579–616, Nov. 1991.

[15] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

[16] P. Kuppinger, G. Durisi, and H. Bölcskei, "Uncertainty relations and sparse signal recovery for pairs of general signal sets," *IEEE Trans. Inf. Theory*, vol. 58, no. 1, pp. 263–277, Jan. 2012.

[17] K. Falconer, *Fractal Geometry: Mathematical Foundations and Applications*, 2nd ed. John Wiley & Sons, 2004.

[18] G. B. Folland, *Real Analysis*, 2nd ed., ser. Pure and Applied Mathematics. New York, NY: Wiley, 1999.

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

[20] E. Riegler and G. Tauböck, "Almost lossless analog compression without phase information," *Proc. IEEE Int. Symp. Inf. Theory*, pp. 999–1003, Jun. 2015.

[21] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, 3rd ed. Heidelberg, Germany: Springer, 2009.

[22] G. Alberti, H. Bölcskei, C. De Lellis, G. Koliander, and E. Riegler, "Lossless analog compression," *IEEE Trans. Inf. Theory,* in preparation.

[23] P. Mattila, *Geometry of Sets and Measures in Euclidean Space: Fractals and Rectifiability*. Cambridge, UK: Cambridge Univ. Press, 1999.

[24] R. Durrett, *Probability: Theory and Examples*, 4th ed. Cambridge University Press, 2010.

[25] R. G. Bartle, *The Elements of Integration and Lebesgue Measure*. New York, NY: Wiley, 1995.

[26] G. J. Minty, "On the extension of Lipschitz, Lipschitz-Hölder continuous, and monotone functions," *Bulletin of the American Mathematical Society*, vol. 76, no. 2, pp. 334–339, Mar. 1970.

[27] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[28] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[29] J. A. Tropp, "On the linear independence of spikes and sines," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 838–858, 2008.