ISIT'2013 1569718163

Polarization of the Rényi Information Dimension for Single and Multi Terminal Analog Compression

Saeid Haghighatshoar*, Emmanuel Abbe†
*EPFL, Lausanne, Switzerland, saeid.haghighatshoar@epfl.ch
†Princeton University, Princeton, NJ, USA, eabbe@princeton.edu

***Abstract*—This paper shows that the Rényi information dimension (RID) of an i.i.d. sequence of mixture random variables polarizes to the extremal values of 0 and 1 (fully discrete and continuous distributions) when transformed by an Hadamard matrix. This provides a natural counter-part over the reals of the entropy polarization phenomenon over finite fields. It is further shown that the polarization pattern of the RID is equivalent to the BEC polarization pattern, which admits a closed form expression. These results are used to construct universal and deterministic partial Hadamard matrices for analog to analog (A2A) compression of memoryless sources. In addition, a framework for the A2A compression of multi-terminal correlated sources is developed, providing a first counter-part of the Slepian-Wolf coding problem in the A2A setting.**

***Index Terms*—Rényi information dimension, Polarization, Information preserving matrices, Analog compression, Distributed analog compression, Compressed sensing.**

I. Introduction

Analog to analog (A2A) compression of signals has recently gathered interest in information theory [3]–[6]. In A2A compression, a high dimensional analog signal $x^n \in \mathbb{R}^n$ is encoded into a lower dimensional analog signal $y^m = f_n(x^n) \in \mathbb{R}^m$. The goal is to design the encoding so as to preserve in y^m all the information about x^n, and to recover x^n from y^m up to an acceptable distortion. It is worth mentioning that when the alphabet of x and y is finite, this framework falls into traditional topics of information theory such as lossless and lossy data compression. The novelty of A2A compression is to consider x and y to be real-valued and to impose regularity constraints, for example linearity of the encoder as motivated by compressed sensing [1], [2].

The challenge and practicality of A2A compression is to obtain dimensionality reduction, i.e., $m/n \ll 1$, by exploiting a prior knowledge on the signal. For the sparsity prior, it has been shown that for k-sparse signals, and without stability or complexity considerations, the dimensionality reduction is of order k/n. Moreover, a measurement rate of order $k/n \log(n/k)$ is sufficient to get stable recovery with a tractable algorithm such as convex programming (ℓ_1 minimization) [1].

Recently in [3], a foundation of A2A compression has been developed, shifting the attention to probabilistic signal models beyond the sparsity structure. It is shown in [3] that under linear encoding and Lipschitz-continuous decoding, the fundamental limit of A2A compression is the Rényi information dimension (RID). For a nonsingular mixture distribution, the RID is given by the mass on the continuous part, and for the specific case of sparse mixture distributions, this gives the dimensionality reduction k/n. [4] shows that robustness to noise is not a limitation for the framework in [3]. Two other works [5], [6] have corroborated the fact that complexity may not be a limitation either. In [5] spatially-coupled matrices are used for the encoding of the signal, leveraging on the analytical ground of spatially-coupled codes and predictions of [8], proving that the RID is achievable using approximate message passing algorithm for block diagonal Gaussian matrices of increasing block lengths. In [6], based on a EPI result that was further developed in [7], the polarization technique was used to deterministically construct partial Hadamard matrices of measurement rate $o(n)$ in order to compress discrete variables having zero RID. However, the case of mixture distributions was left open in [6].

This paper proposes a new approach to A2A compression by means of a polarization theory over the reals. The use of polarization techniques for sparse recovery was proposed in [9] for discrete signals, relying on coding strategies over finite fields. In this paper, it is shown that using the RID, one obtains a natural counter-part of the entropy polarization phenomenon [10], [11] over the reals. We show that the RID of an i.i.d. sequence of mixture random variables also polarizes to the two extreme values 0 and 1 (discrete and continuous distributions). To get to this result, properties of the RID in vector settings and related information measures are first developed. It is then shown that the RID polarizes by a pattern equivalent to the BEC channel polarization [10]. This is then used to obtain universal A2A compression schemes based on explicit partial Hadamard matrices. The current paper focuses on the encoding strategies and on extracting the RID without specifying the decoding strategy. Numerical simulations provide evidence that efficient algorithms like approximate message passing or ℓ_1-minimization may be used in conjunction to the obtained encoders.

Finally, the paper extends the realm of A2A compression to a multi signal settings. We provide here an information theoretic framework for general distributed A2A compression, as a counter part of the Slepian & Wolf coding problem in source compression [12]. A measurement rate region to extract the RID of correlated signals is obtained and shown to be tight.

Notation: $[n] := \{1, 2, \dots, n\}$. For $A_{m\times n}$ and $K \subset [n]$, A_K is the sub-matrix consisting of the columns in K. For another matrix $B_{m'\times n}$, $[A; B]_{(m+m')\times n}$ is the vertical concatenation of A and B. A random variable X with a mixture distribution $p_X = \delta p_c + (1-\delta) p_d$ will be represented by $X = \Theta U + \bar{\Theta} V$, where $\Theta \in \{0,1\}$, $\bar{\Theta} = 1 - \Theta$, and U, V are the continuous and the discrete components with probability distribution p_c and p_d and independent of Θ. For $x \in \mathbb{R}$ and $q \in \mathbb{N}$, $[x]_q = \frac{\lfloor q\, x \rfloor}{q}$ denotes the quantization of x with spacing $\frac{1}{q}$ and for a sequence x_1^n, $[x_1^n]_q$ denotes the component wise quantization.

II. Rényi information dimension

Let X be a random variable with a probability distribution p_X over $\mathbb{R}$. The RID of this random variable, assuming that it exists, is defined as $\lim_{q\to\infty} \frac{H([X]_q)}{\log_2(q)}$.

By Lebesque or Jordan decomposition theorem, any probability distribution over $\mathbb{R}$ like p_X can be written as a convex combination of a discrete part, a continuous part and a singular part, namely, $p_X = \alpha_d p_d + \alpha_c p_c + \alpha_s p_s$. Rényi, in [13], showed that for nonsingular distributions, $p_X = (1-\delta) p_d + \delta p_c$, the RID is well-defined and is equal to the weight of the continuous part δ. He also showed that for X_1^n a continuous random vector, $\lim_{q\to\infty} \frac{H([X_1^n]_q)}{\log_2(q)} = n$.

Our objective is to extend the definition of RID for vector random variables, which are not necessarily continuous. To do this, we will restrict ourselves to a rich space of random variables with well-defined RID. Over this space, it is possible to give a full characterization of the RID as we will see in a moment.

Definition 1. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a standard probability space. The space $\mathcal{L}(\Omega, \mathcal{F}, \mathbb{P})$ is defined as $\mathcal{L} = \cup_{n=1}^{\infty} \mathcal{L}_n$, where $\mathcal{L}_1$ is the set of all nonsingular random variables and for $n \in \mathbb{N}\backslash\{1\}$, $\mathcal{L}_n$ is the space of n-dimensional random vectors defined as

$$\begin{aligned}\mathcal{L}_n =\{X_1^n :\ & \text{there exist } k \in \mathbb{N}, A \in \mathbb{R}^{n\times k} \text{ and } Z_1^k \\ & \text{independent and nonsingular such that } X_1^n = A Z_1^k\}.\end{aligned}$$

Over the space $\mathcal{L}$, we will extend the definition of the RID along with joint RID, conditional RID and mutual Rényi information for vector random variables defined as follows:

Definition 2. Let (X_1^n, Y_1^m) be a random vector in $\mathcal{L}$. The joint RID, the conditional RID of X_1^n given Y_1^m and the mutual Rényi information between X_1^n and Y_1^m, provided that they exist, are defined as

$$d(X_1^n) = \lim_{q\to\infty} \frac{H([X_1^n]_q)}{\log_2(q)}, d(X_1^n|Y_1^m) = \lim_{q\to\infty} \frac{H([X_1^n]_q|Y_1^m)}{\log_2(q)},$$
$$I_R(X_1^n; Y_1^n) = d(X_1^n) - d(X_1^n|Y_1^m).$$

We will also need the following concepts from the linear algebra of matrices.

Definition 3. Let A and B be two arbitrary matrices of dimension $m_1 \times n$ and $m_2 \times n$. Also, let $K \subset [n]$. The influence of the matrix B on the matrix A and the residual of the matrix A given B over the column set K are defined to be

$$\begin{aligned}I(A;B)[K] &= \text{rank}([A;B]_K) - \text{rank}(A_K),\\ R(A;B)[K] &= \text{rank}([A;B]_K) - \text{rank}(B_K).\end{aligned}$$

Remark 1. It is easy to check that $I(A;B)[K]$ is the amount of increase of the rank of the matrix A_K by adding the rows of the matrix B_K. One can also easily check that $I(A;B)[K] = R(B;A)[K]$.

Theorem 1. *Let (X_1^n, Y_1^m) be a random vector in $\mathcal{L}$, namely, there are i.i.d. nonsingular random variables Z_1^k and matrices A and B of dimension $n \times k$ and $m \times k$ such that $X_1^n = A Z_1^k$ and $Y_1^m = B Z_1^k$. Let $Z_i = \Theta_i U_i + \bar{\Theta}_i V_i$ be the representation for Z_i, where U_i and V_i denote the continuous and the discrete part of Z_i. Then, we have*

- $d(X_1^n) = \mathbb{E}\{\text{rank}(A_{C_\Theta})\}$,
- $d(X_1^n|Y_1^m) = \mathbb{E}\{R(A;B)[C_\Theta]\}$,

where $C_\Theta = \{i \in [k] : \Theta_i = 1\}$ is the random set consisting of the position of continuous components.

Using Theorem 1, we obtain a list of properties of the RID.

Theorem 2. *Let (X_1^n, Y_1^m) be a random vector in $\mathcal{L}$ as in Theorem 1. Then, we have the following properties*

- *$d(X_1^n) = d(M X_1^n)$ for any arbitrary invertible matrix M of dimension $n \times n$.*
- *$d(X_1^n, Y_1^m) = d(X_1^n) + d(Y_1^m|X_1^n)$.*
- *$I_R(X_1^n; Y_1^m) = I_R(Y_1^m; X_1^n)$.*
- *$I_R(X_1^n; Y_1^m) \geq 0$ and $I_R(X_1^n; Y_1^m) = 0$ if and only if X_1^n and Y_1^m are independent after removing discrete common parts, namely, those $Z_i, i \in [k]$ that are discrete.*

III. Main results

This section gives an overview of the results in the paper.

A. Polarization of the RID

Definition 4. Let $\alpha \in [0,1]$. An "erasure process" with initial value α is defined as follows:

- $e^{\emptyset} = \alpha$, $e^+ = 2\alpha - \alpha^2$ and $e^- = \alpha^2$.
- Let $e_n = e^{b_1 b_2 \dots b_n}$, for some $\{+,-\}$-valued sequence $b_1 b_2 \dots b_n$. Define $e_n^+ = e^{b_1 b_2 \dots b_n +} = 2e_n - e_n^2$, and $e_n^- = e^{b_1 b_2 \dots b_n -} = e_n^2$.

Let $\{B_n\}_{n=1}^{\infty}$ be a sequence of i.i.d. uniform $\{+,-\}$-valued random variables. By replacing B_1^n for b_1^n in the definition of the erasure process, we obtain a stochastic process $e_n = e^{B_1 B_2 \dots B_n}$. Let $\mathcal{F}_n$ be the σ-field generated by B_1^n. Using the BEC polarization [10], we have the following results:

- $(e_n, \mathcal{F}_n)$ is a positive bounded martingale.
- e_n converges to $e_\infty \in \{0,1\}$ with $\mathbb{P}(e_\infty = 1) = \alpha$.
- For any $0 < \beta < \frac{1}{2}$, $\liminf_{n\to\infty} \mathbb{P}(e_n \leq 2^{-N^\beta}) = 1 - \alpha$, where $N = 2^n$ is the number of possible values for e_n.

Let $n \in \mathbb{N}$ and $N = 2^n$. Assume that X_1^N is a sequence of i.i.d. nonsingular random variables with a RID equal to $d(X)$ and let $Z_1^N = H_N X_1^N$, where H_N is the Hadamard matrix of order N. For $i \in [N]$, let us define $I_n(i) = d(Z_i|Z_1^{i-1})$. Assume that b_1^n is the binary expansion of $i-1$. By replacing 0 by $+$ and 1 by $-$, we can equivalently represent $I_n(i)$ by a sequence of $\{+,-\}$, namely, $I_n(i) = I^{b_1 b_2 \dots b_n}$. Similar to the erasure process, we can convert I_n to a stochastic process $I_n = I^{B_1 B_2 \dots B_n}$. We obtain the following theorem.

Theorem 3 (Single terminal RID polarization)**.** *$(I_n, \mathcal{F}_n)$ is an erasure stochastic process with initial value $d(X)$ polarizing to $\{0,1\}$.*

Let $N = 2^n$ be as before, and let $\{(X_i, Y_i)\}_{i=1}^N$ be an i.i.d. sequence of random vectors in $\mathcal{L}$, with joint and conditional RID $d(X,Y)$, $d(X|Y)$ and $d(Y|X)$. Assume that $Z_1^N = H_N X_1^N$ and $W_1^N = H_N Y_1^N$. For $i \in [N]$, let us define two processes I_n and J_n as follows:

$$I_n(i) = d(Z_i|Z_1^{i-1}),\ J_n(i) = d(W_i|W_1^{i-1}, Z_1^N).$$

It is again possible to label I_n and J_n by a sequence of b_1^n and to convert them to stochastic processes $I_n = I^{B_1 B_2 \dots B_n}$ and $J_n = J^{B_1 B_2 \dots B_n}$.

Theorem 4 (Multi terminal RID polarization)**.** *$(I_n, \mathcal{F}_n)$ and $(J_n, \mathcal{F}_n)$ are erasure stochastic processes with initial value $d(X)$ and $d(Y|X)$, both polarizing to $\{0,1\}$.*

Remark 2. In the t terminal case $t > 2$ for a t terminal memoryless source $(X_1, X_2, \dots, X_t)$, using a similar method it is possible to construct erasure processes with initial values $d(X_1), d(X_2|X_1), \dots, d(X_t|X_1^{t-1})$, polarizing to $\{0,1\}$.

B. Single terminal A2A compression

In this subsection, we will use the properties of the RID developed in Section II to study the A2A compression of memoryless sources.

Definition 5. Let X_1^N be a sequence of i.i.d. random variables in $\mathcal{L}$ with a mixture probability distribution p_X over $\mathbb{R}$. The family of measurement matrices $\{\Phi_N\}$, of dimension $m_N \times N$, is ϵ-REP(p_X) with measurement rate ρ if

$$\limsup_{q\to\infty} \frac{H([X_1^N]_q|\Phi_N X_1^N)}{H([X_1^N]_q)} \le \epsilon,\ \limsup_{N\to\infty} \frac{m_N}{N} = \rho. \qquad (1)$$

Remark 3. For an intuitive interpretation of ϵ-REP definition, assume that all of the measuring devices have a finite precision of order $\frac{1}{q}$. In that case, the potential information of the source will be $H([X_1^N]_q)$. By writing the ϵ-REP definition in the equivalent form $\frac{I([X_1^N]_q;\Phi_N X_1^N)}{H([X_1^N]_q)} \ge 1-\epsilon$, it is easy to see that an ϵ-REP matrix Φ_N preserves more than $1-\epsilon$ ratio of the information of the signal. Also, notice that the ϵ-REP property generally depends on the distribution of the source p_X.

Remark 4. For a memoryless source with $d(X) > 0$, if we divide the numerator and the denumerator in the expression (1) by $\log_2(q)$, take the limit as q tends to infinity and use the definition of the RID, we get the equivalent form $\frac{d(X_1^N|\Phi_N X_1^N)}{d(X_1^N)} \le \epsilon$, which implies that the information isometry is equivalent to the RID isometry.

We can also extend the definition when the probability distribution of the source is not known exactly but it is known to belong to a given collection of distributions Π.

Definition 6. Suppose $\Pi = \{\pi : \pi \in \Pi\}$ is a class of non-singular probability distributions. The family of measurement matrices $\{\Phi_N\}$ is ϵ-REP(Π) with measurement rate ρ if it is ϵ-REP(π) for every $\pi \in \Pi$ with measurement rate ρ.

We are mostly interested in a characterization of the required measurement rate in order to keep the information isometry. Similar to all theorems in information theory, we do this using the "converse" and "achievability" parts.

Theorem 5 (Converse result)**.** *Let X_1^N be a sequence of i.i.d. random variables in $\mathcal{L}$ with $d(X_1) > 0$. Suppose $\{\Phi_N\}$ is a family of ϵ-REP measurement matrices with measurement rate ρ. Then, $\rho \ge d(X_1)(1-\epsilon)$.*

Remark 5. In lossless source coding, the encoding rate must be greater than the entropy of the source. Theorem 5 shows that in the A2A compression setting the appropriate measure is the RID of the source.

Theorem 6 (Achievability result)**.** *Assume that the conditions of Theorem 5 holds. Then, for any $\epsilon > 0$, there is a family of ϵ-REP partial Hadamard matrices of dimension $m_N \times N$, for $N = 2^n$ with measurement rate $\rho = d(X_1)$.*

We also have the general result in Theorem 7.

Theorem 7 (Achievability result)**.** *Let Π be a family of nonsingular mixture distributions with strictly positive RID. For any $\epsilon > 0$, there is a family of ϵ-REP (Π) partial Hadamard matrices of dimension $m_N \times N$, for $N = 2^n$, with $\rho = \sup_{\pi\in\Pi} d(\pi)$.*

Remark 6. Theorem 7 states that there is a fixed ensemble of measurement matrices capable of capturing the information of all of the distributions in the family Π, however the recovery process might need to know the exact distribution of the signal in order to achieve perfect recovery.

C. Multi terminal A2A compression

In this section, our goal is to extend the A2A compression theory from the single terminal case to the multi terminal case. Here, the idea is that we have a memoryless source which is distributed in more than one terminal and we are going to take linear measurements from different terminals in order to capture the information of the source.

Definition 7. Let $\{(X_i, Y_i)\}_{i=1}^N$ be a two terminal memoryless source with a probability distribution $p_{X,Y}$ over $\mathbb{R}^2$. The family of distributed measurement matrices $\{\Phi_N^x, \Phi_N^y\}$ is ϵ-REP($p_{X,Y}$) with measurement rate (ρ_x, ρ_y) if

$$\limsup_{q\to\infty} \frac{H([X_1^N]_q, [Y_1^N]_q|\Phi_N^x X_1^N, \Phi_N^y Y_1^N)}{H([X_1^N]_q, [Y_1^N]_q)} \le \epsilon, \qquad (2)$$
$$\limsup_{N\to\infty} \frac{m_N^x}{N} = \rho_x, \quad \limsup_{N\to\infty} \frac{m_N^y}{N} = \rho_y.$$

Remark 7. If (X, Y) is a random vector in $\mathcal{L}$ with $d(X,Y) > 0$, an argument similar to Remark 4 shows that (2) can be equivalently stated as $\frac{d(X_1^N, Y_1^N|\Phi_N^x X_1^N, \Phi_N^y Y_1^N)}{d(X_1^N, Y_1^N)} \le \epsilon$.

Definition 8. Let (X, Y) be a random vector in $\mathcal{L}$ with a probability distribution $p_{X,Y}$. The Rényi information region of $p_{X,Y}$ is the set of all $(\rho_x, \rho_y) \in [0,1]^2$ satisfying

$$\rho_x \ge d(X|Y),\ \rho_y \ge d(Y|X),\ \rho_x + \rho_y \ge d(X,Y).$$

Definition 9. Assume that Π is a class of nonsingular probability distributions over $\mathbb{R}^2$. The Rényi information region of the class Π is the intersection of the Rényi information region of the distributions in class Π.

Similar to the single terminal case, we are interested in the rate region of the problem. We obtain the following converse and achievability results.

Theorem 8 (Converse result)**.** *Let $\{(X_i, Y_i)\}_{i=1}^N$ be a two-terminal memoryless source with (X_1, Y_1) being in $\mathcal{L}$. Assume that the family of measurement matrices $\{\Phi_N^x, \Phi_N^y\}$ is ϵ-REP with measurement rate (ρ_x, ρ_y). Then,*

$$\rho_x + \rho_y \geq d(X, Y)(1-\epsilon),$$
$$\rho_x \geq d(X|Y) - \epsilon\, d(X, Y),\ \ \rho_y \geq d(Y|X) - \epsilon\, d(X, Y).$$

Remark 8. This rate region is very similar to the rate region of the distributed source coding (Slepian & Wolf) problem with the only difference that the discrete entropy has been replaced by the RID.

Theorem 9 (Achievability result)**.** *Let $\{(X_i, Y_i)\}_{i=1}^N$ be a two-terminal memoryless source with (X_1, Y_1) being in $\mathcal{L}$. Given any $(\rho_x, \rho_y) \in [0, 1]^2$ satisfying*

$$\rho_x + \rho_y \geq d(X_1, Y_1), \rho_x \geq d(X_1|Y_1), \rho_y \geq d(Y_1|X_1),$$

there is a family of ϵ-REP partial Hadamard matrices with measurement rate (ρ_x, ρ_y).

We have also the general result in Theorem 10, which implies that we can construct a family of truncated Hadamard matrices, which is ϵ-REP for a class of distributions.

Theorem 10 (Achievability result)**.** *Let Π be a family of distributions over $\mathbb{R}^2$ belonging to $\mathcal{L}$. For any (ρ_x, ρ_y) in the Rényi information region of Π, there is a family of ϵ-REP(Π) partial Hadamard matrices with measurement rate (ρ_x, ρ_y).*

IV. Proof techniques

A. Polarization of the RID

In this part, we will prove the polarization of the RID in the single terminal case. The proof for the multi terminal case is very similar.

Proof of Theorem 3: For the initial value, we have $I_0(1) = d(X_1)$. Let $n \in \mathbb{N}$ and $N = 2^n$. To simplify the proof, instead of Hadamard matrices, H, we will use shuffled Hadamard matrices, $\tilde{H}$, constructed as follows: $\tilde{H}_1 = H_1$ and $\tilde{H}_{2N}$ is constructed from $\tilde{H}_N$ as follows

$$\begin{pmatrix} \tilde{h}_1 \\ \vdots \\ \tilde{h}_N \end{pmatrix} \to \begin{pmatrix} \tilde{h}_1 & , & \tilde{h}_1 \\ \tilde{h}_1 & , & -\tilde{h}_1 \\ \vdots & , & \vdots \\ \tilde{h}_i & , & \tilde{h}_i \\ \tilde{h}_i & , & -\tilde{h}_i \\ \vdots & , & \vdots \end{pmatrix},$$

where $\tilde{h}_i$, $i \in [N]$, denotes the i-th row of $\tilde{H}_N$. Let X_1^N be as in Theorem 3 and suppose $\tilde{Z}_1^N = \tilde{H}_N X_1^N$, where H_N has been replaced by $\tilde{H}_N$. Also, let $\tilde{I}_n(i) = d(\tilde{Z}_i|\tilde{Z}_1^{i-1})$, $i \in [N]$. We first prove that $\tilde{I}$ is also an erasure process with initial value $d(X_1)$ and evolves as follows

$$\tilde{I}_n(i)^+ = \tilde{I}_{n+1}(2i-1) = 2\tilde{I}_n(i) - \tilde{I}_n(i)^2$$
$$\tilde{I}_n(i)^- = \tilde{I}_{n+1}(2i) = \tilde{I}_n(i)^2.$$

Let $\tilde{H}^{i-1}$ and $\tilde{H}^i$ denote the first $i-1$ and the first i rows of $\tilde{H}_N$. Also, let $\tilde{h}_i$ denote the i-th row of $\tilde{H}_N$. Thus, we have $\tilde{Z}_1^i = \tilde{H}^i X_1^N$ and $\tilde{Z}_1^{i-1} = \tilde{H}^{i-1} X_1^N$. As X_1^N are i.i.d. nonsingular random variables, it results that $\tilde{Z}_1^i$ belong to the space $\mathcal{L}$ generated by X_1^N. Notice that using the rank characterization for RID over $\mathcal{L}$, we have

$$\tilde{I}_n(i) = d(\tilde{Z}_i|\tilde{Z}_1^{i-1}) = \mathbb{E}\{I(\tilde{H}^{i-1}; \tilde{h}_i)[C_\Theta]\},$$

where $I(\tilde{H}^{i-1}; \tilde{h}_i)[C_\Theta] \in \{0, 1\}$ is the amount of increase of the rank of $\tilde{H}_{C_\Theta}^{i-1}$ by adding $\tilde{h}_i$ and $C_\Theta = \{i \in [N] : \Theta_i = 1\}$ is the random set consisting of the position of the continuous components. At stage $n+1$, we have the shuffled Hadamard matrix $\tilde{H}_{2N}$. Consider the row i^+ which corresponds to the row $2i-1$ of $\tilde{H}_{2N}$. Now, if we look at the first block of the $\tilde{H}_{2N}$, we notice that adding $\tilde{h}_i$ has the same effect in increasing the rank of this block as it had in $\tilde{H}_N$. A similar argument holds for the second block. Moreover, adding $\tilde{h}_i$ increases the rank of the matrix if it increases the rank of either the first or the second block or both. Let $\mathbf{1}_i(\Theta_1^N) \in \{0, 1\}$ denote the random rank increase of $\tilde{H}^{i-1}$ by adding the i-the row $\tilde{h}_i$, then we have

$$\mathbf{1}_{2i-1}(\Theta_1^{2N}) = \mathbf{1}_i(\Theta_1^N) + \mathbf{1}_i(\Theta_{N+1}^{2N}) - \mathbf{1}_i(\Theta_1^N)\mathbf{1}_i(\Theta_{N+1}^{2N}).$$

Θ_1^N and Θ_{N+1}^{2N} are i.i.d. random variables and a simple check shows that $\mathbf{1}_i(\Theta_1^N)$ and $\mathbf{1}_i(\Theta_{N+1}^{2N})$ are also i.i.d.. Taking the expectation value, we obtain

$$I_n(i)^+ = I_{n+1}(2i-1) = 2I_n(i) - I_n(i)^2. \tag{3}$$

Let $\tilde{W}_1^N = \tilde{H}_N X_{N+1}^{2N}$. By the structure of $\tilde{H}_N$, we obtain

$$\tilde{I}_n(i)^+ = d(\tilde{Z}_i + \tilde{W}_i|\tilde{Z}_1^{i-1}, \tilde{W}_1^{i-1}),$$
$$\tilde{I}_n(i)^- = d(\tilde{Z}_i - \tilde{W}_i|\tilde{Z}_i + \tilde{W}_i, \tilde{Z}_1^{i-1}, \tilde{W}_1^{i-1}).$$

Using the chain rule for the RID, one can check that

$$\frac{\tilde{I}_n(i)^+ + \tilde{I}_n(i)^-}{2} = d(\tilde{Z}_i|\tilde{Z}_1^{i-1}) = \tilde{I}_n(i),$$

which along with (3), implies that $\tilde{I}_n(i)^- = \tilde{I}_n(i)^2$, thus $\tilde{I}$ is an erasure process with initial value $d(X_1)$. It can also be shown that the results are valid if $\tilde{H}_N$ is replaced by H_N. ∎

B. Single terminal A2A compression

In this part, we will overview the techniques used to prove the achievability part. We will give an explicit construction of the measurement ensemble. Let $n \in \mathbb{N}$ and let $N = 2^n$. Assume that X_1^N is an i.i.d. sequence of nonsingular random variables with the RID equal to $d(X)$. Let $Z_1^N = H_N X_1^N$, where H_N is the Hadamard matrix of order N. Suppose $I_n(i) = d(Z_i|Z_1^{i-1})$, $i \in [N]$. From Theorem 3, I is an erasure process with initial value $d(X)$. We will construct the measurement matrix Φ_N by selecting all of the rows of H_N

with the corresponding I_n value greater than $\epsilon\, d(X)$. By doing this for all N's that are a power of 2, we obtain the ensemble $\{\Phi_N\}$. Assume that the dimension of Φ_N is $m_N \times N$.

Proof of Theorem 6: We first show that the family $\{\Phi_N\}$ has measurement rate $d(X)$. Notice that the process I_n converges almost surely. Thus, it also converges in probability. Specifically, by uniform probability assumption

$$\begin{aligned}\limsup_{N\to\infty} \frac{m_N}{N} &= \limsup_{N\to\infty} \frac{\#\{i \in [N] : I_n(i) \geq \epsilon\, d(X)\}}{N}\\ &= \limsup_{n\to\infty} \mathbb{P}(I_n \geq \epsilon\, d(X))\\ &= \mathbb{P}(I_\infty \geq \epsilon\, d(X)) = d(X).\end{aligned}$$

It remains to prove that $\{\Phi_N\}$ is ϵ-REP. Let $S = \{i \in [N] : I_n(i) \geq \epsilon\, d(X)\}$ denote the selected rows to construct Φ_N. It is easy to check that $\Phi_N X_1^N = Z_S$. Let $B_i = S \cap [i-1]$ denote all of the indices in S before i. We have

$$\begin{aligned}d(X_1^N|Z_S) &= d(Z_1^N|Z_S) = d(Z_{S^c}|Z_S) = \sum_{i\in S^c} d(Z_i|Z_{B_i}, Z_S)\\ &\leq \sum_{i\in S^c} d(Z_i|Z_1^{i-1}) = \sum_{i\in S^c} I_n(i) \leq N\epsilon\, d(X),\end{aligned}$$

which shows the ϵ-REP property for $\{\Phi_N\}$. ■

V. Simulation

Up to now, we defined the notion of ϵ-REP for an ensemble of measurement matrices. This definition is what we call an "informational" characterization. Now, we can ask weather this has some "operational" implications, in the sense that after having the linear measurements, is it possible to recover the source up to an acceptable distortion? In particular, is there a computationally feasible algorithm to do that?

For simulations, we use a unit variance sparse distribution $p_X(x) = (1-\delta)\delta_0(x) + \delta p_c(x)$, where $\delta_0(x)$ is the unit delta measure at point zero and p_c is the distribution of the continuous part. We use the MSE (mean square error) as distortion measure. The simulations are done with the Hadamard matrix of order $N = 512$. To build the measurement matrix A, we select all of the rows of H_N with highest conditional RID, as stated in IV-B, until we get acceptable recovery distortion. Figure 1 shows the phase transition (PT) diagram for the ℓ_1-minimization algorithm. The simulations are done with 3 different distributions for p_c: Gaussian, Laplacian and Uniform. The recovery is successful for the measurement rates above the plotted curves. The results show the insensitivity of the PT region to the distribution of the continuous components. We also simulated the AMP algorithm for the Gaussian case,

$$\begin{aligned}z_t &= y - A\hat{x}_t + \frac{1}{\gamma} z_{t-1}\langle \eta'_{t-1}(A^* z_{t-1} + \hat{x}_{t-1})\rangle,\\ \hat{x}_{t+1} &= \eta_t(A^* z_t + \hat{x}_t),\end{aligned}$$

where $y = A\,x$, γ is the measurement rate, $\langle a_1^n\rangle = \sum_{i=1}^n a_i/n$, $\eta_t(u) = (\eta_{t,1}(u_1), \dots, \eta_{t,N}(u_N))$ and $\eta_{t,i}(u_i) = \mathbb{E}\{X|u_i = X + \tau_t N\}$, with $N \sim \mathcal{N}(0,1)$ and τ_t given by the state evolution equation for AMP. Figure 2 compares the PT diagram for AMP and ℓ_1-minimization. Although AMP performs better than ℓ_1-minimization, there is still a gap with the optimal line.

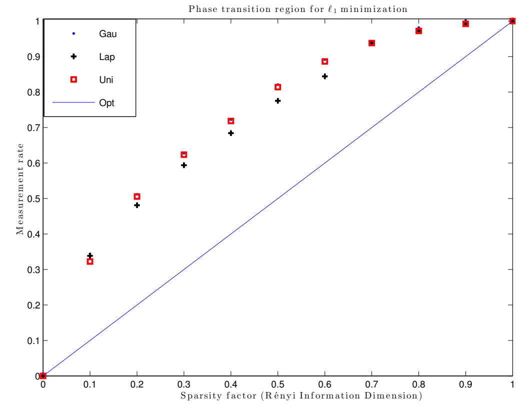

Fig. 1: PT diagram for ℓ_1-minimization

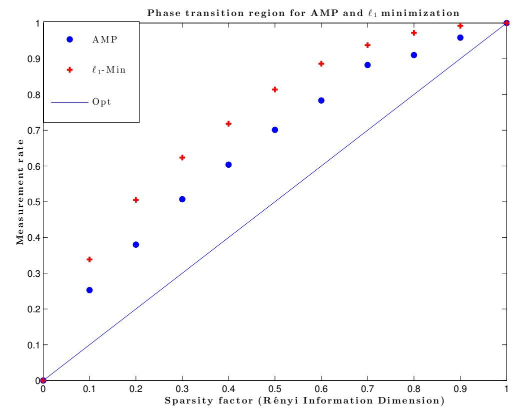

Fig. 2: PT diagram for AMP and ℓ_1-minimization

References

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

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

[3] Y. Wu and S. Verdú, "Rényi Information Dimension: Fundamental Limits of Almost Lossless Analog Compression," *IEEE Transactions on Information Theory*, vol. 56, no. 8, pp. 3721–3747, Aug. 2010.

[4] Y. Wu and S. Verdú, "Optimal Phase Transitions in Compressed Sensing," *IEEE Trans. on Inf. Theory*, vol. 58, no. 10, pp. 6241-6263, Oct. 2012.

[5] D.L. Donoho, A. Javanmard, and A. Montanari, "Information-theoretically optimal compressed sensing via spatial coupling and approximate message passing," submitted to *IEEE Trans. on Inf. Theory*, Dec. 2011.

[6] S. Haghighatshoar, E. Abbe, E. Telatar, "Adaptive sensing using deterministic partial Hadamard matrices," *IEEE International Symposium on Information Theory*, pp. 1842–1846, Jul. 2012.

[7] S. Haghighatshoar, E. Abbe, E. Telatar, "A new entropy power inequality for integer-valued random variables," arxiv.org/abs/1301.4185, Jan. 2013.

[8] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborová, "Statistical physics based reconstruction in compressed sensing," preprint, Nov. 2011, http://arxiv.org/abs/1109.4424.

[9] E. Abbe, "Universal source polarization and sparse recovery," *IEEE Information Theory Workshop (ITW)*, Dublin, Aug. 2010.

[10] E. Arıkan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels," *IEEE Tran. on Inf. Theory*, vol. IT-55, pp. 3051–3073, Jul. 2009.

[11] E. Arıkan, "Source polarization," in *Proc. IEEE Int. Symp. Inf. Theory*, Austin, 2010.

[12] D. Slepian and J.K. Wolf, "Noiseless coding of correlated information sources," *IEEE Trans. on Inf. Theory*, vol. 19, pp. 471–480, Jul. 1973.

[13] A. Rényi,"On the dimension and entropy of probability distributions," *Acta Mathematica Hungarica*, vol. 10, no. 1-2, Mar. 1959.