Adaptive sensing using deterministic partial Hadamard matrices

S. Haghighatshoar, E. Abbe and E. Telatar
EPFL, Switzerland
Email: {saeid.haghighatshoar, emmanuel.abbe, emre.telatar} @epfl.ch

***Abstract*—This paper investigates the construction of deterministic measurement matrices preserving the entropy of a random vector with a given probability distribution. In particular, it is shown that for a random vector with i.i.d. discrete components, this is achieved by selecting a subset of rows of a Hadamard matrix such that (i) the selection is deterministic (ii) the fraction of selected rows is vanishing. In contrast, it is shown that for a random vector with i.i.d. continuous components, no entropy preserving measurement matrix allows dimensionality reduction. These results are in agreement with the results of Wu-Verdu on almost lossless analog compression. This paper is however motivated by the complexity attribute of Hadamard matrices, which allows the use of efficient and stable reconstruction algorithms. The proof technique is based on a polar code martingale argument and on a new entropy power inequality for integer-valued random variables.**

***Index Terms*—Entropy-preserving matrices, Analog compression, Compressed sensing, Entropy power inequality.**

I. INTRODUCTION

Information theory has extensively studied the lossless and lossy compression of discrete time signals into *digital* sequences. These problems are motivated by the model of Shannon, where an analog signal is first acquired, by sampling it at a high enough rate to preserve all of its information (Nyquist-Shannon sampling theorem), and *then* compressed. More recently, it was realized that proceeding to "joint sensing-compression" schemes can be beneficial. In particular, *compressed sensing* introduces the perspective that sparse signals can be compressively sensed to decrease measurement rate. As for joint source-channel coding schemes, one may wonder why this would be useful? Eventually, the signal is represented with the same amount of bits, so why would it be preferable to proceed jointly or separately? In a nutshell, if measurements are expensive (such as for example in certain bio-medical applications), then compressed sensing is beneficial.

From an information-theoretic perspective, compressed sensing can be viewed as a form of analog to analog compression, namely, transforming a higher dimensional discrete time signal into a lower-dimensional one over the reals, without "losing information". The key point being that, since measurements are analog, one may as well pack as much information in each measurement (whereas in the compression of discrete signals, a measurement on a larger alphabet is more expensive than a measurement in bits). However, compressing a vector in $\mathbb{R}^n$ into a vector in $\mathbb{R}^m, m < n$, without regularity constraints is not an interesting problem, since $\mathbb{R}^n$ and $\mathbb{R}^m$ have the same cardinality.

Recently, [1] introduced a more reasonable framework to study analog compression from an information-theoretic perspective. By requiring the encoder to be linear and the decoder to be Lipschitz continuous, the fundamental compression limit is shown to be the Rényi information dimension. The setting of [1] also raises a new interesting problem: in the same way that coding theory aims at approaching the Shannon limit with low-complexity schemes, it is a challenging problem to devise efficient schemes to reach the Rényi dimension. Indeed, in this analog framework, realizing measurements in a low complexity manner is at the heart of the problem: it is rather natural that the Rényi dimension is the fundamental limit irrespective of complexity considerations, but without a low-complexity scheme, one may not have any gain in proceeding with a joint compression-sensing approach. For example in the compressed sensing, with $O(k\log(n/k))$ instead of $O(k)$ measurements, k-sparse signals can be reconstructed using l_1 minimization, which is a convex optimization problem, rather than l_0 minimization, which is intractable [6], [7]. Hence, in general, complexity requirements may raise the measurement rate.

The scope of this paper is precisely to investigate what measurement rates can be achieved by taking into account the complexity of the sensing matrix, which in turn, influences the complexity of the reconstruction algorithm. Our goal is to consider signals that are memoryless and drawn from a probability distribution on $\mathbb{R}$, which may be purely atomic, purely continuous or mixed. It is legitimate to attempt reaching this goal by borrowing tools from coding theory, in particular from codes achieving least compression rates in the discrete setting. Our approach is based on using Hadamard matrices for encoding (taking measurements) and developing a counter-part of the polar technique [2], [3] with arithmetic over $\mathbb{R}$ (or $\mathbb{Z}$ for atomic distributions) rather than $\mathbb{F}_2$ or $\mathbb{F}_q$. The proof technique uses the martingale approach of polar codes and a new form of entropy power inequality for discrete distributions. Rigorous results are obtained and sensing matrix construction is deterministic. A nested property is also investigated which allows one to adapt the measurement rate to the sparsity level of the signal.

Recently, spatially-coupled LDPC codes have allowed to achieve rigorous results in coding theory. This approach has been exploited by [4], [5], which proposes the use of spatially coupled matrices for sensing. In [5], the mixture case is covered and further analysis on the reconstruction algorithm is provided. However, the sensing matrix is still random. It is known that Hadamard matrices truncated randomly afford desirable properties for compressed sensing. We extend this work and show that by knowing signal distribution, Hadamard matrices can be truncated deterministically to achieve a minimal measurement rate.

II. Related Work

Let $X_1, X_2, \ldots, X_N$ be i.i.d. Bernoulli(p) random variables, where $N = 2^n$ for some $n \in \mathbb{Z}_+$. We use the notation a_i^j for the column vector $(a_i, a_{i+1}, \ldots, a_j)^t$ and set a_i^j to null if $j < i$. We also define $[r] = \{i \in Z : 1 \leq i \leq r\}$. Let $G_N = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}^{\otimes n}$, where $\otimes$ denotes the Kronecker product, and let $Y_1^N = B_N G_N X_1^N$, where B_N is a specific shuffling matrix introduced in [3] which changes the order of the rows of G_N. Define $H_i = H(Y_i|Y_1^{i-1})$, $i \in [N]$, to be the conditional entropy of Y_i given Y_1^{i-1}. In [3], Arikan shows that for any $\delta > 0$ and for large N, the values H_i, $i \in [N]$, polarize to 0 or 1. This provides a compression scheme achieving the least compression rate, since for every $\delta \in (0, 1)$

$$\frac{\#\{i \in [N] : H_i \in (1-\delta, 1]\}}{N} \to H(X). \tag{1}$$

From another point of view, every Y_i is associated with a specific row of the matrix G_N and (1) indicates that the "measurement" rate required to extract the informative components is close to the entropy of the source $H(X)$ for large N.

In signal acquisition, measurements are analog. Hence, one could consider $Y_1^N = G_N X_1^N$ with arithmetic over the real field and investigate if any "polarization phenomenon" occurs. The difference is that, in this case, the measurement alphabet is unbounded. In particular, the H_i values are not bounded above.

III. Problem Statement

Definition 1 (Restricted iso-entropy property)**.** Let X_1^N be discrete i.i.d. random variables with a probability distribution p_X supported on a finite set. The family $\{\Phi_N\}$ of measurement matrices, where Φ_N has dimension $m_N \times N$, is ϵ-REP(p_X) with measurement rate ρ if

$$\frac{H(X_1^N|\Phi_N X_1^N)}{N} \leq \epsilon,$$

and $\limsup_{N\to\infty} \frac{m_N}{N} = \rho$.

In general, the labeling N can be any subsequence of $\mathbb{Z}_+$. We will consider $N = 2^n, n \in \mathbb{Z}_+$.

Definition 2. Let X_1^N be continuous (or mixture) random variables with probability distribution p_X. The family of measurement matrices $\{\Phi_N\}$ of dimension $m_N \times N$ is (ϵ, γ)-REP(p_X) with measurement rate ρ if

1) there exists a single letter quantizer $Q : \mathbb{R} \to \mathbb{Z}$ such that M.M.S.E. of X given $Q(X)$ is less that γ,
2) for any N,
$$\frac{H(Q(X_1^N)|\Phi_N X_1^N)}{N} < \epsilon,$$
where $Q(X_1^N) = (Q(X_1), Q(X_2), \ldots, Q(X_N))^t$,
3)
$$\limsup_{N\to\infty} \frac{m_N}{N} = \rho.$$

We address the following questions in this paper:

1) Given a probability distribution p_X over a finite set, and $\epsilon > 0$, is there a family of measurement matrices that is ϵ-REP and has measurement rate ρ? What is the set of all possible (ϵ, ρ) pairs? Is it possible to construct a near optimal family of truncated Hadamard matrices with a minimal measurement rate? How is the truncation adapted to the distribution p_X?
2) Is it possible to obtain an asymptotic measurement rate below 1 for continuous distributions?

Remark 1. The RIP notion, introduced in [6], [7], is useful in compressed sensing, since it guarantees a stable l_2-recovery. We consider truncated Hadamard matrices satisfying ϵ-REP condition and since they have a Kronecker structure, we obtain a low-complexity reconstruction algorithm. However, this part is not emphasized in this paper, and we mainly focus on the construction of the truncated Hadamard matrices. Section VI provides numerical simulations of a divide and conquer ML decoding algorithm and illustrates the robustness of the recovery to noise. In a future work, we will investigate the use of a recovery algorithm à la [5].

IV. Main Results

The main results of this paper are summarized here.

Definition 3. Let $\{J_N = B_N \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix}^{\otimes n}, N = 2^n, n \in \mathbb{Z}_+\}$ be the family of Hadamard matrices, where B_N is the row shuffling matrix introduced in [3]. Suppose X_1^N are i.i.d. random variables with distribution p_X over a finite subset of $\mathbb{Z}$. Let $Y_1^N = J_N X_1^N$ and define $H_i = H(Y_i|Y_1^{i-1})$ and $m_N = \#\{i \in [N] : H_i > \epsilon\}$. The (ϵ, p_X)- truncated Hadamard family $\{\bar{J}_N\}$, is the set of matrices of dimension $m_N \times N$ obtained by selecting those rows of J_N with $H_i > \epsilon$.

Theorem 1 (Absorption phenomenon)**.** *Let X be a discrete random variable with a probability distribution p_X supported on a finite subset of $\mathbb{Z}$. For a fixed $\epsilon > 0$, the family of (ϵ, p_X)-truncated Hadamard matrices $\{\bar{J}_N, N = 2^n, n \in \mathbb{Z}_+\}$ (defined above) are ϵ-REP(p_X) with measurement rate 0. In other words,*

$$\limsup_{N\to\infty} \frac{m_N}{N} = 0.$$

Remark 2. Although all of the measurement matrices $\bar{J}_N$ are constructed by truncating the matrices J_N, the order and number of the selected rows, m_N, to construct $\bar{J}_N$ depends on the distribution p_X.

For continuous distributions, and for any fixed distortion γ, the measurement rate approaches 1 as ϵ tends to 0. This result has been shown in [1] in a more general context. We recover this result in our setting for the case of a uniform distribution over $[-1,1]$.

Lemma 1. *Let p_U be the uniform distribution over $[-1,1]$ and let $Q:[-1,1]\to\{0,1,\dots,q-1\}$ be a uniform quantizer for X with M.M.S.E. less than γ. Assume that $\{\Phi_N\}$ is a family of full rank measurement matrices of dimension $m_N\times N$. If $\{\Phi_N\}$ is (ϵ,γ)-REP(p_U), then the measurement rate, ρ, goes to 1 as ϵ tends to 0.*

Theorem 2 (An EPI over $\mathbb{Z}$)**.** *For every probability distribution p over $\mathbb{Z}$,*

$$H(p\star p)-H(p)\geq g(H(p)), \tag{2}$$

where $g:\mathbb{R}_+\to\mathbb{R}_+$ is strictly increasing, $\lim_{x\to\infty} g(x)=\frac{1}{8\log(2)}$ and $g(x)=0$ if and only if $x=0$.

Remark 3. This theorem complements the work in [8] to obtain an entropy power inequality for discrete random variables.

V. Proof Overview

A. An EPI for Discrete Random Variables

Entropy power inequality for continuous and independent random variables X and Y is

$$2^{2h(X+Y)}\geq 2^{2h(X)}+2^{2h(Y)}, \tag{3}$$

where h denotes the differential entropy. If X and Y have the same density p, then (3) becomes

$$h(p\star p)\geq h(p)+\frac{1}{2},$$

which implies a guaranteed increase of the differential entropy. For this reason, we call (2) an EPI for discrete random variables.

Lemma 2. *Let $c>0$ and suppose p is a probability measure over $\mathbb{Z}$ such that $H(p)=c$. Then, for any $i\in\mathbb{Z}$,*

$$H(p\star p)-c\geq cp_i-(1+p_i)h_2(p_i),$$

where $h_2(x)=-x\log_2(x)-(1-x)\log_2(1-x)$ is the binary entropy function and p_i denotes the probability of i.

Lemma 3. *Let $c>0$, $0<\alpha\leq\frac{1}{2}$ and $n\in\mathbb{Z}$. Assume that p is a probability measure on $\mathbb{Z}$ such that $\alpha\leq p((-\infty,n])\leq 1-\alpha$ and $H(p)=c$, then*

$$\|p\star p_1-p\star p_2\|_1\geq 2\alpha,$$

where $p_1=\frac{1}{p((-\infty,n])}p|_{(-\infty,n]}$ and $p_2=\frac{1}{p([n+1,\infty))}p|_{[n+1,\infty)}$ are scaled restrictions of p to $(-\infty,n]$ and $[n+1,\infty)$ respectively.

Lemma 4. *Assuming the hypotheses of Lemma 3,*

$$H(p\star p)-c\geq\frac{\alpha^2}{2\log(2)}\|p\star p_1-p\star p_2\|_1^2.$$

Lemma 5. *Assuming the hypotheses of Lemma 3,*

$$H(p\star p)-c\geq\frac{2\alpha^4}{\log(2)}.$$

Proof of Theorem 2: Suppose that p is a distribution over $\mathbb{Z}$ with $H(p)=c$. Set $y=\|p\|_\infty$. There is an $\alpha\geq\frac{1-y}{2}$ and an integer n such that $\alpha\leq p((-\infty,n])\leq 1-\alpha$. Using Lemma 2 and Lemma 5, it results that $H(p\star p)-c\geq g(c)$ where

$$g(c)=\min_{y\in[0,1]}\max\Big(\frac{(1-y)^4}{8\log(2)},cy-(1+y)h_2(y)\Big).$$

It is easy to check that $g(c)$ is a continuous function of c. The monotonicity of g follows from the fact that $cy-(1+y)h_2(y)$ is an increasing function of c for every $y\in[0,1]$. For strict positivity, note that $(1-y)^4$ is strictly positive for $y\in[0,1)$ and it is 0 when $y=1$, but $\lim_{y\to1} cy-(1+y)h_2(y)=c$. Hence for $c>0$, $g(c)>0$. If $c=0$ then

$$\max\Big(\frac{(1-y)^4}{8\log(2)},cy-(1+y)h_2(y)\Big)=\frac{(1-y)^4}{8\log(2)},$$

and its minimum over $[0,1]$ is 0. For asymptotic behavior, note that at $y=0$, $cy-(1+y)h_2(y)=0$ and $\frac{(1-y)^4}{8\log(2)}=\frac{1}{8\log(2)}$. Hence, from continuity, it results that $g(c)\leq\frac{1}{8\log(2)}$ for any $c\geq 0$. Also for any $\epsilon>0$ there exists a c_0 such that for any $\epsilon<y\leq 1$, $cy-(1+y)h_2(y)\geq\frac{1}{8\log(2)}$. Thus for any $\epsilon>0$ there is a c_0 such that for $c>c_0$, the outer minimum over y in $g(c)$ is achieved on $[0,\epsilon]$. Hence, for any $c>c_0$, $g(c)\geq\frac{(1-\epsilon)^4}{8\log(2)}$. This implies that for every $\epsilon>0$,

$$\frac{1}{8\log(2)}\geq\limsup_{c\to\infty} g(c)\geq\liminf_{c\to\infty} g(c)\geq\frac{(1-\epsilon)^4}{8\log(2)},$$

and $\lim_{c\to\infty} g(c)=\frac{1}{8\log(2)}$. ■

Figure 1 shows the EPI gap. As expected, for large values of $H(p)$, the gap approaches the asymptotic value $\frac{1}{8\log(2)}$. This is very similar to the EPI bound obtained for continuous random variables.

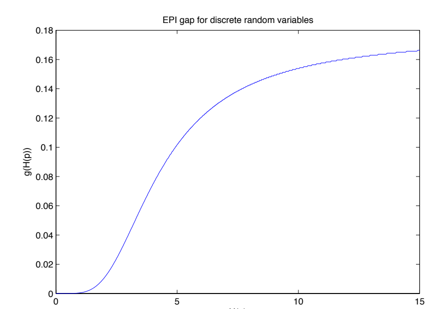


Fig. 1: EPI gap for discrete random variables

B. Conditional Entropy Martingale

Assume that X_1^N, $N=2^n$, $n\in\mathbb{Z}_+$, is a set of i.i.d. random variables with probability distribution p_X over a finite subset of $\mathbb{Z}$. Let $Y_1^N=J_NX_1^N$, where $J_N=B_N\begin{pmatrix}1&1\\-1&1\end{pmatrix}^{\otimes n}$ is the shuffled Hadamard matrix of dimension N and let $H_i=H(Y_i|Y_1^{i-1}), i\in[N]$, be the conditional entropy values.

Lemma 6. *Let X_1^N be as in the previous part and let $Z_1^N = B_N G_N X_1^N$ where G_N and B_N are as before. Assume that $\tilde{H}_i = H(Z_i|Z_1^{i-1})$, $i \in [N]$, then $H_i = \tilde{H}_i$, $i \in [N]$.*

Remark 4. The only point of Lemma 6 is that in application, it is preferred to use J because the rows of J are orthogonal to one another. For simplicity of proofs, we use G matrices and relate to the polar code notations [2], [3].

Notice that we can represent $B_N G_N$ in a recursive way. Let us define two binary operation $\oplus$ and $\ominus$ as follows

$$\ominus(a, b) = a + b$$
$$\oplus(a, b) = b,$$

where $+$ is the usual integer addition. It is easy to see that we can do the multiplication by $B_N G_N$ in a recursive way. Figure 2 shows a simple case for $B_4 G_4$. The $-$ or $+$ sign on an arrow shows that the result for that arrow is obtained by applying a $\ominus$ or $\oplus$ operation to two input operands. If we

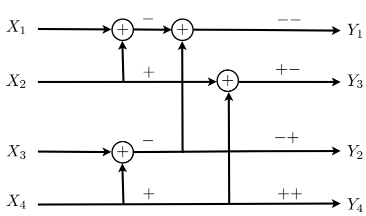


Fig. 2: Recursive structure for multiplication by G_4

consider a special output Y_m, there are a sequence of $\oplus$ and $\ominus$ operations on the input random variables which result in Y_m. An easy way to find this sequence of operations is to write the binary expansion of $m-1$. Then, each 0 in this expansion corresponds to a $\ominus$ operation and each 1 corresponds to a $\oplus$ operation. Using this binary labeling, we define a binary stochastic process. Assume that $\Omega = \{0, 1\}^\infty$, and $\mathcal{F}$ is the σ-algebra generated by the cylindrical sets

$$S_{(i_1, i_2, \cdots, i_s)} = \{\omega \in \Omega \text{ such that } \omega_{i_1} = 1, \cdots, \omega_{i_s} = 1\}$$

for every integer s and $i_1, i_2, \cdots, i_s$. We define $\mathcal{F}_n$ as the σ-algebra generated by the first n coordinates of ω and $\mathcal{F}_0 = \{\emptyset, \Omega\}$ as the trivial σ-algebra. We also define the uniform probability measure μ over the cylindrical sets by $\mu(S_{(i_1, i_2, \cdots, i_n)}) = \frac{1}{2^n}$. This measure can be uniquely extended to $\mathcal{F}$. Let $[\omega]_n = \omega_1\omega_2 \ldots \omega_n$ denote the first n coordinates of $\omega = \omega_1\omega_2 \ldots$ and $Y_{[\omega]_n}$ denote the random variable Y_i, where the binary expansion of $i - 1$ is $[\omega]_n$, and let $Y^{[\omega]_n} = \{Y_{[\eta]_n} : [\eta]_n < [\omega]_n\}$. We also define the random variable I_n by

$$I_n(\omega) = H(Y_{[\omega]_n} | Y^{[\omega]_n}). \tag{4}$$

It is also important to note that

$$I_{n+1}([\omega]_n, 0) = H(Y_{[\omega]_n} + \tilde{Y}_{[\omega]_n} | Y^{[\omega]_n}, \tilde{Y}^{[\omega]_n}) \tag{5}$$

where $\tilde{}$ denotes an independent copy of the corresponding random element.

Theorem 3. *$(I_n, \mathcal{F}_n)$ is a martingale.*

C. Main Theorem

Proof of Theorem 1: Assume that $Y_1^N = J_N X_1^N$, for $N = 2^n, n \in \mathbb{Z}_+$, and $H_i = H(Y_i|Y_1^{i-1})$, $i \in [N]$. Also fix $\epsilon > 0$. Let us define

$$K_n = \{i : i \in [N], H_i > \epsilon\},$$
$$Y_{[K_n]} = \{Y_j : j \in K_n\}.$$

Hence, by Definition 3, $|K_n| = m_N$ and $\bar{J}_N$ is obtained from J_N by selecting the rows with index in K_n. We have

$$\begin{aligned}
&H(X_1^N|\bar{J}_N X_1^N) = H(X_1^N) - I(X_1^N; \bar{J}_N X_1^N) \\
&= H(Y_1^N) - H(Y_{[K_n]}) = H(Y_{[K_n^c]}|Y_{[K_n]}) \\
&\leq \sum_{i \in K_n^c} H(Y_i|Y_1^{i-1}) \leq |K_n^c|\epsilon = (N - m_N)\epsilon,
\end{aligned}$$

which implies that

$$\frac{H(X_1^N|\bar{J}_N X_1^N)}{N} \leq \frac{(N - m_N)\epsilon}{N} \leq \epsilon.$$

This shows that the family $\{\bar{J}_N\}$ is ϵ-REP. Now it remains to show that the measurement rate of this family is 0. To prove this, we use Lemma 6 and construct the martingale I_n by (4). I_n is a positive martingale and converges to a random variable I_∞ almost surely. Our aim is to show that for any two positive numbers a and b where $a < b$, $\mu(I_\infty \in (a, b)) = 0$, which implies that $\mu(I_\infty \in \{0, \infty\}) = 1$. Since I_n is a martingale, $E\{I_n\} = E\{I_0\} = H(X) < \infty$. Using Fatou's lemma we obtain $E\{I_\infty\} \leq \liminf E\{I_n\} = H(X_1) < \infty$, which implies that $\mu(I_\infty = \infty) = 0$. Hence, I_n converges almost surely to 0 and it also converges to 0 in probability. In other words, given $\epsilon > 0$,

$$\limsup_{n\to\infty} \mu(I_n > \epsilon) = \limsup_{n\to\infty} \frac{|K_n|}{2^n} = \limsup_{N\to\infty} \frac{m_N}{N} = 0.$$

This implies that for a fixed $\epsilon > 0$ the measurement rate ρ is 0. Now it remains to prove that for any two positive numbers a and b, where $a < b$, $\mu(I_\infty \in (a, b)) = 0$. Fix a $\delta > 0$ then for every ω in the convergence set there is a n_0 such that for $n > n_0$, $|I_{n+1}(\omega) - I_n(\omega)| < \delta$. This implies that for $n > n_0$

$$|I_{n+1}(\omega) - I_n(\omega)| = |I_{n+1}([\omega]_n, 0) - I_n([\omega]_n)| < \delta.$$

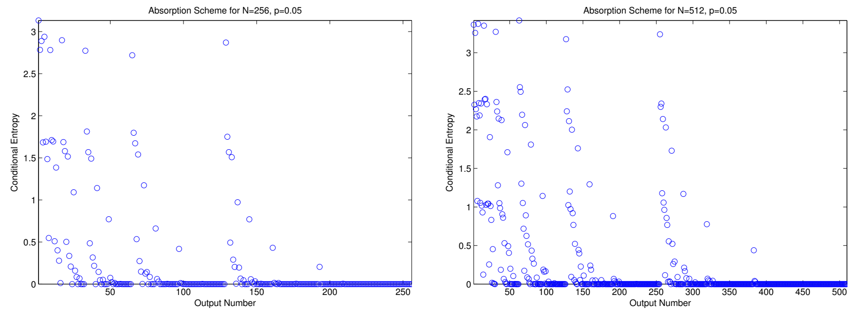
Using (5) and the entropy power inequality (2), it results that $0 \leq I_n(\omega) < \rho(\delta)$ where $\rho(\delta)$ can be obtained from the EPI curve in Figure 1. This implies that I_n must converge to 0 and this completes the proof. ■

VI. Numerical Simulations

For simulation, we use a binary random variable, where $p_X(0) = 1 - p$ for some $0 < p \leq \frac{1}{2}$.

A. Absorption Phenomenon

Figure 3 shows the absorption phenomenon for $p = 0.05$ and $N = 256, 512$.



(a) $N = 256$ (b) $N = 512$

Fig. 3: Absorption trace for $p = 0.05$

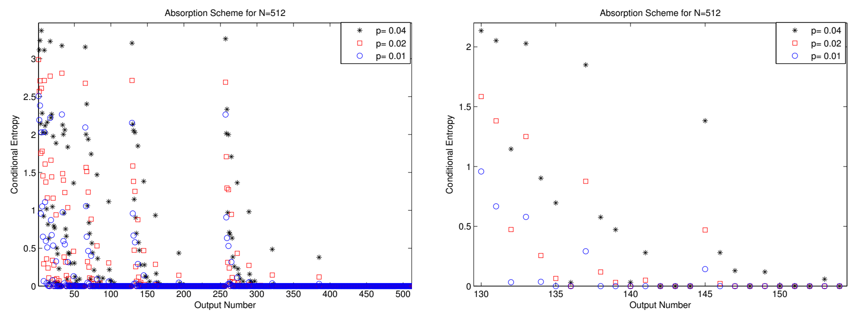


Fig. 4: Nested property for $N = 512$ and different p

B. Nested Property

Absorption phenomenon is shown in Figure 4 for $N = 512$ and different values of p. It is seen that the high entropy indices for smaller p are *included* in the high entropy indices of larger p. We call this the *"nested"* property.

C. Robustness to Measurement Noise

Figure 5 shows the stability analysis of the reconstruction algorithm to i.i.d. $\mathcal{N}(0, \sigma^2)$ measurement noise. For simulation, we used $N = 512$, $p = 0.05$ and a 0.01-REP measurement matrix by keeping all of the rows of the matrix J_N with indices in the set K. For recovery, we used ML decoder which exploits the recursive structure of the polar code. We define the signal to noise ratio at the input and output of the decoder as:

$$\text{SNR}_{\text{in}} = \frac{\sum_{i \in K} E(Y_i^2)}{|K|\,\sigma^2},$$

$$\text{SNR}_{\text{out}} = \frac{\sum_{i=1}^{N} E(X_i^2)}{\sum_{i=1}^{N} E(|X_i - \hat{X}_i|^2)},$$

where $\hat{X}_i$ is the output of the ML decoder. The result shows approximately 4 dB loss in SNR for high SNR regime.

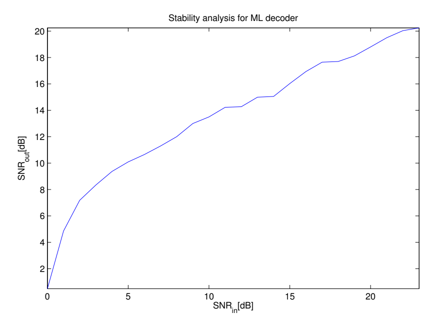


Fig. 5: Stability analysis

ACKNOWLEDGEMENT

E.A. would like to thank M. Madiman and A. Barron for stimulating discussions on this problem at Yale University, and P. Vandergheynst for helpful comments.

REFERENCES

[1] Y. Wu and S. Verdú,"Rényi information dimension: fundamental limits of almost lossless analog compression", IEEE Transaction on Information Theory, vol. 56, no. 8, pp. 3721–3747, August 2010.
[2] E. Arikan, "Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels", IEEE Transaction on Information Theory, vol. 55, p. 3051, July 2009.
[3] E. Arikan, "Source polarization", http://arxiv.org/abs/1001.3087, 2010.
[4] F. Krzakala, M. Mézard, F. Sausset, Y. Sun, and L. Zdeborova, "Statistical physics-based reconstruction in compressed sensing", arXiv:1109.4424, 2011.
[5] D. L. Donoho, A. Javanmard, A. Montanari, "Information-Theoretically optimal compressed sensing via spatial coupling and approximate message passing", http://arxiv.org/abs/1112.0708v1, 2011.
[6] E. J. Candès, T. Tao, "Decoding by linear programming", IEEE Transaction on Information Theory, vol.51, pp. 4203–4215, December 2005.
[7] E. J. Candès, T. Tao, "Near-optimal signal recovery from random projections: universal encoding strategies", IEEE Transaction on Information Theory, vol. 52, pp. 5406–5425, December 2006.
[8] O. Johnson, Y. Yu, "Monotonicity, thinning, and discrete versions of the entropy power inequality", IEEE Transaction on Information Theory, vol. 56, pp. 5387– 5395, November 2010.