Finite-Length Analysis of Lossy Compression via Sparse Regression Codes

Gal Livny

Technion

02/05/2019

Under the supervision of Prof. Igal Sason

Motivation

- Recently, a new capacity-achieving type of codes, sparse regression codes (SPARCs), has been suggested (Joseph and Barron '12) for
  - reliable communication over *memoryless* channels
  - lossy compression of *memoryless and stationary* sources

  with *continuous* alphabets
- Current analysis is mostly asymptotic.
- This work is focused on the finite-length analysis of SPARCs for lossy compression of stationary memoryless sources.

Early Works on SPARCs

- A. Joseph and A. Barron, "Least squares superposition codes of moderate dictionary size are reliable at rates up to capacity," 2012
- A. Joseph and A. Barron, "Fast sparse superposition codes have near exponential error probability for $R < \mathcal{C}$," 2014
- R. Venkataramanan, A. Joseph and S. Tatikonda, "Lossy compression via sparse linear regression: performance under minimum-distance encoding," 2014
- R. Venkataramanan, T. Sarkar and S. Tatikonda, "Lossy compression via sparse linear regression: computationally efficient encoding and decoding," 2014
- J. Barbier and F. Krzakala, "Approximate message-passing decoder and capacity-achieving sparse superposition codes," 2017

Lossy Compression Problem Setup

$$S_1, \ldots, S_n \to \boxed{\text{encoder}} \xrightarrow{\quad R \text{ nats per sample} \quad} \boxed{\text{decoder}} \to \hat{S}_1, \ldots, \hat{S}_n$$

- Mean-squared distortion criterion $\frac{1}{n}\|\mathbf{S} - \hat{\mathbf{S}}\|^2$.
- $\mathbf{S}$ is a memoryless stationary source with continuous alphabet.
- In particular: memoryless Gaussian source, for which $D(R) = \sigma^2 e^{-2R}$.

Sparse Regression Codes

$$\mathbf{A} = \left.\begin{pmatrix} \overbrace{\qquad\qquad}^{\substack{\text{Section 1}\\ M \text{ columns}}} & \Big| & \overbrace{\qquad\qquad}^{\substack{\text{Section 2}\\ M \text{ columns}}} & \Big| & \dots & \Big| & \overbrace{\qquad\qquad}^{\substack{\text{Section L}\\ M \text{ columns}}} \end{pmatrix}\right\} n$$

$$\beta = \begin{pmatrix} c_1 & | & c_2 & | & \dots & | & c_L \end{pmatrix}^\top$$

- Entries of $\mathbf{A}$ are i.i.d. $\sim \mathcal{N}(0, \frac{1}{n})$, known a-priori to encoder+decoder.
- $\mathbf{A}$ consists of L sections with M columns each.
- Entries of β are sparse: exactly one non-zero entry per section.
- Entries of β are pre-fixed.
- Codewords of the SPARC are the linear combinations $\mathbf{A}\beta$.

Sparse Regression Codes

$$
\mathbf{A} = \left.\begin{pmatrix} \overbrace{\qquad\qquad\qquad}^{\substack{\text{Section 1}\\ M\text{ columns}}} & \Bigg| & \overbrace{\qquad\qquad\qquad}^{\substack{\text{Section 2}\\ M\text{ columns}}} & \Bigg| & \dots & \Bigg| & \overbrace{\qquad\qquad\qquad}^{\substack{\text{Section L}\\ M\text{ columns}}} \end{pmatrix}\right\} n
$$

$$
\beta = \begin{pmatrix} c_1 & | & c_2 & | & \dots & | & c_L \end{pmatrix}^\top
$$

- Locations of non-zero entries $\{c_1, c_2, \dots, c_L\}$ in β are determined by the input bits.
- Decoder receives $Y = \mathbf{A}\beta + Z$ and has to find $\hat{\beta}$.
- Total codewords $M^L = \exp(nR) \implies R = \frac{L \log M}{n}$.

Choosing M and L

$$\mathbf{A} = \left.\begin{pmatrix} \overbrace{\qquad\qquad}^{\substack{\text{Section 1}\\ M\text{ columns}}} & \Bigg| & \overbrace{\qquad\qquad}^{\substack{\text{Section 2}\\ M\text{ columns}}} & \Bigg| & \cdots & \Bigg| & \overbrace{\qquad\qquad}^{\substack{\text{Section L}\\ M\text{ columns}}} \end{pmatrix}\right\} n$$

$$\beta = \begin{pmatrix} c_1 & | & c_2 & | & \ldots & | & c_L \end{pmatrix}^\top$$

For set values of n and R there are many valid choices for M and L

Example (Constant $L = 1$)

- Can be reliable for any rate $R < \mathcal{C}$.
- But the size of $\mathbf{A}$ grows exponentially in n.

Choosing M and L

Practical choice: $M = L^b$ for some $b > 0$

In this case,

$$nR = L \log M = bL \log L.$$

To solve L for given n, R and b, we make use of the Lambert W-function, which is the inverse relation of the function $f(z) = ze^z$.

Lambert W-Function

- Applying the Lambert W-function gives

$$L = \exp\left(W\left(\frac{nR}{b}\right)\right)$$

- In our work, we derived new tight upper and lower bounds.

$$W(x) = \log x - \log\log x + O\left(\frac{\log\log x}{\log x}\right).$$

- $M = L^b$ implies that the size of the design matrix is $ML \approx \left(\frac{n}{\log n}\right)^{b+1}$.

Power Allocation

$$\mathbf{A} = \left.\begin{pmatrix} \overbrace{\qquad\qquad}^{\substack{\text{Section 1}\\ M \text{ columns}}} & \Big| & \overbrace{\qquad\qquad}^{\substack{\text{Section 2}\\ M \text{ columns}}} & \Big| & \cdots & \Big| & \overbrace{\qquad\qquad}^{\substack{\text{Section L}\\ M \text{ columns}}} \end{pmatrix}\right\} n$$

$$\beta = \begin{pmatrix} c_1 & | & c_2 & | & \ldots & | & c_L \end{pmatrix}^\top$$

- Non-zero coefficients of β are pre-fixed: $c_i = \sqrt{nP_i}$ for $i \in \{1, \ldots, L\}$.
- Power restriction is met by $\sum_{i=1}^{L} P_i = P$.

Examples

- Flat: $P_i = \frac{P}{n}$
- Exponentially decaying: $P_i \propto e^{-ai/L}$ for some $a > 0$

Lossy Compression with SPARCs

$$
\begin{array}{c}
\begin{array}{ccccc}
\overbrace{\hspace{3em}}^{\substack{\text{Section 1}\\ M \text{ columns}}} & & \overbrace{\hspace{3em}}^{\substack{\text{Section 2}\\ M \text{ columns}}} & & \overbrace{\hspace{3em}}^{\substack{\text{Section L}\\ M \text{ columns}}}
\end{array}\\
\mathbf{A} = \left.\left(\begin{array}{c|c|c|c} \quad & \quad & \cdots & \quad \end{array}\right)\right\} n \\
\beta = \left(\begin{array}{c|c|c|c} c_1 & c_2 & \cdots & c_L \end{array} \right)^\top
\end{array}
$$

- Entries are i.i.d. $\sim \mathcal{N}(0,1)$.
- No power restriction $\Rightarrow$ coefficients $\{c_1, \ldots, c_L\}$ are chosen to minimize distortion.
- $R = \frac{L \log M}{n}$, with our setting where $M = L^b$, $b > 0$.

Optimal Encoding

Optimal Encoder

Minimum distance encoding:

$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{S} - \mathbf{A}\beta\|^2,$$

where $\mathbf{S}$ is the source sequence and $\|\cdot\|$ is the Euclidean norm.

Theorem (Venkataramanan and Tatikonda '17)

For a memoryless Gaussian source $\mathbf{S}$ with zero mean and variance σ^2, there exists a sequence of rate R SPARCs with

$$P\left(\frac{1}{n}\|\mathbf{S} - \mathbf{A}\beta\|^2 > D\right) < e^{-n(E^*(R,D)+O(1))},$$

where $E^*(R,D)$ is the optimal error exponent

- Not feasible since the number of codewords is exponential in n.

Successive Cancellation Encoding

Successive Cancellation Encoder [Venkataramanan et al., '14]

- Set $\mathbf{R}_0 = \mathbf{S}$
- For $i \in \{1, \ldots, L\}$, choose the m_i-th column in section i with

$$m_i \triangleq \underset{j:(i-1)M<j\leq iM}{\arg\max} \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle,$$

  where $\mathbf{A}_j$ is the j's column of $\mathbf{A}$.
- Define recursively the residual

$$\mathbf{R}_i = \mathbf{R}_{i-1} - c_i \mathbf{A}_{m_i}, \quad i \in \{1, \ldots, L\},$$

  with $c_i = \sigma\sqrt{\frac{2R}{L}\left(1 - \frac{2R}{L}\right)^{i-1}}$
- After L steps, β is a sparse vector comprised of values $\{c_i\}_1^L$ in indices $\{m_i\}_1^L$ and zero elsewhere. The codeword is $\mathbf{A}\beta$.

Successive Cancellation Encoding

Theorem (Venkataramanan et al., '14)

For an ergodic source $\mathbf{S}$ *with mean* 0 *and variance* σ^2*, the encoding algorithm produces a codeword* $\mathbf{A}\beta$ *that satisfies the following for large enough* M *and* L*,*

$$P\left(\frac{1}{n}\|\mathbf{S}-\mathbf{A}\beta\|^2 > \sigma^2 e^{-2R}(1+e^R\Delta)^2\right) < p_0 + p_1 + p_2,$$

where $\Delta = \delta_0 + 5R(\delta_1+\delta_2)$ *for any positive constants* δ_0*,* δ_1 *and* δ_2 *such that* $\Delta < \frac{1}{2}$*, and with*

$$p_0 = P\left(\left|\frac{|\mathbf{S}|}{\sigma} - 1\right| > \delta_0\right), \qquad p_2 = \left(\frac{M^{2\delta_2}}{8\log M}\right)^{-L},$$

$$p_1 = 2MLe^{-n\delta_1^2/8},$$

where $|\cdot| = \frac{\|\cdot\|}{\sqrt{n}}$ *is the scaled norm.*

Successive Cancellation Encoding

Corollary

- If $\mathbf{S}$ is a memoryless Gaussian source sequence, the SPARC attains the distortion-rate function.
- For any distortion D greater than $D(R) = \sigma^2 e^{-2R}$, the probability of excess distortion decays exponentially in n.

Complexity

L stages involving M inner products of vectors of length $n \implies O(nML)$
The encoder is polynomial in n

What is missing?

- No finite-length source sequence analysis
- Very loose bound

Preliminaries

Notation

Let $X_1, \ldots, X_M$ be i.i.d. standard Gaussian random variables, then

$$Z_M = \max_{i \in 1,\ldots,M} X_i,$$
$$e_M = E[Z_M]$$

Lemma [Cramér '46]

$$e_M = \sqrt{2 \log M} - \frac{\log \log M + \log 4\pi - 2\gamma}{2\sqrt{2 \log M}} + O\left(\frac{\log \log M}{\log M}\right),$$

where γ is the Euler-Mascheroni constant

Preliminaries (cont.)

Upper Bound on e_M

For all $M \geq 1$,

$$e_M \leq \sqrt{2 \log M}$$

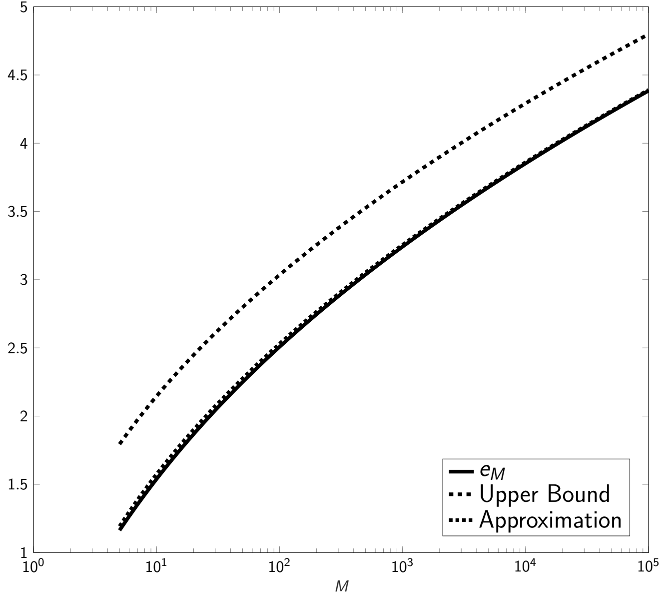
Proof: by invoking Jensen's inequality,

$$\exp(t\, e_M) \leq E\left[\exp\left(t\, Z_M\right)\right] \leq \sum_{i=1}^{M} E\left[\exp\left(t\, X_i\right)\right] = M \exp\left(\tfrac{1}{2}\, t^2\right)$$

$$\Rightarrow e_M \leq \frac{\log M}{t} + \frac{t}{2},$$

for all $t > 0$. Minimization over t yields $t = \sqrt{2 \log M}$, which leads to the required result.

Preliminaries (cont.)



Modification of the Encoder

Reminder

In each step:

$$m_i = \underset{j:(i-1)M<j\leq iM}{\arg\max} \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle,$$

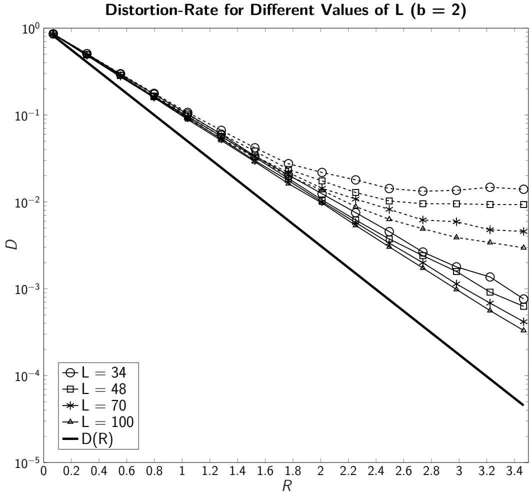
$$\mathbf{R}_i = \mathbf{R}_{i-1} - c_i \mathbf{A}_{m_i},$$

with $c_i = \sigma \sqrt{\frac{2R}{L}\left(1 - \frac{2R}{L}\right)^{i-1}}$

Can we choose better coefficients $\{c_1, \ldots, c_L\}$, so that $|\mathbf{R}_L|^2$ is smaller?

$$|\mathbf{R}_i|^2 \approx |\mathbf{R}_{i-1}|^2 \left(1 + \frac{c_i^2}{|\mathbf{R}_{i-1}|^2} - \frac{2c_i}{|\mathbf{R}_{i-1}|} \cdot \frac{e_M}{\sqrt{n}}\right)$$

Modification of the Encoder



After minimization,

$$c_i = \sqrt{\frac{\sigma^2 e_M^2}{n}\left(1 - \frac{e_M}{n}\right)^{i-1}}$$

$$|\mathbf{R}_i|^2 \approx \sigma^2\left(1 - \frac{e_M^2}{n}\right)^i$$

Remark: replacing

$$e_M \approx \sqrt{2\log M}$$

yields the original $\{c_i\}_{i=1}^L$

New Main Theorem: Definitions

Definition: p_1 and p_2

For arbitrary $\delta_1, \delta_2 > 0$, let

$$p_1 = P\left(\frac{1}{L}\sum_{i=1}^{L}|\gamma_i| > \delta_1\right), \qquad p_2 = P\left(\frac{1}{L}\sum_{i=1}^{L}|\epsilon_i| > \delta_2\right),$$

where $\{\gamma_i\}_{i=1}^{L}$ and $\{\epsilon_i\}_{i=1}^{L}$ are defined by

$$\left\langle \mathbf{A}_{m_i}, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle = e_M(1+\epsilon_i),$$

$$|\mathbf{A}_{m_i}|^2 = 1 + \gamma_i.$$

New Main Theorem

Theorem 1

For an ergodic source $\mathbf{S}$ with mean 0 and variance σ^2, the encoding algorithm produces a codeword $\mathbf{A}\beta$ that satisfies the following for $L \geq 10R$,

$$P\left(\frac{1}{n}\|\mathbf{S} - \mathbf{A}\beta\|^2 > \underbrace{\sigma^2\left(1 - \frac{e_M^2}{n}\right)^L (1 + w^L\Delta)^2}_{\alpha_{L,\Delta}}\right) < p_0 + p_1 + p_2,$$

where $\Delta = \delta_0 + 5R(\delta_1 + \delta_2)$ for any positive constants δ_0, δ_1 and δ_2 such that $\Delta < \frac{1}{2}$, and where

$$w = 1 + \frac{e_M^2}{2(n - e_M^2)}.$$

Discussion

Main Message

- The modified coding scheme performs favorably for finite length codes.
- It is robust in the following sense: for any ergodic source, the proposed encoder achieves the optimal distortion-rate function of an i.i.d Gaussian source with the same variance.

Note

Asymptotically, the difference between the Gaussian distortion-rate function $D(R) = \sigma^2 e^{-2R}$ and $\alpha_{L,\Delta}$, can be arbitrarily small, by choosing proper L and Δ.

Main contributions

- Better performance with the modified encoder for finite-length codes.
- Analysis of the modified encoder with finite L and M.
- The bound on the probability of excess distortion is substantially tighter.

Outline of Proof

Notation

Denote the multiplicative deviation of $|\mathbf{R}_i|^2$ from its approximated value by Δ_i,

$$|\mathbf{R}_i|^2 = \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^i (1 + \Delta_i)^2$$

Idea:

- Find a recursion for Δ_i
- Use this recursion to bound Δ_L by taking the worst case scenario at each step
- Worst case scenario: the deviations Δ_0, $\{\gamma_i\}_{i=1}^L$ and $\{\epsilon_i\}_{i=1}^L$ are maximal and have the same sign

Upper Bound on p_0 (General Case)

For all $\delta \in (0,1)$,

$$p_0 \triangleq P\left(\left|\frac{|\mathbf{S}|}{\sigma} - 1\right| > \delta\right) \leq \inf_{t>0}\left\{e^{-n\left[t(1+\delta)^2 - \log M_{X^2}\left(\frac{t}{\sigma^2}\right)\right]}\right\} + \inf_{t>0}\left\{e^{n\left[t(1-\delta)^2 + \log M_{X^2}\left(-\frac{t}{\sigma^2}\right)\right]}\right\},$$

where:

- X is a random variable with zero mean and finite variance σ^2.
- $M_{X^2}(t)$ is the moment-generating function of X^2.
- $\mathbf{S}$ is an i.i.d. source sequence of length n, with $S_i \sim P_X$.

- Proof: apply the Chernoff Bound twice
- The bound is exponentially decaying in n

Exact Expression for p_0 (Gaussian Case)

Let $\mathbf{S}$ be an i.i.d. source sequence of length n, generated according to the Gaussian distribution $\mathcal{N}(0, \sigma^2)$, and let $\delta \in (0, 1)$. Then,

$$p_0 = 1 - \bar{\gamma}\left(\frac{n}{2}, \frac{n(1+\delta)^2}{2}\right) + \bar{\gamma}\left(\frac{n}{2}, \frac{n(1-\delta)^2}{2}\right)$$

where $\bar{\gamma}$ is the incomplete Gamma function,

$$\bar{\gamma}(a, x) = \frac{1}{\Gamma(a)} \int_0^x t^{a-1} e^{-t} \mathrm{d}t.$$

- Proof: follows from $\frac{\|\mathbf{S}\|^2}{\sigma^2} \sim \chi^2(n)$ (Chi-squared distributed with n degrees of freedom)

Upper Bound on p_1

For all $\delta \in (0,1)$,

$$p_1 \triangleq P\left(\frac{1}{L}\sum_{i=1}^{L}|\gamma_i| > \delta\right)$$
$$\leq \left(\inf_{0<t<\frac{1}{2}}\left\{\left(\frac{e^{t(1-\delta)}}{\sqrt{1+2t}}\right)^n \bar{\gamma}\left(\tfrac{n}{2},\tfrac{n}{2}+nt\right) + \left(\frac{e^{-t(1+\delta)}}{\sqrt{1-2t}}\right)^n \left[1-\bar{\gamma}\left(\tfrac{n}{2},\tfrac{n}{2}-nt\right)\right]\right\}\right)^L$$

Reminders:

- $\gamma_i = |\mathbf{A}_{m_i}|^2 - 1$
- $\bar{\gamma}(\cdot,\cdot)$ denotes the incomplete Gamma function

Upper Bound on p_1 - Proof

Concept of proof:

- The columns of $\mathbf{A}$ are i.i.d.
- Chernoff's bound
- For any $k \in \{1, \ldots, ML\}$, we have $\sum_{j=1}^{n} A_{j,k}^2 \sim \chi^2(n)$, where $A_{j,k}$ is the (j, k) entry in matrix $\mathbf{A}$

Upper Bound on p_2

For all $\delta \in (0,1)$,

$$p_2 \triangleq P\left(\frac{1}{L}\sum_{i=1}^{L}|\epsilon_i| > \delta\right) \le \left(\inf_{t>0}\left\{e^{-te_M\delta}\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right\}\right)^L,$$

where $f_{Z_M}(\cdot)$ is the probability density function of Z_M.

- Reminder: ϵ_i is the random variable which stands for the deviation

$$\max_{(i-1)M<j\le iM}\left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|}\right\rangle = e_M(1+\epsilon_i).$$

- Concept of proof: Chernoff bound after showing that $\{\epsilon_i\}_{i=1}^L$ are i.i.d.

Upper Bound on p_2

Reminder

$$p_2 \leq \left(\inf_{t>0} \left\{ e^{-te_M\delta} \int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z \right\} \right)^L$$

Efficient Computation of the Bound

The infimum in the bound of p_2 is a minimum, and can be obtained numerically by the bisection method in the interval $[0, t_M^*]$, with

$$t_M^* = \frac{1}{e_M} \log \left(\frac{\sqrt{2}M}{\sqrt{\pi} f_{Z_M}(-2e_M)} \right)$$

Concept of proof:

- Convexity
- When $t = t_M^*$, the derivative is positive

The Complete Bound - Comparison

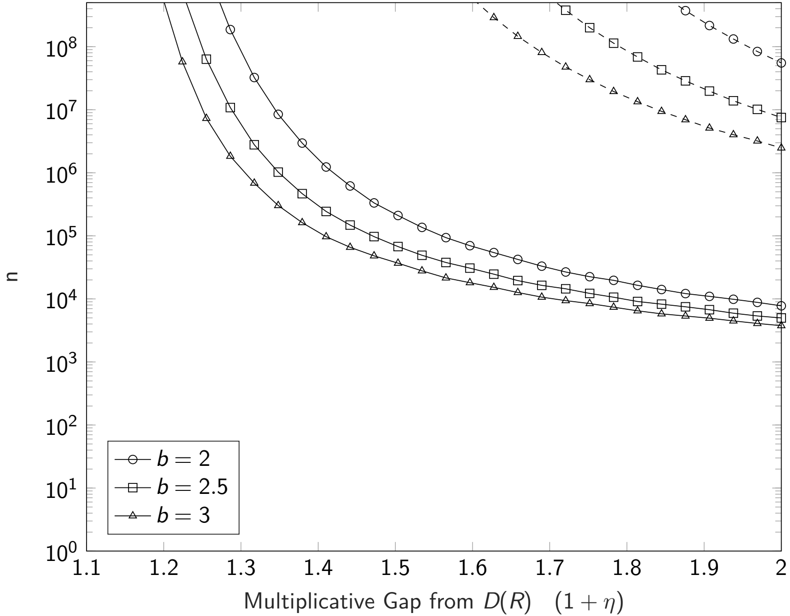
- The simulation was conducted as follows:
  - Fix parameters R, b, σ^2, ϵ
  - Calculate the minimal n such that

$$P\left[|\mathbf{S} - \mathbf{A}\hat{\beta}|^2 > \sigma^2 e^{-2R}(1+\eta)\right] < \epsilon,$$

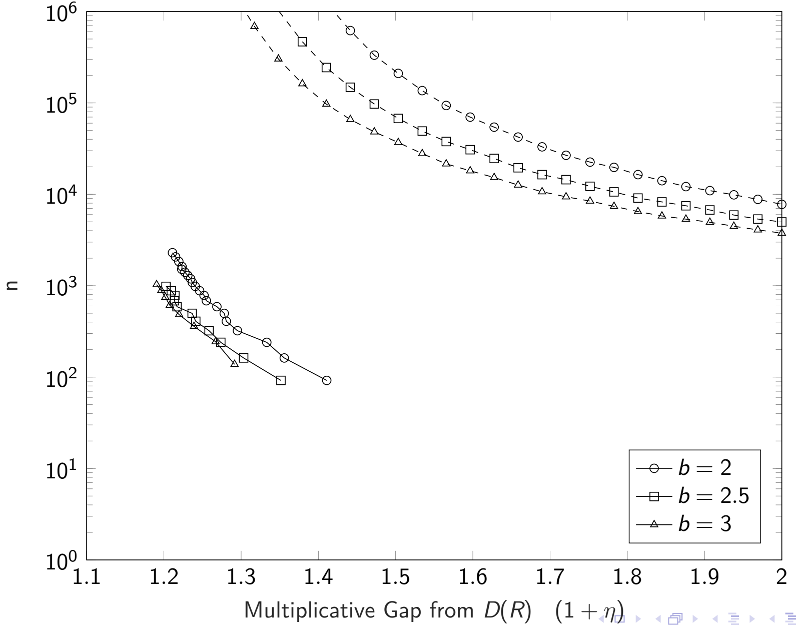
    for a range of values $\eta > 0$

- The following figures use $R = 0.5$, $\sigma^2 = 1$ and $\epsilon = 0.01$

The Complete Bound - Comparison



The Complete Bound - Comparison



Deviation from the Distortion-Rate Function

Bounds on $D_{L,\Delta} \triangleq \left|\alpha_{L,\Delta} - \sigma^2 e^{-2R}\right|$

- For a large enough L,

$$\frac{B \log\log L}{\log L} + O\left(\frac{\epsilon(L)\log\log L}{\log L}\right) \leq D_{L,\Delta},$$

  where $\epsilon(\cdot)$ is a non-negative function with $\epsilon(L) \xrightarrow[L\to\infty]{} 0$ and $B > 0$.
- If $\Delta = \frac{A \log\log L}{b \log L}$ with a constant $A > \frac{1}{4}$, then for a large enough L,

$$D_{L,\Delta} \leq \frac{C \log\log L}{\log L} + O\left(\frac{\epsilon(L)\log\log L}{\log L}\right),$$

  with $C > 0$.

Deviation from the Distortion-Rate Function

Conclusions:

- The deviation from the distortion-rate function is upper bounded by $O\left(\frac{\log\log L}{\log L}\right)$ with probability that tends to 1 as $L \to \infty$.
- Using Theorem 1, no significant improvement can be made to the upper bound.

Note

- The demand for the source sequence to be memoryless Gaussian is necessary to apply the upper bound on p_0.
- For other memoryless source distributions, the same results can be achieved under a suitable condition.

Topics for Future Work

- Further Tightening the bound on the probability of excess distortion.
- Can the lossy compression scheme be generalized, in order to approach the rate-distortion function of non-Gaussian sources?
- Improving the tradeoff between complexity and performance of SPARCs .
- Channel coding via Spatially-coupled SPARCs shows better empirical results than with regular SPARCs. Finite blocklength analysis is currently missing, as well as lossy compression.

Thank You!

Modification to the Encoder

Lemma (Venkataramanan, Sarkar, Tatikonda '14)

Let $\{\mathbf{A}_j\}_{j=1}^N$ be N mutually independent random vectors of length n, and suppose that the components of each vector are i.i.d. standard Gaussian random variables. Let $\mathbf{R}$ be a random vector independent of $\{\mathbf{A}_j\}_{j=1}^N$ whose support lies on the n-dimensional unit sphere, i.e.,

$$\sum_{i=1}^{n} r_i^2 = 1,$$

and let

$$T_j = \langle \mathbf{A}_j, \mathbf{R} \rangle$$

for every $j \in \{1, \ldots, N\}$. Then, $\{T_j\}_{j=1}^N$ are i.i.d. standard Gaussian random variables which are independent of $\mathbf{R}$

Outline of Proof

Notation

Denote the multiplicative deviation of $|\mathbf{R}_i|^2$ from its approximated value by Δ_i,

$$|\mathbf{R}_i|^2 = \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^i (1 + \Delta_i)^2$$

Idea: find a recursion for Δ_i, and by it control the deviation after L steps, Δ_L

$$|\mathbf{R}_i|^2 = |\mathbf{R}_{i-1}|^2 + c_i^2 \left|\mathbf{A}_{m_i}\right|^2 - \frac{2c_i \|\mathbf{R}_{i-1}\|}{n} \left\langle \mathbf{A}_{m_i}, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle$$

$$\vdots$$

$$= \sigma^2 \left(1 - \frac{e_M^2}{n}\right)^i \left[(1 + \Delta_{i-1})^2 + \frac{e_M^2}{n - e_M^2}\Big(\Delta_{i-1}^2 + \gamma_i - 2\epsilon_i(1 + \Delta_{i-1})\Big)\right]$$

Outline of Proof

$$\Rightarrow (1+\Delta_i)^2 = (1+\Delta_{i-1})^2 + \left(\frac{e_M^2}{n-e_M^2}\right)\left(\Delta_{i-1}^2 + \gamma_i - 2\epsilon_i(1+\Delta_{i-1})\right)$$

Next step to bound Δ_L: define event $\mathcal{A}$ to satisfy the following conditions,

- $\left|\frac{|\mathbf{S}|}{\sigma} - 1\right| \leq \delta_0$
- $\frac{1}{L}\sum_{i=1}^{L} |\gamma_i| \leq \delta_1$
- $\frac{1}{L}\sum_{i=1}^{L} |\epsilon_i| \leq \delta_2$

By the union bound,

$$P(\mathcal{A}^c) \leq p_0 + p_1 + p_2$$

What remains: finding a bound on Δ_L conditioning on event $\mathcal{A}$

Outline of Proof

Lemma

For $L \geq 10R$, conditioning on event $\mathcal{A}$,

$$|\Delta_i| \leq |\Delta_0| w^i + \frac{2e_M^2}{n - e_M^2} \sum_{j=1}^{i} w^{i-j}(|\gamma_j| + |\epsilon_j|), \quad i \in \{1, \ldots, L\}$$

The lemma is proved by induction, utilizing the recursion of Δ_i, and taking the worst case scenario: the deviations Δ_0, $\{\gamma_i\}_{i=1}^L$ and $\{\epsilon_i\}_{i=1}^L$ are maximal and have the same sign

$$\begin{aligned}
\Rightarrow \Delta_L &\leq |\Delta_0| w^L + \frac{2e_M^2}{n - e_M^2} \sum_{j=1}^{L} w^{L-j}(|\gamma_j| + |\epsilon_j|) \\
&\leq w^L \left(\delta_0 + \frac{2e_M^2 L}{n - e_M^2} (\delta_1 + \delta_2) \right) \\
&\leq w^L \Delta
\end{aligned}$$

Upper Bound on p_2 - Proof

First show that $\{\epsilon_i\}_{i=1}^{L}$ are i.i.d.
Define

$$T_j^{(i)} \triangleq \left\langle \mathbf{A}_j, \frac{\mathbf{R}_{i-1}}{\|\mathbf{R}_{i-1}\|} \right\rangle, \quad j \in \{(i-1)M+1, \ldots, iM\}$$
$$T^{(i)} \triangleq \left(T^{(i)}_{(i-1)M+1}, \ldots, T^{(i)}_{iM} \right)$$

- The entries of $\mathbf{A}$ are i.i.d. $\Rightarrow$ $\mathbf{A}_j$ is independent of $\mathbf{R}_{i-1}$ conditioned on $\{T^{(i-1)}, \ldots, T^{(1)}, \mathbf{R}_0\}$ for any $j \in \{(i-1)M+1, \ldots, iM\}$
- $\{T_j^{(i)}\}$ are i.i.d. standard Gaussian random variables (previous lemma)

$$\Rightarrow F_{T^{(i)} | T^{(i-1)}, \ldots, T^{(1)}, \mathbf{R}_0} = F_{T^{(i)}} = \prod_{j=(i-1)M+1}^{iM} F_{T_j^{(i)}},$$

and applying recursively,

$$F_{\mathbf{R}_0, T^{(1)}, \ldots, T^{(L)}} = F_{\mathbf{R}_0} \prod_{i=1}^{L} F_{T^{(i)}}$$

Upper Bound on p_2 - Proof

Conclusions:

- $\{T^{(i)}\}$ are i.i.d. Gaussian random vectors
- They are independent of $\mathbf{R}_0$
- Their components are i.i.d. standard Gaussian random variables

The maximal values of $\{T^{(i)}\}$,

$$V_i = \max_{(i-1)M<j\leq iM} T_j^{(i)},$$

are i.i.d. $\sim Z_M$; the deviations $\{\epsilon_i\}_{i=1}^{L}$, given by $\epsilon_i = \frac{V_i}{e_M} - 1$, are i.i.d.

– The Lemma follows from Chernoff's bound

But can this bound be computed efficiently?

Efficient Computation of the Upper Bound on p_2

The bound can be rewritten as

$$p_2 \leq \exp\left(-L \sup_{t>0}\left\{te_M\delta - \log\left(\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right)\right\}\right)$$

Define, for $t > 0$,

$$\begin{aligned} u_{\delta,M}(t) &\triangleq \frac{\mathrm{d}}{\mathrm{d}t}\left(\log\left(\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right) - te_M\delta\right) \\ &= \frac{E\left[|Z_M - e_M|\, e^{t|Z_M-e_M|}\right]}{E\left[e^{t|Z_M-e_M|}\right]} - e_M\delta \end{aligned}$$

$u_{\delta,M}(t)$ is monotonically increasing (Cauchy-Schwarz inequality)
$\Rightarrow$ the supremum can be obtained by the bisection method in the interval $[0, t_M^*]$ for any $t_M^* > 0$ such that $u_{\delta,M}(t_M^*) \geq 0$

But does such t_M^* always exist?

Efficient Computation of the Upper Bound on p_2

If $t_M^* = \frac{1}{e_M} \log\left(\frac{\sqrt{2}M}{\sqrt{\pi} f_{Z_M}(-2e_M)}\right)$, then $u_{\delta,M}(t_M^*) \geq 0$

Proof: it is enough to show that

$$E\left[|Z_M - e_M|\, e^{t_M^*|Z_M - e_M|}\right] - e_M E\left[e^{t_M^*|Z_M - e_M|}\right] \geq 0$$

since $\delta \in (0,1)$. Analysis of the expression yields

$$\begin{aligned} E\left[|Z_M - e_M|\, e^{t_M^*|Z_M - e_M|}\right] - e_M E\left[e^{t_M^*|Z_M - e_M|}\right] \\ &= \int_{-\infty}^{\infty} \left(|z - e_M| - e_M\right) e^{t|z - e_M|} f_{Z_M}(z)\,\mathrm{d}z \\ &\vdots \\ &\geq e_M^2 \left(f_{Z_M}(-2e_M)\, e^{te_M} - M\sqrt{\frac{2}{\pi}}\right) e^{te_M} \end{aligned}$$

First Item

Lemma

Let $\mathbf{S}$ be a memoryless Gaussian source sequence. If $\Delta = \frac{A \log\log L}{b \log L}$ for some $A > \frac{1}{4}$, there exists a division of Δ into δ_0, δ_1 and δ_2 such that

$$\lim_{L\to\infty} p_0 + p_1 + p_2 = 0$$

Previously,

$$p_0 + p_1 \leq 2e^{-\frac{3n\delta_0^2}{4}} + 2MLe^{-\frac{n\delta_1^2}{8}}$$

$\Rightarrow$ If $\delta_0 = \delta_1 = \frac{1}{\log L}$, then $\lim_{L\to\infty} p_0 + p_1 = 0$

For p_2, if $\delta_2 = \frac{A \log\log L}{b \log L}$,

$$p_2 \leq \exp\left[-L \sup_{t>0}\left\{\left(t e_M \delta_2 - \log\left(\int_{-\infty}^{\infty} e^{t|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right)\right)\right\}\right]$$
$$\leq \exp\left[-L\left(\frac{A e_M^2 \log\log M}{\log M} - \log\left(\int_{-\infty}^{\infty} e^{e_M|z-e_M|} f_{Z_M}(z)\,\mathrm{d}z\right)\right)\right]$$

For a large enough M, the integral can be bounded by $C \cdot e_M$, for a constant $C > 0$. From $e_M \leq \sqrt{2 \log M}$,

$$\log(e_M) + C \leq \frac{e_M^2 \log\log M}{4 \log M} + C',$$

and since $A > \frac{1}{4}$,

$$\lim_{L\to\infty} p_2 = 0$$

p_0 Comparison

The bound from previous works for the Gaussian case was

$$p_0 \leq 2e^{-\frac{3n\delta^2}{4}}$$

Comparison:

$(n,\ \delta)$	Exact p_0 (Lemma 2)	Lemma 1 Bound	Previous Bound
$(10,\ 0.1)$	0.66	> 1	> 1
$(1000,\ 0.1)$	$7.86 \cdot 10^{-6}$	$6.19 \cdot 10^{-5}$	$1.1 \cdot 10^{-3}$

p_1 Comparison

The upper bound on p_1 in previous works was

$$p_1 \leq 2MLe^{-\frac{n\delta^2}{8}}$$

Comparison:

(n, L, M, δ)	Lemma 3 Bound	Looser Bound (Remark)	Previous Bound
$(10^2, 10, 10^2, 0.25)$	$1.83 \cdot 10^{-4}$	$3.30 \cdot 10^{-4}$	> 1
$(10^3, 10, 10^3, 0.10)$	$1.57 \cdot 10^{-8}$	$1.83 \cdot 10^{-8}$	> 1

p_2 Comparison

The upper bound on p_2 in previous works was

$$p_2 < \left(\frac{M^{2\delta}}{8\log M}\right)^{-L}$$

Comparison:

$(L,\ M,\ \delta)$	New Bound	Previous Bound
$(10,\ 100,\ 0.25)$	$1.95 \cdot 10^{-2}$	> 1
$(10,\ 1000,\ 0.1)$	$7.96 \cdot 10^{-1}$	> 1
$(10,\ 10^4,\ 0.25)$	$4.15 \cdot 10^{-11}$	$4.7 \cdot 10^{-2}$

Remark

The probability p_2 is defined differently in previous works, since the deviations $\{\epsilon_i\}_{i=1}^{L}$ are defined in relation to a different approximation

Lambert W Function

Known Bounds [Hoorfar and Hassani, '08]

For every $x \geq e$,

$$\log x - \log\log x + \frac{1}{2}\frac{\log\log x}{\log x} \leq W(x) \leq \log x - \log\log x + \frac{e}{e-1}\frac{\log\log x}{\log x}$$

Two bounds on $\exp(W(x))$ can be derived from this bound:

- Applying $\exp(x)$ on both sides of the inequalities
- Using the identity $\exp(W(x)) = \frac{x}{W(x)}$

New Bounds Derived in This Work

For every $x \geq e$,

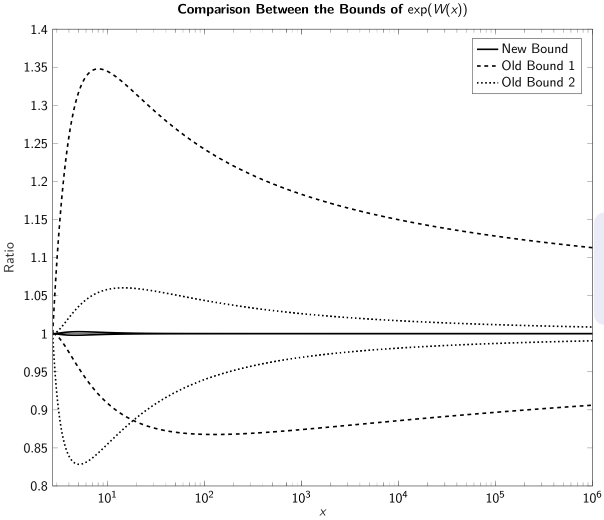
$$s(x) \leq e^{W(x)} \leq t(x),$$

with

$$s(x) \triangleq \frac{x}{\log t(x)}, \qquad t(x) \triangleq \frac{x}{\log x - \log\log\left(\frac{x}{v(x)}\right)},$$

$$v(x) \triangleq \log x - \log\left(\log\left(\frac{x}{\log x}\right) - \log\left(1 - \frac{\log\log x}{1 + \log x}\right)\right)$$

Lambert W Function - Bound Comparison



- Old Bound 1 - exp(bound)
- Old Bound 2 - uses $\exp(W(x)) = \frac{x}{W(x)}$

Decoders

Optimal Decoder

Maximum likelihood decoder:

$$\hat{\beta} = \arg\min_{\beta} \| Y - \mathbf{A}\beta \|,$$

where Y is the output sequence of the channel

- With flat power allocation, for all $R < \mathcal{C}$, the error probability decays exponentially in n [Joseph and Barron '12]
- Impractical: complexity grows exponentially in n

Adaptive Successive Hard-threshold Decoder

Idea: iteratively pick columns of **A** whose inner product with a residual is over a fixed threshold

Complexity

The complexity of the decoder is $O\left(n^2 M\right) \Rightarrow$ Polynomial in n

With exponentially decaying power allocation,

Performance [Joseph and Barron '14]

For $R < \mathcal{C}_M$,

$$P_e \leq e^{-L(\mathcal{C}_M - R)^2 c_1},$$

where

$$\mathcal{C}_M \triangleq \mathcal{C}\left(1 - \frac{c_2}{\log M}\right)$$

Soft Decision Decoders

Two soft-decision decoders:

- Adaptive successive soft-decision decoder [Barron and Cho '12]
- Approximate message-passing (AMP) decoder [Barbier and Krzakala '17], [Rush, Grieg and Venkataramanan '17]

Idea: iteratively update the posterior probabilities of each entry of β being the true non-zero in its section

Performance

With exponential power allocation, for any $R < \mathcal{C}$, the error probability of the soft-decision decoders decays exponentially in $\frac{n}{(\log n)^c}$, where c is a constant that depends on the scheme

Performance is empirically better than the hard-decision decoder

Deviation from the Distortion-Rate Function

Reminder (from Theorem 1)

$$P\left(\frac{1}{n}\|\mathbf{S} - \mathbf{A}\beta\|^2 > \alpha_{L,\Delta}\right) < p_0 + p_1 + p_2$$

Definition

$$D_{L,\Delta} \triangleq \left|\alpha_{L,\Delta} - \sigma^2 e^{-2R}\right|$$

We showed that

$$\lim_{\Delta \to 0} \lim_{L \to \infty} D_{L,\Delta} = 0$$