Almost Lossless Analog Signal Separation
and Probabilistic Uncertainty Relations

David Stotz, Erwin Riegler, Eirikur Agustsson, and Helmut Bölcskei
ETH Zurich, Switzerland

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 subsets 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).

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
I. INTRODUCTION

We consider the following signal separation problem: recover the vectors $y$ and $z$ from the noiseless observation

$$w = Ay + Bz,$$

where $A$ and $B$ 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 corresponding references, we refer the reader to [3, Sec. 1].

The sparse signal recovery literature [3]–[12] provides separation guarantees under sparsity constraints on the vectors $y$ and $z$. Specifically, the sparsity thresholds in [3], [7], [11], [12] are functions of the coherence parameters [3] of the matrices $A$ and $B$ and hold for all $y$ and $z$, but suffer from the “square-root bottleneck” [8], which states that the number of measurements, i.e., the number of entries of $w$, has to scale quadratically in the total number of non-zero entries in $y$ and $z$. For random signals $y$ and $z$, the probabilistic results in [4], [5], [10] overcome the square-root bottleneck, but hold “only” with overwhelming probability. For $B$ the identity matrix and $A$ a random orthogonal matrix it is shown in [9] that the probability of failure of an $\ell_1$-based separation algorithm decays exponentially in the dimension of the ambient space, provided that $y$ and $z$ satisfy certain convex cone conditions.

Contributions. The goal of this paper is to develop an information-theoretic framework for the problem of recovering $y$ and $z$ from $w = Ay + Bz$. Specifically, inspired by the groundbreaking work of Wu and Verdú on analog compression [13], we derive asymptotic (in the dimensions of $w$, $y$, and $z$) recovery guarantees for $y$ and $z$ random, possibly dependent, and of general distributions, i.e., mixtures of discrete, absolutely continuous, and singular distributions. In practical signal separation problems of this form we often encounter a specific structure for one of the matrices (assumed $B$ here, 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 $B$. Our separation guarantees will, indeed, be seen to apply to deterministic $B$ and almost all (a.a.) $A$ (with the set of exceptions for $A$ depending on the specific choice of $B$). Moreover, they do not depend on coherence parameters, are in terms of probability of separation error with respect to the constituents $y$ and $z$, and hence do not provide worst-case guarantees like the coherence-based results in [3], [7], [11], [12]. Specifically, we study the asymptotic setting $\ell, n \to \infty$ where the vectors $y \in \mathbb{R}^{n-\ell}$ and $z \in \mathbb{R}^\ell$ are sections of realizations of
random processes; for each $n$, we let $\ell = \lfloor \lambda n \rfloor$ and $k = \lfloor Rn \rfloor$ 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 in $y$ and $z$ given by $n$. Our first main result, Theorem 1, shows that for each (deterministic) full-rank matrix $B \in \mathbb{R}^{k \times \ell}$ with $k \geq \ell$ and a.a. matrices $A \in \mathbb{R}^{k \times (n-\ell)}$, there exists a measurable 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 $[y^T \; z^T]^T$ as quantified by its Minkowski dimension compression rate $R_B(\varepsilon)$ (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 $B \in \mathbb{R}^{k \times \ell}$ with $k \geq \ell$ and a.a. matrices $A \in \mathbb{R}^{k \times (n-\ell)}$, there exists a $\beta$-Hölder continuous separator achieving error probability $\varepsilon$ provided that $R > R_B(\varepsilon)$ and $\beta < 1 - \frac{R_B(\varepsilon)}{R}$.

In the case of mixed discrete-continuous source distributions a converse matching the general—with respect to the nature of the source distributions—achievability statements in Theorems 1 and 2 can be obtained. This establishes the Minkowski dimension compression rate $R_B(\varepsilon)$ as the critical rate for successful separation when the source distributions are restricted to be mixed discrete-continuous.

In principle one could rewrite (1) in the form $w = [A \; B] \begin{bmatrix} y \\ z \end{bmatrix}$, and consider applying the results in [13] with $H = [A \; B]$. This is, however, not possible as the theory developed in [13] leads to statements that apply to a.a. matrices $H$, whereas here, for reasons mentioned above, we seek statements that apply for a given matrix $B$, and fixing $B$ results in $H = [A \; B]$-matrices supported on a set of Lebesgue measure zero. We therefore have to 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 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]).

**Notation.** For a relation $\# \in \{<,>,\leq,\geq,=,\neq,\in,\notin\}$, we write $f(n) \# g(n)$ if there exists an
$N \in \mathbb{N}$ such that $f(n) \neq g(n)$ holds for all $n \geq N$. Leb$^n$ stands for the $n$-dimensional Lebesgue measure and $\mathcal{B}^\otimes n$ refers to the Borel $\sigma$-algebra on $\mathbb{R}^n$. Matrices are denoted by capital boldface and vectors by lowercase boldface letters. We let $\| \cdot \|$ be the $\ell_2$-norm on $\mathbb{R}^n$ and set $\|A\| := \sup_{\|x\|=1} \|Ax\|$. The $n \times n$ identity matrix is $I_n$ and $F_n$ stands for the $n$-dimensional discrete Fourier transform (DFT) matrix. For $x \in \mathbb{R}^n$ and $T \subseteq \{1, \ldots, n\}$, we let $x_T$ denote the $|T|$-dimensional subvector that consists of the components of $x$ corresponding to the indices in $T$. Sets are represented by calligraphic letters.

$\mathbf{II. STATEMENT OF THE MAIN RESULTS}$

We begin by introducing our information-theoretic framework for signal separation and defining the quantities needed in the statements of the main results. The recovery of the vectors $y$ and $z$ from the noiseless observation $w$ in (1) can be rephrased as the recovery of the vector $x = [y^T \ z^T]^T$ from the linear measurement

$$w = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} y \\ z \end{bmatrix},$$

with measurement matrix $H = [A \ B]$. An information-theoretic framework for analog compression, i.e., for the problem of recovering $x$ from the linear measurements $w = Hx$, 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 the 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 $A$ and $B$ into account. For example, for the recovery of signals corrupted by impulse noise or narrowband interference one of the matrices $A$ and $B$ equals the identity matrix or the DFT matrix, respectively. This will be accounted for by taking $B$ to be deterministic and fixed throughout the paper. As the results in [13] lead to statements that apply to a.a. measurement matrices $H$ and fixing $B$ results in $H = [A B]$-matrices supported on a set of Lebesgue measure zero, it follows that [13] is not applicable here. It turns out that addressing this problem requires new techniques, which lead to 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], that underly much of compressed sensing theory. Our probabilistic uncertainty relations will allow us to make statements that apply to a.a. $A$ for a fixed $B$ (with the set of exceptions for $A$ depending on the specific choice of $B$). Moreover, the new technique also allows for a simplification of the proof of [13, Thm. 18, 1)] and a significant improvement of the statement [13, Thm. 18, 2)].

We next define the specifics of our setup.

Definition 1: Suppose that $(Y_i)_{i \in \mathbb{N}}$ and $(Z_i)_{i \in \mathbb{N}}$ are stochastic processes on $(\mathbb{R}^N, \mathcal{B}^\otimes \mathbb{N})$. Then, for $n \in \mathbb{N}$, we define the concatenated source vector $x$ of dimension $n$ as $x = [X_1 \ldots X_n]^T$ according to

\[
X_i = Y_i, \quad \text{for } i \in \{1, \ldots, n-\ell\} \\
X_{n-\ell+i} = Z_i, \quad \text{for } i \in \{1, \ldots, \ell\},
\]

where $\ell = \lfloor \lambda n \rfloor$ with the parameter $\lambda \in [0, 1]$ representing the asymptotic fraction of components in $x$ 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, absolutely continuous (with respect to Lebesgue measure), and singular distributions, i.e., $\mu = \mu_d + \mu_c + \mu_s$.

The encoding–decoding part of our framework comprises

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

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

December 4, 2015 DRAFT
We will deal with separators $g$ that are measurable (with respect to $B^\otimes k$ and $B^\otimes n$) and with $g$ that are, in addition, $\beta$-Hölder continuous, i.e., for a given $\beta > 0$ they satisfy
\[ \|g(x_1) - g(x_2)\| \leq c\|x_1 - x_2\|^{\beta}, \quad \text{for all } x_1, x_2 \in \mathbb{R}^k, \]
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 $x$ as in Definition $[1]$ and a given measurement matrix $H = [A \ B]$, we say that there exists a (measurable or $\beta$-Hölder continuous) separator that achieves rate $R \in [0, 1]$ with error probability $\varepsilon \in (0, 1)$ if there exists a sequence (with respect to $n$) of (measurable or $\beta$-Hölder continuous) maps $g$ such that $k = \lfloor Rn \rfloor$ and
\[ P[g([A \ B]x) \neq x] \leq \varepsilon. \]

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

Definition 3: (Minkowski dimension, $[17]$). Let $S$ be a non-empty bounded set in $\mathbb{R}^n$. Define the lower and upper Minkowski dimension of $S$ as
\[
\dim_B(S) = \liminf_{\delta \to 0} \frac{\log N_S(\delta)}{\log \frac{1}{\delta}}, \tag{3}
\]
\[
\overline{\dim}_B(S) = \limsup_{\delta \to 0} \frac{\log N_S(\delta)}{\log \frac{1}{\delta}}, \tag{4}
\]
where $N_S(\delta)$ is the covering number of $S$ given by
\[
N_S(\delta) = \min \left\{ m \in \mathbb{N} \mid S \subseteq \bigcup_{i \in \{1, \ldots, m\}} B^n(x_i, \delta), \ x_i \in \mathbb{R}^n \right\}. \tag{5}
\]
If $\dim_B(S) = \overline{\dim}_B(S)$, we define the Minkowski dimension of $S$ as $\dim_B(S) := \dim_B(S) = \overline{\dim}_B(S)$.

Remark 1: It is well-known that the Minkowski dimension can equivalently be defined by replacing $N_S(\delta)$ in (3), (4) with modified covering numbers $[17]$ Equivalent Definitions 3.1. In Lemma $5$ in Appendix $B$ we show that Minkowski dimension can be defined using yet another modification of the covering number, namely $N_S(\delta)$ in (3), (4) can be replaced by
\[
M_S(\delta) = \min \left\{ m \in \mathbb{N} \mid S \subseteq \bigcup_{i \in \{1, \ldots, m\}} B^n(x_i, \delta), \ x_i \in S \right\}. \tag{6}
\]
which is in terms of covering balls with centers in the set \( S \). This equivalent definition is often convenient as the covering ball centers inherit structural properties of the set \( S \).

As our framework involves statements that are asymptotic in the blocklength \( n \), we will need an extension of the concept of Minkowski dimension to random processes. This leads to the notion of Minkowski dimension compression rate:

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

\[
R_B(\epsilon) = \limsup_{n \to \infty} a_n(\epsilon), \quad \text{where}
\]

\[
a_n(\epsilon) = \inf \left\{ \frac{\dim_B(S)}{n} \, \bigg| \, S \subseteq \mathbb{R}^n, \, P[x \in S] \geq 1 - \epsilon \right\},
\]

and

\[
\overline{R}_B(\epsilon) = \limsup_{n \to \infty} \overline{a}_n(\epsilon), \quad \text{where}
\]

\[
\overline{a}_n(\epsilon) = \inf \left\{ \frac{\dim_B(S)}{n} \, \bigg| \, S \subseteq \mathbb{R}^n, \, P[x \in S] \geq 1 - \epsilon \right\}.
\]

If \( R_B(\epsilon) = \overline{R}_B(\epsilon) \), we define the Minkowski dimension compression rate as \( R_B(\epsilon) := R_B(\epsilon) = \overline{R}_B(\epsilon) \).

The following theorem constitutes our first main result.

**Theorem 1**: Let \( x \) be as in Definition 1. Take \( \epsilon > 0 \) and let \( R > R_B(\epsilon) \). Then, for every full-rank matrix \( B \in \mathbb{R}^{k \times \ell} \), with \( k \geq \ell \), and for a.a. (with respect to \( \text{Leb}^k(n-\ell) \)) matrices \( A \in \mathbb{R}^{k \times (n-\ell)} \), where \( k = \lfloor Rn \rfloor \), there exists a measurable separator \( g \) such that

\[
P[g([A \, B]x) \neq x] \leq \epsilon.
\]

**Proof**: See Section IV.

**Remark 2**: The set of exceptions for \( A \) depends on the specific choice of the full-rank matrix \( B \). The proof of Theorem 1 further reveals that the minimum \( N \in \mathbb{N} \) for (11) to hold for all \( n \geq N \) depends on the distribution of \( x \) only and is independent of the matrices \( A \) and \( B \).

**Remark 3**: In [13, Thm. 18, 1]) it was shown—for analog compression—that every rate \( R \) with \( R > \overline{R}_B(\epsilon) \) is achievable for a.a. measurement matrices \( H \in \mathbb{R}^{k \times n} \). As already mentioned, this does, however, not imply that signal separation is possible for a fixed submatrix \( B \) in \( H = [A \, B] \). The proof of [13, Thm. 18, 1]) relies on \( H \) being random and exploits intricate properties of invariant measures on Grassmannian manifolds under the action of the orthogonal group. The corresponding arguments do not carry over to the signal separation setting considered here as our overall measurement matrix \( H = [A \, B] \) has a fixed block \( B \). The new proof technique we develop 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 \(H = [A \ B]\) and the approximate support set \(S\) of the concatenated source vector \(x\) do not intersect if the Minkowski dimension of \(S\) is smaller than \(k\). Underlying this result is the basic idea that two objects—in generic 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 \(H\)) and Minkowski dimension (for the support set \(S\)) are compatible dimensionality notions in this context. 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 a significant improvement 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 \(B\) and a.a. \(A\) to guarantee the existence of measurable separators that are, in addition, Hölder continuous, even though Hölder continuity is a much stronger notion than measurability. It is therefore not surprising that the corresponding statement we obtain is weaker—but only slightly so—than that for measurable separators. Specifically, we establish the existence of a \(\beta\)-Hölder continuous separator with the threshold \(R > R_B(\epsilon)\) instead of \(R > R_B(\epsilon)\), provided that \(\beta < 1 - \frac{R_B(\epsilon)}{R}\).

**Theorem 2:** Let \(x\) be as in Definition 1, \(R > R_B(\epsilon)\), for \(\epsilon > 0\), and fix \(\beta > 0\) such that

\[
\beta < 1 - \frac{R_B(\epsilon)}{R}.
\]

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

\[
\mathbb{P}[g([A \ B]x) \neq x] \leq \epsilon + \kappa,
\]

where \(\kappa > 0\) is an arbitrarily small constant.

**Remark 4:** As in Theorem 1, the set of exceptions for \(A\) depends on the specific choice of the full-rank matrix \(B\). The constant \(\kappa\) honors the fact that we have to excise a small set of points 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 the recovery of \( y \) and \( z \) from \( w \) in (2). Formally, the question of uniqueness boils down to asking whether different pairs \((y, z)\) and \((y', z')\) exist such that

\[
Ay + Bz = Ay' + Bz', \tag{13}
\]

or, equivalently,

\[
A(y - y') = B(z' - z). \tag{14}
\]

In the context of compressed sensing where \( y, y', z, z' \) are sparse signals, the differences \( y - y' \) and \( z' - z \) are sparse as well so that (14) would imply the existence of a non-zero signal \( s := A(y - y') = B(z' - z) \) that can be sparsely represented in both dictionaries \( A \) and \( B \). Uncertainty principles are at the heart of compressed sensing theory and state that no such \( s \) can exist if the signals \( y, y', z, z' \) and hence \( y - y' \) and \( z' - z \) are sufficiently sparse and the dictionaries \( A \) and \( B \) are sufficiently incoherent, thereby guaranteeing that, for a given \( w \), there is a unique pair \((y, z)\) such that \( w = Ay + Bz \). Specifically, the Donoho-Stark uncertainty principle [12] considers the square matrices \( A = I_n \) and \( B = F_n \), and states that there exists no pair of vectors \((p, q) \neq 0\) with \( 2n_p n_q < n \) satisfying

\[
Ap = Bq, \tag{15}
\]

where \( n_p \) and \( n_q \) denote the number of non-zero entries in \( p \) and \( q \), respectively. Elad and Bruckstein [15] generalized the Donoho-Stark uncertainty principle to arbitrary orthonormal bases \( A \) and \( B \) and found that no pair of vectors \((p, q) \neq 0\) with \((n_p + n_q)/2 < 1/\mu\) satisfying (15) exists. Here,

\[
\mu := \sup_{1 \leq i, j \leq n} |(a_i, b_j)|
\]

is the coherence of \( A = [a_1 \ldots a_n] \) and \( B = [b_1 \ldots b_n] \). 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.

The central tool in the proof of our Theorem 1 is the following probabilistic uncertainty relation. If we restrict \((p, q)\) to lie in a set \( S \) of (sufficiently) small Minkowski dimension and fix a matrix \( B \), then for a.a. matrices \( A \) there is no \((p, q) \in S \setminus \{0\}\) such that \( Ap = Bq \) 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 or recovered. The formal statement of our probabilistic uncertainty relation is as follows.

**Proposition 1:** Let \( S \subseteq \mathbb{R}^{n-\ell} \times \mathbb{R}^{\ell} \) be non-empty and bounded such that \( \dim_B(S) < k \), and let \( B \in \mathbb{R}^{k \times \ell} \), with \( k \geq \ell \), be a matrix with rank\((B) = \ell \). Then,

\[
\left\{ \begin{array}{l}
\begin{bmatrix} y \\
z \end{bmatrix} \in S \setminus \{0\} \\
[A \ B] \begin{bmatrix} y \\
z \end{bmatrix} = 0
\end{array} \right\} = \emptyset,
\]

(16)

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

**Proof:** We show that (16) holds with probability 1 for the random matrix \( A = [a_1 \ldots a_k]^T \), where the \( a_i \) are i.i.d. uniform on \( B^{n-\ell}(0, r) \) and \( r > 0 \) is arbitrary. Since \( r \) can, in particular, be chosen arbitrarily large, this establishes that the Lebesgue measure of matrices \( A \) violating (16) is zero. We start by arguing that, thanks to the full-rank assumption on \( B \), it suffices to show that

\[
P \left[ \exists \begin{bmatrix} y \\
z \end{bmatrix} \in S \setminus \{0\} : [A \ B] \begin{bmatrix} y \\
z \end{bmatrix} = 0 \right] = 0,
\]

(17)

for sets \( S \) that have the norm of the \( y \)-parts of their elements bounded away from zero. To see this, we first note that \( B \), by virtue of being full-rank, maps non-zero vectors to non-zero vectors. For \( [y^T z^T]^T \in S \setminus \{0\} \), \( Ay + Bz = 0 \) is therefore possible only when \( y \neq 0 \) as \( y = 0 \) would lead to \( Bz = 0 \) which in turn would result in \( [y^T z^T]^T = 0 \). Therefore, we can rewrite (17) as

\[
P \left[ \exists \begin{bmatrix} y \\
z \end{bmatrix} \in S \setminus \{0\} : [A \ B] \begin{bmatrix} y \\
z \end{bmatrix} = 0 \right] = P \left[ \exists \begin{bmatrix} y \\
z \end{bmatrix} \in S \setminus \{0\}, y \neq 0 : [A \ B] \begin{bmatrix} y \\
z \end{bmatrix} = 0 \right].
\]

(18)

A union bound argument applied to the right-hand side (RHS) of (18) then yields

\[
P \left[ \exists \begin{bmatrix} y \\
z \end{bmatrix} \in S \setminus \{0\} : [A \ B] \begin{bmatrix} y \\
z \end{bmatrix} = 0 \right]
\leq \sum_{m=1}^{\infty} P \left[ \exists \begin{bmatrix} y \\
z \end{bmatrix} \in S \setminus \{0\}, \|y\| \geq \frac{1}{m} : [A \ B] \begin{bmatrix} y \\
z \end{bmatrix} = 0 \right].
\]

(20)

This allows us to conclude that (17) is established by showing that

\[
P \left[ \exists \begin{bmatrix} y \\
z \end{bmatrix} \in S' : [A \ B] \begin{bmatrix} y \\
z \end{bmatrix} = 0 \right] = 0,
\]

(21)

for all non-empty bounded sets \( S' \subseteq S \) with

\[
\inf \left\{ \|y\| \mid \begin{bmatrix} y \\
z \end{bmatrix} \in S' \right\} > 0,
\]

(22)
as this implies that each term in the series in (20) equals zero. Note that we no longer need to excise 0 from \( S' \) in (21), as \( S' \) is guaranteed not to contain 0 by definition, cf. (22).

We next employ a covering argument, which reduces the question of the existence of \( [y^T z^T]^T \in S' \) such that

\[
Ay + Bz = 0
\]

(23)
to the question of the existence of covering ball centers satisfying (23). For reasons that will become clear towards the end of the proof, we employ the modified covering number \( M_{S'}(\delta) \) (defined in (6)), which requires the covering ball centers to lie in \( S' \). This implies that the covering ball centers \( [y_i^T z_i^T]^T \) satisfy

\[
\min_i \|y_i\| \geq \inf \left\{ \|y\| \mid \begin{bmatrix} y \\ z \end{bmatrix} \in S' \right\} > 0. \tag{24}
\]

By definition of \( \dim_B(\cdot) \) in (3) there exists a sequence of covering ball radii \( \delta_j \) tending to zero with corresponding covering ball centers \( x_1, \ldots, x_{M_{S'}(\delta_j)} \in S' \) such that

\[
\frac{\log M_{S'}(\delta_j)}{\log \frac{1}{\delta_j}} \to_{j \to \infty} \dim_B(S'). \tag{25}
\]

Next, we note that

\[
\|[A, B]u\| \leq c(k, r, \|[B]\|)\|u\|, \quad \text{for all } u \in \mathbb{R}^n, \tag{26}
\]

since i) \( \|[A, B]u\| \leq \|[A, B]\|\|u\| \), ii) \( \|[A, B]\| = \sup_{\|[y^T z^T]^T\| = 1} \|Ay + Bz\| \leq \sup_{\|[y^T z^T]^T\| = 1} \left( \|Ay\| + \|Bz\| \right) \leq \sup_{\|[y^T z^T]^T\| = 1} (\|A\| \cdot \|y\| + \|[B]\| \cdot \|z\|) \leq \|A\| + \|[B]\|, \) and iii)

\[
\|Ay\| = \sqrt{\langle a_1, y \rangle^2 + \ldots + \langle a_k, y \rangle^2} \tag{27}
\]

\[
\leq \sqrt{\|a_1\|^2 \|y\|^2 + \ldots + \|a_k\|^2 \|y\|^2} \tag{28}
\]

\[
< r\sqrt{k}\|y\|, \tag{29}
\]

for all \( y \in \mathbb{R}^{n-\ell} \), implying \( \|A\| < r\sqrt{k} \), where in (29) we used \( a_i \in B^{n-\ell}(0, r) \), for \( i = 1, \ldots, k \).
Finally, (33) is by application of the concentration of measure result in Lemma 1 below (for fixed, albeit arbitrarily large, $r$), where (30) follows from a union bound over the covering balls $B^a(x_i, \delta_j)$ of $S'$, in (32) we set $x_i = [y_i^T z_i^T]^T$ and used

$$||Ay_i + Bz_i|| = ||A B||x_i|| < ||A B||x_i - u|| + ||A B||u|| \leq c(k, r, ||B||) \delta_j + ||A B||u||.$$  

Finally, (33) is by application of the concentration of measure result in Lemma 1 below (for fixed, albeit arbitrarily large, $r$), where we used (24) to deduce that $y_i \neq 0$, which, in turn, allows us to absorb the term $1/||y_i||^k$ into the constant $C(n, k, r, ||B||)$. The convergence to 0 in (33) is a consequence of

$$\frac{\log \left( M_{S'}(\delta_j)\delta_j^k \right)}{\log \frac{1}{\delta_j}} = \frac{\log M_{S'}(\delta_j)}{\log \frac{1}{\delta_j}} - k \xrightarrow{j \to \infty} \dim_B(S') - k < 0,$$

where we used (25) together with $\dim_B(S') \leq \dim_B(S)$ thanks to $S' \subseteq S$. Since $\log \frac{1}{\delta_j} \xrightarrow{j \to \infty} \infty$, the convergence to a finite negative number in (34) implies that $\log \left( M_{S'}(\delta_j)\delta_j^k \right) \xrightarrow{j \to \infty} -\infty$ and hence $M_{S'}(\delta_j)\delta_j^k \xrightarrow{j \to \infty} 0$. This concludes the proof. 

Remark 5: A statement similar to Proposition 1 was proved in [14, Lem. 4.3]. Specifically, the result in [14, Lem. 4.3] applies to general linear combinations of Lipschitz mappings, covers the case $\dim_B(S) \geq k$ as well, and gives an upper bound on the lower Minkowski dimension of the set on the left hand side (LHS) in (16). In particular, for $\dim_B(S) < k$, the case considered here, the upper bound in [14, Lem. 4.3] implies (16). The proof of [14, Lem. 4.3] is based on concentration properties of the singular values of $H$ [14, Lem. 4.2]. Our proof is more direct, does not rely on specifics of the distribution of $H = [A B]$, but applies to $\dim_B(S) < k$ only, the case relevant here.

Proposition 1 shows that we can enforce uniqueness of the solution in the recovery of $y$ and $z$ from $w$ in (2) by requiring that the signal $[y^T z^T]^T$ 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 the concatenated
random source vector, 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. This concentration inequality will also turn out instrumental in the proof of Theorem 2.

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

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

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

\[
\mathbb{P}[\|\mathbf{A}\mathbf{u} + \mathbf{v}\| < \delta] = \frac{1}{\alpha(n, r)^k} \text{Leb}^k \{ \mathbf{A} \in B^n(0, r) \times \ldots \times B^n(0, r) \mid \|\mathbf{A}\mathbf{u} + \mathbf{v}\| < \delta \}
\]

\[
\leq \frac{1}{\alpha(n, r)^k} \prod_{i=1}^k \text{Leb}^n \{ \mathbf{a}_i \in B^n(0, r) \mid \|\mathbf{a}_i^T \mathbf{u} + \mathbf{v}_i\| < \delta \}
\]

\[
= \frac{1}{\alpha(n, r)^k} \prod_{i=1}^k \text{Leb}^n \left\{ \mathbf{U}\mathbf{a}_i \in B^n(0, r) \mid \|\mathbf{U}\mathbf{a}_i^T \mathbf{u} + \mathbf{v}_i\| < \delta \left\|\mathbf{u}\right\| \right\}
\]

\[
= \frac{1}{\alpha(n, r)^k} \prod_{i=1}^k \text{Leb}^n \left\{ \mathbf{a}_i \in B^n(0, r) \mid \left\|\mathbf{a}_i^T \mathbf{e}_1 + \mathbf{v}_i\right\| < \delta \left\|\mathbf{u}\right\| \right\}
\]

\[
\leq \frac{(2r)^{k(n-1)}}{\alpha(n, r)^k} \prod_{i=1}^k \text{Leb}^1 \left\{ \mathbf{a}_i \in \mathbb{R} \mid \left\|\mathbf{a}_i + \mathbf{v}_i\right\| < \delta \left\|\mathbf{u}\right\| \right\}
\]

\[
= \frac{(2r)^{k(n-1)}(2\delta)^k}{\alpha(n, r)^k\left\|\mathbf{u}\right\|^k},
\]

where (37) holds by the multiplicativity of Lebesgue measure and since \( \|\mathbf{A}\mathbf{u} + \mathbf{v}\| < \delta \) implies \( \|\mathbf{a}_i^T \mathbf{u} + \mathbf{v}_i\| < \delta \), for all \( i \), (38) follows from \( \mathbf{u} \neq 0 \), the fact that \( \text{Leb}^n \) is invariant under rotations, and we consider a rotation \( \mathbf{U}^T \) that takes \( \mathbf{u}/\|\mathbf{u}\| \) into \( \mathbf{e}_1 = [1 \ 0 \ldots \ 0]^T \). In (39) we relabel \( \mathbf{U}\mathbf{a}_i \in B^n(0, r) \) as \( \mathbf{a}_i \in B^n(0, r) \), and in (40) we denote the first component of the vector \( \mathbf{a}_i \) by \( a_i \), we relax the condition on the magnitude of \( a_i \) to \( a_i \in \mathbb{R} \), and we use the multiplicativity of Lebesgue measure together with the fact that the magnitudes of the remaining components of \( \mathbf{a}_i \) are less than or equal to \( r \). Finally, (41) follows by noting that in (40) we take the product over the Lebesgue measure of intervals of length \( 2\delta/\|\mathbf{u}\| \).

We finally note that the probabilistic uncertainty relation developed here is a quite general tool as it can also be applied to establish information-theoretic limits of matrix completion [18] and of phase.
IV. PROOF OF THEOREM

Since $R > R_B(\varepsilon) = \limsup_{n \to \infty} a_n(\varepsilon)$ and $k = \lfloor Rn \rfloor$, both by assumption, we have
\[
a_n(\varepsilon) < \frac{k}{n},
\]
which, together with the definition of $a_n(\varepsilon)$, implies that there exists a sequence of non-empty bounded sets $U := U_n \subseteq \mathbb{R}^n$ such that
\[
\dim_B(U) < k
\]
and
\[
P[\exists u \in U_x \setminus \{0\} : [A B]u = 0, x \in U] + \varepsilon.
\]
In the remainder of the proof we take $n$ to be sufficiently large for (43) to hold in the $\#$-sense and drop the dot-notation. We formally define the separator (sequence) based on the approximate support set sequence $U$ according to
\[
g(v) = \begin{cases} x, & \text{if } \{u \mid [A B]u = v\} \cap U = \{x\} \\ \text{error}, & \text{else.} \end{cases}
\]
The separator effectively looks for elements $u$ in $U$ that are consistent with the observation $v$ in the sense that $[A B]u = v$. If there is a unique such element in $U$, the separator returns that element. If there are multiple consistent elements in $U$ or none, the separator declares an error. More formally, we have
\[
P[g([A B]x) \neq x] = P[g([A B]x) \neq x, x \in U] + P[g([A B]x) \neq x, x \notin U]
\leq P[g([A B]x) \neq x, x \in U] + \varepsilon
= P[\exists u \in U_x \setminus \{0\} : [A B]u = 0, x \in U] + \varepsilon,
\]
where $U_x := \{v - x \mid v \in U\}$. (47) follows from (44), and in (48) we used the fact that for $x \in U$ there is at least one element in $U$ consistent with the observation, namely $x$, and the existence of $u \in U_x \setminus \{0\}$

1The symbol $U$ actually denotes the sequence $U_n$. We decided to drop the index $n$ for simplicity of exposition.

2Taking “error” to be an arbitrary element of $\mathbb{R}^n$, we obtain a measurable map $g: \mathbb{R}^k \to \mathbb{R}^{n-\ell} \times \mathbb{R}^\ell$ as required in Definition. Specifically, the measurability of $g$ follows by noting that $g$ is a right-inverse of $[A B]$ on $U$ and hence measurable by Sec. 7.2 and a constant (and hence measurable) mapping else. Note that “error” has to be chosen to be an element in $\mathbb{R}^n \setminus U$ as otherwise the step from (47) to (48) could be invalidated.
such that \([A \ B]u = 0\) is equivalent to the existence of multiple elements in \(\mathcal{U}\) consistent with the observation. Since (lower) Minkowski dimension is invariant under translation (as is seen by translating covering balls accordingly), we have \(\dim_{\text{B}}(\mathcal{U}_x) = \dim_{\text{B}}(\mathcal{U}) < k\). We can therefore apply the probabilistic uncertainty relation, Proposition 1, with \(S = \mathcal{U}_x\) and \(\dim_{\text{B}}(\mathcal{U}_x) < k\) to conclude that

\[
\text{Leb}^{k(n-\ell)}\{A \mid \exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0\} = 0, \tag{49}
\]

for all \(x \in \mathbb{R}^n\), i.e., \([A \ B]\) is injective on \(\mathcal{U}\), for a.a. \(A\). In order to show that this implies \(\mathbb{P}[\exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0] = 0\) for a.a. \(A \in \mathbb{R}^{k \times (n-\ell)}\), we employ the following “Fubini-maneuver”, an ingenious idea first suggested in [13] in the context of analog compression. First, by integrating (49) with respect to \(\mu_x(dx)\), we obtain

\[
\int_{\mathbb{R}^n} \text{Leb}^{k(n-\ell)}\{A \mid \exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0\} \mu_x(dx) = 0. \tag{50}
\]

On the other hand, we also have

\[
\begin{align*}
&\int_{\mathbb{R}^n} \text{Leb}^{k(n-\ell)}\{A \mid \exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0\} \mu_x(dx) \\
= &\int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^k} \mathbb{1}_{\{[A, x] \mid \exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0\}}(A, x) \, dA\right] \mu_x(dx) \tag{51} \\
= &\int_{\mathbb{R}^k} \left[\int_{\mathbb{R}^n} \mathbb{1}_{\{[A, x] \mid \exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0\}}(A, x) \, dA\right] \mu_x(dx) \tag{52} \\
= &\int_{\mathbb{R}^k} \mathbb{P}[\exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0] \, dA, \tag{53}
\end{align*}
\]

where in (53) we applied Fubini’s Theorem, whose conditions are satisfied, as we are integrating over \(\sigma\)-finite measure spaces and the indicator function \(\mathbb{1} (\cdot)\) is non-negative [21, Thm. 8.8 (a)], and (54) is by

\[
\mathbb{P}[x \in A] = \mu_x(A) = \int_{\mathbb{R}^n} \mathbb{1}_A(x) \mu_x(dx), \tag{55}
\]

for every Borel set \(A\). Combining (50) with (51)-(54), we get

\[
\int_{\mathbb{R}^k \times (n-\ell)} \mathbb{P}[\exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0] \, dA = 0, \tag{56}
\]

which by non-negativity of probability implies \(\mathbb{P}[\exists u \in \mathcal{U}_x \setminus \{0\} : [A \ B]u = 0] = 0\) for a.a. \(A \in \mathbb{R}^{k \times (n-\ell)}\), i.e., with probability 1 (with respect to drawing \(x\)) there is no \(v \in \mathcal{U} \setminus \{x\}\) resulting in the same observation as \(x\) under \([A \ B]\). When used in (48) this leads to

\[
\mathbb{P}[g([A \ B]x) \neq x] \leq \varepsilon,
\]

for a.a. \(A\), thereby completing the proof. ☐
V. REGULARIZED PROBABILISTIC UNCERTAINTY RELATION

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

Definition 5: For \( A \subseteq \mathbb{R}^n \), \( B \subseteq \mathbb{R}^m \), and \( \beta > 0 \), a map \( f: A \rightarrow B \) is \( \beta \)-Hölder continuous if there exists a constant \( c > 0 \) such that for all \( x_1, x_2 \in A \) we have

\[
\|f(x_1) - f(x_2)\| \leq c\|x_1 - x_2\|^{1/\beta}.
\]

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

\[
c\|x_1 - x_2\|^{1/\beta} \leq \|f(x_1) - f(x_2)\|, \quad \text{for all } x_1, x_2 \in A,
\]

if and only if \( f \) is injective and \( f^{-1}: f(A) \rightarrow A \) is \( \beta \)-Hölder continuous.

Proof: If (57) holds, then \( f \) is injective as for all \( x_1, x_2 \in A \) with \( x_1 \neq x_2 \), we have

\[
\|f(x_1) - f(x_2)\| \geq c\|x_1 - x_2\|^{1/\beta} > 0,
\]

and hence \( f(x_1) \neq f(x_2) \). Therefore, \( f^{-1}: f(A) \rightarrow A \) is well-defined. Moreover, for all \( y_1, y_2 \in f(A) \) we can find \( x_1, x_2 \in A \) such that \( f(x_i) = y_i \) and hence \( \beta \)-Hölder continuity of \( f^{-1} \) follows from

\[
\|f^{-1}(y_1) - f^{-1}(y_2)\| = \|x_1 - x_2\| \leq \frac{1}{c^{1/\beta}}\|f(x_1) - f(x_2)\|^{\beta} = \frac{1}{c^{1/\beta}}\|y_1 - y_2\|^{\beta},
\]

where the inequality is by (57).

Conversely, suppose that \( f \) is injective and \( f^{-1}: f(A) \rightarrow A \) is \( \beta \)-Hölder continuous. Then, for all \( x_1, x_2 \in A \), by \( \beta \)-Hölder continuity of \( f^{-1} \) there exists a constant \( C \) such that

\[
\|f^{-1}(f(x_1)) - f^{-1}(f(x_2))\| \leq C\|f(x_1) - f(x_2)\|^{\beta}.
\] (58)

Since \( f^{-1}(f(x_i)) = x_i \) by injectivity of \( f \), this implies (57) with \( c := 1/C^{1/\beta} \).

For a linear map \( f \), e.g., the map induced by a realization of the random matrix \( H = [A \ B] \), verifying (57) reduces to checking the condition

\[
c\|x\|^{1/\beta} \leq \|f(x)\|, \quad \text{for all } x \in A \odot A.
\] (59)

We next provide a sufficient condition—that is convenient to check—for (59) to hold. For expository simplicity, we formulate this condition for general sets \( S \) in place of \( A \odot A \). In fact, in the proof of Theorem 2 we will verify (59) for \( S = A \odot \{a\} \) in place of \( A \odot A \), for fixed elements \( a \in A \), and employ a Fubini-maneuver similar to the one in the proof of Theorem 1. The sufficient 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 bounded set in $\mathbb{R}^n$, $f: \mathcal{S} \to \mathbb{R}^k$, $\beta \in (0,1)$, and $\delta_j := 2^{-j}$. If there is a $J \in \mathbb{N}$ such that for all $j \geq J$ we have

$$\|f(x)\| \geq \delta_j, \quad \text{for all } x \in \mathcal{S} \setminus B^n(0, \delta_j^\beta), \quad (60)$$

then there exists a constant $c(\mathcal{S}) > 0$ such that

$$c(\mathcal{S})\|x\|^{1/\beta} \leq \|f(x)\|, \quad \text{for all } x \in \mathcal{S}. \quad (61)$$

**Proof:** Let $x \in \mathcal{S}$, $\mathcal{S}_j := \mathcal{S} \setminus B^n(0, \delta_j^\beta)$, and $i_x := \min\{i \in \mathbb{N} \mid x \in \mathcal{S}_i\}$ (see Figure 1 for an illustration). We then find that
∥f(x)∥ \geq \begin{cases} 
\delta_{i_x}, & \text{if } i_x \geq J \\
\delta_{J}, & \text{if } i_x < J 
\end{cases} 
\geq \begin{cases} 
\|x\|^{1/\beta}/2, & \text{if } i_x \geq J \\
2^{-J} \sup_{u \in S} \|u\|^{1/\beta}, & \text{if } i_x < J 
\end{cases} 
\geq c(S)\|x\|^{1/\beta},
(62)
(63)
(64)

where (62) follows, for \(i_x \geq J\), from \(x \in S_i\), together with (60); and for \(i_x < J\), from \(x \in S_J\) and (60) with \(j = J\). In (63) we used \(\|x\| < 2^{-(i_x-1)\beta}\), for \(i_x \geq J\), and for \(i_x < J\) we apply the trivial lower bound \(\|x\| \leq \sup_{u \in S} \|u\|\). Finally, in (64) we set \(c(S) = \min \left\{ \frac{1}{2}, \frac{2^{-J}}{\sup_{u \in S} \|u\|^{1/\beta}} \right\}\), and we note that \(c(S) > 0\) by virtue of \(S\) 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 a fixed \(B\) and a.a. \(A\) there are no non-zero vectors in \(S\) that map to zero under \([A \ B]\) provided that the lower Minkowski dimension of \(S\) is sufficiently small. The regularized version of this result states that the norm of the image of a vector \(x \in S\) under \([A \ B]\) does not become too small relative to \(\|x\|\). This will then allow us to deduce the existence of a Hölder continuous separator in Theorem 2 by applying Lemma 2.

Proposition 2: Let \(B \in \mathbb{R}^{k \times \ell}\), with \(k \geq \ell\), have \(\text{rank}(B) = \ell\), let \(S \subseteq \mathbb{R}^{n-\ell} \times \mathbb{R}^\ell\) be non-empty and bounded, and fix \(\beta \in (0, 1)\) such that

\[1 - \frac{\dim_b(S)}{k} \geq \beta.\]

Then, for a.a. \(A\) there exists a constant \(c(S, A, B) > 0\) such that

\[c(S, A, B) \left\| \begin{bmatrix} y \\
z \end{bmatrix} \right\|^{1/\beta} \leq \|Ay + Bz\|,
(65)
(66)

for all \([y^T \ z^T]^T \in S\), where \(y \in \mathbb{R}^{n-\ell}\) and \(z \in \mathbb{R}^\ell\).

Proof: As in the proof of Proposition 1 we show that for the random matrix \(A = [a_1 \ldots a_k]^T\), with the \(a_i\) i.i.d. uniform on \(B^{n-\ell}(0, r)\) with arbitrary \(r > 0\), with probability 1 there exists a constant \(c(S, A, B) > 0\) satisfying (66) for all \([y^T \ z^T]^T \in S\). Since \(r\) can be taken arbitrarily large, this establishes that the Lebesgue measure of matrices \(A\) for which there is no such constant \(c(S, A, B) > 0\) is zero. As (66) is an inequality of the form (61) we can apply Lemma 3 with \(f = [A \ B]\) and \(x = [y^T \ z^T]^T\) to conclude that the proof can be established by showing the following: Setting \(\delta_j := 2^{-j}\) and \(S_j :=\)
$S \setminus B^n(0, \delta_j)$, with probability 1 (with respect to $A$) there exists a $J \in \mathbb{N}$ such that

$$\|Ay + Bz\| \geq \delta_j, \quad \text{for all } [y^T z^T]^T \in S_j,$$  (67)

and all $j \geq J$. Applying the Borel-Cantelli Lemma [22, Thm. 2.3.1] to the complementary events it follows that it suffices to show that

$$\sum_{j=0}^{\infty} \mathbb{P}[\exists [y^T z^T]^T \in S_j : \|Ay + Bz\| < \delta_j] < \infty. $$  (68)

The basic idea for establishing (68) is to cover $S_j$ with balls of radius $\delta_j$ and to upper-bound the probabilities in (68) by probabilities that are exclusively in terms of the corresponding covering ball centers. Specifically, with the minimum number of balls of radius $\delta_j$ needed to cover $S_j$ denoted by $M_j := M_{S_j}(\delta_j)$ and the corresponding ball centers $x_1^{(j)}, \ldots, x_{M_j}^{(j)} \in S_j$, we establish that

$$\mathbb{P}[\exists [y^T z^T]^T \in S_j : \|Ay + Bz\| < \delta_j] \leq \sum_{i=1}^{M_j} \mathbb{P}[\|Hx_i^{(j)}\| < (L + 1)\delta_j]. $$  (69)

Here $L := c(k, r, \|B\|)$ is the constant in (26) with $u = [y^T z^T]^T$ and $[A \ B] = H$. To prove (69), first note that the existence of an $x = [y^T z^T]^T \in S_j$ such that $\|Hx\| < \delta_j$ implies that $x \in B^n(x_{i_0}^{(j)}, \delta_j)$ for some $i_0 \in \{1, \ldots, M_j\}$, since the balls $B^n(x_{i_0}^{(j)}, \delta_j)$, $i = 1, \ldots, M_j$, cover $S_j$. It then follows that

$$\|Hx_{i_0}^{(j)}\| \leq \|Hx\| + \|H(x_{i_0}^{(j)} - x)\|$$  (70)

$$< \delta_j + L\delta_j = (L + 1)\delta_j, $$  (71)

where we used $\|Hx\| < \delta_j$ and (26). From (70), (71), and a union bound argument we then get (69). We now turn to bounding the terms on the RHS of (69) and will then use these bounds in (68) to establish the final result. Let us start by splitting the covering ball centers

$$x_i^{(j)} = \begin{bmatrix} y_i^{(j)} \\ z_i^{(j)} \end{bmatrix},$$

into two groups

$$\{x_1^{(j)}, \ldots, x_{M_j}^{(j)}\} = \mathcal{X}_1^{(j)} \cup \mathcal{X}_2^{(j)}, $$  (72)

with

$$\mathcal{X}_1^{(j)} := \{x_i^{(j)} \mid \|y_i^{(j)}\| < c\|z_i^{(j)}\|\} $$  (73)

$$\mathcal{X}_2^{(j)} := \{x_i^{(j)} \mid \|y_i^{(j)}\| \geq c\|z_i^{(j)}\|\}. $$  (74)
where the constant \( c > 0 \) will be chosen below. The reasoning behind this splitting is as follows. For ball centers in \( \Lambda_1^{(j)} \), we establish that the corresponding probabilities on the RHS of (69) equal zero for sufficiently large \( j \), whereas for ball centers in \( \Lambda_2^{(j)} \), we use the concentration inequality in Lemma 1 to establish that the corresponding terms in the sum on the RHS of (69) are sufficiently small to result in a finite upper bound as required in (68). We first note that for all ball centers we have

\[
\|x_i^{(j)}\|^2 = \|y_i^{(j)}\|^2 + \|z_i^{(j)}\|^2 \geq \delta_j^{2\beta},
\]

(75)

by virtue of \( x_i^{(j)} \in S_j \). We now turn to the set \( \Lambda_1^{(j)} \). From (73) we get

\[
(c^2 + 1)\|z_i^{(j)}\|^2 \geq \|y_i^{(j)}\|^2 + \|z_i^{(j)}\|^2 \geq \delta_j^{2\beta}.
\]

(76)

This allows us to deduce that

\[
\|Hx_i^{(j)}\| \geq \|Bz_i^{(j)}\| - \|Ay_i^{(j)}\|
\]

(77)

\[
\geq CB\|z_i^{(j)}\| - \|A\|\|y_i^{(j)}\|
\]

(78)

\[
\geq CB\|z_i^{(j)}\| - c\|z_i^{(j)}\|
\]

(79)

\[
\geq \frac{CB - cL}{\sqrt{1 + c^2}} \delta_j^{\beta},
\]

(80)

where in (77) we applied the reverse triangle inequality, for (78) we note that there exists a constant \( C_B > 0 \) such that \( \|Bz\| \geq C_B\|z\| \) for all \( z \in \mathbb{R}^\ell \) as a consequence of \( B \) being full-rank, by assumption, in (79) we used \( x_i^{(j)} \in \Lambda_1^{(j)} \), and in (80) we employed (76) and \( \|A\| = \sup_{\|y\|=1}\|Ay\| = \sup_{\|y^Tz\|_r^r=1}\|Ay + Bz\| \leq \sup_{\|y^Tz\|_r^r=1}\|Ay\| + Bz\| = \|[A\ B]\| \leq L \), where \( L \) was defined right after (69). Since \( \beta \in (0, 1) \), \( \delta_j^{\beta} \) can be made arbitrarily large relative to \( \delta_j \) (i.e., \( \delta_j^{\beta}/\delta_j = 2^{\beta(1-\beta)} \) can be made arbitrarily large) by taking \( j \) sufficiently large. Specifically, choosing \( c > 0 \) such that \( C_B - cL > 0 \), we can find a \( J_1 \in \mathbb{N} \) such that

\[
\frac{CB - cL}{\sqrt{1 + c^2}} \delta_j^{\beta} \geq (L + 1)\delta_j, \quad \text{for all } j \geq J_1.
\]

(81)

By (80) this implies

\[
\|Hx_i^{(j)}\| \geq (L + 1)\delta_j,
\]

(82)

for all \( j \geq J_1 \), and hence establishes that

\[
\mathbb{P}[\|Hx_i^{(j)}\| < (L + 1)\delta_j] = 0,
\]

(83)

\(^3\)This is possible since \( C_B/L > 0 \).
for \( x_i^{(j)} \in X_1^{(j)} \) and \( j \geq J_1 \).

Next, consider \( x_i^{(j)} \in X_2^{(j)} \). From (74) we get
\[
\left(1 + \frac{1}{c^2}\right) \|y_i^{(j)}\|^2 \geq \|y_i^{(j)}\|^2 + \|z_i^{(j)}\|^2 \geq \delta_j^{2\beta},
\]
which, using the concentration inequality Lemma 1, allows us to conclude that
\[
P[\|Ay_i^{(j)} + Bz_i^{(j)}\| < (L + 1)\delta_j] \leq C(n, k, r) \frac{(L + 1)^k \delta_j^k}{\|y_i^{(j)}\|^k}
\]
\[
\leq C(n, k, r, L) \left(\sqrt{1 + \frac{1}{c^2}}\right)^k \frac{2^{-j\beta}}{2^{-\beta jk}}.
\]

Putting things together, we obtain
\[
\sum_{j=0}^{\infty} P[\exists y^T z^T \in S_j : \|Ay + Bz\| < \delta_j] \leq J_1 + \sum_{j=J_1}^{\infty} \sum_{i=1}^{M_j} P[\|Hx_i^{(j)}\| < (L + 1)\delta_j]
\]
\[
= J_1 + \sum_{j=J_1}^{\infty} \sum_{i=1}^{M_j} \sum_{x_i^{(j)} \in X_2^{(j)}} C(n, k, r, L) \left(\sqrt{1 + \frac{1}{c^2}}\right)^k \frac{2^{-j\beta}}{2^{-\beta jk}}
\]
\[
\leq J_1 + C(n, k, r, L, c) \sum_{j=J_1}^{\infty} M_j 2^{-j(1-\beta)}
\]
\[
\leq J_1 + C(n, k, r, L, c, S) \sum_{j=J_1}^{\infty} 2^{d'j} 2^{-j(1-\beta)}
\]
\[
= J_1 + C(n, k, r, L, c, S) \sum_{j=J_1}^{\infty} 2^{-j(1-\frac{d'}{K}-\beta)}
\]
\[
< \infty.
\]

Here, in (87) we upper-bound the probability of the terms for \( j < J_1 \) by 1, where \( J_1 \) was defined in (81), (88) is by (69), (89) follows from (83), in (90) we invoked (86), and (91) holds since \( |X_2^{(j)}| \leq M_j \).

For (92), we set \( d' = \dim_B(S) + \alpha \) with \( \alpha > 0 \) small enough so that \( 1 - \frac{d'}{K} > \beta \), which is possible by
(65), and we used

\[ M_j = M_S(\delta_j) \leq N_S(\delta_j/2) \]
\[ \leq N_S(\delta_j/2) \]
\[ \leq C(S)\delta_j^{-d'}, \quad (95) \]
\[ \leq S_j \subseteq S, \] for all \( j \in \mathbb{N} \), where (95) follows from a triangle inequality argument (cf. (144)), (96) holds as \( S_j \subseteq S \), for all \( j \), and (97) is a consequence of

i) \( \dim_B(S) < d' \) and thus \( N_S(\delta_j/2) \leq (\delta_j/2)^{-d'} = 2^{d}d_j^{-d'} \) for sufficiently large \( j \) by definition of \( \limsup \), and

ii) we take \( C(S) \) to be sufficiently large so that (97) also holds for the (finite number of) \( j \)'s for which \( N_S(\delta_j/2) \leq (\delta_j/2)^{-d'} \) does not hold. Note that \( C(S) \) is guaranteed to be finite as the set \( S \) is bounded and therefore the covering numbers \( N_S(\delta_j/2) \) are finite for all \( j \).

Finally, (94) follows from

\[ 1 - \frac{d'}{k} > \beta, \quad (98) \]
which is by choice of \( d' \). This completes the proof.

\[ \square \]

VI. PROOF OF THEOREM 2

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

\[ 1 - \frac{R_B(\varepsilon) + \alpha}{R} > \beta. \quad (99) \]

Let \( k' := (R_B(\varepsilon) + \alpha)n \). By definition of \( R_B(\varepsilon) \), we can find a sequence of non-empty bounded sets \( U \subseteq \mathbb{R}^n \) such that

\[ \overline{\dim}_B(U) < k' \]
\[ \text{and} \quad \mathbb{P}[x \in U] \geq 1 - \varepsilon. \quad (100) \]

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

$$1 - \frac{k'}{k} = \frac{\lfloor Rn \rfloor - (\overline{R_B}(\varepsilon) + \alpha)n}{\lfloor Rn \rfloor} \quad (102)$$

$$\geq \frac{\lfloor Rn \rfloor - (\overline{R_B}(\varepsilon) + \alpha)n}{Rn} \quad (103)$$

$$> \frac{R - \frac{1}{n} - R_B(\varepsilon) - \alpha}{R} \quad (104)$$

$$> \beta, \quad (105)$$

where (103) follows from $\lfloor Rn \rfloor \leq Rn$ and $\lfloor Rn \rfloor - (\overline{R_B}(\varepsilon) + \alpha)n > (R - \frac{1}{n})n - (\overline{R_B}(\varepsilon) + \alpha)n > 0$ (since $R > \overline{R_B}(\varepsilon) + \alpha$, by choice of $\alpha$), in (104) we used $Rn - 1 < \lfloor Rn \rfloor$, and in (105) we invoked (99). In the remainder of the proof, we take $n$ sufficiently large for (100)–(105) to hold in the $\#$-sense and drop the dot-notation.

Next, we employ the regularized probabilistic uncertainty relation, to find a constant $c(U_x, A, B) > 0$, where $U_x := \{v - x \mid v \in U\}$ for fixed $x \in \mathbb{R}^n$, such that

$$c(U_x, A, B) \|v - x\|^{1/\beta} \leq \|Hv - Hx\|, \quad \text{for all } v \in U.$$

For $H = [A \ B]$, where $A \in \mathbb{R}^{k \times (n-\ell)}$ and $B \in \mathbb{R}^{k \times \ell}$, with $k \geq \ell$, and all $x \in \mathbb{R}^n$, we set\(^5\)

$$c(U_x, A, B) := \inf_{u \in U_x \setminus \{0\}} \frac{\|Hu\|}{\|u\|^{1/\beta}} \quad (106)$$

Since (upper) Minkowski dimension is invariant under translation (as is seen by translating covering balls accordingly), we have $\overline{\dim_B}(U_x) = \overline{\dim_B}(U)$. Using (100) together with (102)–(105) it follows that

$$1 - \frac{\overline{\dim_B}(U_x)}{k} > 1 - \frac{k'}{k} > \beta. \quad (106)$$

Application of the regularized probabilistic uncertainty relation, Proposition 2, to (106) with $H = [A \ B]$, $u = [y^T \ z^T]^T$, and $S = U_x$ then implies that

$$c(U_x, A, B) > 0, \quad (107)$$

for all $x \in \mathbb{R}^n$ and a.a. $A$. Our goal is to find a universal (with respect to $x$) constant $c(U, A, B) > 0$ such that

$$c(U_x, A, B) \geq c(U, A, B) \quad (108)$$

holds for all $x \in U$, as this will allow us to apply Lemma 2 to conclude that $H^{-1}: H(U) \to U$ is $\beta$-Hölder continuous. As we will be able to establish only $c(U_x, A, B) \neq 0$ almost surely, we will have

\(^5\)To avoid the case $c(U_x, A, B) = \infty$, we assume that the non-empty set $U$ is not a singleton, i.e., it does not consist of a single point $U = \{u\}$ only. If $U$ is a singleton, finding a $\beta$-Hölder continuous separator $g$ satisfying (12) is trivial as every $\beta$-Hölder continuous map $g: \mathbb{R}^k \to \mathbb{R}^n$ with $g(Hu) = u$ is a valid separator.
to restrict \(108\) to hold for \(x\) from a subset \(U'\) of \(U\). It turns out, however, that a valid \(U' \subseteq U\) can be found by excising only a small part of \(U\) such that \(\mathbb{P}[x \in U'] \geq 1 - \varepsilon - \kappa\).

We start by employing a Fubini-maneuver akin to that used in the proof of Theorem 1, where we showed that for a.a. \(A\), with probability 1 (with respect to drawing \(x\)) there is no \(v \in U \setminus \{x\}\) resulting in the same observation as \(x\) under \([A, B]\). The difference to the way the Fubini-maneuver is applied in the proof of Theorem 1 is that this uniqueness property is replaced by requiring that there exist a positive lower bound on \(c(U_x, A, B)\) which is universal with respect to \(x\). We first conclude from the argument around (107) that

\[
\text{Leb}^{k(n-\ell)}\{A \mid c(U_x, A, B) = 0\} = 0, \tag{109}
\]

for all \(x \in \mathbb{R}^n\). Integrating (109) with respect to \(\mu_x(dx)\), we obtain

\[
\int_{\mathbb{R}^n} \text{Leb}^{k(n-\ell)}\{A \mid c(U_x, A, B) = 0\} \mu_x(dx) = 0. \tag{110}
\]

On the other hand, we also have

\[
\int_{\mathbb{R}^n} \text{Leb}^{k(n-\ell)}\{A \mid c(U_x, A, B) = 0\} \mu_x(dx) \tag{111}
\]

\[
= \int_{\mathbb{R}^n} \left[ \int_{\mathbb{R}^{k \times (n-\ell)}} 1\{\tilde{A}, \tilde{x} \mid c(U_{x}, \tilde{A}, B) = 0\}(A, x) dA \right] \mu_x(dx) \tag{112}
\]

\[
= \int_{\mathbb{R}^{k \times (n-\ell)}} \left[ \int_{\mathbb{R}^n} 1\{\tilde{A}, \tilde{x} \mid c(U_{x}, \tilde{A}, B) = 0\}(A, x) \mu_x(dx) \right] dA \tag{113}
\]

\[
= \int_{\mathbb{R}^{k \times (n-\ell)}} \mathbb{P}[c(U_x, A, B) = 0] dA, \tag{114}
\]

where in (113) we applied Fubini's Theorem, whose conditions are satisfied, as we are integrating over \(\sigma\)-finite measure spaces and the indicator function \(1(\cdot)\) is non-negative [21, Thm. 8.8 (a)]. Combining (110) with (111)--(114), we can therefore conclude that

\[
\mathbb{P}[c(U_x, A, B) = 0] = 0, \tag{115}
\]

for a.a. \(A\). In order to get a universal (with respect to \(x \in U'\) for a properly chosen subset \(U'\) of \(U\)) constant \(c(U, A, B) > 0\) satisfying (108), we note that (115) implies

\[
\mathbb{P}\left[ x \in U, c(U_x, A, B) > \frac{1}{j} \right] \xrightarrow{j \to \infty} \mathbb{P}[x \in U] \geq 1 - \varepsilon, \tag{116}
\]

for a.a. \(A\), and therefore, for every \(\kappa > 0\), there exists a \(J \in \mathbb{N}\) such that

\[
\mathbb{P}\left[ x \in U, c(U_x, A, B) > \frac{1}{J} \right] \geq 1 - \varepsilon - \kappa. \tag{117}
\]
With $\mathcal{U} := \{x \in \mathcal{U} \mid c(\mathcal{U}_x, A, B) > \frac{1}{J}\}$, we therefore have $\mathbb{P}[x \in \mathcal{U}'] \geq 1 - \varepsilon - \kappa$ and the universal constant $c(\mathcal{U}, A, B) := 1/J$ satisfying

$$
\frac{1}{J} \|v - x\|^{1/\beta} \leq \|Hv - Hx\|, \text{ for all } v, x \in \mathcal{U},
$$

for a.a. $A$ in $H = [A \ B]$. By Lemma 2, $H$ is therefore injective on $\mathcal{U}'$ and its inverse $H^{-1} : H(\mathcal{U}') \to \mathcal{U}'$ is $\beta$-Hölder continuous, and by [23, Thm. 1, ii)] (restated below for completeness) with $\mathcal{V} = \mathbb{R}^k$, $\mathcal{W} = \mathbb{R}^n$, and $A = H(\mathcal{U}')$ the inverse $H^{-1}$ can be extended to a $\beta$-Hölder continuous mapping $g_H : \mathbb{R}^k \to \mathbb{R}^n$.

Finally, thanks to injectivity of $H$ on $\mathcal{U}'$ we have $g_H(Hx) = x$ for all $x \in \mathcal{U}'$, and therefore

$$
\mathbb{P}[g_H(Hx) \neq x] \leq \mathbb{P}[x \notin \mathcal{U}'] < \varepsilon + \kappa.
$$

This completes the proof.

Finally, for the reader’s convenience, we provide the following (reformulated) version of the statement [23, Thm. 1, ii)].

**Theorem 3:** Let $\mathcal{V}, \mathcal{W}$ be Euclidean spaces and let $g : A \to \mathcal{W}$ be $\beta$-Hölder continuous with $0 < \beta < 1$ and $A \subseteq \mathcal{V}$. Then, $g$ can be extended to a $\beta$-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 notion of sparsity as used, e.g., in [3], [7], [12], [24], [25]. Specifically, we consider concatenated source vectors $x$ with independent entries of mixed discrete-continuous distribution and possibly different mixture parameters for the constituent processes $(Y_i)_{i \in \mathbb{N}}$ and $(Z_i)_{i \in \mathbb{N}}$.

**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 $i \in \{1, \ldots, n\}$ are independent and distributed according to

$$
\mu_{X_i} = \begin{cases} 
(1 - \rho_1)\mu_{d_i} + \rho_1\mu_{c_1}, & i \in \{1, \ldots, n - \ell\} \\
(1 - \rho_2)\mu_{d_2} + \rho_2\mu_{c_2}, & i \in \{n - \ell + 1, \ldots, n\},
\end{cases}
$$

(119)

where $0 \leq \rho_i \leq 1$ are mixture parameters, $\ell = \lfloor \lambda n \rfloor$, the $\mu_{d_i}$ are discrete distributions, and the $\mu_{c_i}$ are absolutely continuous (with respect to 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 $\mu_{d_1}, \mu_{d_2}$ are Dirac measures at 0, i.e., $\mu_{d_1} = \mu_{d_2} = \delta_0$, then the non-zero entries of $x$ can be generated only by the continuous parts $\mu_{c_1}, \mu_{c_2}$. With

$$\tilde{Y}_i := 1_{\mathbb{R}\setminus\{0\}}(Y_i), \quad i = 1, \ldots, n - \ell,$$

$$\tilde{Z}_i := 1_{\mathbb{R}\setminus\{0\}}(Z_i), \quad i = n - \ell + 1, \ldots, n,$$

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

$$\frac{1}{n} \left( \sum_{i=1}^{n-\ell} \tilde{Y}_i + \sum_{i=n-\ell+1}^{n} \tilde{Z}_i \right). \quad (120)$$

Letting $n \to \infty$ in (120), we obtain

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

where we used $(n - \ell)/n = (n - \lfloor \lambda n \rfloor)/n \xrightarrow{n \to \infty} (1 - \lambda)$ as $(1 - \lambda)n \leq n - \lfloor \lambda n \rfloor < (1 - \lambda)n + 1$, similarly $\ell/n = \lfloor \lambda n \rfloor/n \xrightarrow{n \to \infty} \lambda$ as $\lambda n - 1 < \lfloor \lambda n \rfloor \leq \lambda n$, and

$$\frac{1}{n - \ell} \sum_{j=1}^{n-\ell} \tilde{Y}_j \xrightarrow{\mathbb{P}} \rho_1 \quad (121)$$

$$\frac{1}{\ell} \sum_{j=n-\ell+1}^{n} \tilde{Z}_j \xrightarrow{\mathbb{P}} \rho_2, \quad (122)$$

by the weak law of large numbers and $\mathbb{E}[\tilde{Y}_i] = \mathbb{P}[\tilde{Y}_i = 1] = \rho_1$, $\mathbb{E}[\tilde{Z}_i] = \mathbb{P}[\tilde{Z}_i = 1] = \rho_2$. This shows that the fraction of non-zero entries in the concatenated source vector $x$ converges—in probability—to $(1 - \lambda)\rho_1 + \lambda\rho_2$. The next result establishes that the Minkowski dimension compression rate $R_B(\varepsilon)$ 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

$$R_B(\varepsilon) = (1 - \lambda)\rho_1 + \lambda\rho_2, \quad (123)$$

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

December 4, 2015 DRAFT
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, Sect. 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 $H = [A \ B] : \mathbb{R}^{n-\ell} \times \mathbb{R}^\ell \rightarrow \mathbb{R}^k$ and a corresponding measurable separator $g : \mathbb{R}^k \rightarrow \mathbb{R}^{n-\ell} \times \mathbb{R}^\ell$, with $k = \lfloor Rn \rfloor$, such that
\[
P[g([A \ B]x) \neq x] \leq \varepsilon,
\]
implies $R \geq R_B(\varepsilon)$.

Proof: The proof does not have to account for the fact that $H = [A \ B]$ contains a fixed block $B$ 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, Sect. III].

Combining the achievability statements in Theorems 1 and 2, and Propositions 3 and 4, we can conclude that, for mixed discrete-continuous sources,
\[
R_B(\varepsilon) = (1 - \lambda)\rho_1 + \lambda\rho_2,
\]
is the critical rate in the following sense:

- For $R > R_B(\varepsilon)$, for every fixed full-rank matrix $B \in \mathbb{R}^{k \times \ell}$, with $k \geq \ell$, and for a.a. $A$ (where the set of exceptions for $A$ depends on the specific choice of $B$) there exists a measurable separator $g$ satisfying (124), as well as a $\beta$-Hölder continuous separator $g$ for fixed $\beta$ with $\beta < 1 - \frac{\tau_0(\varepsilon)}{R}$ satisfying (124) with $\varepsilon$ replaced by $\varepsilon + \kappa$ for arbitrarily small $\kappa > 0$,
- for $R = R_B(\varepsilon)$, we cannot make a general statement on the existence of a separator,
- and for $R < R_B(\varepsilon)$ there does not exist a single pair $(g, \ [A \ B])$ satisfying (124).

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

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 the Elad-Bruckstein [15] uncertainty principles hold for all $y$ and $z$, 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 $A = I_n$ and $B = F_n$ [12].
Relaxing these deterministic thresholds by considering random models for the signals and dictionaries [10], [24]–[26] leads to thresholds that exhibit a “log $n$-factor”. The threshold that follows from our probabilistic uncertainty relations excludes an arbitrarily small set of signals, a set of $A$-matrices of Lebesgue measure zero, is asymptotic in $n$, and suffers neither from the square-root bottleneck nor from the $\log n$-factor. Moreover, it is best possible as the same threshold would be obtained if the support sets of $y$ and $z$ were known and the values of the non-zero entries in the concatenated source vector only were to be recovered. The set of exceptions for $A$ in the “a.a.-statement” in Theorems \[1\] and \[2\] depending on the specific choice of $B$ can be interpreted as a mild incoherence condition between $A$ and $B$ akin to those in \[12\], \[15\]. In fact, we have a phase transition phenomenon, which states that above the critical rate $R_B(\varepsilon)$ a.a. matrices $A$ are “incoherent” to a given matrix $B$, whereas below the critical rate there is not a single pair of matrices $A$ and $B$ that admits separation via a measurable separator $g$. 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 $x$ the Minkowski dimension compression rate is small if the asymptotic fraction $(1 - \lambda)\rho_1 + \lambda\rho_2$ of non-zero entries in $x$ 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 $y$ and $z$. For example, consider the declipping problem \[3\] where a signal that is sparse in the dictionary $A$ is to be recovered from its clipped version. Specifically, we observe $Ay + z$ with $z = g_a(Ay) - Ay$, where $g_a$ denotes entry-wise clipping to the values $\pm a$. If clipping is not too aggressive, the signal $z$ will be sparse in the identity basis $B = I_\ell$ (see Fig. \[2\]). Here $y$ and $z$ are of the same dimension, i.e., $\lambda = 1/2$. Moreover, $z$ is completely determined by $y$, which, as proved in Lemma \[4\] in Appendix \[A\] implies that

$$R_B^x(\varepsilon) = \frac{1}{2}R_B^y(\varepsilon),$$

where $R_B^y(\varepsilon)$ is the Minkowski dimension compression rate of the concatenated source vector $x = [y^T \ z^T]^T$ and $R_B^y(\varepsilon)$ is the Minkowski dimension compression rate of $y$ only. If the components of $y$ are i.i.d. and of discrete-continuous mixture $(1 - \rho)\mu_d + \rho\mu_c$, it follows from Proposition \[3\] that $R_B^y(\varepsilon) = \rho$ and consequently

$$R_B^x(\varepsilon) = \frac{1}{2}\rho.$$  \hspace{1cm} (126)

The description complexity of the concatenated source vector $x$ is therefore determined by the fraction of components in $y$ 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 $y$ and $z$ and $\rho_1 = \rho_2 = \rho$.\hspace{1cm}
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 together with the Fubini-maneuver can be applied to devise a simplification of the proof of [13, Thm. 18, 1]) and a significant improvement of the statement [13, Thm. 18, 2]). We begin by restating [13, Thm. 18, 1]) in our notation and terminology.

Theorem ([13, Thm. 18, 1]) Let \( x = [X_1 \ldots X_n]^T \) be a source vector of dimension \( n \) with underlying stochastic source process \((X_i)_{i \in \mathbb{N}}\) on \((\mathbb{R}^N, \mathcal{B}^\otimes \mathbb{N})\). Take \( \varepsilon > 0 \) and let \( R > R_B(\varepsilon) \). Then, for a.a. \( H \in \mathbb{R}^{k \times n} \), there exists a measurable decoder \( g \) such that

\[
P[g(Hx) \neq x] \leq \varepsilon,
\]

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 \( R > R_B(\varepsilon) \) in [13, Thm. 18, 1]) is replaced by \( R > R_B(\varepsilon) \). For our simplified proof, we start by particularizing the probabilistic uncertainty relation,
Proposition 1 to $\ell = 0$.

**Corollary 1:** Let $S \subseteq \mathbb{R}^n$ be non-empty and bounded such that $\dim_B(S) < k$. Then, we have

$$\{ u \in S \setminus \{0\} \mid Au = 0 \} = \emptyset,$$

(127)

for Lebesgue a.a. $A \in \mathbb{R}^{k \times n}$.

We refer to this result as a probabilistic null-space property as it is a statement on the intersection of the null-space of $A$ with the set $S$. Our alternative, simplified proof of [13, Thm. 18, 1] goes as follows. As in the proof of Theorem 1 we choose a set sequence $U \subseteq \mathbb{R}^n$ satisfying (43) and (44). We define the decoder according to

$$g(v) = \begin{cases} 
  x, & \text{if } \{ u \mid Hu = v \} \cap U = \{ x \} \\
  \text{error}, & \text{else}
\end{cases}$$

(128)

The probability of a decoding error is decomposed as in (46) with $H$ in place of $[A \ B]$. Applying Corollary 1 we find that a.a. matrices $H$ induce mappings that are injective on $U$. Finally, invoking the Fubini-maneuver as in the argument leading to (53) allows us to conclude that the probability of a decoding error is zero when $x \in U$, leaving the total error probability to be smaller than $\varepsilon$ and thereby finishing 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], formulated in our notation and terminology.

**Theorem ([13 Thm. 18, 2]):** Let $x = [X_1 \ldots X_n]^T$ be a source vector of dimension $n$ with underlying stochastic process $(X_i)_{i \in \mathbb{N}}$ on $(\mathbb{R}^N, \mathcal{B}^{\otimes \mathbb{N}})$. Take $\varepsilon > 0$, and let $R > \bar{R}_B(\varepsilon)$ and $\beta > 0$ be fixed such that

$$\beta < 1 - \frac{\bar{R}_B(\varepsilon)}{R}.$$  

(129)

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

$$\mathbb{P}[g(Hx) \neq x] \leq \varepsilon + \kappa,$$

(130)

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 $H$ with a corresponding $g$ satisfying (130), whereas our

Measurability of $g$ follows by the same arguments as in the proof of Theorem 1.
result says that for almost all $H$ there is a corresponding $g$ satisfying (130). 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 $z = g_a(Ay) - Ay$, with $g_a$ denoting entry-wise clipping to the values $\pm a$ for some $a > 0$. We introduce the notation

$$R_B^x(\varepsilon), a_B^x(\varepsilon), R_B^a_x(\varepsilon), a_B^a_x(\varepsilon)$$

$$R_B^y(\varepsilon), a_B^y(\varepsilon), R_B^a_y(\varepsilon), a_B^a_y(\varepsilon)$$

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

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

$$R_B^x(\varepsilon) = \frac{1}{2} R_B^y(\varepsilon) \quad \text{and} \quad R_B^a_x(\varepsilon) = \frac{1}{2} R_B^a_y(\varepsilon). \quad (131)$$

Proof: We only prove the first identity and note that the second is obtained by simply replacing $\dim_B(\cdot)$ by $\dim_B(\cdot)$ in the arguments below. Let us begin by showing

$$R_B^x(\varepsilon) \leq \frac{1}{2} R_B^y(\varepsilon). \quad (132)$$

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

$$T := \{[y^T (g_a(Ay) - Ay)^T]^T | y \in S \} \subseteq \mathbb{R}^n, \quad (133)$$

and note that for all $y_1, y_2 \in \mathbb{R}^{n-\ell}$ we have

$$\|y_1 - y_2\| \leq ||[y_1^T (g_a(Ay_1) - Ay_1)^T]^T - [y_2^T (g_a(Ay_2) - Ay_2)^T]^T|| \quad (134)$$

and

$$\|y_1 - y_2\| + \|g_a(Ay_1) - Ay_1 - (g_a(Ay_2) - Ay_2)\| \leq (1 + \|A\|)\|y_1 - y_2\| + \|g_a(Ay_1) - g_a(Ay_2)\| \quad (136)$$

$$\|y_1 - y_2\| \leq (1 + 2\|A\|)\|y_1 - y_2\|, \quad (138)$$

where (134), (136), and (137) follow from the triangle inequality, and (138) holds as $|g_a(y_1) - g_a(y_2)| \leq |y_1 - y_2|$ for all $y_1, y_2 \in \mathbb{R}$. Combining (134) and (136)–(138), it follows that for $\delta > 0$

$$N_S(\delta) \leq N_T(\delta) \leq N_S((1 + 2\|A\|)\delta), \quad (139)$$

December 4, 2015 DRAFT
which implies \( \dim_B(S) = \dim_B(T) \). Since \( 2(n - \ell) = 2(n - \lfloor n/2 \rfloor) \leq n + 2 \), we obtain
\[
\frac{1}{2} \dim_B(S) \leq \frac{1}{2} \dim_B(T) \geq \dim_B(T) = \frac{n + 2}{n + 2} = \frac{n}{n + 2}.
\]
and therefore \( \frac{1}{2} \alpha^X_{n-\ell}(\varepsilon) \geq \frac{n}{n + 2} \alpha^X_n(\varepsilon) \). As \( n/(n + 2) \xrightarrow{n \to \infty} 1 \), we get (132).

To prove
\[
R^X_B(\varepsilon) \geq \frac{1}{2} R^Y_B(\varepsilon),
\]
we consider a set \( \mathcal{U} \subseteq \mathbb{R}^n \) such that \( P[x \in \mathcal{U}] \geq 1 - \varepsilon \). Setting \( \mathcal{V} := \{ y \in \mathbb{R}^n \mid [y^T (g_a(Ay) - Ay)^T] \in \mathcal{U} \} \), we have \( P[y \in \mathcal{V}] = P[x \in \mathcal{U}] \). For the set \( \tilde{\mathcal{U}} := \{ [y^T (g_a(Ay) - Ay)^T] \mid y \in \mathcal{V} \} \) we have \( \dim_B(\tilde{\mathcal{U}}) = \dim_B(\mathcal{V}) \) by the same arguments as in (134)–(139). Moreover, by definition of \( \mathcal{V} \) we have \( \tilde{\mathcal{U}} \subseteq \mathcal{U} \) and therefore \( \dim_B(\tilde{\mathcal{U}}) \leq \dim_B(\mathcal{U}) \), which implies
\[
\frac{1}{2} \dim_B(\mathcal{V}) \leq \frac{1}{2} \dim_B(\mathcal{U}) \leq \frac{1}{2} \dim_B(\mathcal{U}) \leq \frac{1}{2} \dim_B(\mathcal{U}),
\]
where in the last step we used \( 2(n - \ell) \geq n \). This shows that \( \alpha^X_n(\varepsilon) \geq \frac{1}{2} \alpha^Y_{n-\ell}(\varepsilon) \) which establishes (140) 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) are presented in [17, Equivalent definitions 3.1].

**Lemma 5:** The Minkowski dimension of a non-empty bounded set \( S \subseteq \mathbb{R}^n \) does not change when the covering balls in the definitions (3), (4) are restricted to have their centers inside the set \( S \), that is, we have
\[
\liminf_{\delta \to 0} \frac{\log N_S(\delta)}{\log \frac{1}{\delta}} = \liminf_{\delta \to 0} \frac{\log M_S(\delta)}{\log \frac{1}{\delta}},
\]
\[
\limsup_{\delta \to 0} \frac{\log N_S(\delta)}{\log \frac{1}{\delta}} = \limsup_{\delta \to 0} \frac{\log M_S(\delta)}{\log \frac{1}{\delta}},
\]
where \( N_S(\delta) \) is the covering number of \( S \) given by
\[
N_S(\delta) = \min \left\{ m \in \mathbb{N} \mid S \subseteq \bigcup_{i \in \{1, \ldots, m\}} B^n(x_i, \delta), \ x_i \in \mathbb{R}^n \right\},
\]
and \( M_S(\delta) \) is the covering number of \( S \) with the covering balls centered in \( S \), i.e.,
\[
M_S(\delta) = \min \left\{ m \in \mathbb{N} \mid S \subseteq \bigcup_{i \in \{1, \ldots, m\}} B^n(x_i, \delta), \ x_i \in S \right\}.
\]
Proof: Since $N_S(\delta) \leq M_S(\delta)$, the “$\leq$”-part in (141) and (142) is immediate. To establish the “$\geq$”-part, we consider a set of covering balls of $S$ of radius $\delta/2$ and corresponding centers $x_1, \ldots, x_{N_S(\delta/2)} \in \mathbb{R}^n$. Note that these centers do not necessarily lie in $S$. Since $N_S(\delta/2)$ is the minimum number of balls with radius $\delta/2$ needed to cover $S$, the intersection $B^n(x_i, \delta/2) \cap S$ must be non-empty for all $i = 1, \ldots, N_S(\delta/2)$. We now choose an arbitrary point $y_i \in (B^n(x_i, \delta/2) \cap S)$ and note that
\[ \|u - y_i\| \leq \|u - x_i\| + \|x_i - y_i\| \leq \|u - x_i\| + \delta/2, \quad \text{for all } u \in \mathbb{R}^n, \] (143)
which implies $B^n(x_i, \delta/2) \subseteq B^n(y_i, \delta)$, $i = 1, \ldots, N_S(\delta/2)$. It therefore follows that $B^n(y_i, \delta)$, $i = 1, \ldots, N_S(\delta/2)$, is a covering of $S$ with balls of radius $\delta$ all centered in $S$. This implies
\[ M_S(\delta) \leq N_S(\delta/2), \] (144)
and hence
\[ \frac{\log M_S(\delta)}{\log \frac{1}{\delta}} \leq \frac{\log N_S(\delta/2)}{\log \frac{1}{\delta}} = \frac{\log N_S(\delta/2)}{\log \frac{2}{\delta}} \frac{\log 2}{\log 1}. \] (145)
Taking $\liminf_{\delta \to 0}$ and $\limsup_{\delta \to 0}$ on both sides of (145) yields the “$\geq$”-part in (141) and (142), respectively, noting that
\[ \liminf_{\delta \to 0} \frac{\log N_S(\delta/2)}{\log \frac{1}{\delta}} = \liminf_{\delta \to 0} \frac{\log N_S(\delta)}{\log \frac{1}{\delta}}, \]
\[ \limsup_{\delta \to 0} \frac{\log N_S(\delta/2)}{\log \frac{1}{\delta}} = \limsup_{\delta \to 0} \frac{\log N_S(\delta)}{\log \frac{1}{\delta}}, \]
as replacing $\delta$ by $\delta/2$ simply amounts to a reparametrization. \hfill \blacksquare

REFERENCES


