Almost Lossless Analog Compression without Phase Information

Erwin Riegler
ETH Zurich, 8092 Zurich, Switzerland
eriegler@nari.ee.ethz.ch

Georg Tauböck
Austrian Academy of Sciences, 1040 Vienna, Austria
georg.tauboeck@oeaw.ac.at

Abstract—We propose an information-theoretic framework for phase retrieval. Specifically, we consider the problem of recovering an unknown vector \( x \in \mathbb{R}^n \) up to an overall sign factor from \( m = \lceil Rn \rceil \) phaseless measurements with compression rate \( R \) and derive a general achievability bound for \( R \). Surprisingly, it turns out that this bound on the compression rate is the same as the one for almost lossless analog compression obtained by Wu and Verdú (2010): Phaseless linear measurements are “as good” as linear measurements with full phase information in the sense that ignoring the sign of \( m \) measurements only leaves us with an ambiguity with respect to an overall sign factor of \( x \).

I. INTRODUCTION

In many different areas of science, physical limitations make it impossible to measure the sign (phase in the complex case) of a signal but obtaining amplitudes is relatively easy. Well known examples are X-ray crystallography, astronomy, or diffraction imaging \([1]–[3]\). The problem of retrieving a well known example is X-ray crystallography, astronomy, make it impossible to measure the sign (phase in the complex form \( [x] = |x| e^{-i\arg(x)} \)).

The mapping \( \mathbb{R}^n \rightarrow |Ax| \) is injective without imposing structural assumptions on \( A \). In \([5]\), the authors showed that at least \( 2n-1 \) such measurements are necessary and generically sufficient to guarantee injectivity. Furthermore, it was shown that semidefinite programming can be used to recover \( |x| \) if \( A \) is random with i.i.d. Gaussian entries or with i.i.d. rows that are uniformly distributed on a sphere, as long as \( m \geq c_0 n \) for a sufficiently large constant \( c_0 \) \([6]\). Other phase retrieval methods for which theoretical performance guarantees are available can be found, e.g., in \([7]–[10]\).

Recently, there has been also interest in sparse phase retrieval, where the number \( s \) of nonzero coefficients of the vector \( x \) is much smaller than \( n \). This a-priori knowledge about \( x \) can be used to reduce the number of measurements significantly. For instance, \( O(s \log(n/s)) \) measurements were shown to be sufficient for stable sparse phase retrieval \([11]\).

If the rows of the measurement matrix \( A \) are a generic choice of vectors in \( \mathbb{R}^n \), injectivity of the mapping \( |x| \mapsto |Ax| \) is guaranteed provided that \( m \geq 2s \) \([12]\).

Contributions: Following the approach introduced for compressed sensing \([13]\) and signal separation \([14]\) problems, we formulate phase retrieval as an analog source coding problem. Assuming that the unknown vector \( x \) is random with a certain distribution, we derive asymptotic recovery results for \( |x| \). Our results hold for Lebesgue almost all (a.a.) measurement matrices \( A \). However, our results are in terms of probability of error (with respect to the distribution of \( x \)) and hence do not provide worst-case guarantees. Specifically, we study the asymptotic setting \( n \rightarrow \infty \) where the vector \( x \) is a realization of a random process; for each \( n \), we let \( m = \lceil Rn \rceil \) for a parameter \( R \), which we denote compression rate. In Theorem 1 we show that we can recover \( |x| \) from \( m \) phaseless measurements with arbitrarily small probability of error for a.a. measurement matrices \( A \). Provided that \( n \) is sufficiently large and the compression rate \( R \) is larger than the (lower) Minkowski dimension compression rate (see Definition 3) of \( x \). It is remarkable that the obtained result is identical to the corresponding result in compressive sensing \([13]\) where \( y = Ax \), so that we can conclude that in terms of achievability results, phaseless linear measurements are “as good” as linear measurements with full phase information: Ignoring the sign of \( m \) measurements only leaves us with an ambiguity with respect to an overall sign factor of \( x \).

Notation: Roman letters \( A, B, \ldots \) and \( a, b, \ldots \) designate deterministic matrices and vectors, respectively. Boldface letters \( \mathbf{A}, \mathbf{B}, \ldots \) and \( \mathbf{a}, \mathbf{b}, \ldots \) denote random matrices and random vectors, respectively. For the distribution of a random matrix \( \mathbf{A} \) and a random vector \( \mathbf{u} \), we write \( \mu_{\mathbf{A}} \) and \( \mu_{\mathbf{u}} \), respectively. The \( k \)-th component of the vector \( \mathbf{u} \) (random vector \( \mathbf{u} \)) is \( u_k \) (\( u_k \)). The superscript \( ^T \) stands for transposition. For a matrix \( \mathbf{A} \), \( \text{tr}(\mathbf{A}) \) denotes its trace. The identity matrix of suitable size is denoted by \( I \). For a vector \( \mathbf{u} \), we write \( \|\mathbf{u}\| = \sqrt{\mathbf{u}^T \mathbf{u}} \) for its Euclidean norm. For the Euclidean space \( \mathbb{R}^k \), \( \|\cdot\| \) denotes the open ball of radius \( r \) centered at \( u \in \mathbb{R}^k \) by \( B_k(u,r) \).
\( u \sim v \) means that either \( u = v \) or \( u = -v \) and we write for the corresponding equivalence classes \( [u] = \{ u \} \cup \{-u\} \). For a set \( S \subseteq \mathbb{R}^k \), \( \mathcal{N}_S = \{ [u] : u \in S \} \). The indicator function on a set \( U \) is denoted by \( \chi_U \).

II. MAIN RESULTS

We start by formulating phase retrieval as a source coding problem.

Definition 1. (Source vector) Let \((x_i)_{i \in \mathbb{N}}\) be a stochastic process on \((\mathbb{R}^N, \mathcal{S}_R^N)\). Then, for \( n \in \mathbb{N} \), the source vector \( x \) of length \( n \) is given by \( x = (x_1, \ldots, x_n)^T \in \mathbb{R}^n \).

Definition 2. (Code, achievable rate) For \( x \) as in Definition 1 and \( \varepsilon > 0 \), an \((n, m)\) code consists of

(i) measurements \( |A \cdot x| : \mathbb{R}^n \rightarrow \mathbb{R}^m \);

(ii) a decoder \( g : \mathbb{R}^m_+ \rightarrow \mathbb{R}^n \) that is measurable with respect to \( \mathcal{B}_R^m \) and \( \mathcal{B}_R^n \).

We call \( R \) with \( 0 < R \leq 1 \) an \( \varepsilon \)-achievable rate if there exists an \( N(\varepsilon) \in \mathbb{N} \) and a sequence of \((n, [Rn])\) codes with decoders \( g \) such that

\[ P[g(|A x|) \neq x] \leq \varepsilon \]

for all \( n \geq N(\varepsilon) \).

Next, we introduce the Minkowski dimension compression rate for source vectors.

Definition 3. (Minkowski dimension) Let \( U \) be a nonempty bounded set in \( \mathbb{R}^n \). The lower Minkowski dimension of \( U \) is defined as

\[ \dim_0(U) = \liminf_{\rho \rightarrow 0} \frac{\log N_U(\rho)}{\log \frac{1}{\rho}} \]

and the upper Minkowski dimension of \( U \) is defined as

\[ \dim_1(U) = \limsup_{\rho \rightarrow 0} \frac{\log N_U(\rho)}{\log \frac{1}{\rho}} \]

where \( N_U(\rho) \) is the covering number of \( U \) given by

\[ N_U(\rho) = \min \left\{ k \in \mathbb{N} \mid U \subseteq \bigcup_{i=1}^{k} B_n(u_i, \rho), \, u_i \in \mathbb{R}^n \right\} \]

If \( \dim_0(U) = \dim_1(U) \), we write \( \dim(U) \).

Definition 4. (Minkowski dimension compression rate) For \( x \) from Definition 1 and \( \varepsilon > 0 \), we define the lower Minkowski dimension compression rate as

\[ R_0(\varepsilon) = \limsup_{n \rightarrow \infty} \frac{\dim_0(U)}{n} \]

where

\[ \dim_0(U) = \inf \left\{ \frac{\dim_0(U)}{n} \mid U \subseteq \mathbb{R}^n, \, P[x \in U] \geq 1 - \varepsilon \right\} \]

and the upper Minkowski dimension compression rate as

\[ \overline{R}_0(\varepsilon) = \limsup_{n \rightarrow \infty} \frac{\dim_1(U)}{n} \]

where

\[ \dim_1(U) = \inf \left\{ \frac{\dim_1(U)}{n} \mid U \subseteq \mathbb{R}^n, \, P[x \in U] \geq 1 - \varepsilon \right\} \]

The sets \( U \) in the definitions for \( \underline{m}_n(\varepsilon) \) and \( \overline{m}_n(\varepsilon) \) are assumed to be nonempty and bounded.

Example 1. The source vector \( x \) from Definition 1 has a mixed discrete-continuous distribution if for each \( n \in \mathbb{N} \) the random variables \( x_i, \, i \in \{1, \ldots, n\} \), are independent and distributed according to

\[ \mu_{x_i} = (1 - \lambda) \mu_d + \lambda \mu_c, \quad i \in \{1, \ldots, n\} \]

where \( 0 \leq \lambda \leq 1 \) is the mixing parameter, \( \mu_c \) is a distribution on \((\mathbb{R}, \mathcal{B}_R)\), absolutely continuous with respect to Lebesgue measure, and \( \mu_d \) is a discrete distribution. Then, [13 Th. 15]

\[ \overline{R}_0(\varepsilon) = \overline{R}_1(\varepsilon) = \lambda, \quad 0 < \varepsilon < 1 \]

The following result states that every rate \( R > \overline{R}_0(\varepsilon) \) is \( \varepsilon \)-achievable for Lebesgue a.a. matrices \( A \).

Theorem 1. Let \( 0 < \varepsilon < 1 < \lambda \) and \( x \) as in Definition 1, then, for Lebesgue a.a. matrices \( A \in \mathbb{R}^{n \times n} \) with \( m = [Rn] \), \( R \) is an \( \varepsilon \)-achievable rate provided that \( R > \overline{R}_0(\varepsilon) \).

Proof. Since \( R > \overline{R}_0(\varepsilon) \) and \( m = [Rn] \), Definition 4 implies that there exists a sequence of nonempty bounded sets \( U_n \subseteq \mathbb{R}^n \) and a \( N(\varepsilon) \in \mathbb{N} \) such that

\[ \dim_0(U_n) < m \]

(1)

\[ P[x \in U_n] \geq 1 - \varepsilon \]

(2)

for all \( U = U_n \) with \( n \geq N(\varepsilon) \). In the remainder of the proof we assume that \( n \) is sufficiently large for (1) and (2) to hold. The claim now follows from Proposition 1 below.

Proposition 1. Let \( \varepsilon \geq 0 \), \( x \in \mathbb{R}^n \) a random vector, and \( U \subseteq \mathbb{R}^n \) a nonempty bounded set with \( P[x \in U] \geq 1 - \varepsilon \). Then, for Lebesgue a.a. matrices \( A \in \mathbb{R}^{m \times n} \), there exists a decoder \( g \) with \( P[g(|A x|) \neq x] \leq \varepsilon \) provided that \( \dim_0(U) < m \).

Proof. See Section III.

Remark 1. By [15 Sec. 3.2, Properties (i)-(iii)], the lower Minkowski dimension of any bounded nonempty subset in \( \mathbb{R}^n \) containing only vectors with no more than \( s \) nonzero entries is at most \( s \). Therefore, Proposition 1 implies that any \( s \)-sparse random vector \( x \in \mathbb{R}^n \) can be recovered with arbitrarily small probability of error (by increasing the size of the set \( U \) in Proposition 1), provided that \( m > s \). This result holds for an arbitrary distribution of \( x \) and a.a. matrices \( A \in \mathbb{R}^{m \times n} \). The best known recovery threshold for deterministic \( s \)-sparse vectors is \( m \geq 2s \) [12].

Remark 2. It is worth noting that formally phase retrieval can be formulated as a matrix completion problem with measurements \( y_i = 1(\text{a}_i x x^T) \) using rank-one measurement matrices \( A_i = a_i^T x x^T \), \( i = 1, \ldots, m \). However, compared to the rank-one measurement matrices used in the matrix completion problem [16, 17], the matrices \( a_i^T x \) are symmetric. This complicates the proof of Proposition 1 significantly and forces us to develop a novel concentration of measure result (Lemma 3). On the other hand, in phase retrieval we are interested in recovering symmetric rank-one matrices \( x x^T \) (which is
equivalent to the recovery of \([x]\), whereas matrix completion deals with the recovery of arbitrary low-rank matrices.

In the mixed discrete-continuous case we can strengthen the result of Theorem 1 through the following lemma.

**Lemma 1.** Let \(0 < \epsilon < 1\) and \(x\) be distributed according to the mixed discrete-continuous distribution in Example 2 with mixing parameter \(\lambda\). Then, for Lebesgue a.a. matrices \(A \in \mathbb{R}^{m \times n}\) with \(m = \lfloor Rn \rfloor\), \(R\) is \(\epsilon\)-achievable provided that \(R > \lambda\). Moreover, \(R \geq \lambda\) is also a necessary condition for \(R\) being \(\epsilon\)-achievable.

**Proof.** Achievability: Follows from Theorem 1 and Example 2. Converse: Suppose that a rate \(R < \lambda\) is \(\epsilon\)-achievable for some \(\epsilon\) with \(0 < \epsilon < 1\). This implies that there exists a set \(K \subseteq [n]\) and a matrix \(A \in \mathbb{R}^{m \times n}\) with \(m = \lfloor Rn \rfloor\) such that (a) \(P[\bar{x} \in K] \geq 1 - \epsilon\); (b) \(|A|\) is one-to-one on \(K\).

for \(n\) sufficiently large. From (b) it follows that there can be at most one equivalence class \([u]\) in \(K\). with \(Au = \Lambda(-u) = 0\). Suppose first that there is no equivalence class \([u]\) in \(K\). with \(Au = \Lambda(-u) = 0\) and \(u \neq 0\). Then, (b) implies that \(\Lambda\) is one-to-one on \(K\) which, together with (a) and \(R < \lambda\), leads to a contradiction to the converse part of Theorem 6.

Now suppose that there is an equivalence class \([u]\) and \(u\) is an equivalence class with \(Au = \Lambda(-u) = 0\) and \(u \neq 0\). Let \(\tilde{R}\) be such that \(R < \tilde{R} < \lambda\) and set \(\tilde{n} = \lfloor \tilde{R}n \rfloor\). Then, \(\tilde{n} > m\) for \(n\) sufficiently large. Let \(\tilde{A} = (A^T, u, 0, \ldots, 0)^T \in \mathbb{R}^{m \times n}\). Then, (b) implies that \(\tilde{A}\) is one-to-one on \(K\) which, together with (a) and \(R < \lambda\), leads to a contradiction to the converse part of Theorem 6.

**III. PROOF OF PROPOSITION 1**

Let

\[
\mathcal{F}(y) = \{u \in \mathbb{R}^n | u \in U, |Au| = y\} \\
\cup \{u \in \mathbb{R}^n | u \notin U, |Au| = y\}, \quad y \in \mathbb{R}_+^m.
\]

For a vector \(u \in \mathcal{F}(y) \setminus \{0\}\), let \(\bar{u}\) denote the first nonzero component of \(u\). We then define the reduced set

\[
\hat{\mathcal{F}}(y) = \{u \in \mathcal{F}(y) \setminus \{0\} | |u| = |\bar{u}|\} \cup \{\mathcal{F}(y) \cap \{0\}\}, \quad y \in \mathbb{R}_+^m.
\]

We define the decoder \(g: \mathbb{R}_+^m \to \mathbb{R}^n\) by

\[
g(y) = \begin{cases} u, & \text{if } \hat{\mathcal{F}}(y) = \{u\} \\ e, & \text{else} \end{cases}
\]

where \(e\) is some fixed vector in the complement of \(U\) (used to declare a decoding error). Then, we have

\[
P[g(|Ax|) \neq x] = P[g(|Ax|) \neq x, x \in U] + P[g(|Ax|) \neq x, x \notin U] \\
\leq P[g(|Ax|) \neq x, x \in U] + \epsilon \\
= P[\exists u \in U | u \neq x, |Au| = |Ax|, x \in U] + \epsilon
\]

where (3) follows from the definition of the decoder. Fix an arbitrary \(r > 0\). Suppose that we can show that

\[
P(x) = P[\exists u \in U | u \neq x, |Au| = |Ax|] = 0, \quad x \in U
\]

where \(A \in \mathbb{R}^{m \times n}\) has independent rows that are uniformly distributed on \(B_n(0, r)\). Then,

\[
\int_{A(r)} P[\exists u \in U | u \neq x, |Au| = |Ax|, x \in U] d\mu_A \\
= \int_{A(r)} P[\exists u \in U \text{ with } u \neq x, |Au| = |Ax|] d\mu_x \\
= 0
\]

where we used Fubini’s Theorem and set \(A(r) = B_n(0, r) \times \ldots \times B_n(0, r)\). Since \(r\) is arbitrary, (3) implies that

\[
P[\exists u \in U | u \neq x, |Au| = |Ax|, x \in U] = 0
\]

for Lebesgue a.a. matrices \(A\). Hence, combining (3) and (6) proves the Proposition provided that we can show that (4) holds, which is done in Section IV.

**IV. PROOF OF (4)**

Suppose first that \(x = 0\). Then, \(P(x) = 0\) if and only if

\[
P[\exists u \in U \setminus \{0\} | u \neq 0] = 0.
\]

Since \(\text{dim}_0(U) < m\), (7) follows from [14, Prop. 1]. Therefore, we can assume in what follows that \(x \neq 0\).

We can upper-bound \(P(x) \leq P_1(x) + P_2(x)\) with

\[
P_1(x) = P[\exists u \in U | x, |Au| = |Ax|], \quad i \in \{1, 2\}
\]

where we defined

\[
U_i(x) = \{u \in U | \text{rank}(x, u) = 2\} \\
U_2(x) = \{u \in U | \text{rank}(x, u) = 1\} \setminus \{u \in U | u \sim x\}
\]

We have to show that \(P_i(x) = 0\) for \(i \in \{1, 2\}\). First, we establish \(P_2(x) = 0\). We have (recall that \(x \neq 0\))

\[
P_2(x) = P[\exists u \in U \text{ with } \text{rank}(x, u) = 1, u \neq x, |Au| = |Ax|] \\
= P[|Ax| = 0] \\
= 0
\]

where we used [14, Prop. 1] together with \(\text{dim}_0(\{x\}) = 0\) in the last step. It remains to show that \(P_1(x) = 0\). To this end, we first present an auxiliary lemma.

**Lemma 2.** Let \(r > 0, \emptyset \neq S \subseteq B_n(0, L), \rho > 0, x \in B_n(0, L), \text{ and } A \in \mathbb{R}^{m \times n}\) with independent rows that are uniformly distributed on \(B_n(0, r)\). Then, there exist \(s_i(\rho) \in S_i\), \(l = 1, \ldots, N_S(\rho)\) with \(N_S(\rho)\) being the covering number of \(S\), such that

\[
P[\exists u \in S \text{ with } |Au| - |Ax| \leq \rho] \\
\leq \sum_{l=1}^{N_S(\rho)} P[|X^T s_l(\rho)|^2 \leq 2Lr(2r + 1)\rho]^{m}
\]

(8)
where \( \mathbf{a} \) is uniformly distributed on \( \mathcal{B}_n(0, r) \).

**Proof.** Let \( \mathcal{S} \subseteq \bigcup_{l=1}^{N(\rho)} \mathcal{B}_n(v_l(\rho), \rho) \), \( v_l(\rho) \in \mathbb{R}^n \), be a minimal covering of \( \mathcal{S} \) according to the definition of the covering number, cf. Definition 3. Then, there exist \( s_l(\rho) \in \mathcal{S} \cap \mathcal{B}_n(v_l(\rho), 2\rho) \) for all \( l = 1, ..., N(\rho) \). Hence, the balls \( \mathcal{B}_n(s_l(\rho), 2\rho) \) cover the set \( \mathcal{S} \) and have centers in \( \mathcal{S} \). We can upper bound the lhs in (9) by

\[
P[\exists u \in \mathcal{S} \mid ||\mathbf{Ax}|| \leq \rho] \
\leq \sum_{l=1}^{N(\rho)} P[\exists u \in \mathcal{S} \cap \mathcal{B}_n(s_l(\rho), 2\rho) \mid ||\mathbf{Ax}|| \leq \rho] \
\leq \sum_{l=1}^{N(\rho)} P[\exists u \in \mathcal{S} \cap \mathcal{B}_n(s_l(\rho), 2\rho) \mid ||\mathbf{a}^T\mathbf{u}|| \leq \rho] \]

where (9) follows from the fact that the rows of \( \mathbf{A} \) are independent and uniformly distributed on \( \mathcal{B}_n(0, r) \). Using the triangle inequality we obtain

\[
||\mathbf{a}^T s_l(\rho)|| - ||\mathbf{a}^T x|| \leq ||\mathbf{a}^T x|| - ||\mathbf{a}^T u|| + ||\mathbf{a}^T u|| - ||\mathbf{a}^T s_l(\rho)||.
\]

The second term on the rhs of (10) can be further upper bounded by

\[
||\mathbf{a}^T u|| - ||\mathbf{a}^T s_l(\rho)|| \leq ||\mathbf{a}^T (u - s_l(\rho))|| \
\leq ||\mathbf{a}|| \cdot ||u - s_l(\rho)|| \leq 2\rho
\]

where (11) follows from \( u \in \mathcal{B}_n(s_l(\rho), 2\rho) \) and \( \mathbf{a} \in \mathcal{B}_n(0, r) \). Combining (10) and (11) gives

\[
||\mathbf{a}^T x|| - ||\mathbf{a}^T u|| \geq ||\mathbf{a}^T s_l(\rho)|| - ||\mathbf{a}^T x|| \leq 2\rho
\]

Using (12) in (9) yields

\[
P[\exists u \in \mathcal{S} \mid ||\mathbf{Ax}|| \leq \rho] \
\leq \sum_{l=1}^{N(\rho)} P[||\mathbf{a}^T s_l(\rho)|| - ||\mathbf{a}^T x|| \leq (2\rho + 1)\rho] \]

where (13) follows from \( ||\mathbf{a}^T s_l(\rho)|| - ||\mathbf{a}^T x|| \leq ||\mathbf{a}^T s_l(\rho)|| + ||\mathbf{a}^T s_l(\rho)|| - ||\mathbf{a}^T x|| \leq 2\rho ||\mathbf{a}^T s_l(\rho)|| - ||\mathbf{a}^T x|| \).

We now continue with the proof of \( P_1(x) = 0 \). Since \( \mathcal{U} \) is a bounded set, there exists an \( \lambda \in \mathbb{R} \) such that

\[
||u|| \leq \lambda, \quad u \in \mathcal{U}.
\]

We define the sets \( \mathcal{T}_j(x) \) by

\[
\mathcal{T}_j(x) = \left\{ u \in \mathcal{U}_1(x) \mid \sqrt{||u||^2 - ||\mathbf{a}^T x||^2} > 1/j \right\}, \quad j \in \mathbb{N}.
\]

Since \( P_1(x) \leq \sum_{j \in \mathbb{N}} P[\exists u \in \mathcal{T}_j(x) \mid ||\mathbf{Ax}|| = \mathbf{A}x] \)

it is sufficient to prove that

\[
P_1^{(j)}(x) = P[\exists u \in \mathcal{T}_j(x) \mid ||\mathbf{Au}|| = ||\mathbf{Ax}|| = 0
\]

for all \( j \in \mathbb{N} \). Suppose, by contradiction, that there exists \( j \in \mathbb{N} \) such that \( P_1^{(j)}(x) > 0 \). Then

\[
\liminf_{\rho \to 0} \frac{\log P_1^{(j)}(x)}{\log \frac{1}{\rho}} = 0.
\]

Furthermore, \( \mathcal{T}_j(x) \neq \emptyset \) and by [15 Sec. 3.2, Property (ii)] (recall that \( \mathcal{T}_j(x) \subseteq \mathcal{U}_1(x) \subseteq \mathcal{U} \) ) we get

\[
\dim_\mathbb{B}(\mathcal{T}_j(x)) < m.
\]

We have

\[
\liminf_{\rho \to 0} \frac{\log P_1^{(j)}(x)}{\log \frac{1}{\rho}} = \liminf_{\rho \to 0} \frac{\log P[\exists u \in \mathcal{T}_j(x) \mid ||\mathbf{Au}|| = ||\mathbf{Ax}||]}{\log \frac{1}{\rho}} = \liminf_{\rho \to 0} \frac{\log \left( \sum_{l=1}^{N\mathcal{T}_j(x)(\rho)} \left[ ||\mathbf{a}^T s_l^{(j)}(\rho, x)||^2 - ||\mathbf{a}^T x||^2 \right] \right)^m}{\log \frac{1}{\rho}}
\]

where in (17) we applied Lemma 2 with \( \mathcal{S} = \mathcal{T}_j(x) \) and set \( \tilde{\rho} = 2\lambda (2\rho + 1)\rho \), (18) follows from Lemma 3 below with \( u = s_l^{(j)}(\rho, x) \), \( \mathbf{v} = \mathbf{x} \), and \( \delta = \tilde{\rho} \) where \( f \) is defined in (22). In (19) we used that

\[
f(\tilde{\rho}, \mathbf{v}, s_l^{(j)}(\rho, x), x) \leq \tilde{f}(\tilde{\rho}, \mathbf{v}, L, j) = \frac{2(2\rho)^{n-2}j}{V(n, r)} \left( 1 + \log \left( 2 + \frac{8r^2L^2}{\rho} \right) \right), \quad l = 1, ..., N_{\mathcal{T}_j(x)}(\rho)
\]

which follows from (14) and the fact that \( s_l^{(j)}(\rho, x) \in \mathcal{T}_j(x) \), \( l = 1, ..., N_{\mathcal{T}_j(x)}(\rho) \), and in (20) we applied (16). But (20) is a contradiction to (15). Therefore, \( P_1^{(j)}(x) = 0 \) for all \( j \in \mathbb{N} \), which implies in turn that \( P_1(x) = 0 \) and concludes the proof of (11).

**V. CONCENTRATION OF MEASURE RESULT**

**Lemma 3.** Let \( r > 0 \), \( \mathbf{a} \) be uniformly distributed on \( \mathcal{B}_n(0, r) \), \( \mathbf{C} = \mathbf{u}^T \mathbf{u}^T \) with linearly independent vectors \( \mathbf{u}, \mathbf{v} \in \mathbb{R}^n \), and \( \delta > 0 \). Then

\[
P[||\mathbf{A}\mathbf{C}|| \leq \delta] \leq \delta f(\delta, \rho, \mathbf{u}, \mathbf{v})
\]
with
\[
f(\delta, r, u, v) = \\
\frac{2(2r)^{n-2}}{\sqrt{\|u\|^2\|v\|^2 - |u^T v|^2 V(n, r)}} \left(1 + \log \left(2 + 2^{2n} - \|u\| - \|v\|\right)\right)
\]
(22)

**Proof.** We have
\[
P(|a^T Ca| \leq \delta) \leq \frac{1}{V(n, r)} \int_{B_2(0, r)} X\{a \in \mathbb{R}^n \mid |a^T Ca| \leq \delta\} da
\]
(23)
\[
\leq \frac{1}{V(n, r)} \int_{B_2(0, r)} X\{b \in \mathbb{R}^n \mid |c^T WRJR^T W^T a| \leq \delta\} db
\]
(24)
\[
\leq \frac{1}{V(n, r)} \int_{B_2(0, r)} X\{c \in \mathbb{R} \mid |c^T R^T c| \leq \delta\} dc
\]
(25)
where (23) follows from Lemma 4 with R and J defined in (32) and W defined in (33) and (24) follows from changing variables to \(a = Wb\) with \(W = (W, Z) \in \mathbb{R}^{n \times n}\) where \(Z \in \mathbb{R}^{n \times (n-2)}\) is chosen in such a way that \(W^T W = I\) and \(c = c_1 c_2^T\) with \(c_1 = b_1\) and \(c_2 = b_2\).

The bound (36) on the determinant of the matrix \(RJR^T\) implies that one eigenvalue of \(RJR^T\), say \(\lambda_1\), is positive and the other eigenvalue of \(RJR^T\), say \(-\lambda_2\), is negative. We can assume without loss of generality that \(\lambda_1 \geq \lambda_2\). Using the eigendecomposition \(RJR^T = U \text{diag}(\lambda_1, -\lambda_2) U^T\), where \(U \in \mathbb{R}^{2 \times 2}\) with \(U^T = U\), and changing variables to \(c = Ud\), we can further upper bound (25) by
\[
\frac{(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} X\{t \in \mathbb{R}^2 \mid |t^T R^T c| \leq \delta\} dt
\]
(26)
\[
\leq \frac{(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} X\{s \in \mathbb{R}^2 \mid \lambda_1 t_1^2 - \lambda_2 t_2^2 \leq \delta^2 \}
\times X\{t \in \mathbb{R}^2 \mid |t^T - t_1| \leq \delta\} dt
\times X\{t \in \mathbb{R}^2 \mid |t_2 - t_2| \leq \delta\} dt
\]
(27)
where in (26) we changed variables to \(t = \text{diag}(\sqrt{\lambda_1}, \sqrt{\lambda_2}) d\).

The integral in (27) measures the area that is inside the rectangle \(\{t \mid t_1^2 \leq \lambda_1 r^2, t_2^2 \leq \lambda_2 r^2\}\) and the two hyperbolas

\[\{t \mid t_1^2 - t_2^2 = \pm \delta\}\] (see Figure 1). The bound (24) can then be established by performing the following to steps:

1. deriving an upper bound on the integral in (27),
2. finding an expression of the eigenvalues of \(RJR^T\) in terms of the vectors \(u\) and \(v\),
which will be done next. We have
\[
\frac{(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} X\{t \in \mathbb{R}^2 \mid t_1^2 \leq \lambda_1 r^2, t_2^2 \leq \lambda_2 r^2\}
\times X\{t \in \mathbb{R}^2 \mid |t_1^2 - t_2^2| \leq \delta\} dt
\leq \frac{(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} X\{s \in \mathbb{R}^2 \mid s_1^2 \leq \delta \leq 2\lambda_1 r^2\}
\times X\{s \in \mathbb{R}^2 \mid |s_1^2 - s_2^2| \leq 2\lambda_2 r^2\} dt
dz
\leq \frac{(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} X\{z \in \mathbb{R}^2 \mid |z_1^2 - z_2^2| \leq \delta \leq 2\lambda_1 r^2\}
\times X\{z \in \mathbb{R}^2 \mid |z_1^2 - z_2^2| \leq 2\lambda_2 r^2\} dz
\leq \frac{(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} X\{z \in \mathbb{R}^2 \mid z_1^2 \leq 2\lambda_1 r^2, z_2^2 \leq 2\lambda_2 r^2\}
\times X\{z \in \mathbb{R}^2 \mid |z_1^2 - z_2^2| \leq \delta\} dz
\leq \frac{4(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} X\{z \in \mathbb{R}^2 \mid z_1^2 \leq \sqrt{2 \lambda_1 r^2}, z_2^2 \leq \sqrt{2 \lambda_2 r^2}\}
\times X\{z \in \mathbb{R}^2 \mid |z_1^2 - z_2^2| \leq \delta\} dz
\]
\[
\leq \frac{4(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} \chi\left\{z \in \mathbb{R}^2 \mid z_1 \leq \frac{1}{2 \sqrt{\lambda_1 + \lambda_2}} \right\} dz \\
+ \frac{4(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} \chi\left\{z \in \mathbb{R}^2 \mid z_2 \leq \sqrt{\lambda_1 + \lambda_2} \right\} dz
\]

Moreover, let \(u, v \in \mathbb{R}^n\) be linearly independent and \(C = uu^T - vv^T\). Then,

\[C = WRJRTW^T\]

with

\[J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad R = \begin{pmatrix} ||u|| & \frac{u^Tv}{||u||^2}u \\ 0 & ||v - \frac{u^Tv}{||u||^2}u|| \end{pmatrix}\]

and

\[W = \begin{pmatrix} a/||a|| & b/||b|| \end{pmatrix}\]

where the orthonormal vectors \(a/||a||\) and \(b/||b||\) are defined by

\[a = u\]
\[b = v - \frac{u^Tv}{||u||^2}u\]

Moreover,

\[\det(RJR^T) = ||u^Tv||^2 - ||u||^2||v||^2 < 0\]

\[\text{tr}(RJR^T) = ||u||^2 - ||v||^2\]

\[\sigma_2(RJR^T) = \frac{1}{2}||u + v||||u - v|| - \frac{1}{2}||u||^2 - ||v||^2\]

where \(\sigma_1(RJR^T) \geq \sigma_2(RJR^T)\) are the singular values of \(RJR^T\).

Proof. We can rewrite \(C = A^TA\) with \(A = (u, v)\). Hence, to prove (31), it is sufficient to show that \(A = WR\).

Using the definitions of the vectors \(a\) and \(b\) in (34) and (35), we can rewrite

\[A = (a, \frac{u^Tv}{||u||}a + b)\]
\[= (a, b) \begin{pmatrix} 1 & \frac{u^Tv}{||u||} \\ 0 & 1 \end{pmatrix}\]
\[= \left( \begin{pmatrix} a \\ 0 \end{pmatrix}, \begin{pmatrix} b \\ v \end{pmatrix} \right) \begin{pmatrix} ||u|| & 0 \\ 0 & ||v - \frac{u^Tv}{||u||^2}u|| \end{pmatrix}\]

which proves (31).

The explicit form of the determinant in (36) follows from the fact that

\[\det(RJR^T) = \det(R) \det(J) \det(R^T)\]
\[= -||u||^2\]
\[= -||u||^2v^Tv - \frac{u^Tv}{||u||^2}u^Tv\]
\[< 0\]

where \(\det(A)\) follows from the Cauchy-Schwarz inequality [18] and \(u\) and \(v\) being linearly independent. The expression for the trace (37) follows from \(\text{tr}(RJR^T) = \text{tr}(C)\). Finally, (38) follows from

\[\sigma_2(RJR^T) = \frac{1}{2}(\sigma_1(RJR^T) + \sigma_2(RJR^T))\]
\[= \frac{1}{2}(\sigma_1(RJR^T) - \sigma_2(RJR^T))\]
\[= \frac{1}{2}\sqrt{\text{tr}(RJR^T)^2 - 4\det(RJR^T)} - \frac{1}{2}\text{tr}(RJR^T)|\]
\[= \frac{1}{2}||u + v||||u - v|| - \frac{1}{2}||u||^2 - ||v||^2\]

\[\leq \frac{4(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} \chi\left\{z \in \mathbb{R}^2 \mid z_1 \leq \frac{1}{2 \sqrt{\lambda_1 + \lambda_2}} \right\} dz \\
+ \frac{4(2r)^{n-2}}{\sqrt{\lambda_1 \lambda_2} V(n, r)} \int_{\mathbb{R}^2} \chi\left\{z \in \mathbb{R}^2 \mid z_2 \leq \sqrt{\lambda_1 + \lambda_2} \right\} dz\]

VI. PROPERTIES OF CERTAIN RANK TWO MATRICES

Lemma 4. Let \(u, v \in \mathbb{R}^n\) be linearly independent and \(C = uu^T - vv^T\). Then,

\[C = WRJRTW^T\]

with

\[J = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad R = \begin{pmatrix} ||u|| & \frac{u^Tv}{||u||^2}u \\ 0 & ||v - \frac{u^Tv}{||u||^2}u|| \end{pmatrix}\]

and

\[W = \begin{pmatrix} a/||a|| & b/||b|| \end{pmatrix}\]

where the orthonormal vectors \(a/||a||\) and \(b/||b||\) are defined by

\[a = u\]
\[b = v - \frac{u^Tv}{||u||^2}u\]

Moreover,

\[\det(RJR^T) = ||u^Tv||^2 - ||u||^2||v||^2 < 0\]

\[\text{tr}(RJR^T) = ||u||^2 - ||v||^2\]

\[\sigma_2(RJR^T) = \frac{1}{2}||u + v||||u - v|| - \frac{1}{2}||u||^2 - ||v||^2\]

where \(\sigma_1(RJR^T) \geq \sigma_2(RJR^T)\) are the singular values of \(RJR^T\).

Proof. We can rewrite \(C = A^TA\) with \(A = (u, v)\). Hence, to prove (31), it is sufficient to show that \(A = WR\).
[10] E. J. Candès, X. Li, and M. Soltanolkotab, “Phase retrieval via Wirtinger flow: Theory and algorithms,” *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1985–2007, April 2015.

[11] Y. C. Eldar and S. Mendelson, “Phase retrieval: Stability and recovery guarantees,” *Appl. Comput. Harmon. Anal.*, vol. 36, no. 3, pp. 473–494, 2014.

[12] M. Akçakaya and V. Tarokh, “New conditions for sparse phase retrieval,” 2014, arXiv:1310.1351 [cs.IT].

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

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

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

[17] E. Riegler, D. Stotz, and H. Bölcskei, “Information-theoretic limits of matrix completion,” in *Proc. IEEE Int. Symp. Inf. Theory*, Hong Kong, China, Jun. 2015, arXiv:1504.04970 [cs.IT].

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.