Coding Schemes Using Constrained Subcodes Over Input-Constrained Channels

V. Arvind Rameshwar

Indian Institute of Science, Bangalore

TU Münich

September 2023

Our work was supported by

Qualcomm innovation fellowship

PMRF
Prime Minister's Research Fellowship
Ministry of Human Resource Development
Government of India

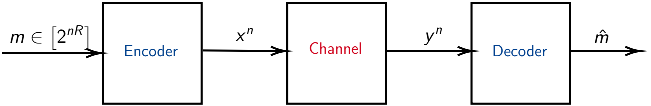
Acknowledgements

Prof. Navin Kashyap, my Ph.D. advisor and collaborator on all the work discussed today

Prof. Henry Pfister (Duke U.), for stimulating discussions

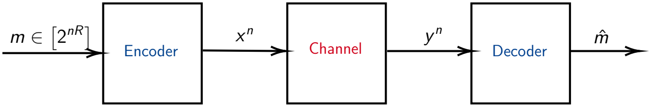
The big picture

- Suppose that we wish to transmit a message m over a noisy medium (or channel):

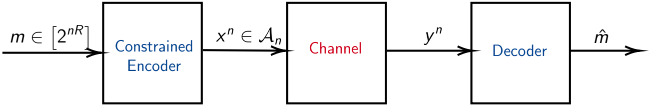


The big picture

- Suppose that we wish to transmit a message m over a noisy medium (or channel):

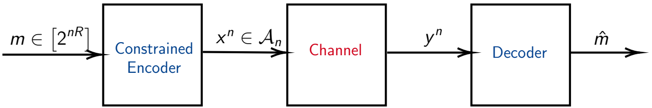


- Suppose also that we require the input x^n to the channel to be "constrained" to lie in $\mathcal{A}_n \subseteq \{0,1\}^n$:



The big picture

- Suppose also that we require the input x^n to the channel to be "constrained" to lie in $\mathcal{A}_n \subseteq \{0,1\}^n$:



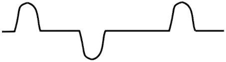
- The broad question to be discussed in this talk:

Q) How does one design good coding schemes over such channels?

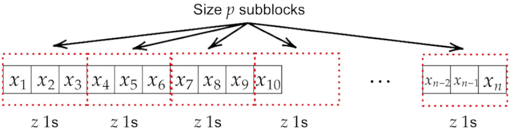
Some input constraints of interest

- Runlength-limited (RLL) constraints: Help alleviate ISI in magneto-optical recording

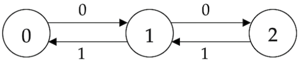
$\ldots 0\,1\,0\,0\,0\,1\,0\,0\,0\,0\,0\,1\,0\,0 \ldots \longleftrightarrow$

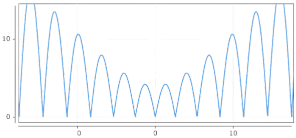


- Subblock composition constraints: Maintain receiver battery levels in energy-harvesting communication



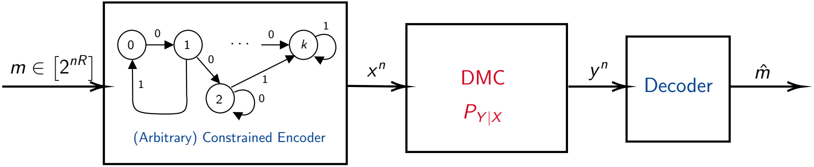
- Charge constraints: Ensure spectral nulls (DC-freeness) in frequency spectrum





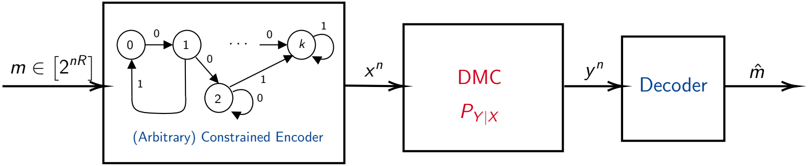
Channel models

Our focus today will be on the class of input-constrained discrete memoryless channels (DMCs)[1]:



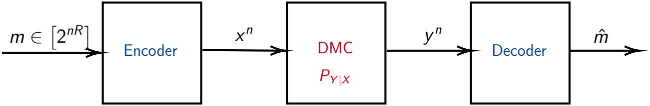
[1] Our bounds for input-constrained combinatorial noise channels can be found in V. A. Rameshwar and N. Kashyap, "Estimating the sizes of binary error-correcting constrained codes," in IEEE JSAIT.

Background on input-constrained memoryless channels



Background on DMCs

- For an unconstrained DMC,



Theorem (Shannon (1948))

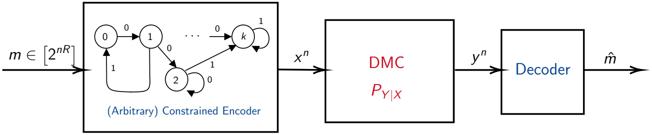
The capacity of an unconstrained DMC is

$$C = \max_{\{P(x)\}} I_P(X;Y).$$ [Single-letter expression!]

Furthermore, explicit capacity-achieving codes such as LDPC codes, RM codes, polar codes, are known.

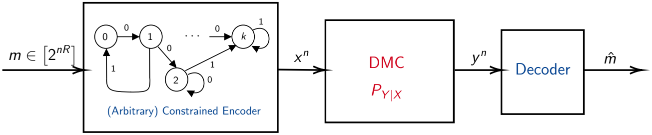
Background on input-constrained DMCs

- We now (re-)introduce our input constraints, represented by (labelled, directed) graphs:



Background on input-constrained DMCs

- We now (re-)introduce our input constraints, represented by (labelled, directed) graphs:



- Such channels form a special class of input-driven finite-state channels (FSCs), with a known initial state.

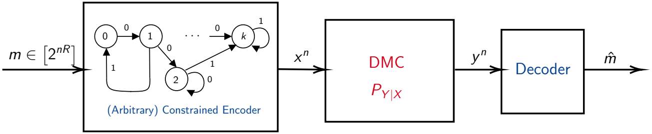
Theorem (Blackwell, Breiman, Thomasian (1958) and Gallager (1968))

The capacity of an input-driven FSC with a fixed, known, initial state s_0 is given by

$$C = \lim_{n\to\infty} \max_{\{P(x^n|s_0)\}} \frac{1}{n} I_P(X^n; Y^n \mid s_0). \qquad \text{[Multi-letter expression!]}$$

Background on input-constrained DMCs

- We now (re-)introduce our input constraints, represented by (labelled, directed) graphs:



- Such channels form a special class of input-driven finite-state channels (FSCs), with a known initial state.
- Explicitly solving for C for general channels is a wide-open problem.
- Evaluating info. rate using simple (Markov) inputs $\equiv$ Computing entropy rate of a Hidden Markov Process [Hard!]

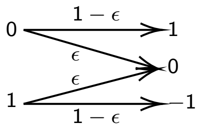
Background on BMS channels

In this talk, we restrict our attention to binary-input memoryless symmetric (BMS) channels:

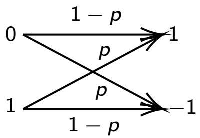
$$Y = (-1)^X \cdot Z,$$

for noise Z independent of X.

Examples:



Binary Erasure Channel (BEC)



Binary Symmetric Channel (BSC)

BMS channels and linear codes

- Suppose that we were to use a linear code $\mathcal{C}$ over the BMS channel.
- Under (optimal) block-MAP decoding, the block-error probabilities are independent of the codeword transmitted.
- Hence, constrained subcodes of $\mathcal{C}$ have the same (average) error probabilities as $\mathcal{C}$ itself!

BMS channels and linear codes

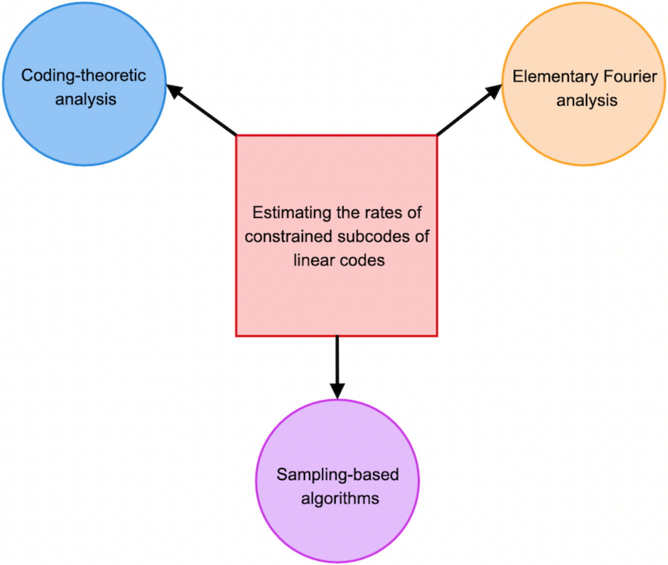
- Suppose that we were to use a linear code $\mathcal{C}$ over the BMS channel.
- Under (optimal) block-MAP decoding, the block-error probabilities are independent of the codeword transmitted.
- Hence, constrained subcodes of $\mathcal{C}$ have the same (average) error probabilities as $\mathcal{C}$ itself!

Our approach: Use constrained subcodes of capacity-achieving codes.

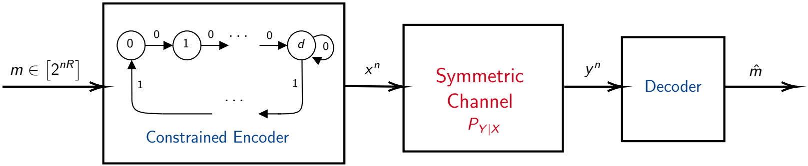
[cf. Reeves & Pfister (2022), Abbe & Sandon (2023), Arikan (2009) ...]

A recurring task: Compute/estimate the rates of constrained subcodes of linear codes.

Agenda


Coding-theoretic analysis
Elementary Fourier analysis
Estimating the rates of constrained subcodes of linear codes
Sampling-based algorithms

Coding Schemes Over (d, ∞)-RLL Input-Constrained BMS Channels Using Reed-Muller (RM) Codes



The (d, ∞)-RLL constraint and RM codes

- A binary sequence satisfies the (d, ∞)-RLL constraint if there exist at least d 0s between every pair of successive 1s.

- For the $(2, \infty)$-RLL constraint,

$$1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1 \quad \checkmark$$

$$1\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 1 \quad \text{X}$$

The r^{th}-order binary RM code $\text{RM}(m, r)$ is defined as the set of binary vectors:

$$\text{RM}(m, r) := \{\text{Eval}(f) : f \in \mathbb{F}_2[x_1, x_2, \ldots, x_m],\ \deg(f) \leq r\},$$

where $\deg(f)$ is the degree of the largest monomial in f and the degree of a monomial $x_S := \prod_{j \in S : S \subseteq [m]} x_j$ is simply $|S|$.

Selected results: explicit linear constrained codes

Theorem

For any $R \in (0,1)$, there exists an explicit sequence of (d,∞)-RLL linear subcodes $\left\{\mathcal{C}_m^{(d)}\right\}_{m\geq 1}$ of a sequence of RM codes of rate R such that

$$\mathit{rate}\left(\mathcal{C}_m^{(d)}\right) \xrightarrow{m\to\infty} R\cdot 2^{-\lceil \log_2(d+1)\rceil}.$$

Selected results: explicit linear constrained codes

Theorem

For any $R \in (0,1)$, there exists an explicit sequence of (d,∞)-RLL linear subcodes $\left\{\mathcal{C}_m^{(d)}\right\}_{m\geq 1}$ of a sequence of RM codes of rate R such that

$$\text{rate}\left(\mathcal{C}_m^{(d)}\right) \xrightarrow{m\to\infty} R \cdot 2^{-\lceil \log_2(d+1) \rceil}.$$

From the previous discussion, a rate of $C \cdot 2^{-\lceil \log_2(d+1) \rceil}$ is achievable, where C is the capacity of the unconstrained BMS channel.

Selected results: explicit linear constrained codes

Theorem
For any $R \in (0,1)$, there exists an explicit sequence of (d,∞)-RLL linear subcodes $\left\{\mathcal{C}_m^{(d)}\right\}_{m\geq 1}$ of a sequence of RM codes of rate R such that

$$\text{rate}\left(\mathcal{C}_m^{(d)}\right) \xrightarrow{m\to\infty} R \cdot 2^{-\lceil \log_2(d+1) \rceil}.$$

Theorem
For any $R \in (0,1)$ and for any sequence of RM codes of rate R, the largest rate of linear (d,∞)-RLL constrained subcodes is $\frac{R}{d+1}$.

Selected results: explicit linear constrained codes

Theorem

For any $R \in (0,1)$, there exists an explicit sequence of (d,∞)-RLL linear subcodes $\left\{\mathcal{C}_m^{(d)}\right\}_{m\geq 1}$ of a sequence of RM codes of rate R such that

$$\text{rate}\left(\mathcal{C}_m^{(d)}\right) \xrightarrow{m\to\infty} R \cdot 2^{-\lceil \log_2(d+1)\rceil}.$$

Theorem

For any $R \in (0,1)$ and for any sequence of RM codes of rate R, the largest rate of linear (d,∞)-RLL constrained subcodes is $\frac{R}{d+1}$.

> The linear constrained subcodes we constructed are hence essentially rate-optimal!

Selected results: existence of nonlinear subcodes

Theorem

For any $R \in (0,1)$*, there exists a sequence of* $(1,\infty)$*-RLL subcodes* $\left\{\hat{\mathcal{C}}_m^{(d)}\right\}_{m\geq 1}$ *of a sequence of RM codes of rate R such that*

$$\textit{rate}\left(\hat{\mathcal{C}}_m^{(d)}\right) \xrightarrow{m\to\infty} \max\left(0, R - \frac{3}{8}\right).$$

Selected results: existence of nonlinear subcodes

Theorem

For any $R \in (0,1)$, there exists a sequence of $(1,\infty)$-RLL subcodes $\left\{\hat{\mathcal{C}}_m^{(d)}\right\}_{m\geq 1}$ of a sequence of RM codes of rate R such that

$$\text{rate}\left(\hat{\mathcal{C}}_m^{(d)}\right) \xrightarrow{m\to\infty} \max\left(0, R-\frac{3}{8}\right).$$

These subcodes are necessarily non-linear for $R > 0.75$, since then $R - \frac{3}{8} > \frac{R}{2}$.

How good are these rate lower bounds?

A benchmark via the probabilistic method

- Consider an $[n, nR]$ *random* linear code obtained via a random parity-check matrix

$$H = \begin{bmatrix} 0 & 1 & 1 & 0 & \dots & 1 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 0 & 1 & \dots & 1 & 1 \end{bmatrix}$$

1. Entries are i.i.d. Ber(0.5)
2. H has $n(1-R)$ rows
3. H is full rank w.h.p.

A benchmark via the probabilistic method

- Consider an $[n, nR]$ *random* linear code obtained via a random parity-check matrix

$$H = \begin{bmatrix} 0 & 1 & 1 & 0 & \dots & 1 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 0 & 1 & \dots & 1 & 1 \end{bmatrix}$$

1. Entries are i.i.d. Ber(0.5)
2. H has $n(1-R)$ rows
3. H is full rank w.h.p.

- Assume that H is full rank. Now, for any $x \in \{0,1\}^n$,

$$\Pr[\boldsymbol{x} \in \mathrm{C}] = \Pr[H \cdot \boldsymbol{x}^T = 0] = \left(\frac{1}{2}\right)^{n(1-R)}.$$

A benchmark via the probabilistic method

- Consider an $[n, nR]$ *random* linear code obtained via a random parity-check matrix

$$H = \begin{bmatrix} 0 & 1 & 1 & 0 & \dots & 1 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 0 & 1 & \dots & 1 & 1 \end{bmatrix}$$

1. Entries are i.i.d. Ber(0.5)
2. H has $n(1-R)$ rows
3. H is full rank w.h.p.

- Assume that H is full rank. Now, for any $\mathrm{x} \in \{0,1\}^n$,

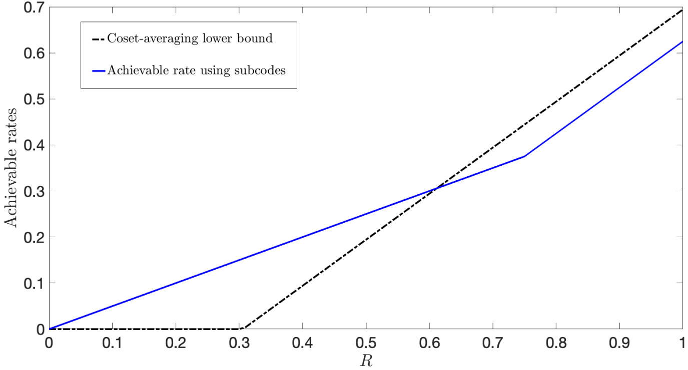
$$\Pr[\boldsymbol{x} \in \mathrm{C}] = \Pr[H \cdot \boldsymbol{x}^T = 0] = \left(\frac{1}{2}\right)^{n(1-R)}.$$

- Hence, the expected number of constrained codewords is

$$\mathbb{E}[N(\mathcal{C}; S^d)] = \sum_{\boldsymbol{x} \in S^d} \mathbb{E}[\mathbb{1}\{\boldsymbol{x} \in \mathcal{C}\}] = |S^d| \cdot 2^{-n(1-R)}.$$

- Since $|S^d| = 2^{n(\kappa_d + o(1))}$, there exist linear codes whose S^d-constrained subcodes are of rate at least $\kappa_d + R - 1$.

Plots and Comparisons - I



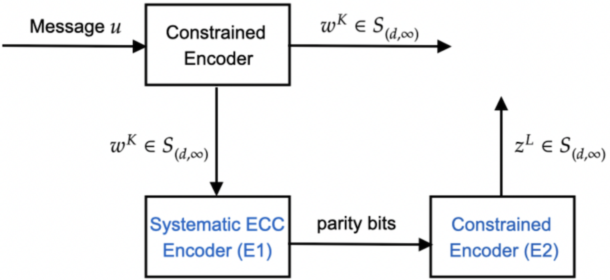
Plot comparing the achievable rates using $(1, \infty)$-RLL RM subcodes with the lower bound via the probabilistic method that is approximately $R - 0.3058$

A concatenated coding scheme

We adopt the “reverse concatenation” strategy of [Bliss (1981)] and [Mansuripur (1991)] that is commonly used to limit error propagation during decoding of constrained codes.

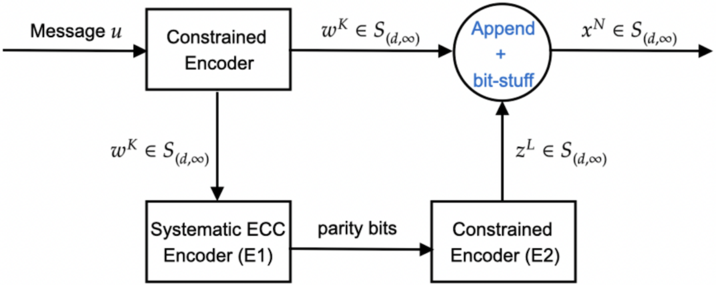
A concatenated coding scheme

Encoding (the Bliss scheme):



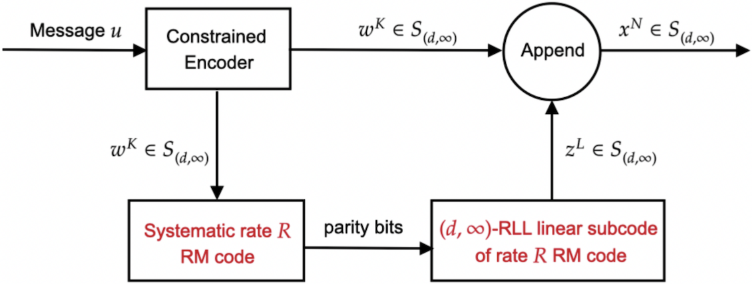
A concatenated coding scheme

Encoding (the Bliss scheme):



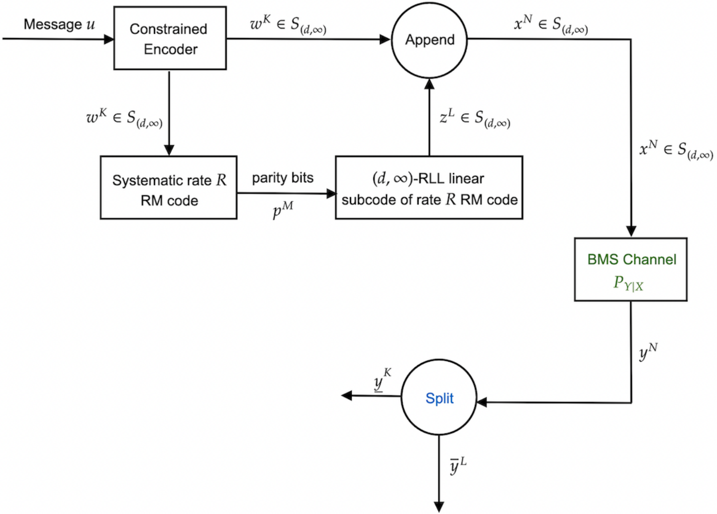
A concatenated coding scheme

Encoding:



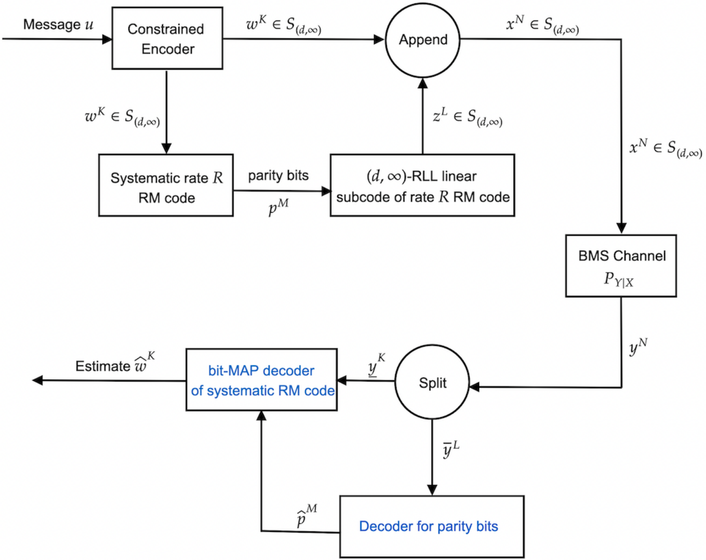
A concatenated coding scheme

Encoding+Decoding:



A concatenated coding scheme

Encoding+Decoding:



Coding theorem

Let C be the capacity of the unconstrained BMS channel.

Slight modifications to the previous concatenated coding scheme yield the following coding theorem:

Coding theorem

Let C be the capacity of the unconstrained BMS channel.

Slight modifications to the previous concatenated coding scheme yield the following coding theorem:

Theorem

For any $R \in (0, C)$, there exists a sequence of (d, ∞)-RLL constrained concatenated codes $\{\mathcal{C}_m^{conc}\}_{m \geq 1}$ that achieves a rate lower bound given by

$$\liminf_{m \to \infty} rate(\mathcal{C}_m^{conc}) \geq \frac{\kappa_d \cdot R^2 \cdot 2^{-\lceil \log_2(d+1) \rceil}}{R^2 \cdot 2^{-\lceil \log_2(d+1) \rceil} + 1 - R + \epsilon},$$

over (d, ∞)-RLL input-constrained BMS channels, where $\epsilon > 0$ can be arbitrarily small.

Plots and Comparisons - II

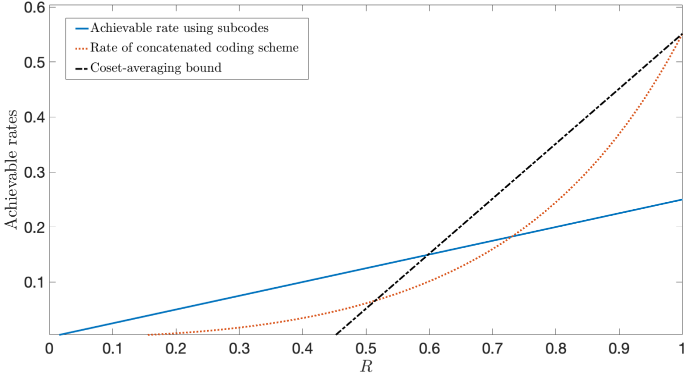


Figure: Plot comparing the achievable rates using $(2, \infty)$-RLL linear RM subcodes with the lower bound via the probabilistic method and the rate achieved by the concatenated coding scheme

Plots and Comparisons - II

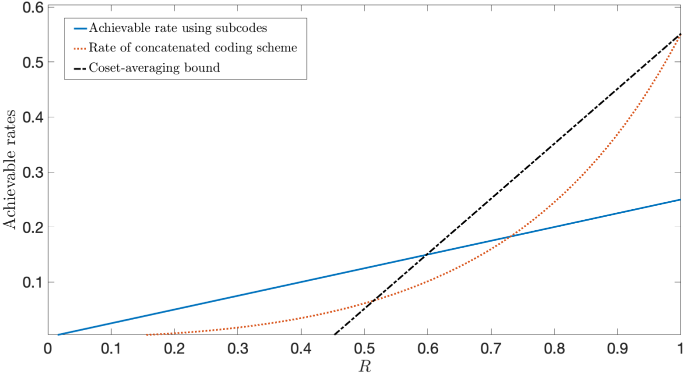
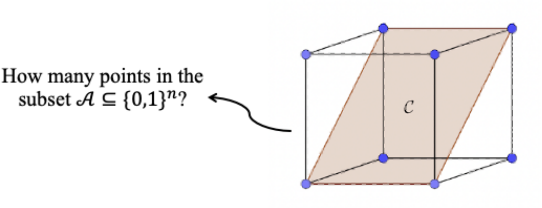


Figure: Plot comparing the achievable rates using $(2, \infty)$-RLL linear RM subcodes with the lower bound via the probabilistic method and the rate achieved by the concatenated coding scheme

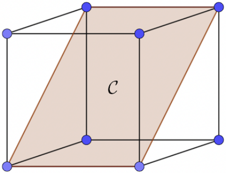
Can we extend our techniques to identifying constrained subcodes of general linear codes, for arbitrary constraints?

Counting constrained codewords in general linear codes



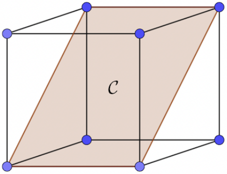
The problem

- Motivated by the previous section, we now consider the problem of computing rates of (arbitrarily-)constrained codewords in general linear codes $\mathcal{C}$.



The problem

- Motivated by the previous section, we now consider the problem of computing rates of (arbitrarily-)constrained codewords in general linear codes $\mathcal{C}$.

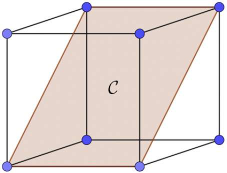


- The problem: Given a set of constrained codewords $\mathcal{A} \subseteq \mathbb{F}_2^n$, we would like to gain insight into

$$N(\mathcal{C};\mathcal{A}) = \sum_{\boldsymbol{x}\in\mathcal{C}} \mathbb{1}\{\boldsymbol{x} \in \mathcal{A}\} = \sum_{\boldsymbol{x}\in\{0,1\}^n} \mathbb{1}\{\boldsymbol{x} \in \mathcal{A}\} \cdot \mathbb{1}\{\boldsymbol{x} \in \mathcal{C}\}.$$

The problem

- Motivated by the previous section, we now consider the problem of computing rates of (arbitrarily-)constrained codewords in general linear codes $\mathcal{C}$.



- The problem: Given a set of constrained codewords $\mathcal{A} \subseteq \mathbb{F}_2^n$, we would like to gain insight into

$$N(\mathcal{C};\mathcal{A}) = \sum_{\boldsymbol{x}\in\mathcal{C}} \mathbb{1}\{\boldsymbol{x} \in \mathcal{A}\} = \sum_{\boldsymbol{x}\in\{0,1\}^n} \mathbb{1}\{\boldsymbol{x} \in \mathcal{A}\} \cdot \mathbb{1}\{\boldsymbol{x} \in \mathcal{C}\}.$$

This looks like an inner product between Boolean functions!

A brief refresher on Fourier analysis on $\mathbb{F}_2^n$

- Given any function $f : \{0,1\}^n \to \mathbb{R}$ and a vector $\boldsymbol{s} = (s_1, \ldots, s_n) \in \{0,1\}^n$, the Fourier coefficient of f at $\boldsymbol{s}$ is

$$\widehat{f}(\boldsymbol{s}) := \frac{1}{2^n} \sum_{\boldsymbol{x} \in \{0,1\}^n} f(\boldsymbol{x}) \cdot (-1)^{\boldsymbol{x} \cdot \boldsymbol{s}}.$$

- The functions $(\chi_{\boldsymbol{s}} : \boldsymbol{s} \in \{0,1\}^n)$, where $\chi_{\boldsymbol{s}}(\boldsymbol{x}) := (-1)^{\boldsymbol{x} \cdot \boldsymbol{s}}$, form an orthonormal basis for the vector space V of functions $f : \{0,1\}^n \to \mathbb{R}$. The inner product $\langle f, g \rangle$ is

$$\langle f, g \rangle := \frac{1}{2^n} \sum_{\boldsymbol{x} \in \{0,1\}^n} f(\boldsymbol{x}) g(\boldsymbol{x}).$$

Theorem (Plancherel's Theorem)

For any $f, g \in \{0,1\}^n \to \mathbb{R}$, we have that

$$\langle f, g \rangle = \sum_{\boldsymbol{s} \in \{0,1\}^n} \widehat{f}(\boldsymbol{s}) \widehat{g}(\boldsymbol{s}).$$

Workhorse

- Observe that

$$N(\mathcal{C};\mathcal{A}) = 2^n \cdot \sum_{\boldsymbol{s}\in\{0,1\}^n} \widehat{\mathbb{1}_{\mathcal{A}}}(\boldsymbol{s}) \cdot \widehat{\mathbb{1}_{\mathcal{C}}}(\boldsymbol{s}).$$

Workhorse

- Observe that

$$N(\mathcal{C};\mathcal{A}) = 2^n \cdot \sum_{\boldsymbol{s}\in\{0,1\}^n} \widehat{\mathbb{1}_{\mathcal{A}}}(\boldsymbol{s}) \cdot \widehat{\mathbb{1}_{\mathcal{C}}}(\boldsymbol{s}).$$

- For linear codes $\mathcal{C}$, it is easy to show that

$$\widehat{\mathbb{1}_{\mathcal{C}}}(\boldsymbol{s}) = \frac{|\mathcal{C}|}{2^n} \cdot \mathbb{1}_{\mathcal{C}^\perp}(\boldsymbol{s}).$$

- Hence,

$$N(\mathcal{C};\mathcal{A}) = |\mathcal{C}| \cdot \sum_{\boldsymbol{s}\in\mathcal{C}^\perp} \widehat{\mathbb{1}_{\mathcal{A}}}(\boldsymbol{s}).$$

Workhorse

- Observe that

$$N(\mathcal{C};\mathcal{A}) = 2^n \cdot \sum_{\boldsymbol{s}\in\{0,1\}^n} \widehat{\mathbb{1}_{\mathcal{A}}}(\boldsymbol{s}) \cdot \widehat{\mathbb{1}_{\mathcal{C}}}(\boldsymbol{s}).$$

- For linear codes $\mathcal{C}$, it is easy to show that

$$\widehat{\mathbb{1}_{\mathcal{C}}}(\boldsymbol{s}) = \frac{|\mathcal{C}|}{2^n} \cdot \mathbb{1}_{\mathcal{C}^\perp}(\boldsymbol{s}).$$

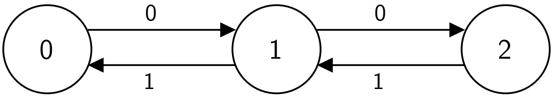
- Hence,

$$N(\mathcal{C};\mathcal{A}) = |\mathcal{C}| \cdot \sum_{\boldsymbol{s}\in\mathcal{C}^\perp} \widehat{\mathbb{1}_{\mathcal{A}}}(\boldsymbol{s}).$$

1. If $\dim(\mathcal{C}) \gg n/2$, then we can employ our insight to count over a low-dimensional space!
2. For many constraints of interest, the Fourier transform above is analytically/numerically computable!

Example 1: 2-charge constraint

- We consider a spectral null constraint, whose sequences in $\{+1, -1\}^n$ have a null at zero frequency.
- We let S_2 denote those sequences in $\{0, 1\}^n$ that can be mapped to 2-charge constrained sequences via the map $x \mapsto (-1)^x$, for $x \in \{0, 1\}$.



Sequences in S_2 can be read off the labels of paths here.

Computation of Fourier coefficients and consequences

Theorem

There exists a vector space V such that for $\boldsymbol{s} \in V$,

$$\widehat{\mathbb{1}_{S_2}}(\boldsymbol{s}) = 2^{\lfloor \frac{n}{2} \rfloor - n} \cdot (-1)^{\gamma(\boldsymbol{s})},$$

where $\gamma : \{0,1\}^n \to \{0,1\}$. Further, for $\boldsymbol{s} \notin V$, we have $\widehat{\mathbb{1}_{S_2}}(\boldsymbol{s}) = 0$.

Computation of Fourier coefficients and consequences

Theorem

There exists a vector space V such that for $\boldsymbol{s} \in V$,

$$\widehat{\mathbb{1}_{S_2}}(\boldsymbol{s}) = 2^{\lfloor \frac{n}{2} \rfloor - n} \cdot (-1)^{\gamma(\boldsymbol{s})},$$

where $\gamma : \{0,1\}^n \to \{0,1\}$. Further, for $\boldsymbol{s} \notin V$, we have $\widehat{\mathbb{1}_{S_2}}(\boldsymbol{s}) = 0$.

We use this theorem to construct a sequence $\{\mathcal{C}^{(n)}\}_{n \geq 1}$ of linear codes of rate R such that the rate of their S_2-constrained subcodes obeys

$$\liminf_{n \to \infty} \text{rate}\left(\mathcal{C}_2^{(n)}\right) > R - \frac{1}{2}.$$

We thus obtain rates better than what is guaranteed via the probabilistic method, using explicit linear codes!

Computation of Fourier coefficients and consequences

Theorem

There exists a vector space V such that for $\boldsymbol{s} \in V$,

$$\widehat{\mathbb{1}_{S_2}}(\boldsymbol{s}) = 2^{\lfloor \frac{n}{2} \rfloor - n} \cdot (-1)^{\gamma(\boldsymbol{s})},$$

where $\gamma : \{0,1\}^n \to \{0,1\}$. Further, for $\boldsymbol{s} \notin V$, we have $\widehat{\mathbb{1}_{S_2}}(\boldsymbol{s}) = 0$.

We can also use the theorem to count S_2-constrained codewords in well-known linear codes:

(m, r)	$(5, 3)$	$(6, 4)$	$(7, 5)$	$(8, 6)$
$N(\mathrm{RM}(m, r); S_2)$	2048	6.711×10^7	1.441×10^{17}	1.329×10^{36}

Some sample numerical values for high rate RM codes

Example 2: (d, ∞)-RLL constraint

- Recall:

$$(d, \infty)\text{-RLL} \equiv \text{at least } d \text{ 0s b/w successive 1s}$$

- Let S^d denote the set of constrained sequences and $\widehat{\mathbb{1}_{S^d}}^{(n)}$ denote the Fourier transform at blocklength $n \geq 1$.

Example 2: (d, ∞)-RLL constraint

- Recall:

$$(d, \infty)\text{-RLL} \equiv \text{at least } d \text{ 0s b/w successive 1s}$$

- Let S^d denote the set of constrained sequences and $\widehat{\mathbb{1}_{S^d}}^{(n)}$ denote the Fourier transform at blocklength $n \geq 1$.

Theorem

For $n \geq d+2$ and for $\mathbf{s} = (s_1, \ldots, s_n) \in \{0,1\}^n$,

$$\widehat{\mathbb{1}_{S^d}}^{(n)}(\mathrm{s}) = 2^{-1} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-1)}(s_2^n) + (-1)^{s_1} \cdot 2^{-(d+1)} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-d-1)}\left(s_{d+2}^n\right).$$

Example 2: (d,∞)-RLL constraint

- Recall:

$$(d,\infty)\text{-RLL} \equiv \text{at least } d \text{ 0s b/w successive 1s}$$

- Let S^d denote the set of constrained sequences and $\widehat{\mathbb{1}_{S^d}}^{(n)}$ denote the Fourier transform at blocklength $n \geq 1$.

Theorem

For $n \geq d+2$ and for $\mathbf{s} = (s_1, \ldots, s_n) \in \{0,1\}^n$,

$$\widehat{\mathbb{1}_{S^d}}^{(n)}(\mathrm{s}) = 2^{-1} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-1)}(s_2^n) + (-1)^{s_1} \cdot 2^{-(d+1)} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-d-1)}\left(s_{d+2}^n\right).$$

The recursive procedure arising from the above theorem is faster for computing Fourier transforms than the Fast Walsh-Hadamard Transform!

Example 2: (d,∞)-RLL constraint

- Recall:

$$(d,\infty)\text{-RLL} \equiv \text{at least } d \text{ 0s b/w successive 1s}$$

- Let S^d denote the set of constrained sequences and $\widehat{\mathbb{1}_{S^d}}^{(n)}$ denote the Fourier transform at blocklength $n \geq 1$.

Theorem

For $n \geq d+2$ and for $\mathbf{s} = (s_1, \ldots, s_n) \in \{0,1\}^n$,

$$\widehat{\mathbb{1}_{S^d}}^{(n)}(\mathrm{s}) = 2^{-1} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-1)}(s_2^n) + (-1)^{s_1} \cdot 2^{-(d+1)} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-d-1)}\left(s_{d+2}^n\right).$$

Similar recurrence relations can also be proved for the flash memory ("no-101") constraint and a version of the even constraint, which requires that the length of every run of 0s be even.

Example 2: (d, ∞)-RLL constraint

- Recall:

$$(d, \infty)\text{-RLL} \equiv \text{at least } d \text{ 0s b/w successive 1s}$$

- Let S^d denote the set of constrained sequences and $\widehat{\mathbb{1}_{S^d}}^{(n)}$ denote the Fourier transform at blocklength $n \geq 1$.

Theorem

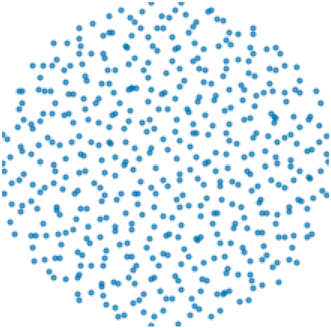
For $n \geq d+2$ and for $\mathbf{s} = (s_1, \ldots, s_n) \in \{0,1\}^n$,

$$\widehat{\mathbb{1}_{S^d}}^{(n)}(\mathrm{s}) = 2^{-1} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-1)}(s_2^n) + (-1)^{s_1} \cdot 2^{-(d+1)} \cdot \widehat{\mathbb{1}_{S^d}}^{(n-d-1)}\left(s_{d+2}^n\right).$$

However, counting in the space of the dual code $\mathcal{C}^\perp$ requires us to store all Fourier coefficients at blocklength n—a task that can quickly become very expensive.

Can we shoot for something less accurate, but more efficient?

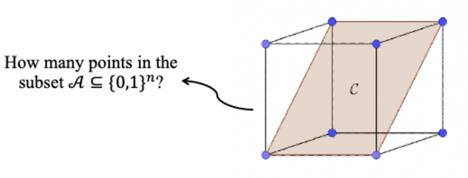
Estimates of the Sizes of Constrained Subcodes of RM Codes via Sampling[2]



[2] With help from Shreyas Jain, IISER Mohali, India

The problem

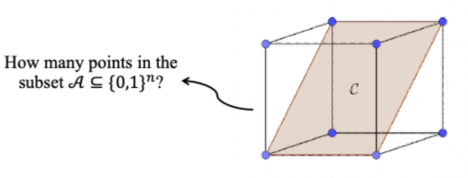
Consider again our recurring (and yet unanswered) question:



- Suppose that $\mathcal{C}$ is $\mathrm{RM}(m, r)$. Can we obtain approximate estimates of $N_{\mathcal{A}} = N(\mathcal{C}; \mathcal{A})$, for arbitrary constraints?

The problem

Consider again our recurring (and yet unanswered) question:



- Suppose that $\mathcal{C}$ is RM(m, r). Can we obtain approximate estimates of $N_{\mathcal{A}} = N(\mathcal{C}; \mathcal{A})$, for arbitrary constraints?
- More precisely, can we efficiently (poly. time?) obtain an estimate $\widehat{N}_{\mathcal{A}}$, such that with high probability,

$$\widehat{N}_{\mathcal{A}} \in [(1-\epsilon)N_{\mathcal{A}}, (1+\epsilon)N_{\mathcal{A}}],$$

  for some arbitrarily small $\epsilon > 0$?

Constraints of interest and a first pass

- We are primarily interested in the (d, ∞)-RLL constraint and in constant-weight constraints.

- Suppose that we try to build an estimator via a simple "rejection sampling" approach:

1. Draw n uniformly random codewords from $\mathrm{RM}(m, r)$.

2. Set $\widehat{N}_{\mathcal{A}} = |\mathrm{RM}(m, r)| \times \left(\frac{\#\{\text{samples in } \mathcal{A}\}}{n}\right)$.

Constraints of interest and a first pass

- We are primarily interested in the (d, ∞)-RLL constraint and in constant-weight constraints.
- Suppose that we try to build an estimator via a simple "rejection sampling" approach:

1. Draw n uniformly random codewords from $\mathrm{RM}(m, r)$.
2. Set $\widehat{N}_{\mathcal{A}} = |\mathrm{RM}(m,r)| \times \left(\frac{\#\{\text{samples in } \mathcal{A}\}}{n}\right)$.

Clearly, for n large, we have that $\widehat{N}_{\mathcal{A}} \in [(1-\epsilon)N_{\mathcal{A}}, (1+\epsilon)N_{\mathcal{A}}]$.

Constraints of interest and a first pass

- We are primarily interested in the (d, ∞)-RLL constraint and in constant-weight constraints.
- Suppose that we try to build an estimator via a simple "rejection sampling" approach:

1. Draw n uniformly random codewords from $\mathrm{RM}(m, r)$.
2. Set $\widehat{N}_{\mathcal{A}} = |\mathrm{RM}(m, r)| \times \left(\frac{\#\{\text{samples in } \mathcal{A}\}}{n}\right)$.

Clearly, for n large, we have that $\widehat{N}_{\mathcal{A}} \in [(1-\epsilon)N_{\mathcal{A}}, (1+\epsilon)N_{\mathcal{A}}]$.

- Fact: For most weights [cf. Rao and Sprumont (2022)] and for $d = 1$ [Rameshwar and Kashyap (2023)], $N_{\mathcal{A}}$ is exponentially smaller than $\mathrm{RM}(m, r)$.

Hence, exponentially many (in n) samples needed!

Constraints of interest and a first pass

- We are primarily interested in the (d, ∞)-RLL constraint and in constant-weight constraints.

- Suppose that we try to build an estimator via a simple "rejection sampling" approach:

> 1. Draw n uniformly random codewords from $\mathrm{RM}(m, r)$.
> 2. Set $\widehat{N}_{\mathcal{A}} = |\mathrm{RM}(m, r)| \times \left(\frac{\#\{\text{samples in } \mathcal{A}\}}{n}\right)$.

Clearly, for n large, we have that $\widehat{N}_{\mathcal{A}} \in [(1-\epsilon)N_{\mathcal{A}}, (1+\epsilon)N_{\mathcal{A}}]$.

- Fact: For most weights [cf. Rao and Sprumont (2022)] and for $d = 1$ [Rameshwar and Kashyap (2023)], $N_{\mathcal{A}}$ is exponentially smaller than $\mathrm{RM}(m, r)$.

Can we design a good estimator that uses only poly. many samples?

Key insight

- Observe that $N_{\mathcal{A}} = Z$, the partition function of the distribution p, where
$$p(\boldsymbol{x}) = \frac{1}{Z} \cdot \mathbb{1}_{\mathcal{C} \cap \mathcal{A}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \{0,1\}^n.$$

Key insight

- Observe that $N_{\mathcal{A}} = Z$, the partition function of the distribution p, where
$$p(\boldsymbol{x}) = \frac{1}{Z} \cdot \mathbb{1}_{\mathcal{C} \cap \mathcal{A}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \{0,1\}^n.$$

- While it is hard to compute Z and indeed even sample from p, consider now the Gibbs distribution p_β, for $\beta > 0$:
$$p_\beta(\boldsymbol{x}) = \frac{1}{Z_\beta} \cdot e^{-\beta \cdot E(\boldsymbol{x})}, \quad \boldsymbol{x} \in \mathcal{C},$$
where $E(\boldsymbol{x}) \geq 0$ with equality iff $\boldsymbol{x} \in \mathcal{A}$, and
$$Z_\beta = \sum_{\mathrm{c} \in \mathcal{C}} e^{-\beta \cdot E(\mathrm{x})}.$$

Key insight

- Observe that $N_{\mathcal{A}} = Z$, the partition function of the distribution p, where

$$p(\boldsymbol{x}) = \frac{1}{Z} \cdot \mathbb{1}_{\mathcal{C} \cap \mathcal{A}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \{0,1\}^n.$$

- While it is hard to compute Z and indeed even sample from p, consider now the Gibbs distribution p_β, for $\beta > 0$:

$$p_\beta(\boldsymbol{x}) = \frac{1}{Z_\beta} \cdot e^{-\beta \cdot E(\boldsymbol{x})}, \quad \boldsymbol{x} \in \mathcal{C},$$

where $E(\boldsymbol{x}) \geq 0$ with equality iff $\boldsymbol{x} \in \mathcal{A}$, and

$$Z_\beta = \sum_{\mathrm{c} \in \mathcal{C}} e^{-\beta \cdot E(\mathrm{x})}.$$

- Examples:
  - (d, ∞)-RLL constraint: $E(\boldsymbol{x}) = \#\{\text{violations of the constraint in } \boldsymbol{x}\}$
  - Constant-weight ω constraint: $E(\boldsymbol{x}) = |w(\boldsymbol{x}) - \omega|$

Key insight

- Observe that $N_{\mathcal{A}} = Z$, the partition function of the distribution p, where
$$p(\boldsymbol{x}) = \frac{1}{Z} \cdot \mathbb{1}_{\mathcal{C}\cap\mathcal{A}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \{0,1\}^n.$$

- While it is hard to compute Z and indeed even sample from p, consider now the Gibbs distribution p_β, for $\beta > 0$:
$$p_\beta(\boldsymbol{x}) = \frac{1}{Z_\beta} \cdot e^{-\beta \cdot E(\boldsymbol{x})}, \quad \boldsymbol{x} \in \mathcal{C},$$
where $E(\boldsymbol{x}) \geq 0$ with equality iff $\boldsymbol{x} \in \mathcal{A}$, and
$$Z_\beta = \sum_{\mathrm{c}\in\mathcal{C}} e^{-\beta \cdot E(\mathrm{x})}.$$

- Note that
$$\lim_{\beta\to\infty} p_\beta(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{and}$$
$$\lim_{\beta\to\infty} Z_\beta = Z.$$

Key insight

- Observe that $N_{\mathcal{A}} = Z$, the partition function of the distribution p, where

$$p(\boldsymbol{x}) = \frac{1}{Z} \cdot \mathbb{1}_{\mathcal{C} \cap \mathcal{A}}(\boldsymbol{x}), \quad \boldsymbol{x} \in \{0,1\}^n.$$

- While it is hard to compute Z and indeed even sample from p, consider now the Gibbs distribution p_β, for $\beta > 0$:

$$p_\beta(\boldsymbol{x}) = \frac{1}{Z_\beta} \cdot e^{-\beta \cdot E(\boldsymbol{x})}, \quad \boldsymbol{x} \in \mathcal{C},$$

where $E(\boldsymbol{x}) \geq 0$ with equality iff $\boldsymbol{x} \in \mathcal{A}$, and

$$Z_\beta = \sum_{\mathrm{c} \in \mathcal{C}} e^{-\beta \cdot E(\mathrm{x})}.$$

- Note that

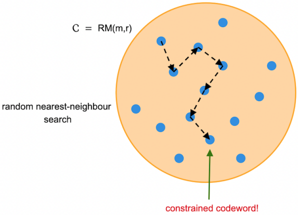
$$\lim_{\beta \to \infty} p_\beta(\boldsymbol{x}) = p(\boldsymbol{x}), \quad \text{and}$$

$$\lim_{\beta \to \infty} Z_\beta = Z.$$

We use Z_β, for large β, as a "good" approximation to $Z = N_{\mathcal{A}}$.

An MCMC scheme to sample from p_β

Before we compute Z_β, we propose an efficient sampler from p_β.



procedure MCMC-SAMPLER

  Start at an arbitrary codeword $\boldsymbol{c}_0 \in \mathcal{C}$.

  Fix a large epoch length τ.

  for $i = 1 : \tau$ **do**

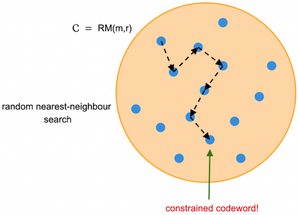
    Sample a uniformly random min.-wt. codeword $\overline{\boldsymbol{c}}$.

    Set $\boldsymbol{c}_i \leftarrow \boldsymbol{c}_{i-1} + \overline{\boldsymbol{c}}$ w.p. $\min\left(1, \exp(-\beta(E(\mathrm{c}) - E(\mathrm{c}^{(i-1)})))\right)$.

  Output $\boldsymbol{c}_\tau$.

An MCMC scheme to sample from p_β

Before we compute Z_β, we propose an efficient sampler from p_β.



1: **procedure** MCMC-SAMPLER
2: Start at an arbitrary codeword $\boldsymbol{c}_0 \in \mathcal{C}$.
3: Fix a large epoch length τ.
4: **for** $i = 1 : \tau$ **do**
5: Sample a uniformly random min.-wt. codeword $\overline{\boldsymbol{c}}$.
6: Set $\boldsymbol{c}_i \leftarrow \boldsymbol{c}_{i-1} + \overline{\boldsymbol{c}}$ w.p. $\min\left(1, \exp(-\beta(E(\mathrm{c}) - E(\mathrm{c}^{(i-1)})))\right)$.
7: Output $\boldsymbol{c}_\tau$.

Facts:

1. We can efficiently draw uniformly random min.-wt. codewords from RM(m, r) using the correspondence with $(m-r)$-dimensional affine subspaces.
2. The Markov chain above is irreducible (min.-wt. codewords span the code!) and has p_β as stationary distribution.

From sampling to counting

Fix a large $\beta^\star > 0$. The following technique to compute $Z_{\beta^\star}$ is well-known [Valleau and Card (1972)].

- Let $\beta^\star = \ell/n$ and fix a "cooling schedule" of β parameters:

$$0 = \beta_0 < \beta_1 < \ldots < \beta_\ell = \beta^\star,$$

where $\beta_i = \beta_{i-1} + 1/n$, for $1 \le i \le \ell$.

From sampling to counting

Fix a large $\beta^\star > 0$. The following technique to compute $Z_{\beta^\star}$ is well-known [Valleau and Card (1972)].

- Let $\beta^\star = \ell/n$ and fix a "cooling schedule" of β parameters:

$$0 = \beta_0 < \beta_1 < \ldots < \beta_\ell = \beta^\star,$$

  where $\beta_i = \beta_{i-1} + 1/n$, for $1 \leq i \leq \ell$.

- Write

$$Z_{\beta^\star} = Z_{\beta_0} \times \prod_{i=1}^{\ell} \frac{Z_{\beta_i}}{Z_{\beta_{i-1}}}. \tag{1}$$

From sampling to counting

Fix a large $\beta^\star > 0$. The following technique to compute $Z_{\beta^\star}$ is well-known [Valleau and Card (1972)].

- Let $\beta^\star = \ell/n$ and fix a "cooling schedule" of β parameters:

$$0 = \beta_0 < \beta_1 < \ldots < \beta_\ell = \beta^\star,$$

  where $\beta_i = \beta_{i-1} + 1/n$, for $1 \leq i \leq \ell$.

- Write

$$Z_{\beta^\star} = Z_{\beta_0} \times \prod_{i=1}^{\ell} \frac{Z_{\beta_i}}{Z_{\beta_{i-1}}}. \tag{1}$$

- Observe that

$$\begin{aligned} \frac{Z_{\beta_i}}{Z_{\beta_{i-1}}} &= \frac{1}{Z_{\beta_{i-1}}} \sum_{\mathrm{c} \in \mathcal{C}} \exp(-\beta_i E(\mathrm{c})) \\ &= \mathbb{E}_{p_{\beta_{i-1}}}[\exp((\beta_{i-1} - \beta_i) E(\mathrm{c}))]. \end{aligned}$$

From sampling to counting

Fix a large $\beta^\star > 0$. The following technique to compute $Z_{\beta^\star}$ is well-known [Valleau and Card (1972)].

- Let $\beta^\star = \ell/n$ and fix a "cooling schedule" of β parameters:
$$0 = \beta_0 < \beta_1 < \ldots < \beta_\ell = \beta^\star,$$
where $\beta_i = \beta_{i-1} + 1/n$, for $1 \leq i \leq \ell$.
- Write
$$Z_{\beta^\star} = Z_{\beta_0} \times \prod_{i=1}^{\ell} \frac{Z_{\beta_i}}{Z_{\beta_{i-1}}}. \tag{1}$$
- Observe that
$$\begin{aligned} \frac{Z_{\beta_i}}{Z_{\beta_{i-1}}} &= \frac{1}{Z_{\beta_{i-1}}} \sum_{\mathrm{c} \in \mathcal{C}} \exp(-\beta_i E(\mathrm{c})) \\ &= \mathbb{E}_{p_{\beta_{i-1}}}[\exp((\beta_{i-1} - \beta_i)E(\mathrm{c}))]. \end{aligned}$$
- Use a sample-average estimator for the expectation above and compute a final estimate $\widehat{Z}_{\beta^\star}$, using (1).

How many samples are enough?

We now provide some theoretical guarantees [cf. Lecture notes on "Partition Functions" by Alistair Sinclair (2020)].

- Firstly, it suffices for $\beta^\star = O(n^2)$ to have

$$(1-\delta_n)Z \le Z_{\beta^\star} \le (1+\delta_n)Z,$$

where $\delta_n \to 0$ exponentially quickly.

How many samples are enough?

We now provide some theoretical guarantees [cf. Lecture notes on "Partition Functions" by Alistair Sinclair (2020)].

- Firstly, it suffices for $\beta^\star = O(n^2)$ to have

$$(1-\delta_n)Z \leq Z_{\beta^\star} \leq (1+\delta_n)Z,$$

where $\delta_n \to 0$ exponentially quickly.

- Secondly, for the number of samples used for each sample average estimator being $\Theta(n^3)$, we have [Dyer and Frieze (1991)]

$$\Pr[(1-\epsilon)Z_{\beta^\star} \leq \widehat{Z}_{\beta^\star} \leq (1+\epsilon)Z_{\beta^\star}] \geq \frac{3}{4}.$$

How many samples are enough?

We now provide some theoretical guarantees [cf. Lecture notes on "Partition Functions" by Alistair Sinclair (2020)].

- Firstly, it suffices for $\beta^\star = O(n^2)$ to have

$$(1 - \delta_n)Z \leq Z_{\beta^\star} \leq (1 + \delta_n)Z,$$

  where $\delta_n \to 0$ exponentially quickly.

- Secondly, for the number of samples used for each sample average estimator being $\Theta(n^3)$, we have [Dyer and Frieze (1991)]

$$\Pr[(1 - \epsilon)Z_{\beta^\star} \leq \widehat{Z}_{\beta^\star} \leq (1 + \epsilon)Z_{\beta^\star}] \geq \frac{3}{4}.$$

- Thus, using only $\Theta(n^6)$ samples overall, we obtain that

$$\Pr[(1 - \epsilon)(1 + \delta_n)N_{\mathcal{A}} \leq \widehat{Z}_{\beta^\star} \leq (1 + \epsilon)(1 + \delta_n)N_{\mathcal{A}}] \geq \frac{3}{4}.$$

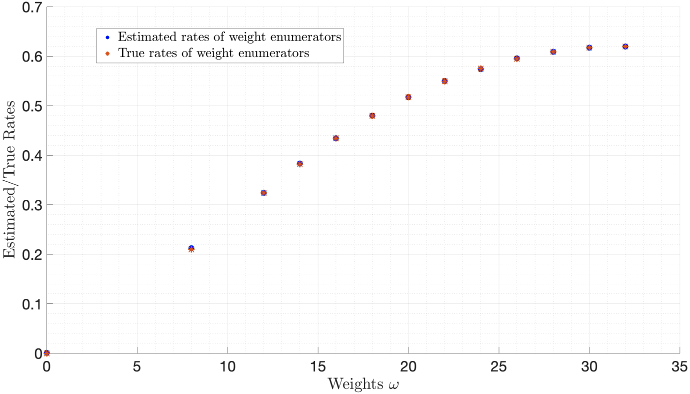
- The constant $\frac{3}{4}$ above can be improved to $1 - \gamma$, for $\gamma > 0$ arbitrarily small, using a "median-of-batches" trick.

Numerical trials - I

m	r	$\frac{\log_2 \widehat{Z}}{2^m}$	$\frac{\log_2 Z}{2^m}$
6	2	0.1557	0.1508
7	2	0.0883	0.0880
8	1	0.0095	0.0078
7	5	0.6340	–
8	3	0.1391	–
8	4	0.3343	–
8	5	0.5520	–

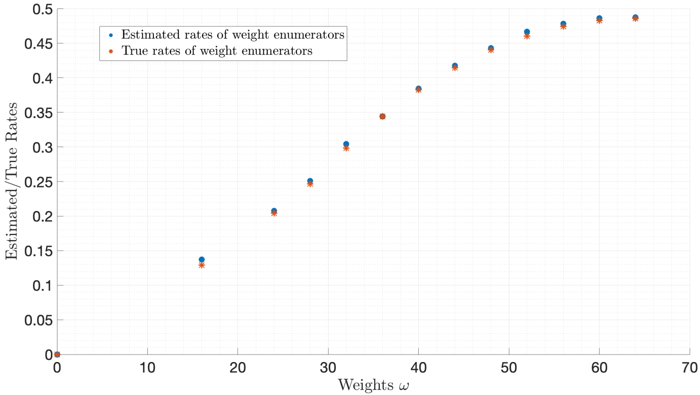
Table: Table of estimated rates rates $\frac{\log_2 \widehat{Z}}{2^m}$ of $(1, \infty)$-RLL constrained codewords in $\mathrm{RM}(m, r)$, for different values $m \geq 1$, $r \leq m$, compared with the true rates $\frac{\log_2 Z}{2^m}$ whenever a brute-force enumeration is tractable.

Numerical trials - II



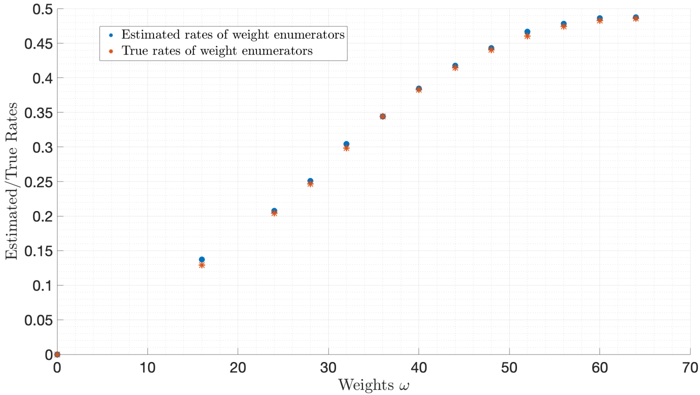
Plot comparing the rate estimates of the weight enumerators with the rates of the true weight enumerators of RM$(6, 3)$.

Numerical trials - III



Plot comparing the rate estimates of the weight enumerators with the rates of the true weight enumerators of RM(7, 3), obtained in [Sugino, Ienaga, Tokura, and Kasami (1971)].

Numerical trials - III



Plot comparing the rate estimates of the weight enumerators with the rates of the true weight enumerators of RM(7, 3), obtained in [Sugino, Ienaga, Tokura, and Kasami (1971)].

We also obtain estimates of the hitherto unknown weight distribution of RM(9, 4), using our techniques.

Open questions for further research

Open questions

- Is it possible to prove (analytically) that the asymptotic rate of (d, k)-RLL constrained subcodes of rate R RM codes is $\kappa \cdot R$? This would then help resolve [Wolf's Conjecture (1988)] for the (d, k)-RLL input-constrained BSC(p):

$$C_{d,k}(p) \geq \kappa_{d,k}(1 - h_b(p)).$$

- Can one design explicit codes over other channels with memory, such as Gilbert-Elliott Channels, using RM/polar codes?

Thank You!