Kitchen Math

Some Fly by Night TCS and Stuffs

Hang Su

Last updated: October 27, 2026

Contents

1 Probability and Computing **4**
1.1 Events and Probability . . . 4
1.2 Discrete Random Variable and Expectation . . . 6
1.3 Moments and Deviations . . . 8
1.3.1 Weak Law of Large Numbers (WLLN) . . . 10
1.4 Chernoff and Hoeffding Bounds . . . 11
1.4.1 Hoeffding's Bound . . . 17
1.4.2 Quicksort Runtime Analysis . . . 19
1.4.3 Network Routing Problems . . . 20
1.5 Balls, Bins, and Random Graphs . . . 24
1.5.1 A few things I should have noted about e . . . 24
1.5.2 Birthday Paradox and Balls-and-Bins model . . . 25
1.5.3 Poisson Distribution . . . 28
1.5.4 Poisson Approximation . . . 29
1.5.5 Recap on Coupon Collector's Problem . . . 32
1.5.6 Hashing and Random Graphs . . . 33
1.5.7 Hamiltonian Cycles in Random Graphs . . . 38
1.6 The Probabilistic Method . . . 41
1.6.1 A few things I should have noted about factorial . . . 41
1.6.2 Basic Counting Argument . . . 43
1.6.3 The Expectation Argument and Derandomization by Conditional Expectation . . . 44
1.6.4 Sample and Modify . . . 47
1.6.5 The Second Moment Method and the Conditional Expectation Inequality . . . 48
1.6.6 Lovász Local Lemma . . . 53
1.6.7 Algorithmic Lovász Local Lemma . . . 58
1.7 Markov Chains and Random Walks . . . 64
1.7.1 Stirling's Formula . . . 64
1.7.2 Markov Chains and Classification of States . . . 66
1.7.3 Strong Law of Large Numbers (SLLN) . . . 75
1.7.4 Stationary Distributions . . . 83
1.7.5 Random Walks on Undirected Graphs . . . 95
1.7.6 Parrondo's Paradox . . . 101
1.8 Continuous Distributions and the Poisson Process . . . 102

1.9 The Normal Distribution . . . 103
1.10 Entropy, Randomness, and Information . . . 104
1.10.1 The Entropy Function . . . 104
1.10.2 Entropy and Binomial Coefficients . . . 107
1.10.3 Entropy: A Measure of Randomness . . . 108
1.10.4 Entropy: A Limit of Data Compression . . . 117
1.10.5 Shannon's Theorem . . . 121
1.11 The Monte Carlo Method . . . 130
1.11.1 The Monte Carlo Method . . . 130
1.11.2 DNF Counting . . . 130
1.11.3 From Approximate Sampling to Approximate Counting . . . 130
1.11.4 The Markov Chain Monte Carlo Method (MCMC) . . . 130
1.12 Coupling of Markov Chains . . . 131
1.12.1 Variation Distance and Mixing Time . . . 131
1.12.2 Coupling . . . 132
1.12.3 Variation Distance is Nonincreasing . . . 139
1.12.4 Geometric Convergence . . . 143
1.12.5 Application in Approximately Sampling Proper Colorings . . . 145
1.12.6 Path Coupling . . . 146
1.13 Martingales . . . 154
1.13.1 Martingales . . . 154
1.13.2 Stopping Times . . . 157
1.13.3 Wald's Equation . . . 160
1.13.4 Azuma-Hoeffding's Inequality . . . 165
1.14 Sample Complexity, VC Dimension, and Rademacher Complexity . . . 174
1.15 Pairwise Independence and Universal Hash Function . . . 175
1.15.1 Pairwise Independence . . . 175
1.15.2 Chebyshev's Inequality for Pairwise Independent Random Variables . . . 175
1.15.3 Universal Families of Hash Functions . . . 175
1.15.4 Application in Finding Heavy Hitters in Data Streams . . . 175
1.16 Power Laws and Related Distributions . . . 176
1.17 Balanced Allocations and Cuckoo Hashing . . . 177

2 Pseudorandomness **178**

3 Probabilitic Method **179**

4 Various Papers **180**
4.1 Leftover Hash Lemma and HILL Entropy . . . 180
4.2 Discrete Gaussian Leftover Hash Lemma . . . 181
4.3 Schwartz-Zippel by Dana Moshkovitz . . . 182
4.4 Ligero Interleaved Linear Code Testing . . . 183
4.5 Interleaved Linear Code Testing with Logarithmic Randomness . . . 186
4.6 Basefold Random Linear Foldable Code and the Distance Proof . . . 187
4.7 Bulletproof . . . 188
4.8 Sublinear Lattice ZKP for Arithmetic Circuits . . . 189
4.9 A Non-PCP Approach to Succinct Quantum-Safe ZK . . . 193
4.10 Short Invertible Polynomial Ring Elements . . . 203
4.11 Lattice-Based Exact Argument of Knowledge . . . 207
4.12 Practical Product Proof from Lattice Commitments . . . 218
4.13 Practical Exact Proofs from Lattices . . . 231
4.14 Lattice Proofs for Exact Euclidean Norm Bounds and Quadratic Relations . . . 239

4.15 Practical Sublinear Proofs for R1CS from Lattices 255
4.16 LaBRADOR 256
4.17 Greyhound 257
4.18 Lattice-Based Succinct Arguments for NP with Polylog Verification 258
4.19 Greco 259

A Degree-Four Bound for Binomial Compression **267**

1 Probability and Computing

1.1 Events and Probability

Exercise 1.1.1 (Exercise 1.6 [MU17])**.** Consider the following balls-and-bin game. We start with one black ball and one white ball in a bin. We repeatedly do the following: choose one ball from the bin uniformly at random, then put the ball back in the bin with another ball of the same color. We repeat until there are n balls in the bin. Show that the number of white balls is equally likely to be any number between 1 and $n-1$.

Proof. We rethink the behavior as: choose a color from the current balls in bin, and add one more ball with same color to the bin. Then we introduce a random variable X_n representing the number of white balls in the bin, where there are n balls in bin in total. The probability of k white balls among n balls can be represented by

$$\begin{aligned}\Pr[X_n = k] &= \Pr[X_n = k \mid X_{n-1} = k-1] \cdot \Pr[X_{n-1} = k-1] \\ &\quad + \Pr[X_n = k \mid X_{n-1} = k] \cdot \Pr[X_{n-1} = k] \\ &= \frac{k-1}{n-1} \cdot \Pr[X_{n-1} = k-1] + \left(1 - \frac{k}{n-1}\right) \cdot \Pr[X_{n-1} = k].\end{aligned}$$

By induction we can have that, the probability of X_n being k is $1/(n-1)$. □

Exercise 1.1.2 (Exercise 1.14 [MU17])**.** Suppose I am playing in a tournament against a player never played before. Consider 3 probabilities for my prior model: we are equally talented, and each of us is equally likely to win each game; I am slightly better, and I win each game independently with probability 0.6; or he is slightly better, and he wins each game independently with probability 0.6. Before we play, I think that each of these 3 probabilities is equally likely.

In the match we play until one player wins three games. I win the second game, but he wins the first, third, and fourth. After this match, with what probability should I believe that my opponent is slightly better than I am?

Proof. This is the Bayesian prior to posterior update.

Let E be the event that "I won the second and he won the first/third/fourth".

Let C_0 be the event that he is better, C_1 be the event that I am better, C_2 be the event that we are equally talented. Then C_0, C_1, C_2 are disjoint events, with prior model on C_i being $\Pr[C_i] = 1/3$. The posterior model on C_0 is derived by the Bayes' law

$$\begin{aligned}\Pr[C_0 \mid E] &= \frac{\Pr[E \mid C_0] \cdot \Pr[C_0]}{\sum_{i\in[0,2]} \Pr[E \mid C_i] \cdot \Pr[C_i]} \\ &= \frac{\Pr[E \mid C_0]}{\sum_{i\in[0,2]} \Pr[E \mid C_i]},\end{aligned}$$

and each $\Pr[E \mid C_i] = p_i \cdot (1-p_i)^3$, where p_i is the probability of me winning a game preconditioned in C_i.

Therefore, the exact probability of he is better preconditioned that E happened is

$$\Pr[C_0 \mid E] = \frac{0.6^3 \cdot 0.4}{0.6^3 \cdot 0.4 + 0.5^4 + 0.4^3 \cdot 0.6} \approx 0.46.$$

□

Exercise 1.1.3 (Exercise 1.18 [MU17])**.** We have function $F : [0, n-1] \to [0, m-1]$, and we know $F((x+y) \bmod n) = (F(x) + F(y)) \bmod m$ for $x, y \in [0, n-1]$. The only way we have for evaluating F is to use a lookup table that stores the values of F. An adversary changed the value of $1/5$ of the table entries.

Describe a simple randomized algorithm that, given an input z, outputs a value that equals $F(z)$ with probability at least $1/2$. The algorithm should be working for every $z \in [0, n-1]$, regardless of what values the adversary changed. The algorithm should use as few lookups and as little computation as possible.

Suppose I allow you to repeat the initial algorithm three times, what is the probability of the enhanced version of algorithm returning the correct answer?

Proof. This looks very much like BLR linearity testing [BLR93]. We construct a random function as follows:

- Sample $\alpha \xleftarrow{R} [0, n-1]$, and query $F(\alpha)$ and $F((z+\alpha) \mod n)$.
- Output $(F((z+\alpha) \mod n) - F(\alpha)) \mod m$.

Since the error probability at a single query is $1/5$, the probability of this algorithm giving right answer is lower bounded by $3/5 \geq 1/2$.

Now that the error probability of our algorithm is upper bounded by $2/5$, if we repeat the algorithm 3 times, the algorithm returning none of them being correct has probability upper bounded by $(2/5)^3$.

- The probability of existing a correct answer is lower bounded by $1 - (2/5)^3$.
- If the majority are correct, then $(3/5)^3 + 3 \times (2/5) \times (3/5)^2 \geq 81/125$.

Try majority vote from 3 repetition as an enhanced version, then the correctness is lower bounded by $81/125$. □

1.2 Discrete Random Variable and Expectation

Exercise 1.2.1 (Exercise 2.2 [MU17])**.** A monkey types on a 26-letter keyboard that has lowercase letters only. Each letter is chosen independently and uniformly at random from the alphabet. If the monkey types 1,000,000 letters, what is the expected number of times the sequence "proof" appears?

Proof. This is exactly a brain teaser in linearity of expectation.

Say X_i are 0-1 random variables indicating if the sequence "proof" appears at i^{th} alphabet. Then $X = \sum X_i$ is the random variable for the number of appearances of the sequence "proof".

By linearity of expectation,

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i\in[1,999996]} X_i\right] = \sum_{i\in[1,999996]} \mathbb{E}[X_i] = 999996 \times \left(\frac{1}{26}\right)^5.$$

□

Exercise 1.2.2 (Exercise 2.12 [MU17])**.** We draw cards uniformly at random with replacement from a deck of n cards. What is the expected number of cards we must draw until we have seen all n cards in the deck? If we draw $2n$ cards, what is the expected number of cards in the deck that are not chosen at all? Chosen exactly once?

Proof. The problem is a combination of "coupon collection" problem and "indicator" problem.

First one is the "coupon collection" problem, then let X_i be the number of draws to draw i^{th} different card, which is a geometrically distributed random variable with parameter $\frac{n-i+1}{n}$, and $X = \sum X_i$ be the total number of draws to draw all n different cards. $\mathbb{E}[X] = n \ln n + \Theta(n)$.

Later, we introduce 0-1 random variables $\{Y_i\}_{i\in[1,n]}$ indicating if the i^{th} cards is not drawn once in the $2n$ draws, then $\Pr[Y_i = 1] = \left(\frac{n-1}{n}\right)^{2n}$. Let $Y = \sum Y_i$ for the number of cards not drawn at all in $2n$ draws, then

$$\mathbb{E}[Y] = \mathbb{E}\left[\sum_{i\in[1,n]} Y_i\right] = \sum_{i\in[1,n]} \mathbb{E}[Y_i] = n \cdot \left(\frac{n-1}{n}\right)^{2n}.$$

We also introduce 0-1 random variables $\{Z_i\}_{i\in[1,n]}$ indicating if the i^{th} card is drawn only once in the $2n$ draws, then $\Pr[Z_i = 1] = 2n \cdot \frac{1}{n} \cdot \left(\frac{n-1}{n}\right)^{2n-1}$. Let $Z = \sum Z_i$ for the number of cards drawn only once in $2n$ draws, then

$$\mathbb{E}[Z] = \mathbb{E}\left[\sum_{i\in[1,n]} Z_i\right] = \sum_{i\in[1,n]} \mathbb{E}[Z_i] = 2n \cdot \left(\frac{n-1}{n}\right)^{2n-1}.$$

□

Exercise 1.2.3 (Exercise 2.16 [MU17])**.** Suppose we flip a coin n times to obtain a sequence of flips $\{X_i\}_{i\in[1,n]}$. A streak of flips is a consecutive subsequence of flips that are all the same.

- Let n be a power of 2. Show that the expected number of streaks of length $\log_2 n + 1$ is $1 - o(1)$.
- Show that, for sufficiently large n, the probability that there is no streak of length at least $\lfloor \log_2 n - 2 \log_2 \log_2 n \rfloor$ is less than $1/n$.

Proof. We first consider the expected number of streaks of length at least $\log_2 n + 1$. This is solved by the linearity of expectation, by introducing $\{X_i\}_{i\in[1,k]}$ 0-1 random variables, where $k = n - \log_2 n$, indicating the i^{th} flip starts a streak of length at least $\log_2 n + 1$. The probability for $X_i = 1$ can be upper bounded as follows

$$\Pr[X_i = 1] = 2 \cdot \frac{1}{2^{\log_2 n+1}} = \frac{1}{n},$$

as streak on either face works. Therefore, number of streaks is $X = \sum_{i\in[1,k]} X_i$, and the expectation is

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}\left[\sum_{i\in[1,k]} X_i\right] = \sum_{i\in[1,k]} \mathbb{E}[X_i] \\ &= \left(n - \log_2 n\right)\mathbb{E}[X_i] \\ &= \frac{n-\log_2 n}{n} = 1 - o(1).\end{aligned}$$

We now show that the probability of no streaks of length at least $\ell = \lfloor \log_2 n - 2\log_2\log_2 n \rfloor$ is less than $1/n$. We consider n/ℓ segments of ℓ length flips, and thus introduce n/ℓ independent 0-1 random variables indicating if the i^{th} segment starts a streak of length at least ℓ. Therefore $\Pr[Y_i = 1] = 2/2^\ell$.

Though n/ℓ segments is less than the total number of potential segments among n flips, it suffices for giving a sufficiently low probability. We now bound the probability of all n/ℓ segments has no streaks of length ℓ by

$$\Pr\left[\bigcap_{i\in[1,n/\ell]} Y_i = 0\right] = \left(1 - \frac{2}{2^\ell}\right)^{n/\ell} = \left(1 - \frac{1}{2^{\ell-1}}\right)^{2^{\ell-1}\cdot n/(\ell\cdot 2^{\ell-1})} \leq \left(\frac{1}{e}\right)^{n/(\ell\cdot 2^{\ell-1})}.$$

Let $\eta = n/\left(\ell\cdot 2^{\ell-1}\right)$. By $\lfloor x \rfloor \leq x$ for any x, we lower bound η by

$$\begin{aligned}\eta &= \frac{n}{\ell\cdot 2^{\ell-1}} = \frac{2n}{\lfloor \log_2 n - 2\log_2\log_2 n\rfloor}\cdot\frac{1}{2^{\lfloor \log_2 n - 2\log_2\log_2 n\rfloor}} \\ &\geq \frac{2n}{\log_2 n - 2\log_2\log_2 n}\cdot\frac{\log_2^2 n}{n} = \frac{2\log_2^2 n}{\log_2 n - 2\log_2\log_2 n} \\ &\geq 2\log_2 n.\end{aligned}$$

Therefore, $\exp(-\eta)$ is at most $1/n$, and thus completes the proof. □

Exercise 1.2.4 (Exercise 2.20 [MU17])**.** A permutation on the numbers $[1, n]$ can be represented as a function $\pi : [1, n] \mapsto [1, n]$. A fixed point of a permutation π is a value for which $\pi(x) = x$. Find the expected number of fixed points for a permutation chosen uniformly at random for all permutations.

Proof. This is another indicator counting problem.

We introduce n 0-1 random variables $\{X_i\}_{i\in[1,n]}$ indicating if i^{th} value is a fixed point for the permutation. Then $\Pr[X_i = 1] = 1/n$, and thus the expectation of $X = \sum X_i$, which is the number of fixed points in π, is

$$\mathbb{E}[X] = \mathbb{E}\left[\sum_{i\in[1,n]} X_i\right] = \sum_{i\in[1,n]} \mathbb{E}[X_i] = 1.$$

□

Theorem 1.2.5 (Jensen's Inequality)**.** *If f is a convex function, then*

$$\mathbb{E}\left[f(X)\right] \geq f(\mathbb{E}[X]).$$

If f is concave, then

$$\mathbb{E}\left[f(X)\right] \leq f(\mathbb{E}[X]).$$

Definition 1.2.6. The expression $\mathbb{E}[Y \mid X]$ is a random variable $f(X)$ that takes on value $\mathbb{E}[Y \mid X = x]$ when $X = x$.

Lemma 1.2.7.

$$\mathbb{E}[Y] = \mathbb{E}\left[\mathbb{E}[Y \mid X]\right].$$

1.3 Moments and Deviations

Exercise 1.3.1 (Exercise 3.9 [MU17]). Let $\{X_i\}_{i\in[1,m]}$ be Bernoulli random variables that are not necessarily mutually independent, and $X = \sum_i X_i$. Then

$$\mathbb{E}\left[X^2\right] = \sum_{i\in[1,m]} \Pr\left[X_i = 1\right]\mathbb{E}\left[X \mid X_i = 1\right].$$

Proof. We can derive the proof by

$$\begin{aligned}\mathbb{E}\left[X^2\right] &= \mathbb{E}\left[\sum_{i\in[1,m]} X_i X\right] = \sum_{i\in[1,m]} \mathbb{E}\left[X_i X\right] \\ &= \sum_{i\in[1,m]} \Pr\left[X_i = 1\right]\mathbb{E}\left[X \mid X_i = 1\right],\end{aligned}$$

where the first two equalities hold from linearity of expectation, and the last one holds by conditional expectation. □

Lemma 1.3.2. *Let $\{X_i\}_{i\in[1,m]}$ be Bernoulli random variables that are not necessarily mutually independent, and $X = \sum_i X_i$. Then*

$$\begin{aligned}\operatorname{Var}[X] &= \sum_{i\in[1,m]} \operatorname{Var}[X_i] + \sum_{\substack{i,j\in[1,m]\\ i\neq j}} \operatorname{Cov}\left[X_i, X_j\right] \\ &= \sum_{i\in[1,m]} \operatorname{Var}[X_i] + 2\cdot\sum_{\substack{i,j\in[1,m]\\ i<j}} \operatorname{Cov}\left[X_i, X_j\right].\end{aligned}$$

Exercise 1.3.3 (Exercise 3.11 [MU17]). Recall the Bubblesort algorithm. Determine the variance of the number of inversions that need to be corrected by Bubblesort.

Proof. We introduce $\{X_{i,j}\}$ 0-1 random variables, where $j \in [1, n]$ and $i \in [1, j-1]$. $X_{i,j}$ is 1 if and only if the i^{th} and the j^{th} element are inverted during the sorting. The variance of $X = \sum_{i,j} X_{i,j}$ can be expressed by lemma 1.3.2

$$\operatorname{Var}[X] = \sum_{\substack{j\in[1,n]\\ i\in[1,j-1]}} \operatorname{Var}\left[X_{i,j}\right] + 2\cdot \sum_{\substack{j_0,j_1\in[1,n]\\ i_0\in[1,j_0-1], i_1\in[1,j_1-1]\\ (i_0,j_0)<(i_1,j_1)}} \operatorname{Cov}\left[X_{i_0,j_0}, X_{i_1,j_1}\right].$$

$X_{i,j} = 1$ if and only if the i^{th} element is greater than j^{th} element, and thus $\Pr\left[X_{i,j} = 1\right] = 1/2$. Since $\mathbb{E}\left[X_{i,j}\right] = 1/2$, therefore $\operatorname{Var}\left[X_{i,j}\right] = 1/4$.

For $\operatorname{Cov}\left[X_{i_0,j_0}, X_{i_1,j_1}\right]$, if the two pairs are disjoint, that $\{i_0, j_0\}\cap\{i_1, j_1\} = \varnothing$, then X_{i_0,j_0} and X_{i_1,j_1} are independent, therefore $\operatorname{Cov}\left[X_{i_0,j_0}, X_{i_1,j_1}\right] = 0$.

On the other hand, if i_0, i_1, j_0, j_1 have index overlap, there are $\binom{n}{3}$ choices, each breaking into 3 categories:

- if $i_0 = i_1$, then $X_{i_0,j_0}\cdot X_{i_1,j_1} = 1$ if and only if i_0^{th} element is the largest among 3, and thus $\mathbb{E}\left[X_{i_0,j_0}\cdot X_{i_1,j_1}\right] = 1/3$.
- if $j_0 = i_1$, then $X_{i_0,j_0}\cdot X_{i_1,j_1} = 1$ if and only if j_0^{th} element is in the middle, and i_0^{th} element is the largest, which halves the probability, and therefore $\mathbb{E}\left[X_{i_0,j_0}\cdot X_{i_1,j_1}\right] = 1/6$.
- if $j_0 = j_1$, then $X_{i_0,j_0}\cdot X_{i_1,j_1} = 1$ if and only if j_0^{th} element is the smallest among 3, and thus $\mathbb{E}\left[X_{i_0,j_0}\cdot X_{i_1,j_1}\right] = 1/3$.

We conclude that

$$\operatorname{Var}[X] = \sum_{j\in[1,n]}\sum_{i\in[1,j-1]} \frac{1}{4} + 2\cdot\binom{n}{3}\cdot\left(\frac{1}{3}+\frac{1}{3}+\frac{1}{6}-\frac{3}{4}\right) = \frac{n(n-1)}{8} + 2\cdot\frac{1}{12}\cdot\binom{n}{3} = \frac{1}{72}n(n-1)(2n+5).$$

□

Exercise 1.3.4 (Exercise 3.13 [MU17])**.** Find an example of random variable with finite j^{th} moment for $j \in [1, k]$, but an unbounded $k+1^{\text{th}}$ moment.

Proof. Let $X \in \mathbb{N}^+$ have $\Pr[X = x] = c \cdot x^{-k-2}$ with $c = (\sum_{x \in \mathbb{N}^+} x^{-k-2})^{-1}$. $\mathbb{E}\left[X^k\right]$ converges, $\mathbb{E}\left[X^{k+1}\right]$ doesn't. □

Remark 1.3.5. Ever heard of Pareto tails?

Theorem 1.3.6. *For any random variable X and finite expectation $\mathbb{E}[X]$, finite $\mathbb{E}\left[X^2\right]$, and finite median m,*

- $\mathbb{E}[X]$ *is the value c that minimizes the expression* $\mathbb{E}\left[(X-c)^2\right]$.
- m *is the value c that minimizes the expression* $\mathbb{E}[|X-c|]$.

Proof. The first one is immediate, as $\mathbb{E}\left[(X-c)^2\right]$ minimizes at $c = \mathbb{E}[X]$ by expanding the expression.

For second case, we assume that there exists c such that $\mathbb{E}[|X-c|] < \mathbb{E}[|X-m|]$, or $\mathbb{E}[|X-m| - |X-c|] > 0$. Since $c > m$ or $c < m$ works the same by symmetry, WLOG we let $c > m$, then

$$|X-m| - |X-c| = \begin{cases} m-c & X \le m \\ 2X-m-c & m \le X \le c \\ c-m & X \ge c \end{cases}.$$

For $m < X < c$, we have $2X - m - c \le c - m$, and for $X \ge c$ the expression equals $c - m$, so

$$\mathbb{E}[|X-m| - |X-c|] \le (m-c)\Pr[X \le m] + (c-m)\Pr[X > m].$$

Since for a median m, $\Pr[X \le m] \ge 1/2$ and $\Pr[X > m] \le 1/2$, the right-hand side is at most 0, contradicting the prior assumption of existence of c, and proves that m minimizes the expression $\mathbb{E}[|X-c|]$. □

Lemma 1.3.7. *Let X be a random variable, with mean μ, median m, and finite standard deviation σ. Then* $|\mu - m| \le \sigma$.

Proof.

$$\begin{aligned} |\mathbb{E}[X] - m| &\le \mathbb{E}[|X-m|] \\ &\le \mathbb{E}[|X - \mathbb{E}[X]|] \\ &\le \sqrt{\mathbb{E}\left[(X-\mathbb{E}[X])^2\right]} = \sigma, \end{aligned}$$

where the first and the third inequality follow from theorem 1.2.5, and the second one follows from theorem 1.3.6. □

Exercise 1.3.8 (Exercise 3.18 [MU17])**.** Show that, for a random variable X with standard deviation $\sigma[X]$ and $t \in \mathbb{R}^+$:

- $\Pr[X - \mathbb{E}[X] \ge t\sigma[X]] \le 1/(1+t^2)$;
- $\Pr[|X - \mathbb{E}[X]| \ge t\sigma[X]] \le 2/(1+t^2)$.

Proof. Cantelli's inequality (or one-sided Chebyshev's inequality) can be proved by:

$$\begin{aligned} \Pr[X - \mathbb{E}[X] \ge t\sigma[X]] &= \Pr[X - \mathbb{E}[X] + u \ge t\sigma[X] + u] \\ &\le \Pr[(X - \mathbb{E}[X] + u)^2 \ge (t\sigma[X] + u)^2] \\ &\le \frac{\mathbb{E}[(X - \mathbb{E}[X] + u)^2]}{(t\sigma[X] + u)^2} = \frac{\mathbb{E}[(X - \mathbb{E}[X])^2] + u^2}{(t\sigma[X] + u)^2} \\ &= \frac{\sigma[X]^2 + u^2}{(t\sigma[X] + u)^2} = \varphi(u). \end{aligned}$$

The last inequality was established by Chebyshev's inequality.

$\varphi(u)$ is minimized at $u = \sigma[X]/t$, and thus the first inequality is upper bounded by $1/(1+t^2)$.

By symmetry, the second inequality is proven. □

Exercise 1.3.9 (Exercise 3.19 [MU17]). Using exercise 1.3.8, show lemma 1.3.7.

Proof. By exercise 1.3.8, we have $\Pr[X \leq \sigma + \mu] \geq 1/2$ and $\Pr[X \geq \mu - \sigma] \geq 1/2$.

By definition of median m, $\Pr[X \geq m] \geq 1/2$ and $\Pr[X \leq m] \geq 1/2$.

Since $\sigma > 0$ is finite, μ is finite, we have $\mu - \sigma \leq m \leq \mu + \sigma$, and thus completes the proof for $|m - \mu| \leq \sigma$. □

Exercise 1.3.10 (Exercise 3.20 [MU17]). Let Y be a nonnegative integer-valued random variable with positive expectation. Prove

$$\frac{\mathbb{E}[Y]^2}{\mathbb{E}[Y^2]} \leq \Pr[Y \neq 0] \leq \mathbb{E}[Y].$$

Proof. By conditional expectation, we have

$$\begin{aligned}\mathbb{E}[Y^2] &= \mathbb{E}[Y^2 \mid Y = 0] \cdot \Pr[Y = 0] + \mathbb{E}[Y^2 \mid Y \neq 0] \cdot \Pr[Y \neq 0] \\ &= \mathbb{E}[Y^2 \mid Y \neq 0] \cdot \Pr[Y \neq 0] \\ &\geq \mathbb{E}[Y \mid Y \neq 0]^2 \cdot \Pr[Y \neq 0],\end{aligned}$$

where the last inequality follows from Jensen's inequality.

Again by conditional expectation, we have $\mathbb{E}[Y] = \mathbb{E}[Y \mid Y \neq 0] \cdot \Pr[Y \neq 0]$, and thus

$$\frac{\mathbb{E}[Y]^2}{\mathbb{E}[Y^2]} \leq \Pr[Y \neq 0].$$

Since Y is an integer-valued random variable,

$$\begin{aligned}\Pr[Y \neq 0] &= 0 \cdot \Pr[Y = 0] + \sum_{y \neq 0} 1 \cdot \Pr[Y = y] \\ &\leq \sum_{y \in \mathbb{N}} y \cdot \Pr[Y = y] = \mathbb{E}[Y].\end{aligned}$$

Another way of seeing it is by a direct application of Markov's inequality, that $\Pr[Y \geq 1] \leq \mathbb{E}[Y]$. □

1.3.1 Weak Law of Large Numbers (WLLN)

Theorem 1.3.11 (WLLN). *If $\{X_i\}_{i \in [1,n]}$ are independent, identically distributed random variables with mean μ and standard deviation σ, then for any constant $\varepsilon > 0$,*

$$\lim_{n \to \infty} \Pr\left[\left|\frac{1}{n}\sum_i X_i - \mu\right| > \varepsilon\right] = 0.$$

1.4 Chernoff and Hoeffding Bounds

Exercise 1.4.1 (Exercise 4.9 [MU17])**.** Suppose that we can obtain independent samples $\{X_i\}_{i\in[1,n]}$ of a random variable X and that we want to use these samples to estimate $\mathbb{E}[X]$. Using t samples, we use $(\sum_{i\in[1,t]} X_i)/t$ for our estimate of $\mathbb{E}[X]$. We want the estimate to be within $\varepsilon\,\mathbb{E}[X]$ from the true value of $\mathbb{E}[X]$ with probability at least $1-\delta$. We may not be able to use Chernoff's bound directly to bound how good our estimate is if X is not a 0-1 random variable, and we do not know its moment generating function. We develop an alternative approach that requires only having a bound on the variance of X. Let $r = \sqrt{\mathrm{Var}[X]}/\mathbb{E}[X]$.

- Show using Chebyshev's inequality that $O(r^2/\varepsilon^2\delta)$ samples are sufficient to solve the problem.
- Suppose that we need only a weak estimate that is within $\varepsilon\,\mathbb{E}[X]$ of $\mathbb{E}[X]$ with probability at least $3/4$. Argue that $O(r^2/\varepsilon^2)$ samples are enough for this weak estimate.
- Show that, by taking the median of $O(\log(1/\delta))$ independent weak estimates, we can obtain an estimate within $\varepsilon\,\mathbb{E}[X]$ of $\mathbb{E}[X]$ with probability at least $1-\delta$. Conclude that we need only $O((r^2\log(1/\delta))/\varepsilon^2)$ samples.

Proof. Let $Y = \sum_{i\in[1,t]} X_i$, then $\mathbb{E}[Y] = t\cdot\mathbb{E}[X]$ and $\mathrm{Var}[Y] = \mathrm{Var}[\sum_{i\in[1,t]} X_i] = t\cdot\mathrm{Var}[X]$.

By Chebyshev's inequality,

$$\Pr\left[\left|\frac{1}{t}\cdot Y - \mathbb{E}[X]\right| \geq \varepsilon\,\mathbb{E}[X]\right] = \Pr\left[|Y-\mathbb{E}[Y]| \geq \varepsilon\,\mathbb{E}[Y]\right]$$
$$\leq \frac{\mathrm{Var}[Y]}{\varepsilon^2\,\mathbb{E}[Y]^2} = \frac{\mathrm{Var}[X]}{t\varepsilon^2\,\mathbb{E}[X]^2} = \frac{r^2}{t\varepsilon^2} \leq \delta.$$

Now that $t \geq r^2/\delta\varepsilon^2$, thus $t = O(r^2/\delta\varepsilon^2)$.

Suppose $\delta \leq 1/4$, then $t \geq 4r^2/\varepsilon^2$, and thus $t = O(r^2/\varepsilon^2)$.

When it comes to using a median of s weak estimates to boost the confidence in estimation, we have a trick called *Median of Mean* (*MoM*) [JVV86]. The key idea is to take advantage of the "structure" of median value, that half of the weak estimates are smaller, while the other half are greater.

Since we hope with probability at least $1-\delta$, the median of weak estimates is within $\varepsilon\,\mathbb{E}[X]$ range, then we say the bad event is the median of weak estimates is out of $\varepsilon\,\mathbb{E}[X]$ range. Equivalently, a bad event occurs when more than half of the weak estimates are above $(1+\varepsilon)\,\mathbb{E}[X]$, or more than half of them are below $(1-\varepsilon)\,\mathbb{E}[X]$.

Form each weak estimate from an independent batch of $O(r^2/\varepsilon^2)$ samples, so the weak estimates are independent. In this sense, we introduce independent 0-1 random variables (indicators) $\{1_{L,i}\}_{i\in[1,s]}$ and $\{1_{R,i}\}_{i\in[1,s]}$ where

- $1_{L,i} = 1$ only when i^{th} weak estimate is smaller than $(1-\varepsilon)\,\mathbb{E}[X]$,
- $1_{R,i} = 1$ only when i^{th} weak estimate is greater than $(1+\varepsilon)\,\mathbb{E}[X]$,

and $1_L = \sum_{i\in[1,s]} 1_{L,i}$, $1_R = \sum_{i\in[1,s]} 1_{R,i}$. Since a weak estimate is within $\varepsilon\,\mathbb{E}[X]$ range with probability at least $3/4$, we have both $\Pr[1_{L,i}=1]$ and $\Pr[1_{R,i}=1]$ upper bounded by $1/4$, and thus both $\mathbb{E}[1_L]$ and $\mathbb{E}[1_R]$ are at most $s/4$.

We first bound 1_L, and derive 1_R by symmetry. By Chernoff's bound,

$$\Pr\left[1_L \geq \frac{s}{2}\right] \leq \mathbb{E}\left[\exp(t\cdot 1_L)\right]/\exp\left(\frac{s}{2}\cdot t\right) \leq \exp\left(\frac{s}{4}\cdot(e^t-1) - \frac{s}{2}\cdot t\right) = \exp(\varphi(t)).$$

Then we have $\varphi'(t) = \frac{s}{4}\cdot e^t - \frac{s}{2}$, and derive that $\varphi(t)$ min by $t = \ln 2$ for a positive t. Now that

$$\exp\left(\left(\frac{1}{4} - \frac{\ln 2}{2}\right)\cdot s\right) \leq \frac{\delta}{2},$$

we have $s \geq ((\ln 2)/2 - 1/4)^{-1}\cdot\ln(2/\delta)$, thus $s = O(\log(1/\delta))$.

We thus conclude that we need $O(r^2\log(1/\delta)/\varepsilon^2)$ samples in total. □

Exercise 1.4.2 (Exercise 4.12 [MU17]). Consider a collection $\{X_i\}_{i\in[1,n]}$ of n independent geometrically distributed random variables with mean 2. Let $X = \sum_{i\in[1,n]} X_i$ and $\delta > 0$.

- Derive a bound on $\Pr[X \geq (1+\delta)\cdot 2n]$ by applying Chernoff's bound to a sequence of $(1+\delta)\cdot 2n$ fair coin tosses.
- Directly derive a Chernoff's bound on $\Pr[X \geq (1+\delta)\cdot 2n]$ using the MGF for geometric random variables.
- Compare these 2 bounds derived above.

Proof. A way to model the "Coupon Collection" alike problem for bounding the sum of geometrically distributed random variables is transforming to an upper bound to the sum of a sequence of Bernoulli trials.

Since a geometrically distributed random variable can be seen as the number of Bernoulli trials until one success, then for event $X \geq (1+\delta)\cdot 2n$, there are at most n successes among $(1+\delta)\cdot 2n$ independent Bernoulli trials (in the worst case, the last trial must be success among n successes in total $(1+\delta)\cdot 2n$ trials).

We write Y_i as i^{th} independent Bernoulli trial with success probability being $1/2$ among $(1+\delta)\cdot 2n$ trials, we write $Y = \sum_{i\in[1,(1+\delta)\cdot 2n]} Y_i$, and $\mathbb{E}[Y] = (1+\delta)\cdot n$. We thus can upper bound $\Pr[X \geq (1+\delta)\cdot 2n]$ by $\Pr[Y \leq n]$, as we can loosen the boundary condition on the worst case, where last trial must be successful among n successes.

By Chernoff's bound for bounding the deviation below the mean, we have

$$\begin{aligned}
\Pr[X \geq (1+\delta)\cdot 2n] &\leq \Pr[Y \leq n] \\
&\leq \frac{\mathbb{E}[\exp(tY)]}{\exp(n\cdot t)} \\
&\leq \frac{\exp((1+\delta)\cdot n\cdot(e^t-1))}{\exp(n\cdot t)} \\
&= \exp((1+\delta)\cdot n\cdot(e^t-1) - n\cdot t) = \exp(\varphi(t))
\end{aligned}$$

with $t < 0$, and min by $t = -\ln(1+\delta)$ deriving a upper bound $((1+\delta)\cdot e^{-\delta})^n$.

If we brute force for a bound on $X = \sum_{i\in[1,n]} X_i$, the MGF of X_i is $M_{X_i}(t) = \sum_{k\geq 1}\exp(kt)\cdot 2^{-k} = \sum_{k\geq 1}(e^t/2)^k$. When $0 < t < \ln 2$, $M_{X_i}(t) = e^t/(2-e^t)$. Again by Chernoff's bound for bounding the deviation above the mean,

$$\begin{aligned}
\Pr[X \geq (1+\delta)\cdot 2n] &\leq \left(\frac{M_{X_i}(t)}{\exp((1+\delta)\cdot 2t)}\right)^n \\
&= \left((2-e^t)\exp((2\delta+1)\cdot t)\right)^{-n} \\
&= \left(2\exp((2\delta+1)\cdot t) - \exp((2\delta+2)\cdot t)\right)^{-n} = \varphi(t)^{-n},
\end{aligned}$$

and $\varphi'(t) = (4\delta+2)\exp((2\delta+1)\cdot t) - (2\delta+2)\exp((2\delta+2)\cdot t)$, and thus $\varphi(t)$ max by $t = \ln(2\delta+1) - \ln(\delta+1)$,

$$\Pr[X \geq (1+\delta)\cdot 2n] \leq \left(\frac{1}{\delta+1}\cdot\left(\frac{2\delta+1}{\delta+1}\right)^{2\delta+1}\right)^{-n} = \left((1+\delta)\cdot\left(\frac{\delta+1}{2\delta+1}\right)^{2\delta+1}\right)^{n}.$$

By section 1.5.1, we know the first one is looser. □

Exercise 1.4.3 (Exercise 4.8 [MU17]). We show how to construct a random permutation π on $[1,n]$, given a black box that outputs numbers independently and uniformly at random from $[1,k]$ where $k \geq n$. If we compute a function $f : [1,n] \to [1,k]$ with $f(i) \neq f(j)$ for $i \neq j$, this yields a permutation, by outputting $[1,n]$ according to the order of $\{f(i)\}_{i\in[1,n]}$. To construct such a function f, do the following for $i \in [1,n]$: choose $f(i)$ by repeatedly drawing number from the black box and set $f(i)$ to the first number that is $f(i) \neq f(j)$ for $j < i$.

Prove that this approach gives a permutation chosen uniformly at random for all permutations. Find the expected number of calls to the black box that are needed when $k = n$ and $k = 2n$. For the case $k = 2n$, argue that the probability that each call to the black box assigns a value of $f(j)$ to some j is at least $1/2$. Based on this, use a Chernoff bound to bound the probability that the number of calls to the black box is at least $4n$.

Proof. At the first sight, this is a "Coupon Collection" problem, where we can introduce n independent geometrically distributed random variables $\{X_i\}_{i\in[1,n]}$ each with success probability $p_i = (k+1-i)/k$, and we write $X = \sum_{i\in[1,n]} X_i$. When $k = 2n$, each $p_i \geq 1/2$ as $p_n = (n+1)/2n$.

For expectation of X, $\mathbb{E}[X] = \sum_{i\in[1,n]} \mathbb{E}[X_i]$ by linearity of expectation. For $\mathbb{E}[X_i]$, by conditional expectation and memoryless of geometric distribution,

$$\begin{aligned}\mathbb{E}[X_i] &= \mathbb{E}[X_i \mid X_{i,0} = 1] \cdot \Pr[X_{i,0} = 1] + \mathbb{E}[X_i \mid X_{i,0} = 0] \cdot \Pr[X_{i,0} = 0] \\ &= 1 \cdot p_i + (1 + \mathbb{E}[X_i]) \cdot (1 - p_i),\end{aligned}$$

$\mathbb{E}[X_i] = p_i^{-1} = k/(k-i+1)$. Thus $\mathbb{E}[X] = n \ln n + \Theta(n)$ when $k = n$, $\mathbb{E}[X] = 2n(\ln 2n - \ln n) + \Theta(n)$ when $k = 2n$.

For upper bounding $\Pr[X \geq 4n]$, we first upper bound by $\Pr[Y \geq 4n]$, where $\{Y_i\}_{i\in[1,n]}$ are n independent geometrically distributed random variables, each with success probability $1/2$, and we write $Y = \sum_{i\in[1,n]} Y_i$. Then we upper bound $\Pr[Y \geq 4n]$ by $\Pr[Z \leq n]$ through the previous trick of turning sum of geometrically distributed random variables into sum of a sequence of Bernoulli trials. We introduce $4n$ independent Bernoulli distributed random variables $\{Z_i\}_{i\in[1,4n]}$ with $\Pr[Z_i = 1] = 1/2$, and we write $Z = \sum_{i\in[1,4n]} Z_i$.

By Chernoff's bound, we have

$$\begin{aligned}\Pr[X \geq 4n] &\leq \Pr[Y \geq 4n] \\ &\leq \Pr[Z \leq n] \\ &\leq \exp(2n(e^t - 1) - nt) = \exp(\varphi(t)),\end{aligned}$$

with $\varphi(t)$ min at $t = -\ln 2$, thus derives an upper bound by $(2/e)^n$. □

Exercise 1.4.4 (Exercise 4.13 [MU17]). Let $\{X_i\}_{i\in[1,n]}$ be independent Poisson trials such that $\Pr[X_i = 1] = p$. Let $X = \sum_{i\in[1,n]} X_i$ so that $\mathbb{E}[X] = pn$. Let

$$F(x,p) = x \ln \frac{x}{p} + (1-x) \ln \frac{1-x}{1-p}.$$

- Show that, for $1 \geq x > p$,
$$\Pr[X \geq xn] \leq e^{-nF(x,p)}.$$
- Show that, when $0 < x, p < 1$, we have $F(x,p) - 2(x-p)^2 \geq 0$.
- Argue that
$$\begin{aligned}\Pr\left[X \geq (p+\varepsilon)n\right] &\leq e^{-2n\varepsilon^2}, \\ \Pr\left[X \leq (p-\varepsilon)n\right] &\leq e^{-2n\varepsilon^2}, \\ \Pr\left[\left|X - pn\right| \geq \varepsilon n\right] &\leq 2e^{-2n\varepsilon^2}.\end{aligned}$$

Proof. For Chernoff's bound on $\Pr[X_i \geq x]$, we typically turn to $1 + kx \leq e^{kx}$ for $k > 0$, then

$$\begin{aligned}\frac{\mathbb{E}\left[\exp(tX_i)\right]}{\exp(xt)} = \frac{p \cdot e^t + 1 - p}{\exp(xt)} &= \frac{1 + p \cdot (e^t - 1)}{\exp(xt)} \\ &\leq \exp(p \cdot (e^t - 1) - xt) = \exp(\varphi(t)),\end{aligned}$$

to derive an upper bound by choosing t. A different bound can be derived by avoiding the trick we often apply,

$$\frac{\mathbb{E}\left[\exp(tX_i)\right]}{\exp(xt)} = p \cdot \exp((1-x) \cdot t) + (1-p) \cdot \exp(-xt) = \varphi(t),$$

then we have

$$\varphi'(t) = p(1-x) \cdot \exp((1-x) \cdot t) - x(1-p) \cdot \exp(-xt).$$

When $t = \ln(1-p) + \ln x - \ln(1-x) - \ln p > 0$, $\varphi(t)$ min by

$$\begin{aligned}\min_t \varphi(t) &= p \cdot \left(\frac{(1-p)x}{(1-x)p}\right)^{1-x} + (1-p) \cdot \left(\frac{(1-p)x}{(1-x)p}\right)^{-x} \\ &= \left(p \cdot \frac{(1-p)x}{(1-x)p} + 1 - p\right) \cdot \left(\frac{(1-p)x}{(1-x)p}\right)^{-x} \\ &= \frac{1-p}{1-x} \cdot \left(\frac{(1-p)x}{(1-x)p}\right)^{-x}.\end{aligned}$$

By taking ln both sides, we have

$$\begin{aligned}\ln \min_t \varphi(t) &= \ln \frac{1-p}{1-x} - x \ln \frac{(1-p)x}{(1-x)p} \\ &= (1-x) \ln \frac{1-p}{1-x} + x \ln \frac{p}{x} = -F(x,p),\end{aligned}$$

thus deriving a bound for $\Pr[X_i \geq x] \leq \exp(-F(x,p))$. Finally, when bounding $\Pr[X \geq xn]$,

$$\begin{aligned}\Pr[X \geq xn] &\leq \frac{\mathbb{E}\left[\exp(tX)\right]}{\exp(xnt)} \\ &= \frac{\prod_{i\in[1,n]} \mathbb{E}\left[\exp(tX_i)\right]}{\exp(xnt)} = \left(\frac{\mathbb{E}\left[\exp(tX_i)\right]}{\exp(xt)}\right)^n \\ &\leq \exp(-nF(x,p)).\end{aligned}$$

Taking partial derivative against x in $F(x,p)$, we have

$$\frac{\partial}{\partial x} F(x,p) = \ln \frac{x}{p} + x \cdot \frac{p}{x} \cdot \frac{1}{p} - \ln \frac{1-x}{1-p} + (1-x) \cdot \frac{1-p}{1-x} \cdot \frac{-1}{1-p} = \ln \frac{x}{p} - \ln \frac{1-x}{1-p},$$

and it is 0 when $x = p$, taking second order partial derivative against x, we have

$$\frac{\partial^2}{\partial x^2} F(x,p) = \frac{1}{x} - \frac{1-p}{1-x} \cdot \frac{-1}{1-p} = \frac{1}{x} + \frac{1}{1-x} \geq 4,$$

and therefore $F(x,p) \geq F(p,p) + 2(x-p)^2 = 2(x-p)^2$ by Taylor expansion.

We can plug in the result $F(x,p) \geq 2(x-p)^2$, and derive

$$\Pr\left[X \geq (p+\varepsilon)n\right] \leq \exp(-nF(p+\varepsilon,p)) \leq \exp(-2n\varepsilon^2).$$

When $x < p$ and we upper bound $\Pr[X_i \leq x]$, $t = \ln(1-p) + \ln x - \ln(1-x) - \ln p < 0$ gives $\ln \min_t \varphi(t) = -F(x,p)$, then $\Pr[X_i \leq x] \leq \exp(-F(x,p))$. Thus,

$$\Pr\left[X \leq (p-\varepsilon)n\right] \leq \exp(-nF(p-\varepsilon,p)) \leq \exp(-2n\varepsilon^2).$$

Eventually, we derive bounds on both sides by

$$\Pr\left[\left|X - pn\right| \geq \varepsilon n\right] \leq 2\exp(-2n\varepsilon^2).$$

Remark 1.4.5. $F(x,p)$ is exactly the KL divergence [KL51] between Bernoulli distributions with parameters x and p.

□

Exercise 1.4.6 (Exercise 4.14 [MU17]). Let $\{X_i\}_{i\in[1,n]}$ be independent Poisson trials such that $\Pr[X_i] = p_i$, and let $\{a_i\}_{i\in[1,n]}$ be real numbers in $[0,1]$. Let $X = \sum_{i\in[1,n]} a_i X_i$ and $\mu = \mathbb{E}[X]$. Then the following Chernoff bound holds: for any $\delta > 0$,

$$\Pr\left[X \geq (1+\delta)\mu\right] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

Prove a similar bound for the probability that $X \leq (1-\delta)\mu$ for $0 < \delta < 1$.

Proof. By Jensen's inequality over concave functions,

$$\mathbb{E}\left[\exp(ta_iX_i)\right] \leq \mathbb{E}\left[\exp(tX_i)\right]^{a_i}$$

for all t, as $f(x) = x^{a_i}$ is concave if $0 \leq a_i \leq 1$.

By Chernoff's upper tail and lower tail bound,

$$\Pr\left[X_i \geq (1+\delta)p_i\right] \leq \frac{\mathbb{E}\left[\exp(tX_i)\right]}{\exp((1+\delta)p_i\cdot t)} \leq \frac{\exp(p_i\cdot(e^t-1))}{\exp((1+\delta)p_i\cdot t)}, \text{ min by } t=\ln(1+\delta)>0 \text{ by } \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{p_i},$$
$$\Pr\left[X_i \leq (1-\delta)p_i\right] \leq \frac{\mathbb{E}\left[\exp(tX_i)\right]}{\exp((1-\delta)p_i\cdot t)} \leq \frac{\exp(p_i\cdot(e^t-1))}{\exp((1-\delta)p_i\cdot t)}, \text{ min by } t=\ln(1-\delta)<0 \text{ by } \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{p_i}.$$

By Chernoff's bound and Jensen's inequality, for upper tail bound, given $t > 0$ and $\delta > 0$,

$$\begin{aligned}\Pr\left[X \geq (1+\delta)\mu\right] &\leq \frac{\mathbb{E}\left[\exp(tX)\right]}{\exp((1+\delta)\mu\cdot t)} = \prod_{i\in[1,n]}\left(\frac{\mathbb{E}\left[\exp(ta_iX_i)\right]}{\exp((1+\delta)a_ip_i\cdot t)}\right)\\ &\leq \prod_{i\in[1,n]}\left(\frac{\mathbb{E}\left[\exp(tX_i)\right]}{\exp((1+\delta)p_i\cdot t)}\right)^{a_i}\\ &\leq \prod_{i\in[1,n]}\left(\frac{\exp(p_i\cdot(e^t-1))}{\exp((1+\delta)p_i\cdot t)}\right)^{a_i} = \left(\frac{\exp(e^t-1)}{\exp((1+\delta)\cdot t)}\right)^\mu,\end{aligned}$$

and right hand side is minimized by

$$\min_{t>0}\left(\frac{\exp(e^t-1)}{\exp((1+\delta)\cdot t)}\right)^\mu = \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^\mu.$$

Symmetrically, for lower tail bound, given $t < 0$ and $0 < \delta < 1$,

$$\begin{aligned}\Pr\left[X \leq (1-\delta)\mu\right] &\leq \frac{\mathbb{E}\left[\exp(tX)\right]}{\exp((1-\delta)\mu\cdot t)} = \prod_{i\in[1,n]}\left(\frac{\mathbb{E}\left[\exp(ta_iX_i)\right]}{\exp((1-\delta)a_ip_i\cdot t)}\right)\\ &\leq \prod_{i\in[1,n]}\left(\frac{\mathbb{E}\left[\exp(tX_i)\right]}{\exp((1-\delta)p_i\cdot t)}\right)^{a_i}\\ &\leq \prod_{i\in[1,n]}\left(\frac{\exp(p_i\cdot(e^t-1))}{\exp((1-\delta)p_i\cdot t)}\right)^{a_i} = \left(\frac{\exp(e^t-1)}{\exp((1-\delta)\cdot t)}\right)^\mu,\end{aligned}$$

with right hand side minimized by

$$\min_{t<0}\left(\frac{\exp(e^t-1)}{\exp((1-\delta)\cdot t)}\right)^\mu = \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^\mu.$$

□

Exercise 1.4.7 (Exercise 4.7 [MU17]). Let $X = \sum_{i\in[1,n]} X_i$, where $\{X_i\}_{i\in[1,n]}$ are independent 0-1 random variables. Let $\mu = \mathbb{E}[X]$. Choose any μ_L and μ_H such that $\mu_L \le \mu \le \mu_H$. Then, for any $\delta > 0$,

$$\Pr\left[X \ge (1+\delta)\mu_H\right] \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu_H}.$$

Similarly, for any $0 < \delta < 1$,

$$\Pr\left[X \le (1-\delta)\mu_L\right] \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_L}.$$

Proof. The MGF of X can be relaxed in following manner

$$\mathbb{E}\left[\exp(tX)\right] = \prod_{i\in[1,n]} \mathbb{E}\left[\exp(tX_i)\right] \le \prod_{i\in[1,n]} \exp\left(p_i\cdot(e^t-1)\right) = \exp\left(\mu\cdot(e^t-1)\right).$$

Given $\mu_L \le \mu \le \mu_H$,

- When $t > 0$, $\exp\left(\mu\cdot(e^t-1)\right) \le \exp\left(\mu_H\cdot(e^t-1)\right)$.
- When $t < 0$, $\exp\left(\mu\cdot(e^t-1)\right) \le \exp\left(\mu_L\cdot(e^t-1)\right)$.

Then for upper tail with $t > 0$ and $\delta > 0$,

$$\Pr\left[X \ge (1+\delta)\mu_H\right] \le \frac{\exp(\mu\cdot(e^t-1))}{\exp((1+\delta)\mu_H\cdot t)} \le \frac{\exp(\mu_H\cdot(e^t-1))}{\exp((1+\delta)\mu_H\cdot t)} = \left(\frac{\exp\left(e^t-1\right)}{\exp\left((1+\delta)\cdot t\right)}\right)^{\mu_H},$$

the right hand side is minimized by

$$\min_{t>0}\left(\frac{\exp(e^t-1)}{\exp((1+\delta)\cdot t)}\right)^{\mu_H} = \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu_H}.$$

For lower tail with $t < 0$ and $0 < \delta < 1$,

$$\Pr\left[X \le (1-\delta)\mu_L\right] \le \frac{\exp(\mu\cdot(e^t-1))}{\exp((1-\delta)\mu_L\cdot t)} \le \frac{\exp(\mu_L\cdot(e^t-1))}{\exp((1-\delta)\mu_L\cdot t)} = \left(\frac{\exp\left(e^t-1\right)}{\exp\left((1-\delta)\cdot t\right)}\right)^{\mu_L},$$

the right hand side is minimized by

$$\min_{t<0}\left(\frac{\exp(e^t-1)}{\exp((1-\delta)\cdot t)}\right)^{\mu_L} = \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_L}.$$

Moreover, for lower tail bound,

$$\Pr\left[X \le (1-\delta)\mu_L\right] \le \Pr\left[X \le (1-\delta)\mu\right] \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \le \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_L}.$$

The first inequality holds from $\mu_L \le \mu$, the second one holds from Chernoff's bound, and the third one holds from the fact that for $0 < \delta < 1$,

$$\frac{e^{-\delta}}{(1-\delta)^{1-\delta}} \le \exp\left(-\frac{\delta^2}{2}\right) \le 1.$$

This chain of inequality comes in handy when we cannot compute the exact expectation, yet we can still derive a lower bound for the expectation, and give a looser version of tail bound. The sandwiched inequality chain can even be applied with previous weighted version of Chernoff bound with weights $\{a_i\}_{i\in[1,n]}$ in $(0,1)$. □

1.4.1 Hoeffding's Bound

Hoeffding's bound says that a sum of independent bounded random variables concentrates around the sum of their expectations. The probability of being far deacays exponentially fast in the squared deviation, namely $\exp\left(-O(\varepsilon^2)\right)$.

Lemma 1.4.8 (Hoeffding's Lemma [Hoe63]). *Let X be a random variable with $X \in [a,b]$ and $\mathbb{E}[X] = 0$, then for any λ,*

$$\mathbb{E}\left[\exp(\lambda X)\right] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

Proof. Since $f(x) = \exp(\lambda x)$ is convex, then

$$f(\alpha a + (1-\alpha)b) \leq \alpha f(a) + (1-\alpha) f(b).$$

For $x \in [a,b]$, let $x = \alpha a + (1-\alpha)b$, then $\alpha = (b-x)/(b-a)$.

Since $\mathbb{E}[X] = 0$, then by linearity of expectation,

$$\mathbb{E}\left[\exp(\lambda X)\right] \leq \mathbb{E}\left[\frac{b-X}{b-a}\exp(\lambda a) + \frac{X-a}{b-a}\exp(\lambda b)\right] = \frac{b}{b-a}\exp(\lambda a) - \frac{a}{b-a}\exp(\lambda b).$$

Let $\theta = -a/(b-a) \in [0,1]$, and

$$\varphi(t) = -\theta t + \ln\left(1 - \theta + \theta\exp(t)\right),$$

then when $t = \lambda(b-a)$,

$$\begin{aligned}
\exp\left(\varphi\left(\lambda\left(b-a\right)\right)\right) &= \exp\left(-\theta\lambda\left(b-a\right)\right)\cdot\left(1-\theta+\theta\exp\left(\lambda(b-a)\right)\right) \\
&= (1-\theta)\cdot\exp\left(-\theta\lambda(b-a)\right) + \theta\cdot\exp\left((1-\theta)\lambda(b-a)\right) \\
&= (1-\theta)\cdot\exp(\lambda a) + \theta\cdot\exp(\lambda b) \\
&= \frac{b}{b-a}\exp(\lambda a) - \frac{a}{b-a}\exp(\lambda b).
\end{aligned}$$

Since $\varphi(0) = 0$,

$$\varphi'(t) = -\theta + \frac{\theta\exp(t)}{1-\theta+\theta\exp(t)}$$

means $\varphi'(0) = 0$, and

$$\varphi''(t) = \frac{\theta\exp(t)}{1-\theta+\theta\exp(t)} - \left(\frac{\theta\exp(t)}{1-\theta+\theta\exp(t)}\right)^2 \leq \frac{1}{4}.$$

By the Taylor expansion and the Taylor remainder theorem in Lagrange form, there is t' between 0 and t such that

$$\varphi(t) = \varphi(0) + \varphi'(0)\cdot t + \frac{\varphi''(t')}{2}\cdot t^2 \leq \frac{t^2}{8}.$$

Therefore,

$$\mathbb{E}\left[\exp(\lambda X)\right] \leq \exp\left(\varphi\left(\lambda(b-a)\right)\right) \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

□

Theorem 1.4.9 (Hoeffding's Bound [Hoe63]). *Let $X_1, \ldots, X_n$ be independent random variables such that for all $1 \leq i \leq n$, $\mathbb{E}[X_i] = \mu$ and $X_i \in [a,b]$. Then*

$$\Pr\left[\left|\frac{1}{n}\sum_{i\in[1,n]} X_i - \mu\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right).$$

Proof. Let $Z_i = (X_i - \mu)/n$, then $\mathbb{E}[Z_i] = 0$ and $Z_i \in [(a-\mu)/n, (b-\mu)/n]$ for $i \in [1,n]$. Let $Z = \sum_{i\in[1,n]} Z_i$. For $\lambda \geq 0$,

$$\begin{aligned}
\Pr[Z \geq \varepsilon] &= \Pr\left[\exp(\lambda Z) \geq \exp(\lambda\varepsilon)\right] \\
&\leq \frac{1}{\exp(\lambda\varepsilon)} \cdot \mathbb{E}\left[\exp(\lambda Z)\right] = \frac{1}{\exp(\lambda\varepsilon)} \cdot \prod_{i\in[1,n]} \mathbb{E}\left[\exp(\lambda Z_i)\right] \\
&\leq \frac{1}{\exp(\lambda\varepsilon)} \cdot \left(\exp\left(\frac{\lambda^2(b-a)^2}{8n^2}\right)\right)^n = \exp\left(\frac{(b-a)^2}{8n} \cdot \lambda^2 - \varepsilon \cdot \lambda\right) \\
&\leq \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right),
\end{aligned}$$

where the second inequality is from Markov inequality over MGF, the third equality is from $X_1, \ldots$ being independent, the fourth inequality is from lemma 1.4.8, and the last inequality is by minimizing at $\lambda = 4n\varepsilon/(b-a)^2$.

The lower bound argument is similar, for $\lambda < 0$,

$$\begin{aligned}
\Pr[Z \leq -\varepsilon] &= \Pr\left[\exp(\lambda Z) \geq \exp(-\lambda\varepsilon)\right] \\
&\leq \frac{1}{\exp(-\lambda\varepsilon)} \cdot \mathbb{E}\left[\exp(\lambda Z)\right] = \exp(\lambda\varepsilon) \cdot \prod_{i\in[1,n]} \mathbb{E}\left[\exp(\lambda Z_i)\right] \\
&\leq \exp\left(\frac{(b-a)^2}{8n} \cdot \lambda^2 + \varepsilon \cdot \lambda\right) \leq \exp\left(-\frac{2n\varepsilon^2}{(b-a)^2}\right),
\end{aligned}$$

where the last inequality is by minimizing at $\lambda = -4n\varepsilon/(b-a)^2$. □

Theorem 1.4.10 (Generalized Hoeffding's Bound [Hoe63]). *Let $X_1, \ldots, X_n$ be independent random variables such that for all $1 \leq i \leq n$, $\mathbb{E}[X_i] = \mu_i$ and $X_i \in [a_i, b_i]$, then*

$$\Pr\left[\left|\sum_{i\in[1,n]} X_i - \sum_{i\in[1,n]} \mu_i\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i\in[1,n]}(b_i - a_i)^2}\right).$$

Proof. Let $Z_i = X_i - \mu_i$, then $\mathbb{E}[Z_i] = 0$ and $Z_i \in [a_i - \mu_i, b_i - \mu_i]$ for $i \in [1,n]$. Let $Z = \sum_{i\in[1,n]} Z_i$. For $\lambda \geq 0$,

$$\begin{aligned}
\Pr[Z \geq \varepsilon] &= \Pr\left[\exp(\lambda Z) \geq \exp(\lambda\varepsilon)\right] \\
&\leq \frac{1}{\exp(\lambda\varepsilon)} \cdot \mathbb{E}\left[\exp(\lambda Z)\right] = \frac{1}{\exp(\lambda\varepsilon)} \cdot \prod_{i\in[1,n]} \mathbb{E}\left[\exp(\lambda Z_i)\right] \\
&\leq \frac{1}{\exp(\lambda\varepsilon)} \cdot \prod_{i\in[1,n]} \exp\left(\frac{\lambda^2(b_i-a_i)^2}{8}\right) = \exp\left(\sum_{i\in[1,n]} \frac{(b_i-a_i)^2}{8} \cdot \lambda^2 - \varepsilon \cdot \lambda\right) \\
&\leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i\in[1,n]}(b_i-a_i)^2}\right),
\end{aligned}$$

where the last inequality is by minimizing at $4\varepsilon/\sum_{i\in[1,n]}(b_i - a_i)^2$.

The lower bound argument is similar, for $\lambda < 0$,

$$\begin{aligned}
\Pr[Z \leq -\varepsilon] &= \Pr\left[\exp(\lambda Z) \geq \exp(-\lambda\varepsilon)\right] \\
&\leq \frac{1}{\exp(-\lambda\varepsilon)} \cdot \mathbb{E}\left[\exp(\lambda Z)\right] = \exp(\lambda\varepsilon) \cdot \prod_{i\in[1,n]} \mathbb{E}\left[\exp(\lambda Z_i)\right] \\
&\leq \exp(\lambda\varepsilon) \cdot \prod_{i\in[1,n]} \exp\left(\frac{\lambda^2(b_i-a_i)^2}{8}\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i\in[1,n]}(b_i-a_i)^2}\right),
\end{aligned}$$

where the last inequality is by minimizing at $-4\varepsilon/\sum_{i\in[1,n]}(b_i - a_i)^2$. □

1.4.2 Quicksort Runtime Analysis

Exercise 1.4.11 (Exercise 4.21 [MU17])**.** We prove that the Randomized Quicksort algorithm sorts a set of n numbers in time $O(n \log n)$ with high probability. Consider the following view of Randomized Quicksort. Every point in the algorithm where it decideds on a pivot element is called a *node*. Suppose the size of the set to be sorted at a particular node is s. The node is called *good* if the pivot element divides the set into two parts, each of size not exceeds $2s/3$. Otherwise the node is called *bad*. The nodes can be thought of as forming a tree in which the root node has the whole set to be sorted and its children have the two sets formed after the first pivot step and so on.

- Show that the number of good nodes in any path from the root to a leaf in this tree is not greater than $c \log_2 n$, where c is some positive constant.
- Show that, w.h.p. (greater than $1-1/n^2$), the number of nodes in a given root to leaf path of the tree is not greater than $c' \log_2 n$, where c' is another constant.
- Show that, w.h.p. (greater than $1 - 1/n$), the number of nodes in the longest root to leaf path is not greater than $c' \log_2 n$.
- Show that the running time of Quicksort is $O(n \log n)$ with probability at least $1 - 1/n$.

Proof. Suppose a given node with all the internal nodes being good nodes, then the longest such path gives the upper bound on the number of good nodes in any path of the tree. This can be proven by contradiction.

Suppose the path with the largest number of good nodes is not with all good nodes, looking like $\mathbf{u} = \mathbf{v}_0 \parallel c \parallel \mathbf{v}_1$, where c is a not good node. Then the suffix path $c \parallel \mathbf{v}_1$ has to be the path with the largest number of good nodes among the subtree with root c; otherwise there exists a path $\mathbf{s}$ in subtree with root c with more good nodes, then $\mathbf{u}$ has less good nodes than $\mathbf{v}_0 \parallel \mathbf{s}$, contradicting with $\mathbf{u}$ being the path with the most good nodes among the tree.

We assume $c \parallel \mathbf{v}_1$ being the path with the largest number of good nodes among the subtree rooted by c, then the largest number of good nodes under subtree rooted by c is $|\mathbf{v}_1|$, so is the subtree rooted by $v_{1,0}$. Since the subtree rooted by c as strictly more nodes than the subtree rooted by $v_{1,0}$, then the largest number of good nodes of subtree rooted by c is lower bounded by $|\mathbf{v}_1|$. Thus we complete the proof that the largest number of good nodes from a tree is from the longest path with all nodes being good nodes.

The longest path full of good nodes has length $\log_{\frac{3}{2}} n = \log_2 n \cdot (\log_2 3 - 1)^{-1} = c \log_2 n$, by partition on boundary of good node condition, and choose the big piece side.

One observation: a path has at most $c \log_2 n$ good nodes.

Let Y be the length of a path, and we want to upper bound $\Pr[Y \geq c' \log_2 n]$. By prior observation, the path can have at most $c \log_2 n$ good nodes, and we can turn the bound on path length into the bound on the number of good nodes. If the path survives ℓ levels, then among those first ℓ levels we have seen at most $c \log_2 n$ good nodes (since too many of good nodes stops a path from growing). Therefore $\Pr[Y \geq c' \log_2 n] \leq \Pr[X \leq c \log_2 n]$, where $\{X_i\}_{i \in [1, c' \log_2 n]}$ are independent Bernoulli trials with parameter $1/3$, and $X = \sum_{i \in [1, c' \log_2 n]} X_i$.

We want $c < c'/3$ for lower tail bounding, and c' sufficiently large. By Chernoff's bound,

$$\begin{aligned} \Pr\left[Y \geq c' \log_2 n\right] &\leq \Pr\left[X \leq c \log_2 n\right] \\ &\leq \exp\left(-\frac{1}{2} \cdot \mu \cdot \delta^2\right) = n^{-\Theta(1)}, \end{aligned}$$

where $\mu = \frac{1}{3} c' \log_2 n$ and $\delta = 1 - \frac{3c}{c'}$.

For bounding the longest path is no longer than $c' \log_2 n$, we union bound against all leaves (at most n), that the probability of existing a path longer than $c' \log_2 n$ is at most $n \cdot 1/n^2$, hence the path is at most $c' \log_2 n$ long with probability $1 - 1/n^2 \cdot n = 1 - 1/n$.

Each row of partition is dominated by runtime in $O(n)$, and the tree height is at most $c' \log_2 n$ with probability at least $1 - 1/n$, hence that the runtime of the algorithm is $O(n \log n)$ with probability greater than $1 - 1/n$. □

1.4.3 Network Routing Problems

Exercise 1.4.12 (Exercise 4.22 [MU17])**.** Consider the bit-fixing routing algorithm for routing a permutation on the n-cube. Suppose that n is even. Write each source node as $a_s \parallel b_s$, with a_s and b_s of length $n/2$. Let the destination of s's packet be $b_s \parallel a_s$. Show that this permutation causes the bit-fixing routing algorithm to take $\Omega(\sqrt{N})$ steps.

Proof. Since the permutation is $a_s \parallel b_s \mapsto b_s \parallel a_s$, we can view the algorithm as 2 phases by fixing a_s and b_s. WLOG we analyze the property from phase 1, as the 2 phases are symmetric.

We write address format $a_s \parallel b_s$ in $a_{s,0} \parallel a_{s,1} \parallel b_{s,0} \parallel b_{s,1}$, where $a_s = a_{s,0} \parallel a_{s,1}$ and $b_s = b_{s,0} \parallel b_{s,1}$. During the bit fixing, the packet from $a_s \parallel b_s$ traverses to $b_{s,0} \parallel a_{s,1} \parallel b_{s,0} \parallel b_{s,1}$. Eventually, when all a_s bits are fixed to b_s, the packet arrives $b_s \parallel b_s$.

We notice that there are up to $\sqrt{N}$ such intermediate addresses $b_s \parallel b_s$, yet in the n-cube there are $N = 2^n$ packets. Thus, nodes with addresses $b_s \parallel b_s$ will be traversed by $\sqrt{N}$ packets in total.

Another interesting observation is: the packet routing in such permutation is not "load-balancing".

View each bit being fixed as a packet traversing from one half-hypercube to another half-hypercube. By prior discussion, we noticed that the first phase bit fixing is "localized", namely all $\sqrt{N}$ packets from $* \parallel b_s$ go to $b_s \parallel b_s$, and it takes no other packets. To be more generalized, all packets from $* \parallel a_{s,1} \parallel b_{s,0} \parallel b_{s,1}$ go to $b_{s,0} \parallel a_{s,1} \parallel b_{s,0} \parallel b_{s,1}$. The earlier the bit being fixed, the less packets traversing through half-hypercubes. With more bits fixed in the front ($b_{s,0}$) and less bits flexible in the back ($a_{s,1}$), more packets aggregated from sub-hypercubes are forced to take 1 path to merge sub-hypercubes (Think of 8 packets over 3 bit hypercube, the $110 - 111$ edge will take 4 packages). Therefore, right before n^{th} bit being fixed, where $n = |a_s|$, there are $\sqrt{N}/2$ packets aggregated for 1 bit fix to $b_s \parallel b_s$, that the edge will be used by $\sqrt{N}/2$ packets, and thus the edge has an $\Omega(\sqrt{N})$ delay.

We conclude that the deterministic bit-fixing algorithm takes $\Omega(\sqrt{N})$ steps to route in this permutation. □

Exercise 1.4.13 (Exercise 4.23 [MU17])**.** Consider the following modification to the bit-fixing routing algorithm for routing a permutation on the n-cube. Instead of fixing the bits in order from 1 to n, each packet chooses a random order (independent of other packet's choices) and fixes the bits in that order. Show that there is a permutation for which this algorithm requires $2^{\Omega(n)}$ steps with high probability.

Proof. We use the previous permutation $a_s \parallel b_s \mapsto b_s \parallel a_s$, and $2\sqrt{N}$ special cases, $a_s \parallel 0$ and $0 \parallel b_s$, are interesting. For these permutations, we would like to lower bound the number of packets that traverse $0 \parallel 0$.

To reach $0 \parallel 0$, the non-zero side should be fixed to zero first, before any bit on zero side being fliped to non-zero. WLOG we discuss the case that non-zero on the left side, the non-zero on right side follows from symmetry. We introduce 0-1 independent random variables $\{X_{i,j}\}$ where $i \in [1, \frac{n}{2}]$ and $j \in [1, \binom{n/2}{i}]$, indicating if j^{th} packet with i ones traverses $0 \parallel 0$. The routing should fix i one bits first, thus $\Pr\left[X_{i,j} = 1\right] = \binom{2i}{i}^{-1}$.

Before moving to the lower bound the packet numbers, we need to derive the expectation of $X = \sum X_{i,j}$,

$$\mathbb{E}[X] = \sum_{i \in [1, \frac{n}{2}]} \sum_{j \in [1, \binom{n/2}{i}]} \mathbb{E}\left[X_{i,j}\right] = \sum_{i \in [1, \frac{n}{2}]} \frac{\binom{n/2}{i}}{\binom{2i}{i}} \geq \sum_{i \in [1, \frac{n}{2}]} \frac{1}{4^i} \cdot \binom{n/2}{i} = \left(\frac{5}{4}\right)^{n/2} - 1 = \mu_L,$$

the relaxation on $\binom{2i}{i}$ follows

$$\binom{2i}{i} \leq \sum_{k \in [0,2i]} \binom{2i}{k} = 2^{2i} = 4^i.$$

The previous lower tail bound inequality chain comes in handy, that

$$\Pr\left[X \leq (1-\delta) \cdot \mu\right] \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu} \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu_L}.$$

Since $\mu_L = 2^{\Omega(n)}$, we conclude that the algorithm on this permutation has runtime $2^{\Omega(n)}$ time steps. □

Exercise 1.4.14 (Exercise 4.24 [MU17]). Assume we use the randomized routing algorithm for n-cube network to route a total of up to $p2^n$ packets, where each node is the source of no more than p packets, and each node is the destination of no more than p packets.

- Give a high probability bound on the runtime of the algorithm.
- Give a high probability bound on the maximum number of packets at any nodes at any step of the execution of the routing algorithm.

Proof. The analysis follows from the runtime analysis for randomized bit-fixing routing [VB81], but the number of packets increase from 2^n to $p \cdot 2^n$. There are 2 phases, by first routing $p \cdot 2^n$ packets to random intermediate nodes, then route these nodes to their destinations. WLOG we analyze the first phase's runtime, as they are symmetric.

Follow from the proof structure, we first bound the number of "active packets" over certain path, then upper bound the probability of the path with not many "active packets" yet runs slow, finally we derive a probability upper bound by union bounding all paths over the n-cube.

We recall the notion of "active packets", that a packet is "active" at a node v_{i-1} on the path. If v_{i-1} and v_i are adjacent on path, and diff by j^{th} bit, in order for a packet to be "active", its bits should be fixed before j^{th} bit when it reaches v_{i-1}. With that said, there are $p2^{j-1}$ packets sharing same suffix as v_{i-1} from j^{th} bit, while the probability of sampling a random intermediate address with prefixed $j-1$ bits matching v_{i-1} is $2^{-(j-1)}$. Thus, the expected number of "active packets" at a node on an edge is p.

Given a path of length m, the path is at most n as there are n bits to fix. Let X be the number of "active packets" on this path, then the expected number of "active packets" is upper bounded by $\mathbb{E}[X] = mp \leq np$. By Chernoff's bound, the probability of having too many "active packets" is upper bounded as follows

$$\Pr\left[X \geq 6np \geq 6mp\right] \leq 2^{-6np}.$$

By conditional probability, we derive the following upper bound by sums of (conditional) probabilities

$$\begin{aligned}\Pr[A] &= \Pr[A \mid B] \cdot \Pr[B] + \Pr[A \mid \neg B] \cdot \Pr[\neg B] \\ &\leq \Pr[B] + \Pr[A \mid \neg B].\end{aligned}$$

We upper bound the probability of the runtime of the path being at least $30np$ time steps by either "too many active packets" or "not many active packets but running slow" as follows

$$\begin{aligned}\Pr\left[T \geq 30np\right] &= \Pr\left[T \geq 30np \mid X \geq 6np\right] \cdot \Pr\left[X \geq 6np\right] + \Pr\left[T \geq 30np \mid X < 6np\right] \cdot \Pr\left[X < 6np\right] \\ &\leq \Pr\left[X \geq 6np\right] + \Pr\left[T \geq 30np \mid X < 6np\right].\end{aligned}$$

For "not many active packets but running slow" case, the runtime of the path is measuring how many packets stay on the path, turns into the time steps used to route all the packets. Only "active packets" can stay on the path, with probability $1/2$ to branch into the next node on path. An "active packet"'s stay on the path can be modeled by the number of failures in a geometrically distributed random variable, and it leaves the path on a successful trial with probability $1/2$. Upper bound on the sum of geometrically distributed random variables can be relaxed by upper bounding the sum of a serial of Bernoulli trials with parameter $1/2$ here

$$\Pr\left[T \geq 30np \mid X < 6np\right] \leq \Pr\left[\sum_{i \in [1, 36np]} Y_i < 6np\right] \leq \exp\left(-\frac{1}{2} \cdot 18np \cdot \left(\frac{2}{3}\right)^2\right) = \exp\left(-4np\right) \leq 2^{-3np-1}.$$

Therefore, the probability for the path's runtime being at least $30np$ is upper bounded by 2^{-3np}

$$\Pr\left[T \geq 30np\right] \leq 2^{-6np} + 2^{-3np-1} \leq 2^{-3np}.$$

Union bounding over 2^{2n} paths, the probability of the first phase runtime exceeding $30np$ is at most 2^{2n-3np}.

On the number of packets at any nodes at any time step, we can upper bound by the number of packets that will traverse the node. Let X_i be the number of packets arriving the node on i^{th} bit, $X = \sum_{i\in[1,n]} X_i$ be the number of packets using the node. Since $\mathbb{E}[X_i] = p$ by $p2^i$ source packets and 2^{-i} probability in random address prefix matching the node's address's prefix, by linearity of expectation, $\mathbb{E}[X] = np$.

Once we have an expectation, by Chernoff's bound we have $\Pr\left[X \geq 6np\right] \leq 2^{-6np}$ for this node at a certain step. By union bound over all nodes and time steps, the queue exceeding $6np$ has probability at most $2^{-6np} \cdot 2^n \cdot O(np)$. □

Exercise 1.4.15 (Exercise 4.26 [MU17])**.** Given a network that is an undirected graph G, where nodes represent processors and the edges between the nodes represent wires. We are also given a set of N packets to route. For each packet we are given a source node, a destination node, and the exact route that the packet should take from source to destination. In each time step, at most one packet can traverse an edge. A packet can wait at any node during any time step, and we assume unbounded queue sizes at each node.

A schedule for a set of packets specifies the timing for the movement of packets along their respective routes. That is, it specifies which packet should move and which should wait at each time step. Our goal is to produce a schedule for the packets that tries to minimize the total time and the maximum queue size needed to route all the packets to their destination.

- The dilation d is the maximum distance traveled by any packet. The congestion c is the maximum number of packets that must traverses a single edge during the entire course of the routing. Argue that the time required for any schedule should be at least $\Omega(c + d)$.
- Consider the following unconstrained schedule, where many packets may traverse an edge during a single time step. Assign each packet an integral delay x, chosen randomly, independently, and uniformly from the interval $\left[1, \lceil \alpha c/\log(Nd)\rceil\right]$, where α is a constant. A packet waits in its source node for x time steps, then it moves on to its final destination through its specified route without ever stopping. Give an upper bound on the probability that more than $O(\log(Nd))$ packets use a particular edge e at a particular time step t.
- Again using the unconstrained schedule, show that the probability that more than $O(\log(Nd))$ packets pass through any edge at any time step is at most $1/(Nd)$ for a positive constant α.
- Use the unconstrained schedule to devise a simple randomized algorithm that, with high probability, produces a schedule of length $O(c + d\log(Nd))$ using queues of size $O(\log(Nd))$ and following the constraint that at most one packet crosses an edge per time step.

Proof. The length of the schedule is the runtime of the routing algorithm, and it is lower bounded by: the maximum distance a packet can traverse in the network, and the maximum congestion that delays the package traversal. Therefore, let T denotes the time steps of the routing algorithm, and $T \geq \max(c, d)$, and we conclude that $T = \Omega(c+d)$.

Considering how many packets traverse edge e at the t^{th} time step, we introduce $\{X_i\}_{i\in[1,N]}$ independent 0-1 random variables, where $X_i = 1$ stands for i^{th} packet traverses e at t^{th} time step. Since there are at most $c \leq N$ packets traverse e, and if they traverse e at the t^{th} time step, they had to be starting by one of the $\lceil \alpha c/\log Nd\rceil$ delays, we here upper bound the expectation of $X = \sum_{i\in[1,N]} X_i$ by

$$\mathbb{E}[X] \leq c \cdot \frac{\log(Nd)}{\alpha c} = \frac{\log(Nd)}{\alpha}.$$

By exercise 1.4.7, given $\mu_H \geq \mu$,

$$\Pr\left[X \geq (1+\delta)\mu_H\right] \leq \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{\mu_H} \leq \exp(-\mu_H\delta^2/3),$$

we have $X \geq (1+\delta)\log(Nd)/\alpha$ probability upper bounded by $(Nd)^{-\delta^2/3\alpha}$.

Bounding the number of packets traverses any edge at any time step is equivalent to bounding the congestion of the schedule, which can be achieved by union bounding through all the time steps and all the edges. Each packet can go as far as d edges, and the packet arrives destination with at most $d + \lceil \alpha c / \log(Nd) \rceil$ time steps (by assumption that it does not stop once). If $\delta = 1$ and $\alpha < 1/9$, the upper bounded probability is

$$\begin{aligned} Nd \cdot \left(d + \left\lceil \frac{\alpha c}{\log Nd} \right\rceil \right) \cdot (Nd)^{-\frac{\delta^2}{3\alpha}} &\leq Nd \cdot Nd \cdot (Nd)^{-\frac{\delta^2}{3\alpha}} \\ &\leq (Nd)^2 \cdot (Nd)^{-3} = (Nd)^{-1}. \end{aligned}$$

If the tuned algorithm has schedule of length $O(d \log(Nd) + c)$ with queue size $O(\log(Nd))$, try scale the schedule by $2\log(Nd)/\alpha$. By previous result, with the assumption that multiple packets crossing an edge at a time with no stop, the probability of more than $2\log(Nd)/\alpha$ packets passing through any edge at any time step is at most $(Nd)^{-1}$. In the prior model, each time step the edge let packets crossing in a batch, and the batch is at most $2\log(Nd)/\alpha$.

Now removing the assumption of multiple packets crossing an edge at a time, by scaling $2\log(Nd)/\alpha$ to the schedule and queue of size $O(\log(Nd))$, given an edge allowing only 1 packet at a time, the batched moving can be simulated with the devised algorithm. □

Remark 1.4.16. The intuition is: we first let everyone run non-stop with random starts, count how bad the pileups get, then spread those pileups thin by stretching time proportionally, and leading to an average case non stop routing. This is the Leighton-Maggs-Rao trick [LMR94].

1.5 Balls, Bins, and Random Graphs

1.5.1 A few things I should have noted about e

We begin with the definition of e and e^{-1}

$$\lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = e, \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n = \frac{1}{e}.$$

Both $f(n) = (1+1/n)^n$ and $g(n) = (1-1/n)^n$ are monotone, proven as follows:

- Let $\varphi_f(n) = n\ln(1+1/n)$, then the first order derivative

  $$\varphi_f'(n) = \ln\left(1+\frac{1}{n}\right) + n\cdot\frac{1}{1+1/n}\cdot\left(-\frac{1}{n^2}\right) = \ln\left(1+\frac{1}{n}\right) - \frac{1}{1+n}.$$

  Let $t = 1+1/n > 1$, then

  $$\varphi_f'(n) = \ln\left(1+\frac{1}{n}\right) - \frac{1}{1+n} = \ln\left(1+\frac{1}{n}\right) + \frac{n}{1+n} - 1 = \ln t + \frac{1}{t} - 1.$$

  As $t > 1$ and by integral,

  $$\ln t = \int_1^t \frac{1}{x}dx > (t-1)\cdot\frac{1}{t},$$

  we have $\varphi_f'(n) > 0$, and thus $\varphi_f(n)$ is monotone increasing.
- Similarly, let $\varphi_g(n) = n\ln(1-1/n)$, then the first order derivative

  $$\varphi_g'(n) = \ln\left(1-\frac{1}{n}\right) + n\cdot\frac{1}{1-1/n}\cdot\left(-\frac{-1}{n^2}\right) = \ln\left(1-\frac{1}{n}\right) + \frac{1}{n-1}.$$

  Let $t = 1-1/n$ that $0 < t < 1$, then

  $$\varphi_g'(n) = \ln\left(1-\frac{1}{n}\right) + \frac{1}{n-1} = \ln\left(1-\frac{1}{n}\right) + \frac{n}{n-1} - 1 = \ln t + \frac{1}{t} - 1.$$

  As $0 < t < 1$ and by integral again,

  $$\ln t = \int_1^t \frac{1}{x}dx = -\int_t^1 \frac{1}{x}dx > -(1-t)\cdot\frac{1}{t},$$

  we have $\varphi_g'(n) > 0$, and thus $\varphi_g(n)$ is monotone increasing.

Moreover, by the Taylor expansion

$$e^x = \sum_{i\geq 0}\frac{x^i}{i!},$$

then $e^x \geq 1+x$ for all x. We derive following handy inequality from the Taylor expansion.

Lemma 1.5.1. *For integer $x \geq 0$ and $n \geq 0$, we have*

$$\frac{1}{x!} \leq \frac{e^n}{n^x}.$$

Proof. By Taylor expansion, since

$$\frac{n^x}{x!} \leq \sum_{m\geq 0}\frac{n^m}{m!} = e^n,$$

therefore

$$\frac{1}{x!} \leq \frac{e^n}{n^x}.$$

□

Exercise 1.5.2 (Exercise 5.7 [MU17])**.** Use the Taylor expansion

$$\ln(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots$$

to prove that, for any $|x| \le 1$, $e^x(1-x^2) \le 1 + x \le e^x$.

Proof. We prove the left side $e^x(1-x^2) \le 1+x$ first. Take ln on both sides, we have left side $x + \ln(1+x) + \ln(1-x)$.

Now by the Taylor expansion, for $|x| \le 1$, we have

$$\begin{aligned} x + \ln(1+x) + \ln(1-x) &= x + x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} \cdots - x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} \cdots \\ &= x - 2 \cdot \frac{x^2}{2} - 2 \cdot \frac{x^4}{4} \cdots \\ &\le x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots = \ln(1+x). \end{aligned}$$

The right side holds from the Taylor expansion of e^x, and we conclude that for $|x| \le 1$, $e^x(1-x^2) \le 1+x \le e^x$. □

Lemma 1.5.3. *For any integer $k \ge 1$, by concavity of $\ln x$,*

$$\frac{1}{2}(\ln k + \ln(k+1)) \le \int_k^{k+1} \ln x dx.$$

Lemma 1.5.4 (Lemma 5.8 [MU17])**.**

$$n! \le e\sqrt{n}\left(\frac{n}{e}\right)^n.$$

Proof. Since $\ln n! = \sum_{i\in[1,n]} \ln i$, and we have the following from lemma 1.5.3

$$\sum_{i\in[1,n]} \ln i \le \int_1^n \ln x dx + \frac{1}{2}\ln n = n\ln n - n + 1 + \frac{1}{2}\ln n,$$

which completes the proof by exponenting over e on both sides. □

Corollary 1.5.5. *By lemma 1.5.1 and lemma 1.5.4, we have*

$$\left(\frac{n}{e}\right)^n \le n! \le e\sqrt{n}\cdot\left(\frac{n}{e}\right)^n \text{ and } \frac{1}{e\sqrt{n}}\cdot\left(\frac{e}{n}\right)^n \le \frac{1}{n!} \le \left(\frac{e}{n}\right)^n.$$

1.5.2 Birthday Paradox and Balls-and-Bins model

Suppose there are n possible birthdays, and there are m people in a room. Then, the probability that all m are having different birthdays is

$$\prod_{j\in[0,m-1]}\left(1-\frac{j}{n}\right) \le \prod_{j\in[0,m-1]} \exp\left(-\frac{j}{n}\right) \le \exp\left(-\frac{m^2}{2n}\right).$$

Hence, when $m = \sqrt{2n\ln 2}$, the probability goes to $1/2$.

But through union bound, we can also give a loose bound as follows. Consider event E_k being the k^{th} people failed to have distinct birthday from all previous $k-1$ people. Then $\Pr[E_k] = (k-1)/n$. Through union bound, the probability of the first k people failed to have distinct birthdays is upper bounded by

$$\Pr\left[\bigcup_{i\in[1,k]} E_i\right] \le \sum_{i\in[1,k]} \Pr[E_i] = \frac{k(k-1)}{2n}.$$

The upper bound indicates that when $k = \lceil \sqrt{n} \rceil$, the probability of having distinct birthdays is at most $1/2$.

Moreover, assuming that first $\lceil \sqrt{n} \rceil$ birthdays are distinct, then the next $\lceil \sqrt{n} \rceil$ birthdays are different from the first $\lceil \sqrt{n} \rceil$ birthdays with probability upper bounded by

$$\left(1 - \frac{\lceil \sqrt{n} \rceil}{n}\right)^{\lceil \sqrt{n} \rceil} \leq \frac{1}{e}.$$

Therefore, once there are $2\lceil \sqrt{n} \rceil$ people out there, the probability of having all distinct birthdays is at most $1/e < 1/2$.

Remark 1.5.6. The idea of union bound is used throughout the Poisson approximation and balls-and-bins model later on, after introduction of theorem 1.5.15.

Exercise 1.5.7 (Exercise 5.4 [MU17])**.** Consider the extended birthday paradox, but with no 3 having same birthdays.

Proof. Let there be k people, and there be n days in a year.

Suppose there are j pairs of same birthdays, then we first consider the number of pair choices out of k people. We can choose 2 people continuously from k people, then rule out all the duplicate choices, as we don't care about the ordering, and thus we have number of choices in total

$$\binom{k}{2}\binom{k-2}{2}\cdots\binom{k-2j+2}{2} \cdot \frac{1}{j!} = \frac{k!}{2^j(k-2j)!} \cdot \frac{1}{j!} = \binom{k}{2;\ldots;2} \cdot \frac{1}{j!}.$$

For each choice, there are $k - j$ days to be distinct, the rest j people should have same birthdays as their pairs, leading to probability

$$\frac{1}{n^j} \cdot \prod_{i \in [1, k-j]} \left(1 - \frac{i-1}{n}\right).$$

Now that the probability p_j with j pairs of same birthdays is

$$p_j = \frac{k!}{2^j(k-2j)!} \cdot \frac{1}{j!} \cdot \frac{1}{n^j} \cdot \prod_{i \in [1, k-j]} \left(1 - \frac{i-1}{n}\right),$$

summing over p_j over $j \in \left[0, \left\lfloor \frac{k}{2} \right\rfloor\right]$ is the exact probability. □

Exercise 1.5.8 (Exercise 5.8 [MU17])**.** n balls are thrown independently and uniformly at random into n bins.

- Find the conditional probability that bin 1 has one ball given that exactly one ball fell into the first three bins.
- Find the conditional expectation of the number of balls in bin 1 under the condition that bin 2 received no balls.
- Write an expression for the probability that bin 1 receives more balls than bin 2.

Proof. If conditioned that there are only 1 ball in first 3 bins, then the probability that a ball is in bin 1 is $1/3$.

We introduce independent Bernoulli trials $\{X_i\}_{i \in [1,n]}$ with probability $1/n$ indicating if the i^{th} ball falls in bin 1. Then the conditional probability of $X_i = 1$ given that the i^{th} ball does not fall into bin 2 is

$$\Pr\left[X_i = 1 \mid i^{\text{th}} \text{ ball does not fall in bin 2}\right] = \frac{1}{n-1}.$$

Therefore, the conditional expectation of $X = \sum_{i \in [1,n]} X_i$ under the condition that bin 2 has no balls is

$$\begin{aligned}\mathbb{E}\left[X \mid \text{bin 2 has no balls}\right] &= \sum_{i \in [1,n]} \mathbb{E}\left[X_i \mid \text{bin 2 has no balls}\right] \\ &= \sum_{i \in [1,n]} 1 \cdot \Pr\left[X_i = 1 \mid i^{\text{th}} \text{ ball does not fall in bin 2}\right] = \frac{n}{n-1}.\end{aligned}$$

There are only 3 cases related to balls in bin 1 and bin 2, one of more, less, or equal. By symmetry, former 2 cases are equal. Therefore, we need only the equal case to derive the exact result.

We choose $k \in \left[0, \left\lfloor \frac{n}{2} \right\rfloor\right]$ balls each for bin 1 and bin 2, and the probability is given by

$$p_{e,k} = \binom{n}{k;k} \left(\frac{1}{n}\right)^k \left(\frac{1}{n}\right)^k \left(1 - \frac{2}{n}\right)^{n-2k} = \binom{n}{k;k} \left(\frac{1}{n}\right)^{2k} \left(1 - \frac{2}{n}\right)^{n-2k}.$$

Therefore, the exact probability for the equal case is $p_e = \sum_k p_{e,k}$, and the desired probability is $(1 - p_e)/2$. □

Lemma 1.5.9 (Lemma 5.1 [MU17])**.** *When n balls are thrown independently and uniformly at random into n bins, the probability that the maximum load is more than $3 \ln n / \ln \ln n$ is at most $1/n$ for n sufficiently large.*

Proof. Suppose for the i^{th} bin receives at least M balls, let it be event E_i, then the probability is

$$\Pr[E_i] = \binom{n}{M} \frac{1}{n^M}.$$

If any bins are receiving with at least M balls, by union bound, the probability is upper bounded by

$$\Pr\left[\bigcup_{i \in [1,n]} E_i\right] \leq \sum_{i \in [1,n]} \Pr[E_i] = n \binom{n}{M} \frac{1}{n^M}.$$

Since

$$\binom{n}{M} \frac{1}{n^M} \leq \frac{1}{M!} \leq \left(\frac{e}{M}\right)^M,$$

where second inequality follows from lemma 1.5.1, the probability is then upper bounded by $n\,(e/M)^M$.

Choose $M = 3 \ln n / \ln \ln n$, then when n is sufficiently large, the probability upper bound is at most $1/n$. □

Exercise 1.5.10 (Exercise 5.12 [MU17])**.** The following problem models a simple distributed system wherein agents contend for resources but "back off" in the face of contention. Balls represent agents, and bins represent resources.

The system evolves over rounds. Every round, balls are thrown independently and uniformly at random into n bins. Any ball that lands in a bin by itself is served and removed from consideration. The remaining balls are thrown again in the next round. We begin with n balls in the first round, and we wish when every ball is served.

- If there are b balls at the start of a round, what is the expected number of balls at the start of the next round?
- Suppose that every round the number of balls served was exactly the expected number of balls to be served. Show that all the balls would be served in $O(\log \log n)$ rounds. (Hint: If x_j is the expected number of balls left after j rounds, show and use that $x_{j+1} \leq x_j^2/n$.)

Proof. We first consider the expected number of remaining balls. Only the bin with one and only one ball can serve the ball, and thus we can disregard the ball onwards. We introduce n independent 0-1 random variables $\{X_i\}_{i \in [1,n]}$ for i^{th} bin has one and only one ball to serve. Therefore, the probability for a bin with only one ball is

$$\Pr[X_i = 1] = \binom{b}{1} \frac{1}{n} \left(1 - \frac{1}{n}\right)^{b-1} = \frac{b}{n} \cdot \left(1 - \frac{1}{n}\right)^{b-1},$$

where b is the number of balls remaining by now.

Let $X = \sum_{i \in [1,n]} X_i$ be the number of bins can serve (or balls to disregard), by linearity of expectation, we have

$$\mathbb{E}[X] = n \cdot \frac{b}{n} \cdot \left(1 - \frac{1}{n}\right)^{b-1} = b \cdot \left(1 - \frac{1}{n}\right)^{b-1}.$$

Therefore, the remaining balls next round is $b - b \cdot (1 - 1/n)^{b-1}$.

Let b be the remaining balls to serve after round j, then $x_{j+1} = b - b \cdot (1 - 1/n)^{b-1}$. Since

$$\left(1 - \frac{1}{n}\right)^{b-1} \geq 1 - \frac{b-1}{n},$$

we have $x_{j+1} \leq b(b-1)/n$, and therefore $x_{j+1} \leq x_j^2/n$.

Now we write $y_j = x_j/n$, and therefore $y_{j+1} \leq y_j^2$, and we want to know the number of rounds to reach a $y = 1/n$. Suppose $y_j = 1/2$ for some constant number of rounds j, we count the number of additional rounds r until we have

$$y_j^{2^r} = \left(\frac{1}{2}\right)^{2^r} \leq \frac{1}{n}.$$

Therefore $r = O(\log \log n)$, that all balls will be served in $O(\log \log n)$ rounds. □

1.5.3 Poisson Distribution

Exercise 1.5.11 (Exercise 5.6 [MU17])**.** Let X be a Poisson random variable with mean μ, representing the number of a page of this book. Each error is independently a grammatical error with probability p and a spelling error with probability $1 - p$. If Y and Z are random variables representing the number of grammatical and spelling error (respectively) on a page of this book, prove that Y and Z are Poisson random variables with means μp and $\mu(1-p)$, respectively. Also, prove that Y and Z are independent.

Proof. The sum of a finite number of independent Poisson random variables is a Poisson random variable, by looking at the MGFs. Now think in a reverse way, that a Poisson random variable can be thinned down to a finite number of independent Poisson random variables, which is the Poisson Thinning (Splitting) Property.

Let a Poisson variable X with mean μ represent the errors of a page, then Y be the random variable represent the grammatical errors with probability p in these errors. The probability of $Y = k$ is the summation of all conditional probability conditioned over X by

$$\begin{aligned}
\Pr[Y = k] &= \sum_{x \geq k} \Pr[Y = k \cap X = x] \\
&= \sum_{x \geq k} \Pr[Y = k \mid X = x] \cdot \Pr[X = x] \\
&= \sum_{x \geq k} \binom{x}{k} p^k (1-p)^{x-k} \cdot \frac{e^{-\mu}\mu^x}{x!} = \frac{e^{-\mu}p^k\mu^k}{k!} \sum_{x \geq k} (1-p)^{x-k} \cdot \frac{\mu^{x-k}}{(x-k)!} \\
&= \frac{e^{-\mu}p^k\mu^k}{k!} \cdot e^{(1-p)\mu} = \frac{e^{-p\mu}(p\mu)^k}{k!},
\end{aligned}$$

which is the probability of a Poisson random variable with mean $p\mu$ being k. By symmetry, Z is a Poisson random variable with mean $(1-p)\mu$ by substituting p with $1 - p$.

For Y and Z being independent, we prove by the property of independence as follows:

$$\begin{aligned}
\Pr\left[Y = y \cap Z = z\right] &= \Pr\left[Y = y \cap Z = z \mid X = y + z\right] \cdot \Pr\left[X = y + z\right] \\
&= \binom{y+z}{y} p^y (1-p)^z \cdot \frac{e^{-\mu}\mu^{y+z}}{(y+z)!} = p^y(1-p)^z \cdot \frac{e^{-\mu}\mu^{y+z}}{y!z!} \\
&= \frac{e^{-p\mu}p^y\mu^y}{y!} \cdot \frac{e^{-(1-p)\mu}(1-p)^z\mu^z}{z!} \\
&= \Pr\left[Y = y\right] \cdot \Pr[Z = z].
\end{aligned}$$

□

Lemma 1.5.12 (Exercise 5.14 [MU17]). *If Z is a Poisson random variable of mean $\mu \geq 1$ integer, then $\Pr\left[Z \geq \mu\right] \geq 1/2$.*

- *Show that $\Pr\left[Z = \mu + h\right] \geq \Pr\left[Z = \mu - h - 1\right]$ for $0 \leq h \leq \mu - 1$.*
- *Argue that $\Pr\left[Z \geq \mu\right] \geq 1/2$.*

Proof. We write $\Pr\left[Z = \mu + h\right] = \delta_h \cdot \Pr\left[Z = \mu - h - 1\right]$, where $h \in \left[0, \mu - 1\right]$. Therefore,

$$\delta_h = \prod_{j \in [\mu - h, \mu - 1]} \frac{\mu}{j} \cdot \prod_{j \in [\mu + 1, \mu + h]} \frac{\mu}{j} = \prod_{k \in [1,h]} \frac{\mu^2}{(\mu - k)(\mu + k)} = \prod_{k \in [1,h]} \alpha_k.$$

As $\alpha_k \geq 1$ for $k \in [1, h]$, therefore $\delta_h \geq 1$. Thus $\Pr\left[Z \leq \mu - 1\right] \leq \Pr\left[Z \geq \mu\right]$, and $\Pr\left[Z \geq \mu\right] \geq 1/2$.

Moreover, since $\delta_0 = 1$, therefore $\Pr\left[Z = \mu - 1\right] = \Pr\left[Z = \mu\right]$. □

Theorem 1.5.13 (Poisson Median Range Theorem [Cho94]). *Let Z be a Poisson random variable of mean $\mu > 0$, and m be the median of Z. Then*

$$\mu - \ln 2 \leq m \leq \mu + \frac{1}{3}.$$

If μ is an integer such that $\mu \geq 1$, then $\mu = m$, and thus $\Pr\left[Z \geq \mu\right] \geq 1/2$ and $\Pr\left[Z \leq \mu\right] \geq 1/2$.

1.5.4 Poisson Approximation

Theorem 1.5.14 (Theorem 5.6 [MU17]). *Let $\{X_i^{(m)}\}_{i \in [1,n]}$ be the number of balls in each bin of the m balls and n bins model, and $\{Y_i^{(m)}\}_{i \in [1,n]}$ be independent Poisson random variables of mean m/n. The distribution of $\left(X_1^{(k)}, \ldots, X_n^{(k)}\right)$ is the same as the distribution of $\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)$ conditioned on $Y^{(m)} = \sum_{i \in [1,n]} Y_i^{(m)} = k$, regardless of the value of m.*

Theorem 1.5.15 (Theorem 5.7 [MU17]). *Let $f(x_1, \ldots, x_n)$ be a nonnegative function. Then*

$$\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \leq e\sqrt{m} \cdot \mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right].$$

Proof. By conditional expectation and theorem 1.5.14, we have

$$\begin{aligned}
\mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right] &= \sum_{k \geq 0} \mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right) \,\middle|\, Y^{(m)} = k\right] \Pr\left[Y^{(m)} = k\right] \\
&= \sum_{k \geq 0} \mathbb{E}\left[f\left(X_1^{(k)}, \ldots, X_n^{(k)}\right)\right] \Pr\left[Y^{(m)} = k\right] \\
&\geq \mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \Pr\left[Y^{(m)} = m\right].
\end{aligned}$$

By definition of the Poisson random variable, $Y^{(m)}$ is a Poisson random variable of mean $\mu = m$, thus

$$\Pr\left[Y^{(m)} = m\right] = e^{-m} \cdot \frac{m^m}{m!}.$$

By lemma 1.5.4, $\Pr\left[Y^{(m)} = m\right] \geq \left(e\sqrt{m}\right)^{-1}$. Therefore, $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \leq e\sqrt{m} \cdot \mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right]$. □

Corollary 1.5.16. *Let E be an event. If E has probability p in the Poisson case, where there are n independent Poisson random variables of mean m/n, then E has probability at most $e\sqrt{m} \cdot p$ in the exact m balls and n bins case.*

Proof. Immediately from theorem 1.5.15. □

Theorem 1.5.17 (Theorem 5.10 [MU17]). *Let $f(x_1, \ldots, x_n)$ be a nonnegative function such that $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ is either monotonically increasing or monotonically decreasing in m. Then*

$$\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \leq 2\,\mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right].$$

Exercise 1.5.18 (Exercise 5.15 [MU17]). Prove that if $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ is monotonically increasing in m, then

$$\mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right] \geq \mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \Pr\left[Y^{(m)} \geq m\right],$$

again under the condition that f is nonnegative.

Make a similar statement for the case when $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ is monotonically decreasing in m.

Moreover, prove theorem 1.5.17 for the case that $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ is monotonically increasing in m.

Proof. We derive the chain of inequalities as follows

$$\begin{aligned}
\mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right] &= \sum_{k\geq 0} \mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right) \,\middle|\, Y^{(m)} = k\right] \Pr\left[Y^{(m)} = k\right] \\
&= \sum_{k\geq 0} \mathbb{E}\left[f\left(X_1^{(k)}, \ldots, X_n^{(k)}\right)\right] \Pr\left[Y^{(m)} = k\right] \\
&\geq \sum_{k\geq m} \mathbb{E}\left[f\left(X_1^{(k)}, \ldots, X_n^{(k)}\right)\right] \Pr\left[Y^{(m)} = k\right] \\
&\geq \sum_{k\geq m} \mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \Pr\left[Y^{(m)} = k\right] \\
&= \mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \Pr\left[Y^{(m)} \geq m\right].
\end{aligned}$$

The first equality holds from conditional expectation, the second equality holds from theorem 1.5.14, the third holds from nonnegativity of f, the fourth holds from $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ being monotonically increasing in m.

Similarly, if f is nonnegative and $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ is monotonically decreasing in m, the following holds

$$\mathbb{E}\left[f\left(Y_1^{(m)}, \ldots, Y_n^{(m)}\right)\right] \geq \mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right] \Pr\left[Y^{(m)} \leq m\right].$$

By lemma 1.5.12, theorem 1.5.13, and $Y^{(m)}$ being a Poisson random variable of mean $\mu = m \geq 1$ integer, therefore we prove theorem 1.5.17 for the $\mathbb{E}\left[f\left(X_1^{(m)}, \ldots, X_n^{(m)}\right)\right]$ being monotonically increasing/decreasing in m. □

Corollary 1.5.19 (Corollary 5.11 [MU17]). *Let E be an event whose probability is monotonically increasing or decreasing in the number of balls. If E has probability p in the Poisson case, then E has probability at most $2p$ in the exact case.*

Proof. Immediately follow from theorem 1.5.17. □

Lemma 1.5.20 (Lemma 5.12 [MU17]). *When n balls are thrown independently and uniformly at random into n bins, the max load is at least $\ln n / \ln \ln n$ with probability at least $1 - 1/n$ for n sufficiently large.*

Proof. In the Poisson case, there are n independent Poisson random varables $\{Z_i\}_{i\in[1,n]}$ of mean 1.

If $Z_i \geq M$ for some load M, then

$$\Pr[Z_i < M] < 1 - \Pr[Z_i = M] = 1 - \frac{1}{eM!}.$$

Considering all n Poisson random variables, then the probability of all loads being lower than M is

$$\Pr\left[\bigcap_{i\in[1,n]} Z_i < M\right] < \left(1 - \frac{1}{eM!}\right)^n \leq \exp\left(-\frac{n}{eM!}\right).$$

From corollary 1.5.16 or corollary 1.5.19, we want the probability upper bound to be less than $1/n^2$, as $e\sqrt{n}/n^2 = o(1)$. By lemma 1.5.4 and relax by sufficiently large n and M by

$$M! \leq e\sqrt{M}\left(\frac{M}{e}\right)^M \leq M\left(\frac{M}{e}\right)^M,$$

therefore let $M = \ln n / \ln\ln n$, with less than $1/n$ probability in exact case the max load is no greater than M. □

Remark 1.5.21. For n balls and n bins, the maximum load is $\Theta(\log n/\log\log n)$ from lemma 1.5.9 and lemma 1.5.20 with probability approaches 1 as $n \to \infty$.

Exercise 1.5.22 (Exercise 5.13 [MU17])**.** Suppose that we vary the balls-and-bins process as follows. For convenience let the bins be numbered from 0 to $n-1$. There are $\log_2 n$ players. Each player choose a starting position $\ell \xleftarrow{R} [0, n-1]$, then places one ball in each of the bins numbered $\ell \mod n$ to $\ell + n/\log_2 n - 1 \mod n$. Argue that the maximum load in this case is only $O(\log\log n/\log\log\log n)$ with probability that approaches 1 as $n \to \infty$.

Proof. For the j^{th} player and the i^{th} bin, we introduce an independent 0-1 random variable $X_{i,j}$, indicating if the j^{th} player puts ball in the i^{th} bin. $\Pr\left[X_{i,j} = 1\right] = 1/\log_2 n$, as the j^{th} player starts at most $n/\log_2 n$ bins before the i^{th} bin. Therefore, $X_i = \sum_{j\in[1,\log_2 n]} X_{i,j}$ is the load of the i^{th} bin, and by linearity of expectation, $\mathbb{E}[X_i] = \sum_j \mathbb{E}\left[X_{i,j}\right] = 1$. We can apply Chernoff's tail bound to upper bound the probability for $X_i \geq 1 + \delta$ by

$$\Pr[X_i \geq 1 + \delta] \leq \frac{e^\delta}{(1+\delta)^{1+\delta}}.$$

By union bound, if there exists bins with load exceeding $1 + \delta$, the probability is upper bounded by $n \cdot e^\delta/(1+\delta)^{1+\delta}$.

Let $t = 1 + \delta$ be the load bound, then the probability upper bound is $n/e \cdot (e/t)^t$. We want to have

$$\lim_{n\to\infty} \frac{n}{e} \cdot \left(\frac{e}{t}\right)^t = 0$$

to bound all of the bin loads. Taking ln on both sides, $t\ln t - t = \omega(\ln n)$, giving only $t = O(\log n/\log\log n)$.

But there are at most $\log_2 n$ players and n balls, such result follow from vanilla remark 1.5.21. The problem lies in that we did not utilize the structure of the player model, giving a loose bound.

Looking at the balls-and-bins structure, we can divide the circle into $\log_2 n$ intervals, each has length $n/\log_2 n$. Let $\{Y_i\}_{i\in[1,\log_2 n]}$ be independent random variables for number of players in the i^{th} interval, following $\log_2 n$ balls and $\log_2 n$ bins model. An observation is that, for a bin in the i^{th} interval, its load is at most $Y_i + Y_{i-1}$, relaxed by

$$Y_i + Y_{i-1} \leq 2\max(Y_i, Y_{i-1}).$$

Therefore, the max load of a bin is $O(Y_i)$.

We continue with Chernoff tail bound for Y_i, then union bound over $\log_2 n$ intervals, rather than over all n bins. Since $\mathbb{E}[Y_i] = 1$, the probability for existing $Y_i \geq 1 + \delta$ is $\log_2 n \cdot e^\delta/(1+\delta)^{1+\delta}$.

Again we write $t = 1 + \delta$ to be the number of players in an interval, and we want

$$\lim_{n\to\infty} \frac{\log_2 n}{e} \cdot \left(\frac{e}{t}\right)^t = 0,$$

and therefore $t\ln t - t = \omega(\log\log n)$. We conclude that $t = O(\log\log n/\log\log\log n)$, and the load is $O(t)$.

This $\log n$ balls and $\log n$ bins max load result matches remark 1.5.21. □

1.5.5 Recap on Coupon Collector's Problem

Theorem 1.5.23 (Theorem 5.13 [MU17]). *Let X be the number of coupons observed before obtaining one of each of n types of coupons. Then, for a constant c,*

$$\lim_{n\to\infty} \Pr[X > n\ln n + cn] = 1 - e^{-e^{-c}}.$$

Proof. We can view this problem in balls-and-bins model: If balls are thrown independently and uniformly at random into bins, how many balls are thrown until all bins have at least one ball? We begin with a Poisson approximation, and later demonstrate that the Poisson approximation gives the right answer in the limit.

We begin by Poisson approximation, supposing $\{X_i\}_{i\in[1,n]}$ independent Poisson random variable of mean $\ln n + c$, then the expected number of balls is $m = n\ln n + cn$. Let E be the event that no bin is empty, and let $X = \sum_{i\in[1,n]} X_i$ be the number of balls in bin. Since $\Pr[X_i = 0] = (ne^c)^{-1}$, then

$$\lim_{n\to\infty} \Pr[E] = \lim_{n\to\infty} \left(1 - \frac{1}{ne^c}\right)^n = e^{-e^{-c}}.$$

Let $b = \sqrt{2m\ln m}$. By conditional probability, we have

$$\begin{aligned}\Pr[E] = \Pr[E \mid |X-m| > b]\Pr[|X-m| > b] \\ + \Pr[E \mid |X-m| \le b]\Pr[|X-m| \le b].\end{aligned}$$

Lemma 1.5.24. $\Pr[|X-m| > b] = o(1)$.

Proof. We can use the Chernoff bound such that $\Pr[|X-m| > b] < 2\exp(-2\ln m/3) = 2m^{-2/3} = o(1)$. □

Lemma 1.5.25. $|\Pr[E \mid |X-m| \le b] - \Pr[E \mid X = m]| = o(1)$.

Proof. Since $\Pr[E \mid X = k]$ is monotonically increasing in k, by conditional probability

$$\begin{aligned}\Pr[E \mid |X-m| \le b] &= \sum_{k\in[-b,b]} \Pr[E \mid X = m+k]\cdot\Pr[X = m+k \mid |X-m| \le b] \\ &\le \sum_{k\in[-b,b]} \Pr[E \mid X = m+b]\cdot\Pr[X = m+k \mid |X-m| \le b] \\ &= \Pr[E \mid X = m+b].\end{aligned}$$

Therefore, we relax the term by

$$|\Pr[E \mid |X-m| \le b] - \Pr[E \mid X = m]| \le \Pr[E \mid X = m+b] - \Pr[E \mid X = m-b],$$

which is interpreted as: when $m-b$ balls are thrown, there exists empty bins; after $m+b$ balls, all bins are filled.

Let A be the event that $m-b$ balls thrown with empty bins, S be the set for all possible empty bins $B \subseteq [1,n]$ after $m-b$ initial throws, and F_B be the event that B empty bins are filled after another $2b$ balls. The exact term follows

$$\begin{aligned}\Pr[E \mid X = m+b] - \Pr[E \mid X = m-b] &= \Pr[A]\cdot\sum_{B\in S}\left(\Pr\left[\text{empty bins } B \subseteq [1,n] \mid A\right]\cdot\Pr[F_B]\right) \\ &\le \sum_{B\in S}\left(\Pr\left[\text{empty bins } B \subseteq [1,n] \mid A\right]\cdot\Pr[F_B]\right) \\ &\le \max_{B\in S}\Pr[F_B].\end{aligned}$$

Since $\Pr[F_B] \le \Pr[F_r]$ for some $r \in B$ by $F_B \subseteq F_r$, the probability for all empty bins being filled is upper bounded by the probability of one of the empty bins in B being filled, we upper bound $\Pr[F_r] \le 2b/n = o(1)$ by union bound.

Therefore, we conclude the resulting difference by $o(1)$. □

By lemma 1.5.24 and lemma 1.5.25, we conclude that

$$\begin{aligned}\lim_{n\to\infty}\Pr[E] &= \lim_{n\to\infty} o(1) + (\Pr[E \mid X = m] + o(1)) \cdot (1 - o(1)) \\ &= \lim_{n\to\infty}\Pr[E \mid X = m],\end{aligned}$$

indicating as $n \to \infty$, for m balls n bins exact case, all n bins are filled with probability $e^{-e^{-c}}$. □

1.5.6 Hashing and Random Graphs

Exercise 1.5.26 (Exercise 5.16 [MU17])**.** We consider another way to obtain Chernoff-like bounds in the settings of balls and bins without using theorem 1.5.15. Consider n balls and n bins model. Let $X_i = 1$ iff the i^{th} bin is empty, and $X = \sum_i X_i$. Let $\{Y_i\}_{i\in[1,n]}$ be independent Bernoulli random variables with probability $(1-1/n)^n$, and $Y = \sum_i Y_i$.

- Show that $\mathbb{E}\left[\prod_{i\in[1,k]} X_i\right] \le \mathbb{E}\left[\prod_{i\in[1,k]} Y_i\right]$ for any $k \ge 1$.
- Show that $\mathbb{E}\left[\exp(tX)\right] \le \mathbb{E}\left[\exp(tY)\right]$ for any $t \ge 0$.
- Derive a Chernoff bound for $\Pr[X \ge (1+\delta)\,\mathbb{E}[X]]$.

Proof. Considering the event that first k bins are empty in n balls and n bins model, the probability is $(n-k)^n/n^n$. On the other hand, since each Y_i are independent, then $\prod_{i\in[1,k]} Y_i = 1$ has probability $(1-1/n)^{nk}$.

Since $(1-1/n)^k \ge 1 - k/n$, then $\mathbb{E}\left[\prod_{i\in[1,k]} X_i\right] \le \mathbb{E}\left[\prod_{i\in[1,k]} Y_i\right]$ for any $k \ge 1$.

Expanding the prior claim, for any subsets of indices $S \subseteq [1,n]$, $\Pr[\prod_{i\in S} X_i = 1] \le \Pr[\prod_{i\in S} Y_i = 1]$. We expand MGF function in its Taylor expansion, then we derive the following

$$\begin{aligned}\mathbb{E}\left[\exp(tX)\right] &= \mathbb{E}\left[\sum_{k\ge0}\frac{(tX)^k}{k!}\right] = \sum_{k\ge0}\frac{t^k}{k!}\cdot\mathbb{E}\left[X^k\right] = \sum_{k\ge0}\frac{t^k}{k!}\cdot\mathbb{E}\left[\left(\sum_{i\in[1,n]} X_i\right)^k\right] \\ &= \sum_{k\ge0}\frac{t^k}{k!}\cdot\sum_{S\subseteq[1,n],|S|\le k}\Pr\left[\prod_{i\in S} X_i = 1\right] \\ &\le \sum_{k\ge0}\frac{t^k}{k!}\cdot\sum_{S\subseteq[1,n],|S|\le k}\Pr\left[\prod_{i\in S} Y_i = 1\right] = \mathbb{E}\left[\exp(tY)\right].\end{aligned}$$

Since each Y_i are independent, let $p = (1-1/n)^n$, and therefore

$$\mathbb{E}\left[\exp(tY)\right] = \prod_{i\in[1,n]}\mathbb{E}\left[\exp(tY_i)\right] = \left(p\cdot\left(e^t - 1\right) + 1\right)^n.$$

By linearity of expectation, we have $\mathbb{E}[X] = \sum_{i\in[1,n]}\mathbb{E}[X_i] = np$. Therefore, the Chernoff bound is derived as follows

$$\Pr\left[X \ge (1+\delta)\,np\right] \le \frac{\mathbb{E}\left[\exp(tX)\right]}{\exp\left((1+\delta)\,npt\right)} \le \frac{\mathbb{E}\left[\exp(tY)\right]}{\exp\left((1+\delta)\,npt\right)} \le \left(\frac{e^\delta}{(1+\delta)^{1+\delta}}\right)^{np}.$$

□

Remark 1.5.27. This trick in balls-and-bins model is analogous to the Poisson approximation. The spirit is the same: both are replacing dependent counts with independent random variables whose marginal behaviors matches that of the exact model. Poisson approximation captures more than just the expected load per bin, it captures the correct distributional shape and small-count probabilities, and approximate the joint distribution via independence, giving a limit theorem for the exact case. The Bernoulli trick preserves the expected emptiness per bin via independent random variables, allowing Chernoff to be applied, but it is more of a bounding device than a limit theorem.

Remark 1.5.28. This trick can be extended to m balls and n bins, and be used in Bloom Filter analysis on the number of zeroes (k hash functions, m disallowed passwords, and n bits to store, yielding km balls and n bins model).

Exercise 1.5.29 (Exercise 5.23 [MU17])**.** Bloom filter can be used to estimate set differences. Suppose there are sets X and Y both with n elements. Create Bloom filters for X and Y, using the same number of bits m and the same k hash functions. Determine the expected number of bits where Bloom filters differ as a function of m, n, k, and $|X \cap Y|$.

Proof. Let $A = X \setminus Z$, $B = Y \setminus Z$, $Z = X \cap Y$, and $|Z| = c$. We want to know the probability of some bit being set by only one of A or B, and not by Z. Therefore, we define following events

- $E_{i,0}$ be the event that i^{th} bit set by A not by B,
- $E_{i,1}$ be the event that i^{th} bit set by B not by A,
- $E_{i,2}$ be the event that i^{th} bit not set by Z,

and let $\{W_i\}_{i\in[1,m]}$ be 0-1 random variables such that $W_i = 1$ if and only if $(E_{i,0} \cup E_{i,1}) \cap E_{i,2}$, and $W = \sum_i W_i$.

Since $E_{i,0}$ and $E_{i,1}$ are symmetric, WLOG we analyze $E_{i,0}$. Assuming hash function set a random bit to 1, then

$$\Pr[E_{i,0}] = \left(1 - \frac{1}{m}\right)^{k(n-c)} \cdot \left(1 - \left(1 - \frac{1}{m}\right)^{k(n-c)}\right).$$

On the other hand, not being set by Z for i^{th} bit is event $E_{i,2}$, which follow $\Pr[E_{i,2}] = (1 - 1/m)^{kc}$.

Therefore, $\Pr[W_i = 1] = 2\Pr[E_{i,0}]\Pr[E_{i,2}]$, and by linearity of expectation,

$$\mathbb{E}\left[W = \sum_{i\in[1,m]} W_i\right] = 2m\left(1 - \frac{1}{m}\right)^{kn} \cdot \left(1 - \left(1 - \frac{1}{m}\right)^{k(n-c)}\right).$$

□

Definition 1.5.30. In $G_{n,p}$ model we consider all undirected graphs on n distinct vertices, an edge is connected between 2 distinct vertices with probability p, thus a graph with a given set of m edges has probability $p^m(1-p)^{\binom{n}{2}-m}$. The expected number of edges in the graph is $p\binom{n}{2}$, and each vertex has expected degree $(n-1)p$.

In $G_{n,N}$ model we consider all undirected graphs on n distinct vertices with exact N edges. There are $\binom{\binom{n}{2}}{N}$ distinct graphs to select with equal probability.

Remark 1.5.31. The relation between $G_{n,p}$ and $G_{n,N}$ has following properties: Let $p = N/\binom{n}{2}$, the number of edges in $G \xleftarrow{R} G_{n,p}$ is concentrated around N. Conditioned on $G \xleftarrow{R} G_{n,p}$ having N edges, then G is uniform over $G_{n,N}$.

Such relation is similar to the one between Poisson approximation and the balls-and-bins model.

Lemma 1.5.32. *The distribution of $G_{n,p}$ conditioned on the graph sampled has N edges is the same as $G_{n,N}$ regardless of N.*

Lemma 1.5.33. *Let E be an event. If $\Pr[E] = t$ in the $G_{n,p}$ model, then $\Pr[E] = O(Nt)$ in the $G_{n,N}$ model where $N = p\binom{n}{2}$.*

Proof. By the probability lower bounding from conditional probability (trick used in theorem 1.5.15), we have

$$\begin{aligned}\Pr\left[E \text{ in } G_{n,p}\right] &= \sum_{k\geq 0} \Pr\left[E \text{ in } G_{n,p} \mid \text{graph sampled in } G_{n,p} \text{ has } k \text{ edges}\right] \cdot \Pr\left[\text{graph sampled in } G_{n,p} \text{ has } k \text{ edges}\right] \\ &\geq \Pr[E \text{ in } G_{n,N}] \cdot \Pr\left[\text{graph sampled in } G_{n,p} \text{ has } N \text{ edges}\right].\end{aligned}$$

Let $M = (1-p)\binom{n}{2}$. To lower bound the probability of graph sampled in $G_{n,p}$ with N edges, the probability follows

$$\Pr\left[\text{graph sampled in } G_{n,p} \text{ has } N \text{ edges}\right] = \binom{\binom{n}{2}}{N} \cdot p^N (1-p)^M = (N+M)! \cdot \frac{p^N}{N!} \cdot \frac{(1-p)^M}{M!}.$$

By result from corollary 1.5.5, we can lower bound components as follows:

$$\begin{aligned}(N+M)! &\geq \left(\frac{N+M}{e}\right)^{N+M}\\ \frac{p^N}{N!} &\geq \frac{p^N}{e\sqrt{N}}\cdot\left(\frac{e}{N}\right)^N = \frac{e^N}{e\sqrt{N}}\cdot(N+M)^{-N}\\ \frac{(1-p)^M}{M!} &\geq \frac{(1-p)^M}{e\sqrt{M}}\cdot\left(\frac{e}{M}\right)^M = \frac{e^M}{e\sqrt{M}}\cdot(N+M)^{-M},\end{aligned}$$

The probability is lower bounded by $O(N^{-1})$, therefore the probability for the event happening in $G_{n,N}$ is $O(Nt)$. □

Remark 1.5.34. Using Stirling bound, then the probability for the event to happen in $G_{n,N}$ can be improved to $O(nt)$.

Definition 1.5.35. A graph property is a property that holds for a graph, and all the isomorphisms of the graph. We say a graph property is monotone increasing if whenever the property holds for $G = (V, E)$, it holds for any graph $G' = (V', E')$ with $E \subseteq E'$. Monotone decreasing property is defined similarly, that whenever the property holds for $G = (V, E)$, it holds for any graph $G' = (V', E')$ with $E' \subseteq E$.

Lemma 1.5.36 (Lemma 5.14 [MU17])**.** *For a given monotone increasing graph property, let $P_{n,N}$ be the probability that the property holds for a graph in $G_{n,N}$ and $P_{n,p}$ be the probability that the property holds for a graph in $G_{n,p}$. Let $p^+ = (1+\varepsilon)N/\binom{n}{2}$ and $p^- = (1-\varepsilon)N/\binom{n}{2}$ for a constant $\varepsilon \in (0,1)$. Then*

$$P_{n,p^-} - e^{-O(N)} \leq P_{n,N} \leq P_{n,p^+} + e^{-O(N)}.$$

Proof. The main theme of the proof is by relaxation from conditional probability, then apply Chernoff bound.

Let X be the random variable for the number of edges of a graph that is sampled from G_{n,p^-}. We relax P_{n,p^-} by

$$\begin{aligned}P_{n,p^-} &= \sum_{k\leq N} P_{n,k}\cdot\Pr[X=k] + \sum_{k>N} P_{n,k}\cdot\Pr[X=k]\\ &\leq P_{n,N}\cdot\Pr[X\leq N] + \sum_{k>N} P_{n,k}\cdot\Pr[X=k]\\ &\leq P_{n,N} + \Pr[X>N]\end{aligned}$$

The first equality follow from conditional probability, the second inequality holds from monotone increasing property has $P_{n,N} \geq P_{n,k}$ for all $k \leq N$, and the last inequality holds from bounding valid probability by 1.

Now we bound $\Pr[X > N]$ by Chernoff bound, as X is the sum of $\binom{n}{2}$ independent Bernoulli random variables with probability p^-, then

$$\Pr[X>N] = \Pr\left[X > \frac{1}{1-\varepsilon}p^-\binom{n}{2}\right] < \Pr\left[X > (1+\varepsilon)\,p^-\binom{n}{2}\right] \leq \exp\left(-\frac{1}{3}\varepsilon^2 p^-\binom{n}{2}\right) = \exp\left(-\frac{\varepsilon^2(1-\varepsilon)}{3}N\right),$$

as $1/(1-\varepsilon) > 1+\varepsilon$ for $0 < \varepsilon < 1$, which proves the $P_{n,N} \geq P_{n,p^-} - e^{-O(N)}$.

Similarly, let X be the random variable for the number of edges of a graph that is sampled from G_{n,p^+}. Then

$$\begin{aligned}P_{n,p^+} &= \sum_{k<N} P_{n,k}\cdot\Pr[X=k] + \sum_{k\geq N} P_{n,k}\cdot\Pr[X=k]\\ &\geq P_{n,N}\cdot\Pr[X\geq N]\\ &\geq P_{n,N} - \Pr[X<N].\end{aligned}$$

Again by Chernoff bound,

$$\Pr[X<N] = \Pr\left[X < \frac{1}{1+\varepsilon}p^+\binom{n}{2}\right] < \Pr\left[X < \left(1-\frac{\varepsilon}{2}\right)p^+\binom{n}{2}\right] \leq \exp\left(-\frac{\varepsilon^2}{8}p^+\binom{n}{2}\right) = \exp\left(-\frac{\varepsilon^2(1-\varepsilon)}{8}N\right),$$

as $1/(1+\varepsilon) < 1-\varepsilon/2$ for $0 < \varepsilon < 1$, which proves the $P_{n,N} \leq P_{n,p^+} + e^{-O(N)}$, and we complete the proof. □

Remark 1.5.37. The spirit is similar to the Poisson approximation theorem 1.5.17. In the n balls m bins model, we replace dependent bins with independent m Poisson random variables of mean n/m. In the random graph case, we replace dependent N edges among all possible $\binom{n}{2}$ edges with each lined up with probability $N/\binom{n}{2}$ independently.

Lemma 1.5.38. *For a given monotone increasing graph property, let $P_{n,N}$ and $P_{n,p}$ be notions follow from lemma 1.5.36. Let $N = p\binom{n}{2}$, $N^+ = (1+\varepsilon)N$, and $N^- = (1-\varepsilon)N$ for a constant $\varepsilon \in (0,1)$. Then*

$$P_{n,N^-} - e^{-O(N)} \le P_{n,p} \le P_{n,N^+} + e^{-O(N)}.$$

Proof. We prove with a similar strategy to lemma 1.5.36.

Let X be the random variable for the number of edges of a graph sampled from $G_{n,p}$. X can be seen as the sum of $\binom{n}{2}$ independent Bernoulli random variables with probability p. By conditional probability,

$$\begin{aligned} P_{n,p} &= \sum_{k \le N^+} P_{n,k} \cdot \Pr[X = k] + \sum_{k > N^+} P_{n,k} \cdot \Pr[X = k] \\ &\le P_{n,N^+} \cdot \Pr[X \le N^+] + \sum_{k > N^+} P_{n,k} \cdot \Pr[X = k] \\ &\le P_{n,N^+} + \Pr[X > N^+]. \end{aligned}$$

By Chernoff bound, the upper tail of X can be bounded as follows

$$\Pr[X > N^+] \le \exp\left(-\frac{1}{3}\varepsilon^2 N\right) = \exp(-O(N)).$$

Similarly, again by conditional probability,

$$\begin{aligned} P_{n,p} &= \sum_{k < N^-} P_{n,k} \cdot \Pr[X = k] + \sum_{k \ge N^-} P_{n,k} \cdot \Pr[X = k] \\ &\ge P_{n,N^-} \cdot \Pr[X \ge N^-] \\ &\ge P_{n,N^-} - \Pr[X < N^-]. \end{aligned}$$

By Chernoff bound, the lower tail bound of X can be bounded as follows

$$\Pr[X < N^-] \le \exp\left(-\frac{1}{2}\varepsilon^2 N\right) = \exp(-O(N)).$$

Therefore, we complete the proof on both sides. □

Exercise 1.5.39 (Exercise 5.19 [MU17])**.** An undirected graph on n vertices is disconnected if there exists a set of $k < n$ vertices such that there is no edge between this set and the rest of the graph. Otherwise, the graph is connected. Show that there exists a constant c such that if $N \ge cn \log n$ then, with probability $1 - o(1)$, $G \overset{R}{\leftarrow} G_{n,N}$ is connected.

Proof. Let $P_{n,N}$ be the probability of $G \overset{R}{\leftarrow} G_{n,N}$ being connected, $P_{n,p}$ be the probability of $G \overset{R}{\leftarrow} G_{n,p}$ being connected, where $p = N/\binom{n}{2}$. The proof strategy follows: Since connectedness is a monotone increasing graph property, we apply lemma 1.5.36 such that we bound $P_{n,N}$ by

$$P_{n,p^-} - \exp(-O(N)) \le P_{n,N} \le P_{n,p^+} + \exp(-O(N)),$$

where $p^+ = (1+\varepsilon)p$ and $p^- = (1-\varepsilon)p$. Suppose P_{n,p^+} and P_{n,p^-} are both $1 - o(1)$, then we complete the proof.

We continue by analyzing $G_{n,p}$ model. For disconnected subset, the vertex size ranges in $\left[1, \left\lfloor \frac{n}{2} \right\rfloor\right]$ by symmetry. Let E be the event that $G \overset{R}{\leftarrow} G_{n,p}$ being disconnected. Then $\Pr[E]$ is union bounded by

$$\Pr[E] \le \sum_{k \in \left[1, \left\lfloor \frac{n}{2} \right\rfloor\right]} \binom{n}{k} (1-p)^{k(n-k)} = \sum_{k \in \left[1, \left\lfloor \frac{n}{2} \right\rfloor\right]} a_k.$$

We want to show that $\Pr[E] = o(1)$ as $n \to \infty$. Observing that

$$\frac{a_{k+1}}{a_k} = \frac{n-k}{k+1} \cdot (1-p)^{n-2k-1},$$

then for a fixed k, as $n \to \infty$, $a_{k+1}/a_k \to 0$. Suppose for some K, we have $\frac{n}{2} \cdot (1-p)^{n-2K-1} < 1$, then $\{a_k\}_{k\in[1,K]}$ are monotone decreasing. We thus separate into 2 cases:

- $\sum_{k\le K} a_k \le Ka_1 = Kn(1-p)^{n-1}$, and the upper bound approaches $Kn\exp\left(-p(n-1)\right)$ as $n \to \infty$.
- $\sum_{k>K} a_k < 2^n(1-p)^{\alpha n^2}$ by binomial coefficients for some α, and the upper bound approaches 0 when $n \to \infty$.

Let $p = O(\log n/n)$ with a sufficiently large constant, then the former bound approaches 0 when $n \to \infty$.

The rest follow from both $P_{n,p^+} = P_{n,p^-} = 1 - o(1)$ when ε is sufficiently small. Bounding by lemma 1.5.36, then when $N \ge cn\log n$ we complete the proof. □

Theorem 1.5.40 (Theorem 5.15 [MU17])**.** *Let $N = \frac{1}{2}(n\ln n + cn)$. Then the probability that there are no isolated vertices (vertices with degree 0) in $G_{n,N}$ converges to $e^{-e^{-c}}$ as $n \to \infty$.*

Proof. The main idea is utilizing the balls-and-bins model to model adding edges in the $G_{n,N}$ model.

We introduce edges to the graph $G = (V, E)$, with $|V| = n$ and $E = \varnothing$ on initialization.

Repeat $\ell = N + \sqrt{n}$ times, each time toss 2 balls into n bins at random.

- If they gets into the i^{th} and j^{th} bin, with $i \ne j$, edge $(i, j) \notin E$, and $|E| \le N$, introduce the edge.
- Otherwise, take back the 2 balls from the bins.

Conditioning on having N edges, the distribution of the resulting G is equivalent to sampling from $G_{n,N}$.

Lemma 1.5.41. *After repeating ℓ times, as $n \to \infty$, $\Pr[|E| = N] = 1 - o(1)$.*

Proof. There are 2 mutually exclusive cases that an edge introduction trial get rejected: Either the 2 balls are tossed into the same bin, with probability $r_0 = n/\left(n + \binom{n}{2}\right)$; or rejected by prior introduction, upper bounded by $r_1 = N/\binom{n}{2}$.

Let $\{X_i\}_{i\in[1,\ell]}$ be independent 0-1 random variables of probability $r = r_0 + r_1$, and $X = \sum_i X_i$. Then X is a random variable upwards relaxing the number of rejections in the edge introduction trials.

Since by linearity of expectation $\mathbb{E}[X] = \ell r = O(\log^2 n)$, then by Markov/Chernoff concentration inequality, we have $\Pr\left[X > \sqrt{n}\right] = o(1)$, which completes the proof. □

Let E_i be the event that G has no isolated vertices after the i^{th} edge introduction trial. Next corollary is immediate.

Corollary 1.5.42. *As $n \to \infty$, $\Pr[E_\ell] = \Pr[E_\ell \mid |E| = N]$.*

Let $M_{n,m}$ be the event that n balls and m bins model having no empty bin.

On one hand, $\Pr[E_\ell]$ is upper bounded by the 2ℓ balls and n bins model without empty bins, as:

- Repeated edge introduction does not matter in bin emptiness.
- 2 balls into the same bin means bin filled in the balls-bins model, rather than being rejected in the $G_{n,N}$ model.

A direct application of theorem 1.5.23 implies that as $n \to \infty$, $\Pr[E_\ell] \le \Pr[M_{2\ell,n}] = e^{-e^{-c}}$.

On the other hand, $\Pr[E_\ell \mid |E| = N]$ is lower bounded by the $2N$ balls and n bins model with no empty bin, as

- Repeated edge introduction is allowed in the balls-and-bins model.
- 2 balls tossed into the same bin fills 1 bin rather than 2.

A direct application of corollary 1.5.42 and theorem 1.5.23 implies that as $n \to \infty$, $\Pr[E_\ell] \geq \Pr[M_{2N,n}] = e^{-e^{-c}}$.

Since we have sandwiched $\Pr[E_\ell] = e^{-e^{-c}}$ as $n \to \infty$, therefore we conclude the proof [1]. □

Theorem 1.5.43. *Let $p = (\ln n + c)/n$. The probability that there is no isolated vertices in $G_{n,p}$ converges to $e^{-e^{-c}}$ as $n \to \infty$.*

Proof. This is immediate after theorem 1.5.40 and lemma 1.5.38. □

1.5.7 Hamiltonian Cycles in Random Graphs

Definition 1.5.44. Let G be an undirected graph. Suppose that a path $P = v_1, \ldots, v_k$ is a simple path in G, and that (v_k, v_i) is an edge in G. Then

$$P = v_1, \ldots, v_i, v_k, v_{k-1}, \ldots, v_{i+1}$$

is also a simple path, which we refer to as the rotation of P with the rotation edge (v_k, v_i).

On input a graph $G = (V, E)$ with associated "used/unused-edges" lists for each vertex, the modified Hamiltonian cycle finding algorithm follows:

- Start with a random vertex as the head of the path.
- Repeat the following until a Hamiltonian cycle is closed, or the "unused-edges" of the head vertex is empty.
  - Let the current path be $P = v_1, \ldots, v_k$ with v_k being the head.
  - Execute one of the following 3 cases with probabilities specified as follows.
    - With probability $1/n$, reverse the path and make v_1 the new head.
    - With probability $\left|\text{used-edges}(v_k)\right|/n$, sample $(v_k, v_i) \xleftarrow{R}$ used-edges(v_k), rotate P with edge (v_k, v_i), and let v_{i+1} be new head (if the edge is (v_k, v_{k-1}), take no action).
    - Otherwise, sample $(v_k, u) \xleftarrow{R}$ unused-edges(v_k). If u is not on the path, make u the new head by $v_{k+1} = u$; or $u = v_i$, then rotate P by (v_k, v_i) and make v_{i+1} the new head (if the edge is (v_k, v_{k-1}), take no action).
  - Update "unused-edges" and "used-edges" list accordingly.

We construct a special random graph model similar to $G_{n,p}$. Assuming each of the $n-1$ possible edges connected to a vertex v is initially on the "unused-edges" list for vertex v independently with probability q. We also assume the order of these edges are in random order.

Remark 1.5.45. One way of looking at it is, before beginning the Hamiltonian cycle finding algorithm, we create the "unused-edges" list for each vertex v by inserting each possible edge (u, v) with probability q. The corresponding graph G is the graph including all edges that were inserted into some "unused-edges" list. But notice that an edge (u, v) could initially be on the "unused-edges" list for v but not for u. The independence among each "unused-edges" list is helpful for analysis onwards.

The following lemma says, regardless of the viewed vertices, the next vertex to view is still uniformly random, if the "unused-edges" list is not empty in the head vertex.

Lemma 1.5.46. *Suppose the modified Hamiltonian cycle algorithm runs on the random graph model as described. Let V_i be the head vertex after the i^{th} time step. Suppose for any vertex u, as long as the i^{th} time step the unused edges list is not empty for the head vertex,*

$$\Pr[V_{i+1} = u \mid V_i = v_i, V_{i-1} = v_{i-1}, \ldots, V_1 = v_1] = \frac{1}{n}.$$

[1] The proof is inspired by the math stackexchange link. More specifically, I was thinking of sandwiching by $G_{n,p}$, but this solution suggested a way of sampling $G_{n,N}$ graph from the balls-and-bins model, and inspired me sandwiching by different number of balls.

Proof. The first 2 cases are both with probability $1/n$ to make u new head of path.

For the last case, there are at most $p = n - \left|\text{used-edges}(v_i)\right| - 1$ options for v_i's unused edges, at the i^{th} time step. To make u the new head, u has to be in the previously sampled unused-edges(v_k), and the probability follows

$$\Pr\left[V_{i+1} = u \,\middle|\, \bigcap_{j\in[1,i]} V_j = v_j \cap \text{third case}\right] = q \cdot \sum_{k\in[0,p-1]} \left(\binom{p-1}{k} q^k (1-q)^{p-1-k} \cdot \frac{1}{k+1}\right) - (1-q)^{p-1} = \frac{1}{n}.$$

(or actually we can prove by symmetry, that for any possible option, the probability is the same, and it splits up the whole sample space.) Thus we complete the proof. □

Now the problem looks exactly like the coupon collector's problem, that the probability of finding a new vertex to add to the path, when there are k vertices are left to be added, is k/n.

Theorem 1.5.47. *Suppose the input to the modified Hamiltonian cycle algorithm initially has the probability of adding edges to the unused edges list $q \geq 20 \ln n/n$. Then the algorithm managed to find a Hamiltonian cycle in $O(n \ln n)$ time steps with probability $1 - O(n^{-1})$.*

Proof. We start by upper bounding the event $\mathcal{E}$ that the algorithm failed to find such cycle, splitting by:

- The algorithm ran for $3n \ln n$ steps with no empty unused edges list, and failed to close a Hamiltonian cycle.
- The algorithm drained some unused edges list in the first $3n \ln n$ steps, and failed to close a Hamiltonian cycle.

Let the first event be $\mathcal{E}_1$. We relax the second event by lifting the constraint of "failed to close a Hamiltonian cycle", and we let the relaxed event be $\mathcal{E}_2$. We thus have $\Pr[\mathcal{E}] \leq \Pr[\mathcal{E}_1] + \Pr[\mathcal{E}_2]$.

We start with $\mathcal{E}_1$ bounding first. Failing to close a Hamiltonian cycle can be considered as either failing to traverse all vertices, or failing to close by traversing back to starting vertex. We model the vertices traversal conditioned on no unused edges list empty by balls-and-bins model in coupon collector's problem, and let $\mathcal{E}_{1,a}$ be the event that all coupons are collected before $2n \ln n$ time steps. By conditional probability, we have

$$\begin{aligned}\Pr[\mathcal{E}_1] &= \Pr[\mathcal{E}_1 \mid \mathcal{E}_{1,a}] \Pr[\mathcal{E}_{1,a}] + \Pr[\mathcal{E}_1 \mid \neg\mathcal{E}_{1,a}] \Pr[\neg\mathcal{E}_{1,a}] \\ &\leq \Pr[\mathcal{E}_1 \mid \mathcal{E}_{1,a}] + \Pr[\neg\mathcal{E}_{1,a}],\end{aligned}$$

where the first term means not closing the Hamiltonian cycle in at least $n \ln n$ steps after traversing all vertices, and the second term means the coupon collector's problem does not stop before $2n \ln n$ steps.

The first term is upper bounded by

$$\Pr[\mathcal{E}_1 \mid \mathcal{E}_{1,a}] \leq \left(1 - \frac{1}{n}\right)^{n \ln n} \leq \frac{1}{n}.$$

The second term is upper bounded by union bounding over all n vertices not being picked after $2n \ln n$ trials

$$\Pr[\neg\mathcal{E}_{1,a}] \leq n \left(1 - \frac{1}{n}\right)^{2n \ln n} \leq \frac{1}{n}.$$

Therefore, $\Pr[\mathcal{E}_1] \leq 2/n = O(n^{-1})$.

Now we bound $\Pr[\mathcal{E}_2]$. In this case, either all vertices have at least $10 \ln n$ unused edges initially, but some gets hit hard; or some vertices have too few unused edges on initialization. Let $\mathcal{E}_{2,a}$ be the event that some vertices have too few unused edges. By conditional probability

$$\begin{aligned}\Pr[\mathcal{E}_2] &= \Pr[\mathcal{E}_2 \mid \mathcal{E}_{2,a}] \Pr[\mathcal{E}_{2,a}] + \Pr[\mathcal{E}_2 \mid \neg\mathcal{E}_{2,a}] \Pr[\neg\mathcal{E}_{2,a}] \\ &\leq \Pr[\mathcal{E}_2 \mid \neg\mathcal{E}_{2,a}] + \Pr[\mathcal{E}_{2,a}].\end{aligned}$$

The first term can be upper bounded by some vertex being visited too often. An unused edges list can be emptied after at least $10 \ln n$ visits. Let X be the number of visits, which is upper bounded by a sum of $3n \ln n$ independent Bernoulli trials of probability $1/n$. By Chernoff bound,

$$\Pr[X \geq 10 \ln n] \leq \frac{1}{n^2}.$$

By union bounding over all n vertices, the $\Pr[\mathcal{E}_2 \mid \neg\mathcal{E}_{2,a}] \leq 1/n$.

The second term can be bounded similarly. Let Y be the initial number of unused edges. Then

$$\mathbb{E}[Y] = (n-1)q \geq 19 \ln n$$

for a sufficiently large n. By Chernoff bound for $\Pr[Y \leq 10 \ln n]$, then union bounding over n vertices, $\Pr[\mathcal{E}_{2,a}] \leq 1/n$.

Therefore, $\Pr[\mathcal{E}] \leq 4/n = O(n^{-1})$, and we complete the proof. □

Remark 1.5.48. The operation of rotating in definition 1.5.44 is derived from [Pós76], which is often referred to as the Pósa's rotation-extension technique.

1.6 The Probabilistic Method

1.6.1 A few things I should have noted about factorial

We bounded around factorial in corollary 1.5.5. We can derive a tighter lower bound from following lemmas.

Lemma 1.6.1. *For any integer $k \geq 1$, by concavity of $\ln x$,*

$$\ln k \geq \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \ln x dx.$$

Lemma 1.6.2.

$$n! \geq \sqrt{2n} \left(\frac{n}{e}\right)^n.$$

Proof. Since $\ln n! = \sum_{i\in[1,n]} \ln i$, and we have the following from lemma 1.6.1

$$\begin{aligned}\sum_{i\in[1,n]} \ln i &\geq \int_{\frac{1}{2}}^{n+\frac{1}{2}} \ln x dx = (x \ln x - x)\Big|_{\frac{1}{2}}^{n+\frac{1}{2}} \\ &= \left(\left(n+\frac{1}{2}\right)\ln\left(n+\frac{1}{2}\right) - \left(n+\frac{1}{2}\right)\right) - \left(\frac{1}{2}\ln\frac{1}{2} - \frac{1}{2}\right) \\ &\geq \left(n+\frac{1}{2}\right)\ln n - n + \frac{1}{2}\ln 2 = n\ln n - n + \frac{1}{2}\left(\ln(2n)\right),\end{aligned}$$

which completes the proof by exponenting over e on both sides. □

Corollary 1.6.3. *By lemma 1.5.4 and lemma 1.6.2, we have*

$$\sqrt{2n}\cdot\left(\frac{n}{e}\right)^n \leq n! \leq e\sqrt{n}\cdot\left(\frac{n}{e}\right)^n \text{ and } \frac{1}{e\sqrt{n}}\cdot\left(\frac{e}{n}\right)^n \leq \frac{1}{n!} \leq \frac{1}{\sqrt{2n}}\cdot\left(\frac{e}{n}\right)^n.$$

Lemma 1.6.4. *By corollary 1.6.3, we have*

$$\binom{2n}{n} = \Theta\left(\frac{1}{\sqrt{n}}\cdot 2^{2n}\right).$$

Proof. Since $\binom{2n}{n} = (2n)!/(n!)^2$, then we have

$$\begin{aligned}\binom{2n}{n} &\leq e\sqrt{2n}\cdot\left(\frac{2n}{e}\right)^{2n}\cdot\left(\frac{1}{\sqrt{2n}}\cdot\left(\frac{e}{n}\right)^n\right)^2 = \frac{e}{\sqrt{2n}}\cdot 2^{2n} \\ \binom{2n}{n} &\geq 2\sqrt{n}\cdot\left(\frac{2n}{e}\right)^{2n}\cdot\left(\frac{1}{e\sqrt{n}}\cdot\left(\frac{e}{n}\right)^n\right)^2 = \frac{2}{e^2\sqrt{n}}\cdot 2^{2n},\end{aligned}$$

which completes the proof. □

Lemma 1.6.5. *For $s = O(\sqrt{n})$, $(n-s)!\cdot(n+s)! = \Theta(1)\cdot(n!)^2$.*

Proof.

$$\frac{(n-s)!}{n!}\cdot\frac{(n+s)!}{n!} = \prod_{i\in[1,s]}\frac{n-s+i}{n+i} = \prod_{i\in[1,s]}\left(1-\frac{s}{n+i}\right) = \Theta\left(\exp\left(-\frac{s^2}{n}\right)\right) = \exp(-\Theta(1)).$$

□

Corollary 1.6.6. *Immediate from lemma 1.6.5, for $s = O(\sqrt{n})$,*

$$\binom{2n}{n+s} = \Theta(1)\cdot\binom{2n}{n}.$$

Corollary 1.6.7. *Immediate from lemma 1.6.5, for $s = O(\sqrt{n})$,*

$$\binom{2n+s}{n}\binom{2n-s}{n} = \Theta(1)\cdot\binom{2n}{n}^2.$$

Exercise 1.6.8 (Exercise 6.16.b [MU17])**.** Let $\{Y_i\}_{i\in[1,2n]}$ be independent random variables uniform over $\{0,1\}$. Show that there exists a positive constant c such that, for sufficiently large n,

$$\Pr\left[\left|\sum_{i\in[1,n]} Y_i - \sum_{i\in[n+1,2n]} Y_i\right| > c\sqrt{n}\right] > \frac{1}{2}.$$

Proof. Let $Y = \sum_i Y_i$, $A = \sum_{i\in[1,n]} Y_i$, and $B = \sum_{i\in[n+1,2n]} Y_i$. Let E_t be the event $|A-B| > c\sqrt{n}$ when $Y = t$.

Let k be a positive constant. By conditional probability (similar to lemma 1.5.38),

$$\Pr\left[|A-B| > c\sqrt{n}\right] \geq \Pr\left[|A-B| > c\sqrt{n} \mid |Y-n| \leq k\sqrt{n}\right] - \Pr\left[|Y-n| > k\sqrt{n}\right],$$

where $\Pr\left[|Y-n| > k\sqrt{n}\right] < 2\exp\left(-k^2/3\right)$ by Chernoff bound.

We continue by conditional probability, that

$$\begin{aligned}\Pr\left[|A-B| > c\sqrt{n} \mid |Y-n| \leq k\sqrt{n}\right] &= \sum_t \Pr\left[E_t \mid Y = t\right]\cdot\Pr\left[Y = t \mid |Y-n| \leq k\sqrt{n}\right] \\ &= 1 - \sum_t \Pr\left[\neg E_t \mid Y = t\right]\cdot\Pr\left[Y = t \mid |Y-n| \leq k\sqrt{n}\right].\end{aligned}$$

Event $\neg E_t$ is equivalent to $A \in \left[(t - c\sqrt{n})/2, (t + c\sqrt{n})/2\right] = R_t$ when $Y = t$. Therefore,

$$\begin{aligned}\Pr\left[\neg E_t \mid Y = t\right] &= \sum_{a\in R_t} \Pr\left[A = a \mid Y = t\right] \\ &\leq (c\sqrt{n}+1)\cdot\max_a \Pr\left[A = a \mid Y = t\right] \\ &= (c\sqrt{n}+1)\cdot\max_a \binom{n}{a}\binom{n}{t-a}\binom{2n}{t}^{-1} = (c\sqrt{n}+1)\cdot p(a).\end{aligned}$$

By $p(a+1)/p(a)$,

$$\frac{p(a+1)}{p(a)} = \frac{(a+1)(n-t+a+1)}{(n-a)(t-a)},$$

$p(a)$ peaks at $a = \left\lfloor \frac{t}{2} \right\rfloor$ giving $\binom{n}{\lfloor \frac{t}{2} \rfloor}^2 \binom{2n}{t}^{-1}$. On the other hand, by lemma 1.6.4, the following is immediate

$$p\left(\frac{n}{2}\right) = \binom{n}{\frac{n}{2}}^2\binom{2n}{n}^{-1} = \Theta\left(\frac{1}{\sqrt{n}}\right).$$

By corollary 1.6.6, for $t \in [n - k\sqrt{n}, n + k\sqrt{n}]$, we have $p(a) = \Theta\left(\frac{1}{\sqrt{n}}\right)$.

Choose c sufficiently small such that $\Pr\left[\neg E_t \mid Y = t\right] \leq 1/2$ for all $t \in [n - k\sqrt{n}, n + k\sqrt{n}]$, then

$$\sum_t \Pr\left[\neg E_t \mid Y = t\right]\cdot\Pr\left[Y = t \mid |Y-n| \leq k\sqrt{n}\right] \leq \frac{1}{2},$$

which completes the proof. □

Exercise 1.6.9 (Exercise 6.16.c [MU17])**.** Let $\{Y_i\}_{i\in[1,2n]}$ be independent random variables uniform over $\{0,1\}$, and let $\{b_i\}_{i\in[1,2n]}$ each be either +1 or −1. Show that there exists a positive constant c such that, for sufficiently large n,

$$\Pr\left[\left|\sum_{i\in[1,2n]} b_i Y_i\right| > c\sqrt{n}\right] > \frac{1}{2}.$$

Proof. We claim that either the $\{b_i\}$ are inbalanced, that $|B| = |\sum_i b_i| > \ell\sqrt{n}$; or they are balanced. WLOG let $B > 0$.

For the inbalanced case, let $Z = \sum_i b_i Y_i$. By linearity of expectation, $\mathbb{E}[Z] = B/2$. Since $\{Y_i\}$ are independent, by Hoeffding's bound

$$\Pr\left[Z \le \mathbb{E}[Z] - \sqrt{2n}\right] \le \exp\left(-\frac{2(\sqrt{2n})^2}{\sum_i b_i^2}\right) = \exp(-2),$$

as $\sum_i b_i^2 = 2n$. Since $\Pr\left[Z \le \mathbb{E}[Z] - \sqrt{2n}\right] \le e^{-2}$, then for $\ell \ge 2\left(c + \sqrt{2}\right)$,

$$\Pr\left[Z \ge c\sqrt{n}\right] \ge \Pr\left[Z \ge \mathbb{E}[Z] - \sqrt{2n}\right] > 1 - e^{-2} > 1/2$$

the inbalanced base is proved.

Now for the balanced case, we reuse the structure and notion in exercise 1.6.8 (the same window argument) that

$$\begin{aligned}\Pr[\neg E_t \mid Y = t] &\le (c\sqrt{n}+1)\cdot \max_a \Pr[A = a \mid Y = t] \\ &= (c\sqrt{n}+1)\cdot \max_a \binom{n+B/2}{a}\binom{n-B/2}{t-a}\binom{2n}{t}^{-1} = (c\sqrt{n}+1)\cdot p(a).\end{aligned}$$

By $p(a+1)/p(a)$,

$$\frac{p(a+1)}{p(a)} = \frac{(n+B/2-a)(t-a)}{(a+1)(n-B/2-t+a+1)},$$

where $p(a)$ peaks at $\left\lfloor t\cdot\frac{2n+B}{4n}\right\rfloor$.

On the other hand, by corollary 1.6.6 and corollary 1.6.7, the max of $p(a)$ is still $\Theta\left(\frac{1}{\sqrt{n}}\right)$, completing the proof. □

1.6.2 Basic Counting Argument

Exercise 1.6.10 (Exercise 6.10 [MU17]). A family of subsets $\mathcal{F}$ of $[1, n]$ is called an antichain if there is no pair of sets A and B in $\mathcal{F}$ satisfying $A \subset B$.

- Let f_k be the number of sets in $\mathcal{F}$ with size k. Show that

$$\sum_{k\in[0,n]} f_k \cdot \binom{n}{k}^{-1} \le 1.$$

- Argue that $|\mathcal{F}| \le \binom{n}{\lfloor \frac{n}{2}\rfloor}$ for any antichain $\mathcal{F}$.

Proof. We start choosing a random permutation of $[1, n]$. Let $X_k = 1$ if the first k numbers in the permutation yield a set in $\mathcal{F}$, and let $X = \sum_{i\in[1,n]} X_i$. Then X is at most 1 by antichain property, as no larger subset containing all prior elements, and X can be 0 as no element in $\mathcal{F}$ exists on the random enumeration starting from 0.

Therefore, for a $\mathcal{F}$, when we are iterating over all permutations, if a k sized subset is in $\mathcal{F}$, we will encounter $k!(n-k)!$ permutations containing such subset, and we derive

$$\sum_{k\in[0,n]} f_k \cdot k!(n-k)! \le n!,$$

as a permutation can have at most 1 element in $\mathcal{F}$, and the first result is immediate.

Since $\binom{n}{k} \le \binom{n}{\lfloor \frac{n}{2}\rfloor}$ for all $k \in [0, n]$, then

$$|\mathcal{F}| = \sum_{k\in[0,n]} f_k \le \sum_{k\in[0,n]} \binom{n}{\lfloor \frac{n}{2}\rfloor} \cdot f_k \cdot \binom{n}{k}^{-1} \le \binom{n}{\lfloor \frac{n}{2}\rfloor}.$$

□

Remark 1.6.11. This is a result from [Spe28] with a Lubell [Lub66] flavored proof.

Another way of seeing it is, a k sized subset in $\mathcal{F}$ appears in a random permutation over $[1, n]$ with probability $k!(n-k)!/n! = \binom{n}{k}^{-1}$. By linearity of expectation,

$$\mathbb{E}[X] = \sum_{k\in[0,n]} \mathbb{E}[X_k] = \sum_{k\in[0,n]} \frac{f_k}{\binom{n}{k}}.$$

On the other hand, X is a 0-1 random variable, hence $\mathbb{E}[X] \le 1$, which completes the proof.

1.6.3 The Expectation Argument and Derandomization by Conditional Expectation

Lemma 1.6.12 (The Expectation Argument)**.** *For random variable X with $\mathbb{E}[X] = \mu$, $\Pr[X \ge \mu] > 0$ and $\Pr[X \le \mu] > 0$.*

Lemma 1.6.13. *Let $\{X_i\}_{i\in[1,n]}$ be 0-1 random variables, $X = \sum_i X_i$ with $\mathbb{E}[X] = m \cdot 2^{-k}$ for some integer $m \ge 1$ and $k \ge 0$. Then for all $X < \mathbb{E}[X]$, $X \le \mathbb{E}[X] - 2^{-k}$; for all $X > \mathbb{E}[X]$, $X \ge \mathbb{E}[X] + 2^{-k}$.*

Lemma 1.6.14. *Let $\{X_i\}_{i\in[1,n]}$ be 0-1 random variables, $X = \sum_i X_i$ with $\mathbb{E}[X] = m \cdot 2^{-k}$ for some integer $m \ge 1$ and $k \ge 0$. Then*

$$\Pr[X \ge \mathbb{E}[X]] \ge \frac{1}{2^k(n - \mathbb{E}[X]) + 1},$$
$$\Pr[X \le \mathbb{E}[X]] \ge \frac{1}{2^k \mathbb{E}[X] + 1}.$$

Proof. Let $p = \Pr[X \ge \mathbb{E}[X]]$. By conditional expectation and lemma 1.6.13,

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i<\mathbb{E}[X]} i \Pr[X = i] + \sum_{i\ge\mathbb{E}[X]} i \Pr[X = i] \\ &\le (\mathbb{E}[X] - 2^{-k})(1-p) + np,\end{aligned}$$

yielding $p \ge 1/(2^k(n - \mathbb{E}[X]) + 1)$.

Let $p = \Pr[X \le \mathbb{E}[X]]$. By conditional expectation and lemma 1.6.13,

$$\begin{aligned}\mathbb{E}[X] &= \sum_{i\le\mathbb{E}[X]} i \Pr[X = i] + \sum_{i>\mathbb{E}[X]} i \Pr[X = i] \\ &\ge (\mathbb{E}[X] + 2^{-k})(1-p),\end{aligned}$$

yielding $p \ge 1/(2^k \mathbb{E}[X] + 1)$. □

Exercise 1.6.15 (Exercise 6.1 [MU17])**.** Consider an SAT instance with m clauses and each clause has exactly k literals. Give a Las Vegas algorithm that finds an assignment satisfying at least $m(1-2^{-k})$ clauses, and its expected runtime. Give a derandomization of the randomized algorithm using the method of conditional expectation.

Proof. For a k-literal SAT clause, it is satisfied with probability $1 - 2^{-k}$. Let X_i be the 0-1 random variable indicating if the i^{th} clause is satisfied, and $X = \sum_i X_i$ be the random number of satisfied clauses. By linearity of expectation, $\mathbb{E}[X] = m(1-2^{-k})$.

The Las Vegas algorithm searching for assignment satisfying at least $m(1-2^{-k})$ clauses works by keep sampling assignments in $O(km)$ time, and checks if the assignment satisfies sufficient number of clauses in $O(km)$ time.

By lemma 1.6.14, $\Pr[X \ge \mathbb{E}[X]] \ge (2^k(m - \mathbb{E}[X]) + 1)^{-1}$.

The expected number of sampling is at most $2^k(m - \mathbb{E}[X]) + 1$, and the expected runtime is $O(km2^k(m - \mathbb{E}[X]))$.

Let there be $n = O(km)$ variables for the SAT instance, σ be a random permutation of x_i, and x_i be the choice of the i^{th} variable. The derandomized algorithm from conditional expectation wants to prove

$$\mathbb{E}[X] \leq \mathbb{E}[X \mid x_1, \ldots, x_n],$$

the right hand side is the number of satisfied clauses after all x_i are decided. We prove by induction as follows.

The base case $\mathbb{E}[X] = \mathbb{E}[X \mid x_1]$ is immediate by symmetry.

The intermediate step $\mathbb{E}[X \mid x_1, \ldots, x_i] \leq \mathbb{E}[X \mid x_1, \ldots, x_{i+1}]$ can be shown by conditional expectation, as

$$\mathbb{E}[X \mid x_1, \ldots, x_i] = \frac{1}{2} \sum_{x_{i+1} \in \{0,1\}} \mathbb{E}[X \mid x_1, \ldots, x_{i+1}] \leq \max_{x_{i+1} \in \{0,1\}} \mathbb{E}[X \mid x_1, \ldots, x_{i+1}].$$

We can greedily choose the next x_{i+1} that satisfies more clauses.

The n step greedy algorithm suffices to find an assignment satisfying at least $m(1-2^{-k})$ clauses by our analysis. □

Exercise 1.6.16 (Exercise 6.2 [MU17])**.** Prove that for every integer n, there exists a 2-coloring of the edges of K_n so that the total number of monochromatic K_4 is at most $\binom{n}{4}2^{-5}$, and show the following

- A randomized algorithm to find a coloring with at most $\binom{n}{4}2^{-5}$ monochromatic K_4 in expected poly time in n.
- A deterministic polynomial time algorithm to find such a coloring by derandomization.

Proof. There are $\binom{n}{4}$ copies of K_4 in a K_n, and a K_4 is 2-colored monochromatic with probability 2^{-5}.

Let X_i be the 0-1 random variable indicating the i^{th} K_4 being monochromatic, and $X = \sum_i X_i$ be the number of monochromatic K_4 in K_n. By linearity of expectation, $\mathbb{E}[X] = \binom{n}{4}2^{-5}$. Hence, the existence of 2-coloring with at most $\binom{n}{4}2^{-5}$ monochromatic K_4 is immediate by the expectation argument.

We can keep randomly sampling 2-coloring K_n until we see at most $\mathbb{E}[X]$ copies of monochromatic K_4.

By lemma 1.6.14, $\Pr[X \leq \mathbb{E}[X]] \geq (2^5 \mathbb{E}[X] + 1)^{-1}$.

The expected sampling time is at most $2^5 \mathbb{E}[X] + 1$, and expected runtime is $O\left(\binom{n}{4}^2\right) = O(n^8)$.

Let x_i be the i^{th} edge's color choice among all $N = \binom{n}{2}$ edges. The derandomized algorithm wants to show

$$\mathbb{E}[X] \geq \mathbb{E}[X \mid x_1, \ldots, x_N],$$

right hand side is the number of monochromatic K_4 after fixing all vertices' colors. We prove by induction as follows.

The base case $\mathbb{E}[X] = \mathbb{E}[X \mid x_1]$ is immediate by symmetry.

The intermediate step $\mathbb{E}[X \mid x_1, \ldots, x_i] \geq \mathbb{E}[X \mid x_1, \ldots, x_{i+1}]$ can be shown by conditional expectation, as

$$\mathbb{E}[X \mid x_1, \ldots, x_i] = \frac{1}{2} \sum_{x_{i+1} \in \{0,1\}} \mathbb{E}[X \mid x_1, \ldots, x_{i+1}] \geq \min_{x_{i+1} \in \{0,1\}} \mathbb{E}[X \mid x_1, \ldots, x_{i+1}].$$

We can greedily choose the next x_{i+1} that bring less monochromatic K_4.

The N step greedy algorithm suffices to 2-color K_n with at most $\binom{n}{4}2^{-5}$ monochromatic K_4 by our analysis. □

Remark 1.6.17. This appears in The Probabilistic Method [AS16], and uses the "pessimistic estimator" idea [Rag88].

Exercise 1.6.18 (Exercise 6.3 [MU17])**.** Given a n-vertex undirected graph, consider the following method of generating an independent set. Given a permutation σ of V, define $S(\sigma) \subset V$ as follows: for each vertex i, $i \in S(\sigma)$ if and only if no neighbor j of i precedes i in the permutation σ.

- Show $\mathbb{E}[|S(\sigma)|] = \sum_i 1/(d_i + 1)$, where $\{d_i\}_{i \in [1,n]}$ denotes the degree of vertices.
- Prove that G has an independent set of size at least $\sum_i 1/(1 + d_i)$.

Proof. Given a random σ, the resulting $S(\sigma)$ has each vertex with no neighbors, which is an independent set.

Let $\{X_i\}_{i\in[1,n]}$ be 0-1 random variables indicating if the i^{th} vertex is the first among all its neighbors, and $X = \sum_i X_i$ is the random variable for $|S(\sigma)|$. The i^{th} vertex is the first among all d_i neighbors and itself with probability $1/(1+d_i)$. Therefore, by linearity of expectation, $\mathbb{E}[X] = \sum_i 1/(1+d_i)$.

The existence of an independent set at least $\sum_i 1/(1+d_i)$ large is immediate by the expectation argument. □

Exercise 1.6.19 (Exercise 6.5 [MU17])**.** We have shown using the probabilistic method that, if a graph G has n nodes and m edges, then there exists a partition of the n nodes into 2 sets such that at least $m/2$ edges cross the partition. Improve this result slightly: show that there exists a partition such that at least $mn/(2n-1)$ edges cross the partition.

Proof. The $m/2$ result derive from 2-coloring the vertices each with probability $1/2$, and same color vertices goes to the same set. In this way, edges with two ends colored differently are included in the cut, with a probability of $1/2$. Existence of cut with at least $m/2$ edges is immediate by linearity of expectation and the expectation argument.

On the other hand, n vertices can be partitioned into sets with $\lfloor \frac{n}{2} \rfloor$ and $\lceil \frac{n}{2} \rceil$ vertices each, and improve by the expectation argument. Let the event of choosing 2 vertices with different colors be E, the probability is

$$\Pr[E] = \frac{2\lfloor \frac{n}{2} \rfloor \lceil \frac{n}{2} \rceil}{n(n-1)}.$$

- When n even, the probability is $n/(2n-2) > n/(2n-1)$.
- When n odd, the probability is $(n+1)/2n > n/(2n-1)$.

By linearity of expectation, the expected size of the cut is $m\Pr[E] \geq mn/(2n-1)$. Hence, the existence of a cut with at least $mn/(2n-1)$ is immediate by the expectation argument. □

Exercise 1.6.20 (Exercise 6.6 [MU17])**.** A k-cut is a partition of the vertices into k disjoint sets, and the value of a cut is the weight of all edges crossing from one of the k sets to another. Show that any graph G with m edges has a k-cut with value at least $(k-1)m/k$. Show how to use derandomization to deterministically find such a cut.

Proof. We assign one out of the k colors to each of the n vertices with probability $1/k$. The probability of 2 vertices having same color is $1/k$, and therefore an edge being included into a k-cut has probability $(k-1)/k$.

By linearity of expectation, the expected k-cut size is $m(k-1)/k$. Hence, the existence of k-cut with size at least $m(k-1)/k$ is immediate by the expectation argument.

When it comes to the derandomized deterministic algorithm for finding a k-cut that is larger than $m(k-1)/k$, we let $\{S_i\}_{i\in[1,k]}$ be the sets of partitioned vertices, and we write $C(\{S_i\}_{i\in[1,k]})$ for size of the k-cut. We pick an arbitrary emulation of all vertices $\{v_i\}_{i\in[1,n]}$, and write $\{x_i\}_{i\in[1,n]}$ for each vertex's color choice. We prove by induction for

$$\mathbb{E}\left[C(\{S_i\}_{i\in[1,k]})\right] \leq \mathbb{E}\left[C(\{S_i\}_{i\in[1,k]}) \mid x_1, \ldots, x_n\right],$$

where the right hand side is the k-cut size determined by the derandomized algorithm, where the color choices are $\{x_i\}_{i\in[1,n]}$, hence the algorithm gives a cut whose size is at least $\mathbb{E}\left[C(\{S_i\}_{i\in[1,k]})\right] = m(k-1)/k$.

The base case $\mathbb{E}\left[C(\{S_i\}_{i\in[1,k]})\right] \leq \mathbb{E}\left[C(\{S_i\}_{i\in[1,k]}) \mid x_1\right]$ is immediate by symmetry.

The intermediate step

$$\mathbb{E}\left[C(\{S_i\}_{i\in[1,k]}) \mid x_1, \ldots, x_i\right] \leq \mathbb{E}\left[C(\{S_i\}_{i\in[1,k]}) \mid x_1, \ldots, x_{i+1}\right]$$

can be shown by conditional expectation, that

$$\begin{aligned}\mathbb{E}\left[C(\{S_i\}_{i\in[1,k]}) \mid x_1, \ldots, x_i\right] &= \frac{1}{k}\sum_{x_{i+1}\in[1,k]} \mathbb{E}\left[C(\{S_i\}_{i\in[1,k]}) \mid x_1, \ldots, x_{i+1}\right] \\ &\leq \max_{x_{i+1}\in[1,k]} \mathbb{E}\left[C(\{S_i\}_{i\in[1,k]}) \mid x_1, \ldots, x_{i+1}\right].\end{aligned}$$

In fact, we greedily find the next x_{i+1} without computing all conditional expectations on x_{i+1}'s choice. We choose the x_{i+1} such that v_{i+1} contribute most edges to the k-cut. Equivalently, we put v_{i+1} to the partition with the fewest v_{i+1}'s colored neighbors, such that v_{i+1} introduces most k-cut edges with colored neighbors.

This n steps greedy searching algorithm suffices to determine a k-cut at least $m(k-1)/k$ large by our analysis. □

Exercise 1.6.21 (Exercise 6.9 [MU17])**.** A tournament is a graph on n vertices with exactly one directed edge between each pair of vertices. If vertices represent players, then each edge can be thought of as the result of a match between the two players: the edge points to the winner. A ranking is an ordering of the n players from best to worst (ties are not allowed). Given the outcome of a tournament, one might wish to determine a ranking of the players. A ranking is said to disagree with a directed edge from y to x if y is ahead of x in the ranking.

- Prove that, for every tournament, there exists a ranking that disagrees with at most 50% of the edges.
- Prove that, for sufficiently large n, there exists a tournament such that every ranking disagrees with at least 49% of the edges in the tournament.

Proof. Fix a tournament, and let X_i be the 0-1 random variable indicating if a random ranking disagrees with the i^{th} of all $\binom{n}{2}$ edges of the tournament, then $X = \sum_i X_i$ is the total number of disagreement.

For any unordered pair (x, y), each ordering is equally likely in a random ranking. Therefore, $\Pr[X_i = 1] = 1/2$. By linearity of expectation,

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = \frac{1}{2}\binom{n}{2}.$$

By the expectation argument, there exists a ranking disagreeing at most 50% of the edges, proving the first part.

Fix a ranking. Let $Y = \sum_j Y_j$ be the random variable for the number of disagreement with a random tournament, and Y_j be the 0-1 random variable indicating if the ranking disagrees with the j^{th} edge of the random tournament.

Let the fixed ranking be the k^{th} among all $n!$ ranking, and let E_k be the event that the fixed k^{th} ranking has at most 49% of disagreement against a random tournament, namely $Y \le (1-\delta)\mathbb{E}[Y]$, where $\delta = 1/50$. Since $\{Y_j\}$ are independent, we can use concentration inequality like Chernoff bound to upper bound the lower tail of Y as follows

$$\Pr[Y \le (1-\delta)\mathbb{E}[Y]] \le \exp\left(-\delta^2\mu/2\right) = \exp\left(-\frac{1}{50^2}\cdot\frac{1}{2}\binom{n}{2}\cdot\frac{1}{2}\right).$$

Union bounding over all rankings yields

$$\Pr\left[\bigcup_{k\in[1,n!]} E_k\right] \le \sum_{k\in[1,n!]} \Pr[E_k] = n!\cdot\exp\left(-\frac{1}{100^2}\binom{n}{2}\right) \le n^n\cdot\exp\left(-\frac{1}{100^2}\binom{n}{2}\right) = \exp\left(n\ln n - \frac{1}{100^2}\binom{n}{2}\right),$$

and the bound is less than 1 if n is sufficiently large. Thus, for a tournament, the probability of existing a ranking permutation with at most 49% disagreement is strictly less than 1. Hence,

$$\Pr\left[\bigcap_{k\in[1,n!]} \neg E_k\right] = 1 - \Pr\left[\bigcup_{k\in[1,n!]} E_k\right] > 0.$$

□

1.6.4 Sample and Modify

Exercise 1.6.22 (Exercise 6.7 [MU17])**.** A Hypergraph $H = (V, E)$, where V is the set of vertices and E is the set of hyperedges, and every hyperedge in E is a subset of V. In particular, r-uniform hypergraph is one where the size of each hyperedge is r. A standard graph is a 2-uniform hypergraph. A dominanting set in a hypergraph H is a set of vertices $S \subset V$ such that $e \cap S \neq \emptyset$ for every $e \in E$, namely S hits every hyperedge in H.

Let $|V| = n$ and $|E| = m$, show that there is a dominanting set S with size at most $np + (1-p)^r m$ for $0 \le p \le 1$.

Proof. Given a hypergraph $H = (V, E)$ with $|V| = n$ and $|E| = m$, we first sample across V with each vertices to be chosen with probability p, then check if any hyperedges are not covered, choose a vertex from each of the hyperedge. The sampling stage has expected chosen vertices np, leaving expected unchosen hyperedges $m(1-p)^r$ to be fixed. □

Exercise 1.6.23 (Exercise 6.8 [MU17])**.** Prove that, for every integer n, there is a way to 2-color the edges of K_x such that there is no monochromatic cliques of size k when

$$x = n - \binom{n}{k} 2^{1-\binom{k}{2}}.$$

Proof. We start 2-coloring K_n, keep removing vertices and attached edges until there is no monochromatic K_k.

Let each edge in K_n be 2-colored with probability 1/2, and let X_i be the 0-1 random variable indicating the i^{th} K_k being monochromatic. There are $\binom{n}{k}$ K_k's in a K_n, hence $X = \sum_i X_i$ is the number of monochromatic K_k's in the K_n.

For a K_k, the probability of being monochromatic is $2^{1-\binom{k}{2}}$ as it can be one of 2 colors. By linearity of expectation,

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = \binom{n}{k} 2^{1-\binom{k}{2}}.$$

By expectation argument, there exists a 2-coloring such that $X \leq \mathbb{E}[X]$.

To modify up to no monochromatic K_k, we can remove a vertex and the attached edges in each K_k, therefore we need to remove at most X vertices from K_n. Since there exists a 2-coloring for K_n containing at most $\mathbb{E}[X]$ K_k's, then there exists a 2-coloring with no monochromatic K_k for K_m that $m \geq n - \mathbb{E}[X]$.

"no monochromatic K_k" is a monotonely decreasing graph property, therefore any subgraph of such 2-colored K_m has this property. Therefore, a K_x in this K_m has no monochromatic K_k, which completes the proof. □

1.6.5 The Second Moment Method and the Conditional Expectation Inequality

Theorem 1.6.24 (The Second Moment Method (Chebyshev's version))**.** *Let X be a random variable, then*

$$\Pr[X = 0] \leq \frac{\operatorname{Var}[X]}{\mathbb{E}[X]^2}.$$

Proof. Since

$$\Pr[X = 0] \leq \Pr[|X - \mathbb{E}[X]| \geq \mathbb{E}[X]] \leq \frac{\operatorname{Var}[X]}{\mathbb{E}[X]^2},$$

which is immediate by Chebyshev's inequality. □

Lemma 1.6.25. *For integer valued non-negative random variable X, $\Pr[X > 0] \leq \mathbb{E}[X]$ by Markov's inequality.*

Theorem 1.6.26 (Cauchy-Schwarz)**.** *For random variables X and Y (that may not be independent),*

$$\mathbb{E}[XY]^2 \leq \mathbb{E}\left[X^2\right] \cdot \mathbb{E}\left[Y^2\right].$$

Proof. Let $Z = X - t \cdot Y$, then the second moment of Z is

$$\mathbb{E}\left[Z^2\right] = \mathbb{E}\left[X^2\right] + t^2 \cdot \mathbb{E}\left[Y^2\right] - 2t \cdot \mathbb{E}[XY] \geq 0,$$

which holds for any t. Minimizing over t, and we have $\mathbb{E}\left[X^2\right] \cdot \mathbb{E}\left[Y^2\right] \geq \mathbb{E}[XY]^2$. □

Corollary 1.6.27. *For random variable X with $\mathbb{E}[X] \geq 0$, $0 \leq t \leq \mathbb{E}[X]$, and $1_t = 1$ if and only if $X \geq t$, by theorem 1.6.26*

$$\Pr[X \geq t] \geq \frac{\mathbb{E}[X \cdot 1_t]^2}{\mathbb{E}[X^2]}.$$

Theorem 1.6.28 (The Second Moment Method (Cauchy-Schwarz's version))**.** *For random variable $X \geq 0$,*

$$\mathbb{E}[X] \geq \Pr[X > 0] \geq \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Proof. Let $1_+ = 1$ if and only if $X > 0$, then by theorem 1.6.26 and corollary 1.6.27,

$$\mathbb{E}\left[X^2\right]\mathbb{E}\left[1_+^2\right] \geq \mathbb{E}\left[X^2\right]\Pr[X > 0] \geq \mathbb{E}[X \cdot 1_+]^2 \geq \mathbb{E}[X]^2,$$

and the last inequality holds by non-negativity of X. □

Remark 1.6.29. A weaker version of theorem 1.6.28 is in exercise 1.3.10. We can relax non-negative X to $\mathbb{E}[X] \geq 0$.

Theorem 1.6.30 (Paley-Zygmund)**.** *For random variable $X \geq 0$ with finite $\mathbb{E}\left[X^2\right]$, and $0 \leq \alpha < 1$. Then*

$$\Pr[X \geq \alpha\,\mathbb{E}[X]] \geq (1-\alpha)^2 \cdot \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

Proof. Let $1_\alpha = 1$ if and only if $X \geq \alpha\,\mathbb{E}[X]$, and $1_{\overline{\alpha}} = 1$ if and only if $X < \alpha\,\mathbb{E}[X]$, then

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[X \cdot 1_\alpha] + \mathbb{E}[X \cdot 1_{\overline{\alpha}}] \\ &\leq \mathbb{E}[X \cdot 1_\alpha] + \alpha \cdot \mathbb{E}[X],\end{aligned}$$

and therefore $\mathbb{E}[X \cdot 1_\alpha] \geq (1-\alpha) \cdot \mathbb{E}[X]$. By theorem 1.6.26 and corollary 1.6.27, we have

$$\mathbb{E}[X \cdot 1_\alpha]^2 \leq \mathbb{E}\left[X^2\right] \cdot \mathbb{E}[1_\alpha] = \mathbb{E}\left[X^2\right] \cdot \Pr[X \geq \alpha\,\mathbb{E}[X]].$$

Hence, the result is immediate by

$$\Pr[X \geq \alpha\,\mathbb{E}[X]] \geq \frac{\mathbb{E}[X \cdot 1_\alpha]^2}{\mathbb{E}[X^2]} \geq (1-\alpha)^2 \cdot \frac{\mathbb{E}[X]^2}{\mathbb{E}[X^2]}.$$

□

Remark 1.6.31. The theorem 1.6.30 can be relaxed from non-negative X to $\mathbb{E}[X] \geq 0$.

Exercise 1.6.32 (Exercise 6.11 [MU17])**.** Prove a threshold for the existence of triangles in the $G_{n,p}$. Let $\{X_i\}$ each be the random value for the i^{th} triangle appearing in the graph among the $\binom{n}{3}$ triplets of vertices, and let $X = \sum_i X_i$.

- Show $\mathbb{E}[X]$, and show if $pn \to 0$, then $\Pr[X > 0] \to 0$.
- Show $\operatorname{Var}[X_i] \leq p^3$.
- Show that $\operatorname{Cov}\left[X_i, X_j\right] = p^5 - p^6$ for $O(n^4)$ pairs of triangle triplets $i \neq j$, otherwise 0.
- Show $\operatorname{Var}[X] = O(n^3p^3 + n^4(p^5 - p^6))$.
- Show if $pn \to \infty$ then $\Pr[X = 0] \to 0$.

Proof. By linearity of expectation, $\mathbb{E}[X] = \binom{n}{3}p^3 = O(n^3p^3)$. If $pn \to 0$, then $\mathbb{E}[X] \to 0$.

For variance, we have $\operatorname{Var}[X_i] = \mathbb{E}\left[X_i^2\right] - \mathbb{E}[X_i]^2 = p^3 - p^6 \leq p^3$.

For covariance, when the triplets overlap up to 1 vertex, no edges are shared, and $\mathbb{E}\left[X_iX_j\right] = \mathbb{E}[X_i]\mathbb{E}\left[X_j\right]$, hence $\operatorname{Cov}\left[X_i, X_j\right] = 0$. Otherwise, when 2 vertices are shared, $\mathbb{E}\left[X_iX_j\right] = p^5$, and there are in total $\binom{n}{2;1;1}$ csaes.

Therefore, the variance for X is derived as

$$\begin{aligned}\operatorname{Var}[X] &= \sum_i \operatorname{Var}[X_i] + \sum_{i \neq j} \operatorname{Cov}\left[X_i, X_j\right] = \binom{n}{3}(p^3 - p^6) + \binom{n}{2;1;1}(p^5 - p^6) \\ &= O(n^3p^3 + n^4(p^5 - p^6)).\end{aligned}$$

When $pn \to \infty$, by theorem 1.6.24, $\Pr[X = 0] \leq O((pn)^{-1}) \to 0$. □

Exercise 1.6.33 (Exercise 6.12 [MU17]). Use exercise 1.3.1 to derive the variance of the number of K_4 in $G_{n,p}$.

Proof. Let $\{X_i\}$ be the 0-1 random variable for the existence of the i^{th} K_4 among all possible $\binom{n}{4}$ K_4, and let $X = \sum_i X_i$. By exercise 1.3.1,

$$\mathbb{E}\left[X^2\right] = \sum_i \mathbb{E}\left[X \mid X_i = 1\right] \Pr\left[X_i = 1\right].$$

We know $\Pr[X_i = 1] = p^6$, and among all $\binom{n}{4}$ possible K_4, if $X_i = 1$,

- If there are no vertex overlap, then $\binom{n-4}{4}$ K_4 with probability p^6.
- If there is 1 vertex overlapped, then $\binom{n-4}{3}\binom{4}{1}$ K_4 with probability p^6.
- If there are 2 vertices overlapped, then $\binom{n-4}{2}\binom{4}{2}$ K_4 with probability p^5.
- If there are 3 vertices overlapped, then $\binom{n-4}{1}\binom{4}{3}$ K_4 with probability p^3.
- If there are 4 vertices overlapped, then it has to be i^{th} K_4 with probability 1.

By summation,

$$\mathbb{E}\left[X \mid X_i\right] = \binom{n-4}{4}p^6 + \binom{n-4}{3}\binom{4}{1}p^6 + \binom{n-4}{2}\binom{4}{2}p^5 + \binom{n-4}{1}\binom{4}{3}p^3 + 1,$$

therefore,

$$\begin{aligned}\mathbb{E}\left[X^2\right] &= \binom{n}{4}p^6\left(\binom{n-4}{4}p^6 + \binom{n-4}{3}\binom{4}{1}p^6 + \binom{n-4}{2}\binom{4}{2}p^5 + \binom{n-4}{1}\binom{4}{3}p^3 + 1\right)\\ &= \binom{n}{4;4}p^{12} + \binom{n}{3;3;1}p^{12} + \binom{n}{2;2;2}p^{11} + \binom{n}{3;1;1}p^9 + \binom{n}{4}p^6,\end{aligned}$$

hence

$$\begin{aligned}\operatorname{Var}[X] &= \left(\binom{n}{4;4} + \binom{n}{3;3;1} - \binom{n}{4}^2\right)p^{12} + \binom{n}{2;2;2}p^{11} + \binom{n}{3;1;1}p^9 + \binom{n}{4}p^6\\ &= O(n^7p^{12}) + O(n^6p^{11}) + O(n^5p^9) + O(n^4p^6).\end{aligned}$$

When $p = o(n^{-2/3})$, by lemma 1.6.25, that $\Pr[X \geq 1] \leq \mathbb{E}[X] = o(1)$.
When $p = \omega(n^{-2/3})$, by theorem 1.6.24 and $\mathbb{E}[X]^2 = O(n^8p^{12}) = \omega(1)$, we have

$$\Pr[X = 0] \leq \frac{\operatorname{Var}[X]}{\mathbb{E}[X]^2} = o(1).$$

□

Theorem 1.6.34 (Conditional Expectation Inequality). *Let $\{X_i\}_{i\in[1,n]}$ be 0-1 random variables, and $X = \sum_i X_i$. Then*

$$\Pr[X > 0] \geq \sum_{i\in[1,n]} \frac{\Pr[X_i = 1]}{\mathbb{E}[X \mid X_i = 1]}.$$

Proof. Let $Y = 1/X$ when $X \neq 0$, and 0 otherwise. Then $\mathbb{E}[XY] = \Pr[X > 0]$, and

$$\begin{aligned}\mathbb{E}[XY] &= \sum_{i\in[1,n]} \mathbb{E}[Y \mid X_i = 1]\Pr[X_i = 1]\\ &\geq \sum_{i\in[1,n]} \frac{\Pr[X_i = 1]}{\mathbb{E}[X \mid X_i = 1]},\end{aligned}$$

where the first equality holds from exercise 1.3.1, and the last inequality holds from theorem 1.2.5. □

Remark 1.6.35. Previous exercise 1.6.33 $p = \omega(n^{-2/3})$ case can be solved with theorem 1.6.34 by

$$\begin{aligned}\Pr[X > 0] &\geq \sum_i \frac{\Pr[X_i = 1]}{\mathbb{E}[X \mid X_i = 1]} \\ &= \frac{\binom{n}{4}p^6}{\binom{n-4}{4}p^6 + \binom{n-4}{3}\binom{4}{1}p^6 + \binom{n-4}{2}\binom{4}{2}p^5 + \binom{n-4}{1}\binom{4}{3}p^3 + 1}.\end{aligned}$$

When $p = \omega(n^{-2/3})$, both $\binom{n}{4}p^6$ and $\binom{n-4}{4}p^6$ are $\omega(1)$, while the others are $o(1)$, proving the second part of the proof.

Exercise 1.6.36 (Exercise 6.13 [MU17])**.** Consider the problem of whether graphs in $G_{n,p}$ have K_k of constant size k. Suggest a threshold function and generalize the argument for K_4 using either theorem 1.6.24 or theorem 1.6.34, to prove that your threshold function is correct for K_5.

Proof. We reuse notions in exercise 1.6.33, let X_i be the random variable for the existence of the i^{th} K_k in $G_{n,p}$, and let $X = \sum_i X_i$. We guess the threshold $p = n^{-2/(k-1)}$, which balances both the dominanting numerator and denumerator terms in theorem 1.6.34, and convert $\mathbb{E}[X] = \binom{n}{k}p^{k(k-1)/2} = O(n^k p^{k(k-1)/2})$ into constant for lemma 1.6.25.

When $p = o(n^{-1/2})$, by linearity of expectation, $\mathbb{E}[X] = \binom{n}{5}p^{10} = o(1)$. By lemma 1.6.25, $\Pr[X > 0] \leq \mathbb{E}[X] = o(1)$.

When $p = \omega(n^{-1/2})$, we apply the theorem 1.6.34. Among all $\binom{n}{5}$ possible K_5, conditioned that $X_i = 1$,

- If there are no vertex overlap, then $\binom{n-5}{5}$ K_5 with probability p^{10}.
- If there is 1 vertex overlapped, then $\binom{n-5}{4}\binom{5}{1}$ K_5 with probability p^{10}.
- If there are 2 vertices overlapped, then $\binom{n-5}{3}\binom{5}{2}$ K_5 with probability p^9.
- If there are 3 vertices overlapped, then $\binom{n-5}{2}\binom{5}{3}$ K_5 with probability p^7.
- If there are 4 vertices overlapped, then $\binom{n-5}{1}\binom{5}{4}$ K_5 with probability p^4.
- If there are 5 vertices overlapped, then it has to be the i^{th} K_5 with probability 1.

By summation,

$$\mathbb{E}[X \mid X_i = 1] = \binom{n-5}{5}p^{10} + \binom{n-5}{4}\binom{5}{1}p^{10} + \binom{n-5}{3}\binom{5}{2}p^9 + \binom{n-5}{2}\binom{5}{3}p^7 + \binom{n-5}{1}\binom{5}{4}p^4 + 1.$$

Since $\binom{n-5}{5}p^{10}$ and $\binom{n}{5}p^{10}$ are $\omega(1)$, while all other terms are $o(1)$, therefore by theorem 1.6.34, $\Pr[X > 0] \to 1$. □

Exercise 1.6.37 (Exercise 6.14 [MU17])**.** Consider $G \in G_{n,p}$ with $p = c \ln n/n$. Use theorem 1.6.24 or theorem 1.6.34 to prove that if $c < 1$ then, for any constant $\varepsilon > 0$ and for n sufficiently large, the graph has isolated vertices with probability at least $1 - \varepsilon$.

Proof. Let X_i be random variable for the i^{th} vertex being isolated in the graph, and $X = \sum_i X_i$.

By theorem 1.6.34, among n vertices, conditioned that $X_i = 1$,

- If a vertex is not the i^{th} vertex, it has probability $(1-p)^{n-2}$ to be isolated.
- Otherwise, the i^{th} vertex is isolated with probability 1 by conditioning.

By summing,

$$\lim_{n\to\infty} \mathbb{E}[X \mid X_i = 1] = \lim_{n\to\infty} 1 + (n-1)(1-p)^{n-2} = 1 + \frac{n}{n^c}.$$

Since $\Pr[X_i = 1] = (1-p)^{n-1}$, which is $1/n^c$ as $n \to \infty$, therefore by theorem 1.6.34,

$$\lim_{n\to\infty} \Pr[X > 0] \geq \frac{n/n^c}{1 + n/n^c} = \frac{n}{n^c + n},$$

which approaches 1 if $c < 1$. □

Exercise 1.6.38 (Exercise 6.15 [MU17])**.** Consider a graph in $G_{n,p}$ where $p = 1/n$. Let X be the number of triangles in the graph. Show that

$$\Pr[X \geq 1] \leq \frac{1}{6},$$

and that

$$\lim_{n\to\infty} \Pr[X \geq 1] \geq \frac{1}{7}.$$

Proof. There are $\binom{n}{3}$ triangles in K_n, while a triangle exists with probability p^3. Let $\{X_i\}$ be the set of 0-1 random variables indicating if the i^{th} triangle exists in the graph, then $X = \sum_i X_i$. By linearity of expectation,

$$\mathbb{E}[X] = \sum_i \mathbb{E}[X_i] = \binom{n}{3} p^3 = \frac{(n-1)(n-2)}{6n^2} \leq \frac{1}{6}.$$

On the other hand, by lemma 1.6.25

$$\mathbb{E}[X] = \sum_{x \in \left[0, \binom{n}{3}\right]} x \Pr[X = x] \geq \sum_{x \in \left[1, \binom{n}{3}\right]} \Pr[X = x] = \Pr[X \geq 1],$$

which completes the proof for the first part.

On the other hand, by theorem 1.6.34 we have

$$\begin{aligned}\lim_{n\to\infty} \Pr[X \geq 1] &\geq \lim_{n\to\infty} \sum_{i \in \left[1, \binom{n}{3}\right]} \frac{\Pr[X_i = 1]}{\mathbb{E}[X \mid X_i = 1]} \\ &= \lim_{n\to\infty} \frac{\binom{n}{3} p^3}{\mathbb{E}[X \mid X_i = 1]}.\end{aligned}$$

For all $\binom{n}{3}$ possible triangles,

- If there are no vertex overlap, then $\binom{n-3}{3}$ triangles with probability p^3.
- If there are 1 vertex overlap, then $\binom{n-3}{2}\binom{3}{1}$ triangles with probability p^3.
- If there are 2 vertex overlap, then $\binom{n-3}{1}\binom{3}{2}$ triangles with probability p^2.
- If there are 3 vertex overlap, then 1 triangle with probability 1.

Summing, we have

$$\begin{aligned}\mathbb{E}[X \mid X_i] &= \sum_{j \in \left[1, \binom{n}{3}\right]} \mathbb{E}\left[X_j \mid X_i = 1\right] \\ &= 1 + \binom{n-3}{1}\binom{3}{2} p^2 + \binom{n-3}{2}\binom{3}{1} p^3 + \binom{n-3}{3} p^3.\end{aligned}$$

When $n \to \infty$, $\binom{n}{3} p^3 \to 1/6$ and $\mathbb{E}[X \mid X_i = 1] \to 7/6$, which proves the second part. □

Remark 1.6.39. We attempted exercise 1.6.8 from theorem 1.6.30, which is inspired by [AS16] to use the second/fourth moment method, just to avoid the bounding over binomial coefficients, but it did not bound as good.

Let $\{Y_i\}_{i\in[2n]}$ be independent Bernoulli trials from even coins, $\{Z_i\}_{i\in[n]}$ each be $Y_i - Y_{i+n}$, and $Z = \sum_i Z_i$. Since Z is symmetric over 0, then event $|Z| > c\sqrt{n}$ is equivalent to $Z^2 > c'n$.

For random variable Z^2, we want to prove $\Pr\left[Z^2 > c'n\right] \geq 1/2$ for some constant $c' > 0$, which fits theorem 1.6.30:

- The second moment is $\mathbb{E}\left[Z^2\right] = \sum_i \mathbb{E}\left[Z_i^2\right] = n/2$.
- The fourth moment is $\mathbb{E}\left[Z^4\right] = \sum_i \mathbb{E}\left[Z_i^4\right] + \sum_{i\neq j} \mathbb{E}\left[Z_i^2 Z_j^2\right] = \sum_i \mathbb{E}\left[Z_i^4\right] + \sum_{i\neq j} \mathbb{E}\left[Z_i^2\right] \mathbb{E}\left[Z_j^2\right] = n/2 + \binom{n}{2}\binom{4}{2}/4$.

Therefore, even when $\Pr\left[Z^2 > 0\right] < 1/3$ by Paley-Zygmund theorem 1.6.30, which is too loose.

1.6.6 Lovász Local Lemma

The symmetric Lovász Local Lemma requires that the bad events are all upper bounded by a probability p, and each bad event depends on at most d others.

By definition, an event E_{n+1} is mutually independent of $\{E_i\}_{i\in[1,n]}$, if for any subset $S \subseteq [1,n]$,

$$\Pr\left[E_{n+1} \,\middle|\, \bigcap_{i\in S} E_i\right] = \Pr\left[E_{n+1}\right].$$

Onwards, we denote E_S for $\bigcap_{i\in S} E_i$, and $\overline{E}_S$ for $\bigcap_{i\in S} \neg E_i$.

Definition 1.6.40 (Dependency Graph)**.** A dependency graph for events $\{E_i\}$ is a graph $G = (V, E)$, such that E_i is independent of the events $\{E_j \mid (i,j) \notin E\}$. The degree of the dependency graph is the maximum degree of any vertex in the graph.

Lemma 1.6.41 (Symmetric Lovász Local Lemma [EL75])**.** *Let $\{E_i\}_{i\in[1,n]}$ be a set of events, and the following holds:*

- $\Pr[E_i] \le p$ *for all* $i \in [1,n]$,
- *The degree of the dependency graph given by $\{E_i\}$ is at most d,*
- $4dp \le 1$,

then $\Pr\left[\overline{E}_{[1,n]}\right] > 0$.

Proof. The proving strategy is proving by induction the following lemma for all possible $S \subseteq [1,n]$.

Lemma 1.6.42. *For all $S \subseteq [1,n]$ with $k \notin S$, we have both following being true*

$$\Pr\left[E_k \,\middle|\, \overline{E}_S\right] \le 2p \text{ and } \Pr\left[\overline{E}_S\right] \ge (1-2p)^{|S|}.$$

Proof. We prove by induction over all possible $|S|$.

When $|S| = 0$, this is vacuously true.

When $|S| = 1$, let $S = \{i\}$. Then $\Pr[\neg E_i] \ge 1 - p \ge 1 - 2p$. For $\Pr[E_k \mid \neg E_i]$:

- If E_k and E_i mutually independent, namely $(i,j) \notin E$, then

$$\Pr[E_k \mid \neg E_i] = \Pr[E_k] \le p \le 2p.$$

- Otherwise $(i,k) \in E$, then

$$\Pr[E_k \mid \neg E_i] = \frac{\Pr[E_k \cap \neg E_i]}{\Pr[\neg E_i]} \le \frac{\Pr[E_k]}{\Pr[\neg E_i]} \le \frac{p}{1-p} \le 2p,$$

where the last inequality holds as $4dp \le 1$, meaning $p \le 1/4$.

WLOG let $S' = [1, s-1]$, $S = [1,s]$, and $s \ge 2$. Supposing the lemma holds for S' with $|S'| = s-1$, then for S,

$$\begin{aligned}\Pr\left[\overline{E}_S\right] &= \Pr\left[\neg E_s \,\middle|\, \overline{E}_{S'}\right] \cdot \Pr\left[\overline{E}_{S'}\right] \\ &= \left(1 - \Pr\left[E_s \,\middle|\, \overline{E}_{S'}\right]\right) \cdot \Pr\left[\overline{E}_{S'}\right] \\ &\ge (1-2p)\cdot(1-2p)^{s-1} = (1-2p)^s,\end{aligned}$$

where the lower bound follows immediately from the induction hypothesis.

Remark 1.6.43. Alternatively, we have

$$\begin{aligned}\Pr\left[\overline{E}_S\right] &= \Pr\left[\neg E_s \,\middle|\, \overline{E}_{S'}\right] \cdot \Pr\left[\overline{E}_{S'}\right] \\ &= \Pr\left[\neg E_s \,\middle|\, \overline{E}_{S'}\right] \cdot \prod_{i\in[1,s-1]} \Pr\left[\neg E_i \,\middle|\, \overline{E}_{[1,i-1]}\right] \\ &= \prod_{i\in[1,s]} \left(1 - \Pr\left[E_i \,\middle|\, \overline{E}_{[1,i-1]}\right]\right) \\ &\geq (1-2p)^s,\end{aligned}$$

where the last inequality holds from induction hypothesis.

Let $A \subseteq S$ such that $\forall i \in A$, $(i, s+1) \in E$, and $B = S \setminus A$. If $B = S$, then E_{s+1} is mutually independent of $\{E_i\}_{i\in S}$,

$$\Pr\left[E_{s+1} \,\middle|\, \overline{E}_S\right] = \Pr\left[E_{s+1}\right] \leq p \leq 2p.$$

Otherwise, by Bayes' rule,

$$\begin{aligned}\Pr\left[E_{s+1} \,\middle|\, \overline{E}_S\right] &= \frac{\Pr\left[E_{s+1} \cap \overline{E}_S\right]}{\Pr\left[\overline{E}_S\right]} = \frac{\Pr\left[E_{s+1} \cap \overline{E}_A \cap \overline{E}_B\right]}{\Pr\left[\overline{E}_A \cap \overline{E}_B\right]} \\ &= \frac{\Pr\left[E_{s+1} \cap \overline{E}_A \,\middle|\, \overline{E}_B\right]}{\Pr\left[\overline{E}_A \,\middle|\, \overline{E}_B\right]}.\end{aligned}$$

For numerator, since E_{s+1} is mutually independent of $\{E_i\}_{i\in B}$ by the dependency graph, we have

$$\Pr\left[E_{s+1} \cap \overline{E}_A \,\middle|\, \overline{E}_B\right] \leq \Pr\left[E_{s+1} \,\middle|\, \overline{E}_B\right] = \Pr\left[E_{s+1}\right] \leq p.$$

For denumerator,

$$\begin{aligned}\Pr\left[\overline{E}_A \,\middle|\, \overline{E}_B\right] &= 1 - \Pr\left[\bigcup_{i\in A} E_i \,\middle|\, \overline{E}_B\right] \\ &\geq 1 - \sum_{i\in A} \Pr\left[E_i \,\middle|\, \overline{E}_B\right] \\ &\geq 1 - d \cdot 2p \\ &\geq \frac{1}{2},\end{aligned}$$

where the first inequality holds from union bound, the second inequality holds from the induction hypothesis, and $|A| \leq d$ by the degree of the dependency graph, and the last inequality holds from the lemma's requirement. Hence, $\Pr\left[E_{s+1} \,\middle|\, \overline{E}_S\right] \leq 2p$. □

Since lemma 1.6.42 holds for any S, which finishes the proof immediately. □

Exercise 1.6.44 (Exercise 6.17 [MU17])**.** Use lemma 1.6.41 to show that, if

$$4 \cdot \binom{k}{2}\binom{n}{k-2} \cdot 2^{1-\binom{k}{2}} \leq 1,$$

then it is possible to 2-color the edges of K_n so that it has no monochromatic K_k subgraph.

Proof. Let E_i be the bad event that the i^{th} K_k being monochromatic.

To upper bound the dependency degree d, we can pick an edge from the i^{th} K_k, then uniformly randomly sample the rest of the vertices from all possible vertices. Hence

$$d \leq \binom{k}{2}\binom{n}{k-2}.$$

The probability of K_k being monochromatic is $2^{1-\binom{k}{2}}$. Therefore by lemma 1.6.41, if

$$4 \cdot \binom{k}{2}\binom{n}{k-2} \cdot 2^{1-\binom{k}{2}} \leq 1,$$

then it is possible to 2-color K_n edges such that there is no monochromatic K_k. □

The symmetric Lovász Local Lemma lemma 1.6.41 works best when all events are upper bounded equally, and the dependency degrees are also upper bounded equally. The asymmetric Lovász Local Lemma has likely or rare events, where likely events have smaller dependency degrees, while rare events could depend on many other events.

Onwards we denote $\Gamma_i = \{j \mid (i,j) \in E\}$ for all the neighbors of E_i in the dependency graph.

Lemma 1.6.45 (Asymmetric Lovász Local Lemma [EL75])**.** *Let $\{E_i\}_{i\in[1,n]}$ be a set of events. If $\{x_i\}_{i\in[1,n]}$ in $[0,1]$ has*

$$\Pr[E_i] \leq x_i \cdot \prod_{j\in\Gamma_i}(1-x_j),$$

then

$$\Pr\left[\overline{E}_{[1,n]}\right] \geq \prod_{i\in[1,n]}(1-x_i).$$

Proof. The proving strategy is similar to lemma 1.6.41 by introducing following lemma for all possible $S \subseteq [1,n]$.

Lemma 1.6.46. *For all $S \subseteq [1,n]$ with $k \notin S$, we have both following being true*

$$\Pr\left[E_k \,\middle|\, \overline{E}_S\right] \leq x_k \text{ and } \Pr\left[\overline{E}_S\right] \geq \prod_{i\in S}(1-x_i).$$

Proof. We prove by induction over all possible $|S|$.

When $|S| = 0$, this is vacuously true.

When $|S| = 1$, let $S = \{i\}$. Then

$$\Pr[\neg E_i] \geq 1 - x_i \cdot \prod_{j\in\Gamma_i}(1-x_j) \geq 1 - x_i.$$

For $\Pr[E_k \mid \neg E_i]$,

- If E_k and E_i are mutually independent, namely $(i,k) \notin E$, then

$$\Pr[E_k \mid \neg E_i] = \Pr[E_k] \leq x_k \cdot \prod_{j\in\Gamma_k}(1-x_j) \leq x_k.$$

- Otherwise $(i,k) \in E$, then

$$\Pr[E_k \mid \neg E_i] = \frac{\Pr[E_k \bigcap \neg E_i]}{\Pr[\neg E_i]} \leq \frac{\Pr[E_k]}{\Pr[\neg E_i]} \leq x_k \cdot \prod_{j\in\Gamma_k\setminus\{i\}}(1-x_j) \leq x_k,$$

  where the second inequality holds from $\Pr[\neg E_i] \geq 1 - x_i$.

WLOG let $S' = [1, s-1]$, $S = [1, s]$, and $s \geq 2$. Supposing the lemma holds for any S' with $|S'| = s-1$, then for S,

$$\Pr\left[\overline{E}_S\right] = \Pr\left[\neg E_s \,\middle|\, \overline{E}_{S'}\right] \cdot \Pr\left[\overline{E}_{S'}\right] \geq (1-x_s) \cdot \prod_{i\in S'} (1-x_i) = \prod_{i\in S} (1-x_i)\,,$$

which is immediate from induction hypothesis. Alternatively, the same procedure in remark 1.6.43 applies.

Let $A \subseteq S$ such that $\forall i \in A$, $(i, s+1) \in E$, and $B = S \setminus A$. If $B = S$, then E_{s+1} is mutually independent of $\{E_i\}_{i\in S}$,

$$\Pr\left[E_{s+1} \,\middle|\, \overline{E}_S\right] = \Pr\left[E_{s+1}\right] \leq x_{s+1} \cdot \prod_{j\in\Gamma_{s+1}} \left(1-x_j\right) \leq x_{s+1}.$$

Otherwise, by Bayes' rule,

$$\begin{aligned}\Pr\left[E_{s+1} \,\middle|\, \overline{E}_S\right] &= \frac{\Pr\left[E_{s+1} \cap \overline{E}_S\right]}{\Pr\left[\overline{E}_S\right]} = \frac{\Pr\left[E_{s+1} \cap \overline{E}_A \cap \overline{E}_B\right]}{\Pr\left[\overline{E}_A \cap \overline{E}_B\right]} \\ &= \frac{\Pr\left[E_{s+1} \cap \overline{E}_A \,\middle|\, \overline{E}_B\right]}{\Pr\left[\overline{E}_A \,\middle|\, \overline{E}_B\right]}.\end{aligned}$$

For numerator, since E_{s+1} is mutually independent of $\{E_i\}_{i\in B}$ by the dependency graph, we have

$$\Pr\left[E_{s+1} \cap \overline{E}_A \,\middle|\, \overline{E}_B\right] \leq \Pr\left[E_{s+1} \,\middle|\, \overline{E}_B\right] = \Pr\left[E_{s+1}\right] \leq x_{s+1} \cdot \prod_{j\in\Gamma_{s+1}} \left(1-x_j\right).$$

For denumerator, let $|A| = t$, $A = \{a_i\}_{i\in[1,t]}$, and $A_\ell = \{a_i\}_{i\in[\ell,t]}$, then

$$\begin{aligned}\Pr\left[\overline{E}_A \,\middle|\, \overline{E}_B\right] &= \prod_{i\in[1,t]} \Pr\left[\neg E_{a_i} \,\middle|\, \overline{E}_{A_{i+1}} \cap \overline{E}_B\right] \\ &= \prod_{i\in[1,t]} \left(1 - \Pr\left[E_{a_i} \,\middle|\, \overline{E}_{A_{i+1}} \cap \overline{E}_B\right]\right) \\ &\geq \prod_{i\in A} (1-x_i) \\ &\geq \prod_{i\in\Gamma_{s+1}} (1-x_i),\end{aligned}$$

where the first inequality holds from induction hypothesis, and the second inequality holds from $A \subseteq \Gamma_{s+1}$. Hence $\Pr\left[E_{s+1} \,\middle|\, \overline{E}_S\right] \leq x_{s+1}$. □

Since lemma 1.6.46 holds for any S, we finish the proof. □

Lemma 1.6.47. *For integer $x \geq 0$,*

$$g(x) = \left(1 - \frac{1}{1+x}\right)^x$$

is monotone decreasing and lower bounded by e^{-1}.

Proof. Let $\varphi(x) = x \ln\left(1 - 1/(1+x)\right)$, then the first order derivative

$$\varphi'(x) = \ln\left(1 - \frac{1}{1+x}\right) + x \cdot \frac{1}{1-1/(1+x)} \cdot \left(-\frac{-1}{(1+x)^2}\right) = \ln\left(1 - \frac{1}{1+x}\right) + \frac{1}{1+x}.$$

Let $t = 1 - 1/(1+x)$, then $\varphi'(x) = \ln t + 1 - t$.

Since $\ln t \leq t - 1$ for all t, then $\varphi'(x) \leq 0$, $g(x)$ is monotone decreasing. As $x \to \infty$, $g(x) \to e^{-1}$. □

Exercise 1.6.48 (Exercise 6.18 [MU17]). Use the asymmetric LLL lemma 1.6.45 to show we can improve the symmetric LLL lemma 1.6.41 by replacing the condition $4pd \le 1$ to $ep(d+1) \le 1$.

Proof. Let each $x_i = 1/(d+1)$, then

$$x_i \cdot \prod_{j \in \Gamma_i} (1 - x_j) \ge x_i \cdot (1 - x_j)^d = \frac{1}{d+1} \cdot \left(1 - \frac{1}{d+1}\right)^d \ge \frac{1}{e(d+1)} \ge p,$$

where the second inequality holds from lemma 1.6.47.

p is directly applied in lemma 1.6.41, and the weakened condition is immediate. □

Remark 1.6.49. Spencer proved this strengthened result in [Spe77].

Exercise 1.6.50 (Exercise 6.19 [MU17]). Let $G = (V, E)$ be an undirected graph and suppose each $v \in V$ is associated with a set $S(v)$ of $8r$ colors, where $r \ge 1$. Suppose, in addition, that for each $v \in V$ and $c \in S(v)$, there are at most r neighbors u of v with c lies in $S(u)$. Prove that there exists a coloring of G assigning to each vertex v a color from $S(v)$ such that, for any edge $(u, v) \in E$, the color assigned to u and v are different. Hint: use $A_{u,v,c}$ to be the event that u and v are both colored with c, then consider the family of such events.

Proof. Let X_u be the random number for vertex u's color, and $c \in S(u)$, then $\Pr[X_u = c] = 1/8r$.

Let v be a neighboring vertex of u, since v is also associated with a set $S(v)$ of $8r$ colors, then $\Pr[X_v = c] \le 1/8r$, as c might not be in $S(v)$. Therefore,

$$\Pr[X_u = c \wedge X_v = c] = \Pr[A_{u,v,c}] \le \frac{1}{64r^2}.$$

For the dependency graph formed by events $\{A_{u,v,c}\}$, fixing a color $c \in S(u) \cap S(v)$ and an edge $(u, v) \in E$, the event $A_{u,v,c}$ is dependent on all $A_{u',v',c'}$ where either $u' = u$ or $v' = v$. Since u has at most $8r$ colors, and at most $8r^2$ neighbors may share colors with u, therefore the dependency graph degree $d \le 16r^2$ by symmetry on v side.

By lemma 1.6.41, by $4dp \le 4 \cdot 16r^2 \cdot 1/64r^2 \le 1$, there is probability that all edges are colored differently by

$$\Pr\left[\bigcap_{u,v,c} \neg A_{u,v,c}\right] > 0.$$

□

Exercise 1.6.51 (Exercise 6.20 [MU17]). A k-uniform hypergraph is an ordered pair $G = (V, E)$, but edges consist of sets of k (distinct) vertices, instead of just 2. (So a 2-uniform hypergraph is just what we normally call a graph.) A hypergraph is k-regular if all vertices have degree k; that is, they are in k hypergraph edges.

Show that for sufficiently large k, the vertices of a k-uniform, k-regular hypergraph can be 2-colored so that no edge is monochromatic. What's the smallest value of k you can achieve?

Proof. We form a dependency graph for events of a hypergraph edge being monochromatic. Since the hypergraph is k-regular, then each vertex is appearing in k hypergraph edges. Excluding the hypergraph edge we are discussing, there are $k - 1$ other hypergraph edges for the vertex, and therefore the dependency graph degree $d \le k(k-1)$.

On the other hand, a k-uniform hypergraph edge is monochromatic with probability 2^{1-k}.

- We can try lemma 1.6.41 and by $4dp \le 1$, then $k \ge 10$.
- We can also try the Spencer result in exercise 1.6.48 that $ep(d+1) \le 1$, then $k \ge 9$.

□

1.6.7 Algorithmic Lovász Local Lemma

Theorem 1.6.52 (Beck's k-Satisfiability Algorithm [Bec91])**.** *Consider a k-SAT formula with m clauses, where $k \geq 12$ is an even constant, and each variable appears in up to $2^{\alpha k}$ clauses for a sufficiently small constant $\alpha > 0$, then there is an algorithm finding a satisfying assignment for the formula in expected polynomial time in m.*

Proof. Let $\{x_i\}_{i\in[1,\ell]}$ be the variables, and $\{C_i\}_{i\in[1,m]}$ be the m clauses.

The algorithm has 2 phases: some variables are fixed in the first phase, and the remaining variables are deferred to the second phase. During the first phase variable fixing, a clause C_i is considered *dangerous* if C_i is not yet satisfied, and $k/2$ variables in the clause have been fixed.

In the first phase, the algorithm iterates through all variables, if the variable is not in any dangerous clause, assign it independently and uniformly a value from $\{0, 1\}$.

After the first phase, a clause is *surviving* if it is not yet satisfied, and such surviving clause has no more than $k/2$ variables fixed. A *deferred* variable is a variable not yet fixed in the first phase. In phase 2, we use exhaustive search to assign values to the deferred variables and so to complete a satisfying assignment for the formula.

Remark 1.6.53. The algorithm assigns a subset of variables in the first phase, and the remaining variables are deferred to the second phase. The subset of variables with values assigned in the first phase are chosen such that

- By the Local Lemma, the random partial solution fixed in phase 1 can be extended to a full satisfying assignment without modifying any phase 1 variable assignments.
- The dependency graph H' defined by the deferred variables in phase 2 is with high probability containing small connected subcomponents.

When the dependency graph consists only of small connected subcomponents, a solution for the variables of one component can be found independently of the other components. Therefore, such 2 phase algorithm first breaks the problem into smaller subproblems, and each smaller subproblem can be solved independently with exhaustive search in the phase 2.

We introduce following 2 lemmas for the 2 points in the prior remark.

Lemma 1.6.54. *There is an assignment to the deferred variables such that all the surviving clauses are satisfied.*

Proof. Let $H = (V, E)$ be the dependency graph on m nodes, where $V = [1, m]$, and $(i, j) \in E$ if and only if $C_i \cap C_j \neq \emptyset$. H is the dependency graph for the original problem. Let $H' = (V', E')$ be the dependency graph with $V' \subseteq V, E \subseteq E$, with $i \in V'$ if and only if C_i is a surviving clause, and $(i, j) \in E'$ if and only if C_i and C_j share deferred variables.

Since a surviving clause has at least $k/2$ deferred variables, then the satisfying probability is at most $2^{-k/2}$. On the other hand, a variable appears up to $2^{\alpha k}$ clauses, then the degree of the dependency graph H' is $d \leq k2^{\alpha k}$.

By symmetric Local Lemma lemma 1.6.41, given constant $\alpha > 0$ that is sufficiently small, if $4 \cdot k2^{\alpha k} \cdot 2^{-k/2} \leq 1$, there exists an assignment that all surviving clauses are satisfied. □

Lemma 1.6.55. *All connected components in H' are of size $O(\log m)$ with probability $1 - o(1)$.*

Proof. We start by bounding the probability of a given clause surviving phase 1. For a given node, it is dangerous with probability at most $2^{-k/2}$, that exactly $k/2$ variables in phase 1 are fixed but the clause is not satisfying. A clause is surviving, either itself is dangerous, or it is sharing at least a deferred variable with a dangerous clause, namely it is neighboring a dangerous clause in H'. Union bounding over the clause and its neighbors in H', which is at most the dependency graph degree $d \leq k2^{\alpha k}$, a given node surviving phase 1 with probability at most $(d + 1)2^{-k/2}$.

The survival of individual clauses are not independent of each other, as 2 neighboring surviving clauses share at least a deferred variable. We consider a connected component $R \subseteq H$ of r vertices, then we identify a subset of vertices in R such that the survival of the clauses represented by these vertices are independent events.

Definition 1.6.56. A 4-tree S of a connected component R in H is defined as follows:

- S is a rooted tree.
- Any 2 nodes in S is at least distance 4 in H.
- There can be an edge in S only between 2 vertices of distance exactly 4 between them in H.
- Any node in R is either in S or in distance at most 3 from a node in S.

The motivation of introducing 4-tree is that, for any 2 clauses on S, the survival of the clauses are independent. Since any 2 neighboring clauses $u, v \in S$ are of distance exactly 4 in H, if both u and v are surviving, the clause causing u to survive is at least distance 2 in H from the clause causing v to survive. If 2 clauses are at least distance 2 in H, they are not sharing any variables, and their survival are surely independent. Supposing the survival of a large 4-tree is with low probability, this implies the survival of a large connected subcomponent with low probability.

For a 4-tree to survive, the probability is at most $\left((d+1)2^{-k/2}\right)^{|S|}$. For a clause in a 4-tree, there are at most

$$d + d(d-1) + d(d-1)^2 \leq d^3 - 1$$

clauses in H that are within distance 3. Immediate by the expectation argument, there exists a 4-tree S in R with $|R| = r$ such that $|S| \geq r/d^3$.

The next step is to union bound over all possible sufficiently large 4-tree, such that we can derive an upper bound for the probability of existence of a sufficiently large 4-tree. A crude way to upper bound the number of 4-trees is, for each root, traverse the 4-tree in Euler cycle by visiting each path in 2 directions. There are at most $m\left(d^4\right)^{2r/d^3}$ number of traversals by picking a vertex every 4 vertices until $2r/d^3$ vertices are picked, over m roots.

The union bound is

$$md^{8r/d^3} \cdot \left((d+1)2^{-k/2}\right)^{r/d^3},$$

when $r \geq c \log m$ for a sufficiently large constant $c > 0$ and a sufficiently small constant $\alpha > 0$, the bound is $o(1)$. □

Therefore, the proof is immediate by lemma 1.6.54 and lemma 1.6.55, that phase 1 partitions the problem into subproblems with $O(k \log m)$ variables with probability $1 - o(1)$, then we just need to run constant number of times of phase 1 before we obtain a good partition. □

We now discuss the constructive proof by Moser [Mos09]. A simpler version was presented in his STOC'09 talk, which was based on the entropy compression argument. The following analysis follows the "entropy compression" idea, which appears explicitly in his doctoral thesis [Mos12] (see the "Incompressibility" section), rather than being the primary framing in the STOC'09 paper.

At the core of the "entropy compression" argument, it is arguing that the algorithm's expected runtime must be small, otherwise we are able to compress the random bits used by the algorithm significantly.

We begin with a useful fact on the incompressibility of random strings.

Fact 1.6.57 (Incompressibility Principle)**.** For injective $f : \{0,1\}^t \mapsto \{0,1\}^*$, if $\mathbf{s} \xleftarrow{R} \{0,1\}^t$, then for any integer c,

$$\Pr\left[\text{the length of } f(\mathbf{s}) \leq t - c\right] \leq 2^{1-c}.$$

Proof. There are in total $2^{t-c+1} - 1$ binary strings of length at most $t - c$, while there are 2^t possible inputs in $\{0,1\}^t$. Hence, if $\mathbf{s} \xleftarrow{R} \{0,1\}^t$, then the probability of f outputting at most $t - c$ bits is at most 2^{1-c}. □

There are $2^t - 1$ binary strings of length at most $t - 1$, and the average length is

$$\sum_{i \in [0,t-1]} i \cdot \frac{2^i}{2^t - 1} \geq \sum_{i \in [0,t-1]} i \cdot \frac{2^i}{2^t} \geq t - 2.$$

Therefore, the expected compression length over $\mathbf{s} \xleftarrow{R} \{0,1\}^t$ is lower bounded by $t - 2$.

Theorem 1.6.58 (Moser's k-Satisfiability Algorithm [Mos12]). *For a k-SAT formula with m clauses, and each clause shares variables with at most $2^{k-3}-1$ other clauses. Then a satisfying assignment can be found in expected polynomial time in m.*

The "entropy compression" argument is rather clean in the specific setting of k-SAT.

We let $\{x_i\}_{i\in[1,\ell]}$ be the ℓ variables, $\{C_i\}_{i\in[1,m]}$ be the m clauses, Γ_i be the set of indices of clauses sharing variables with the clause C_i, Γ_i^+ be the inclusive neighborhood $\Gamma_i \cup \{i\}$, and V_i be the set of indices of variables in C_i.

Informally, the algorithm picks a uniformly random assignment, then look for an unsatisfied C_i. If such C_i exists, we sample a new assignment for the variables in V_i uniformly at random. Doing so may fix C_i, but it may end up unsatisfying C_j for some $j \in \Gamma_i^+$. The algorithm recursively fix the neighboring clauses, and by the end of the recursion, C_i is satisfied with no neighboring clauses damaged. Therefore, the situation is improved by satisfying at least 1 previously unsatisfied clause. Iterating through all m clauses, the algorithm finishes the fixing.

This leads to the question that when will the recursion stop, and is solved by the "entropy compression" argument. We describe the algorithm formally as follows, and bound the expected runtime.

The Moser's algorithm has 2 functions: `solve` and `local-fix`.

- `solve`
  - Sample a uniformly random variable assignment for $\{x_i\}_{i\in[1,\ell]}$.
  - While there is C_i unsatisfied
    - Choose the unsatisfied C_i with the smallest i.
    - Enter i into the execution log with $\lceil \log_2 m \rceil$ bits.
    - Call `local-fix` on the clause C_i.
- `local-fix`
  - Sample a uniformly random variable assignment for $\{x_j\}_{j\in\Gamma_i}$.
  - While there is $k \in \Gamma_i^+$ such that C_k is unsatisfied
    - Choose the unsatisfied C_k with the smallest k.
    - Enter "0" together with k in $k-3$ bits into the execution log.
    - Call `local-fix` on the clause C_k.
    - Enter "1" into the execution log.

Proof. There are 2 distinct ways of describing how the algorithm works.

We can think of the algorithm as being described by the random string of bits it used. It takes ℓ bits for the initial variable assignment at random, then it takes k bits to resample variables for a clause each time `local-fix` is called. We refer to each time `local-fix` is called as a round. Hence, a way to describe the algorithm's action for j rounds is with the random strings of $\ell + jk$ bits used by the algorithm.

We can also think of the algorithm as being described as the history, the bits in its execution log. It includes the list of clauses called in the main routine and recursion of `local-fix`. For recursion, the log starts with a bit 0 and ends with a bit 1 to mark the start and the end of of recursion calls. For the index of the clause in the recursion call, we use $k-3$ bits instead of $\lceil \log_2 m \rceil$ bits, as $\left|\Gamma_i^+\right| \leq 2^{k-3}$. The ℓ bits of final variable assignments are also included in the history, serving as an updatable state: each time a clause's variables being resampled, the state is updated. Hence, the output execution log for the algorithm's action for j rounds takes at most $\ell + m\lceil \log_2 m \rceil + j(k-1)$, as a round of clause variable resampling outputs $k-1$ bits: 2 bits for start/end marking, and $k-3$ bits for clause index.

Given $\ell + kj$ resampling bits used by the algorithm, and $\ell + m\lceil \log_2 m \rceil + j(k-1)$ bits of output execution log, we can view the algorithm as a compression algorithm over random bits. We now show how to recover the resampling bits from the execution log. The execution log determines a sequence of clauses visited by the algorithm, and each visited clause must be unsatisfied before resampling. Only 1 setting of variables unsatisfy a clause. Hence, we work backwards from the final variable assignments. On each clause, we update with unsatisfying variable assignment, such that we know the resampling variables, and therefore we recover the $\ell + jk$ bits of resampling bits.

After proving the compression is injective, now we use the compression argument to bound the expected runtime. Let J be the random variable for the number of rounds in the algorithm. The algorithm takes $\ell + Jk$ uniformly random bits as inputs, outputting at most $\ell + m\lceil \log_2 m \rceil + J(k-1)$ bits of history. By fact 1.6.57,

$$\ell + m\lceil \log_2 m \rceil + J(k-1) \geq \ell + kJ - 1,$$

and when $J = m\lceil \log_2 m \rceil + 1 + i$,

$$\Pr\left[\text{output length} \leq \ell + m\lceil \log_2 m \rceil + J(k-1)\right] \leq 2^{-i}.$$

Therefore, $\mathbb{E}[J]$ can be bounded as follows

$$\mathbb{E}[J] \leq m\lceil \log_2 m \rceil + 1 + \sum_{i \geq 0} \frac{i}{2^i} = m\lceil \log_2 m \rceil + 3.$$

Therefore, the expected runtime of the algorithm is $O(m \log m)$. □

Exercise 1.6.59 (Exercise 6.20 [MU17]). In the k-Satisfiability Algorithm using the algorithmic Lovász Local Lemma, we used $\lceil \log_2 m \rceil$ bits in history to represent each clause called in the main routine. Instead, we could record in the history which clauses are initially unsatisfied with an array of m bits. Explain any other changes to the algorithm to properly record a history that one can "reverse" to obtain the initial assignment, and explain how this allows one to modify the proof of theorem 1.6.58 so that only $O(m)$ rounds are needed in expectation.

Proof. Observing each run of `local-fix`, the subroutine visits an unsatisfied clause, resamples its variables, then fix neighboring unsatisfied clauses recursively. The invariant is, all of the visited clauses and their neighboring clauses are satisfied by the end of the recursion.

If we put m bits for the initially unsatisfied clauses, then we know the first unsatisfied clause visited by `local-fix`. We maintain a list of satisfied clauses by the first call to the `local-fix`, update the m bits for unsatisfied clauses, and move on to the remaining unsatisfied clause with smallest index. Repeating $O(m)$ times, we have a full list of visited clauses from history.

Therefore, we use m bits in history for visited unsatisfied clauses from the main routine rather than $m\lceil \log_2 m \rceil$, and therefore replacing all appearances of $m\lceil \log_2 m \rceil$ in the argument, we have the expected runtime being $O(m)$. □

The Algorithmic Lovász Local Lemma by Moser and Tardos [MT10] states that: in the variable-based LLL setting, an LLL existence proof is automatically (Las Vegas) constructive via resampling, with expected polynomial runtime. We describe the asymmetric algorithmic LLL, and the symmetric algorithmic LLL follows from exercise 1.6.48.

Theorem 1.6.60 (Asymmetric Algorithmic Lovász Local Lemma [MT10]). *Let $\{E_i\}_{i\in[1,n]}$ be a set of events in arbitrary probability space that are determined by mutually independent random variables $\{y_i\}_{i\in[1,\ell]}$, and let $G = (V, E)$ be the dependency graph for the events. Suppose there are $\{x_i\}_{i\in[1,n]}$ in $[0,1]$ such that for $i \in [1,n]$,*

$$\Pr[E_i] \leq x_i \cdot \prod_{j\in\Gamma_i}(1 - x_j),$$

then there exists an assignment $\{y_i\}$ such that $\overline{E}_{[1,n]}$ holds, and a resampling algorithm whose expected number of resampling E_i in finding an assignment is at most $x_i/(1-x_i)$. Hence, the expected runtime is at most $\sum_i x_i/(1-x_i)$.

Corollary 1.6.61 (Symmetric Algorithmic Lovász Local Lemma [MT10]). *Let $\{E_i\}_{i\in[1,n]}$ be a set of events in arbitrary probability space that are determined by mutually independent random variables $\{y_i\}_{i\in[1,\ell]}$, and let $G = (V, E)$ be the dependency graph for the events. Supposing the degree of G is at most d, $\Pr[E_i] \leq p$ for $i \in [1,n]$, and $ep(d+1) \leq 1$, then there exists an assignment $\{y_i\}$ such that $\overline{E}_{[1,n]}$ holds, and a resampling algorithm whose expected number of resampling E_i in finding an assignment is at most $1/d$. Hence, the expected runtime is at most n/d.*

The Moser-Tardos algorithmic LLL is even simpler than Moser's k-SAT algorithm, and the algorithm is described as follows. Onwards we say each event E_i as bad events, and the algorithms avoids all the bad events, eventually.

- `Moser-Tardos`
  - Sample a uniformly random variable assignments for $\{y_i\}$.
  - While there is E_i happening
    - Randomly pick an happening E_i.
    - Resample all y_i determines this E_i.

We define $\mathbf{c} \in [1, n]^k$ as the execution log, where the algorithm has k rounds of resampling over events $\{E_i\}$, and each c_i is an index for n events $\{E_i\}$. At the core of the argument is the construction and analysis of a witness tree from resampling each event E_{c_i} in the execution log.

Definition 1.6.62 (Witness Tree)**.** A witness tree is constructed from execution log $\mathbf{c}$ as follows.

- Let c_k be the root.
- Let $T_{j,k}$ be a witness tree from $c_j, \ldots, c_k$, and let j goes backwards from $k-1$ to 1.
  - If $c_j \in \Gamma_o^+$ for some $o \in T_{j+1,k}$, then add c_j as the children of the deepest occurrence of such event.
  - If E_{c_j} do not share variables with any event in $T_{j+1,k}$, then drop c_j and move on by $T_{j,k} \leftarrow T_{j+1,k}$.
- Eventually, $T_k \leftarrow T_{1,k}$.

The following facts are immediate by the construction of the witness tree.

Fact 1.6.63. For any events E_i and E_j sharing variables, they are appearing in the same depth of a witness tree.

Fact 1.6.64. A variable x_i is not appearing twice in the same depth of a witness tree.

Intuitively, a witness tree is a tree justifying how we end up selecting E_{c_i} to resample all variables in V_{c_i}.

We introduce an additional fact, that the witness trees from different prefixes of an execution log are different.

Fact 1.6.65. For execution logs $\mathbf{c}_0$ of length k_0 and $\mathbf{c}_1 = \mathbf{c}_0 \parallel \mathbf{c}'$ of length k_1, the execution trees T_{k_0} and T_{k_1} are different.

Proof. Let the roots from $\mathbf{c}_0$ and $\mathbf{c}_1$ be r_0 and r_1.

If $r_0 \neq r_1$, then $T_{k_0} \neq T_{k_1}$ by root. Otherwise, $\mathbf{c}_1$ has at least one more occurrence of r_0, resulting in $T_{k_0} \neq T_{k_1}$. □

Hence, the next lemma for counting the occurrence of resampling E_i in $\mathbf{c}$ is immediate by fact 1.6.65.

Lemma 1.6.66. *Let X_i be the random variable of occurrence of resampling E_i in $\mathbf{c}$, and $X_{i,j}$ be the 0-1 random variable indicating the witness tree T_j from $\mathbf{c}_{[1,j]}$, the prefix of $\mathbf{c}$, has root i. Then*

$$X_i = \sum_{j \in [1,k]} X_{i,j}.$$

Moreover, let W_i be the set of witness trees rooted by i, then by linearity of expectation,

$$\mathbb{E}[X_i] = \sum_{j \in [1,k]} \mathbb{E}\left[X_{i,j}\right] = \sum_{j \in [1,k]} \sum_{T \in W_i} \Pr\left[T = T_j\right] = \sum_{T \in W_i} \sum_{j \in [1,k]} \Pr\left[T = T_j\right].$$

We want to bound $\Pr\left[T_j = T\right]$. We say T appears in $\mathbf{c}$, if $T = T_j$ for T_j constructed from some $\mathbf{c}_{[1,j]}$ prefix of $\mathbf{c}$. Hence, fixing a T, since $\{T_j\}$ are all distinct by fact 1.6.65, $T = T_j$ are all independent, and

$$\Pr\left[T \text{ appears in } \mathbf{c}\right] = \sum_{j \in [1,k]} \Pr\left[T = T_j\right].$$

Lemma 1.6.67.

$$\Pr\left[T \text{ appears in } \mathbf{c}\right] \leq \prod_{j \in T} \Pr\left[E_j\right],$$

where the T in the right hand side production is a multiset, such that we capture multiple occurrence of an event being resampled.

Proof. If we traverse the witness tree T in reverse BFS method, namely layering by depth bottom up, then we observe earlier events resampling induces later events resampling, following the Moser-Tardos algorithm's execution order.

For variables $\{y_i\}$, each resampling creates a sequence of assignments, say $y_i^1, \ldots, y_i^s$. By fact 1.6.64, we know that y_i^a and y_i^b with $a \neq b$ should be on different depth.

Hence, fixing the randomness from Moser-Tardos algorithm, which is fixing the assignment sequences $\{\{y_i^s\}_s\}_i$, we form a checking algorithm for T as follows. The check algorithm traverses in described way, and checks if events in the current layer are happening: aborts if any event is not happening. Before traversing the next depth, the check algorithm resamples all the variables of the events in the current layer, by moving each y_i^s to y_i^{s+1}.

It becomes obvious that, on input the witness tree T from a successful run of Moser-Tardos algorithm, by fixing the assignment sequences for $\{\{y_i^s\}_s\}_i$, the tree checking algorithm always passes. Therefore

$$\Pr\left[T \text{ appears in } \mathbf{c}\right] \leq \Pr\left[\text{check}(T) \text{ succeeds}\right].$$

To bound the probability of a tree being valid in the checking algorithm, we use fact 1.6.63 and fact 1.6.64, that the events are not sharing variables in a same depth, such that events in a layer are independent, and each layer is independent of each other. Hence, a tree is valid in the checking algorithm with probability $\prod_{j \in T} \Pr\left[E_j\right]$.

Thus, the inequality holds by the coupling arguments above. □

Now we have derived

$$\mathbb{E}\left[X_i\right] = \sum_{T \in W_i} \Pr\left[T \text{ appears in } \mathbf{c}\right] \leq \sum_{T \in W_i} \prod_{j \in T} \Pr\left[E_j\right]$$

by lemma 1.6.66 and lemma 1.6.67, the final step would be bounding the right hand side sum of product. We derive the upper bound by Galton-Watson process, that each children of E_i in Γ_i^+ spawns with probability x_i independently.

Lemma 1.6.68. *Deriving T rooted with i with Galton-Watson process has probability*

$$p_T = \frac{1 - x_i}{x_i} \cdot \prod_{j \in T} \Pr\left[E_j\right].$$

Proof. Let each node E_j has a set $S_j \subseteq \Gamma_j^+$ for the indices of children in T, and the probability derivation follows

$$\begin{aligned} p_T &= \prod_{j \in T} \left(\prod_{k \in S_j} x_k \cdot \prod_{k \in \Gamma_j^+ \setminus S_j} (1 - x_k) \right) = \prod_{j \in T} \left(\prod_{k \in S_j} \frac{x_k}{1 - x_k} \cdot \prod_{k \in \Gamma_j^+} (1 - x_k) \right) \\ &= \frac{1 - x_i}{x_i} \cdot \prod_{j \in T} \left(x_j \cdot \prod_{k \in \Gamma_j} (1 - x_k) \right) = \frac{1 - x_i}{x_i} \cdot \prod_{j \in T} \Pr\left[E_j\right], \end{aligned}$$

where the third equality holds from moving all the $x_k/(1 - x_k)$ terms for $k \in S_j$ from E_j to all the children in S_j. □

The proof for theorem 1.6.60 follows immediately:

Proof. By lemma 1.6.68, we have

$$\mathbb{E}\left[X_i\right] = \sum_{T \in W_i} \Pr\left[T \text{ appears in } \mathbf{c}\right] \leq \sum_{T \in W_i} \prod_{j \in T} \Pr\left[E_j\right] = \frac{x_i}{1 - x_i} \cdot \sum_{T \in W_i} p_T \leq \frac{x_i}{1 - x_i},$$

and $\sum_T p_T \leq 1$ as we are summing over disjoint events. The rest is immediate by the linearity of expectation. □

1.7 Markov Chains and Random Walks

1.7.1 Stirling's Formula

After previous bounds for factorial by corollary 1.5.5 and corollary 1.6.3, we reach Stirling's Formula.

Lemma 1.7.1 (Stirling's Formula)**.** *For $n > 0$,*

$$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n (1 \pm o(1)).$$

In particular, for $n > 0$,

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \le n! \le \left(\sqrt{2\pi n} + 1\right)\left(\frac{n}{e}\right)^n.$$

Proof. We start from the fact that factorial is the discrete form of the Gamma function, by

$$n! = \Gamma(n+1) = \int_0^\infty t^n e^{-t}\, dt.$$

Let $t = n(1+x)$ where $x > -1$, then

$$\begin{aligned}
\Gamma(n+1) &= \int_0^\infty t^n e^{-t}\, dt \\
&= \int_{-1}^\infty (n(1+x))^n \cdot e^{-n(1+x)} \cdot n\, dx \\
&= n\left(\frac{n}{e}\right)^n \int_{-1}^\infty (1+x)^n e^{-nx}\, dx \\
&= n\left(\frac{n}{e}\right)^n \int_{-1}^\infty \exp\left(n\left(\ln(1+x) - x\right)\right)\, dx \\
&= n\left(\frac{n}{e}\right)^n \int_{-1}^\infty \exp\left(n \cdot f(x)\right)\, dx,
\end{aligned}$$

where $dt = n\, dx$, and we let $f(x) = \log(x+1) - x$. We continue by series expansion:

- $f'(x) = (1+x)^{-1} - 1 = -x/(1+x)$, and $f'(0) = 0$;
- $f''(x) = -(1+x)^{-2}$, and $f''(0) = -1$.

Hence, $f(x) = -x^2/2 + O(x^3)$. We let $-z^2/2 = f(x)$, where

$$z = \operatorname{sgn}(x) \cdot \sqrt{2\left(x - \ln(x+1)\right)},$$

then by $df(x) = f'(x)\, dx = -z\, dz$, $dx = (z + z/x)\, dz$, and by replacing dx with dz,

$$\Gamma(n+1) = n\left(\frac{n}{e}\right)^n \int_{-\infty}^\infty \exp\left(-\frac{nz^2}{2}\right) \cdot \left(z + \frac{z}{x}\right)\, dz.$$

We continue by bounding the term z/x with results from [Top07].

Lemma 1.7.2 ($\log(x+1)$ bounds [Top07])**.** *For $x > -1$,*

$$\frac{x(6+x)}{6+4x} \ge \ln(x+1) \ge \begin{cases} x - \dfrac{x^2}{2} & x \ge 0 \\ \dfrac{x(2+x)}{2+2x} & x \le 0 \end{cases}.$$

With lemma 1.7.2, we need to prove the following bound to complete the proof.

Corollary 1.7.3. *If* $x > -1$, *then* $1 + 2z/3 \le z + z/x \le \max(1, 1+z)$.

Proof. First, $z(1+1/x) > 0$, since $-1 < x \le 0$, both z and $1+1/x$ are less than 0. Hence, we can bound by $z^2(1+1/x)^2$.

When $-1 < x \le 0$, by lemma 1.7.2

$$z^2 = 2\left(x - \ln(x+1)\right) \le \frac{x^2}{1+x} \le \left(\frac{x}{1+x}\right)^2,$$

and therefore $z^2(1+1/x)^2 \le 1$. Hence, $z + z/x \le 1$.

When $x \ge 0$,

$$z^2 = 2\left(x - \ln(x+1)\right) \le x^2,$$

and therefore $z^2/x^2 \le 1$. Hence, $z + z/x \le 1 + z$.

Hence, the upper bound is $\max(1, 1+z)$.

For lower bound, we have

$$z^2 = 2\left(x - \ln(x+1)\right) \ge \frac{3x^2}{3+2x},$$

and

$$\frac{3x^2}{3+2x} \cdot \left(\frac{1}{3} + \frac{1}{x}\right)^2 = \frac{x^2}{9+6x} + 1 \ge 1.$$

Hence, the lower bound is $1 + 2z/3$, which completes the proof. □

Since

$$\int_{-\infty}^{\infty} \exp\left(-z^2\right)\, dz = \sqrt{\pi} \text{ and } \int_0^{\infty} z \cdot \exp\left(-z^2\right)\, dz = \frac{1}{2},$$

then

$$\int_{-\infty}^{\infty} \exp\left(-\frac{nz^2}{2}\right)\, dz = \sqrt{\frac{2\pi}{n}} \text{ and } \int_0^{\infty} z \cdot \exp\left(-\frac{nz^2}{2}\right)\, dz = \frac{1}{n},$$

and therefore the upper bound is derived by,

$$\begin{aligned}\Gamma(n+1) &= n\left(\frac{n}{e}\right)^n \int_{-\infty}^{\infty} \exp\left(-\frac{nz^2}{2}\right) \cdot \left(z + \frac{z}{x}\right)\, dz \\ &\le n\left(\frac{n}{e}\right)^n \cdot \left(\sqrt{\frac{2\pi}{n}} + \frac{1}{n}\right) = \left(\sqrt{2\pi n} + 1\right)\left(\frac{n}{e}\right)^n,\end{aligned}$$

and the lower bound is derived by

$$\begin{aligned}\Gamma(n+1) &= n\left(\frac{n}{e}\right)^n \int_{-\infty}^{\infty} \exp\left(-\frac{nz^2}{2}\right) \cdot \left(z + \frac{z}{x}\right)\, dz \\ &\ge n\left(\frac{n}{e}\right)^n \int_{-\infty}^{\infty} \exp\left(-\frac{nz^2}{2}\right) \cdot \left(1 + \frac{2z}{3}\right)\, dz \\ &= n\left(\frac{n}{e}\right)^n \int_{-\infty}^{\infty} \exp\left(-\frac{nz^2}{2}\right)\, dz \\ &= n\left(\frac{n}{e}\right)^n \cdot \sqrt{\frac{2\pi}{n}} = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n,\end{aligned}$$

which completes the proof [2]. □

[2] This derivation is based on Thomas Ahle's post in https://mathoverflow.net/a/484270.

1.7.2 Markov Chains and Classification of States

Definition 1.7.4. The stochastic process $\mathbf{X} = \{X_t\}_{t\in T}$ is a collection of random variables. The index t represents time, and the process $\mathbf{X}$ models how a random variable changes over time.

X_t is the state of the process at the t^{th} time step.

- If X_t takes values from a countably infinite set, $\mathbf{X}$ is a discrete-state process.
- If X_t takes values from a finite set, then $\mathbf{X}$ is finite-state.
- If T is a countably infinite set, $\mathbf{X}$ is a discrete-time process.

A discrete stochastic process $X_0, X_1, \ldots$ is a time-homogeneous Markov chain if

$$\Pr[X_t = a_t \mid X_{t-1} = a_{t-1}, \ldots, X_0 = a_0] = \Pr[X_t = a_t \mid X_{t-1} = a_{t-1}] = P_{a_{t-1},a_t}.$$

This is called the Markov property, or memoryless property, or Markovian.

Exercise 1.7.5 (Exercise 7.2 [MU17])**.** Consider the 2-state Markov chain with the following transition matrix

$$\mathbf{P} = \begin{bmatrix} p & 1-p \\ 1-p & p \end{bmatrix}.$$

Find the minimum expression for $P^t_{0,0}$ in $\mathbf{P}^t$.

Proof. We can derive the following recurrence relation

$$P^t_{0,0} = \langle \left[P^{t-1}_{0,0}, 1 - P^{t-1}_{0,0}\right]^{\mathsf{T}}, \left[p, 1-p\right]^{\mathsf{T}} \rangle = (2p-1) \cdot P^{t-1}_{0,0} - (p-1).$$

as the $\mathbf{P}^{t-1}$ is diagonally symmetric with same diagonal values. Therefore, $P^t_{0,0} = \left(1 + (2p-1)^t\right)/2$. □

We present a 2-SAT randomized algorithm from Markov chain by [Pap91] that is similar to theorem 1.6.60:

- Start with a random truth assignment.
- Repeat up to $m \cdot 2n^2$ times, terminating on all clauses satisfied:
  - Choose an arbitrary unsatisfied clause.
  - Choose uniformly at random a variable in the clause, and flip it.
- If the loop ends with a satisfying assignment, return it; otherwise, abort.

Lemma 1.7.6 (2-SAT randomized algorithm expected runtime [Pap91])**.** *Assume the 2-SAT formula is satisfiable, and the 2-SAT randomized algorithm is allowed to run until it finds a satisfying assignment. Then the expected number of steps until the algorithm finds a satisfying assignment is at most n^2.*

Proof. One can view the state transition graph as a line of $n+1$ vertices, from 0 to n.

Suppose the algorithm outputs a satisfying assignment S, and we let the initial random assignment be A_0, A_t be the assignment on the t^{th} step. We define the process $\{X_t\}_{t\in T}$ over the number of matches of A_t with S.

- When $X_{t-1} = 0$, any flip of variable can make A_{t-1} have more match, and thus $\Pr[X_t = 1 \mid X_{t-1} = 0] = 1$.
- When $X_{t-1} = j$, then
  - If the clause picked has 1 unmatched variable, then the probability of making situation better is $1/2$.
  - If the clause picked has 2 unmatched variables, then any flip can make situation better.

  Hence, $\Pr\left[X_t = j+1 \mid X_{t-1} = j\right] \geq 1/2$.

But this goes against the Markovian property, namely memoryless, as the state transition probability should only be tied to the time step and the states. But in the current model, the probability varies on the number of disagreeing variables in a chosen clause. Hence, we define the following process $\{Y_t\}_{t\in T}$ that $Y_0 = X_0$, with

$$\begin{aligned}\Pr[Y_t = 1 \mid Y_{t-1} = 0] &= 1,\\ \Pr\left[Y_t = j+1 \mid Y_{t-1} = j\right] &= 1/2,\\ \Pr\left[Y_t = j-1 \mid Y_{t-1} = j\right] &= 1/2.\end{aligned}$$

Let $\{Z_i\}_{i\in[0,n]}$ be the random variables for the number of steps to reach n from i matches in process $\{Y_t\}_{t\in T}$. The following claims are immediate:

$$\begin{aligned}\mathbb{E}[Z_n] &= 0,\\ \mathbb{E}\left[Z_j\right] &= \mathbb{E}\left[Z_j \mid Y_{t+1} = j-1\right]\cdot P_{j,j-1} + \mathbb{E}\left[Z_j \mid Y_{t+1} = j+1\right]\cdot P_{j,j+1}\\ &= \left(1+\mathbb{E}\left[Z_{j-1}\right]\right)/2 + \left(1+\mathbb{E}\left[Z_{j+1}\right]\right)/2\\ &= 1 + \left(\mathbb{E}\left[Z_{j-1}\right] + \mathbb{E}\left[Z_{j+1}\right]\right)/2,\\ \mathbb{E}[Z_0] &= \mathbb{E}[Z_1] + 1.\end{aligned}$$

It is immediate that $\mathbb{E}[Z_1] = \mathbb{E}[Z_2] + 3$. By induction, we derive the following recurrence relation

$$\mathbb{E}\left[Z_j\right] = \mathbb{E}\left[Z_{j+1}\right] + (2j+1).$$

By summation, $\mathbb{E}[Z_0] = n^2$, and $\mathbb{E}\left[Z_j\right] = n^2 - j^2$. □

The following theorem is immediate from lemma 1.7.6.

Theorem 1.7.7. *The 2-SAT algorithm always returns correct answer on an unsatisfiable formula. If the formula is satisfiable, with probability at least $1 - 2^{-m}$ the algorithm gives a satisfiable assignment.*

Proof. One can segment the $m \cdot 2n^2$ steps in the algorithm into m chunks of $2n^2$ steps, running independently.

Let Z be the random variable for the steps the algorithm needs to take from i matches. By lemma 1.7.6, $\mathbb{E}[Z] \le n^2$. Immediate from Markov's inequality,

$$\Pr\left[Z \ge 2n^2\right] \le \frac{\mathbb{E}[Z]}{2n^2} \le \frac{1}{2}.$$

Hence, the probability that the algorithm fails to find a satisfying assignment after m segments is at most 2^{-m}. □

Exercise 1.7.8 (Exercise 7.6 [MU17])**.** Consider a partially reflecting boundary at 0, rather than a complete reflecting boundary at 0 in lemma 1.7.6. At position 0, with probability $1/2$ the walk moves to position 1 and with probability $1/2$ the walk stays at 0. Everywhere else is the same. Find the expected number of moves to reach n from position i.

Proof. We continue with reusing $\{Z_i\}_{i\in[0,n]}$ and process $\{Y_t\}_{t\in T}$. The following claims are immediate:

$$\begin{aligned}\mathbb{E}[Z_n] &= 0,\\ \mathbb{E}\left[Z_j\right] &= \mathbb{E}\left[Z_j \mid Y_{t+1} = j+1\right]\cdot P_{j,j+1} + \mathbb{E}\left[Z_j \mid Y_{t+1} = j-1\right]\cdot P_{j,j-1}\\ &= 1 + \left(\mathbb{E}\left[Z_{j+1}\right] + \mathbb{E}\left[Z_{j-1}\right]\right)/2,\\ \mathbb{E}[Z_0] &= \mathbb{E}[Z_0 \mid Y_{t+1} = 0]\cdot P_{0,0} + \mathbb{E}[Z_0 \mid Y_{t+1} = 1]\cdot P_{0,1}\\ &= 1 + \left(\mathbb{E}[Z_0] + \mathbb{E}[Z_1]\right)/2.\end{aligned}$$

It is immediate that $\mathbb{E}[Z_1] = \mathbb{E}[Z_2] + 4$. By induction, we derive the following recurrence relation

$$\mathbb{E}\left[Z_j\right] = \mathbb{E}\left[Z_{j+1}\right] + (2j+2).$$

By summation, $\mathbb{E}[Z_n] = n(n+1)$, and $\mathbb{E}\left[Z_j\right] = n(n+1) - j(j+1)$. □

Exercise 1.7.9 (Exercise 7.7 [MU17]). Find the expected runtime of the 2-SAT randomized algorithm for lemma 1.7.6, where the input is sampled uniformly at random.

Proof. We follow the $\{Z_j\}_{j\in[0,n]}$ notions of lemma 1.7.6, and let Y be the random variable for the number of mismatches in the initial assignment. Then by conditional expectation,

$$\begin{aligned}
\mathbb{E}[Z] &= \sum_{j\in[0,n]} \mathbb{E}\left[Z \mid Y = j\right] \cdot \Pr\left[Y = j\right] = \sum_{j\in[0,n]} \mathbb{E}\left[Z_j\right] \cdot \binom{n}{j} \cdot \frac{1}{2^n} = \sum_{j\in[0,n]} \left(n^2 - j^2\right) \cdot \binom{n}{j} \cdot \frac{1}{2^n} \\
&= n^2 - \frac{1}{2^n} \sum_{j\in[0,n]} j^2 \cdot \binom{n}{j} = n^2 - \frac{n}{2^n} \sum_{j\in[1,n]} j \cdot \binom{n-1}{j-1} \\
&= n^2 - \frac{n}{2} - \frac{n}{2^n} \sum_{k\in[0,n-1]} k \cdot \binom{n-1}{k} = n^2 - \frac{n}{2} - \frac{n(n-1)}{2^n} \sum_{\ell\in[0,n-2]} \binom{n-2}{\ell} \\
&= n^2 - \frac{n}{2} - \frac{n(n-1)}{4} = \frac{3}{4}n^2 - \frac{1}{4}n.
\end{aligned}$$

□

After lemma 1.7.6, we present a 3-SAT randomized algorithm from Markov chain. A prototype as follows:

- Pick a random truth assignment.
- Repeat up to m times, terminating on all clauses satisfied:
  - Choose an arbitrary unsatisfied clause.
  - Choose uniformly at random a variable in the clause, and flip it.
- If the loop ends with a satisfying assignment, return it; otherwise, abort.

Again after lemma 1.7.6, we define the Markov chain by the following process $\{Y_t\}_{t\in T}$ by

$$\begin{aligned}
\Pr\left[Y_t = 1 \mid Y_{t-1} = 0\right] &= 1, \\
\Pr\left[Y_t = j+1 \mid Y_{t-1} = j\right] &= 1/3, \\
\Pr\left[Y_t = j-1 \mid Y_{t-1} = j\right] &= 2/3.
\end{aligned}$$

We proceed by defining the random variables $\{Z_i\}_{i\in[0,n]}$ for the number of steps to reach n matches for the current assignment with a satisfying assignment S, where the current assignment has i matches. By conditional expectation,

$$\begin{aligned}
\mathbb{E}\left[Z_n\right] &= 0, \\
\mathbb{E}\left[Z_j\right] &= \mathbb{E}\left[Z_j \mid Y_{t+1} = j-1\right] \cdot P_{j,j-1} + \mathbb{E}\left[Z_j \mid Y_{t+1} = j+1\right] \cdot P_{j,j+1} \\
&= 1 + 2/3 \cdot \mathbb{E}\left[Z_{j-1}\right] + 1/3 \cdot \mathbb{E}\left[Z_{j+1}\right], \\
\mathbb{E}\left[Z_0\right] &= 1 + \mathbb{E}\left[Z_1\right].
\end{aligned}$$

By induction, $\mathbb{E}\left[Z_j\right] = \mathbb{E}\left[Z_{j+1}\right] + 2^{j+2} - 3$. But this results in $\mathbb{E}\left[Z_0\right] = O(2^n)$, where trying all n variables is also 2^n, which is not impressive. The following 2 observations are helpful:

- It is more likely to make more violation when choosing a random unsatisfied clause and flip a variable at random. Hence, on a random initial assignment, it seems better to try a smaller number of repetition, as it is more likely to drift to a larger number of matches, than a rather large number of repetition.
- Initial assignment of variables with uniform randomness can be viewed as independent Bernoulli trials. Hence, we have some chance of sampling an initial assignment with matches higher than $n/2$.

Based on the observation, it seems better to try for a smaller amount of repetitions on a random initial assignment, and try over a large variety of different random initial assignments.

Given an initial assignment with j mismatches, we let q_j be the probability that the random walk finds a satisfying assignment from this initial assignment in $3n$ steps. The random walk always needs $2k+j$ steps to reach n matches. To lower bound the probability that a satisfying assignment can be found within $2k+j$ steps, we consider $k+j$ steps drifting to more matches, while k steps drifting to less matches, by

$$p_k = \binom{2k+j}{k}\left(\frac{1}{3}\right)^{k+j}\left(\frac{2}{3}\right)^{k} = \binom{2k+j}{k}\cdot\left(\frac{1}{3}\right)^{2k+j}\cdot 2^k.$$

It turned out that for any k, $p_k \le q_j$, as

- All exact $2k+j$ steps cases are captured.
- All cases reaching n matches earlier, should drift around the n matches, with a probability at most 1.
- All other cases reaching n matches later, are omitted, while their probability is non-negative.

Since $q_j \ge \max_{2k+j\le 3n} p_k$, then surely the following holds:

$$q_j \ge \binom{3j}{j}\cdot\frac{2^j}{3^{3j}}.$$

By lemma 1.7.1, we have

$$\begin{aligned} q_j &\ge \binom{3j}{j}\cdot\frac{2^j}{3^{3j}} = \frac{(3j)!}{(2j)!\cdot j!}\cdot\frac{2^j}{3^{3j}} \\ &\ge \frac{\sqrt{2\pi\cdot 3j}\cdot(3j)^{3j}}{\left(2\sqrt{2\pi\cdot 2j}\cdot(2j)^{2j}\right)\cdot\left(2\sqrt{2\pi\cdot j}\cdot j^j\right)}\cdot\frac{2^j}{3^{3j}} \\ &= \frac{\sqrt{3}}{8\sqrt{\pi}}\cdot\frac{1}{2^j\sqrt{j}} = \frac{c}{2^j\sqrt{j}}. \end{aligned}$$

Therefore, the lower bound for q, the probability for finding a satisfying assignment within $3n$ steps, is

$$\begin{aligned} q &= \sum_{j\in[0,n]} q_j\cdot\Pr\left[\text{initial assignment has } j \text{ mismatches}\right] \\ &\ge \sum_{j\in[0,n]}\binom{3j}{j}\cdot\frac{2^j}{3^{3j}}\cdot\binom{n}{j}\cdot\frac{1}{2^n} \ge \frac{c}{2^n}\sum_{j\in[0,n]}\frac{1}{2^j\sqrt{j}}\cdot\binom{n}{j} \\ &\ge \frac{c}{2^n\sqrt{n}}\sum_{j\in[0,n]}\binom{n}{j}\cdot\frac{1}{2^j} = \frac{c}{\sqrt{n}}\cdot\left(\frac{3}{4}\right)^n. \end{aligned}$$

We now formulate the 3-SAT randomized algorithm from random walk as follows:

- Repeat up to m times:
  - Pick a random truth assignment.
  - Repeat up to $3n$ times, terminating on all clauses satisfied:
    - Choose an arbitrary unsatisfied clause.
    - Choose uniformly at random a variable in the clause, and flip it.
- If the loop ends with a satisfying assignment, return it; otherwise, abort.

The expected number of rounds is at most $1/q = O(\sqrt{n}\,(4/3)^n)$, and the expected runtime is $O(n^{3/2}\,(4/3)^n)$.

Similar to the 2-SAT randomized algorithm, we can make

$$m = b \cdot \frac{2\sqrt{n}}{c} \cdot \left(\frac{4}{3}\right)^n,$$

such that we have b segments, while one segment failing to find a satisfying assignment can be viewed as

$$\Pr\left[Z > \frac{2\sqrt{n}}{c} \cdot \left(\frac{4}{3}\right)^n\right] < \mathbb{E}[Z] \cdot \frac{c}{2\sqrt{n}} \cdot \left(\frac{3}{4}\right)^n \leq \frac{1}{2},$$

by Markov inequality, where Z is a geometrically distributed random variable with success probability q.

Hence, the failing probability of finding a satisfying assignment for 3-SAT is at most 2^{-b}.

Exercise 1.7.10 (Exercise 7.8 [MU17])**.** Generalize the 3-SAT randomized algorithm to k-SAT algorithm, and write the expected runtime of the algorithm as a function of k.

Proof. We reuse most of the notions in 3-SAT randomized algorithm, where S is a satisfying assignment, $\{A_t\}_{t\in T}$ is the assignment at t^{th} time step, $\{Y_t\}_{t\in T}$ is the Markov chain for the number of matches that

$$\begin{aligned}\Pr[Y_t = 1 \mid Y_{t-1} = 0] &= 1,\\ \Pr\left[Y_t = j+1 \mid Y_{t-1} = j\right] &= 1/k,\\ \Pr\left[Y_t = j-1 \mid Y_{t-1} = j\right] &= 1 - 1/k,\end{aligned}$$

and $\{Z_i\}_{i\in[0,n]}$ are the random variables for the steps to satisfy all clauses, from an initial assignment of i mismatches.

For $k \geq 3$, the random walk is always more likely to drift to less satisfied variables than more satisfied variables. Hence, the prior strategy of trying over a large variety of initial assignments, then repeat a small number of times each for random flipping, remains to apply. Therefore, the 3-SAT algorithm should still apply for k-SAT case.

Let q_i be the probability for finding satisfying assignment from i mismatches in $3n$ steps of random walk. Then

$$q_i \geq \max_{2\ell+i\leq 3n} \binom{2\ell+i}{\ell} \left(\frac{1}{k}\right)^{\ell+i} \left(\frac{k-1}{k}\right)^{\ell},$$

and therefore by lemma 1.7.1,

$$\begin{aligned}q_i &\geq \binom{3i}{i} \left(\frac{1}{k}\right)^{2i} \left(\frac{k-1}{k}\right)^{i}\\ &\geq \frac{\sqrt{2\pi\cdot 3i}\cdot(3i)^{3i}}{\left(2\sqrt{2\pi\cdot 2i}\cdot(2i)^{2i}\right)\cdot\left(2\sqrt{2\pi i}\cdot i^i\right)} \cdot \frac{(k-1)^i}{k^{3i}}\\ &= \frac{\sqrt{3}}{8\sqrt{\pi}} \cdot \frac{1}{\sqrt{i}} \cdot \left(\frac{27(k-1)}{4k^3}\right)^i = \frac{c}{\sqrt{i}} \cdot \left(\frac{27(k-1)}{4k^3}\right)^i\end{aligned}$$

Let q be the probability for sampling a random initial assignment, then

$$q = \sum_{i\in[0,n]} q_i \cdot \binom{n}{i} \cdot \frac{1}{2^n} \geq \frac{c}{2^n\sqrt{n}} \sum_{i\in[0,n]} \binom{n}{i} \cdot \left(\frac{27(k-1)}{4k^3}\right)^i = \frac{c}{\sqrt{n}} \left(\frac{1}{2} + \frac{27(k-1)}{8k^3}\right)^n.$$

Hence, the expected runtime for the k-SAT algorithm is

$$\frac{3n}{q} \leq \frac{\sqrt{n}}{c} \left(\frac{8k^3}{4k^3 + 27(k-1)}\right)^n \cdot 3n = a \cdot 3n,$$

where a is the expected number of trials to the first satisfying assignment.

We set $m = 2ab$ for at least $1 - 2^{-b}$ probability of outputting satisfying assignment on a satisfiable formula. □

Exercise 1.7.11 (Exercise 7.9 [MU17])**.** The prior 3-SAT randomized algorithm analysis made pessimistic assumption that the current assignment A_t and the satisfying assignment S differ in one variable in the chosen unsatisfied clause. Instead, suppose that the 2 assignments disagree on 1 variable with probability p, and disagree on at least 2 variables with probability $1-p$. What is the largest p, such that one can prove that the expected number of steps before the randomized 3-SAT stops is polynomial in n?

Proof. We reuse most of the notions in 3-SAT randomized algorithm, where S is a satisfying assignment, $\{A_t\}_{t\in T}$ is the assignment at t^{th} time step, $\{Y_t\}_{t\in T}$ is the modified Markov chain for the number of matches that

$$\begin{aligned}
\Pr\left[Y_t = 1 \mid Y_{t-1} = 0\right] &= 1,\\
\Pr\left[Y_t = j+1 \mid Y_{t-1} = j\right] &= \frac{1}{3}\cdot p + \frac{2}{3}\cdot(1-p) = \frac{2}{3} - \frac{1}{3}\cdot p,\\
\Pr\left[Y_t = j-1 \mid Y_{t-1} = j\right] &= \frac{2}{3}\cdot p + \frac{1}{3}\cdot(1-p) = \frac{1}{3} + \frac{1}{3}\cdot p,
\end{aligned}$$

and $\{Z_i\}_{i\in[0,n]}$ are the random variables for the steps to satisfy all clauses, from an initial assignment of i mismatches. Then q_j, the probability of finding a satisfying assignment within $3n$ steps for initial j mismatches, has

$$\begin{aligned}
q_j &\geq \binom{3j}{j}\left(\frac{2}{3} - \frac{1}{3}\cdot p\right)^{2j}\left(\frac{1}{3} + \frac{1}{3}\cdot p\right)^{j}\\
&\geq \frac{\sqrt{2\pi\cdot 3j}\cdot(3j)^{3j}}{\left(2\sqrt{2\pi\cdot 2j}\cdot(2j)^{2j}\right)\cdot\left(2\sqrt{2\pi j}\cdot j^{j}\right)}\cdot\left(\frac{(2-p)^2(1+p)}{27}\right)^{j}\\
&= \frac{\sqrt{3}}{8\sqrt{\pi}}\cdot\frac{1}{\sqrt{j}}\cdot\left(\frac{(2-p)^2(1+p)}{4}\right)^{j} = \frac{c}{\sqrt{j}}\cdot\left(\frac{(2-p)^2(1+p)}{4}\right)^{j}.
\end{aligned}$$

Now the probability q for finding a satisfying assignment within $3n$ steps is

$$q = \sum_{j\in[0,n]} q_j\cdot\binom{n}{j}\cdot\frac{1}{2^n} \geq \frac{c}{2^n\sqrt{n}}\sum_{j\in[0,n]}\binom{n}{j}\cdot\left(\frac{(2-p)^2(1+p)}{4}\right)^{j} = \frac{c}{\sqrt{n}}\left(\frac{1}{2} + \frac{(2-p)^2(1+p)}{8}\right)^{n}.$$

In this case, only when $p = 0$ can make $1/q = \mathsf{poly}(n)$. □

Exercise 1.7.12 (Exercise 7.10 [MU17])**.** A coloring of a graph is an assignment of a color to each of its vertices. A graph is k-colorable if there exists coloring such that no 2 adjacent vertices have the same color. Let G be 3-colorable.

- Show that there exists a 2-coloring for G such that no triangle is monochromatic.
- Consider the following 2-coloring randomized algorithm for G such that there is no monochromatic triangle. Start from any 2-coloring. If there are any monochromatic triangles, choose a random one and flip the color of a random vertex in the triangle. Find an upper bound for the expected number of steps in the algorithm.

Proof. Let the 3-colorable graph G be given a proper 3-coloring such that no pair of adjacent vertices has the same color. Then any triangle in G must be colored in 3 distinct colors; otherwise at least one pair of adjacent vertices would have the same color, contradicting that G is 3-colorable. We randomly merge 2 colors from the 3-coloring into 1 color, then all the triangles are 2-colored, such that none of them are monochromatic.

The algorithm runtime analysis is similar to the one for the 3-SAT randomized algorithm. We define the satisfying coloring S, $\{A_t\}_{t\in T}$ as the coloring at time step t, $\{Y_t\}_{t\in T}$ as the Markov chain for the number of matches that

$$\begin{aligned}
\Pr\left[Y_t = 1 \mid Y_{t-1} = 0\right] &= 1,\\
\Pr\left[Y_t = j+1 \mid Y_{t-1} = j\right] &= 1/3,\\
\Pr\left[Y_t = j-1 \mid Y_{t-1} = j\right] &= 2/3,
\end{aligned}$$

and $\{Z_i\}_{i\in[0,n]}$ as the random variables for the number of steps to match S fully from an initial coloring of i mismatches. We adopt the analysis for the unsuccessful 3-SAT randomized algorithm, and by conditional expectation,

$$\begin{aligned}
\mathbb{E}[Z_n] &= 0,\\
\mathbb{E}\left[Z_j\right] &= \mathbb{E}\left[Z_j \mid Y_{t+1} = j-1\right]\cdot P_{j,j-1} + \mathbb{E}\left[Z_j \mid Y_{t+1} = j+1\right]\cdot P_{j,j+1}\\
&= 1 + 2/3\cdot\mathbb{E}\left[Z_{j-1}\right] + 1/3\cdot\mathbb{E}\left[Z_{j+1}\right],\\
\mathbb{E}[Z_0] &= 1 + \mathbb{E}[Z_1].
\end{aligned}$$

By induction, $\mathbb{E}\left[Z_j\right] = \mathbb{E}\left[Z_{j+1}\right] + 2^{j+2} - 3$, and eventually

$$\mathbb{E}\left[Z_j\right] = \sum_{k\in[j,n-1]}\left(2^{k+2}-3\right) = 4\left(2^n - 2^j\right) - 3(n-j) = \left(2^{n+2}-3n\right) - \left(2^{j+2}-3j\right).$$

Let Z be the random variable for the number of steps in the algorithm, and Y be the random variable for the number of mismatches in the initial assignment, then by conditional expectation,

$$\begin{aligned}
\mathbb{E}[Z] &= \sum_{j\in[0,n]}\mathbb{E}\left[Z \mid Y = j\right]\cdot\Pr\left[Y = j\right] = \sum_{j\in[0,n]}\mathbb{E}\left[Z_j\right]\cdot\binom{n}{j}\cdot\frac{1}{2^n}\\
&= \left(2^{n+2}-3n\right) - \frac{1}{2^n}\sum_{j\in[0,n]}\left(2^{j+2}-3j\right)\cdot\binom{n}{j}\\
&= \left(2^{n+2}-3n\right) - 4\left(\frac{3}{2}\right)^n + \frac{3}{2^n}\sum_{j\in[0,n]} j\cdot\binom{n}{j}\\
&= \left(2^{n+2}-3n\right) - 4\left(\frac{3}{2}\right)^n + \frac{3n}{2} = 2^{n+2} - 4\left(\frac{3}{2}\right)^n - \frac{3n}{2}.
\end{aligned}$$

□

Theorem 1.7.13 (Chapman-Kolmogorov Equation)**.** *For a time-homogeneous Markov chain,*

$$P^{m+n}_{i,j} = \sum_k P^m_{i,k}\cdot P^n_{k,j}.$$

Proof.

$$\begin{aligned}
P^{m+n}_{i,j} &= \Pr\left[X_{m+n} = j \mid X_0 = i\right]\\
&= \sum_k \Pr\left[X_{m+n} = j \mid X_m = k\right]\cdot\Pr\left[X_m = k \mid X_0 = i\right]\\
&= \sum_k \Pr\left[X_n = j \mid X_0 = k\right]\cdot\Pr\left[X_m = k \mid X_0 = i\right] = \sum_k P^n_{k,j}\cdot P^m_{i,k}.
\end{aligned}$$

□

Definition 1.7.14. State j is accessible from state i if there exists some integer $n \geq 0$, $P^n_{i,j} > 0$. If state i, j are accessible from each other, we say they communicate by $i \leftrightarrow j$. The communicating relation defines an equivalence relation:

- Reflexive: For any state i, $i \leftrightarrow i$.
- Symmetric: If $i \leftrightarrow j$, then $j \leftrightarrow i$.
- Transitive: If $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$.

The communication relation partitions the states into disjoint equivalence classes, which are referred to as communicating classes. It might be possible to move from one class to another, but not in the other way around.

Definition 1.7.15. A Markov chain is irreducible if all states belong to one communicating class.

Hence, a Markov chain is irreducible if for all ordered pair of states (i, j), the probability that state i can reach state j is nonzero. The next lemma is immediate.

Lemma 1.7.16. *A finite Markov chain is irreducible iff its graph representation is a strongly connected graph.*

Definition 1.7.17. Let $r_{i,j}^t$ be the probability that, the first transition from state i to state j happens on the t^{th} step:

$$r_{i,j}^t = \Pr\left[X_t = j \wedge \bigwedge_{k\in[1,t-1]} X_k \neq j \,\middle|\, X_0 = i\right].$$

A state is transient if $\sum_{t\geq 1} r_{i,i}^t < 1$, and it is recurrent if $\sum_{t\geq 1} r_{i,i}^t = 1$.

Definition 1.7.18. Let $T_{i,j}$ be the random variable for the steps to transition from i to j, and $T_{i,i}$ be the random variable for the steps to return to i, such that $T_{i,i} \geq 1$.

Exercise 1.7.19 (Exercise 7.15 [MU17])**.** Let $P_{i,i}^t$ be the probability that a Markov chain returns to state i when started in state i after t steps. Prove that $\sum_{t\geq 1} P_{i,i}^t$ is unbounded iff state i is recurrent.

Proof. We can express $P_{i,i}^t$ in the following recursive way:

$$P_{i,i}^t = \sum_{k\in[1,t]} r_{i,i}^k \cdot P_{i,i}^{t-k},$$

where $P_{i,i}^0 = 1$. Then the summation can be rearranged to

$$\begin{aligned}\sum_{t\geq 1} P_{i,i}^t &= \sum_{t\geq 1}\sum_{k\in[1,t]} r_{i,i}^k \cdot P_{i,i}^{t-k} \\ &= \left(\sum_{t\geq 1} r_{i,i}^t\right)\cdot\left(\sum_{k\geq 0} P_{i,i}^k\right) = \left(\sum_{t\geq 1} r_{i,i}^t\right)\cdot\left(\sum_{k\geq 1} P_{i,i}^k\right) + \left(\sum_{t\geq 1} r_{i,i}^t\right) \\ &= \left(\sum_{t\geq 1} r_{i,i}^t\right)\cdot\left(1 + \sum_{k\geq 1} P_{i,i}^k\right).\end{aligned}$$

In this way, $\sum_{t\geq 1} P_{i,i}^t$ being unbounded is equivalent to the state i being recurrent. □

Corollary 1.7.20. *If a state in a communicating class is transient (recurrent), then all states in the class are transient (recurrent).*

Proof. Suppose state i, j are in the same communicating class, and with integers $c_0, c_1 \geq 1$ such that $P_{i,j}^{c_0} > 0$, $P_{j,i}^{c_1} > 0$. By theorem 1.7.13 and omitting the first $c_0 + c_1$ terms,

$$\sum_{\tau\geq 1} P_{i,i}^\tau \geq \sum_{\tau\geq 1+c_0+c_1} P_{i,i}^\tau = \sum_{t\geq 1}\sum_{u,v} P_{i,u}^{c_0}\cdot P_{u,v}^t \cdot P_{v,i}^{c_1} \geq \sum_{t\geq 1}\sum_{k} P_{i,k}^{c_0}\cdot P_{k,k}^t \cdot P_{k,i}^{c_1} = \sum_k P_{i,k}^{c_0}\cdot P_{k,i}^{c_1}\cdot\sum_{t\geq 1} P_{k,k}^t \geq P_{i,j}^{c_0}\cdot P_{j,i}^{c_1}\cdot\sum_{t\geq 1} P_{j,j}^t,$$

where the fourth equality for changing summation sequence holds by probability being non-negative.

Then by exercise 1.7.19, if state i is transient, then $\sum_{t\geq 1} P_{j,j}^t$ has to be bounded, that state j is transient. Conversely, if state j is recurrent, then $\sum_{t\geq 1} P_{i,i}^t$ has to be unbounded, that state i is recurrent. The intuition follows.

Suppose state i is transient, then all the other states in the same communicating class have nonzero probability to reach i, and get out of the communicating class from i with nonzero probability, hence all of the states are transient.

On the other hand, supposing state i is recurrent but another state j is transient in the same communicating class, this is a contradiction as state i can reach state j and never get back to itself. Contradiction remains until no state in the communicating class is transient. □

Definition 1.7.21. Let $h_{i,i} = \sum_{t\geq 1} t \cdot r_{i,i}^t$ be the expected time to return to i when starting at state i, and $h_{i,j} = \sum_{t\geq 1} t \cdot r_{i,j}^t$ be the expected time to reach j from i. A recurrent state i is positive recurrent if $h_{i,i} < \infty$, otherwise it is null recurrent.

Corollary 1.7.22. *If i is transient, then $h_{i,i} = \infty$.*

Proof. Since i may never return by $\sum_{t\geq 1} r_{i,i}^t < 1$, hence $\Pr[T_{i,i} = \infty] > 0$. □

Corollary 1.7.23. *If i is recurrent, then $\Pr[T_{i,i} < \infty] = 1$.*

Proof. Suppose $\Pr[T_{i,i} = \infty] > 0$ for recurrent i, then it contradicts with recurrence: $\sum_{t\geq 1} r_{i,i}^t = 1$. □

Remark 1.7.24. The convergence of $h_{i,i}$ differs positive recurrence from null recurrence, and $\Pr[T_{i,i} = \infty] = 0$ differs recurrence from transience.

Example 1.7.25. Consider a finite Markov chain of states $[1, 2]$ with $P_{i,j} = 1/2$ for all $i, j \in [1, 2]$. The number of steps for "reaching state j" can be viewed as a geometrically distributed random variable, and $h_{i,j} = 2$.

However, consider a Markov chain over states $\mathbb{Z}^+$, that $P_{i,1} + P_{i,i+1} = 1$ and $P_{i,1} = (1 + i)^{-1}$. Then

$$r_{1,1}^t = P_{t,1} \cdot \prod_{i\in[1,t-1]} P_{i,i+1} = \frac{1}{t+1} \cdot \prod_{i\in[1,t-1]} \frac{i}{i+1} = \frac{1}{t+1} \cdot \frac{1}{t} = \frac{1}{t} - \frac{1}{t+1}.$$

Hence, $\sum_{t\geq 1} r_{1,1}^t = 1$, yet $h_{1,1} = \sum_{t\geq 1} t \cdot r_{1,1}^t = \infty$.

Theorem 1.7.26. *Let i be recurrent, and suppose i can access j. Then j is recurrent, and both $T_{i,j}$ and $T_{j,i}$ are bounded with probability 1.*

Proof. If i can access j, then with non-zero probability i can transition to j. Supposing j is transient, by corollary 1.7.20 i, j are in different communicating classes, thus $i \nleftrightarrow j$, hence $P_{j,i}^\tau = 0$ for any $\tau \geq 1$. Let $P_{i,j}^t > 0$ for some $t \geq 1$, then i returns to itself with probability at most $1 - P_{i,j}^t$, contradicting with the recurrence of i. Thus, j is recurrent.

Since i and j communicate, for one path from i to j of length ℓ, let the probability of taking this path be q, then returning to i before reaching j has probability at most $1 - q$. By corollary 1.7.23, $\Pr[T_{i,i} < \infty] = 1$. Hence, we model this with a geometrically distributed random variable: on a coin flip, either success with probability at least q to reach j in n steps, or fail with probability at most $1 - q$ to return to i in $T_{i,i}$ steps. Hence, $\Pr\left[T_{i,j} < \infty\right] = 1$.

Suppose $\Pr\left[T_{j,i} = \infty\right] = p > 0$, then $\Pr[T_{i,i} = \infty] \geq p \cdot q > 0$, which is contradicting to i being recurrent. □

Lemma 1.7.27. *In a finite Markov chain, there is at least one recurrent state, and all recurrent states are positive recurrent.*

Proof. A finite directed graph G can be decomposed into a finite number of strongly connected components (SCCs). Contracting SCCs gives a finite DAG, which has at least one sink SCC. Hence, there must exist a closed SCC.

Such closed SCC with 0 out degree corresponds to a closed communicating class S, such that the subchain is a finite irreducible Markov chain, and the probability of transitioning out is 0.

For $i \in S$, $\sum_{t\geq 1} r_{i,i}^t$ is the probability of i returning to itself. Let m be the integer such that $\Pr[T_{x,i} \leq m] > 0$ for all $x \in S$ by the irreducibility and finiteness of S, then

$$\Pr[T_{i,i} > m] = 1 - \Pr[T_{i,i} \leq m] \leq 1 - \min_{x\in S} \Pr[T_{x,i} \leq m] = 1 - \delta.$$

Moreover, by the Markovian property,

$$\Pr[T_{i,i} > (r + 1)m \mid T_{i,i} > rm] \leq 1 - \delta.$$

By the closure of S, we have $\Pr[T_{i,i} > rm] \leq (1-\delta)^r$ after repeating r blocks of m steps, converging to 0 when $r \to \infty$, yielding $\sum_{t\geq 1} r_{i,i}^t = 1$. By corollary 1.7.20, all $i \in S$ are recurrent.

For any $i \in S$,

$$\mathbb{E}[T_{i,i}] = \sum_{t\geq 1} t \cdot r_{i,i}^t \leq \sum_{r\geq 0} m \Pr[T_{i,i} > rm] \leq m \sum_{r\geq 0} (1 - \delta)^r \leq \frac{m}{\delta}.$$

Such a bound applies to all sink SCCs, so all recurrent states are positive recurrent. □

1.7.3 Strong Law of Large Numbers (SLLN)

We now introduce the Kolmogorov Strong Law of Large Numbers (SLLN), in contrast to theorem 1.3.11.

Theorem 1.7.28 (Kolmogorov SLLN). *Let $\{X_i\}_{i\geq 1}$ be i.i.d. random variables with $\mathbb{E}[|X_1|] < \infty$, and let $\mu = \mathbb{E}[X_1]$.*

$$\Pr\left[\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[1,n]} X_i = \mu\right] = 1.$$

We adopt the proof structure from [Ete81] and first prove a useful lemma.

Definition 1.7.29. Let $\{A_n\}_{n\geq 1}$ be a sequence of subsets of Ω. The supremum of $\{A_n\}_{n\geq k}$ is

$$\sup_{n\geq k} A_n = \bigcup_{n\geq k} A_n.$$

The limit superior of $\{A_n\}_{n\geq 1}$ is

$$\limsup_{n\to\infty} A_n = \bigcap_{n\geq 1} \bigcup_{k\geq n} A_k.$$

Equivalently, $\limsup_{n\to\infty} A_n$ is the event that "for every cutoff n, there exists $N \geq n$ such that A_N holds". In other words, there are infinitely many indices n such that A_n holds, or equivalently, "A_n occurs infinitely often".

Lemma 1.7.30 (Borel-Cantelli Lemma [Fel68a]). *Let $\{A_i\}_{i\geq 1}$ be an infinite sequence of events. If $\sum_{i\geq 1} \Pr[A_i] < \infty$,*

$$\Pr\left[\limsup_{n\to\infty} A_n\right] = 0. \tag{1.1}$$

Proof. We notice a non-increasing sequence of sets by

$$\limsup_{n\to\infty} A_n \subseteq \ldots \subseteq \bigcup_{k\geq N+1} A_k \subseteq \bigcup_{k\geq N} A_k \subseteq \ldots \subseteq \bigcup_{k\geq 2} A_k \subseteq \bigcup_{k\geq 1} A_k.$$

Hence, on an arbitrary N, we upper bound the probability by

$$\Pr\left[\limsup_{n\to\infty} A_n\right] \leq \Pr\left[\bigcup_{k\geq N} A_k\right] \leq \sum_{k\geq N} \Pr[A_k],$$

the second inequality holds from union bounding, and the right hand side summation approaches to 0 as $N \to \infty$. □

Remark 1.7.31. The event $\limsup_{n\to\infty} A_n$ means that no matter how far into the sequence we go, there is always a later A_k that occurs. In other words, A_n keep occurring arbitrarily far out, or equivalently, occur infinitely often.

Hence, eq. (1.1) means that with probability 1, there exists a finite cutoff after which none of the events A_n occur. Equivalently, only finitely many of them occur.

Now we begin the proof for theorem 1.7.28.

Proof. WLOG we let $\mu = \mathbb{E}[X_1] = 0$, $Y_n = X_n \cdot 1_{\leq n}$ by $Y_n = 0$ if $|X_n| > n$, otherwise $Y_n = X_n$, and $Z_n = X_n \cdot 1_{>n} = X_n - Y_n$.

Since $\mathbb{E}[|X_1|] < \infty$, and by $\{X_i\}$ are i.i.d., we have

$$\infty > \mathbb{E}[|X_1|] \geq \sum_{t\geq 1} \Pr[|X_1| \geq t] = \sum_{t\geq 1} \Pr[|X_t| \geq t] \geq \sum_{t\geq 1} \Pr[|X_t| > t]. \tag{1.2}$$

By lemma 1.7.30 and remark 1.7.31, the eq. (1.2) means with probability 1, only finitely many indices i have $|X_i| > i$. Equivalently, with probability 1, there exists a finite cutoff N such that $|X_n| > n$ does not occur for all $n > N$.

Hence, with probability 1, we have $Z_n = 0$ for all $n > N$, so only finitely many Z_n are nonzero. Therefore,

$$\Pr\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i\in[1,n]} Z_i = 0\right] = 1,$$

and what remains to show is

$$\Pr\left[\lim_{n\to\infty}\frac{1}{n}\sum_{i\in[1,n]} Y_i = 0\right] = 1.$$

Let $W_i = Y_i - \mathbb{E}[Y_i]$, $S_n = \sum_{i\in[1,n]} W_i$, and $m \in \mathbb{N}^+$, by Chebyshev's inequality and the independence of $\{X_i\}$,

$$\begin{aligned}\Pr\left[\left|\frac{S_{2^m}}{2^m}\right| \geq \varepsilon\right] &\leq \frac{\operatorname{Var}[S_{2^m}]}{\varepsilon^2\cdot 2^{2m}}\\ &= \frac{1}{\varepsilon^2\cdot 2^{2m}}\sum_{i\in[1,2^m]}\operatorname{Var}[W_i] = \frac{1}{\varepsilon^2\cdot 2^{2m}}\sum_{i\in[1,2^m]}\operatorname{Var}[Y_i]\\ &\leq \frac{1}{\varepsilon^2\cdot 2^{2m}}\sum_{i\in[1,2^m]}\mathbb{E}\left[Y_i^2\right].\end{aligned}$$

We want to take advantage of Borel-Cantelli lemma 1.7.30 again by showing if

$$\sum_{m\geq 1}\Pr\left[\left|\frac{S_{2^m}}{2^m}\right| \geq \varepsilon\right] \leq \sum_{m\geq 1}\frac{1}{\varepsilon^2\cdot 2^{2m}}\sum_{i\in[1,2^m]}\mathbb{E}\left[Y_i^2\right] < \infty,$$

then

$$\Pr\left[\limsup_{m\to\infty}\left|\frac{S_{2^m}}{2^m}\right| \geq \varepsilon\right] = 0,$$

that only a finite number of $m \geq 1$ has $|S_{2^m}| \geq \varepsilon\cdot 2^m$. There are 3 things remains to be proved:

Lemma 1.7.32.

$$\sum_{m\geq 1}\frac{1}{2^{2m}}\sum_{i\in[1,2^m]}\mathbb{E}\left[Y_i^2\right] < \infty.$$

Lemma 1.7.33.

$$\Pr\left[\limsup_{m\to\infty}\max_{2^m<n\leq 2^{m+1}}\left|\frac{S_n - S_{2^m}}{2^m}\right| \geq \varepsilon\right] = 0.$$

Lemma 1.7.34.

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i\in[1,n]}\mathbb{E}[Y_i] = 0.$$

We begin with proof for lemma 1.7.34.

Proof. We derive $\mathbb{E}[Y_n] = \mathbb{E}[X_1\cdot 1_{\leq n}] = -\mathbb{E}[X_1\cdot 1_{>n}]$, as we simplify by $\mathbb{E}[X_1] = 0$. Hence, by Jensen's inequality,

$$|\mathbb{E}[Y_n]| \leq |\mathbb{E}[X_1\cdot 1_{>n}]| \leq \mathbb{E}[|X_1|\cdot 1_{>n}].$$

By monotone convergence, we have $\mathbb{E}[|X_1|\cdot 1_{\leq n}] \to \mathbb{E}[|X_1|]$ as $n\to\infty$. Since

$$\mathbb{E}[|X_1|] = \mathbb{E}[|X_1|\cdot 1_{>n}] + \mathbb{E}[|X_1|\cdot 1_{\leq n}] < \infty$$

and $|X_1|\cdot 1_{\leq n} \geq 0$, we have $\mathbb{E}[|X_1|\cdot 1_{>n}] \to 0$ as $n\to\infty$.

By Cesàro summation, that if $a_n \to 0$ as $n\to\infty$, then $1/n\cdot\sum a_n \to 0$, thus

$$0 \leq \left|\frac{1}{n}\sum_{i\in[1,n]}\mathbb{E}[Y_i]\right| \leq \frac{1}{n}\sum_{i\in[1,n]}|\mathbb{E}[Y_i]| \leq \frac{1}{n}\sum_{i\in[1,n]}\mathbb{E}[|X_1|\cdot 1_{>i}] \to 0,$$

which completes the proof. □

Now we move on by proving lemma 1.7.32.

Proof. Noticing that for real valued random variable X_k,

$$Y_k^2 = X_1^2 \cdot 1_{\leq k} \leq \sum_{m\in[1,k]} (2m-1)\cdot 1_{>m-1},$$

since when $|X_1| \leq k$, the right hand side is $\lceil |X_1| \rceil^2 \geq X_1^2$, and when $|X_1| > k$, the left hand side is 0. Hence,

$$\begin{aligned}\mathbb{E}\left[Y_k^2\right] &= \mathbb{E}\left[X_1^2 \cdot 1_{\leq k}\right] \\ &\leq \sum_{m\in[1,k]} (2m-1)\cdot \Pr\left[|X_1| > m-1\right].\end{aligned} \tag{1.3}$$

Relaxing the summation, then let p_t be the first p such that $2^p \geq t$,

$$\begin{aligned}\sum_{m\geq 1}\frac{1}{2^{2m}}\sum_{t\in[1,2^m]}\mathbb{E}\left[Y_t^2\right] &= \sum_{t\geq 1}\mathbb{E}\left[Y_t^2\right]\sum_{m:2^m\geq t}\frac{1}{2^{2m}} \\ &= \frac{4}{3}\sum_{t\geq 1}\mathbb{E}\left[Y_t^2\right]\cdot\frac{1}{2^{2p_t}} \\ &\leq \frac{4}{3}\sum_{t\geq 1}\mathbb{E}\left[Y_t^2\right]\cdot\frac{1}{t^2}.\end{aligned} \tag{1.4}$$

Applying the eq. (1.3) to eq. (1.4),

$$\begin{aligned}\sum_{t\geq 1}\frac{1}{t^2}\cdot\mathbb{E}\left[Y_t^2\right] &\leq \sum_{t\geq 1}\frac{1}{t^2}\sum_{m\in[1,t]}(2m-1)\cdot\Pr\left[|X_1| > m-1\right] \\ &= \sum_{m\geq 1}(2m-1)\cdot\Pr\left[|X_1| > m-1\right]\cdot\sum_{t\geq m}\frac{1}{t^2}.\end{aligned}$$

By integration,

$$\sum_{t\geq m}\frac{1}{t^2} \leq \int_{m-\frac{1}{2}}^{\infty}\frac{dt}{t^2} = \frac{2}{2m-1},$$

then

$$\begin{aligned}\sum_{m\geq 1}(2m-1)\cdot\Pr\left[|X_1| > m-1\right]\cdot\sum_{t\geq m}\frac{1}{t^2} &\leq 2\sum_{m\geq 1}\Pr\left[|X_1| > m-1\right] \\ &\leq 2\left(1+\mathbb{E}\left[|X_1|\right]\right) < \infty,\end{aligned}$$

which completes the proof. □

We introduce Etemadi Inequality [Ete81] for proof of lemma 1.7.33.

Lemma 1.7.35 (Etemadi Inequality [Ete81])**.** *Let $\{X_i\}_{i\in[1,n]}$ be i.i.d. real-valued random variables defined over some common probability space, and let $\alpha \geq 0$, then*

$$\Pr\left[\max_{k\in[1,n]}\left|\sum_{i\in[1,k]}X_i\right| \geq 3\alpha\right] \leq 3\max_{k\in[1,n]}\Pr\left[\left|\sum_{i\in[1,k]}X_i\right| \geq \alpha\right].$$

For lemma 1.7.33, the proof follows.

Proof. We show only finite many m have $\max_{2^m<n\le 2^{m+1}} |S_n - S_{2^m}| \ge 3\varepsilon \cdot 2^m$ by lemma 1.7.30, then by lemma 1.7.35,

$$\sum_{m\ge 1} \Pr\left[\max_{2^m<n\le 2^{m+1}} \left|\frac{S_n - S_{2^m}}{2^m}\right| \ge 3\varepsilon\right] \le 3\sum_{m\ge 1} \max_{2^m<n\le 2^{m+1}} \Pr\left[\left|\frac{S_n - S_{2^m}}{2^m}\right| \ge \varepsilon\right]. \tag{1.5}$$

Again by Chebyshev's inequality,

$$\begin{aligned}
\max_{2^m<n\le 2^{m+1}} \Pr\left[\left|\frac{S_n - S_{2^m}}{2^m}\right| \ge \varepsilon\right] &\le \max_{2^m<n\le 2^{m+1}} \frac{\operatorname{Var}[S_n - S_{2^m}]}{\varepsilon^2 \cdot 2^{2m}} \\
&= \max_{2^m<n\le 2^{m+1}} \frac{1}{\varepsilon^2 \cdot 2^{2m}} \sum_{i\in[2^m+1,n]} \operatorname{Var}[W_i] \\
&= \frac{1}{\varepsilon^2 \cdot 2^{2m}} \sum_{i\in[1,2^m]} \operatorname{Var}[W_{2^m+i}] \\
&\le \frac{1}{\varepsilon^2 \cdot 2^{2m}} \sum_{i\in[1,2^m]} \mathbb{E}\left[Y_{2^m+i}^2\right].
\end{aligned} \tag{1.6}$$

The second equality holds by linearity of variance of independent random variables, the third equality holds by the non-negativity of variance, and the last inequality holds by

$$\operatorname{Var}[W_i] = \operatorname{Var}[Y_i] \le \mathbb{E}\left[Y_i^2\right].$$

By lemma 1.7.32, the right hand side upper bound of eq. (1.6) is summable by

$$\sum_{m\ge 1} \frac{1}{\varepsilon^2 \cdot 2^{2m}} \sum_{i\in[1,2^m]} \mathbb{E}\left[Y_{2^m+i}^2\right] \le \sum_{m\ge 1} \frac{4}{\varepsilon^2 \cdot 2^{2(m+1)}} \sum_{i\in[1,2^{m+1}]} \mathbb{E}\left[Y_i^2\right] < \infty,$$

finishes upper bounding the left hand side of eq. (1.5),

$$\sum_{m\ge 1} \Pr\left[\max_{2^m<n\le 2^{m+1}} \left|\frac{S_n - S_{2^m}}{2^m}\right| \ge 3\varepsilon\right] < \infty,$$

then we conclude lemma 1.7.33 by lemma 1.7.30, that

$$\Pr\left[\limsup_{m\to\infty} \max_{2^m<n\le 2^{m+1}} \left|\frac{S_n - S_{2^m}}{2^m}\right| \ge \varepsilon\right] = 0.$$

□

Corollary 1.7.36. *For every $\varepsilon > 0$,*

$$\Pr\left[\limsup_{n\to\infty} \left|\frac{S_n}{n}\right| \ge \varepsilon\right] = 0.$$

Proof. By Chebyshev's inequality,

$$\Pr\left[\left|\frac{S_{2^m}}{2^m}\right| \ge \frac{\varepsilon}{2}\right] \le \frac{4}{\varepsilon^2 \cdot 2^{2m}} \sum_{i\in[1,2^m]} \mathbb{E}\left[Y_i^2\right].$$

Hence,

$$\sum_{m\ge 1} \Pr\left[\left|\frac{S_{2^m}}{2^m}\right| \ge \frac{\varepsilon}{2}\right] < \infty$$

by lemma 1.7.32. Therefore, by lemma 1.7.30,

$$\Pr\left[\limsup_{m\to\infty} \left|\frac{S_{2^m}}{2^m}\right| \ge \frac{\varepsilon}{2}\right] = 0.$$

Moreover, for $2^m < n \leq 2^{m+1}$, by triangle inequality

$$\left|\frac{S_n}{2^m}\right| \leq \left|\frac{S_n - S_{2^m}}{2^m}\right| + \left|\frac{S_{2^m}}{2^m}\right|,$$

so

$$\left\{\limsup_{\substack{m,n\to\infty \\ 2^m<n\leq 2^{m+1}}} \left|\frac{S_n}{2^m}\right| \geq \varepsilon\right\} \subseteq \left\{\limsup_{m\to\infty} \max_{2^m<n\leq 2^{m+1}} \left|\frac{S_n - S_{2^m}}{2^m}\right| \geq \frac{\varepsilon}{2}\right\} \cup \left\{\limsup_{m\to\infty} \left|\frac{S_{2^m}}{2^m}\right| \geq \frac{\varepsilon}{2}\right\}.$$

Since $2^m < n$ implies $|S_n/n| \leq |S_n/2^m|$, we have

$$\left\{\limsup_{n\to\infty} \left|\frac{S_n}{n}\right| \geq \varepsilon\right\} \subseteq \left\{\limsup_{\substack{m,n\to\infty \\ 2^m<n\leq 2^{m+1}}} \left|\frac{S_n}{2^m}\right| \geq \varepsilon\right\}.$$

Therefore, by lemma 1.7.33,

$$\Pr\left[\limsup_{n\to\infty} \left|\frac{S_n}{n}\right| \geq \varepsilon\right] = 0.$$

□

To complete the proof for Kolmogorov SLLN theorem 1.7.28, by corollary 1.7.36,

$$\Pr\left[\limsup_{n\to\infty} \left|\frac{S_n}{n}\right| \geq \varepsilon\right] = 0.$$

Since $\varepsilon > 0$ was arbitrary, we conclude that

$$\Pr\left[\lim_{n\to\infty} \frac{S_n}{n} = 0\right] = 1.$$

On the other hand, by lemma 1.7.34,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[1,n]} \mathbb{E}[Y_i] = 0.$$

Therefore, on the event $\{\lim_{n\to\infty} S_n/n = 0\}$, we have

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[1,n]} Y_i = \lim_{n\to\infty} \left(\frac{S_n}{n} + \frac{1}{n} \sum_{i\in[1,n]} \mathbb{E}[Y_i]\right) = 0,$$

so

$$\Pr\left[\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[1,n]} Y_i = 0\right] = 1.$$

Together with

$$\Pr\left[\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[1,n]} Z_i = 0\right] = 1,$$

and $X_i = Y_i + Z_i$, we conclude that

$$\Pr\left[\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[1,n]} X_i = 0\right] = 1.$$

□

Definition 1.7.37 (Stopping Time)**.** Let $\{X_t\}_{t\geq 0}$ be a discrete-time stochastic process. A non-negative integer-valued random variable T is a stopping time if for every $n \geq 0$, whether event $T = n$ occurs is determined only by $X_0, \ldots, X_n$.

Definition 1.7.38 (Strong Markov Property for Discrete-Time Markov Chains)**.** Let $\{X_t\}_{t\geq 0}$ be a discrete-time Markov chain. We say that $\{X_t\}_{t\geq 0}$ satisfies the strong Markov property if for every stopping time T with $\Pr[T < \infty] = 1$, conditional on $T = n$ and the history

$$X_0 = x_0, X_1 = x_1, \ldots, X_{n-1} = x_{n-1}, X_n = i,$$

the shifted process $\{X_{T+m}\}_{m\geq 0}$ has the same distribution as the original chain started from the state i.

Theorem 1.7.39. *Any discrete-time, time-homogeneous Markov chain satisfies the strong Markov property.*

Proof. Take any states $x_0, \ldots, x_{n-1}, i, j_1, \ldots, j_r$, and any time steps $n \geq 0$ and $0 \leq m_1 < \cdots < m_r$. We show that conditional on $T = n$ and the history $X_0 = x_0, \ldots, X_n = i$, the shifted process has the same finite-dimensional distributions as the original chain started from state i. Since event $T = n$ is in the condition,

$$\begin{aligned}&\Pr\left[X_{T+m_1} = j_1, \ldots, X_{T+m_r} = j_r \mid T = n, X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i\right] \\ &= \Pr\left[X_{n+m_1} = j_1, \ldots, X_{n+m_r} = j_r \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i\right].\end{aligned}$$

By the Markov property applied successively,

$$\begin{aligned}&\Pr\left[X_{n+m_1} = j_1, \ldots, X_{n+m_r} = j_r \mid X_0 = x_0, \ldots, X_{n-1} = x_{n-1}, X_n = i\right] \\ &= \Pr\left[X_{n+m_1} = j_1 \mid X_n = i\right] \cdot \prod_{\ell\in[2,r]} \Pr\left[X_{n+m_\ell} = j_\ell \mid X_{n+m_{\ell-1}} = j_{\ell-1}\right].\end{aligned}$$

By time-homogeneity in definition 1.7.4, this equals

$$\Pr\left[X_{m_1} = j_1 \mid X_0 = i\right] \cdot \prod_{\ell\in[2,r]} \Pr\left[X_{m_\ell - m_{\ell-1}} = j_\ell \mid X_0 = j_{\ell-1}\right],$$

which is exactly

$$\Pr\left[X_{m_1} = j_1, \ldots, X_{m_r} = j_r \mid X_0 = i\right].$$

Therefore, conditional on $T = n$ and the history up to time n, the shifted process $\{X_{T+m}\}_{m\geq 0}$ has the same finite-dimensional distributions as the original chain started from state i. Hence it has the same distribution, and the chain satisfies the strong Markov property. □

Theorem 1.7.40 (SLLN for Markov chains)**.** *Let $\{X_t\}_{t\in T}$ be a Markov chain, with $X_0 = i$, and states i, j communicating. Let $1^t_{i,j}$ be the indicator random variable for the event $X_t = j$ given $X_0 = i$. Then the limiting fraction of time spent in state j is $h^{-1}_{j,j}$, that is,*

$$\Pr\left[\lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} 1^t_{i,j} = \frac{1}{h_{j,j}}\right] = 1.$$

Proof. If i is transient, and i, j are in same communicating class, then by corollary 1.7.20, j is also transient. Intuitively, there are only finitely many times that i can reach j. We can show by a similar trick in corollary 1.7.20, and let n be a number of steps such that $P^n_{j,i} > 0$, then

$$\infty > \sum_{\tau\geq 1} P^\tau_{i,i} \geq \sum_{\tau\geq n+1} P^\tau_{i,i} = \sum_{t\geq 1}\sum_k P^t_{i,k} \cdot P^n_{k,i} \geq \sum_{t\geq 1} P^t_{i,j} \cdot P^n_{j,i} = P^n_{j,i} \cdot \sum_{t\geq 1} P^t_{i,j} = P^n_{j,i} \cdot \sum_{t\geq 1} \mathbb{E}\left[1^t_{i,j}\right].$$

Since $\sum_t \mathbb{E}\left[1^t_{i,j}\right] < \infty$, then $\sum_t 1^t_{i,j} = \infty$ with probability 0, hence the left hand side is 0. By corollary 1.7.22, $h_{j,j} = \infty$, thus the right hand side is 0. Hence, the transient case is proved.

If i is recurrent, then i, j are both recurrent by corollary 1.7.20.

We first handle the positive recurrent case, and start with the special case when $i = j$. Suppose that j is positive recurrent, so $h_{j,j} < \infty$. Let $R_0 = 0$, and for each $c \geq 1$, let R_c be the time of the c^{th} visit to state j, where

$$R_c = \min\{t > R_{c-1} \mid X_t = j\},$$

Then the cycle lengths are

$$C_c = R_c - R_{c-1},$$

and consequently, R_c is the total length of the first c cycles:

$$R_c = \sum_{k \in [1,c]} C_k.$$

One checks that each R_c is a stopping time by definition 1.7.37. Then by the strong Markov property theorem 1.7.39 at the successive return times to j and by time-homogeneity, the random variables $\{C_c\}_{c \geq 1}$ are i.i.d., with $\mathbb{E}[C_1] = h_{j,j}$.

Let k_n be the largest number of cycles such that $R_{k_n} \leq n$ by

$$k_n = \max\{k \mid R_k \leq n\}.$$

Hence, $R_{k_n} \leq n < R_{k_n+1}$ by definition, and consequently,

$$\frac{R_{k_n}}{k_n} \leq \frac{n}{k_n} \leq \frac{R_{k_n+1}}{k_n}.$$

Since j is recurrent, we have $k_n \to \infty$ as $n \to \infty$. Then, by SLLN theorem 1.7.28,

$$\Pr\left[\lim_{k\to\infty} \frac{R_k}{k} = \lim_{k\to\infty} \frac{1}{k} \sum_{c \in [1,k]} C_c = h_{j,j}\right] = 1,$$

and

$$\lim_{k\to\infty} \frac{R_{k+1}}{k} = \lim_{k\to\infty} \frac{R_{k+1}}{k+1} \cdot \frac{k+1}{k} = \lim_{k\to\infty} \frac{R_{k+1}}{k+1}. \tag{1.7}$$

Hence, when $i = j$,

$$\Pr\left[\lim_{n\to\infty} \frac{k_n}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{t \in [1,n]} 1^t_{j,j} = \frac{1}{h_{j,j}}\right] = 1.$$

We next handle the null recurrent case when $i = j$. Suppose that j is null recurrent, so $h_{j,j} = \infty$. Reuse the same notations $\{R_c\}_{c \geq 0}$, $\{C_c\}_{c \geq 1}$, and $\{k_n\}_{n \geq 1}$. For each positive integer m, define the truncated cycle lengths

$$C_c^{(m)} = \min(C_c, m).$$

Since $\{C_c\}_{c \geq 1}$ are i.i.d., the bounded random variables $\{C_c^{(m)}\}_{c \geq 1}$ are also i.i.d. Hence, by SLLN theorem 1.7.28,

$$\Pr\left[\lim_{k\to\infty} \frac{1}{k} \sum_{c \in [1,k]} C_c^{(m)} = \mathbb{E}\left[C_1^{(m)}\right]\right] = 1.$$

Since $\mathbb{E}\left[C_1^{(m)}\right]$ increases monotonically as $m \to \infty$, by monotone convergence, $\mathbb{E}\left[C_1^{(m)}\right] \to \mathbb{E}[C_1] = h_{j,j} = \infty$ by null recurrence. Then, for any positive integer α, there exists a sufficiently large m such that $\mathbb{E}\left[C_1^{(m)}\right] > \alpha$.

$$\Pr\left[\lim_{k\to\infty} \frac{1}{k} \sum_{c \in [1,k]} C_c^{(m)} > \alpha\right] = 1.$$

Since for each fixed m, $C_c^{(m)} \leq C_c$ pointwise, then

$$R_k = \sum_{c\in[1,k]} C_c \geq \sum_{c\in[1,k]} C_c^{(m)}.$$

Hence, for any fixed positive integer α,

$$\Pr\left[\liminf_{k\to\infty} \frac{R_k}{k} > \alpha\right] = 1.$$

Since this holds for every positive integer α, we conclude that

$$\Pr\left[\lim_{k\to\infty} \frac{R_k}{k} = \infty\right] = 1.$$

Since j is recurrent, we still have $k_n \to \infty$ as $n \to \infty$, and $n \geq R_{k_n}$. By prior squeezing argument eq. (1.7), when $i = j$,

$$\Pr\left[\lim_{n\to\infty} \frac{k_n}{n} = 0\right] = 1.$$

For the general recurrent case when $i \neq j$, let $T_{i,j}$ be the first hitting time from i to j. By the strong Markov property at time $T_{i,j}$, the path after time $T_{i,j}$ behaves like a fresh copy of the chain started from j. Thus the path consists of an initial segment of length $T_{i,j}$, after which the chain proceeds through successive j-cycles. Let $k_n = \sum_{t\in[1,n-T_{i,j}]} 1_{j,j}^t$ for the cycles before $n - T_{i,j}$ steps, then

$$\frac{R_{k_n} + T_{i,j}}{k_n} \leq \frac{n}{k_n} < \frac{R_{k_n+1} + T_{i,j}}{k_n}.$$

By theorem 1.7.26, $\Pr\left[T_{i,j} < \infty\right] = 1$, and thus

$$\Pr\left[\lim_{k\to\infty} \frac{T_{i,j}}{k} = 0\right] = 1,$$

hence the same squeezing argument eq. (1.7) as in the $i = j$ case yields

$$\Pr\left[\lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} 1_{i,j}^t = \frac{1}{h_{j,j}}\right] = 1.$$

□

Theorem 1.7.41 (Bounded Convergence Theorem)**.** *For a sequence of random variables $\{X_i\}_{i\geq 1}$, if*

$$\Pr\left[\lim_{n\to\infty} X_n = X\right] = 1$$

and there exists a finite b such that $|X_n| \leq b$ for all n, then

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \mathbb{E}[X].$$

Corollary 1.7.42 (Cesàro Limit of Return Probability)**.** *Let $\{X_t\}_{t\in T}$ be an irreducible Markov chain, then for any state i, j,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} P_{i,j}^t = \frac{1}{h_{j,j}}.$$

Proof.

$$\lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} P_{i,j}^t = \lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} \mathbb{E}\left[1_{i,j}^t\right] = \lim_{n\to\infty} \mathbb{E}\left[\frac{1}{n} \sum_{t\in[1,n]} 1_{i,j}^t\right] = \mathbb{E}\left[\lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} 1_{i,j}^t\right] = \frac{1}{h_{j,j}},$$

where the first 2 equalities are from linearity of expectation, and the third equality is from theorem 1.7.41. □

Lemma 1.7.43. *Positive recurrence and null recurrence are class properties.*

Proof. By corollary 1.7.20, states in a communicating class S are either all recurrent or all transient. Hence, we focus on recurrent class. Let state i, j be in S, and on $t_0, t_1 > 0$ we have $P_{i,j}^{t_0}, P_{j,i}^{t_1} > 0$. By corollary 1.7.42 and theorem 1.7.13,

$$\begin{aligned}
\frac{1}{h_{i,i}} = \lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} P_{i,i}^t &\geq \lim_{n\to\infty} \frac{1}{n} \sum_{t\in[t_0+t_1+1,n]} P_{i,i}^t \\
&\geq P_{i,j}^{t_0} \cdot P_{j,i}^{t_1} \cdot \lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n-t_0-t_1]} P_{j,j}^t \\
&= P_{i,j}^{t_0} \cdot P_{j,i}^{t_1} \cdot \lim_{n\to\infty} \frac{n-t_0-t_1}{n} \cdot \frac{1}{n-t_0-t_1} \sum_{t\in[1,n-t_0-t_1]} P_{j,j}^t \\
&= P_{i,j}^{t_0} \cdot P_{j,i}^{t_1} \cdot \lim_{n\to\infty} \frac{1}{n} \sum_{t\in[1,n]} P_{j,j}^t = P_{i,j}^{t_0} \cdot P_{j,i}^{t_1} \cdot \frac{1}{h_{j,j}}.
\end{aligned}$$

The proof idea goes similar to corollary 1.7.20. If i is null recurrent by $h_{i,i}^{-1} = 0$, then j has to be null recurrent; conversely, if j is positive recurrent with $h_{j,j}^{-1} > 0$, then $h_{i,i}^{-1} > 0$ means i is positive recurrent. □

1.7.4 Stationary Distributions

Definition 1.7.44. A state j in a discrete time Markov chain $\mathbf{X} = \{X_t\}$ is periodic if there exists an integer $\Delta > 1$

$$\Pr\left[X_{t+s} = j \mid X_t = j\right] = 0$$

unless $\Delta \mid s$. $\mathbf{X}$ is periodic if $\exists\Delta > 1$ such that all states have period Δ. Otherwise, a state or chain is aperiodic.

Example 1.7.45. For a discrete time Markov chain defined over $\mathbb{Z}$, that i transitions to $i \pm 1$ with probability $1/2$ each, the discrete time Markov chain is periodic as $\Delta = 2$.

Definition 1.7.46. An aperiodic, positive recurrent state is ergodic. A Markov chain is ergodic if all states are ergodic.

Corollary 1.7.47. *Any finite, irreducible, aperiodic Markov chain is an ergodic Markov chain.*

Lemma 1.7.48 (Ergodic Limit of Return Probability)**.** *Let $\{X_t\}_{t\in T}$ be an irreducible ergodic Markov chain, the $\lim_{t\to\infty} P_{i,i}^t$ exists for any state i, and*

$$\lim_{t\to\infty} P_{i,i}^t = h_{i,i}^{-1}.$$

Definition 1.7.49 (Discrete Renewal Equation)**.** Discrete convolution for sequences $\mathbf{a} = \{a_i\}_{i\geq 0}$ and $\mathbf{b} = \{b_i\}_{i\geq 0}$ is

$$(\mathbf{a} * \mathbf{b})_n = \sum_{k\in[0,n]} a_k \cdot b_{n-k}.$$

A renewal kernel is a probability mass function $\mathbf{f} = \{f_i\}_{i\geq 1}$ on positive integer such that each $f_i \geq 0$ and $\sum_{i\geq 1} f_i = 1$. Given a forcing sequence $\mathbf{g} = \{g_i\}_{i\geq 0}$, a sequence $\mathbf{u} = \{u_i\}_{i\geq 0}$ solves the renewal equation if

$$u_n = g_n + \sum_{i\in[1,n]} f_i \cdot u_{n-i} = g_n + (\mathbf{f} * \mathbf{u})_{n-1}.$$

Theorem 1.7.50 (Discrete Key Renewal Theorem [Fel68b])**.** *Let $\mu = \sum_{t\geq 1} t \cdot f_t$, and the span for $\mathbf{f}$ be $d = \gcd\{f_k > 0\}_{k\geq 1}$. If $d = 1$ then $\mathbf{f}$ is aperiodic. Supposing $\mathbf{f}$ is aperiodic and $\mu < \infty$, and $\mathbf{g} = \{g_i\}_{i\geq 0}$ is absolutely summable by $\sum_{i\geq 0} |g_i| < \infty$. Let $\mathbf{u} = \{u_i\}_{i\geq 0}$ be the unique bounded solution to the renewal equation by $u_n = g_n + (\mathbf{f} * \mathbf{u})_{n-1}$, then the limit exists by*

$$\lim_{n\to\infty} u_n = \frac{1}{\mu} \sum_{i\geq 0} g_i.$$

The proof for lemma 1.7.48 is immediate by theorem 1.7.50.

Proof. Given an irreducible ergodic Markov chain, suppose $X_0 = i$, then $P^n_{i,i}$ can be derived as a renewal equation by

$$P^n_{i,i} = \sum_{t\in[1,n]} r^t_{i,i} \cdot P^{n-t}_{i,i} = \left(\{r^t_{i,i}\}_{t\geq 1} * \{P^t_{i,i}\}_{t\geq 0}\right)_{n-1},$$

and the "forcing sequence" is $r^0_{i,i} = 1$ follows 0s, as $P^0_{i,i} = r^0_{i,i} = 1$, thus the absolute sum is 1. Since the Markov chain is ergodic, then $\{r^t_{i,i}\}_{t\geq 1}$ is aperiodic and $\mu = h_{i,i} < \infty$. Hence,

$$\lim_{n\to\infty} P^n_{i,i} = \frac{1}{h_{i,i}}.$$

□

Example 1.7.51. Consider a sequence of fair, independent gambling games between 2 gamblers. One wins or loses a dollar with probability 1/2. The state of the chain at time t is the number of dollars won. Start from initial state 0.

We assume that a player loses by $-\ell_1$ dollars and wins by ℓ_2 dollars. Hence, these are the only 2 recurrent states, while the others are transitioning out to either $-\ell_1$ or ℓ_2. We write P^t_i for the probability of i dollars at the t^{th} steps. Let s be one of the transient states, then $\lim_{t\to\infty} P^t_i = 0$. Hence, let $q = \lim_{t\to\infty} P^t_{-\ell_1}$, then $\lim_{t\to\infty} P^t_{\ell_2} = 1 - q$.

Let W^t be the random variable for the gain of the player at the t^{th} step, where the game has not stopped yet. On one hand, the games are fair, then by linearity of expectation, $\mathbb{E}\left[W^t\right] = 0$. On the other hand,

$$\mathbb{E}\left[W^t\right] = \sum_{i\in[-\ell_1,\ell_2]} i \cdot P^t_i = -\ell_1 \cdot q + \ell_2 \cdot (1-q) = 0.$$

Therefore, $q = \ell_2/(\ell_1 + \ell_2)$.

Definition 1.7.52. A stationary (equilibrium) distribution of a Markov chain is a distribution $\mathcal{D}_\pi$ such that $\boldsymbol{\pi} = \mathbf{P}^\mathsf{T}\boldsymbol{\pi}$.

Theorem 1.7.53. *Any finite, irreducible, ergodic Markov chain has the following properties:*

- *The chain has a stationary distribution $\mathcal{D}_\pi$ such that $\boldsymbol{\pi} = [\pi_1, \dots, \pi_n]^\mathsf{T}$.*
- *For all $i, j \in [1, n]$, the $\lim_{t\to\infty} P^t_{j,i}$ exists, and is independent of j.*
- *$\pi_i = \lim_{t\to\infty} P^t_{j,i} = h^{-1}_{i,i}$ for all $j \in [1, n]$.*

Remark 1.7.54. The theorem gives 2 interpretations for the stationary distribution $\mathcal{D}_\pi$:

- If we run the finite, irreducible, ergodic Markov chain sufficiently long, then the initial state is forgotten, and the probability of being in state i is given by π_i.
- Given the prior observation of forgetting the initial state, then the limiting probability of $P^t_{i,i}$ being $h^{-1}_{i,i}$ becomes reasonable, that any initial state ends up in state i at the t^{th} step has same probability of $\pi_i = P^t_{i,i}$. Since $h_{i,i}$ is the average returning time from state i, then we should expect to be in state i with probability $h^{-1}_{i,i}$.

The proof for theorem 1.7.53 is immediate by lemma 1.7.48, together with showing $\lim_{t\to\infty} P^t_{j,i} = \lim_{t\to\infty} P^t_{i,i}$.

Proof. Given a state j that $j \neq i$, then similar to exercise 1.7.19, we write

$$P^t_{j,i} = \sum_{k\in[1,t]} r^k_{j,i} \cdot P^{t-k}_{i,i}.$$

Then for a $t_1 \leq t$, we have

$$\lim_{t\to\infty} P^t_{j,i} = \lim_{t\to\infty} \sum_{k\in[1,t]} r^k_{j,i} \cdot P^{t-k}_{i,i} \geq \lim_{t\to\infty} \sum_{k\in[1,t_1]} r^k_{j,i} \cdot P^{t-k}_{i,i} = \lim_{t\to\infty} P^t_{i,i} \sum_{k\in[1,t_1]} r^k_{j,i} \geq (1-\varepsilon) \cdot \lim_{t\to\infty} P^t_{i,i}.$$

On the other hand, since $\sum_{k\in[1,t_1]} r_{j,i}^k \geq 1-\varepsilon$, which leads to

$$\begin{aligned}\lim_{t\to\infty} P_{j,i}^t &= \lim_{t\to\infty} \sum_{k\in[1,t_1]} r_{j,i}^k \cdot P_{i,i}^{t-k} + \lim_{t\to\infty} \sum_{k\in[t_1+1,t]} r_{j,i}^k \cdot P_{i,i}^{t-k} \\ &\leq \lim_{t\to\infty} \sum_{k\in[1,t_1]} r_{j,i}^k \cdot P_{i,i}^{t-k} + \lim_{t\to\infty} \sum_{k\in[t_1+1,t]} r_{j,i}^k \\ &\leq \lim_{t\to\infty} P_{i,i}^t \sum_{k\in[1,t_1]} r_{j,i}^k + \varepsilon \\ &\leq \lim_{t\to\infty} P_{i,i}^t + \varepsilon.\end{aligned}$$

Since $\varepsilon \geq \sum_{k>t_1} r_{j,i}^k$, as $t_1 < t$ approaches ∞, $\varepsilon \to 0$, which proves the limiting probability of $P_{j,i}^t$ is independent of j.

We now show $\sum \pi_i = 1$ and $\langle \boldsymbol{\pi}, [P_{1,i}, \ldots, P_{n,i}]^\mathsf{T} \rangle = \pi_i$ for all $i \in [1,n]$ to complete the proof. Fix a state $j \in [1,n]$,

$$\sum_{i\in[1,n]} \pi_i = \sum_{i\in[1,n]} \lim_{t\to\infty} P_{i,i}^t = \sum_{i\in[1,n]} \lim_{t\to\infty} P_{j,i}^t = 1,$$

and by theorem 1.7.13,

$$\sum_{k\in[1,n]} \pi_k \cdot P_{k,i} = \sum_{k\in[1,n]} \lim_{t\to\infty} P_{k,k}^t \cdot P_{k,i} = \sum_{k\in[1,n]} \lim_{t\to\infty} P_{j,k}^t \cdot P_{k,i} = \lim_{t\to\infty} P_{j,i}^{t+1} = \lim_{t\to\infty} P_{i,i}^{t+1} = \pi_i.$$

Together with lemma 1.7.48, we complete the proof. □

Remark 1.7.55. Being aperiodic in being ergodic is not necessary for the existence of the stationary distribution, but in this case, it is not the limiting probability, but rather the long term frequency of visiting states.

Theorem 1.7.56. *Let S be a set of states in a finite, irreducible, ergodic Markov chain. In the stationary distribution, the probability of leaving the state is equal to the probability of entering the state.*

Proof. Let there be a state $i \in [1,n]$ for the finite, irreducible, ergodic Markov chain, then by stationary distribution

$$\langle \boldsymbol{\pi}, [P_{1,i}, \ldots, P_{n,i}]^\mathsf{T} \rangle = \pi_i = \pi_i \cdot \sum_{j\in[1,n]} P_{i,j},$$

that the invariant reaches by

$$\sum_{j\neq i} \pi_j \cdot P_{j,i} = \sum_{j\neq i} \pi_i \cdot P_{i,j}.$$

By generalizing to the subset S, the proof is completed. □

Remark 1.7.57. The theorem 1.7.56 gives another way of deriving stationary distribution of $\mathbf{P}$ on a graph perspective. The cut-set around state i holds an invariant of $\sum_{j\neq i} \pi_j \cdot P_{j,i} = \sum_{j\neq i} \pi_i \cdot P_{i,j}$. If C is a cut-set around $S \subseteq [1,n]$, then in stationary distribution the probability crossing the cut in one direction is equivalent to the probability crossing the cut in the other direction.

Theorem 1.7.58. *Consider a finite, irreducible, ergodic Markov chain with a transition matrix $\mathbf{P}$. If there are nonnegative $\boldsymbol{\pi}$ such that $\sum_i \pi_i = 1$, and if any pairs $i, j \in [1,n]$ such that $\pi_j \cdot P_{j,i} = \pi_i \cdot P_{i,j}$, then $\boldsymbol{\pi}$ is a stationary distribution $\mathcal{D}_{\boldsymbol{\pi}}$ for $\mathbf{P}$. Chains satisfying the condition are time reversible.*

Theorem 1.7.59. *An irreducible, aperiodic Markov chain belongs to one of the following two categories:*

- *The chain is ergodic: For any pair of states i, j, $\lim_{t\to\infty} P_{j,i}^t$ exists and is independent of j, and $\mathcal{D}_{\boldsymbol{\pi}}$ has $\pi_i = \lim_{t\to\infty} P_{j,i}^t > 0$.*
- *The chain has no positive recurrent state: For any state i, $\lim_{t\to\infty} P_{i,i}^t = 0$, and the chain has no $\mathcal{D}_{\boldsymbol{\pi}}$.*

Proof. Immediate by lemma 1.7.43, states in an irreducible ergodic Markov chain are either all positive recurrent or all null recurrent. Hence, by lemma 1.7.48, if the irreducible aperiodic Markov chain is positive recurrent,

$$\lim_{t\to\infty} P_{i,i}^t = \frac{1}{h_{i,i}}.$$

Showing the limiting probability $P_{j,i}^t$ is independent of j can be done by the same sandwiching trick in theorem 1.7.53.

Otherwise, the limiting recurrent probability of a null recurrent Markov chain has to be 0. By corollary 1.7.42,

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[1,n]} P_{j,i}^t = 0,$$

since $h_{i,i} = \infty$ for null recurrent states. $\lim_{t\to\infty} P_{j,i}^t > 0$ contradicts the Cesàro summation, hence $\lim_{t\to\infty} P_{j,i}^t = 0$. □

Remark 1.7.60. Irreducibility and ergodicity leads to convergence, existence and uniqueness of stationary distribution. For both finite and countably infinite Markov chains, both needs to be irreducible and aperiodic, but in finite case, positive recurrence derive from the finiteness, hence the ergodicity derive from aperiodicity and positive recurrence, while the countably infinite case need to decide if the states are positive recurrent.

Exercise 1.7.61 (Exercise 7.11 [MU17])**.** An $n \times n$ matrix $\mathbf{P}$ is called stochastic if all entries are nonnegative and the sum of each row is 1. It is doubly stochastic if the sum of each column is 1. Show that the uniform distribution is a stationary distribution for any Markov chain represented by a doubly stochastic matrix.

Proof. Supposing

$$\sum_{j\in[1,n]} P_{i,j} = \sum_{j\in[1,n]} P_{j,i} = 1$$

for a doubly stochastic matrix, then the stationary distribution $\{\pi_i\}_{i\in[1,n]}$ has

$$\sum_{j\in[1,n]} \pi_j \cdot P_{j,i} = \frac{1}{n} = \pi_i$$

for an arbitrary $i \in [1, n]$, which completes the proof. □

Exercise 1.7.62 (Exercise 7.12 [MU17])**.** Let X_n be the sum of n independent fair dice rolls. Show that for any $k \geq 2$,

$$\lim_{n\to\infty} \Pr[X_n \equiv 0 \pmod k] = \frac{1}{k}.$$

Proof. Let the current accumulation be X_n, $Y_n = X_n \pmod k$ be the accumulation modulo k, and $P_{i,j}$ be the transition probability over k states of Y_n, then $P_{i,j}$ can be represented as follows

$$P_{i,j} = \sum_{\substack{\alpha\in[1,6] \\ i+\alpha\equiv j \pmod k}} \frac{1}{6}.$$

Hence, $P_{a,b} = P_{b,c}$ if $a + \alpha \equiv b \pmod k$ and $b + \alpha \equiv c \pmod k$. Moreover, we discover the invariant of

$$\sum_{j\in[0,k-1]} P_{j,i} = \sum_{j\in[0,k-1]} P_{0,j} = 1,$$

making the uniform distribution the stationary distribution by exercise 1.7.61.

The chain is irreducible, as any 2 states can be reached from each other. The chain is also aperiodic, by existence of path of length $k-1$: rolling 1 for exactly $k-2$ times and 2 for a time, and path of length k: rolling 1 for k times. Since $\gcd(k, k-1) = 1$, the chain is aperiodic, and hence it is ergodic. By theorem 1.7.53,

$$\lim_{n\to\infty} \Pr[X_n \equiv 0 \pmod k] = \lim_{n\to\infty} P_{0,0}^n = \frac{1}{k}.$$

□

Exercise 1.7.63 (Exercise 7.13 [MU17]). Consider a finite n state Markov chain with stationary distribution $\boldsymbol{\pi}$ and transition probability $P_{i,j}$. Starting at time 0 and running for m states, obtaining a sequence of states $X_0, X_1, \ldots, X_m$. Consider the states in reverse order, $X_m, \ldots, X_1, X_0$.

- Argue that given X_{t+1}, X_t is independent of $X_{t+2}, \ldots, X_m$, that the reverse sequence is Markovian.
- Argue that for the reverse sequence, the transition probability is given by

$$Q_{i,j} = \frac{\pi_j \cdot P_{j,i}}{\pi_i}.$$

- Show that if the original Markov chain is time reversible, that $\pi_i \cdot P_{i,j} = \pi_j \cdot P_{j,i}$, then $Q_{i,j} = P_{i,j}$: the state follows the same transition probability, whether viewed in forward order or reverse order.

Proof. We already know the Markovian by

$$\Pr[X_{t+1} = s_{t+1} \mid X_t = s_t, \ldots, X_1 = s_1, X_0 = s_0] = \Pr[X_{t+1} = s_{t+1} \mid X_t = s_t].$$

By Bayes' rule,

$$\Pr[X_t = s_t \mid X_{t+1} = s_{t+1}, \ldots, X_m = s_m] = \frac{\Pr[X_t = s_t, \ldots, X_m = s_m]}{\Pr[X_{t+1} = s_{t+1}, \ldots, X_m = s_m]},$$

the numerator can be decomposed by

$$\Pr[X_t = s_t, \ldots, X_m = s_m] = \Pr[X_t = s_t] \cdot \Pr[X_{t+1} = s_{t+1} \mid X_t = s_t] \cdots \Pr[X_m = s_m \mid X_{m-1} = s_{m-1}],$$

and the denominator can be decomposed by

$$\Pr[X_{t+1} = s_{t+1}, \ldots, X_m = s_m] = \Pr[X_{t+1} = s_{t+1}] \cdot \Pr[X_{t+2} = s_{t+2} \mid X_{t+1} = s_{t+1}] \cdots \Pr[X_m = s_m \mid X_{m-1} = s_{m-1}].$$

Hence, we prove the independence by

$$\Pr[X_t = s_t \mid X_{t+1} = s_{t+1}, \ldots, X_m = s_m] = \frac{\Pr[X_t = s_t, X_{t+1} = s_{t+1}]}{\Pr[X_{t+1} = s_{t+1}]} = \Pr[X_t = s_t \mid X_{t+1} = s_{t+1}].$$

Given 2 states i, j, when $t \to \infty$,

$$\lim_{t\to\infty} \Pr\left[X_t = j \mid X_{t+1} = i\right] = \lim_{t\to\infty} \frac{\Pr\left[X_t = j, X_{t+1} = i\right]}{\Pr[X_{t+1} = i]} = \lim_{t\to\infty} \frac{\Pr\left[X_{t+1} = i \mid X_t = j\right] \cdot \Pr\left[X_t = j\right]}{\Pr[X_{t+1} = i]},$$

hence

$$Q_{i,j} = \frac{\pi_j \cdot P_{j,i}}{\pi_i}.$$

If the chain is time reversible, then it is straightforward that $Q_{i,j} = P_{i,j}$ by swapping in $\pi_j \cdot P_{j,i} = \pi_i \cdot P_{i,j}$. □

Exercise 1.7.64 (Exercise 7.18 [MU17]). Consider a random walk over 1/2/3-dimensional integer lattice, is each state transient, null recurrent, or positive recurrent? [3]

Proof. Consider the 1-dimensional random walk, and the particle returns to 0 only on even length steps. Hence,

$$P_{0,0}^{2m} = \frac{1}{2^{2m}} \cdot \binom{2m}{m} \geq \frac{1}{2^{2m}} \cdot \frac{\sqrt{2\pi \cdot 2m} \cdot (2m)^{2m}}{\left(\sqrt{2\pi \cdot m} + 1\right)^2 \cdot m^{2m}} \geq \frac{1}{\sqrt{2\pi \cdot 2m}},$$

[3] A drunk man will find his way home, but a drunk bird may get lost forever. — Shizuo Kakutani

where the second inequality holds from lemma 1.7.1, and the third inequality holds from $\sqrt{2\pi \cdot 2m} > \sqrt{2\pi \cdot m} + 1$. By exercise 1.7.19, that

$$\sum_{t\geq 1} P^t_{0,0} = \sum_{m\geq 1} P^{2m}_{0,0} \geq \frac{1}{2\sqrt{\pi}} \cdot \sum_{m\geq 1} \frac{1}{\sqrt{m}}$$

is unbounded, and hence the 1-dimensional random walk is at least recurrent by corollary 1.7.20.

When we expand to 2-dimensional random walk, we want to show the probability of $P^{2m}_{0^2,0^2}$ from $P^{2t}_{0,0}$ by

$$\begin{aligned}
P^{2m}_{0^2,0^2} &= \sum_{a+b=m} \frac{1}{2^{2m}} \cdot \binom{2m}{2a} \cdot P^{2a}_{0,0} \cdot P^{2b}_{0,0} \\
&= \sum_{a+b=m} \frac{1}{2^{2m}} \cdot \binom{2m}{2a} \cdot \frac{1}{2^{2m}} \cdot \binom{2a}{a}\binom{2b}{b} \\
&= \frac{1}{4^{2m}} \cdot \sum_{a+b=m} \frac{(2m)!}{(2b)! \cdot (2a)!} \cdot \frac{(2a)!}{(a!)^2} \cdot \frac{(2b)!}{(b!)^2} \\
&= \frac{1}{4^{2m}} \cdot \sum_{a+b=m} \binom{2m}{m}\binom{m}{a}\binom{m}{b} = \frac{1}{4^{2m}} \cdot \binom{2m}{m}^2 = \left(P^{2m}_{0,0}\right)^2,
\end{aligned}$$

hence the 2-dimensional random walk is also at least recurrent by corollary 1.7.20 and

$$\sum_{t\geq 1} P^t_{0^2,0^2} = \sum_{m\geq 1} P^{2m}_{0^2,0^2} = \sum_{m\geq 1} \left(P^{2m}_{0,0}\right)^2 \geq \frac{1}{4\pi} \cdot \sum_{m\geq 1} \frac{1}{m} = \infty.$$

For a random walk of dimension $d \geq 3$, the exact probability of recurrence after $2m$ moves is

$$\begin{aligned}
P^{2m}_{0^d,0^d} &= \sum_{a_1+\cdots+a_d=m} \frac{1}{d^{2m}} \cdot \binom{2m}{2a_1} \cdots \binom{2a_{d-1}+2a_d}{2a_{d-1}} \cdot P^{2a_1}_{0,0} \cdots P^{2a_d}_{0,0} \\
&= \frac{1}{(2d)^{2m}} \cdot \sum_{a_1+\cdots+a_d=m} \binom{2m}{2a_1;\cdots;2a_{d-1}} \cdot \binom{2a_1}{a_1} \cdots \binom{2a_d}{a_d} \\
&= \frac{1}{(2d)^{2m}} \cdot \sum_{a_1+\cdots+a_d=m} \frac{(2m)!}{(a_1!)^2 \cdots (a_d!)^2}
\end{aligned}$$

that is choosing each operation out of $2m$ moves over d dimensions, then each dimension contains only even moves, with the probability of returning to 0 after the even moves; or equivalently, for $2m$ moves, each choose 1 out of $2 \times d$ directions, and each opposite directions should have same amount of moves. We simplify further by noticing

$$\sum_{a_1+\cdots+a_d=m} \binom{m}{a_1;\cdots;a_d} = d^m,$$

and hence

$$P^{2m}_{0^d,0^d} = \frac{1}{2^{2m}} \cdot \binom{2m}{m} \sum_{a_1+\cdots+a_d=m} \binom{m}{a_1;\cdots;a_d}^2 \cdot \frac{1}{d^{2m}} \leq \frac{1}{2^{2m}} \cdot \binom{2m}{m} \max_{a_1+\cdots+a_d=m} \binom{m}{a_1;\cdots;a_d} \cdot \frac{1}{d^m}.$$

For fixed dimension d, the multinomial coefficient is maximized when the a_i are as balanced as possible, and

$$\max_{a_1+\cdots+a_d=m} \binom{m}{a_1;\cdots;a_d} = \Theta\left(d^m m^{-(d-1)/2}\right),$$

which follows from lemma 1.7.1[4]. Hence,

$$P^{2m}_{0^d,0^d} = \Theta\left(m^{-d/2}\right).$$

[4] https://www.statslab.cam.ac.uk/~jrn10//Markov/s16.pdf

Hence 3-dimensional (and above) random walk is transient by corollary 1.7.20.

To decide the 1-dimensional and 2-dimensional random walks are null recurrent or positive recurrent, recall the computation of $P_{0,0}^{2m}$ from exercise 1.7.19, that

$$P_{0,0}^{2m} = \sum_{k\in[1,m]} r_{0,0}^{2k} \cdot P_{0,0}^{2m-2k},$$

we can (no I can't) construct as follows

$$p(x) - 1 = \sum_{m\geq 1} P_{0,0}^{2m} \cdot x^m = \sum_{m\geq 1}\sum_{k\in[1,m]} r_{0,0}^{2k} \cdot x^k \cdot P_{0,0}^{2m-2k} \cdot x^{m-k} = \left(\sum_{k\geq 1} r_{0,0}^{2k} \cdot x^k\right) \cdot \left(\sum_{t\geq 0} P_{0,0}^{2t} \cdot x^t\right) = r(x) \cdot p(x),$$

then $r(x) = 1 - 1/p(x)$. Since

$$\frac{1}{\sqrt{1-4x}} = \sum_{t\geq 0} \binom{2t}{t} x^t, \tag{1.8}$$

then $p(x) = (1-x)^{-1/2}$, and hence $r(x) = 1 - \sqrt{1-x}$, yielding

$$\begin{aligned} r_{0,0}^{2m} &= -\binom{1/2}{m} \cdot (-1)^m = (-1)^{m-1} \cdot \frac{(1/2)\cdots(1/2-(m-1))}{m!} \\ &= \frac{1}{2^m} \cdot \frac{1\cdots(2m-3)}{m!} = \frac{1}{2^m} \cdot \frac{(2m-2)!}{2^{m-1}\cdot(m-1)!\cdot m!} \\ &= \frac{1}{2^{2m}} \cdot \frac{2}{2m-1} \cdot \binom{2m-1}{m} = \frac{1}{2^{2m}} \cdot \frac{1}{2m-1} \cdot \binom{2m}{m} \end{aligned}$$

for $m \geq 1$, and

$$h_{0,0} = \sum_{m\geq 1} r_{0,0}^{2m} \cdot 2m \geq \sum_{m\geq 1} \frac{1}{2^{2m}} \cdot \binom{2m}{m} \geq \frac{1}{2\sqrt{\pi}} \cdot \sum_{m\geq 1} \frac{1}{\sqrt{m}},$$

which is unbounded, meaning the whole 1-dimensional random walk is null recurrent by lemma 1.7.43.

The 2-dimensional random walk is also null recurrent. If it is positive recurrent, then by lemma 1.7.48, the limit of $P_{0^2,0^2}^{2m}$ should be positive, yet $\left(P_{0,0}^{2m}\right)^2$ approaches 0, which is a contradiction, hence it is null recurrent. □

After exercise 1.7.64 and by lemma 1.7.1, we derive a finer bound than corollary 1.6.3.

Lemma 1.7.65.

$$\frac{2}{\sqrt{\pi n}} \cdot 2^{2n} \geq \binom{2n}{n} \geq \frac{1}{2\sqrt{\pi n}} \cdot 2^{2n}.$$

Exercise 1.7.66 (Exercise 7.17 [MU17])**.** Consider the Markov chain over $\mathbb{N}$: For $i \geq 1$, the particle move to $i+1$ with probability p, and to $i-1$ with probability $1-p$ for $i \geq 1$, and from 0 to 1 with probability 1.

Show that the chain is transient when $p > 1/2$, positive recurrent when $p < 1/2$, and null recurrent when $p = 1/2$.

Proof. The particle returns to 0 only on even number of steps. Hence, we want to count all possible $2m$ steps random walk over $\mathbb{Z}$, such that the particle never go to negative side, to derive $r_{0,0}^{2m}$ and $P_{0,0}^{2m}$.

There are 2 relevant combinatorics tools: Catalan number and Dyck path. A Dyck path is a series of ups and downs. The path will begin and end on the same level; and as the path moves from left to right it will rise and fall, never dipping below the height it began on. Hence, the Dyck path models the desired random walk well.

Catalan number C_n counts unique Dyck paths of n ups and n downs. Noticing that for n ups and n downs, C_{n-1} is the number of unique Dyck paths that first returns after $2n$ steps. Hence, we can count C_n by

$$C_n = \sum_{k\in[1,n]} C_{k-1}C_{n-k},$$

that leads by first returns from 2 to $2n$ steps, and $C_0 = 1$. We derive C_i with generating function method in the same way as exercise 1.7.64, by constructing

$$\frac{1}{x}\left(f(x)-1\right) = \sum_{n\geq 1} C_n \cdot x^{n-1} = \sum_{n\geq 1}\sum_{k\in[1,n]} C_{k-1}\cdot x^{k-1}\cdot C_{n-k}\cdot x^{n-k} = \left(\sum_{t\geq 0} C_t \cdot x^t\right)^2 = f(x)^2,$$

Then $f(x) = \left(1-\sqrt{1-4x}\right)\cdot(2x)^{-1}$, as $f(x)$ approaches 1 when $x = 0$. By previous eq. (1.8) in exercise 1.7.64,

$$1-\sqrt{1-4x} = -\sum_{n\geq 1}\binom{1/2}{n}\cdot(-4x)^n = \sum_{n\geq 1}\frac{1}{2n-1}\binom{2n}{n}x^n = \sum_{n\geq 0}\frac{2}{n+1}\binom{2n}{n}x^{n+1} = 2x\sum_{n\geq 0}\frac{1}{n+1}\binom{2n}{n}x^n = 2x\sum_{n\geq 0} C_n x^n.$$

Hence, $r_{0,0}^{2n} = C_{n-1}\cdot p^{n-1}\cdot(1-p)^n$, and we march on to derive $h_{0,0}$:

$$\begin{aligned} h_{0,0} &= \sum_{n\geq 1} C_{n-1}\cdot(1-p)^n\cdot p^{n-1}\cdot 2n \\ &= \sum_{n\geq 0} C_n\cdot(1-p)^{n+1}\cdot p^n\cdot(2n+2) \\ &= \sum_{n\geq 0}\frac{1}{n+1}\binom{2n}{n}\cdot(1-p)^{n+1}\cdot p^n\cdot(2n+2) \\ &= 2(1-p)\sum_{n\geq 0}\binom{2n}{n}(1-p)^n p^n. \end{aligned}$$

Again by eq. (1.8),

$$h_{0,0} = \frac{2-2p}{\sqrt{1-4p(1-p)}} = \frac{2-2p}{|1-2p|}.$$

- When $p < 1/2$, $h_{0,0}$ is bounded, and the chain is positive recurrent.
- When $p = 1/2$, $h_{0,0}$ is infinity, and the chain is null recurrent.
- When $p > 1/2$, $h_{0,0}$ is not valid, and the chain is transient.

Another way of showing this is to use the cut-set way by theorem 1.7.56. [5] Assuming $\boldsymbol{\pi}$ exists, then we can derive

$$\begin{aligned} \pi_0 &= (1-p)\pi_1, \\ \pi_i &= \pi_{i-1} + (1-p)\pi_{i+1}, \end{aligned}$$

and we have $(1-p)\pi_{i+1} = \pi_i$.

- When $p < 1/2$, $\pi_0 = (1-2p)/(2-2p)$, hence the chain is positive recurrent.
- When $p = 1/2$, $\sum_i \pi_i = 1$ but π_i are 0, by lemma 1.7.48 the chain is null recurrent.
- When $p > 1/2$, $\sum_i \pi_i$ diverges for any positive π_0, hence no stationary distribution exists.

□

Lemma 1.7.67. *A Dyck path is a series of ups and downs. The path will begin and end on the same level; and the path never dips below the height it began on. Catalan number C_n counts Dyck paths of n ups and n downs by*

$$C_n = \frac{1}{n+1}\binom{2n}{n}$$

with the property $C_0 = 1$ and

$$C_n = \sum_{k\in[1,n]} C_{k-1}C_{n-k}.$$

[5] For further read on other methods, https://chihaozhang.com/teaching/SP2025/lec4.pdf

Exercise 1.7.68 (Exercise 7.19 [MU17])**.** Consider example 1.7.51. Show that the expected number of games is $\ell_1\ell_2$.

Proof. Let X_j be the random variable for the number of steps to be absorbed for a particle to start at state j. Then

$$\begin{aligned}\mathbb{E}\left[X_j\right] &= \mathbb{E}\left[X_j \mid \text{ move to } j+1\right]\cdot r^1_{j,j+1} + \mathbb{E}\left[X_j \mid \text{ move to } j-1\right]\cdot r^1_{j,j-1}\\ &= \frac{1}{2}\mathbb{E}\left[X_{j+1}\right] + \frac{1}{2}\mathbb{E}\left[X_{j-1}\right] + 1\end{aligned}$$

for $j \in [-\ell_1+1, \ell_2-1]$, and $\mathbb{E}[X_{-\ell_1}] = \mathbb{E}[E_{\ell_2}] = 0$. Now we have the linear system for $k \in [-\ell_1+1, \ell_2-1]$

$$2\,\mathbb{E}[E_k] = \mathbb{E}[X_{k-1}] + \mathbb{E}[X_{k+1}] + 2$$

and solve by

$$\begin{aligned}\mathbb{E}[X_{-\ell_1+1}] &= \frac{1}{2}\mathbb{E}[X_{-\ell_1+2}] + 1\\ &\cdots\\ \mathbb{E}[X_0] &= \frac{\ell_1}{\ell_1+1}\mathbb{E}[X_1] + \ell_1\\ &\cdots\\ \mathbb{E}[X_{\ell_2-1}] &= \frac{\ell_1+\ell_2-1}{\ell_1+\ell_2}\mathbb{E}[X_{\ell_2}] + (\ell_1+\ell_2-1)\end{aligned}$$

Hence, for $i \in [-\ell_2, \ell_1]$, we have $\mathbb{E}[X_i] = (\ell_1+i)\cdot(\ell_2-i)$ to complete the proof. □

Exercise 1.7.69 (Exercise 7.20 [MU17])**.** Consider the gambler's ruin where games are unfair with losing probability 2/3. Suppose one starts with i dollars by either reaching n or 0. Let W_t be the amount one gained after t games.

- Show that $\mathbb{E}\left[2^{W_{t+1}}\right] = \mathbb{E}\left[2^{W_t}\right]$.
- Determine the probability of finishing with 0 and the probability of finishing with n when starting at position i.
- Generalize when losing probability is $p > 1/2$.

Proof. We write p^j_t to denote the probability of being at position j at t^{th} step. Moreover, $r^1_{k,k-1} = 2/3$ and $r^1_{k,k+1} = 1/3$ for each $k \in [1, n-1]$, we have the following invariant

$$2^k = \frac{2}{3}\cdot 2^{k-1} + \frac{1}{3}\cdot 2^{k+1} = r^1_{k,k-1}\cdot 2^{k-1} + r^1_{k,k+1}\cdot 2^{k+1},$$

and hence $\mathbb{E}\left[2^{W_t}\right] = \mathbb{E}\left[2^{W_{t+1}}\right]$ by

$$\mathbb{E}\left[2^{W_{t+1}}\right] = 2^0\cdot p^0_t + \sum_{k\in[1,n-1]}\left(r^1_{k,k-1}\cdot 2^{k-1} + r^1_{k,k+1}\cdot 2^{k+1}\right)\cdot p^k_t + 2^n\cdot p^n_t = \sum_{k\in[0,n]} 2^k\cdot p^k_t = \mathbb{E}\left[2^{W_t}\right].$$

Since $\mathbb{E}\left[2^{W_0}\right] = \mathbb{E}\left[2^{W_t}\right]$ for arbitrary $t > 0$, then $\mathbb{E}\left[2^{W_0}\right] = 2^i = \mathbb{E}\left[2^{W_t}\right]$.

Let q_i be the probability of absorbing to n from state i, then

$$\lim_{t\to\infty}\mathbb{E}\left[2^{W_t}\right] = q_i\cdot 2^n + (1-q_i)\cdot 2^0 = 2^i,$$

hence $q_i = \left(1-2^i\right)/(1-2^n)$.

In general, for losing probability p, we want to find similar invariant $\mathbb{E}\left[q^{W_t}\right] = \mathbb{E}\left[q^{W_{t+1}}\right]$ such that

$$p\cdot q^{t-1} + (1-p)\cdot q^{t+1} = q^t,$$

and solve $(1-p)x^2 - x + p = 0$ for $x = p/(1-p)$, which yields

$$q_k = \frac{1-x^k}{1-x^n}.$$

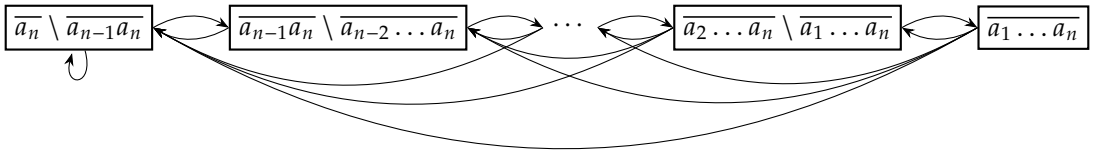
When $p = 1/2$, $q_k = k/n$ as in example 1.7.51. □

Exercise 1.7.70 (Exercise 7.27 [MU17]). Suppose that we are given n records, $R_1, R_2, \ldots, R_n$ kept in some order. The cost of accessing the j^{th} record in the order is j. Thus, if we had four records ordered as R_2, R_4, R_3, R_1, then the cost of accessing R_4 would be 2 and the cost of accessing R_1 would be 4.

Suppose further that, at each step, record R_j is accessed with probability p_j, with each step being independent of other steps. If we knew the values of the p_j in advance, we would keep the R_j in decreasing order with respect to p_j. But if we don't know the p_j in advance, we might use the "move to front" heuristic: at each step, put the record that was accessed at the front of the list. We assume that moving the record can be done with no cost and that all other records remain in the same order. For example, if the order was R_2, R_4, R_3, R_1 before R_3 was accessed, then the order at the next step would be R_3, R_2, R_4, R_1.

The order of the records can be thought of as the state of a Markov chain. Give the stationary distribution of this chain. Also, let X_k be the cost for accessing the k^{th} requested record. Determine an expression for $\lim_{k\to\infty} \mathbb{E}[X_k]$.

Proof. By the building process, we observe a pattern: To build towards a target $\overline{a_1 \ldots a_n}$, we build from $\overline{a_n}$ to $\overline{a_{n-1}a_n}$, then to $\overline{a_k \ldots a_n}$, finally $\overline{a_1 \ldots a_n}$. We illustrate here by enforcing each state to contain substring $\overline{a_k \ldots a_n}$ at most, and

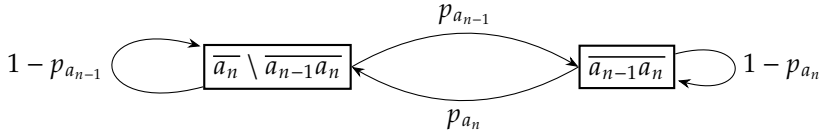


notice the absorbing relation

$$\overline{a_k \ldots a_n} = \overline{a_1 \ldots a_n} \cup \bigcup_{i\in[2,k]} \left(\overline{a_i \ldots a_n} \setminus \overline{a_{i-1} \ldots a_n}\right),$$

then we can use cutset to inductively derive the stationary distribution for $\overline{a_1 \ldots a_n}$. For simplicity we say $\pi_k = \pi_{\overline{a_k \ldots a_n}}$.

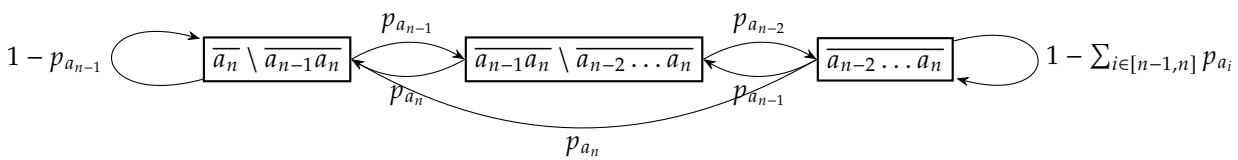
The simplest case is as follows,



and hence the stationary distribution of $\overline{a_{n-1}a_n}$ is

$$\pi_{n-1} = \frac{p_{a_{n-1}}}{p_{a_{n-1}} + p_{a_n}} \cdot \pi_n = \frac{p_{a_{n-1}}}{p_{a_{n-1}} + p_{a_n}}.$$

Relaxing $\overline{a_{n-1}a_n}$ to $\overline{a_{n-1}a_n} \setminus \overline{a_{n-2} \ldots a_n}$ and $\overline{a_{n-2} \ldots a_n}$, we have



and hence the stationary distribution for $\overline{a_{n-2} \ldots a_n}$ is

$$\pi_{n-2} = \frac{p_{a_{n-2}}}{p_{a_{n-2}} + \cdots + p_{a_n}} \cdot \pi_{n-1}.$$

By cutset, for $\overline{a_k \dots a_n}$, the inflowing probability is its stationary distribution, then

$$p_{a_k} \cdot (\pi_{k+1} - \pi_k) + \left(1 - \sum_{i \in [k+1,n]} p_{a_i}\right) \cdot \pi_k = \pi_k,$$

and therefore

$$\pi_k = \frac{p_{a_k}}{p_{a_k} + \cdots + p_{a_n}} \cdot \pi_{k+1}.$$

Hence, by induction, the resulting stationary distribution is

$$\pi_1 = p_{a_1} \cdot \frac{p_{a_2}}{1 - p_{a_1}} \cdots \frac{p_{a_{n-1}}}{1 - \sum_{i \in [1,n-2]} p_i} \cdot \frac{p_{a_n}}{1 - \sum_{i \in [1,n-1]} p_i} = \prod_{j \in [1,n]} p_j \cdot \prod_{k \in [1,n]} \left(\sum_{i \in [k,n]} p_{a_i}\right)^{-1}.$$

To derive the $\mathbb{E}[X_k]$ as $k \to \infty$, we eventually reach the stationary distribution after infinity steps, and hence we want to derive the expected distance to the first place in the list, for each record. By prior cutset method and state building, the probability of having i^{th} record, and j^{th} record in any position after i^{th} record, is

$$\pi_{\overline{ij}} = \frac{p_i}{p_i + p_j}.$$

Let Y_i be random variable for the order of i^{th} record in the queue, then by linearity of expectation,

$$\mathbb{E}[Y_i] = 1 + \sum_{j \neq i} \pi_{\overline{ji}} = 1 + \sum_{j \neq i} \frac{p_j}{p_i + p_j}.$$

Hence, by conditional expectation, we have

$$\lim_{k \to \infty} \mathbb{E}[X_k] = \lim_{k \to \infty} \sum_{i \in [1,n]} p_i \cdot \mathbb{E}\left[X_k \mid k^{\text{th}} \text{ query is on } i^{\text{th}} \text{ record}\right] = \sum_{i \in [1,n]} p_i \cdot \mathbb{E}[Y_i] = 1 + \sum_{i \in [1,n]} p_i \sum_{j \neq i} \frac{p_j}{p_i + p_j}.$$

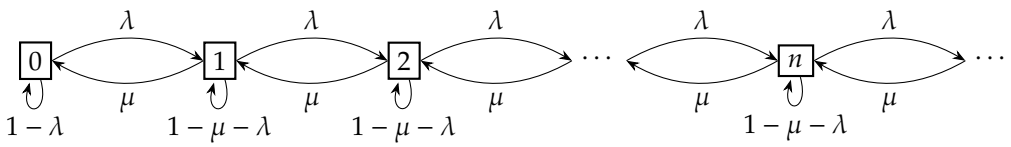
□

Remark 1.7.71. Buddy, why torture me with Tsetlin Library exercise 1.7.70 [Tse63].

Exercise 1.7.72 (Exercise 7.28 [MU17])**.** Consider the following variation of the discrete time queue. Time is divided into fixed-length steps. At the beginning of each time step, a customer arrives with probability λ. At the end of each time step, if the queue is nonempty then the customer at the front of the line completes service with probability μ.

- Explain under what conditions a stationary distribution exists, and find the stationary distribution when it exists.
- Now consider the variation where we change the order of incoming arrivals and service. That is: at the beginning of each time step, if the queue is nonempty then a customer is served with probability μ; and at the end of a time step a customer arrives with probability λ. How does this change the stationary distribution?

Proof. We show the vanilla Markov chain of the infinite discrete queue that each step only 1 of the 2 things happens.

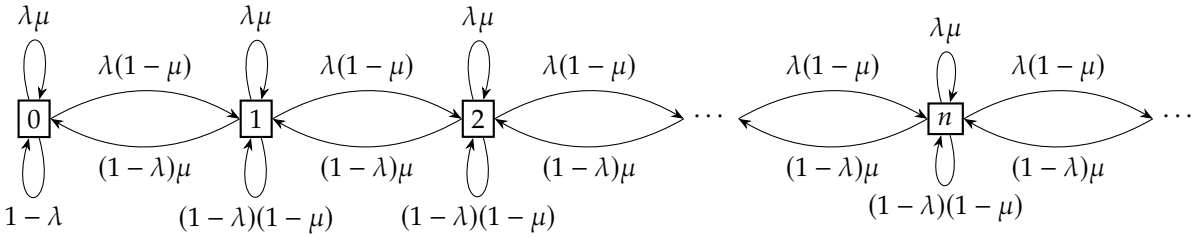


By cutset $\lambda \pi_i = \mu \pi_{i+1}$, $\boldsymbol{\pi}$ exists only on $\lambda < \mu$ when the chain is positively recurrent,

$$\pi_i = \left(1 - \frac{\lambda}{\mu}\right) \cdot \left(\frac{\lambda}{\mu}\right)^i,$$

otherwise when $\lambda = \mu$, the chain is null recurrent, $\lambda > \mu$ the chain is transient.

When the chain introduces a new customer at the beginning of a time step, and processes a customer at the end of the time step, the Markov chain follows, where remaining unchanged can be either no customer coming and no customer processed, or 1 customer coming and 1 customer processed. Otherwise, transitioning to the next state means 1 customer coming, and transitioning to the previous state means 1 customer being processed.

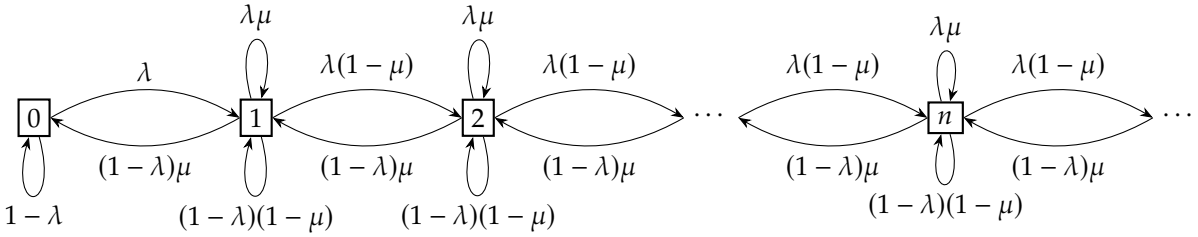


By cutset $\lambda(1-\mu)\pi_i = (1-\lambda)\mu\pi_{i+1}$, $\boldsymbol{\pi}$ exists only on $\lambda < \mu$ when the chain is positively recurrent,

$$\pi_i = \left(1 - \frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right) \cdot \left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i,$$

and the chain being null recurrent on $\lambda = \mu$, being transient on $\lambda > \mu$.

When we reverse the order, say on new time step with probability μ the nonempty queue a customer is processed, and a customer will arrive with probability λ at the end of the time step, the new Markov chain follows.



The only difference lies in the empty queue, that follows the vanilla discrete time queue behavior, as the empty queue does not have customers to consume, and either with probability λ to accept new customer at the end of the time step, or remain empty with probability $1 - \lambda$.

By cutset, we have

$$\lambda\pi_0 = (1-\lambda)\mu\pi_1$$
$$\lambda(1-\mu)\pi_i = (1-\lambda)\mu\pi_{i+1},$$

hence for $i \geq 1$,

$$\pi_i = \frac{\pi_0}{1-\mu} \cdot \left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i,$$

and by $\sum_{i\geq 0} \pi_i = 1$, we have

$$\pi_0 = 1 - \frac{\lambda}{\mu}.$$

Therefore, for $i \geq 1$,

$$\pi_i = \frac{\pi_0}{1-\mu} \cdot \left(\frac{\lambda(1-\mu)}{\mu(1-\lambda)}\right)^i,$$

with the existence condition for $\boldsymbol{\pi}$ being $\lambda < \mu$, in which case the chain is positively recurrent. □

1.7.5 Random Walks on Undirected Graphs

Definition 1.7.73. A random walk on G is a Markov chain defined by the sequence of moves of a particle between vertices of G. The position of the particle at a particular time step is the state of the chain. If the particle is at vertex i with $d(i)$ outgoing edges, then the probability that the particle follows edge (i, j) to neighboring vertex j is $d(i)^{-1}$.

Lemma 1.7.74. *A random walk on an undirected graph G is aperiodic iff G is not bipartite, which means there is no 2-coloring such that the endpoints of each edge are colored differently.*

Proof. A graph G being not bipartite means it has cycles of odd length. If a graph G is bipartite and undirected, then the length of the path going back to itself is always even, such that the Markov chain is of period 2. Otherwise, undirected G is not bipartite, each vertex has an odd-length path back to itself, then the Markov chain is aperiodic. □

Similar to theorem 1.7.53, A random walk on a finite, undirected, connected, and non-bipartite graph G defines a Markov chain that converges to a stationary distribution that is determined by the degree sequences of the graph.

Theorem 1.7.75. *A random walk on a finite, undirected, connected, and non-bipartite G has a stationary distribution $\boldsymbol{\pi}$ where*

$$\pi_v = \frac{d(v)}{2|E|}.$$

Proof. Since $\sum_{v\in V} d(v) = 2|E|$, then $\boldsymbol{\pi}$ defines a proper distribution of particle appearing over V.

On the other hand, by definition of stationary distribution in theorem 1.7.56, for $u \in V$ and its neighbors $\Gamma(u)$,

$$\sum_{v\in\Gamma(u)} \frac{d(v)}{2|E|}\cdot\frac{1}{d(v)} = \frac{d(u)}{2|E|} = \frac{d(u)}{2|E|}\cdot\sum_{v\in\Gamma(u)}\frac{1}{d(u)},$$

where the second equality holds from $|\Gamma(u)| = d(u)$. □

Exercise 1.7.76 (Exercise 7.14 [MU17])**.** Show that the Markov chain defined by a random walk on a finite, undirected, non-bipartite, connected graph G is time reversible.

Proof. When G is finite, then theorem 1.7.75 kicks in, by definition of time reversibility in exercise 1.7.63,

$$\frac{d(v)}{2|E|}\cdot\frac{1}{d(v)} = \frac{1}{2|E|} = \frac{d(u)}{2|E|}\cdot\frac{1}{d(u)}.$$

□

Lemma 1.7.77 (Commute Time)**.** *For $G = (V, E)$, if $(u, v) \in E$, the commute time $h_{u,v} + h_{v,u} \le 2|E|$.*

Proof. Let D be the set of directed edges: for every $(u, v) \in E$, we have 2 directed edges $u \to v$ and $v \to u$ in D.

We can view the random walk over G as a Markov chain of state space D, where the state of the Markov chain at the t^{th} time step is the direct edge taken by the random walk in its t^{th} transition. The Markov chain has $2|E|$ states, and the stationary distribution $\pi_{u\to v} = (2|E|)^{-1}$ for all $(u, v) \in E$, by cutset theorem 1.7.56

$$\pi_{u\to v} = \frac{1}{d(u)}\sum_{w\in\Gamma(u)}\pi_{w\to u}$$

for all $(u, v) \in E$ and $\Gamma(u)$ is the set of neighboring vertices of u, $\pi_{u\to v} = (2|E|)^{-1}$ is a solution. By theorem 1.7.53,

$$\pi_{u\to v} = h^{-1}_{u\to v,u\to v} = (2|E|)^{-1},$$

hence $h_{u\to v,u\to v} = 2|E|$: after taking the directed path $u \to v$, the expected number of steps to retake $u \to v$ is $2|E|$. By Markovian (memoryless) of the random walk, once it reaches v, it forgets about it reaches by $u \to v$, hence the expected time to start from v and reach u and take (u, v) to v is $2|E|$.

Since $h_{u,v}$ will not rule out the walk reaching v without taking (u, v), then it will not force the walk to go back to u and take (u, v), even if it reached v before, hence $h_{v,u} + h_{u,v} \le h_{u\to v,u\to v} = 2|E|$. □

Exercise 1.7.78 (Exercise 7.22 [MU17]). A cat and a mouse each independently take a random walk on a connected, undirected, non-bipartite graph G. They start at the same time on different nodes, and each makes 1 transition at each time step. The cat eats the mouse if they are ever at the same node at a time step. Let $|V| = n$ and $|E| = m$. Show an upper bound of $O(m^2 n)$ on the expected time before the cat eats the mouse.

Proof. Construct a graph G^2 from G by making vertices $V^2 = (V, V)$, and edges E^2 for all $((u_0, u_1), (v_0, v_1))$ such that $(u_0, v_0) \in E$ and $(u_1, v_1) \in E$. Counting the degrees of each vertices $(a, b) \in V^2$ double counts $\left|E^2\right|$, then

$$2\left|E^2\right|^2 = \sum_{(a,b)\in V^2} d(a,b) = \sum_{a\in V} d(a) \sum_{b\in V} d(b) = 4\left|E\right|^2,$$

such that $\left|E^2\right| = 2\left|E\right|^2$.

We show that, for every $(a, b) \in V^2$, there exists paths of $O(|V|)$ length to a $(v, v) \in V^2$ in the connected, undirected, non-bipartite graph. First, for every $u, v \in V$, the path in between over G is $O(|V|)$ length. Then, for paths in between a, v and b, v, we use ℓ_1, ℓ_2 to denote the paths lengths,

- If both ℓ_1, ℓ_2 are even or odd, then there exists path in between $(a, b), (v, v)$ of length $\max(\ell_1, \ell_2) = O(|V|)$, as the shorter one can keep looping in between v and $u \in \Gamma(v)$ for $2k$ steps, such that they eventually meet at (v, v).
- Otherwise, since G is non-bipartite, there exists odd length cycle in G, such that the odd length path can be transformed into an even length path, and the 2 lengths can be calibrated to be the same as previously described. The cycle is of $O(|V|)$ length, and original paths are $O(|V|)$ lengths, hence the resulting length is $O(|V|)$.

After obtaining such path $\mathbf{s}$ from (a, b) to (v, v) of length $\ell = O(|V|)$, we bound $h_{(a,b),(v,v)}$ by

$$h_{(a,b),(v,v)} \le h_{(a,b),(v,v)} + h_{(v,v),(a,b)} \le \sum_{i\in[1,\ell-1]} \left(h_{(s_i,s_{i+1})} + h_{(s_{i+1},s_i)}\right) \le \ell \cdot 2\left|E^2\right| = 4\ell\left|E\right|^2.$$

where the second inequality holds by enforcing the path $\mathbf{s}$ from (a, b) to (v, v), similar to the idea of lemma 1.7.77, and the third inequality is immediate by lemma 1.7.77. Hence, the expected time $h_{(a,b),(v,v)} = O(|E|^2 |V|)$. □

Definition 1.7.79 (Covering Time). The covering time of $G = (V, E)$ is the max of all $v \in V$ of the expected time to visit all vertices by a random walk starting from v.

Lemma 1.7.80 (Covering Time Upper Bound). *The covering time C_G of $G = (V, E)$ is upper bounded by $2|E|(|V| - 1)$.*

Proof. One can consider traversing all vertices with a MST in G, such that there are $|V| - 1$ directed paths in the MST. Then we bound similarly in lemma 1.7.77 and exercise 1.7.78 by enforcing the paths

$$C_G \le \sum_{i\in[1,|V|-1]} \left(h_{v_i,v_{i+1}} + h_{v_{i+1},v_i}\right) \le 2|E|(|V| - 1).$$

□

We introduce Matthew's theorem, connecting the hitting time $h_{u,v}$ to the covering time. We write $H(n) = \sum_{i\le n} 1/i$.

Lemma 1.7.81 (Matthews' Theorem [Mat88]). *The covering time C_G of $G = (V, E)$ with $|V| = n$ is bounded by*

$$H(n-1) \min_{u,v\in V, u\neq v} h_{u,v} \le C_G \le H(n-1) \max_{u,v\in V, u\neq v} h_{u,v}.$$

Proof. The proof utilizes balls-and-bins model, and the path enforcing technique of lemma 1.7.77 and exercise 1.7.78.

We begin by enforcing a traversal order $\{Z_i\}_{i\in[1,n]}$ over V sampled from uniform permutation over $[1, n]$, write B as the maximal hitting time $h_{u,v}$ for convenience, T_i be the first time step when all the first i vertices $Z_1, \ldots, Z_i$ have been visited in the order [6], and X_i be the state of the Markov chain at the i^{th} time step.

[6] If Z_j are visited for $j \in [1, k] \setminus \{i\}$, then when Z_i is visited, $T_i = T_{i+1} = \cdots = T_k$.

Given all the prior setup, we are interested in the time intervals $\{T_i - T_{i-1}\}_{i\in[2,n]}$. Given the random walk history,

$$Y_i = \mathbb{E}\left[T_i - T_{i-1} \mid Z_1, \ldots, Z_i; X_1, \ldots, X_{T_{i-1}}\right]$$

for $i \in [2, n]$ [7], and the expected time to cover the graph starting from u is

$$C_G = \sum_{i\in[2,n]} Y_i + \mathbb{E}[T_1].$$

For T_1, one has $1/n$ probability having $Z_1 = u$, otherwise u is reached by random walk. By conditional expectation,

$$\begin{aligned}\mathbb{E}[T_1] &= \mathbb{E}[T_1 \mid Z_1 = u]\Pr[Z_1 = u] \\ &\quad + \mathbb{E}[T_1 \mid Z_1 \neq u]\Pr[Z_1 \neq u] \\ &\leq \left(1 - \frac{1}{n}\right) B,\end{aligned}$$

where the second inequality holds by B upper bounding the expectation of hitting time, and $T_1 = 0$ when $Z_1 = u$.

For Y_i, if i has already been visited previously, then $Y_i = 0$ as $T_i = T_{i-1}$, otherwise $Y_i \leq B$ regardless.

Since the permutation $\{Z_i\}_{i\in[1,n]}$ is sampled uniformly random and independent of the random walk, then any one of $\{Z_k\}_{i\in[1,k]}$ can be the last visited for first k entries with probability $1/k$ [8], and hence

$$\begin{aligned}\sum_{i\in[2,n]} Y_i + \mathbb{E}[T_i] &\leq \sum_{i\in[2,n]} \frac{1}{i} B + \left(1 - \frac{1}{n}\right) B \\ &= \sum_{i\in[1,n-1]} \frac{1}{i} B \\ &= H(n-1)B.\end{aligned}$$

The lower bound is proven in the same way, by setting B to be the minimal of the expected hitting time. □

Exercise 1.7.82 (Exercise 7.24 [MU17]). A lollipop graph on n vertices is a clique $K_{n/2}$ connected to a $n/2$ long path. u is the node being both clique and path, and v be the other end of the path. Show the C_G starting from v is $\Theta(n^2)$, and starting from u is $\Theta(n^3)$.

Proof. We bound the covering time from v by enforcing the path from v to u, then to other $x \in K_{n/2}$. By lemma 1.7.80,

$$C_G \leq h_{v,u} + \sum_{x\in K_{n/2}, x\neq u} (h_{x,u} + h_{u,x}).$$

For $h_{v,u}$, let $i \in [1, n/2 - 1]$ be vertices on the (u, v) path, we have

$$\begin{aligned}h_{1,u} &= \frac{1}{2}\cdot 1 + \frac{1}{2}\cdot(1 + h_{2,u}) \\ &\cdots \\ h_{n/2-1,u} &= \frac{1}{2}\cdot\left(1 + h_{n/2-2,u}\right) + \frac{1}{2}\cdot(1 + h_{v,u}) \\ h_{v,u} &= 1 + h_{n/2-1,u},\end{aligned}$$

hence $h_{i,u} = i\cdot(n-i)$, and $h_{v,u} = n^2/4$.

[7] Coupon-collector on the first k entries, that does not strictly enforces traverse by the permutation, but enforces by visiting first k entries.

[8] Z_k being last visited w.r.t. the fixed random walk is only related to $\{Z_i\}_{i\in[1,k]}$, by symmetry of random permutation, the probability is $1/k$.

For $h_{u,x}$ for some $x \in K_{n/2}$, let $k = n/2$ and $x' \neq x$ that $x' \in K_{n/2}$, then

$$\begin{aligned}
h_{u,x} &= \frac{1}{k} \cdot 1 + \frac{k-2}{k} \cdot (1 + h_{x',x}) + \frac{1}{k} \cdot (1 + h_{1,x}) \\
h_{x',x} &= \frac{1}{k-1} \cdot 1 + \frac{k-3}{k-1} \cdot (1 + h_{x',x}) + \frac{1}{k-1} \cdot (1 + h_{u,x}) \\
h_{1,x} &= \frac{1}{2} \cdot (1 + h_{u,x}) + \frac{1}{2} \cdot (1 + h_{2,x}) \\
&\cdots \\
h_{v,x} &= 1 + h_{n/2-1,x},
\end{aligned}$$

then $h_{u,x} = (2k^2 + 3)/2k = n + 3/n$.

For the rest of hitting time from $x \neq u$, $x \in K_{n/2}$, there is a vanilla way of computing by

$$\begin{aligned}
h_{x,u} &= \frac{1}{k-1} \cdot 1 + \frac{k-2}{k-1} \cdot (1 + h_{x',u}) \\
h_{x',u} &= \frac{1}{k-1} \cdot 1 + \frac{k-2}{k-1} \cdot (1 + h_{x',u}),
\end{aligned}$$

then $h_{x,u} = h_{x',u} = k - 1$.

Another way of deriving $h_{x,u}$ is by taking advantage of theorem 1.7.75, that the lollipop graph is non-bipartite as $K_{n/2}$ has odd length cycle, then

$$\pi_{u,u}^{-1} = \frac{2|E|}{d(u)} = h_{u,u} = \frac{1}{d(u)} \sum_{w \in \Gamma(u)} (h_{w,u} + 1) = 1 + \frac{1}{k} \sum_{w \in \Gamma(u)} h_{w,u},$$

where $d(u) = |\Gamma(u)| = k$, hence $\sum_{w \in \Gamma(u)} h_{w,u} = 2|E| - k$. Since $|E| = k(k+1)/2$, and $h_{1,u} = n - 1 = 2k - 1$, hence

$$\sum_{w \in \Gamma(u), w \neq 1} h_{w,u} = 2|E| - k - h_{1,u} = (k-1)^2.$$

Therefore, the covering time from v is $\Theta(n^2)$.

To derive the C_G from u, $h_{u,v}$ is upper bounding the covering time for the path part, then

$$C_G \leq h_{u,v} + \sum_{x \in K_{n/2}, x \neq u} (h_{x,u} + h_{u,x}),$$

and we have the formulas

$$\begin{aligned}
h_{x,v} &= \frac{1}{k-1} \cdot (1 + h_{u,x}) + \frac{k-2}{k-1} \cdot (1 + h_{x,v}) \\
h_{u,v} &= \frac{1}{k} \cdot (1 + h_{1,v}) + \frac{k-1}{k} \cdot (1 + h_{x,v}) \\
h_{1,v} &= \frac{1}{2} \cdot (1 + h_{u,v}) + \frac{1}{2} \cdot (1 + h_{2,v}) \\
&\cdots \\
h_{n/2-1,v} &= \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot \left(1 + h_{n/2-2,v}\right),
\end{aligned}$$

then $h_{k-i,v} = i \cdot (k^2 + k - i)$, and hence $h_{u,v} = k^3$, which explains C_G from u is $\Theta(n^3)$. □

Exercise 1.7.83 (Exercise 7.26 [MU17]). Let n equidistant points be marked on a circle. WLOG, think of the points as being labeled clockwise from 0 to $n - 1$. Initially, a wolf begins at 0 and there is one sheep at each of the remaining $n - 1$ points. The wolf takes a random walk on the circle. For each step, it moves with probability $1/2$ to one neighboring point and with probability $1/2$ to the other neighboring point. At the first visit to a point, the wolf eats a sheep if there is still one there. Which sheep is most likely to be the last eaten?

Proof. Brain teaser. Say for any sheep $j \in [1, n-1]$ is eaten last, then view this as a Gambler's Ruin example 1.7.51 by states $j, j+1 \cdots n-1, 0, 1, \cdots, j-1, j$ and absorb at both j's. Now the requirement is, visit all states before absorbing.

Visiting $j+1$ first is mutually exclusive with visiting $j-1$ first, and hence the rest is visiting $j+1$ or $j-1$, then absorb to the j on the other end, which is of probability $1/(n-1)$ in total.

Such $1/(n-1)$ probability holds for any $j \in [1, n-1]$, and hence all sheeps eaten last is equal probability. □

Exercise 1.7.84 (Exercise 7.30 [MU17])**.** Let (u, v) be an edge in a hypercube of $N = 2^n$ nodes.

- Prove the expected time between the traversal of (u, v) edge is nN.
- We consider the time between transitions from u to v in a different way. After moving from u to v, the walk must first return to u. When it returns to u, the walk might next move to v, or it might move to another neighbor of u, in which case it must return to u again before moving to v for there to be a transition from u to v.

  Use symmetry and the above description to prove the following recurrence:

$$Nn = \sum_{i \geq 1} \frac{1}{n} \left(\frac{n-1}{n} \right)^{i-1} (i(h_{u,v} + 1)) = n(h_{u,v} + 1).$$

- Conclude that $h_{u,v} = N - 1$.
- Using the result on the hitting time of adjacent vertices and Matthews' theorem, show the cover time is $O(Nn^2)$.
- As a much more challenging problem, try to prove that the maximum hitting time between any two vertices for the random walk on the hypercube is $O(N)$, and that the cover time is correspondingly $O(Nn)$.

Proof. Each vertex in an n-dimensional hypercube has n edges, then $2|E| = nN$. By lemma 1.7.77,

$$h_{u \to v, u \to v} = 2|E| = nN.$$

To derive $h_{u \to v, u \to v}$ again from geometric distribution, consider u going to v, or $n-1$ other neighbors in $\Gamma(u) \setminus \{v\}$. It takes one step to go from u to any $w \in \Gamma(u)$. To get back to u from $w \in \Gamma(u)$, $h_{w,u} = h_{v,u} = h_{u,v}$ by symmetry.

Hence, a trial contains going back from neighbor $w \in \Gamma(u)$, then depart from u to one of the neighbors in $\Gamma(u)$. By Markovian of random walk, the probability of going from u to v is n^{-1}, hence the expected steps to take (u, v) is

$$\sum_{i \geq 1} \frac{1}{n} \left(\frac{n-1}{n} \right)^{i-1} \cdot i(h_{u,v} + 1) = n(h_{u,v} + 1).$$

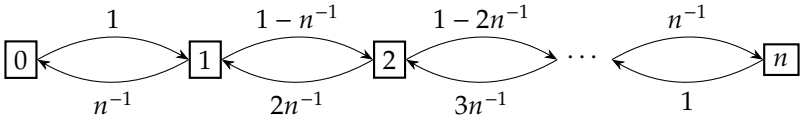
By prior result on $h_{u \to v, u \to v}$ by lemma 1.7.77, we conclude that $h_{u,v} = N - 1$.

Consider any $p, q \in V$ for the hypercube, that $(p, q) \notin E$. Similar to exercise 1.7.78, we upper bound by

$$h_{p,q} \leq \sum_{(a,b) \in P} h_{a,b} = O(Nn),$$

where there is a path of length $O(n)$ connecting p and q, as there are at most $O(n)$ bits different in between p and q. By lemma 1.7.81, $H(N-1) = O(n)$, hence the covering time is upper bounded by $O(n^2 N)$.

For departing from u, we can model the hitting time to v by the Hamming distance as follows.



Then the linear system for the hitting time is

$$\begin{aligned} h_{0,n} &= 1 + h_{1,n} \\ h_{i,n} &= 1 + \frac{i}{n} \cdot h_{i-1,n} + \frac{n-i}{n} \cdot h_{i+1,n} \\ h_{n-1,n} &= 1 + \frac{n-1}{n} \cdot h_{n-2,n}, \end{aligned}$$

and let $d_i = h_{i,n} - h_{i+1,n}$ with $d_{n-1} = h_{n-1,n}$, then

$$\begin{aligned} d_0 &= 1 \\ (n-i)d_i &= n + id_{i-1}. \end{aligned}$$

Since

$$\begin{aligned} (n-i)\binom{n-1}{i-1} &= i\binom{n-1}{i} \\ n\binom{n-1}{i} &= (n-i)\binom{n}{i}, \end{aligned}$$

then

$$\begin{aligned} (n-i)\binom{n-1}{i}d_i &= n\binom{n-1}{i} + i\binom{n-1}{i}d_{i-1} \\ &= (n-i)\binom{n}{i} + (n-i)\binom{n-1}{i-1}d_{i-1}, \end{aligned}$$

and eventually

$$s_i = \binom{n-1}{i}d_i = \binom{n}{i} + \binom{n-1}{i-1}d_{i-1} = \binom{n}{i} + s_{i-1}.$$

Since $s_0 = 1$, hence

$$s_k = \sum_{i\in[0,k]}\binom{n}{i} = 2^n - \sum_{i\in[k+1,n]}\binom{n}{k},$$

and

$$d_k = \binom{n-1}{k}^{-1}\sum_{i\in[0,k]}\binom{n}{i} = \binom{n-1}{k}^{-1}\left(2^n - \sum_{i\in[k+1,n]}\binom{n}{k}\right),$$

and finally

$$\begin{aligned} h_{0,n} &= \sum_{k\in[0,n-1]} d_k \\ &= \sum_{k\in[0,n-1]}\binom{n-1}{k}^{-1}\sum_{i\in[0,k]}\binom{n}{i}. \end{aligned}$$

Since $d_{n-1} = 2^n - 1$, $d_{n-2} \le 2^n/(n-1)$, and

$$\begin{aligned} \sum_{k\in[2,n-3]}\binom{n-1}{k}^{-1} &\le \sum_{k\in[2,n-3]}\binom{n-1}{2}^{-1} \\ &= \frac{2(n-4)}{(n-1)(n-2)} = O(n^{-1}). \end{aligned}$$

then $h_{0,n} = O(N)$. By lemma 1.7.81 again, the covering time is $O(nN)$. □

1.7.6 Parrondo's Paradox

Parrondo's paradox shows that, sometimes 2 losing games combined together makes a winning game.

- Consider a game A with a biased coin A, winning with probability p_A and losing with probability $1 - p_A$.
- Consider a game B, that the coin flipped depends on the state of the game.

  Let w be the number of games won so far, and ℓ be the number of games lost so far, then $w - \ell$ is the winning.

  - If $w - \ell \equiv 0 \mod 3$, then use coin B winning with probability p_B and losing with probability $1 - p_B$.
  - Otherwise, use coin C winning with probability p_C and losing with probability $1 - p_C$.

  To show whether game B is winning or losing, we use the following lemma.

Lemma 1.7.85. *For any sequence s of moves that starts at 0 and ends at 3 before reaching -3, let $f(s)$ be a 1-to-1 mapping to s such that $f(s)$ negates every number starting from the last 0 in the sequence s, then*

$$\frac{\Pr[s \text{ occurs}]}{\Pr[f(s) \text{ occurs}]} = \frac{p_B p_C^2}{(1-p_B)(1-p_C)^2}.$$

Proof. We can view the game B by a Markov chain over $[-3, 3]$ that starts from 0, with ± 3 being 2 absorbing states. Similar to the Gambler's ruin argument in example 1.7.51, if it is more likely to reach -3 before reaching 3, then the game is a losing game.

Now consider the 1-to-1 mapping. Let

- a_1 be the number of moves $0 \to 1$,
- a_2 be the number of moves from $0 \to -1$,
- a_3 be the number of moves $1 \to 2$, $2 \to 3$, $-2 \to -1$, $-1 \to 0$,
- a_4 be the number of moves $2 \to 1$, $1 \to 0$, $-1 \to -2$, $-2 \to -3$

in s. Then

$$\Pr[s \text{ occurs}] = p_B^{a_1}(1-p_B)^{a_2} p_C^{a_3}(1-p_C)^{a_4}.$$

Consider $f(s)$ reaching -3 before reaching 3 that is 1-to-1 mapped to s.

- Since the first move after last 0 in s is $0 \to 1$, then the first move after last 0 in $f(s)$ is $0 \to -1$, hence $a_1' = a_1 - 1$. By symmetry of $a_1 + a_2 = a_1' + a_2'$, we have $a_2' = a_2 + 1$.
- After the last $0 \to 1$ in s, there is no $1 \to 0$ moves. The remaining moves in s are $1 \to 2$, $2 \to 1$, and $2 \to 3$. For $f(s)$, the remaining moves are $-1 \to -2$, $-2 \to -1$, and $-2 \to -3$.

  All the moves $-1 \to -2$ then $-2 \to -1$ in $f(s)$ are self-cancelling, symmetric to the $1 \to 2$ then $2 \to 1$ on s side.

  What remains is $1 \to 2 \to 3$ in s, and $-1 \to -2 \to -3$ in $f(s)$. Hence, $a_3' = a_3 - 2$, and $a_4' = a_4 + 2$.

Therefore,

$$\Pr[f(s) \text{ occurs}] = p_B^{a_1 - 1}(1-p_B)^{a_2+1} p_C^{a_3-2}(1-p_C)^{a_4+2},$$

and the ratio $\eta = p_B p_C^2 (1-p_B)^{-1}(1-p_C)^{-2}$. □

Suppose $p_A = 0.49$, $p_B = 0.09$, $p_C = 0.74$, then both game A and game B are both losing, as $\eta \approx 0.8$.

However, with a fair coin D with $p_D = 0.5$, we combine the game by winning on coin D plays game A, and losing on coin D plays game B. In this case, $p_B' = (p_A + p_B)/2$ and $p_C' = (p_A + p_C)/2$, and eventually $\eta > 1$.

The difference lies in the fair coin D breaks the structure of game B, that is rather likely to lose on any game with $w - \ell \equiv 0 \mod 3$. If one managed to get over the barrier in game B with game A, which is close to a fair game, then it is likely to win the next 2 games.

1.8 Continuous Distributions and the Poisson Process

TODO: ...

1.9 The Normal Distribution

TODO: ...

1.10 Entropy, Randomness, and Information

1.10.1 The Entropy Function

Definition 1.10.1 (Discrete Entropy)**.** The entropy in bits of a discrete random variable X is

$$H(X) = -\sum_x \Pr[X = x] \cdot \log_2 \Pr[X = x],$$

where the summation is over the support of X, namely over all values x in the range of X. Equivalently,

$$H(X) = \mathbb{E}\left[\log_2 \frac{1}{\Pr[X]}\right].$$

Definition 1.10.2 (Binary Entropy)**.** The binary entropy function $H(p)$ for a random variable that assumes only two possible outcomes, one of which occurs with probability p, is

$$H(p) = -p \log_2 p - (1-p)\log_2(1-p).$$

Exercise 1.10.3 (Exercise 10.5 [MU17])**.** If p is chosen uniformly at random from real interval $[0, 1]$, derive $\mathbb{E}\left[H(p)\right]$.

Proof. Since $H(p) = -p \log_2 p - (1-p)\log_2(1-p)$ is symmetric, hence we derive the part on the left hand side by

$$-\int_0^1 p \log_2 p \, dp = -\frac{1}{\ln 2}\int_0^1 p \ln p \, dp = -\frac{1}{\ln 2}\left(\frac{1}{2}p^2 \ln p \Big|_0^1 - \int_0^1 \frac{1}{2}p^2 \cdot \frac{1}{p}\, dp\right) = \frac{1}{4\ln 2},$$

where the second equality holds by integration by parts. Therefore, $\mathbb{E}\left[H(p)\right] = 1/(2\ln 2)$. □

Exercise 1.10.4 (Exercise 10.2 [MU17])**.** Consider an n-sided die, where the i^{th} face comes up with probability p_i. Show that the entropy of a die roll is maximized when each face comes up with equal probability.

Proof. Let X be the random variable for the face showing up for the n-sided die. The probability of the face showing up can be decided by a set of probabilities $\{p_i\}_{i\in I}$ where $|I| = n-1$, such that $k \notin I$ and $I \cup \{k\} = [1, n]$.

On an index $j \in I$, take partial derivative over p_j for $H(X)$,

$$\frac{\partial}{\partial p_j}H(X) = -\frac{\partial}{\partial p_j}\left(p_j \log_2 p_j\right) - \frac{\partial}{\partial p_j}\left(p_k \log_2 p_k\right) = -\frac{1}{\ln 2}\left((\ln p_j + 1) - (\ln p_k + 1)\right) = \log_2 \frac{p_k}{p_j},$$

where $p_k = 1 - \sum_{i\in I} p_i$. By symmetry, k can be arbitrary over $[1, n]$, hence if the partial derivative is 0 for any pair of $j \neq k$, the only solution would be $p_i = 1/n$ for any $i \in [1, n]$. □

Remark 1.10.5. For a fair n-faced die, the entropy is $\log_2 n$ bits of randomness.

Exercise 1.10.6 (Exercise 10.4 [MU17])**.** Let $p_k = \left(k \ln^2 k\right)^{-1}$, and show that $S = \sum_{k\geq 2} p_k$ is finite. Let integer-valued discrete random variable X has distribution $\Pr[X = k] = p_k/S$ for $k \geq 2$, show that $H(X)$ is unbounded.

Proof. Let $g(k) = -(\ln k)^{-1}$, then $g'(k) = \left(k \ln^2 k\right)^{-1}$. Hence,

$$S \leq \int_{\frac{3}{2}}^{\infty} \frac{1}{k \ln^2 k}\, dk = \frac{-1}{\ln k}\Big|_{\frac{3}{2}}^{\infty} = \frac{1}{\ln 3 - \ln 2}, \tag{1.9}$$

which is upper bounded by a bounded value. For the entropy, we have

$$H(X) = \sum_{k\geq 2} \frac{\log_2\left(Sk \ln^2 k\right)}{Sk \ln^2 k} = \frac{1}{\ln 2}\sum_{k\geq 2} \frac{\ln\left(Sk \ln^2 k\right)}{Sk \ln^2 k} = \frac{1}{\ln 2}\sum_{k\geq 2}\left(\boxed{\frac{\ln S}{Sk \ln^2 k}}_1 + \boxed{\frac{1}{Sk \ln k}}_2 + \boxed{\frac{2\ln\ln k}{Sk \ln^2 k}}_3\right).$$

The sum over the first term is bounded by eq. (1.9) or by definition of X. For the sum over the third term, the sum over the first few terms with $2 \le k < 6$ is bounded. For $k \ge 6$, the sum over the third term is bounded by

$$\sum_{k\ge 6} \frac{2\ln\ln k}{Sk\ln^2 k} \le \frac{2}{S}\int_5^\infty \frac{\ln\ln k}{k\ln^2 k}\,dk = \frac{2}{S}\int_{\ln 5}^\infty \frac{\ln u}{u^2}\,du = \frac{2}{S}\left(\frac{-\ln u}{u}\Big|_{\ln 5}^\infty + \int_{\ln 5}^\infty \frac{1}{u^2}\,du\right) = -\frac{2}{S}\cdot\frac{\ln u+1}{u}\Big|_{\ln 5}^\infty,$$

where $u = \ln k$, and the summation is finitely bounded. The sum over the second term is

$$\sum_{k\ge 2} \frac{1}{Sk\ln k} \ge \frac{1}{S}\int_2^\infty \frac{1}{k\ln k}\,dk = \frac{1}{S}\int_{\ln 2}^\infty \frac{1}{u}\,du = \frac{1}{S}\ln u\Big|_{\ln 2}^\infty,$$

which is unbounded. We conclude that $H(X)$ is unbounded. □

Lemma 1.10.7. *Let X_1, X_2 be independent random variables, and let $Y = (X_1, X_2)$. Then $H(Y) = H(X_1) + H(X_2)$.*

Proof. Let X_1 has support Ω_1, and X_2 has support Ω_2. We show by

$$\begin{aligned}
H(Y) &= -\sum_{x_1\in\Omega_1, x_2\in\Omega_2} \Pr[X_1 = x_1, X_2 = x_2]\cdot\log_2\Pr[X_1 = x_1, X_2 = x_2] \\
&= -\sum_{x_1\in\Omega_1, x_2\in\Omega_2} \Pr[X_1 = x_1]\cdot\Pr[X_2 = x_2]\cdot\left(\log_2\Pr[X_1 = x_1] + \log_2\Pr[X_2 = x_2]\right) \\
&= -\sum_{x_1\in\Omega_1, x_2\in\Omega_2} \left(\Pr[X_1 = x_1]\cdot\log_2\Pr[X_1 = x_1]\cdot\Pr[X_2 = x_2] + \Pr[X_2 = x_2]\cdot\log_2\Pr[X_2 = x_2]\cdot\Pr[X_1 = x_1]\right) \\
&= -\sum_{x_1\in\Omega_1}\Pr[X_1 = x_1]\cdot\log_2\Pr[X_1 = x_1] - \sum_{x_2\in\Omega_2}\Pr[X_2 = x_2]\cdot\log_2\Pr[X_2 = x_2] = H(X_1) + H(X_2),
\end{aligned}$$

where the second and the last equality holds by the independence between X_1 and X_2. □

Exercise 1.10.8 (Exercise 10.6 [MU17])**.** The conditional entropy $H(Y \mid X)$ is defined by

$$H(Y \mid X) = -\sum_{x,y}\Pr\left[X = x, Y = y\right]\cdot\log_2\Pr\left[Y = y \mid X = x\right].$$

If $Z = (X, Y)$, show that $H(Z) = H(X) + H(Y \mid X)$.

Proof. This can be seen as an extended version of lemma 1.10.7. We show by

$$\begin{aligned}
H(Z) &= -\sum_{x,y}\Pr\left[X = x, Y = y\right]\cdot\log_2\Pr\left[X = y, Y = y\right] \\
&= -\sum_{x,y}\Pr\left[X = x, Y = y\right]\cdot\left(\log_2\Pr[X = x] + \log_2\Pr\left[Y = y \mid X = x\right]\right) \\
&= -\sum_{x,y}\left(\Pr[X = x]\cdot\log_2\Pr[X = x] + \Pr\left[X = x, Y = y\right]\cdot\log_2\Pr\left[Y = y \mid X = x\right]\right) \\
&= H(X) + H(Y \mid X),
\end{aligned}$$

where the third equality holds by summing $\Pr\left[X = x, Y = y\right]$ over y preserves the marginal probability of X. □

Corollary 1.10.9 (Chain Rule of Conditional Entropy)**.**

$$H(X_1, \ldots, X_n) = \sum_{i\in[1,n]} H(X_i \mid X_1, \ldots, X_{i-1}).$$

Corollary 1.10.10 (Bayes' Rule of Conditional Entropy)**.**

$$H(Y \mid X) = H(X \mid Y) + H(Y) - H(X).$$

Definition 1.10.11 (Mutual Information)**.** The mutual information of a pair of random variables (X, Y) is defined by

$$I(X;Y) = H(X) - H(X \mid Y) = H(Y) - H(Y \mid X).$$

Remark 1.10.12. For mutual information in definition 1.10.11, we can expand the expression by

$$\begin{aligned}
H(Y) - H(Y \mid X) &= -\sum_{y} \Pr[Y=y] \cdot \log_2 \Pr[Y=y] + \sum_{x,y} \Pr[Y=y, X=x] \cdot \log_2 \Pr[Y=y \mid X=x] \\
&= -\sum_{x,y} \Pr[Y=y, X=x] \cdot \log_2 \Pr[Y=y] + \sum_{x,y} \Pr[Y=y, X=x] \cdot \log_2 \Pr[Y=y \mid X=x] \\
&= -\sum_{x,y} \Pr[Y=y, X=x] \cdot \left(\log_2 \Pr[Y=y] - \log_2 \Pr[Y=y \mid X=x]\right) \\
&= -\sum_{x,y} \Pr[Y=y, X=x] \cdot \left(\log_2 \Pr[Y=y] + \log_2 \Pr[X=x] - \log_2 \Pr[Y=y, X=x]\right) \\
&= \sum_{x,y} \Pr[Y=y, X=x] \cdot \log_2 \frac{\Pr[Y=y, X=x]}{\Pr[Y=y] \cdot \Pr[X=x]}.
\end{aligned}$$

Definition 1.10.13 (KL Divergence [KL51])**.** The Kullback-Leibler (KL) divergence is a statistical distance defined as

$$D_{\mathsf{KL}}(P \parallel Q) = \sum_{x} P(x) \cdot \log_2 \frac{P(x)}{Q(x)}.$$

Remark 1.10.14 (Mutual Information as KL Divergence). In KL divergence by definition 1.10.13 and mutual information with remark 1.10.12, if we consider $\mathcal{D}_{(X,Y)}$, that is the joint distribution of (X, Y), and $\mathcal{D}_X \times \mathcal{D}_Y$, which is the marginal distribution of (X, Y), mutual information fits into KL divergence by

$$D_{\mathsf{KL}}\left(\mathcal{D}_{(X,Y)} \parallel \mathcal{D}_X \times \mathcal{D}_Y\right) = I(X;Y).$$

Moreover, we can again expand the expression in remark 1.10.12 by

$$\begin{aligned}
H(Y) - H(Y \mid X) &= \sum_{x,y} \Pr[Y=y, X=x] \cdot \log_2 \frac{\Pr[Y=y, X=x]}{\Pr[Y=y] \cdot \Pr[X=x]} \\
&= \sum_{x,y} \Pr[Y=y \mid X=x] \cdot \Pr[X=x] \cdot \log_2 \frac{\Pr[Y=y \mid X=x]}{\Pr[Y=y]} \\
&= \sum_{x} \Pr[X=x] \cdot \left(\sum_{y} \Pr[Y=y \mid X=x] \cdot \log_2 \frac{\Pr[Y=y \mid X=x]}{\Pr[Y=y]} \right) \\
&= \mathbb{E}\left[D_{\mathsf{KL}}\left(\mathcal{D}_{Y|X} \parallel \mathcal{D}_Y\right)\right],
\end{aligned}$$

where $\mathcal{D}_{Y|X}$ is conditional distribution of Y given X with distribution $\mathcal{D}_X$.

Lemma 1.10.15 (Gibbs' Inequality)**.** *Given 2 distributions P and Q over the same support,*

$$D_{\mathsf{KL}}(P \parallel Q) \geq 0.$$

Proof. By Jensen's inequality theorem 1.2.5,

$$\sum_{x \in \Omega} P(x) \cdot \log_2 \frac{Q(x)}{P(x)} \leq \log_2 \left(\sum_{x \in \Omega} P(x) \cdot \frac{Q(x)}{P(x)} \right) = 0.$$

□

Corollary 1.10.16 (Monotonicity of Conditional Entropy)**.**

$$H(Y \mid X) \leq H(Y).$$

1.10.2 Entropy and Binomial Coefficients

For lemma 1.7.1, we give a tighter version as follows, to be used soon.

Lemma 1.10.17 (Tighter Stirling's Formula)**.**

$$\sqrt{2\pi n}\left(\frac{n}{e}\right)^n \le n! \le \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \cdot \exp\left(\frac{1}{12n}\right).$$

Lemma 1.10.18. *For an integer $nq \in [0, n]$, we have*

$$\frac{2^{nH(q)}}{2\sqrt{n}} \le \binom{n}{nq} \le 2^{nH(q)}.$$

Proof. We begin with the upper bound. Since

$$1 = \big(q + (1-q)\big)^n \ge \binom{n}{nq} q^{nq}(1-q)^{n(1-q)},$$

then

$$\log_2 \binom{n}{nq} \le -nq\log_2 q - n(1-q)\log_2(1-q) = nH(q).$$

The lower bound is trivial when $n = 1$, and it is also trivial when $q = 0$ or 1 for $n \ge 2$. We derive the lower bound for $n \ge 2$ and $q \in (0, 1)$ using the following upper and lower bounds by lemma 1.10.17:

$$\begin{aligned}
n! &\ge \sqrt{2\pi n}\left(\frac{n}{e}\right)^n \\
(nq)! &\le \sqrt{2\pi nq}\cdot\left(\frac{nq}{e}\right)^{nq}\cdot\exp\left(\frac{1}{12nq}\right) \\
\big(n(1-q)\big)! &\le \sqrt{2\pi n(1-q)}\cdot\left(\frac{n(1-q)}{e}\right)^{n(1-q)}\cdot\exp\left(\frac{1}{12n(1-q)}\right),
\end{aligned}$$

and therefore

$$\begin{aligned}
\binom{n}{nq} &\ge \frac{1}{\sqrt{2\pi nq(1-q)}}\cdot\exp\left(-\frac{1}{12nq(1-q)}\right)\cdot\left(\frac{1}{q}\right)^{nq}\cdot\left(\frac{1}{1-q}\right)^{n(1-q)} \\
&= \frac{2^{nH(q)}}{\sqrt{2\pi nq(1-q)}}\cdot\exp\left(-\frac{1}{12nq(1-q)}\right).
\end{aligned} \tag{1.10}$$

Let $k = q(1-q)$ such that $0 < k \le 1/4$, then

$$\frac{d}{dk}\left(\frac{1}{\sqrt{k}}\cdot\exp\left(-\frac{1}{12nk}\right)\right) = -\frac{1}{2k^{3/2}}\cdot\exp\left(-\frac{1}{12nk}\right) + \frac{1}{\sqrt{k}}\cdot\exp\left(-\frac{1}{12nk}\right)\cdot\frac{1}{12nk^2} = \frac{1}{k^{3/2}}\cdot\exp\left(-\frac{1}{12nk}\right)\cdot\left(\frac{1}{12nk} - \frac{1}{2}\right),$$

which is negative for $k \ge (6n)^{-1}$, and $k \ge (n-1)/n^2 \ge (6n)^{-1}$ when $q \in (0, 1)$, so eq. (1.10) is minimized when $k = 1/4$, giving

$$\binom{n}{nq} \ge \frac{2^{nH(q)}}{\sqrt{\pi n/2}}\cdot\exp\left(-\frac{1}{3n}\right).$$

Moreover,

$$\left(\frac{2\sqrt{n}}{\sqrt{\pi n/2}}\right)\cdot\exp\left(-\frac{1}{3n}\right) = \sqrt{\frac{8}{\pi}}\cdot\exp\left(-\frac{1}{3n}\right) > 1$$

for all $n \ge 1$, then

$$\binom{n}{nq} \ge \frac{2^{nH(q)}}{2\sqrt{n}}.$$

□

Corollary 1.10.19. *When* $0 \le q \le 1/2$,

$$\frac{2^{nH(q)}}{2\sqrt{n}} \le \binom{n}{\lceil nq \rceil}, \quad \binom{n}{\lfloor nq \rfloor} \le 2^{nH(q)}.$$

When $1/2 \le q \le 1$,

$$\frac{2^{nH(q)}}{2\sqrt{n}} \le \binom{n}{\lfloor nq \rfloor}, \quad \binom{n}{\lceil nq \rceil} \le 2^{nH(q)}.$$

1.10.3 Entropy: A Measure of Randomness

Definition 1.10.20 (Variable-Length Fair-Bit Extraction Function)**.** Let $|\mathbf{y}|$ be the number of bits in a sequence of bits $\mathbf{y}$. An extraction function Ext takes as input the value of a random variable X, and outputs a sequence of bits. For every sequence of bits $\mathbf{y}$ with $|\mathbf{y}| = k$,

$$\Pr\left[\mathsf{Ext}\left(X\right) = \mathbf{y} \mid |\mathsf{Ext}(X)| = k\right] = 2^{-k},$$

whenever $\Pr\left[|\mathsf{Ext}(X)| = k\right] > 0$.

Remark 1.10.21. We can view the input X as the outcomes of n biased coin flips, and the output $\mathbf{y}$ as k fair coin flips. In this sense, if the biased coin comes up heads with probability p, the entropy of the input is $nH(p)$, which is the maximum possible expected number of fair bits that any extraction function can output.

There is no guarantee that Ext is efficient (runs in polynomial time).

Example 1.10.22. Consider a construction of Ext for X uniformly distributed over $[0, 11]$. The elements in $[0, 7]$ can be represented in 3 bits, while we can associate the 4 probabilities in $[8, 11]$ with 2 bits. Associate each input $[0, 11]$ to each 2 or 3 bit output, such that they output with the same probability. Hence, the probability of outputting 3 bits is $2/3$, while the probability of outputting 2 bits is $1/3$, and

$$\Pr\left[\mathsf{Ext}\left(X\right) = \mathbf{y} \mid |\mathsf{Ext}(X)| = k\right] = 2^{-k}$$

for k being 2 or 3, such that definition 1.10.20 is satisfied.

Theorem 1.10.23. *Suppose the value of a random variable X is uniform over $[0, m-1]$, such that $H\left(X\right) = \log_2 m$, then there is an extraction function for X that outputs on average at least $\lfloor \log_2 m \rfloor - 1 = \lfloor H(X) \rfloor - 1$ independent and unbiased bits.*

Proof. Suppose m is a power of 2, the extraction function trivially outputs the value's $\lfloor \log_2 m \rfloor$ bits representation.

In the general case where m is not a power of 2, we first show in an inductive way. Let $\alpha = \lfloor \log_2 m \rfloor$, Y be the random variable for the length of output bits, and for the elements in $[2^\alpha, m-1]$, the theorem 1.10.23 still holds, that

$$\mathbb{E}\left[Y\right] \ge \frac{2^\alpha}{m}\alpha + \frac{m-2^\alpha}{m}\left(\left\lfloor \log_2\left(m-2^\alpha\right)\right\rfloor - 1\right) = \alpha - \frac{m-2^\alpha}{m}\left(\alpha - \left\lfloor \log_2\left(m-2^\alpha\right)\right\rfloor + 1\right).$$

Let $\beta = \left\lfloor \log_2\left(m-2^\alpha\right)\right\rfloor$, then since

$$\frac{2^{\beta+1}}{2^{\beta+1}+2^\alpha} \ge \frac{m-2^\alpha}{m} \ge \frac{2^\beta}{2^\beta+2^\alpha},$$

we have

$$\mathbb{E}\left[Y\right] \ge \alpha - \frac{m-2^\alpha}{m}\left(\alpha-\beta+1\right) \ge \alpha - \frac{2^{\beta+1}}{2^{\beta+1}+2^\alpha}\left(\alpha-\beta+1\right) = \alpha - \frac{1}{1+2^{\alpha-\beta-1}}\left(\alpha-\beta+1\right) \ge \alpha - 1,$$

as $2^{\alpha-\beta-1} \ge \alpha - \beta$ holds for any $\alpha > \beta$.

We show in an explicit construction. Consider k distinct indices $\{\beta_i\}_{i\in[1,k]} \subseteq [0, \alpha]$ in descending order, satisfying

$$m = \sum_{i\in[1,k]} 2^{\beta_i}, \quad s_i = \sum_{j\in[1,i]} 2^{\beta_j},$$

such that $[0, m-1]$ is partitioned into k subsets, and the i^{th} subset is $[s_{i-1}, s_i - 1]$, of size 2^{β_i}, then

$$\mathbb{E}[Y] = \sum_{i\in[1,k]} \frac{2^{\beta_i}}{m}\beta_i,$$

and since $\beta_1 = \alpha$,

$$\mathbb{E}[Y] - (\alpha - 1) = \frac{2^{\beta_1}}{m} + \sum_{i\in[2,k]} \frac{2^{\beta_i}}{m}\left(\beta_i - (\alpha-1)\right) = \frac{1}{m}\left(2^{\beta_1} - \sum_{i\in[2,k]} 2^{\beta_i}\left((\alpha-1) - \beta_i\right)\right).$$

Considering

$$2^{\beta_1} - \sum_{i\in[2,k]} 2^{\beta_i}\left((\alpha-1)-\beta_i\right) \geq 2^{\alpha} - \sum_{i\in[0,\alpha-2]} 2^i \cdot (\alpha - 1 - i) = 2^\alpha - ((2^\alpha - 1) - \alpha) = \alpha + 1 > 0,$$

then $\mathbb{E}[Y] > \alpha - 1$. □

Exercise 1.10.24 (Exercise 10.9 [MU17]). We have shown in theorem 1.10.23 that an extraction function for a random variable X uniform over $[0, m-1]$ extracts on average at least $\lfloor H(X)\rfloor - 1$ unbiased bits. Show that there is an extraction function for $\{X_i\}_{i\in[1,k]}$ independent random variables uniform over $[0, m-1]$ extracting on average at least $k\lfloor H(X)\rfloor - 1$ unbiased bits.

Proof. Apparently, one can invoke k times the extraction function in theorem 1.10.23, and by linearity of expectation, outputting on average at least $k\lfloor H(X)\rfloor - k$ unbiased bits. The idea is to construct W over $[0, m^k - 1]$ by

$$W = X_1 + mX_2 + m^2X_3 + \cdots + m^{k-1}X_k,$$

and $H(W) = H(X_1, \ldots, X_k) = k\log_2 m$. The average number of output bits by theorem 1.10.23 is

$$\left\lfloor k\log_2 m\right\rfloor - 1 \geq k\left\lfloor \log_2 m\right\rfloor - 1.$$

□

Exercise 1.10.25 (Exercise 10.10 [MU17]). Suppose we have an oracle of generating independent fair coin flips, and m is not a power of 2.

- Show that one cannot uniformly sample over $[0, m-1]$ with exactly k fair coin flips, for any fixed k.
- Show that one cannot uniformly sample over $[0, m-1]$ with at most k fair coin flips, for any fixed k.
- Give an algorithm using fair coins to sample uniformly over $[0, m-1]$, with expected flips at most $2\left\lceil\log_2 m\right\rceil$.

Proof. We first rule out the case that $2^k < m$, as it is impossible to associate 2^k outcomes with m numbers.

If one tries to uniformly sample over $[0, m-1]$ with exactly k fair coins, $m \nmid 2^k$ as m is not a power of 2, and one cannot associate m numbers with 2^k outcomes in an equal-sized partition manner.

In the at most k flips case, pad each execution to exactly k flips: if the algorithm stops after ℓ flips with $\ell < k$, then all $2^{k-\ell}$ possible continuations are counted as producing the same outcome. Hence each output a corresponds to some integer number N_a of the 2^k length-k bit strings, so

$$\Pr\left[\text{outputting } a\right] = N_a \cdot 2^{-k}.$$

Uniformity requires $N_a/2^k = 1/m$ for every a, so $m \mid 2^k$, which is impossible because m is not a power of 2, failing in the same manner as the exact k tosses.

An algorithm sampling uniformly over $[0, m-1]$ is rejection sampling: keep tossing $\left\lceil\log_2 m\right\rceil$ fair coins, until the outputting integer is within range $[0, m-1]$. A round being rejected has probability at most $1/2$, and the expected number of rounds is therefore at most 2, thus the expected runtime is at most $2\left\lceil\log_2 m\right\rceil$. □

Exercise 1.10.26 (Exercise 10.11 [MU17])**.** Suppose we have an oracle of generating independent fair coin flips.

- Give an algorithm using the fair coin that simulates flipping a biased coin that comes up heads with probability p. The expected number of flips should be at most 2.
- Give an algorithm using the coin to sample uniformly over $[0, m-1]$ with expected flips at most $\lceil \log_2 m \rceil + 2$.

Proof. Write p into its binary form, such that there exists bits $\{b_i\}_{i\geq 1}$, to binary decompose p into the form $\sum_{i\geq 1} b_i/2^i$. Keep tossing independent fair coins $\{X_i\}_{i\geq 1}$ until $X_i \neq b_i$. If $X_i > b_i$, the biased coin comes up tails, otherwise heads. Since $\Pr[X_i \neq b_i] = 1/2$, the expected number of sampling is at most 2, as there might be finite number of $\{b_i\}_{i\geq 1}$.

A first attempt at sampling uniformly over $[0, m-1]$ is an extension of the biased coin sampling, with probability $m/2^\alpha$ the biased coin comes up heads, where $\alpha = \lceil \log_2 m \rceil$. In each attempt, fair coins are tossed sequentially, and they are used both to simulate the biased coin, and maintained as the prefix of an α-bit integer.

- If the biased coin comes up tails, then any completion of the current prefix ties to an integer at least m, so restart.
- If the biased coin comes up heads, the generated number is within $[0, m-1]$; keep tossing until α fair coins have been sampled, and output the resulting α-bit integer.

Let $A_i = (Y_i, Z_i)$ be the random variable for the i^{th} biased coin simulation, where Y_i is the number of fair coins tossed, and Z_i is the outcome of the biased coin simulation. Let T be a stopping time for $A_1, A_2, \ldots$ on the first heads of the biased coin. Then the total number of fair coin flips is

$$W = \sum_{i\in[1,T]} Y_i + (\alpha - Y_T) = \sum_{i\in[1,T-1]} Y_i + \alpha.$$

Since Z_i is heads with probability $m/2^\alpha$, then $\mathbb{E}[T] = 2^\alpha/m \leq 2$. Moreover, for each Y_i,

$$\mathbb{E}[Y_i] = \sum_{k\in[1,\alpha]} \frac{k}{2^k} + \frac{\alpha}{2^\alpha} = 2 - \frac{2}{2^\alpha}.$$

Since $\mathbb{E}[T]$ and $\mathbb{E}[Y_i]$ are both bounded, by theorem 1.13.18,

$$\mathbb{E}\left[\sum_{i\in[1,T]} Y_i\right] = \mathbb{E}[Y_i]\cdot\mathbb{E}[T] = \frac{2^\alpha}{m}\cdot\left(2 - \frac{2}{2^\alpha}\right).$$

By linearity of expectation,

$$\mathbb{E}[W] = \frac{2^\alpha}{m}\cdot\left(2 - \frac{2}{2^\alpha}\right) + \alpha - \mathbb{E}[Y_T] \leq \alpha + 2\cdot\frac{2^\alpha - 1}{m} - 1 \leq \alpha + 3$$

where the second inequality holds by $\mathbb{E}[Y_T] \geq 1$, and the last inequality holds by $2^{\alpha-1} < m \leq 2^\alpha$.

This falls short of the desired $\alpha+2$ bound, because rejected attempts discard all previously sampled bits. We now give a recycling version that keeps the remaining randomness after rejection. Consider the following algorithm:

- Let $c \leftarrow 0$ and $v \leftarrow 1$. Keep looping:
  - Sample $b \xleftarrow{R} \{0,1\}$, and update $c \leftarrow c\times 2 + b$, $v \leftarrow v \times 2$.
  - Continue looping if $v < m$, otherwise if $v \geq m$:
    - If $c \geq m$, update $c \leftarrow c - m$, $v \leftarrow v - m$.
    - Otherwise, output c.

The invariant here is c is always uniform over $[0, v-1]$, and the algorithm is rejection sampling over $[0, m-1]$.

Indeed, if c is uniform over $[0, v-1]$, then after sampling $b \xleftarrow{R} \{0,1\}$, the value $2c+b$ is uniform over $[0, 2v-1]$. On $v \geq m$, conditioning on $c < m$ gives a uniform output over $[0, m-1]$; conditioning on $c \geq m$ makes $c-m$ uniform over $[0, v-m-1]$, preserving the invariant.

The new algorithm always runs α steps up to the first comparison against m, then let T be the step for the output. We have the following observation: At the t^{th} step, we have 2^t possible bit strings, separated into 2 classes:

- Resolved prefixes: On reading the prefix of a t-bit string, the algorithm stops at some bit and outputs a sample. Any continuation for a resolved prefix is considered stopped.
- Unresolved prefixes: The algorithm does not output after consuming all of the t bits as inputs.

Lemma 1.10.27. *After the first α bits, the unresolved prefixes at the t^{th} step form a possibly empty interval $\left[\left\lfloor 2^t/m \right\rfloor m, 2^t - 1\right]$.*

Proof. We prove by induction. Let $q_t = \left\lfloor 2^t/m \right\rfloor$. When $t = \alpha$, the unresolved prefixes are $[m, 2^\alpha - 1]$, since $q_\alpha = 1$.

Suppose the claim holds for $t = t'$. After tossing one more fair coin, each unresolved prefix has two continuations, so before the next rejection step the unresolved interval doubles to

$$\left[2q_{t'}m, 2^{t'+1} - 1\right].$$

Once the algorithm sees an entire block of size m in the interval at the bottom of the interval, the block is removed, as that goes to the new resolved prefixes. Hence the unresolved interval that remains is

$$\left[q_{t'+1}m, 2^{t'+1} - 1\right].$$

This proves the induction step. □

With lemma 1.10.27, we upper bound the number of unresolved prefixes at the t^{th} step v_t by m, and thus

$$\begin{aligned}
\mathbb{E}[T] &= \sum_{t\geq 0} \Pr[T > t] = \alpha + \sum_{t \geq \alpha} \Pr[T > t] \\
&= \alpha + \Pr[T > \alpha] + \sum_{t > \alpha} \Pr[T > t] \\
&\leq \alpha + \frac{2^\alpha - m}{2^\alpha} + \sum_{t \geq \alpha+1} \frac{m}{2^t} = \alpha + 1.
\end{aligned}$$

The last equality can be roughly relaxed to an upper bound of $\alpha + 2$, as $m/2^t \leq 1/2$ for $t > \alpha$, and the sum of the last 2 terms is upper bounded by 2. □

Theorem 1.10.28. *Suppose a coin comes up heads with probability $p > 1/2$. For any constant $\delta > 0$ and sufficiently large n,*

- *There exists an extraction function that on input sequences of n independent flips, outputs an average of at least $(1-\delta)nH(p)$ independent random bits.*
- *The average number of bits output by any extraction function on an input of n independent flips is at most $nH(p)$.*

Proof. We begin by describing an extraction function that extracts on average at least $(1-\delta)nH(p)$ random bits from n flips of the biased coin. After n flips, the number of heads coming up concentrates around np, which means the outcome is most likely to be one of the $\binom{n}{np}$ sequences with np heads up.

Conditioned that k heads coming up, the sequence with k heads is uniformly distributed over all $\binom{n}{k}$ sequences, which can be bijectively mapped to $\left[0, \binom{n}{k} - 1\right]$. We apply the randomness extraction function for uniform random variable in theorem 1.10.23 to extract on average $\left\lfloor \log_2 \binom{n}{k} \right\rfloor - 1$ fair bits. Moreover, by lemma 1.10.18, we bound the number of fair bits within the range of $\left[(1-\delta)nH(q), nH(q)\right]$, where $q = k/n$.

We now formalize the derivation of the bounds for the number of fair bits extracted from n flips of biased coins with probability p coming up heads. Let Z be a random variable representing the number of heads flipped, and let B be the random variable representing the bits extracted by the extraction function, then by conditional expectation,

$$\mathbb{E}[B] = \sum_{x\in[0,n]} \Pr[Z = x] \cdot \mathbb{E}[B \mid Z = x],$$

and by theorem 1.10.23,

$$\mathbb{E}[B \mid Z = x] \geq \left\lfloor \log_2 \binom{n}{x} \right\rfloor - 1.$$

Now consider the fact that the number of heads concentrates around np, to lower bound the average number of fair bits extracted, we consider only $x \in \left[\lceil (p-\varepsilon)n \rceil, \lfloor (p+\varepsilon)n \rfloor\right]$, where $\varepsilon \leq \min(p - 1/2, 1 - p)$. By monotonicity,

$$\mathbb{E}[B \mid Z = x] \geq \left\lfloor \log_2 \binom{n}{x} \right\rfloor - 1 \geq \left\lfloor \log_2 \binom{n}{\lfloor n(p+\varepsilon) \rfloor} \right\rfloor - 1 = \mathbb{E}\left[B \mid Z = \lfloor n(p+\varepsilon) \rfloor\right]. \tag{1.11}$$

By standard Chernoff's bound, let $\gamma = \varepsilon/p$ and $\mu = np$,

$$\Pr\left[\left|Z - np\right| \geq \varepsilon n\right] \leq \left(\frac{e^{-\gamma}}{(1-\gamma)^{1-\gamma}}\right)^{\mu} + \left(\frac{e^{\gamma}}{(1+\gamma)^{1+\gamma}}\right)^{\mu} \leq \exp\left(-\frac{n\varepsilon^2}{3p}\right) + \exp\left(-\frac{n\varepsilon^2}{2p}\right).$$

Alternatively, by KL Divergence driven Chernoff's bound in exercise 1.4.4,

$$\Pr\left[\left|Z - np\right| \geq \varepsilon n\right] \leq 2\exp\left(-2n\varepsilon^2\right).$$

Moreover, we can further lower bound eq. (1.11) by theorem 1.10.23, that

$$\begin{aligned} \mathbb{E}\left[B \mid Z = \lfloor n(p+\varepsilon) \rfloor\right] &\geq \left\lfloor \log_2 \binom{n}{\lfloor (p+\varepsilon)n \rfloor} \right\rfloor - 1 \geq \log_2 \binom{n}{\lfloor (p+\varepsilon)n \rfloor} - 2 \\ &\geq \log_2 \frac{2^{nH(p+\varepsilon)}}{2\sqrt{n}} - 2 = nH(p+\varepsilon) - \frac{1}{2}\log_2 n - 3. \end{aligned} \tag{1.12}$$

Therefore,

$$\begin{aligned} \mathbb{E}[B] &\geq \sum_{x\in\left[\lceil (p-\varepsilon)n \rceil, \lfloor (p+\varepsilon)n \rfloor\right]} \Pr[Z = x] \cdot \mathbb{E}[B \mid Z = x] \\ &\geq \Pr\left[\left|Z - np\right| \leq \varepsilon n\right] \cdot \mathbb{E}\left[B \mid Z = \lfloor (p+\varepsilon)n \rfloor\right] \\ &\geq \left(1 - 2\exp\left(-2n\varepsilon^2\right)\right) \cdot \left(nH(p+\varepsilon) - \frac{1}{2}\log_2 n - 3\right), \end{aligned}$$

where the last inequality holds by eq. (1.12). We conclude that, for any $\delta > 0$, we have

$$\mathbb{E}[B] \geq (1-\delta)\, nH(p)$$

for sufficiently small ε and sufficiently large n.

We now show no extraction function on average obtains more than $nH(p)$ fair bits. Suppose input bits $\mathbf{x}$ occurs with probability q, and the extraction function outputs $\mathbf{y}$ corresponding to input $\mathbf{x}$, then by definition 1.10.20,

$$\Pr[\mathsf{Ext}(X) = \mathbf{y} \mid |\mathsf{Ext}(X)| = k] = 2^{-k},$$

that conditioned the output number of bits from the extraction function is k, the probability of outputting $\mathbf{y}$ is 2^{-k}. Hence, for a fixed $\mathbf{y}$ with k bits, the probability of outputting $\mathbf{y}$ is immediate by the probability of outputting k bits,

$$\Pr[\mathsf{Ext}(X) = \mathbf{y}] = \Pr[\mathsf{Ext}(X) = \mathbf{y}, |\mathsf{Ext}(X)| = k] = 2^{-k} \cdot \Pr[|\mathsf{Ext}(X)| = k] \leq 2^{-k}.$$

Moreover, since

$$\Pr[\mathrm{Ext}(X) = \mathbf{y}] \geq \Pr[X = \mathbf{x}] = q,$$

then $q \leq 2^{-k}$, giving an upper bound on $k \leq -\log_2 q$. Therefore,

$$\mathbb{E}[B] = \sum_{\mathbf{x}} \Pr[X = \mathbf{x}] \cdot \mathbb{E}[B \mid X = \mathbf{x}] \leq -\sum_{\mathbf{x}} \Pr[X = \mathbf{x}] \cdot \log_2 \Pr[X = \mathbf{x}] = H(X) = nH(p),$$

where the second inequality is by upper bounding the output length, and the last equality holds by lemma 1.10.7. □

Exercise 1.10.29 (Exercise 10.12 [MU17])**.** Consider the extraction function $\mathcal{A}$ whose input is a sequence of n independent flips of a coin that comes up heads with probability $p > 1/2$. Break the sequence $X_1, \ldots, X_n$ into $\lfloor n/2 \rfloor$ pairs, such that $A_i = (X_{2i-1}, X_{2i})$, and consider the pairs in order. If A_i is heads and tails, it outputs a 0; if A_i is tails and heads, it outputs a 1; otherwise, move up to the next pair.

- Show that the bits extracted by $\mathcal{A}$ are independent and unbiased, and the expected number of bits extracted is

$$\lfloor n/2 \rfloor 2p(1-p) \approx np(1-p).$$

- We derive another set of flips $Y_1, \ldots$ from $X_1, \ldots, X_n$: Let j, k be 1, and repeat until $j > \lfloor n/2 \rfloor$: If A_j is both heads, Y_k is heads and increment j and k; if A_j is both tails, Y_k is tails and increment j and k; otherwise increment j.

  The intuition here is to take some randomness $\mathcal{A}$ cannot use effectively and reuse it. Show that the bits produced by running $\mathcal{A}$ over $Y_1, \ldots$ are independent and unbiased, and further argue that they are independent of those produced by running $\mathcal{A}$ over $X_1, \ldots, X_n$.
- We derive another set of flips $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$ from $X_1, \ldots, X_n$: Z_i is heads if A_i is both heads or tails, otherwise tails. Show that the bits produced by running $\mathcal{A}$ over $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$ are independent and unbiased, and further argue that they are independent of those produced by running $\mathcal{A}$ over $X_1, \ldots, X_n$, and $Y_1, \ldots$.
- After we run $\mathcal{A}$ over $Y_1, \ldots$ and $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$, we can recursively derive two further sequences from each of the sequences, namely $Y_1, \ldots$ and $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$, in the same way, run $\mathcal{A}$ on those, and so on.

  Let $A(p)$ be the average number of bits extracted for each flip in the sequence $X_1, \ldots, X_n$, in the limit as the length of the sequence $X_1, \ldots, X_n$ goes to infinity. Let $q = 1 - p$, and argue that $A(p)$ satisfies the recurrence

$$A(p) = pq + \frac{p^2 + q^2}{2} A\left(\frac{p^2}{p^2 + q^2}\right) + \frac{1}{2} A\left(p^2 + q^2\right).$$

- Show that $H(p)$ satisfies the same recurrence as $A(p)$.

Proof. First, for the bits directly extracted by $\mathcal{A}$, since the A_i are mutually independent, and for $A_i = (X_{2i-1}, X_{2i})$, the biased coin tosses are mutually independent, each output bit by $\mathcal{A}$ is independent of the others. The probability of a bit sampled by $\mathcal{A}$ being 1 is the same as the probability of a bit sampled by $\mathcal{A}$ being 0; both are $p(1-p)$, requiring a head and a tail in A_i, so the bits are unbiased. Since a bit is output with probability $2p(1-p)$ by $\mathcal{A}$, by linearity of expectation, the expected number of bits extracted is $\lfloor n/2 \rfloor 2p(1-p)$.

To show the bits produced by $\mathcal{A}$ over $Y_1, \ldots$ are independent and unbiased, we first show $Y_1, \ldots$ are independent. Since $X_1, \ldots, X_n$ are independent, then $A_1, \ldots, A_{\lfloor n/2 \rfloor}$ are independent. Since the retained A_i are both heads or both tails, then $Y_1, \ldots$ are independent, so are the output by $\mathcal{A}$ over $Y_1, \ldots$. To show the outputs by $\mathcal{A}$ from $Y_1, \ldots$ are unbiased, since each Y_i can be modeled by a biased coin with probability $p^2/(p^2 + q^2)$ coming up heads, applying the same argument for $\mathcal{A}$ producing unbiased bits over $X_1, \ldots, X_n$, the outputs by $\mathcal{A}$ over $Y_1, \ldots$ are unbiased.

To show the bits produced by $\mathcal{A}$ over $Y_1, \ldots$ are independent of the bits produced by $\mathcal{A}$ over $X_1, \ldots, X_n$, observe that the former uses only the same-faced A_i, while the latter uses only the distinct-faced A_i. In the latter case, HT and TH are equally likely, and the corresponding A_i are independent across index i. In the former case, the same-faced

A_i are independent across index i, and the values of these A_i are independent of the orders of the distinct-faced A_i. Hence, the bits extracted by $\mathcal{A}$ over $Y_1, \ldots$ are independent of the bits extracted by $\mathcal{A}$ over $X_1, \ldots, X_n$.

To show the bits produced by $\mathcal{A}$ over $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$ are independent and unbiased, we show $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$ are independent of each other, as $A_1, \ldots, A_{\lfloor n/2 \rfloor}$ are independent. Each Z_i can be modeled by a biased coin coming up heads with probability $p^2 + q^2$, then the bits extracted by $\mathcal{A}$ are independent and unbiased.

To show the bits produced by $\mathcal{A}$ over $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$ are independent of the bits from $\mathcal{A}$ over $X_1, \ldots, X_n$, for a bit produced by $\mathcal{A}$ over $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$, supposing it is extracted through (Z_i, Z_{i+1}), which are determined by (A_i, A_{i+1}), then it is possible that the bit overlaps with the bits extracted over A_i and/or A_{i+1}. Since the A_i are independent of each other, then all the other bits extracted over $X_1, \ldots, X_n$ are independent of the bit extracted over (Z_i, Z_{i+1}). Conditioned that the bit extracted over (Z_i, Z_{i+1}) overlaps with the bit extracted from A_i, then A_i can be HT or TH, the bit extracted from A_i can be 0 or 1, and Z_i is tails regardless, and thus the bit extracted over (Z_i, Z_{i+1}) is 1, which is independent of the bit extracted from A_i. Symmetrically, if the bit extracted over (Z_i, Z_{i+1}) overlaps with the bit extracted over A_{i+1}, the same argument applies, that the bit extracted over (Z_i, Z_{i+1}) is 0.

To show the bits produced by $\mathcal{A}$ over $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$ are independent of the bits from $\mathcal{A}$ over $Y_1, \ldots$, the strategy is similar to the last independence argument. Supposing a bit is extracted through (Z_i, Z_{i+1}), determined by (A_i, A_{i+1}), then by A_i being independent of each other, it suffices to only argue that the bit is independent of the bit extracted from $Y_1, \ldots$ using a Y_j that comes from A_i or A_{i+1}. Supposing A_j is overlapped, regardless of A_j being HH or TT, Z_j is heads, and the bit extracted from (Z_j, Z_{j+1}) is 0, independent of the bit extracted using Y_j. Symmetrically, if the overlap is with the second coordinate, then that coordinate is heads, and the bit extracted from the Z pair is 1.

For $Y_1, \ldots$, the biased coin comes up heads with probability $p^2/(p^2 + q^2)$, and the expected number of retained Y_i is $(p^2 + q^2)n/2$. For $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$, the biased coin has probability $p^2 + q^2$ of coming up heads. For $\mathcal{A}$ directly extracting from $X_1, \ldots, X_n$, the expected number of bits extracted is $\lfloor n/2 \rfloor 2pq \approx npq$. Thus, when $n \to \infty$, the average number of bits extracted per flip approaches

$$pq + \frac{p^2 + q^2}{2} A\left(\frac{p^2}{p^2 + q^2}\right) + \frac{1}{2} A\left(p^2 + q^2\right).$$

To show that $A(p) = H(p)$, we have

$$\frac{p^2 + q^2}{2} H\left(\frac{p^2}{p^2 + q^2}\right) = -p^2 \log_2 p - q^2 \log_2 q + \frac{p^2 + q^2}{2} \log_2\left(p^2 + q^2\right),$$

$$\frac{1}{2} H\left(p^2 + q^2\right) = -\frac{p^2 + q^2}{2} \log_2\left(p^2 + q^2\right) - pq \log_2(2pq),$$

and finally

$$\begin{aligned} A(p) &= pq - p^2 \log_2 p - q^2 \log_2 q - pq \log_2(2pq) \\ &= -p^2 \log_2 p - q^2 \log_2 q - pq \log_2 pq \\ &= -(p^2 + pq) \log_2 p - (q^2 + pq) \log_2 q \\ &= -p \log_2 p - q \log_2 q = H(p). \end{aligned}$$

□

Remark 1.10.30. Exercise 1.10.29 is called the Peres extractor [von51, Eli72, Per92]. The von Neumann method [von51] extracts the bits pairwise from $X_1, \ldots, X_n$, but throws away the randomness from the structure of the bit string, and there are two leftover sources of randomness: One is the structure of both heads or both tails, which gives $Y_1, \ldots$, and the other is the structure of being equal or not, which gives $Z_1, \ldots, Z_{\lfloor n/2 \rfloor}$. The von Neumann method took care of the symmetry case of a head and a tail, and the two leftovers are taken care of in the Peres extractor.

Exercise 1.10.31 (Exercise 10.13 [MU17])**.** Suppose we only have a biased 6-sided die with entropy $h > 0$ instead of a fair coin. Modify the extraction function in theorem 1.10.28 so that it extracts, on average, almost h random bits per roll from a sequence of die rolls.

Proof. We consider a variant construction based on theorem 1.10.28, where we roll the die n times. Conditioned on seeing n_k occurrences of the k^{th} face, the sequence is uniformly distributed over all

$$N = \binom{n}{n_1; n_2; n_3; n_4; n_5; n_6}$$

sequences. We apply the theorem 1.10.23 extractor to extract the random variable uniform over $[0, N-1]$.

Let Z_k be the random variable for the number of times the k^{th} face comes up, p_k be the probability that the k^{th} face comes up, and B be the random variable for the bits extracted. Define

$$\mathcal{T}_\varepsilon = \left\{ (n_1, \ldots, n_6) \ : \ \sum_{k\in[1,6]} n_k = n, \text{ and } \forall k \in [1,6], \left|n_k - np_k\right| \leq \varepsilon n \right\}.$$

By conditional expectation,

$$\begin{aligned}
\mathbb{E}[B] &= \sum_{n_1,\ldots,n_6} \mathbb{E}[B \mid Z_1 = n_1, \ldots, Z_6 = n_6] \cdot \Pr[Z_1 = n_1, \ldots, Z_6 = n_6] \\
&\geq \min_{(n_1,\ldots,n_6)\in\mathcal{T}_\varepsilon} \mathbb{E}[B \mid Z_1 = n_1, \ldots, Z_6 = n_6] \cdot \left(1 - \sum_{k\in[1,6]} \Pr\left[\left|Z_k - np_k\right| \geq \varepsilon n\right]\right) \\
&\geq \min_{(n_1,\ldots,n_6)\in\mathcal{T}_\varepsilon} \left(\left\lfloor \log_2 \binom{n}{n_1; \ldots; n_6} \right\rfloor - 1\right) \cdot \left(1 - \sum_{k\in[1,6]} \Pr\left[\left|Z_k - np_k\right| \geq \varepsilon n\right]\right),
\end{aligned}$$

where the second inequality holds by union bounding, and the third inequality holds by theorem 1.10.23.

We now consider a variant of lemma 1.10.18 for the biased 6-sided die.

Lemma 1.10.32. *For integers $\{nq_i\}_{i\in[1,6]}$ in $[0,n]$ and $\sum_{i\in[1,6]} q_i = 1$, we have*

$$\frac{1}{2n^{5/2}} \cdot 2^{nH(q_1,\ldots,q_6)} \leq \binom{n}{nq_1; \ldots; nq_6} \leq 2^{nH(q_1,\ldots,q_6)},$$

where $H(q_1, \ldots, q_6) = -\sum_{i\in[1,6]} q_i \log_2 q_i$.

Proof. We begin with the upper bound. Since

$$1 = (q_1 + \ldots + q_6)^n \geq \binom{n}{nq_1; \ldots; nq_6} \prod_{i\in[1,6]} q_i^{nq_i},$$

then we complete the upper bound proof by

$$\log_2 \binom{n}{nq_1; \ldots; nq_6} \leq -n \sum_{i\in[1,6]} q_i \log_2 q_i = nH\left(q_1, \ldots, q_6\right).$$

The lower bound is trivial for $n \in [1,6]$. Let $I \subseteq [1,6]$ be the subset of indices such that $nq_i > 0$. By lemma 1.10.17,

$$n! \geq \sqrt{2\pi n}\left(\frac{n}{e}\right)^n, \quad (nq_i)! \leq \sqrt{2\pi nq_i}\left(\frac{nq_i}{e}\right)^{nq_i} \exp\left(\frac{1}{12nq_i}\right),$$

and therefore

$$\binom{n}{nq_1; \ldots; nq_6} \geq \left(\frac{1}{\sqrt{2\pi n}}\right)^{|I|-1} \cdot \prod_{i\in I}\left(\frac{1}{\sqrt{q_i}} \cdot \exp\left(-\frac{1}{12nq_i}\right)\right) \cdot \prod_{i\in I} q_i^{-nq_i} = \frac{2^{nH(q_1,\ldots,q_6)}}{(2\pi n)^{(|I|-1)/2}} \cdot \prod_{i\in I}\left(\frac{1}{\sqrt{q_i}} \cdot \exp\left(-\frac{1}{12nq_i}\right)\right).$$

By AM-GM inequality,

$$\prod_{i\in I} q_i \le \left(\frac{1}{|I|}\sum_{i\in I} q_i\right)^{|I|} \le |I|^{-|I|},$$

then

$$\binom{n}{nq_1;\ldots;nq_6} \ge \frac{2^{nH(q_1,\ldots,q_6)}}{(2\pi n)^{(|I|-1)/2}} \cdot \prod_{i\in I}\left(\frac{1}{\sqrt{q_i}}\cdot\exp\left(-\frac{1}{12nq_i}\right)\right) \ge \frac{2^{nH(q_1,\ldots,q_6)}\cdot|I|^{|I|/2}}{(2\pi n)^{(|I|-1)/2}} \cdot \prod_{i\in I}\exp\left(-\frac{1}{12nq_i}\right).$$

On the other hand, since each $nq_i \ge 1$ for $i \in I$, then

$$\begin{aligned}\binom{n}{nq_1;\ldots;nq_6} &\ge \frac{2^{nH(q_1,\ldots,q_6)}\cdot|I|^{|I|/2}}{(2\pi n)^{(|I|-1)/2}} \cdot \prod_{i\in I}\exp\left(-\frac{1}{12nq_i}\right)\\ &\ge \frac{2^{nH(q_1,\ldots,q_6)}\cdot|I|^{|I|/2}}{(2\pi n)^{(|I|-1)/2}} \cdot \exp\left(-\frac{|I|}{12}\right)\\ &= \frac{2^{nH(q_1,\ldots,q_6)}}{n^{(|I|-1)/2}}\cdot\sqrt{2\pi}\cdot\left(\frac{|I|}{2\pi\exp(1/6)}\right)^{|I|/2},\end{aligned}$$

and for $|I| \in [1,6]$ the coefficient

$$\sqrt{2\pi}\cdot\left(\frac{|I|}{2\pi\exp(1/6)}\right)^{|I|/2} \ge \frac{1}{2},$$

which completes the proof. □

By the KL-divergence-driven Chernoff bound in exercise 1.4.4,

$$\Pr\left[\left|Z_k - np_k\right| \ge \varepsilon n\right] \le 2\exp\left(-2n\varepsilon^2\right).$$

Together with lemma 1.10.32,

$$\begin{aligned}\mathbb{E}[B] &\ge \min_{(n_1,\ldots,n_6)\in\mathcal{T}_\varepsilon}\left(\left\lfloor\log_2\binom{n}{n_1;\ldots;n_6}\right\rfloor - 1\right)\cdot\left(1-\sum_{k\in[1,6]}\Pr\left[\left|Z_k-np_k\right|\ge\varepsilon n\right]\right)\\ &\ge \min_{(n_1,\ldots,n_6)\in\mathcal{T}_\varepsilon}\left(nH(q_1,\ldots,q_6) - \frac{5}{2}\log_2 n - 3\right)\cdot\left(1-12\exp\left(-2n\varepsilon^2\right)\right),\end{aligned}$$

where $q_k = n_k/n$. Since H is continuous over $[0,1]^6$, there exists a constant $\gamma_\varepsilon \to 0$ as $\varepsilon \to 0$, such that

$$\begin{aligned}\mathbb{E}[B] &\ge \min_{(n_1,\ldots,n_6)\in\mathcal{T}_\varepsilon}\left(nH(q_1,\ldots,q_6) - \frac{5}{2}\log_2 n - 3\right)\cdot\left(1-12\exp\left(-2n\varepsilon^2\right)\right)\\ &\ge \left(n(h-\gamma_\varepsilon) - \frac{5}{2}\log_2 n - 3\right)\cdot\left(1-12\exp\left(-2n\varepsilon^2\right)\right),\end{aligned}$$

proving the lower bound.

To show an extraction function extracts on average at most nh fair bits, the idea is similar to the theorem 1.10.28. Let X be the random variable of the outcome of n die rolls, then for an input-output pair $(\mathbf{x},\mathbf{y})$ with $|\mathbf{y}| = k$,

$$\Pr[X=\mathbf{x}] \le \Pr[\mathsf{Ext}(X)=\mathbf{y}] \le 2^{-k},$$

the second inequality was discussed in theorem 1.10.28 from the conditional probability and definition 1.10.20, then

$$k = \mathbb{E}[B \mid X=\mathbf{x}] \le -\log_2\Pr[X=\mathbf{x}].$$

By conditional expectation,

$$\mathbb{E}[B] = \sum_{\mathbf{x}}\Pr[X=\mathbf{x}]\cdot\mathbb{E}[B\mid X=\mathbf{x}] \le -\sum_{\mathbf{x}}\Pr[X=\mathbf{x}]\cdot\log_2\Pr[X=\mathbf{x}] = H(X) = nh,$$

where the last equality is from lemma 1.10.7, and we complete the proof. □

1.10.4 Entropy: A Limit of Data Compression

Example 1.10.33. Consider a biased coin that comes up heads with probability 3/4. Let HH map to 0, HT map to 10, TH map to 110, and TT map to 111. Then on average, the number of bits we use for each pair of flips is

$$1 \cdot \frac{9}{16} + 2 \cdot \frac{3}{16} + 3 \cdot \frac{3}{16} + 3 \cdot \frac{1}{16} = \frac{27}{16} < 2.$$

This is an example of compression, as on average a biased coin flip can be represented with less than 1 fair coin flip.

Moreover, it is worth noting that this construction allows for breaking a sequence of biased coin flips into pairs, and concatenating the compression results. The result can be uniquely decoded by simply parsing from left to right. For example, 011110 stands for HHTTHT. The construction preserves the property that the bit representation of any pair of coin flips is not a prefix of the bit representation of any other pair of coin flips.

Representations with this property are called prefix codes.

Exercise 1.10.34 (Exercise 10.15 [MU17])**.** We wish to compress a sequence of i.i.d. random variables $X_1, \ldots$, where each X_j takes on one of n possible values, into a prefix code like example 1.10.33. We map the i^{th} value among all n values to a codeword, which is a sequence of ℓ_i bits. Prove that the ℓ_i must satisfy

$$\sum_{i \in [1,n]} 2^{-\ell_i} \leq 1.$$

Proof. Consider a binary tree of depth $\ell_{\max} = \max\{\ell_i\}_{i \in [1,n]}$, then the length of a codeword is the distance to the root, and a codeword is a prefix of the path to leaves: 0 stands for the left subtree, and 1 stands for the right subtree. For a codeword of length ℓ_i, there are $2^{\ell_{\max} - \ell_i}$ leaves whose root-to-leaf paths begin with that codeword. These sets of leaves are disjoint, since otherwise one codeword would be a prefix of another codeword, contradicting the prefix code property.

Now that every leaf in the binary tree is owned by at most 1 codeword, the number of owned leaves are

$$\sum_{i \in [1,n]} 2^{\ell_{\max} - \ell_i} \leq 2^{\ell_{\max}},$$

which completes the proof. □

Remark 1.10.35. The exercise 1.10.34 is also called Kraft-McMillan inequality [Kra49, McM56].

Definition 1.10.36. A compression function Com takes as input a sequence of n coin flips, given as an element of $\{H, T\}^n$, and outputs a sequence of bits, such that each input sequence of n flips yields distinct output sequences.

We consider the case of compressing the outcome of a sequence of biased coin flips, similar to theorem 1.10.28.

Theorem 1.10.37. *Suppose a coin comes up heads with probability $p > 1/2$. For any constant $\delta > 0$ and n sufficiently large,*

- *There exists a compression function* Com *such that the expected number of output bits on an input sequence of n independent coin flips is at most $(1 + \delta)nH(p)$.*
- *The expected number of output bits by any compression function on an input sequence of n independent coin flips is at least $(1 - \delta)nH(p)$.*

Proof. We begin with the upper bound, with an explicit construction of a compression function.

Let $\varepsilon > 0$ be a sufficiently small constant with $p - \varepsilon > 1/2$. On input a sequence of independent coin flips, output the first bit as a flag bit, 0 when the input has at least $\lceil n(p - \varepsilon) \rceil$ heads, 1 otherwise. When the first flag bit is 1, the compression function translates a head to a 1, and a tail to a 0, which requires $n + 1$ bits to output. When the first flag bit is 0, let the number of coin flip sequences be m, and map a sequence to a value in $[0, m - 1]$.

If there are less than $\lceil n(p-\varepsilon)\rceil$ heads, we upper bound the probability using Chernoff's bound by

$$\Pr\left[X - np < -n\varepsilon\right] \leq \left(\frac{e^{-\gamma}}{(1-\gamma)^{1-\gamma}}\right)^{\mu} \leq \exp\left(-\frac{\mu\gamma^2}{2}\right) = \exp\left(-\frac{n\varepsilon^2}{2p}\right),$$

where $\gamma = \varepsilon/p$, and $\mu = np$.

If there are at least $\lceil n(p-\varepsilon)\rceil$ heads, we upper bound the number of coin flip sequences by

$$\sum_{k\in\left[\lceil n(p-\varepsilon)\rceil, n\right]} \binom{n}{k} \leq \sum_{k\in\left[\lceil n(p-\varepsilon)\rceil, n\right]} \binom{n}{\lceil n(p-\varepsilon)\rceil} \leq \frac{n}{2}\binom{n}{\lceil n(p-\varepsilon)\rceil} \leq \frac{n}{2}\cdot 2^{nH(p-\varepsilon)} = 2^{nH(p-\varepsilon)+\log_2 n - 1} \leq 2^{\lfloor nH(p-\varepsilon)+\log_2 n\rfloor},$$

where the second and the third inequality holds by $p - \varepsilon > 1/2$, and the fourth inequality holds by corollary 1.10.19. Including the flag bit, it takes at most $nH(p-\varepsilon) + \log_2 n + 1$ bits for the sequences with at least $\lceil n(p-\varepsilon)\rceil$ heads.

Let B be the number of output bits of the compression function. The upper bound of its expectation is

$$\begin{aligned}
\mathbb{E}[B] &\leq \exp\left(-\frac{n\varepsilon^2}{2p}\right)\cdot(n+1) + \left(1-\exp\left(-\frac{n\varepsilon^2}{2p}\right)\right)\cdot\left(nH(p-\varepsilon)+\log_2 n + 1\right)\\
&= 1 + \exp\left(-\frac{n\varepsilon^2}{2p}\right)\cdot n + \left(1-\exp\left(-\frac{n\varepsilon^2}{2p}\right)\right)\cdot\left(nH(p)+n\gamma_\varepsilon+\log_2 n\right)\\
&= 1 + nH(p) + \boxed{\exp\left(-\frac{n\varepsilon^2}{2p}\right)\cdot n\left(1-H(p)\right)} + \boxed{\left(1-\exp\left(-\frac{n\varepsilon^2}{2p}\right)\right)\cdot\left(n\gamma_\varepsilon+\log_2 n\right)},
\end{aligned}$$

where the γ_ε is a constant similar to the one in exercise 1.10.31, since H is continuous and monotone over $[1/2, 1]$, $\gamma_\varepsilon \to 0$ as $\varepsilon \to 0$. For ε sufficiently small, when n is sufficiently large, the first boxed term approaches to 0, and the second boxed term is upper bounded by $\delta nH(p)$ for some constant δ.

We now show the lower bound. It suffices to consider a compression function minimizing the expected output number of bits. We observe a fact: if an input string $\mathbf{s}_1$ is more likely than $\mathbf{s}_2$, then the output number of bits on input $\mathbf{s}_2$ should be at least as long as the one of $\mathbf{s}_1$. Otherwise, swapping the two outputs decreases the expectation. Since the probability of deriving a specific sequence with k heads is $p^k(1-p)^{n-k}$, it is more likely for a string to have more heads. Hence, by Chernoff's bound,

$$\Pr\left[X > \lfloor n(p+\varepsilon)\rfloor\right] \leq \Pr\left[X - np > n\varepsilon - 1\right] \leq \left(\frac{e^{\gamma}}{(1+\gamma)^{1+\gamma}}\right)^{\mu} \leq \exp\left(-\frac{\mu\gamma^2}{3}\right) = \exp\left(-\frac{n\left(\varepsilon - 1/n\right)^2}{3p}\right),$$

where $\mu = np$, $\gamma = \varepsilon/p - 1/np$, and ε sufficiently small so that $p + \varepsilon < 1$.

Another fact: for a random variable uniform over $[1, m]$, on average any compression function outputs at least $\log_2 m - 3$ bits. By mapping to $[0, m-1]$, and let ℓ be the first integer that $2+2+4+\ldots+2^{\ell} \geq m$, hence $\ell \geq \log_2 m - 1$. Hence, the average output bit of the compression function is at least $\ell - 2$. Averaged over sequences with $\lfloor n(p+\varepsilon)\rfloor$ heads, the compression function outputs at least

$$\log_2\binom{n}{\lfloor n(p+\varepsilon)\rfloor} - 3 \geq \log_2 \frac{2^{nH(p+\varepsilon)}}{2\sqrt{n}} - 3 = nH(p+\varepsilon) - \frac{1}{2}\log_2 n - 4$$

bits. By the ordering of an optimal compression function, sequences with fewer heads have average output length at least this large. The expected number of output bits of the compression function is lower bounded by ignoring the strings with more than $\lfloor n(p+\varepsilon)\rfloor$ heads, then

$$\left(1-\exp\left(-\frac{n(\varepsilon-1/n)^2}{3p}\right)\right)\cdot\left(nH(p+\varepsilon) - \frac{1}{2}\log_2 n - 4\right).$$

By choosing a sufficiently small ε, and then a sufficiently large n, we can lower bound the expected number of output bits to $(1-\delta)nH(p)$. □

Exercise 1.10.38 (Exercise 10.14 [MU17])**.** Suppose we only have a biased 6-sided die with entropy $h > 0$ instead of a fair coin. Modify the compression function in theorem 1.10.37 so that it compresses a sequence of die rolls to almost nh bits on average.

Proof. We begin with the lower bound by reusing lemma 1.10.32 and the ideas in exercise 1.10.31 and theorem 1.10.37. Let Z_i be the random variable for the number of times the i^{th} face comes up, p_i be the probability that the i^{th} face comes up, and B be the random variable for the bits of the compression outcome. Recall the $\mathcal{T}_\varepsilon$ defined in exercise 1.10.31, by union bound over Chernoff's bound in exercise 1.4.4,

$$\Pr\left[(Z_1,\ldots,Z_6)\in\mathcal{T}_\varepsilon\right] = 1-\Pr\left[(Z_1,\ldots,Z_6)\notin\mathcal{T}_\varepsilon\right] \geq 1-\sum_{i\in[1,6]}\Pr\left[\left|Z_i-np_i\right|\geq n\varepsilon\right] \geq 1-12\exp\left(-2n\varepsilon^2\right).$$

Moreover, by previous discussion in theorem 1.10.37, a compression function minimizing the expected output bits outputs fewer bits on inputs that are more likely than the others, then let $(n_1,\ldots,n_6)$ be the face counts that

$$\prod_{i\in[1,6]} p_i^{n_i} = \max_{(k_1,\ldots,k_6)\in\mathcal{T}_\varepsilon}\prod_{i\in[1,6]} p_i^{k_i},$$

and conditioned on $(k_1,\ldots,k_6)$, the inputs are uniform over $\binom{n}{k_1;\ldots;k_6}$ sequences, hence within the $\mathcal{T}_\varepsilon$, the compression function outputs at least

$$\mathbb{E}\left[B \mid Z_1=k_1,\ldots,Z_6=k_6\right] \geq \log_2\binom{n}{n_1;\ldots;n_6} - 3 \geq \log_2\left(\frac{1}{2n^{5/2}}\cdot 2^{nH(q_1,\ldots,q_6)}\right) - 3 = nH(q_1,\ldots,q_6) - \frac{5}{2}\log_2 n - 4,$$

where the second inequality holds by lemma 1.10.32, and $q_i = n_i/n$. Since this choice maximizes the probability of each individual sequence with these counts, every other sequence in $\mathcal{T}_\varepsilon$ is less likely, so its output length is at least the average output length over this type. Then

$$\begin{aligned}
\mathbb{E}[B] &\geq \sum_{(k_1,\ldots,k_6)\in\mathcal{T}_\varepsilon}\mathbb{E}\left[B \mid Z_1=k_1,\ldots,Z_6=k_6\right]\cdot\Pr\left[Z_1=k_1,\ldots,Z_6=k_6\right]\\
&\geq \left(nH(q_1,\ldots,q_6)-\frac{5}{2}\log_2 n-4\right)\cdot\left(1-12\exp\left(-2n\varepsilon^2\right)\right)\\
&= \left(n(h-\gamma_\varepsilon)-\frac{5}{2}\log_2 n-4\right)\cdot\left(1-12\exp\left(-2n\varepsilon^2\right)\right),
\end{aligned}$$

where γ_ε is a constant such that $\gamma_\varepsilon\to 0$ when $\varepsilon\to 0$ as H is continuous over $[0,1]^6$. Hence, for a sufficiently small ε, and a sufficiently large n, there exists δ such that the lower bound $(1-\delta)nh$ is proved.

Now for the upper bound, we follow the main architecture of theorem 1.10.28, by introducing a flag bit separating the case where the dice rolling sequence is not in $\mathcal{T}_\varepsilon$ and the case where the dice rolling sequence is in $\mathcal{T}_\varepsilon$. If some count deviates too far, then output in its vanilla form where each face is represented in 3 bits. Otherwise, we count the number of possible sequences in $\mathcal{T}_\varepsilon$ to be m, and map a sequence to a value in $[0,m-1]$, and we have

$$m = \sum_{(k_1,\ldots,k_6)\in\mathcal{T}_\varepsilon}\binom{n}{k_1;\ldots;k_6} \leq (2\varepsilon n+1)^5\binom{n}{n_1;\ldots;n_6} \leq (2\varepsilon n+1)^5\cdot 2^{nH(q_1,\ldots,q_6)},$$

where $n_1,\ldots,n_6$ maximizes $\binom{n}{k_1;\ldots;k_6}$ over $\mathcal{T}_\varepsilon$, the last inequality holds by lemma 1.10.32, and $q_i = n_i/n$. Hence,

$$\begin{aligned}
\mathbb{E}[B] &\leq 1+12\exp\left(-2\varepsilon^2 n\right)\cdot 3n+\left(\log_2\left((2\varepsilon n+1)^5\cdot 2^{nH(q_1,\ldots,q_6)}\right)+1\right)\\
&= 2+12\exp\left(-2\varepsilon^2 n\right)\cdot 3n+5\log_2(2\varepsilon n+1)+nH(q_1,\ldots,q_6)\\
&= 2+12\exp\left(-2\varepsilon^2 n\right)\cdot 3n+5\log_2(2\varepsilon n+1)+n(h+\gamma_\varepsilon),
\end{aligned}$$

where γ_ε is a constant such that $\gamma_\varepsilon\to 0$ when $\varepsilon\to 0$ as H is continuous over $[0,1]^6$. Hence, for a sufficiently small ε, and a sufficiently large n, there exists δ such that the upper bound $(1+\delta)nh$ is proved. □

Exercise 1.10.39 (Exercise 10.16 [MU17]). We wish to compress a sequence of i.i.d. random variables $X_1, \ldots$, where each X_j takes on one of n values. The i^{th} value occurs with probability p_i, where $p_1 \geq p_2 \geq \ldots \geq p_n$. The compressed result follows. Let $T_i = \sum_{j \in [1,i-1]} p_j$, and let the i^{th} codeword to be the first $\ell_i = \lceil \log_2(1/p_i) \rceil$ bits of T_i. Show that it is a prefix code example 1.10.33. Let z be the average number of bits used for each X_j. Show that $H(X) \leq z \leq H(X) + 1$.

Proof. We observe that, for the binary representation of p_i, the first positive bit is the ℓ_i^{th} bit. Since the first positive bit is the i^{th} bit for any value in $\left[2^{-i}, 2^{-i+1}\right)$, proving the claim. Moreover, as $p_1 \geq \ldots \geq p_n$, then $\ell_1 \leq \ldots \leq \ell_n$, and it suffices to show that any pair of neighboring adjacent codewords does not violate the prefix code property.

On a pair of T_i and $T_{i+1} = T_i + p_i$, since $T_i \leq T_i + 2^{-\ell_i} \leq T_i + p_i$, the first ℓ_i bits are distinct for the i^{th} and $(i+1)^{\text{th}}$ codeword, proving the prefix code property.

Now for the average number of bits output by the compression function,

$$z = \sum_{i \in [1,n]} p_i \lceil \log_2(1/p_i) \rceil \geq - \sum_{i \in [1,n]} p_i \log_2 p_i = H(X),$$

as $\lceil \log_2(1/p_i) \rceil \geq -\log_2 p_i$. Moreover, by $\lceil \log_2(1/p_i) \rceil \leq -\log_2 p_i + 1$, then

$$z = \sum_{i \in [1,n]} p_i \lceil \log_2(1/p_i) \rceil \leq \sum_{i \in [1,n]} p_i \left(-\log_2 p_i + 1\right) = H(X) + 1,$$

and we proved $H(X) \leq z \leq H(X) + 1$. □

Exercise 1.10.40 (Exercise 10.3 [MU17]). Flip a fair coin repeatedly X times until the first heads occurs. Find $H(X)$.

Now a friend flips the fair coin repeatedly until the first heads occurs. One wants to determine how many flips required, and is allowed to ask a series of yes-no questions of the following form: give the friend a set of integers, and the friend answers "yes" iff the number of flips is in the set, and "no" otherwise. Find a strategy such that the expected number of questions asked before determining the number of flips is $H(X)$. Give an intuitive explanation of why one cannot come up with a strategy that would ask fewer than $H(X)$ questions on average.

Proof. X follows geometrically distributed random variable with success probability $1/2$, then $\mathbb{E}[X] = 2$, and

$$H(X) = -\sum_{i \geq 1} \Pr[X = i] \cdot \log_2 \Pr[X = i] = \sum_{i \geq 1} \frac{1}{2^i} \cdot i = 2.$$

Intuitively, on the k^{th} question, ask if $X = k$, then the number of questions is same distributed as X. For a strategy, it is impossible to have a same terminating transcript of the yes-no answers from distinct values of X. Moreover, a terminating transcript cannot be continued, as it converges to a single value, ruling out other values, namely other suffixes. Hence, they form a prefix code. By exercise 1.10.39, at least $H(X) = 2$ questions are needed on average. □

Exercise 1.10.41 (Exercise 10.17 [MU17]). Arithmetic coding is a standard compression method. In the case where the string to be compressed is a sequence of biased coin flips, it can be described as follows. For a sequence of i.i.d. Bernoulli trials $X_1, \ldots, X_n$ with success probability $1 - p$. The sequences can be ordered lexicographically, so that for $\mathbf{x}, \mathbf{y} \in \{0,1\}^n$, we say $\mathbf{x} < \mathbf{y}$ if $x_i = 0$ and $y_i = 1$ in the first coordinate i that $x_i \neq y_i$. If $z_{\mathbf{x}}$ is the number of zeroes in the string $\mathbf{x}$, define $p(\mathbf{x}) = p^{z_{\mathbf{x}}}(1-p)^{n - z_{\mathbf{x}}}$, and $q(\mathbf{x}) = \sum_{\mathbf{y} < \mathbf{x}} p(\mathbf{y})$.

- Suppose we are given $\mathbf{x}$ sequentially, explain how to compute $q(\mathbf{x})$ in $O(n)$ time.
- Argue that $\left[q(\mathbf{x}), q(\mathbf{x}) + p(\mathbf{x})\right)$ are disjoint subintervals over $[0,1)$, show that $\overline{X} = (X_1, \ldots, X_n)$ can be represented by any point in $\left[q(\overline{X}), q(\overline{X}) + p(\overline{X})\right)$, and a codeword can be chosen in the interval by the first $\left\lceil \log_2(1/p(\overline{X})) \right\rceil + 1$ bits, such that the codeword is a prefix code.
- On a codeword, show how to decompress to determine the corresponding $\overline{X}$.
- Using a Chernoff's bound, argue that $\log_2(1/p(\overline{X}))$ is close to $nH(p)$ with high probability.

Proof. To derive $q(\mathbf{x})$ in $O(n)$, count $i \in [1, n]$: when $x_i = 0$, move forward; otherwise, add $p \prod_{j\in[1,i-1]} q_j$ to $q(\mathbf{x})$, where q_j is the probability corresponding to x_j. Eventually, one sub-hypercube at a time, and

$$q(\mathbf{x}) = p \sum_{i\in[1,n]} x_i \prod_{j\in[1,i-1]} q_j.$$

Let $\mathbf{x}+1$ be defined as adding by 1 to the number with binary representation $\mathbf{x}$, then $q(\mathbf{x})$ is monotonically increasing for each $\mathbf{x} \in [0, 2^n - 1]$, such that $\big[q(\mathbf{x}), q(\mathbf{x}) + p(\mathbf{x})\big)$ are disjoint, as $q(\mathbf{x}+1) = q(\mathbf{x}) + p(\mathbf{x})$, and $q(\mathbf{1}) + p(\mathbf{1}) = 1$.

By exercise 1.10.39, the first positive bit of $p(\mathbf{x})$ is $\ell_{\mathbf{x}} = \lceil \log_2(1/p(\mathbf{x})) \rceil$, then there exists a dyadic interval of length $2^{-(\ell_{\mathbf{x}}+1)}$ in $\big[q(\mathbf{x}), q(\mathbf{x}) + p(\mathbf{x})\big)$, as $2^{-\ell_{\mathbf{x}}} \le p(\mathbf{x})$. The codeword can be chosen by the lower bound of the dyadic interval, requiring $\ell_{\mathbf{x}} + 1$ bits. If a codeword $\mathbf{c}_0$ is a prefix of a codeword $\mathbf{c}_1$, then $\mathbf{c}_1$ in a dyadic interval of $\mathbf{c}_0$, while the exterior intervals are disjoint, which is a contradiction.

Decoding the codeword c is in the reverse order of deriving $q(\mathbf{x})$: Count $i \in [1, n]$, and let the left endpoint $a = 0$. If $c \ge a + p \prod_{j\in[1,i-1]} q_j$, the i^{th} bit is 1, $q_i \leftarrow 1 - p$, and add $p \prod_{j\in[1,i-1]} q_j$ to a; otherwise, the i^{th} bit is 0 and $q_i \leftarrow p$.

Let Z be the number of tails in $[0, n]$, then $\mathbb{E}[Z] = np$, and

$$Y = \log_2(1/p(\overline{X})) = -\log_2 p \cdot Z - \log_2(1-p) \cdot (n - Z) = -n \log_2(1-p) + Z \log_2 \frac{1-p}{p},$$

such that $\mathbb{E}[Y] = nH(p)$. If $p = 1/2$, $Y = nH(p) = n$ deterministically. Otherwise, as $Z = \sum_{i\in[1,n]} Z_i$ for independent Z_i being 1 on the i^{th} flip, we can apply Chernoff's bound, such that

$$\Pr\left[\left|Y - nH(p)\right| \ge \delta np \left| \log_2 \frac{1-p}{p} \right|\right] = \Pr\left[\left|Z - np\right| \ge \delta np\right] \le \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{np} + \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{np} \le 2\exp\left(-\frac{\delta^2}{3} np\right).$$

□

1.10.5 Shannon's Theorem

Consider the following type of channel.

Definition 1.10.42. The input to a binary symmetric channel with parameter p is a sequence of bits $x_1, \ldots$ and the output is a sequence of bits $y_1, \ldots$ such that $\Pr\left[x_i = y_i\right] = 1 - p$ independently for each i.

Consider encoding functions that bring redundancy to help protect against the introduction of errors.

Definition 1.10.43. A (k, n) encoding function $\text{Enc} : \{0,1\}^k \to \{0,1\}^n$ takes as input a sequence of k bits and outputs a sequence of n bits. Conversely, a (k, n) decoding function $\text{Dec} : \{0,1\}^n \to \{0,1\}^k$ takes as input a sequence of n bits and outputs a sequence of k bits.

Exercise 1.10.44 (Exercise 10.18 [MU17])**.** Alice wants to send Bob the result of a fair coin flip over a binary symmetric channel that flips each bit with probability $p < 1/2$. To avoid errors in transmission, she encodes heads as a sequence of $2k + 1$ zeroes and tails as a sequence of $2k + 1$ ones.

- Consider the case where $k = 1$. For each possible received sequence of 3 bits, determine the probability that Alice flipped a heads conditioned on Bob receiving that sequence.
- Bob decodes by examining the 3 bits. If two or three of the bits are 0, Bob decides the corresponding coin flip was heads. Prove that this rule minimizes the probability of error for each flip.
- Argue that, for general k, Bob minimizes the probability of error by deciding the flip was heads if at least $k + 1$ of the bits are 0.
- Give a formula for the probability that Bob makes an error that holds for general k.

Proof. Let $n = 2k+1$, $\mathbf{y} \in \{0,1\}^n$, $\text{nzero}(\mathbf{y})$ be the number of zeroes of the $\mathbf{y}$, X be a fair coin outcome, tails on $X = 0$, otherwise heads, and $\overline{Y} = (Y_1, \ldots, Y_n)$ be the received bits. Hence,

$$\Pr\left[\overline{Y} = \mathbf{y} \,\middle|\, X = 1\right] = (1-p)^{\text{nzero}(\mathbf{y})} \cdot p^{n-\text{nzero}(\mathbf{y})},$$
$$\Pr\left[\overline{Y} = \mathbf{y} \,\middle|\, X = 0\right] = p^{\text{nzero}(\mathbf{y})} \cdot (1-p)^{n-\text{nzero}(\mathbf{y})},$$

and by Bayes' Rule,

$$\Pr\left[X = 1 \,\middle|\, \overline{Y} = \mathbf{y}\right] = \frac{(1-p)^{\text{nzero}(\mathbf{y})} \cdot p^{n-\text{nzero}(\mathbf{y})}}{(1-p)^{\text{nzero}(\mathbf{y})} \cdot p^{n-\text{nzero}(\mathbf{y})} + p^{\text{nzero}(\mathbf{y})} \cdot (1-p)^{n-\text{nzero}(\mathbf{y})}}.$$

Consider decoding the flip: if at least m bits are 0, then we have heads; otherwise we have tails. Then

$$\begin{aligned}
\Pr[\text{error}] &= \Pr\left[X = 0 \wedge \text{nzero}\left(\overline{Y}\right) \geq m\right] + \Pr\left[X = 1 \wedge \text{nzero}\left(\overline{Y}\right) < m\right] \\
&= \frac{1}{2} \sum_{i \in [m,n]} \binom{n}{i} \cdot p^i \cdot (1-p)^{n-i} + \frac{1}{2} \sum_{i \in [0,m-1]} \binom{n}{i} \cdot (1-p)^i \cdot p^{n-i} \\
&= \frac{1}{2} + \frac{1}{2} \sum_{i \in [m,n]} \binom{n}{i} \cdot \left(p^i \cdot (1-p)^{n-i} - (1-p)^i \cdot p^{n-i}\right).
\end{aligned}$$

Since $p < 1/2$, then $p^i \cdot (1-p)^{n-i} < (1-p)^i \cdot p^{n-i}$ for $i \geq k+1$, which proves the error minimizing when $m = k+1$. □

Lemma 1.10.45. *For a binary symmetric channel with parameter $p < 1/2$ and for any constant $\delta, \gamma > 0$, when n is sufficiently large, for any $k \leq n(1 - H(p) - \delta)$ and a uniformly distributed k-bit input message, there exist (k, n) encoding and decoding functions such that the probability the receiver fails to obtain the correct message is at most γ.*

Proof. We begin by describing a probabilistic algorithm constructing 2^k codewords in $\{0,1\}^n$, one for each message, and the inefficient encoding and decoding functions.

Let $\{X_i\}_{i \in [0, 2^k - 1]}$ be the distinct codewords sampled uniformly at random from $\{0,1\}^n$. The encoding function $\text{Enc} : \{0,1\}^k \to \{0,1\}^n$ is a lookup table mapping the k-bit inputs to their corresponding codewords. The decoding function $\text{Dec} : \{0,1\}^n \to \{0,1\}^k \cup \{\perp\}$ maintains the look up table, iterates through the codewords, and checks if a codeword differing from the received n bits in between $\left[\left\lceil n(p-\varepsilon)\right\rceil, \left\lfloor n(p+\varepsilon)\right\rfloor\right]$ bits. If there is only one codeword in this interval, output the corresponding k-bit message; otherwise, decoding function fails and outputs $\perp$. Since the decoding runtime is exponential in k, the decoding is not efficient, and we are not requiring it to be efficient.

When it comes to decoding failure or failure to obtain the correct message, it falls into 2 cases:

- The noisy channel made less than $n(p-\varepsilon)$ or more than $n(p+\varepsilon)$ errors.
- There are other codeword(s) that are differing from the received n bits in between $\left[\left\lceil n(p-\varepsilon)\right\rceil, \left\lfloor n(p+\varepsilon)\right\rfloor\right]$ bits.

The goal is to prove the existence of a set of codewords, whose probability of failures above is at most γ. The strategy here is probabilistic method, via averaging argument lemma 1.6.12, if probability of failures above on average among all $\{X_i\}_{i \in [0, 2^k - 1]}$ is at most γ, then there must exist a set of codewords whose failure probability is at most γ. The first failure event can be upper bounded via concentration inequalities, while the second failure event can be upper bounded via counting argument.

We continue by introducing notation and random variable formulations. Let $\Delta(\mathbf{s}_0, \mathbf{s}_1)$ be the Hamming weight of $\mathbf{s}_0 \oplus \mathbf{s}_1$ for the number of positions $\mathbf{s}_0$ and $\mathbf{s}_1$ differing from each other. We say $(\mathbf{s}_0, \mathbf{s}_1)$ pair has weight

$$w(\mathbf{s}_0, \mathbf{s}_1) = p^{\Delta(\mathbf{s}_0, \mathbf{s}_1)} \cdot (1-p)^{n - \Delta(\mathbf{s}_0, \mathbf{s}_1)},$$

corresponding to the probability of $\mathbf{s}_0$ being altered into $\mathbf{s}_1$ by the binary symmetric channel. Let $\{S_i\}_{i\in[0,2^k-1]}$ be the sets of n bits that decodes to X_i, which are n bits that are close to X_i, and are far from other X_j. Let $\{W_i\}_{i\in[0,2^k-1]}$ be the set of probabilities that X_i fails in decoding correctly, where

$$W_i = \sum_{\mathbf{r}\notin S_i} w(X_i, \mathbf{r}).$$

It can also be expressed in a way that, let $1_{i,\mathbf{r}}$ be an indicator $0/1$ random variable that is 1 iff $\mathbf{r} \notin S_i$, then

$$W_i = \sum_{\mathbf{r}\in\{0,1\}^n} 1_{i,\mathbf{r}} \cdot w(X_i, \mathbf{r}).$$

Note that, both $\{S_i\}_{i\in[0,2^k-1]}$ and $\{W_i\}_{i\in[0,2^k-1]}$ are *random variables* dependent on $\{X_i\}_{i\in[0,2^k-1]}$.

WLOG, we analyze with respect to X_0, as by symmetry the following results applies to other X_i. Write

$$W_0 = \sum_{\mathbf{r}\in T_1} 1_{0,\mathbf{r}} \cdot w(X_0, \mathbf{r}) + \sum_{\mathbf{r}\in T_2} 1_{0,\mathbf{r}} \cdot w(X_0, \mathbf{r}), \tag{1.13}$$

where $T_1 = \{\mathbf{s} \;:\; \left|\Delta(X_0, \mathbf{s}) - np\right| > n\varepsilon\}$, and $T_2 = \{\mathbf{s} \;:\; \left|\Delta(X_0, \mathbf{s}) - np\right| \le n\varepsilon\}$.

For the first term, let $\{Z_i\}_{i\in[1,n]}$ be independent Bernoulli trials with success probability p, then $Z = \sum_{i\in[1,n]} Z_i$ is the random variable for the number of bits flipped by the channel, and by Chernoff's bound, for $0 < \varepsilon < p$,

$$\sum_{\mathbf{r}\in T_1} 1_{0,\mathbf{r}} \cdot w(X_0, \mathbf{r}) = \sum_{\mathbf{r}\in T_1} w(X_0, \mathbf{r}) = \Pr\left[\left|Z - np\right| > n\varepsilon\right] \le \left(\frac{e^{\theta}}{(1+\theta)^{1+\theta}}\right)^{np} + \left(\frac{e^{-\theta}}{(1-\theta)^{1-\theta}}\right)^{np} \le 2\exp\left(-\frac{\mu\theta^2}{3}\right), \tag{1.14}$$

where $\theta = \varepsilon/p$ and $\mu = np$. For any threshold $0 < \varepsilon < p$, there is n sufficiently large such that $2\exp\left(-\mu\theta^2/3\right) \le \gamma/2$.

For the second term, consider the following symmetry: On a set of codewords $\{X_i\}_{i\in[0,2^k-1]}$, we have $\{X'_i\}_{i\in[0,2^k-1]}$ such that $X'_i \leftarrow X_i \oplus X_0$, and $X'_0 = \mathbf{0}$. Then $\{S'_i\}_{i\in[0,2^k-1]}$ can be obtained by XORing X_0 with elements in $\{S_i\}_{i\in[0,2^k-1]}$, and the W_0 can be expressed by

$$W_0 = \sum_{\mathbf{r}\in T'_1} 1_{0,\mathbf{r}} \cdot w(\mathbf{0}, \mathbf{r}) + \sum_{\mathbf{r}\in T'_2} 1_{0,\mathbf{r}} \cdot w(\mathbf{0}, \mathbf{r}).$$

where $T'_1 = \{\mathbf{s} \;:\; \left|\Delta(\mathbf{0}, \mathbf{s}) - np\right| > n\varepsilon\}$, and $T'_2 = \{\mathbf{s} \;:\; \left|\Delta(\mathbf{0}, \mathbf{s}) - np\right| \le n\varepsilon\}$. The symmetry holds for the claim in the first term, while the symmetry allows us to argue the second term with respect to $X'_0 = \mathbf{0}$, then

$$\sum_{\mathbf{r}\in T'_2} 1_{0,\mathbf{r}} \cdot w(\mathbf{0}, \mathbf{r})$$

is a random variable depending on $\{X'_i\}_{i\in[1,2^k-1]}$. Moreover, for a $\mathbf{r} \in T'_2$, averaging over $\{X'_i\}_{i\in[1,2^k-1]}$,

$$\mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(\mathbf{0}, \mathbf{r})\right] = w(\mathbf{0}, \mathbf{r}) \cdot \Pr\left[\exists i \in [1, 2^k - 1], \left|\Delta(X'_i, \mathbf{r}) - np\right| \le n\varepsilon\right].$$

Supposing we have

$$\Pr\left[\exists i \in [1, 2^k - 1], \left|\Delta(X'_i, \mathbf{r}) - np\right| \le n\varepsilon\right] \le \frac{\gamma}{2},$$

then the second term can be upper bounded by

$$\sum_{\mathbf{r}\in T'_2} \mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(\mathbf{0}, \mathbf{r})\right] \le \sum_{\mathbf{r}\in T'_2} w(\mathbf{0}, \mathbf{r}) \cdot \frac{\gamma}{2} \le \frac{\gamma}{2}.$$

We continue by upper bounding the number of possible n bits around $\mathbf{r}$, such that we upper bound the second term:

$$\sum_{k\in\left[\left\lceil n(p-\varepsilon)\right\rceil, \left\lfloor n(p+\varepsilon)\right\rfloor\right]} \binom{n}{k} \le (2\varepsilon n + 1) \cdot \binom{n}{\left\lceil n(p+\varepsilon)\right\rceil} \le (2\varepsilon n + 1) \cdot 2^{nH(p+\varepsilon)},$$

threshold ε is chosen such that $\lceil n(p+\varepsilon) \rceil \leq n/2$, and the second inequality holds by corollary 1.10.19. The probability of a particular codeword X_i' with $i > 0$ having a Hamming distance $\left[\lceil n(p-\varepsilon) \rceil, \lfloor n(p+\varepsilon) \rfloor\right]$ to $\mathbf{r}$ causing a decoding failure, when $\mathbf{0}$ is sent is at most

$$(2\varepsilon n + 1) \cdot 2^{n(H(p+\varepsilon)-1)} = (2\varepsilon n + 1) \cdot 2^{n(H(p)+\gamma_\varepsilon - 1)},$$

where γ_ε is a constant similar to the one appearing in theorem 1.10.28 and theorem 1.10.37. Union bounding over all other codewords, the probability of other codeword failing decoding $\mathbf{0}$ is upper bounded by

$$\sum_{\mathbf{r} \in T_2'} \mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(\mathbf{0}, \mathbf{r})\right] \leq (2^k - 1) \cdot (2\varepsilon n + 1) \cdot 2^{n(H(p)+\gamma_\varepsilon - 1)} \leq (2\varepsilon n + 1) \cdot 2^{n(\gamma_\varepsilon - \delta)}, \tag{1.15}$$

where the last inequality holds by $k \leq n(1 - H(p) - \delta)$. Since eq. (1.15) approaches 0 as $n \to \infty$ with $\delta > \gamma_\varepsilon$, then it is at most $\gamma/2$ on a sufficiently large n, and by symmetry, we conclude

$$\sum_{\mathbf{r} \in T_2} \mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(X_0, \mathbf{r})\right] \leq \frac{\gamma}{2}. \tag{1.16}$$

Combining eqs. (1.13), (1.14) and (1.16), on a sufficiently large n, the probability of decoding incorrectly X_0 is

$$\mathbb{E}[W_0] = \sum_{\mathbf{r} \in \{0,1\}^n} \mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(X_0, \mathbf{r})\right] \leq \gamma.$$

Since the k-bit message is uniform over $\{0,1\}^k$, the probability of decoding incorrectly is averaging over all $\mathbb{E}[W_i]$ by

$$2^{-k} \sum_{i \in [0, 2^k - 1]} \mathbb{E}[W_i] \leq \gamma.$$

By an averaging argument lemma 1.6.12, there must exist a set of codewords $\{\mathbf{x}_i\}_{i \in [0, 2^k-1]}$, such that

$$\sum_{i \in [0, 2^k - 1]} (W_i \mid X_0 = \mathbf{x}_0, \ldots, X_{2^k-1} = \mathbf{x}_{2^k-1}) \leq 2^k \gamma.$$

□

The decoding failure probability in lemma 1.10.45 is obtained by averaging over all k-bit messages and all possible codeword samplings. We will obtain a stronger result than lemma 1.10.45 in theorem 1.10.46: for all $i \in [0, 2^k - 1]$,

$$(W_i \mid X_0 = \mathbf{x}_0, \ldots, X_{2^k-1} = \mathbf{x}_{2^k-1}) \leq \gamma$$

holds simultaneously.

Theorem 1.10.46 (Shannon's Theorem [Sha48a, Sha48b])**.** *For a binary symmetric channel with parameter $p < 1/2$ and for any constants $\delta, \gamma > 0$, when n is sufficiently large:*

- *for any $k \leq n(1 - H(p) - \delta)$, there exist (k, n) encoding and decoding functions such that the probability the receiver fails to obtain the correct message is at most γ for every k-bit input message; and*
- *there are no (k, n) encoding and decoding functions with $k \geq n(1 - H(p) + \delta)$ such that the probability decoding correctly is at least γ for a k-bit input message chosen uniformly at random.*

Proof. WLOG we let $\{\mathbf{x}_i\}_{i \in [0, 2^k-1]}$ be sorted in increasing order of W_i. Hence, for $i \in [0, 2^{k-1} - 1]$, each $\mathbf{x}_i$ has $W_i \leq 2\gamma$, or otherwise we contradict by

$$\sum_{i \in [2^{k-1}, 2^k - 1]} (W_i \mid X_0 = \mathbf{x}_0, \ldots, X_{2^k-1} = \mathbf{x}_{2^k-1}) \geq 2\gamma \cdot 2^{k-1} = 2^k \gamma.$$

In this way, we prove there exist encoding and decoding functions for $k-1$ bits messages over 2^{k-1} codewords, and the probability of each codeword being decoded incorrectly is simultaneously at most 2γ, when $k \leq n(1-H(p)-\delta)$. Since δ and γ can be any constant, the first part of the proof is completed.

Having the first part of the proof finished, we move on to the second part of the proof, which is the converse of lemma 1.10.45. We begin by an observation in lemma 1.10.45 to build an intuition: The decoding function maps the received n bits to the only one codeword within $\left[\left\lceil n(p-\varepsilon)\right\rceil, \left\lfloor n(p+\varepsilon)\right\rfloor\right]$ many errors, and the number of such n bits is lower bounded by

$$\sum_{k\in\left[\left\lceil n(p-\varepsilon)\right\rceil,\left\lfloor n(p+\varepsilon)\right\rfloor\right]} \binom{n}{k} \geq \binom{n}{\lceil np\rceil} \geq \frac{2^{nH(p)}}{2\sqrt{n}},$$

where the last inequality holds by corollary 1.10.19. Since there are 2^k possible messages, and $k \geq n(1-H(p)+\delta)$, on a sufficiently large n,

$$2^k \cdot \frac{2^{nH(p)}}{2\sqrt{n}} \geq \frac{2^{n\delta}}{2\sqrt{n}} \cdot 2^n \geq 2^n.$$

Hence, given $k \geq n(1-H(p)+\delta)$ and n sufficiently large, conditioned that the number of bits flipped by the channel is within $\left[\left\lceil n(p-\varepsilon)\right\rceil, \left\lfloor n(p+\varepsilon)\right\rfloor\right]$, there is no such set of codewords always decode successfully, namely it is getting too dense, such that in the "unique decoding" range defined in lemma 1.10.45, there are more than 1 valid codeword.

More specifically, we can show how to rule out lemma 1.10.45. Let C_i be the set of received n-bit strings that are correctly decoded to the codeword $\mathbf{x}_i$ in lemma 1.10.45, then the probability of successful decoding is

$$\frac{1}{2^k} \sum_{i\in[0,2^k-1]} \sum_{\mathbf{r}\in C_i} w(\mathbf{x}_i, \mathbf{r})$$

by averaging over all 2^k messages. Since for any $\mathbf{r} \in C_i$, it is within $\left[\left\lceil n(p-\varepsilon)\right\rceil, \left\lfloor n(p+\varepsilon)\right\rfloor\right]$ many bit flips from $\mathbf{x}_i$,

$$w(\mathbf{x}_i, \mathbf{r}) = p^{\Delta(\mathbf{x}_i,\mathbf{r})} \cdot (1-p)^{n-\Delta(\mathbf{x}_i,\mathbf{r})} \leq p^{n(p-\varepsilon)} \cdot (1-p)^{n(1-p+\varepsilon)} = \left(\frac{1-p}{p}\right)^{n\varepsilon} \cdot 2^{-nH(p)},$$

then the averaging probability is upper bounded by

$$\begin{aligned} \frac{1}{2^k} \sum_{i\in[0,2^k-1]} \sum_{\mathbf{r}\in C_i} w(\mathbf{x}_i, \mathbf{r}) &\leq \frac{1}{2^k} \sum_{i\in[0,2^k-1]} \left(\frac{1-p}{p}\right)^{n\varepsilon} \cdot 2^{-nH(p)} \cdot |C_i| \\ &\leq \frac{1}{2^k} \cdot \left(\frac{1-p}{p}\right)^{n\varepsilon} \cdot 2^{-nH(p)} \cdot 2^n = \left(\left(\frac{1-p}{p}\right)^{\varepsilon} \cdot 2^{-\delta}\right)^n, \end{aligned} \tag{1.17}$$

where the last inequality holds by $k \geq n(1-H(p)+\delta)$, so conditioned that

$$\left(\frac{1-p}{p}\right)^{\varepsilon} \cdot 2^{-\delta} < 1, \tag{1.18}$$

then when n is sufficiently large, the probability is upper bounded by γ.

We proceed by analyzing the general case. For the set of codewords $\{\mathbf{x}_i\}_{i\in[0,2^k-1]}$, let S_i be the set of n bits that are uniquely decoded to $\mathbf{x}_i$, and $S_{i,1} \cap S_{i,2} = \varnothing$, $S_i = S_{i,1} \cup S_{i,2}$ be defined as follows: $S_{i,1} = \{\mathbf{r} \in S_i \ : \ \left|\Delta(\mathbf{x}_i, \mathbf{r}) - pn\right| > \varepsilon n\}$, and $S_{i,2} = \{\mathbf{r} \in S_i \ : \ \left|\Delta(\mathbf{x}_i, \mathbf{r}) - pn\right| \leq \varepsilon n\}$. Then the probability of successful decoding is

$$\frac{1}{2^k} \sum_{i\in[0,2^k-1]} \sum_{\mathbf{r}\in S_i} w(\mathbf{x}_i, \mathbf{r}) = \frac{1}{2^k} \sum_{i\in[0,2^k-1]} \sum_{\mathbf{r}\in S_{i,1}} w(\mathbf{x}_i, \mathbf{r}) + \frac{1}{2^k} \sum_{i\in[0,2^k-1]} \sum_{\mathbf{r}\in S_{i,2}} w(\mathbf{x}_i, \mathbf{r}),$$

where the first term is upper bounded by Chernoff's bound with the same bound in eq. (1.14), and the second term is upper bounded in the same procedure of eq. (1.17), conditioned that eq. (1.18) holds. □

Exercise 1.10.47 (Exercise 10.19 [MU17]). Consider the following channel. The sender can send a symbol from $[0, 4]$, the channel introduces errros: when the symbol k is sent, the recipient receives $k + 1 \mod 5$ with probability $1/2$, or $k - 1 \mod 5$ with probability $1/2$. The errors on each symbol are mutually independent of each other.

Define the encoding and decoding functions for this channel by: A (j, n) encoding function maps a number in $[0, j - 1]$ into a sequence in $[0, 4]^n$. A (j, n) decoding function maps a sequence in $[0, 4]^n$ into a number in $[0, j - 1]$.

There are $(1, 1)$ encoding and decoding functions with zero probability of error, as the decoder maps 1 and 4 to 0, and maps 0 and 2 to 1. Hence at least 1 bit can be sent without error per channel use. Show that:

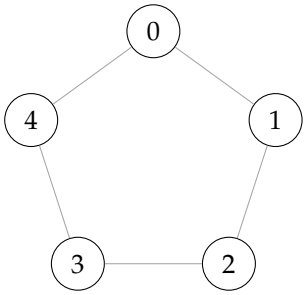
- There are $(5, 2)$ encoding and decoding functions with zero probability of error. Argue that more than one bit of information can be sent per use of the channel.
- If there exists (j, n) encoding and decoding functions with zero probability of error, then $n \geq \log_2 j/(\log_2 5 - 1)$.

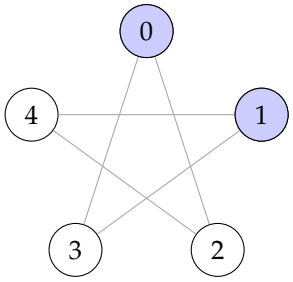
Proof. Consider the $(5, 2)$ encoding and decoding function. The $(5, 2)$ encoder encodes a with a pair $(a, 2a \mod 5)$. The $(5, 2)$ decoder performs by the following lookup table

encode	received	decode to
$(0, 0)$	$(1, 1), (1, 4), (4, 1), (4, 4)$	0
$(1, 2)$	$(2, 3), (2, 1), (0, 3), (0, 1)$	1
$(2, 4)$	$(3, 0), (3, 3), (1, 0), (1, 3)$	2
$(3, 1)$	$(4, 2), (4, 0), (2, 2), (2, 0)$	3
$(4, 3)$	$(0, 4), (0, 2), (3, 4), (3, 2)$	4

The encoding scheme can also be instantiated by $(a, 3a \mod 5)$ [9].

In fact, the channel operates over a cycle graph C_5, mapping a to $a \pm 1 \mod 5$ each with probability $1/2$. Moreover, since we want a set of codewords avoiding overlapping received symbols on the decoding side, we can construct the following graph for the single coordinate $(1, 1)$ encoding and decoding, by pairing up symbols bringing "confusion" after transmitted through the channel. Hence, $\{0, 1\}$ is an independent set of size 2 over the C_5 illustrated as folows, and it happens to be the codewords for $(1, 1)$ encoding and decoding.



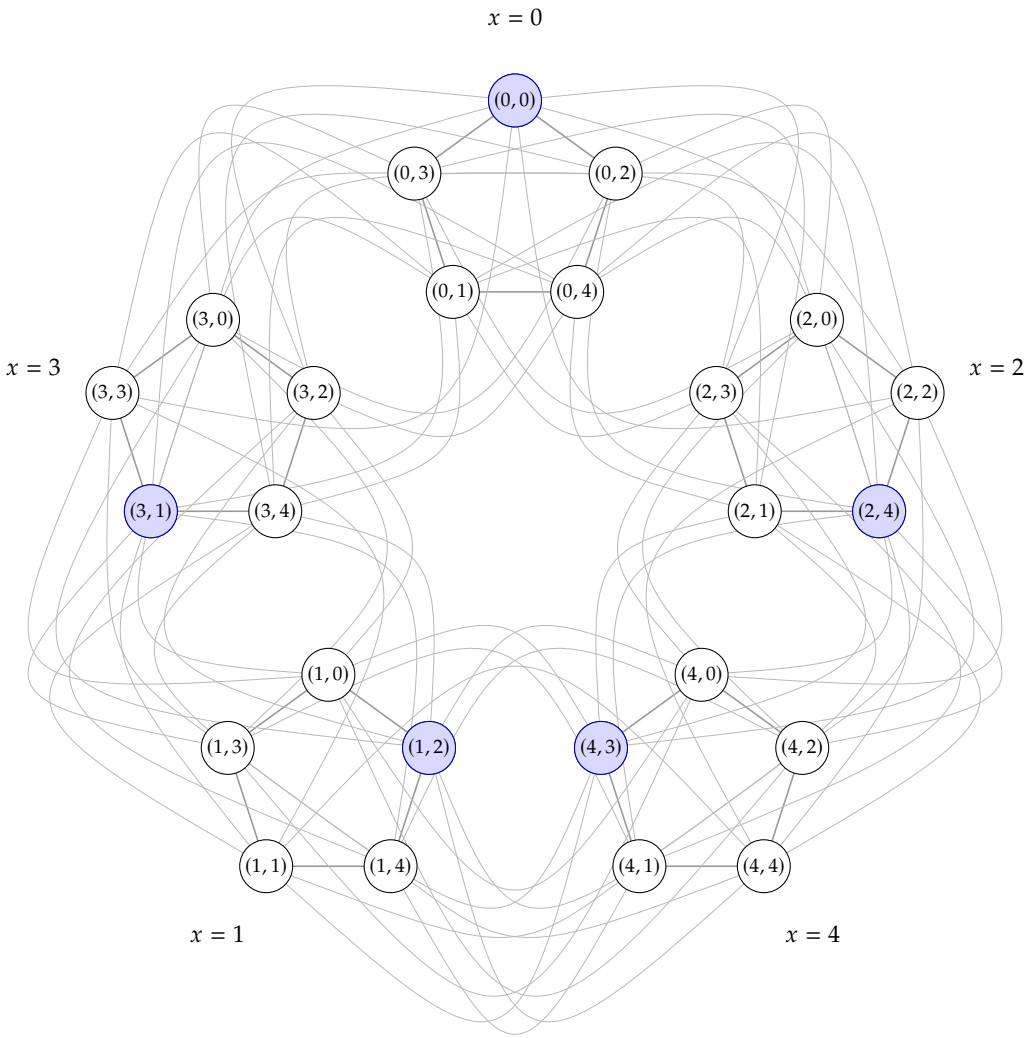


On a pair of C_5 confusion graphs, we want to define a graph product such that any neighboring (i, j) brings overlaps on the decoding side. On a pair of (u, v) and (u', v'), they are adjacent iff

- $u = u'$, v and v' are adjacent on C_5 confusion graph;
- $v = v'$, u and u' are adjacent on C_5 confusion graph;
- u and u', v and v' are both adjacent on C_5 confusion graph.

We then define the symbol $\boxtimes$ such that $C_5 \boxtimes C_5$ is illustrated as follows:

[9] The story is, I brute-forced adding constant, additive inverse, multiplicative inverse, and finally scalar multiplication.



The new codewords form an independent set of size 5 on $C_5 \boxtimes C_5$, each C_5 is ordered as $0, 2, 4, 1, 3$ so that adjacency corresponds to difference $\pm 2 \mod 5$.

When it comes to general (j, n) encoding and decoding functions, supposing it exists, then there must be $j \cdot 2^n \leq 5^n$, such that for each of the j values, the decoding boundary is 2^n, and the total area of decoding is no more than 5^n, or confusion must exists, and hence

$$n \geq \frac{\log_2 j}{\log_2 5 - 1}.$$

□

Remark 1.10.48. The exercise 1.10.47 is the classical zero-error capacity viewpoint of Shannon [Sha56]. The pentagon case C_5 was solved by Lovász [Lov79].

Exercise 1.10.49 (Exercise 10.20 [MU17])**.** A binary erasure channel transfers a sequences of n bits. Each bit either arrives successfully without error or fails to arrive, and is replaced with a "?" symbol. Failure occur independently with probability p. Define (k, n) encoding and decoding functions for BEC similar to ones in the BSC, except that $\mathrm{Dec} : \{0, 1, ?\}^n \to \{0, 1\}^k \cup \{\perp\}$. Prove that, for any $p > 0$ and any constants $\delta, \gamma > 0$, if n is sufficiently large, then there exists (k, n) encoding and decoding functions with $k \leq n(1 - p - \delta)$ such that the probability that the receiver fails to obtain the correct message is at most γ for every possible k-bit message.

Proof. The proof architecture is similar to lemma 1.10.45 and theorem 1.10.46, with a same setup of the randomized codewords, and the same encoding function. The decoder finds a codeword that has within $\left[\left\lceil n(p-\varepsilon)\right\rceil, \left\lfloor n(p+\varepsilon)\right\rfloor\right]$ number of "?", and the other symbols must match. If there is one such codeword, output the corresponding message; otherwise, reject by $\perp$.

We reuse a few notations in lemma 1.10.45: Applying Hamming weight Δ over $\mathbf{x}_i \in \{0,1\}^n$ and $\mathbf{s}_i \leftarrow \mathbf{x}_i \oplus \mathbf{r}_i$ where $\mathbf{r}_i \in \{0,?\}^n$, and $\oplus$ overrides ? to a symbol in $\{0,1\}$. Then the weight defined in between $\mathbf{x}_i$ and $\mathbf{s}_i$ is

$$w(\mathbf{x}_i, \mathbf{s}_i) = p^{\Delta(\mathbf{x}_i,\mathbf{s}_i)} \cdot (1-p)^{n-\Delta(\mathbf{x}_i,\mathbf{s}_i)}.$$

Let $\{S_i\}_{i\in[0,2^k-1]}$ be the set of $\{0,1,?\}^n$ decodable to X_i, $\{W_i\}_{i\in[0,2^k-1]}$ be the set of probabilities that X_i fails in decoding correctly, where

$$W_i = \sum_{\mathbf{r}\in\{0,?\}^n, X_i\oplus\mathbf{r}\notin S_i} w(X_i, X_i\oplus\mathbf{r}).$$

It can also be expressed in the way that, let $1_{i,\mathbf{r}}$ be an indicator 0/1 random variable that is 1 iff $X_i \oplus \mathbf{r} \notin S_i$, then

$$W_i = \sum_{\mathbf{r}\in\{0,?\}^n} 1_{i,\mathbf{r}} \cdot w(X_i, X_i\oplus\mathbf{r}) = \sum_{\mathbf{r}\in T_1} 1_{i,\mathbf{r}} \cdot w(X_i, X_i\oplus\mathbf{r}) + \sum_{\mathbf{r}\in T_2} 1_{i,\mathbf{r}} \cdot w(X_i, X_i\oplus\mathbf{r}),$$

by letting $T_1 = \{\mathbf{r}\in\{0,?\}^n \ : \ \left|\Delta(\mathbf{0},\mathbf{r}) - np\right| > n\varepsilon\}$ and $T_2 = \{\mathbf{r}\in\{0,?\}^n \ : \ \left|\Delta(\mathbf{0},\mathbf{r}) - np\right| \le n\varepsilon\}$.

WLOG we analyze X_0, as by symmetry the same analysis hold for all other X_i.

The first term is bounded in the same way as eq. (1.14) via Chernoff's bound. Let $\{Z_i\}_{i\in[1,n]}$ be the independent Bernoulli trials with success probability p, then $Z = \sum_{i\in[1,n]} Z_i$ is the number "?"s, for $0 < \varepsilon < p$,

$$\sum_{\mathbf{r}\in T_1} 1_{0,\mathbf{r}} \cdot w(X_0, X_0\oplus\mathbf{r}) = \sum_{\mathbf{r}\in T_1} w(X_0, X_0\oplus\mathbf{r}) = \Pr\left[\left|Z-np\right| > n\varepsilon\right] \le \left(\frac{e^\theta}{(1+\theta)^{1+\theta}}\right)^{np} + \left(\frac{e^{-\theta}}{(1-\theta)^{1-\theta}}\right)^{np} \le 2\exp\left(-\frac{\mu\theta^2}{3}\right),$$

For any threshold $0 < \varepsilon < p$, there is n sufficiently large such that $2\exp\left(-\mu\theta^2/3\right) \le \gamma/2$.

It suffices to argue with respect to $X_0 = \mathbf{0}$ for the second term by the same symmetry discussed in lemma 1.10.45, then the random variable

$$\sum_{\mathbf{r}\in T_2} 1_{0,\mathbf{r}} \cdot w(\mathbf{0},\mathbf{r})$$

is dependent on $\{X_i\}_{i\in[1,2^k-1]}$. For $\mathbf{r}\in T_2$, averaging over $\{X_i\}_{i\in[1,2^k-1]}$,

$$\mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(\mathbf{0},\mathbf{r})\right] = w(\mathbf{0},\mathbf{r}) \cdot \Pr\left[\exists i\in[1,2^k-1], \left|\Delta(X_i,\mathbf{r}) - np\right| \le n\varepsilon\right].$$

Supposing we have

$$\Pr\left[\exists i\in[1,2^k-1], \left|\Delta(X_i,\mathbf{r}) - np\right| \le n\varepsilon\right] \le \frac{\gamma}{2},$$

then the second term can be upper bounded by

$$\sum_{\mathbf{r}\in T_2} \mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(\mathbf{0},\mathbf{r})\right] \le \sum_{\mathbf{r}\in T_2} w(\mathbf{0},\mathbf{r}) \cdot \frac{\gamma}{2} \le \frac{\gamma}{2}.$$

We continue by upper bounding the number of possible n bits to derive $\mathbf{r}$ by $2^{\Delta(\mathbf{0},\mathbf{r})} \le 2^{n(p+\varepsilon)}$, then the probability of a particular codeword X_i with $i > 0$ being able to produce $\mathbf{r}$ is at most $2^{n(p+\varepsilon-1)}$. Union bounding over all other codewords, the probability of other codeword failing decoding $\mathbf{0}$ is upper bounded by

$$(2^k-1)\cdot 2^{n(p+\varepsilon-1)} \le 2^k \cdot 2^{n(p+\varepsilon-1)} \le 2^{n(\varepsilon-\delta)},$$

where the last inequality is by $k \le n(1-p-\delta)$. Since $2^{n(\varepsilon-\delta)} \to 0$ as $n\to\infty$ if $\delta > \varepsilon$, then it is at most $\gamma/2$ on an n. By symmetry, we conclude

$$\sum_{\mathbf{r}\in T_2} \mathbb{E}\left[1_{0,\mathbf{r}} \cdot w(X_0, X_0\oplus\mathbf{r})\right] \le \frac{\gamma}{2}.$$

Combining the separated cases, we have

$$\mathbb{E}[W_0] = \sum_{\mathbf{r}\in T_1} \mathbb{E}[1_{0,\mathbf{r}} \cdot w(X_0, X_0 \oplus \mathbf{r})] + \sum_{\mathbf{r}\in T_2} \mathbb{E}[1_{0,\mathbf{r}} \cdot w(X_0, X_0 \oplus \mathbf{r})] \leq \gamma.$$

By symmetry and linearity of expectation,

$$\sum_{i\in[0,2^k-1]} \mathbb{E}[W_i] \leq 2^k\gamma.$$

By the same ranking argument in theorem 1.10.46, there exists a set of codewords with at least 2^{k-1} codewords with $W_i \leq 2\gamma$, and we choose the half of the set of codewords with small W_i, and we complete the proof. □

Exercise 1.10.50 (Exercise 10.21 [MU17])**.** In lemma 1.10.45 and theorem 1.10.46, we let decoder find a codeword with Hamming distance to the bits received within $\lceil n(p-\varepsilon), n(p+\varepsilon)\rceil$. Instead, let decoder find the codeword by looking for the codeword with least number of differences, and break ties arbitrarily. Show how to modify the proof for lemma 1.10.45 and theorem 1.10.46 to obtain a similar result.

Proof. Just use the Hamming ball of radius $\lfloor n(p+\varepsilon)\rfloor$ around the codeword. The "too far" Chernoff case still decays exponentionally fast, and the "in radius" case should approach 0 as $n \to \infty$. □

1.11 The Monte Carlo Method

1.11.1 The Monte Carlo Method

1.11.2 DNF Counting

1.11.3 From Approximate Sampling to Approximate Counting

1.11.4 The Markov Chain Monte Carlo Method (MCMC)

TODO: ...

1.12 Coupling of Markov Chains

The motivation of coupling is to decide the rate of convergence to an ergodic Markov chain's stationary distribution.

1.12.1 Variation Distance and Mixing Time

Definition 1.12.1 (Variation Distance)**.** The variation distance between two distributions $\mathcal{D}_1$ and $\mathcal{D}_2$ on a countable state space S is given by

$$d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) = \frac{1}{2}\sum_{x\in S} |\mathcal{D}_1(x) - \mathcal{D}_2(x)| .$$

The factor $1/2$ in definition 1.12.1 guarantees that the variation distance is in $[0, 1]$.

Lemma 1.12.2. *For any $A \subseteq S$, let $\mathcal{D}(A) = \sum_{x\in A} \mathcal{D}(x)$, then*

$$d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) = \max_{A\subseteq S} |\mathcal{D}_1(A) - \mathcal{D}_2(A)| .$$

Proof. Let S^+ be all the $x \in S$ that $\mathcal{D}_1(x) \geq \mathcal{D}_2(x)$, and $S^- = S \setminus S^+$, such that $\forall x \in S^-, \mathcal{D}_1(x) < \mathcal{D}_2(x)$. Then

$$(\mathcal{D}_1(S^+) - \mathcal{D}_2(S^+)) + (\mathcal{D}_2(S^-) - \mathcal{D}_1(S^-)) = \sum_{a\in S} |\mathcal{D}_1(a) - \mathcal{D}_2(a)| = 2d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) .$$

Since $\mathcal{D}_1(S^+) + \mathcal{D}_1(S^-) = \mathcal{D}_2(S^+) + \mathcal{D}_2(S^-) = 1$,

$$\mathcal{D}_1(S^+) - \mathcal{D}_2(S^+) = \mathcal{D}_2(S^-) - \mathcal{D}_1(S^-) = d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) .$$

Moreover, by the definition of S^+ and S^-, such that $\forall x \in S^+, \mathcal{D}_1(x) \geq \mathcal{D}_2(x)$, then $\forall A \subseteq S$,

$$\mathcal{D}_1(A) - \mathcal{D}_2(A) \leq \mathcal{D}_1(S^+) - \mathcal{D}_2(S^+).$$

Since $\mathcal{D}_1(S^+) - \mathcal{D}_2(S^+)$ is maximized, so is $\mathcal{D}_2(S^-) - \mathcal{D}_1(S^-)$ by symmetry, the proof is concluded. □

Remark 1.12.3. The alternative characterization of variation distance by lemma 1.12.2 gives a strong statement, that supposing $d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) < \varepsilon$, then for any $x \in S$, $|\mathcal{D}_1(x) - \mathcal{D}_2(x)| < \varepsilon$.

Corollary 1.12.4. *Let $m(a) = \min(\mathcal{D}_1(a), \mathcal{D}_2(a))$ for all $a \in S$, then*

$$d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) = 1 - m(S).$$

Proof. The result is immediate by $d_{\mathrm{TV}}(\mathcal{D}_1, \mathcal{D}_2) = \mathcal{D}_1(S^+) - \mathcal{D}_2(S^+)$ and $m(S) = \mathcal{D}_1(S^-) + \mathcal{D}_2(S^+)$. □

Definition 1.12.5 (Mixing Time)**.** Let π be the stationary distribution of an ergodic Markov chain with state space S. Let p_x^t represent the distribution of the state of the chain starting at state x after t steps. We define

$$\Delta_x(t) = d_{\mathrm{TV}}\left(\mathcal{D}_\pi, p_x^t\right)$$
$$\Delta(t) = \max_{x\in S} \Delta_x(t) = \max_{x\in S} d_{\mathrm{TV}}\left(\mathcal{D}_\pi, p_x^t\right),$$

that is, $\Delta_x(t)$ is the variation distance between the stationary distribution and p_x^t, and $\Delta(t)$ is the maximum of $\Delta_x(t)$ over all states x. We also define

$$\tau_x(\varepsilon) = \min\{t \mid \Delta_x(t) \leq \varepsilon\}$$
$$\tau(\varepsilon) = \max_{x\in S} \tau_x(\varepsilon) = \max_{x\in S} \min\{t \mid \Delta_x(t) \leq \varepsilon\},$$

that is, $\tau_x(\varepsilon)$ is the first step t, at which the variation distance between p_x^t and the stationary distribution is at most ε, and $\tau(\varepsilon)$ is the maximum of $\tau_x(\varepsilon)$ over all states x.

While $\tau(\varepsilon)$ is a function of ε, it is generally called the mixing time of the Markov chain.

Definition 1.12.6 (Rapidly Mixing [JS89])**.** A Markov chain with state space S is rapidly mixing if its mixing time $\tau(\varepsilon)$ is $\mathrm{poly}(\log(1/\varepsilon), \log|S|)$. We view $\log|S|$ as the problem size, since a state can be encoded using $O(\log|S|)$ bits.

1.12.2 Coupling

Coupling of Markov chains is a general technique for bounding the mixing time of a Markov chain.

Definition 1.12.7 (Markov Chain Coupling)**.** A coupling of a Markov chain M_t with state space S is a Markov chain $Z_t = (X_t, Y_t)$ on state space $S \times S$ such that

$$\Pr\left[X_{t+1} = x' \mid Z_t = (x, y)\right] = \Pr\left[M_{t+1} = x' \mid M_t = x\right];$$
$$\Pr\left[Y_{t+1} = y' \mid Z_t = (x, y)\right] = \Pr\left[M_{t+1} = y' \mid M_t = y\right].$$

Remark 1.12.8. A coupling consists of two copies of the Markov chain M_t running simultaneously, each copy behaves exactly like the original Markov chain in terms of its transition probabilities. The joint distribution preserves the marginal distribution of each copy, but the two copies may be correlated.

Remark 1.12.9. To show a chain is close to stationarity, it suffices to couple it with a copy started from the stationary distribution. When the two chains reach the same state (the *coupling time* T), they can be made to evolve identically thereafter. In particular, if they have coupled by time T, then the first chain is exactly equal to a stationary sample at time T. Thus, the only obstruction to being close to stationarity is the probability that they have not yet met.

Lemma 1.12.10 (Mixing-Time Coupling Lemma)**.** *Let $Z_t = (X_t, Y_t)$ be a coupling for a Markov chain M_t on a state space S. Suppose that there exists a T such that, for every $x, y \in S$,*

$$\Pr\left[X_T \neq Y_T \mid X_0 = x, Y_0 = y\right] \leq \varepsilon, \tag{1.19}$$

then $\tau(\varepsilon) \leq T$, that the variation distance between the distribution over S after T steps and $\mathcal{D}_\pi$ is at most ε.

Proof. Consider the coupling when initial state x is fixed and $y \leftarrow \mathcal{D}_\pi$. We first prove the following core lemma.

Lemma 1.12.11. *In the coupling $Z_t = (X_t, Y_t)$, where initial state x is fixed and $y \leftarrow \mathcal{D}_\pi$, then*

$$\Delta_x(T) \leq \Pr\left[X_T \neq Y_T\right].$$

Proof. We begin by lower bounding $p_x^T(A)$, where $A \subseteq S$, by

$$\begin{aligned} p_x^T(A) &= \Pr\left[X_T \in A\right] \\ &= \Pr\left[X_T \in A \wedge Y_T = X_T\right] + \Pr\left[X_T \in A \wedge Y_T \neq X_T\right] \\ &\geq \Pr\left[X_T \in A \wedge Y_T = X_T\right] = \Pr\left[Y_T \in A \wedge Y_T = X_T\right], \end{aligned}$$

where the last equality holds by the same event. We continue the lower bounding by

$$\begin{aligned} p_x^T(A) &\geq \Pr\left[Y_T \in A \wedge Y_T = X_T\right] \\ &= 1 - \Pr\left[Y_T \notin A \vee Y_T \neq X_T\right] \\ &\geq 1 - \Pr\left[Y_T \notin A\right] - \Pr\left[Y_T \neq X_T\right] = \Pr\left[Y_T \in A\right] - \Pr\left[Y_T \neq X_T\right] \\ &= \mathcal{D}_\pi(A) - \Pr\left[Y_T \neq X_T\right], \end{aligned}$$

where the third inequality holds by union bound. Symmetric to the prior derivation, we have

$$p_x^T(S \setminus A) \geq \mathcal{D}_\pi(S \setminus A) - \Pr\left[Y_T \neq X_T\right],$$

or equivalently,

$$\mathcal{D}_\pi(A) \geq p_x^T(A) - \Pr\left[Y_T \neq X_T\right].$$

Now that for all $A \subseteq S$, we have

$$\left|p_x^T(A) - \mathcal{D}_\pi(A)\right| \leq \Pr\left[X_T \neq Y_T\right],$$

then by the alternative characterization of variation distance lemma 1.12.2, we complete the proof by

$$\Delta_x(T) = d_{\mathrm{TV}}\left(p_x^T, \mathcal{D}_\pi\right) = \max_{A \subseteq S} \left|p_x^T(A) - \mathcal{D}_\pi(A)\right| \leq \Pr\left[X_T \neq Y_T\right].$$

□

Since $\Pr[X_T \neq Y_T] \leq \varepsilon$ by eq. (1.19), then by lemma 1.12.11, $\Delta_x(T) \leq \varepsilon$ for all $x \in S$. Hence, by definition 1.12.5,

$$T \geq \tau_x(\varepsilon) = \min\{t \mid \Delta_x(t) \leq \varepsilon\}$$

holds for all $x \in S$, and therefore $\tau(\varepsilon) \leq T$. □

Example 1.12.12 (Coupling for Card Shuffling)**.** A card-shuffling Markov chain for n cards is defined over $n!$ states, where a state transitions to another by choosing a position $j \overset{R}{\leftarrow} [1, n]$, and moving the j^{th} card to the deck top. It is obvious that the stationary distribution is the uniform distribution.

To analyze the mixing time $\tau(\varepsilon)$ of the chain by lemma 1.12.10, we obtain 2 copies of the chain, and the coupling is constructed by sampling a card C uniformly at random, and moving C to the deck top in both copies. The coupling is valid: recall remark 1.12.8, that the marginal distribution is preserved, while the joint distribution may be correlated.

Moreover, when C is moved to the deck top, then it is always in the same position in both copies, and the copies are coupled once every card has been moved to the top for at least once. Let $Z_t = (X_t, Y_t)$ be the coupling, then $\Pr[X_T \neq Y_T]$ is the probability that after T steps, there exists cards still not chosen.

We continue by analyzing the coupon collector's problem: with probability $(1 - n^{-1})^N$, a specific card is not chosen after N trials. By theorem 1.5.23, let $N = n \ln n + cn$, and by union bounding over all n cards,

$$\Pr[X_N \neq Y_N] \leq n\left(1 - \frac{1}{n}\right)^N \leq n \exp\left(-\frac{N}{n}\right) = e^{-c},$$

where the second inequality holds by $1 + x \leq e^x$. When $N = n \ln(n/\varepsilon)$, we have $\Pr[X_N \neq Y_N] \leq \varepsilon$.

We conclude by lemma 1.12.10, that the variation distance between the uniform distribution and the distribution of the state of the chain after $n \ln(n/\varepsilon)$ steps is at most ε, and $\tau(\varepsilon) \leq n \ln(n/\varepsilon)$.

Exercise 1.12.13 (Exercise 12.2 [MU17])**.** Considering the Markov chain for shuffling n cards as in example 1.12.12. Suppose that, instead of running the chain for a fixed number of steps, we stop the chain at the first step where every card has been moved to the top at least once. Show that, at the stopping time, the state of the chain is uniformly distributed over $n!$ possible permutations of the cards.

Proof. By the coupling in example 1.12.12, fix an initial state x for X_0 and let Y_0 be sampled uniformly from $n!$ decks. Let N be the first time that every card has been moved to the top at least once, then $X_N = Y_N$. We conclude that X_N is uniformly distributed over $n!$ possible decks. □

Exercise 1.12.14 (Exercise 12.3 [MU17])**.** Considering the Markov chain for shuffling n cards as in example 1.12.12. Show that if the chain runs for only $(1 - \varepsilon)n \ln n$ steps for some constant $\varepsilon > 0$, then the variation distance is $1 - o(1)$.

Proof. We started by lower bounding $d_{\text{TV}}(p_x^N, \mathcal{D}_\pi)$, where $N = (1 - \varepsilon)n \ln n$, x be an initial state. Since

$$d_{\text{TV}}(p_x^N, \mathcal{D}_\pi) = \max_{A \subseteq S} \left|p_x^N(A) - \mathcal{D}_\pi(A)\right| \geq \left|p_x^N(B) - \mathcal{D}_\pi(B)\right|$$

for all $B \subseteq S$, the lower bound can be derived from lower bounding $\left|p_x^N(B) - \mathcal{D}_\pi(B)\right|$.

Let the cards be initially unshuffled and ordered, then we define a subset of states by A_ℓ such that the last ℓ cards preserve the relative order. We let X_i be the 0-1 indicator random variable for the i^{th} card not being picked, and $X = \sum_i X_i$ be the random variable for the number of untouched cards.

Since $X \geq \ell$ leads to A_ℓ occurring, then $\{X \geq \ell\} \subseteq A_\ell$, and

$$\Pr[X \geq \ell] \leq p_x^N(A_\ell).$$

By the linearity of expectation,

$$\mathbb{E}[X] = \sum_{i \in [1,n]} \mathbb{E}[X_i] = n\left(1 - \frac{1}{n}\right)^N \leq n \cdot \exp\left(-(1 - \varepsilon)\ln n\right) = n^\varepsilon.$$

We let $\ell = n^{\varepsilon/2}$, and hence for any sufficiently large n,

$$\Pr[X < \ell] \le \Pr\left[X \le \frac{1}{2}n^{\varepsilon}\right].$$

Moreover, for $i \neq j \in [1, n]$,

$$\mathrm{Cov}\left[X_i, X_j\right] = \mathbb{E}\left[X_i X_j\right] - \mathbb{E}[X_i]\mathbb{E}\left[X_j\right] = \left(1 - \frac{2}{n}\right)^N - \left(1 - \frac{1}{n}\right)^{2N} < 0,$$

hence

$$\mathrm{Var}[X] = \sum_{i\in[1,n]} \mathrm{Var}[X_i] + 2\sum_{j<k\in[1,n]} \mathrm{Cov}\left[X_j, X_k\right] \le \sum_{i\in[1,n]} \mathrm{Var}[X_i] \le \sum_{i\in[1,n]} \mathbb{E}[X_i] = \mathbb{E}[X],$$

then by Chebyshev's inequality,

$$\Pr[X < \ell] \le \Pr\left[X \le \frac{1}{2}n^{\varepsilon}\right] \le \frac{4\mathrm{Var}[X]}{n^{2\varepsilon}} \le \frac{4}{n^{\varepsilon}}.$$

On the other hand, $\mathcal{D}_\pi(A_\ell) = (\ell!)^{-1} = \left(n^{\varepsilon/2}!\right)^{-1}$. Hence,

$$d_{\mathrm{TV}}(p_x^N, \mathcal{D}_\pi) = \max_{A\subseteq S}\left|p_x^N(A) - \mathcal{D}_\pi(A)\right| \ge \left|p_x^N(A_\ell) - \mathcal{D}_\pi(A_\ell)\right| \ge 1 - \frac{4}{n^{\varepsilon}} - \frac{1}{n^{\varepsilon/2}!},$$

which completes the proof for $d_{\mathrm{TV}}(p_x^N, \mathcal{D}_\pi) = 1 - o(1)$. □

Example 1.12.15 (Coupling for Random Walks on the Hypercube)**.** Consider the following Markov chain defined on a n-dimensional hypercube. At each step, it picks a coordinate $i \xleftarrow{R} [1, n]$. The new state $\mathbf{x}'$ is obtained from the current state $\mathbf{x}$ by keeping all coordinates the same, except possibly for x_i. x_i is set to 0 with probability $1/2$, and to 1 with probability $1/2$. The stationary distribution of the Markov chain is uniform distribution.

A coupling for the Markov chain can be obtained by enforcing the same move on the two copies: both changing x_i to 0 or 1. The marginal probability of the coupling still follows the transition probability of the chain.

Moreover, once both copies make the same move on x_i, they are surely agreeing on the i^{th} coordinate, and they will couple after all n coordinates have each been chosen at least once. Immediate from the coupon collector's problem, and by the same analysis in example 1.12.12, $\tau(\varepsilon) \le n\ln(n/\varepsilon)$.

Exercise 1.12.16 (Exercise 12.11 [MU17])**.** Show that the Markov chain for sampling all independent sets of fixed size $k \le n/(3\Delta + 3)$ in a graph of n vertices and maximum degree Δ is ergodic and has a uniform stationary distribution. A move is made from the independent set X_t by choosing a vertex $v \xleftarrow{R} X_t$ and $w \xleftarrow{R} V$. If $w \notin X_t$ and $X_t \setminus \{v\} \cup \{w\}$ is an independent set, then $X_{t+1} = X_t \setminus \{v\} \cup \{w\}$; otherwise, $X_{t+1} = X_t$.

Proof. Consider an independent set $A \subseteq V$ with $|A| = k$, and let $\Gamma^+(A)$ be the inclusive neighborhood of A, such that

$$|\Gamma^+(A)| \le k(\Delta + 1).$$

Suppose we want to transition from A to B, another independent set of fixed size with k, we know that

$$|\Gamma^+(A \cup B)| \le |\Gamma^+(A)| + |\Gamma^+(B)| \le 2k(\Delta + 1).$$

Since $n \ge 3k(\Delta + 1)$, then $|V \setminus \Gamma^+(A \cup B)| \ge k(\Delta + 1)$.

We can greedily construct an independent set $C \subseteq V \setminus \Gamma^+(A \cup B)$: at each step choose a vertex, put it into C, and delete it together with all its neighbors. Since each step deletes at most $\Delta + 1$ vertices, we can greedily select at least k vertices from $V \setminus \Gamma^+(A \cup B)$ into C. Hence there exists an independent set C of size k.

Then for any transition in between states of independent sets of fixed size, one can first transition to an independent set of fixed size $C \subseteq V \setminus \Gamma^+(A \cup B)$, then transition to B. Hence, the irreducibility is proved.

Positive recurrence is immediate by finiteness and irreducibility, and aperiodicity is immediate by the existence of odd length loop, namely the loop back to a state itself. Hence, the ergodicity is proved by corollary 1.7.47. □

Example 1.12.17 (Independent Sets of Fixed Size)**.** For exercise 1.12.16, consider a coupling $Z_t = (X_t, Y_t)$, requiring a perfect matching M between vertices in $X_t \setminus Y_t$ and $Y_t \setminus X_t$ at each step. We choose a transition for X_t by sampling $v \xleftarrow{\text{R}} X_t$ and $w \xleftarrow{\text{R}} V$ and perform the move. If $v \in Y_t$, then perform the move over v, w, and Y_t; otherwise, perform the move over $M(v)$, w, and Y_t.

An alternative way of coupling is established by sampling $v \xleftarrow{\text{R}} X_t$ and $w \xleftarrow{\text{R}} V$ and perform the move for X_t, if $v \in Y_t$, perform the move over v, w, and Y_t; otherwise, sample $v' \xleftarrow{\text{R}} Y_t \setminus X_t$ and perform the move over v', w, and Y_t.

The coupling still remains valid, by

$$\Pr\left[v' \text{ chosen} \wedge v' \in X_t \cap Y_t\right] = \Pr\left[v' \text{ chosen} \mid v \xleftarrow{\text{R}} X_t \text{ and } v \in Y_t\right] \Pr\left[v \xleftarrow{\text{R}} X_t \text{ and } v \in Y_t\right] = \frac{1}{|X_t \cap Y_t|} \cdot \frac{|X_t \cap Y_t|}{k},$$

$$\Pr\left[v' \text{ chosen} \wedge v' \in Y_t \setminus X_t\right] = \Pr\left[v' \text{ chosen} \mid v \xleftarrow{\text{R}} X_t \text{ and } v \notin Y_t\right] \Pr\left[v \xleftarrow{\text{R}} X_t \text{ and } v \notin Y_t\right] = \frac{1}{|Y_t \setminus X_t|} \cdot \frac{|X_t \setminus Y_t|}{k},$$

such that $v \in Y_t$ and $w \in V$ are still chosen with probability $1/kn$. Note that, in the coupling, it is possible to transition in one chain and stay put in the other.

Let $D_t = |X_t \setminus Y_t|$ be the random variable for the difference size between X_t and Y_t, and D_t can change by at most 1 at a step. We want to show D_t is more likely to decrease, and we want to upper bound the probability of $D_t > 0$ for a sufficiently large t.

Suppose $D_t > 0$. In order to make $D_{t+1} = D_t + 1$, the vertex v to be removed has to be sampled from $X_t \cap Y_t$, while the vertex w to be introduced can only cause one of the two states to transition, such that [10]

$$w \in \Gamma^+\left((X_t \setminus Y_t) \cup (Y_t \setminus X_t)\right).$$

Then it follows that

$$\Pr\left[D_{t+1} = D_t + 1 \mid D_t > 0\right] \leq \frac{k - D_t}{k} \cdot \frac{2D_t(\Delta + 1)}{n}.$$ [11]

In order to make $D_{t+1} = D_t - 1$, the vertex v to be removed has to be sampled from $X_t \setminus Y_t$, while the vertex w to be introduced must cause both states to transition, such that

$$w \notin \Gamma^+\left(X_t \cup Y_t \setminus \{v, v'\}\right),$$

hence

$$\Pr\left[D_{t+1} = D_t - 1 \mid D_t > 0\right] \geq \frac{D_t}{k} \cdot \frac{n - (\Delta + 1)(k + D_t - 2)}{n}.$$

We thus have, for $D_t > 0$,

$$\begin{aligned}
\mathbb{E}\left[D_{t+1} \mid D_t\right] &= D_t + \Pr\left[D_{t+1} = D_t + 1 \mid D_t > 0\right] - \Pr\left[D_{t+1} = D_t - 1 \mid D_t > 0\right] \\
&\leq D_t + \frac{k - D_t}{k} \cdot \frac{2D_t(\Delta + 1)}{n} - \frac{D_t}{k} \cdot \frac{n - (\Delta + 1)(k + D_t - 2)}{n} \\
&= D_t \left(1 - \frac{n - (\Delta + 1)(3k - D_t - 2)}{kn}\right) \\
&\leq D_t \left(1 - \frac{n - (\Delta + 1)(3k - 3)}{kn}\right),
\end{aligned}$$

and once $D_t = 0$, $\mathbb{E}\left[D_{t+1} \mid D_t\right] = 0$. By conditional expectation,

$$\mathbb{E}\left[D_{t+1}\right] = \mathbb{E}\left[\mathbb{E}\left[D_{t+1} \mid D_t\right]\right] \leq \mathbb{E}\left[D_t\right] \left(1 - \frac{n - (\Delta + 1)(3k - 3)}{kn}\right).$$

[10] w can be strengthened from causing one of the two states to transition, by observing if $w \in (X_t \setminus Y_t) \cup (Y_t \setminus X_t)$, the transition state's shared component with the state staying put is still of size $k - D_t$, by removing 1 shared element and reintroducing 1 shared element. If w is sampled over the neighbors of $(X_t \setminus Y_t) \cup (Y_t \setminus X_t)$, then the transition state is surely decreasing the size of the shared component with the state staying put.

[11] If we strengthen the condition for elements to introduce, then the $\Delta + 1$ should be Δ by considering only the neighboring elements.

We then conclude that

$$\begin{aligned}\Pr[D_t \geq 1] \leq \mathbb{E}[D_t] &\leq \mathbb{E}[D_0]\left(1 - \frac{n-(\Delta+1)(3k-3)}{kn}\right)^t \\ &\leq k\left(1 - \frac{n-(\Delta+1)(3k-3)}{kn}\right)^t \\ &\leq k \cdot \exp\left(-t \cdot \frac{n-(\Delta+1)(3k-3)}{kn}\right),\end{aligned}$$

the first inequality holds from Markov inequality, the third holds by $\mathbb{E}[D_0] \leq k$, and the last holds by $1 + x \leq e^x$. It converges to 0 when $k \leq n/(3\Delta+3)$, then in this case,

$$\tau(\varepsilon) \leq \frac{kn\ln(k/\varepsilon)}{n-(\Delta+1)(3k-3)}.$$

It turns out that $\tau(\varepsilon)$ is polynomial in n (which is logarithmic in the number of independent sets of fixed size k) and $\ln(\varepsilon^{-1})$, hence it is rapidly mixing.

Exercise 1.12.18 (Exercise 12.8 [MU17])**.** Consider the random walk on a non-bipartite, connected graph with $|V| = n$, where each vertex has the same degree $n > d > n/2$. Show that

$$\tau(\varepsilon) \leq \frac{\ln\varepsilon}{\ln(1-(2d-n)/d)}.$$

Proof. By theorem 1.7.75, the stationary distribution of such a graph is uniform.

A coupling can be established by sampling one copy's initial state from the stationary distribution, while the other's initial state is any fixed state. Suppose they are on states $u \neq v$, then on transition, construct a perfect matching M between $\Gamma(u)$ and $\Gamma(v)$, such that $\Gamma(u) \cap \Gamma(v)$ are matched to themselves, while the other neighboring vertices are randomly 1-to-1 matched. We randomly choose 1 out of d options in M, and since $|\Gamma(u) \cap \Gamma(v)| \geq 2d - n$, then

$$\Pr\left[X_{t+1} = Y_{t+1} \mid X_t \neq Y_t, X_0 = x, Y_0 = y\right] \geq \frac{2d-n}{d}$$

for all initial states $x \neq y$, hence by induction

$$\Pr\left[X_t \neq Y_t \mid X_0 = x, Y_0 = y\right] \leq \left(1 - \frac{2d-n}{d}\right)^t = \varepsilon,$$

then

$$t = \frac{\ln\varepsilon}{\ln(1-(2d-n)/d)}.$$

By lemma 1.12.10, $\Delta_x(t) \leq \Pr[X_t \neq Y_t]$, then it is immediate that $\tau(\varepsilon) \leq t$. □

Exercise 1.12.19 (Exercise 12.9 [MU17])**.** Consider a Markov chain on n points $[0, n-1]$ lying in order on a circle. At each step, the chain stays at the current point with probability $1/2$, or moves to the next point in the clockwise direction with probability $1/2$. Show that for any $\varepsilon > 0$, the mixing time $\tau(\varepsilon)$ is $O(n^2 \ln(1/\varepsilon))$.

Proof. It is clear by the cutset argument that the stationary distribution is the uniform distribution.

Consider a coupling where each copy operates oppositely, namely if one stays, the other moves to the next point. Let the distance be the number of steps to take clockwise from a chain state to the other, then it can be modeled by the gambler's ruin example 1.7.51: states are $[0, n]$, absorbing states are 0 and n, moving ± 1 with probability $1/2$.

By exercise 1.7.68, if the current distance is i, the expected number of moves is $i \cdot (n-i) \leq n^2/4$. Let T be the random variable for the number of steps before coupling, then by Markov's inequality, for all initial states $x \neq y$,

$$\Pr\left[T \geq \frac{n^2}{2} \,\middle|\, X_0 = x, Y_0 = y\right] \leq \frac{1}{2}.$$

By Markov property,

$$\Pr\left[T \geq n^2 \,\middle|\, T \geq \frac{n^2}{2}, X_0 = x, Y_0 = y\right] \leq \frac{1}{2},$$

and by induction,

$$\Pr\left[T \geq \frac{k}{2}n^2 \,\middle|\, X_0 = x, Y_0 = y\right] \leq 2^{-k}.$$

Hence,

$$\Pr\left[X_T \neq Y_T \mid X_0 = x, Y_0 = y\right] = \Pr\left[T \geq \frac{\log\left(\varepsilon^{-1}\right)}{2}n^2 \,\middle|\, X_0 = x, Y_0 = y\right] \leq \varepsilon.$$

Immediate from lemma 1.12.10, we have $\tau(\varepsilon) \leq T = \log(\varepsilon^{-1})n^2/2$, hence $\tau(\varepsilon) = O(\ln(\varepsilon^{-1})n^2)$. □

Exercise 1.12.20 (Exercise 12.12 [MU17])**.** We improve the coupling in example 1.12.17: if an attempt is made to move $v \in X_t \setminus Y_t$ to a vertex w, then the same attempt is made with the matched vertex in the other chain; if, however, if an attempt is made to move $v \in X_t \cap Y_t$, we no longer attempt to make the same move.

- Assume there are sets $S_1 = \Gamma^+(X_t \setminus Y_t)$, and $S_2 = \Gamma^+(Y_t \setminus X_t)$, each of exactly $D_t(\Delta + 1)$ distinct vertices. Assume further that $S_1 \cap S_2 = \varnothing$. Suppose that we match up elements in S_1 and S_2 in a 1-to-1 fashion.

  Argue that the moves can be coupled that, when one chain attempts and fails to move v to a vertex in S_1, it also attempts and fails to move v to the matching vertex in S_2 on the other chain.

  Similarly, argue that the moves can be coupled that, when one chain attempts and succeeds in moving v to a vertex in S_1, it also attempts and succeeds in moving v to the matching vertex in S_2 on the other chain.

  Show that the coupling gives

$$\Pr\left[D_{t+1} = D_t + 1\right] \leq \frac{k - D_t}{k} \cdot \frac{D_t(\Delta + 1)}{n}. \tag{1.20}$$

- In the general case, S_1 and S_2 are not necessarily disjoint or of same size. Show that in the general case, by pairing up failing moves as much as possible, the number of choices for w that can increase D_t is $\max(|S_1|, |S_2|) \leq D_t(\Delta + 1)$. Then argue that eq. (1.20) holds in all cases.
- Use this coupling to derive a polynomial bound on $\tau(\varepsilon)$ that holds for any $k \leq n/(2\Delta + 2)$.

Proof. Consider in the optimal case, where S_1 and S_2 are $D_t(\Delta + 1)$ large, $S_1 \cap S_2 = \varnothing$, and M is the 1-to-1 mapping. Let M' be the 1-to-1 mapping between $X_t \setminus Y_t$ and $Y_t \setminus X_t$. The chain is coupled by sampling $v \xleftarrow{R} X_t$ and $w \xleftarrow{R} V$:

- If $v \in X_t \setminus Y_t$, then let $v' = M'(v)$ for Y_t; otherwise, let $v' = v$.
- If $w \in S_1$, then let $w' = M(w)$ for Y_t; if $w \in S_2$, then let $w' = M^{-1}(w)$ for Y_t; otherwise, let $w' = w$.

Perform the move removing v for w on X_t and removing v' for w' on Y_t.

When we want to remove $v \in X_t \cap Y_t$ for chain X_t, the transition fails when $w \in \Gamma^+(X_t \setminus \{v\})$, hence the transitions for X_t fails when $w \in S_1 = \Gamma^+(X_t \setminus Y_t)$. Symmetrically, by the mapping M, if $w' = M(w) \in S_2$, transition for Y_t fails.

Conversely, when $v \in X_t \cap Y_t$, for chain X_t, transition succeeds when $w \in V \setminus \Gamma^+(X_t \setminus \{v\})$, hence the transition for X_t succeeds when $w \in S_2$. Symmetrically, by mapping M, if $w' = M^{-1}(w) \in S_1$, transition for Y_t succeeds.

In order for $D_{t+1} = D_t + 1$, it must be at time t the vertex v chosen from $X_t \cap Y_t$, and w must be chosen from S_2, or otherwise either $w = w'$, or $w \in S_1$ fails to transition. Hence, by the same upper bound argument in example 1.12.17,

$$\Pr\left[D_{t+1} = D_t + 1\right] \leq \frac{|X_t \cap Y_t|}{k} \cdot \frac{|S_2|}{n} = \frac{k - D_t}{k} \cdot \frac{D_t(\Delta + 1)}{n}.$$

Moving on the general case. We reuse previous S_1, S_2, M', and let M be the mapping between S_1 and S_2, such that $S_1 \cap S_2$ are mapped to itself, bijectively map elements between S_1 and S_2 as many as possible, and the remaining elements in the larger set map to itself.

In this way, if $v \in X_t \cap Y_t$, failing moves of moving v for $w \in S_1$ in X_t and moving v for $w' \in S_2$ in Y_t are maximally paired. WLOG let S_2 be the larger set, as the symmetric case can be analyzed in the same way.

- If $w \in M(S_1 \setminus S_2)$, then the successful moves are also maximally paired.
- If $w \in (S_2 \setminus S_1) \setminus M(S_1 \setminus S_2)$, then w is mapped to itself because S_2 is larger, only X_t transitions successfully.
- Otherwise, for $w \in V \setminus (S_1 \cup S_2)$, then either both chains transition to same w, or neither chain transitions.

D_t increases only in the first 2 cases for at most $\max(|S_1|, |S_2|) \leq D_t(\Delta + 1)$ elements. Hence eq. (1.20) holds by

$$\Pr[D_{t+1} = D_t + 1] \leq \frac{|X_t \cap Y_t|}{k} \cdot \frac{\max(|S_1|, |S_2|)}{n} = \frac{k - D_t}{k} \cdot \frac{D_t(\Delta + 1)}{n}.$$

To decrease D_t, it has to be $u \in X_t \setminus Y_t$ moving to $w \in V \setminus \Gamma^+(X_t \cup Y_t \setminus \{u, u'\})$ in X_t. Continuing from example 1.12.17,

$$\Pr[D_{t+1} = D_t - 1 \mid D_t > 0] \geq \frac{D_t}{k} \cdot \frac{n - (\Delta + 1)(k + D_t - 2)}{n}.$$

By the same analysis,

$$\begin{aligned}
\mathbb{E}[D_{t+1} \mid D_t] &= D_t + \Pr[D_{t+1} = D_t + 1 \mid D_t > 0] - \Pr[D_{t+1} = D_t - 1 \mid D_t > 0] \\
&\leq D_t + \frac{k - D_t}{k} \cdot \frac{D_t(\Delta + 1)}{n} - \frac{D_t}{k} \cdot \frac{n - (\Delta + 1)(k + D_t - 2)}{n} \\
&= D_t \left(1 - \frac{n - (\Delta + 1)(2k - 2)}{kn}\right),
\end{aligned}$$

then by the same Markov inequality argument in example 1.12.17,

$$\begin{aligned}
\Pr[D_t \geq 1] \leq \mathbb{E}[D_t] &= \mathbb{E}[D_0] \left(1 - \frac{n - (\Delta + 1)(2k - 2)}{kn}\right)^t \\
&\leq k \left(1 - \frac{n - (\Delta + 1)(2k - 2)}{kn}\right)^t \\
&\leq k \cdot \exp\left(-t \cdot \frac{n - (\Delta + 1)(2k - 2)}{kn}\right),
\end{aligned}$$

which converges to 0 when $k \leq n/(2\Delta + 2)$, and we have

$$\tau(\varepsilon) \leq \frac{kn \ln(k/\varepsilon)}{n - (\Delta + 1)(2k - 2)}.$$

□

Remark 1.12.21. The exercise 1.12.20 is a finer analysis of example 1.12.17 derived from [BD97].

Exercise 1.12.22 (Exercise 12.14 [MU17])**.** Consider the following variant of shuffling a deck of n cards: at each step, two cards are chosen independently and uniformly at random from the deck, and their positions are exchanged.

- Argue the following being equivalent: at each step, a specific card is chosen uniformly at random from the deck, and a position $i \xleftarrow{R} [1, n]$ sampled independently. The i^{th} card is exchanged with the specific chosen card.
- Consider the coupling where the two choices of card and position are the same for both copies of the chain. Let X_t be the number of cards whose positions differ in the two copies of the chain. Show X_t is nonincreasing,

$$\Pr[X_{t+1} \leq X_t - 1 \mid X_t > 0] \geq \left(\frac{X_t}{n}\right)^2,$$

and the expected time until $X_t = 0$ is $O(n^2)$, regardless of the initial states of the two chains.

Proof. It suffices to show that, for a specific card, sampling $i \xleftarrow{R} [1, n]$ and the i^{th} card in the deck being this specific card has the same probability as choosing this card out of uniform randomness.

For every coupled move,

- If the chosen card is in the same position in both decks, then $X_{t+1} = X_t$.
- Otherwise, if the chosen card is in different positions in both decks,
  - If the i^{th} position cards are the same in both decks, then $X_{t+1} = X_t$.
  - Otherwise, the coupled move at least introduces 1 correction, hence $X_{t+1} \le X_t - 1$.

Therefore the coupled move makes X_t nonincreasing over time, and correction only happens in both the card chosen and the position chosen being in disagreeing section, hence

$$\Pr[X_{t+1} \le X_t - 1 \mid X_t > 0] \ge \left(\frac{X_t}{n}\right)^2.$$

Conditioned on $X_t = k$, the steps to decrease X_t is geometrically distributed, with success probability $(k/n)^2$. Let Y_k be the steps to decrease from k differences, and $Y = \sum_k Y_k$ be the steps before $X_t = 0$,

$$\mathbb{E}[Y] \le \sum_{k \in [1,n]} \left(\frac{n}{k}\right)^2 = O(n^2),$$

which completes the proof. □

1.12.3 Variation Distance is Nonincreasing

Lemma 1.12.23 (Optimal Coupling Lemma)**.** *Given distribution σ_X and σ_Y on a state space S, let $Z = (X, Y)$ be a random variable on $S \times S$, where X is distributed according to σ_X and Y is distributed according to σ_Y. Then*

$$\Pr[X \ne Y] \ge d_{\mathsf{TV}}(\sigma_X, \sigma_Y).$$

Moreover, there exists a joint distribution $Z = (X, Y)$, where X is distributed according to σ_X and Y is distributed according to σ_Y, for which the equality holds.

Proof. The proof idea of the first part is the same as lemma 1.12.11, where we replace $\mathcal{D}_\pi$ with σ_Y.

Alternatively, we can prove by

$$\begin{aligned}\Pr[X = x] &= \Pr[X = x \wedge X = Y] + \Pr[X = x \wedge X \ne Y] \\ &= \Pr[Y = x \wedge X = Y] + \Pr[X = x \wedge X \ne Y] \\ &\le \Pr[Y = x] + \Pr[X = x \wedge X \ne Y].\end{aligned}$$

By symmetry, we have

$$\begin{aligned}\Pr[X = x] - \Pr[Y = x] &\le \Pr[X = x \wedge X \ne Y], \\ \Pr[Y = x] - \Pr[X = x] &\le \Pr[Y = x \wedge X \ne Y].\end{aligned}$$

Let $S^+, S^- \subseteq S$ be the sets, such that $S^- = S \setminus S^+$, $\forall a \in S^+$, $\sigma_X(a) \ge \sigma_Y(a)$, and $\forall a \in S^-$, $\sigma_X(a) < \sigma_Y(a)$, then

$$\begin{aligned}2d_{\mathsf{TV}}(\sigma_X, \sigma_Y) &= (\sigma_X(S^+) - \sigma_Y(S^+)) + (\sigma_Y(S^-) - \sigma_X(S^-)) \\ &\le \sum_{a \in S^+} \Pr[X = a \wedge X \ne Y] + \sum_{a \in S^-} \Pr[Y_t = a \wedge X \ne Y] \\ &\le \Pr[X \ne Y] + \Pr[X \ne Y].\end{aligned}$$

Or alternatively, by corollary 1.12.4, for each $a \in S$, both X and Y are a with probability at most $m(a)$

$$\Pr[X = Y = a] \leq m(a),$$

where the equality holds by matching X and Y as much as possible, hence we complete the proof by corollary 1.12.4,

$$\Pr[X = Y] \leq m(S) = 1 - d_{\mathsf{TV}}(\sigma_X, \sigma_Y).$$

The equality case follows by the construction matching X and Y as much as possible in the joint distribution. If $m(S) = 1$, the equality holds by X and Y are same distribution. Otherwise, let $Z = (X, Y)$,

$$\Pr[X = x, Y = y] = \begin{cases} m(x) & x = y, \\ \dfrac{(\sigma_X(x) - m(x))(\sigma_Y(y) - m(y))}{1 - m(S)} & x \neq y. \end{cases}$$

The equality holds by corollary 1.12.4,

$$\Pr[X \neq Y] = 1 - m(S) = d_{\mathsf{TV}}(\sigma_X, \sigma_Y).$$

The marginal distribution is preserved as follows: if $m(x) = \sigma_X(x)$, then $\Pr[X = x] = \sigma_X(x)$; otherwise $m(x) = \sigma_Y(x)$,

$$\Pr[X = x] = m(x) + \frac{\sigma_X(x) - m(x)}{1 - m(S)} \sum_{y \neq x} (\sigma_Y(y) - m(y)) = m(x) + \sigma_X(x) - m(x) = \sigma_X(x).$$

The case for Y holds by symmetry. □

Corollary 1.12.24. *Let (X_t, Y_t) be two copies of a Markov chain M_t on a state space S. For any initial states $x, y \in S$ and any time $t \geq 0$, there exists a coupling of the two copies, such that*

$$d_{\mathsf{TV}}\left(p_x^t, p_y^t\right) = \Pr[X_t \neq Y_t].$$

Proof. There exists a coupled pair of random variables (U, V) of marginal distributions p_x^t and p_y^t

$$\Pr[U \neq V] = d_{\mathsf{TV}}\left(p_x^t, p_y^t\right)$$

by lemma 1.12.23. We now extend this optimal coupling of the endpoints to a coupling of full trajectories.

First sample the endpoint pair (U, V). Conditioned on $U = u$, sample a path

$$X_0, \ldots, X_{t-1}, X_t$$

by the transition probability of M_t, conditioned on the events $X_t = u$ and $X_0 = x$.

Similarly, conditioned on $V = v$, sample a path

$$Y_0, \ldots, Y_{t-1}, Y_t$$

conditioned on the events $Y_t = v$ and $Y_0 = y$.

For any path $X_0 = x, \ldots, X_{t-1} = x_{t-1}, X_t = u$, the conditional probability is

$$\Pr[X_1 = x_1, \ldots, X_{t-1} = x_{t-1} \mid X_0 = x, X_t = u] = \frac{1}{p_x^t(u)} \prod_{i \in [1,t]} P_{x_{i-1}, x_i}.$$

Since $p_x^t(u) > 0$ by U having marginal distribution p_x^t, the conditional probability is well-defined.

By construction, the marginal distribution of $(X_0, \ldots, X_t)$ is that of the chain M_t started from x, and the marginal distribution of $(Y_0, \ldots, Y_t)$ is that of the chain M_t started from y.

Moreover, since $X_t = U$ and $Y_t = V$, then

$$\Pr[U \neq V] = \Pr[X_t \neq Y_t] = d_{\mathsf{TV}}\left(p_x^t, p_y^t\right),$$

and we have constructed a coupling of two copies of the Markov chain reaching optimal coupling at time t. □

Lemma 1.12.25. *For any ergodic Markov chain,* $\Delta(T+1) \le \Delta(T)$.

Proof. Keep in mind that $\Delta_x(t)$ is a concrete deterministic value of variation distance between p_x^t and $\mathcal{D}_\pi$. The proof idea is to couple optimally at time T, then show that $\Delta_x(T+1)$ has a lower upper bound than $\Pr[X_T \neq Y_T]$.

Fix an initial state $x \in S$. If we sample $y \leftarrow \mathcal{D}_\pi$ for the initial state of p_y^T, then

$$\Delta_x(T) = d_{\mathrm{TV}}\left(p_x^T, p_y^T\right).$$

By corollary 1.12.24, there exists a coupling for the chains such that they are optimally coupled at time T, we have

$$\Pr[X_T \neq Y_T] = \Delta_x(T).$$

Next, we couple the chains X_T and Y_T by: if $X_T = Y_T$, for all $s \ge T$, $X_s = Y_s$; otherwise, they transition independently. By the sticky step-by-step coupling, since $\{X_T = Y_T\} \subseteq \{X_{T+1} = Y_{T+1}\}$,

$$\Pr[X_T \neq Y_T] \ge \Pr[X_{T+1} \neq Y_{T+1}].$$

Hence, $\Delta_x(T)$ is non-increasing by

$$\Delta_x(T) = \Pr[X_T \neq Y_T] \ge \Pr[X_{T+1} \neq Y_{T+1}] \ge \Delta_x(T+1),$$

where the last inequality holds by lemma 1.12.11. Since it holds for all $x \in S$, we have $\Delta(T) \ge \Delta(T+1)$. □

Exercise 1.12.26 (Exercise 12.13 [MU17])**.** Consider a Markov chain with state space S and a stationary distribution $\mathcal{D}_\pi$. For any nonnegative integer t define

$$\overline{\Delta}(t) = \max_{x,y \in S} d_{\mathrm{TV}}\left(p_x^t, p_y^t\right).$$

- For any positive s and t, show $\overline{\Delta}(s+t) \le \overline{\Delta}(s) \cdot \overline{\Delta}(t)$.
- For any positive s and t, show $\Delta(s+t) \le \Delta(s) \cdot \overline{\Delta}(t)$.
- For any positive t, show $\Delta(t) \le \overline{\Delta}(t) \le 2\Delta(t)$.

Proof. Fix initial states $(X_0, Y_0) = (x, y)$. By corollary 1.12.24, there exists a coupling for the chains such that they are optimally coupling at time t, then

$$\Pr[X_t \neq Y_t] = d_{\mathrm{TV}}\left(p_x^t, p_y^t\right).$$

Consider any coupling for the chains in such a way that, whenever $X_t = Y_t$, then $X_{t+i} = Y_{t+i}$ for all $i \ge 1$. Then

$$\begin{aligned}\Pr[X_{t+s} \neq Y_{t+s}] &= \Pr[X_{t+s} \neq Y_{t+s} \mid X_t \neq Y_t] \Pr[X_t \neq Y_t] + \Pr[X_{t+s} \neq Y_{t+s} \mid X_t = Y_t] \Pr[X_t = Y_t] \\ &= \Pr[X_{t+s} \neq Y_{t+s} \mid X_t \neq Y_t] \Pr[X_t \neq Y_t].\end{aligned}$$

Conditioned on $(X_t, Y_t) = (x', y')$, again by corollary 1.12.24, there exists a coupling for the chains such that they are optimally coupling at time $t+s$, then

$$\Pr\left[X_{t+s} \neq Y_{t+s} \mid X_t = x', Y_t = y'\right] = d_{\mathrm{TV}}\left(p_{x'}^s, p_{y'}^s\right) \le \max_{x' \neq y' \in S} d_{\mathrm{TV}}\left(p_{x'}^s, p_{y'}^s\right) = \overline{\Delta}(s), \tag{1.21}$$

Hence,

$$\Pr[X_{t+s} \neq Y_{t+s} \mid X_t \neq Y_t] \le \overline{\Delta}(s). \tag{1.22}$$

Therefore, we prove $\overline{\Delta}(t+s) \le \overline{\Delta}(t) \cdot \overline{\Delta}(s)$ by showing the following holds for all initial states x and y,

$$d_{\mathrm{TV}}\left(p_x^{t+s}, p_y^{t+s}\right) \le \Pr[X_{t+s} \neq Y_{t+s}] \le \overline{\Delta}(s) \cdot d_{\mathrm{TV}}\left(p_x^t, p_y^t\right) \le \overline{\Delta}(s) \cdot \overline{\Delta}(t).$$

Fix initial states $(X_0, Y_0) = (x, y)$, where $y \leftarrow \mathcal{D}_\pi$. Let (X_s, Y_s) be optimally coupled at time s by corollary 1.12.24, then

$$\Pr[X_s \neq Y_s] = \Delta_x(s).$$

Continue with the same sticky coupling after time s. Conditioned on $(X_s, Y_s) = (x', y')$, by corollary 1.12.24, there exists an optimal coupling for the next t steps. Then, again by corollary 1.12.24,

$$\Pr[X_{t+s} \neq Y_{t+s} \mid X_s \neq Y_s] \leq \overline{\Delta}(t)$$

by eq. (1.21) and eq. (1.22). Hence, $\Delta(t+s) \leq \Delta(s) \cdot \overline{\Delta}(t)$ holds by the same argument.

It is immediate from the triangle inequality that $\overline{\Delta}(t) \leq 2\Delta(t)$. To see $\Delta(t) \leq \overline{\Delta}(t)$, by theorem 1.7.13, for any t

$$\mathcal{D}_\pi(a) = \sum_{y \in S} \mathcal{D}_\pi(y) \cdot p_y^t(a),$$

then by lemma 1.12.2, let $S^+ \subseteq S$ be the set of all elements in S such that $p_x^t(a) \geq \mathcal{D}_\pi(a)$, we have

$$\begin{aligned}
\Delta(t) &= p_x^t(S^+) - \mathcal{D}_\pi(S^+) \\
&= p_x^t(S^+) - \sum_{y \in S} \mathcal{D}_\pi(y) \cdot p_y^t(S^+) = \sum_{y \in S} \mathcal{D}_\pi(y) \cdot \left(p_x^t(S^+) - p_y^t(S^+) \right) \\
&\leq \sum_{y \in S} \mathcal{D}_\pi(y) \cdot d_{\mathsf{TV}}\left(p_x^t, p_y^t \right) \leq \overline{\Delta}(t).
\end{aligned}$$

Alternatively, by definition 1.12.1,

$$\begin{aligned}
\Delta(t) &= \frac{1}{2} \sum_{a \in S} \left| p_x^t(a) - \mathcal{D}_\pi(a) \right| \\
&= \frac{1}{2} \sum_{a \in S} \left| p_x^t(a) - \sum_{y \in S} \mathcal{D}_\pi(y) \cdot p_y^t(a) \right| = \frac{1}{2} \sum_{a \in S} \left| \sum_{y \in S} \mathcal{D}_\pi(y) \cdot \left(p_x^t(a) - p_y^t(a) \right) \right| \\
&\leq \frac{1}{2} \sum_{a \in S} \sum_{y \in S} \left| \mathcal{D}_\pi(y) \cdot \left(p_x^t(a) - p_y^t(a) \right) \right| = \sum_{y \in S} \mathcal{D}_\pi(y) \cdot \frac{1}{2} \sum_{a \in S} \left| p_x^t(a) - p_y^t(a) \right| \\
&= \sum_{y \in S} \mathcal{D}_\pi(y) \cdot d_{\mathsf{TV}}\left(p_x^t, p_y^t \right) \leq \overline{\Delta}(t).
\end{aligned}$$

□

Remark 1.12.27. The lemma 1.12.25 and exercise 1.12.26 [12] are deriving from [Dob40], and well noted in [LP17].

Lemma 1.12.28.

$$d_{\mathsf{TV}}\left(p_x^{t+s}, p_y^{t+s} \right) \leq d_{\mathsf{TV}}\left(p_x^t, p_y^t \right).$$

Proof. Let $A \subseteq S$ be all the elements $a \in S$ such that $p_x^{t+s}(a) \geq p_y^{t+s}(a)$. By theorem 1.7.13, we have

$$d_{\mathsf{TV}}\left(p_x^{t+s}, p_y^{t+s} \right) = p_x^{t+s}(A) - p_y^{t+s}(A) = \sum_{c \in S} \left(p_x^t(c) - p_y^t(c) \right) p_c^s(A).$$

Let $B \subseteq S$ be all the elements $a \in S$ such that $p_x^t(a) \geq p_y^t(a)$, then

$$d_{\mathsf{TV}}\left(p_x^{t+s}, p_y^{t+s} \right) \leq \sum_{c \in B} \left(p_x^t(c) - p_y^t(c) \right) p_c^s(A) \leq p_x^t(B) - p_y^t(B) = d_{\mathsf{TV}}\left(p_x^t, p_y^t \right),$$

where the second inequality holds by $p_c^s(A) \leq 1$, and the last equality holds by lemma 1.12.2. □

[12] Quite a bit of the unsticking came from reading Daskalakis's https://people.csail.mit.edu/costis/6896sp11/.

1.12.4 Geometric Convergence

Theorem 1.12.29 (Geometric Convergence)**.** *Let* $\mathbf{P}$ *be the transition matrix for a finite, irreducible, aperiodic Markov chain,* m_j *be the smallest entry in the* j^{th} *column of* $\mathbf{P}$*, and* $m = \sum_j m_j$*. Then, for all* $x \in S$ *and* t*,*

$$\Delta_x(t) = d_{\mathsf{TV}}\left(p_x^t, \mathcal{D}_{\boldsymbol{\pi}}\right) \leq (1-m)^t .$$

Proof. Given m_j is the smallest entry in the j^{th} column, then in one step the chain reaches state j with probability at least m_j from every state. Consider a coupling similar to the Optimal Coupling lemma 1.12.23, such that the two copies of the chain both move to state j with probability m_j. Since $\sum_j P_{i,j} = 1$, then $m \leq 1$.

Concretely, a possible construction for $m < 1$ follows

$$\Pr\left[X_{t+1} = x, Y_{t+1} = y \mid X_t = x', Y_t = y'\right] = \begin{cases} m_x & x = y, \\ \dfrac{(P_{x',x} - m_x)(P_{y',y} - m_y)}{1-m} & x \neq y. \end{cases}$$

The chains couple with probability m at a step, and the probability they haven't coupled after t steps is $(1-m)^t$. By lemma 1.12.11, $\Delta_x(t) \leq \Pr[X_t \neq Y_t] = (1-m)^t$. □

A side product for theorem 1.12.29 is the following corollary.

Corollary 1.12.30. *If* $m = 1$ *in theorem 1.12.29, then all rows in* $\mathbf{P}$ *are identical, and* $p_x^t = \mathcal{D}_{\boldsymbol{\pi}}$ *for all* x *and* $t \geq 1$*.*

Proof. Since m_j is the smallest entry in the j^{th} column, then for each column, there is no $P_{i,j}$ that is smaller than m_j.

On the other hand, by $m = \sum_j m_j = 1$ and $\sum_j P_{i,j} = 1$ for any i, there is no m_j that is smaller than $P_{i,j}$.

Therefore, any $m_j = P_{i,j}$ for all i, which proves all rows in $\mathbf{P}$ are identical.

Moreover, when $m_j = p$, then by the cutset, the outflowing probability is $(1-p)\pi_j$, while inflowing probability is $(1-\pi_j)p$, and we have $\pi_j = p$. Since any state reaches state j with probability $P_{i,j} = m_j = p$, this applies to any target state $j \in S$, and hence reaching stationarity is only 1 step. □

Theorem 1.12.31 (Frobenius's Coin Problem)**.** *Given* $m, n \in \mathbb{Z}^+$ *with* $\gcd(m,n) = 1$*, every integer* $k > mn - m - n$ *can be represented by* $am + bn$ *where* $a, b \in \mathbb{N}$*.*

Proof. We first show that $mn - m - n$ cannot be represented by $am + bn$ where $a, b \in \mathbb{N}$. Suppose there are $a, b \in \mathbb{N}^+$, such that $mn - m - n = am + bn$, then

$$(a+1)m = (m-1-b)n. \tag{1.23}$$

But we notice that $mn - m - n < mn$, then $a < n$ and $b < m$. Since m and n are coprime, then there is no a and b such that $a + 1 \leq n$, $m > m - 1 - b \geq 0$, and eq. (1.23) is satisfied.

We now show for every $k > mn - m - n$, there is a non-negative linear combination representation $am + bn$. [13] Since m and n are coprime, then $\{cn\}_{c \in [0, m-1]}$ represents different residue classes modulo m, and $(m-1)n$ is the first element among the linear representation $am + bn$ that

$$(m-1)n \equiv -n \pmod{m}.$$

Hence, for any $q \not\equiv -n \pmod{m}$, there always exists $b \in [0, m-2]$ such that $bn \equiv q \pmod{m}$, and $bn < (m-1)n$. Moreover, the elements $d \in [(m-1)n - m + 1, (m-1)n - 1]$ lie in distinct residue classes, so for each d, there is always $bn \equiv d \pmod{m}$. Since $d < (m-1)n$, then each $d = am + bn$ for some a.

Since every $d \in [(m-1)n - m + 1, (m-1)n]$ is representable, and adding m preserves representability, any integer $k > mn - m - n$ is representable. □

[13] The idea derives from https://math.stackexchange.com/a/66978/794321.

Exercise 1.12.32 (Exercise 12.6 [MU17]). The theorem 1.12.29 is useful only if there exists at least one column j in the transition matrix with $m_j > 0$. Argue that for any finite, aperiodic, irreducible Markov chain, there exists a time T such that every entry of $\mathbf{P}^T$ is nonzero.

Proof. Since the chain is aperiodic, then for any state j, there are 2 paths back to itself of length ℓ_{j_0} and ℓ_{j_1}, such that

$$\gcd\left(\ell_{j_0}, \ell_{j_1}\right) = 1.$$

By theorem 1.12.31, let $t_j = (\ell_{j_0} - 1)(\ell_{j_1} - 1)$, then there always exist loops starting and ending in j of any length $\ell \geq t_j$.

Since the chain is irreducible, then all states are in the same communicating class, such that for any pair of $i \neq j$, there is a time $t_{i,j}$ such that i reaches j, and $P_{i,j}^{t_{i,j}} > 0$.

Given the threshold t_j for self looping path length and $t_{i,j}$ for i to reach j, then there must exist a time

$$T > \max_{i,j}(t_{i,j} + t_j),$$

such that all entries in $\mathbf{P}^T$ are positive. □

A more general result than theorem 1.12.29 is the following: suppose we upper-bound $\tau(c)$ for some constant $c \leq 1/2$, then we can bootstrap a bound for $\tau(\varepsilon)$ for any $\varepsilon > 0$.

Theorem 1.12.33. *Let M_t be a finite, irreducible, aperiodic Markov chain with $\tau(c) \leq T$ for some $c \leq 1/2$, then for M_t,*

$$\tau(\varepsilon) \leq \left\lceil \frac{\ln \varepsilon}{\ln(2c)} \right\rceil T.$$

Proof. The proving strategy is similar to exercise 1.12.26. Since for all initial states $x \in S$, we have

$$\Delta_x(T) = d_{\mathrm{TV}}\left(p_x^T, \mathcal{D}_\pi\right) \leq c,$$

then by the triangle inequality, $d_{\mathrm{TV}}\left(p_x^T, p_y^T\right) \leq 2c$ for all $x, y \in S$, and hence $\overline{\Delta}(T) \leq 2c$.

Therefore, on a pair of initial states x, y, we couple the copies of M_t such that (X_T, Y_T) are optimally coupled by

$$\Pr\left[X_T \neq Y_T\right] = d_{\mathrm{TV}}\left(p_x^T, p_y^T\right)$$

from lemma 1.12.23.

Then we couple the chains such that: if $X_T = Y_T$, they stay together onwards; otherwise, (X_{2T}, Y_{2T}) are optimally coupled conditioned that $(X_T, Y_T) = (x', y')$ by

$$\Pr\left[X_{2T} \neq Y_{2T} \mid X_T = x', Y_T = y'\right] = d_{\mathrm{TV}}\left(p_{x'}^T, p_{y'}^T\right).$$

By induction, we have

$$\Pr\left[X_{kT} \neq Y_{kT} \mid X_0 = x, Y_0 = y\right] \leq (2c)^k,$$

then $\Delta_x(kT) \leq (2c)^k$ for all $x \in S$ by lemma 1.12.10. If $(2c)^k = \varepsilon$, then

$$k \leq \left\lceil \frac{\ln \varepsilon}{\ln(2c)} \right\rceil,$$

and we have $\tau(\varepsilon) \leq kT$ by lemma 1.12.10. □

1.12.5 Application in Approximately Sampling Proper Colorings

We now present a Markov chain Monte Carlo (MCMC) process that generates an almost uniformly at random sample of a proper coloring of a graph, then use a coupling technique to show that it is rapidly mixing.

The Markov chain on proper coloring works as follows: At each step, choose a vertex v and a color ℓ uniformly at random. Recolor vertex v with color ℓ if the new coloring is proper, that v is not neighboring a vertex colored ℓ, or otherwise let the state of the chain be unchanged.

The Markov chain is aperiodic, as there exists loop of length 1 back to itself; and it is irreducible if $c \geq \Delta + 2$, as recoloring in any order can be unblocked by recoloring a later unrecolored neighbor to an unconflicting color: the degree of the graph is at most Δ, then $|\Gamma^+(v)| \leq \Delta + 1$ for every vertex, such that the recoloring is always unblocked. Therefore, the Markov chain is ergodic and has stationary distribution being uniform distribution.

When $c \geq 4\Delta + 1$, we can use a trivial coupling for (X_t, Y_t) by choosing the same vertex and color at each step, allowing us to almost uniformly sample colorings efficiently.

Theorem 1.12.34. *For an n-vertex graph with maximum degree Δ, the mixing time of the graph-coloring Markov chain satisfies*

$$\tau(\varepsilon) \leq \left\lceil \frac{nc}{c - 4\Delta} \ln\left(\frac{n}{\varepsilon}\right) \right\rceil,$$

provided that $c \geq 4\Delta + 1$.

Proof. The proof strategy is similar to example 1.12.17. Fix initial states x, y be distinct c-colorings. Let D_t be the set of vertices with different colors in the two chains at time t, and let $d_t = |D_t|$.

Consider any vertex $v \in D_t$, but colored same after a move. There are at least $c - 2\Delta$ colors that have not appeared on the at most 2Δ neighbors among 2 chains, then

$$\Pr[d_{t+1} = d_t - 1 \mid d_t > 0] \geq \frac{d_t}{n} \cdot \frac{c - 2\Delta}{c}.$$

Consider any vertex $v \in V \setminus D_t$, but colored differently after a move. It has to be the case that the move succeeded in one chain, but not the other. Moreover, v must have at least one neighbor in D_t, such that the move's color may work in one chain, but not the other. Hence, for every $w \in D_t$, it can affect at most Δ neighbors with 2 colors, and

$$\Pr[d_{t+1} = d_t + 1 \mid d_t > 0] \leq \frac{\Delta d_t}{n} \cdot \frac{2}{c}.$$

By the linearity of expectation,

$$\mathbb{E}[d_{t+1} \mid d_t] \leq d_t + \frac{\Delta d_t}{n} \cdot \frac{2}{c} - \frac{d_t}{n} \cdot \frac{c - 2\Delta}{c} = d_t \cdot \left(1 - \frac{c - 4\Delta}{nc}\right),$$

then by the Markov inequality argument,

$$\Pr[d_t \geq 1] \leq \mathbb{E}[d_t] \leq \mathbb{E}[d_0] \cdot \left(1 - \frac{c - 4\Delta}{nc}\right)^t \leq n\left(1 - \frac{c - 4\Delta}{nc}\right)^t \leq n \exp\left(-\frac{c - 4\Delta}{nc} \cdot t\right).$$

The proof is immediate by lemma 1.12.10. □

We then show how to improve the coupling to reduce the number of colors necessary to $2\Delta + 1$. On one hand, for $v \in D_t$, if some neighboring vertices are same colored in both chains, then the color options should be much greater than $c - 2\Delta$. On the other hand, for $v \in V \setminus D_t$, we can decrease the number of bad moves of increasing differences by a careful coupling.

Theorem 1.12.35. *For an n-vertex graph with maximum degree Δ, the mixing time of the graph-coloring Markov chain satisfies*

$$\tau(\varepsilon) \leq \left\lceil \frac{nc}{c - 2\Delta} \ln\left(\frac{n}{\varepsilon}\right) \right\rceil,$$

provided that $c \geq 2\Delta + 1$.

Proof. The proof strategy is similar to exercise 1.12.20.

We reuse all notions in theorem 1.12.34, and let $A_t = V \setminus D_t$ be the set of vertices that are colored same in both chains. For $v \in A_t$, $d'(v) = |\Gamma(v) \cap D_t|$, that is the number of neighboring vertices colored differently in both chains; for $w \in D_t$, $d'(w) = |\Gamma(w) \cap A_t|$, that is the number of neighboring vertices colored same in both chains.

Hence, we explore the invariant that

$$m = \sum_{v \in A_t} d'(v) = \sum_{w \in D_t} d'(w),$$

which can be viewed as the number of edges between vertices in D_t and vertices in A_t.

For $v \in D_t$, a move on v sends v to A_{t+1} if both chains recolor v to the same valid color. Taking into account the neighbors of v that are colored the same in both chains, the forbidden color options for v have union size at most $2\Delta - d'(v)$, since those $d'(v)$ neighbors are double-counted. Therefore,

$$\Pr\left[d_{t+1} = d_t - 1 \mid d_t > 0\right] \geq \frac{1}{n} \sum_{v \in D_t} \frac{c - 2\Delta + d'(v)}{c} = \frac{1}{n} \cdot \frac{d_t(c - 2\Delta) + m}{c}.$$

For $w \in A_t$, but colored distinctly after move by $w \in D_{t+1}$, consider the coupling in a similar way in exercise 1.12.20.

Suppose in the particular case where $d'(w) = 1$, then instead of assigning w the same color and causing 2 potential bad moves, on c_1 and c_2 being the disagreeing colors of neighbor v of w, if w is assigned one of the two colors in X_t, assign the other color to w in Y_t. In this case, 2 bad moves are halved down to 1, as either they both fail, or they both move through and cause 1 more difference.

Moving to the general case. Let $S_1(w)$ be the colors of neighboring vertices of w over X_t that are in D_t, and $S_2(w)$ be the colors of neighboring vertices of w over Y_t that are in D_t. Using the same strategy in exercise 1.12.20, we pair the distinct colors between $S_1(w)$ and $S_2(w)$ as much as possible: if $c_1 \in S_1(w)$ is chosen for one chain, $c_2 \in S_2(w)$ is chosen for the other; otherwise they are mapped to the same color. In this way, the bad moves are reduced down to $\max\left(|S_1(w)|, |S_2(w)|\right) \leq d'(w)$. As a result, we have

$$\Pr\left[d_{t+1} = d_t + 1 \mid d_t > 0\right] \leq \frac{1}{n} \sum_{w \in A_t} \frac{d'(w)}{c} = \frac{m}{cn}.$$

By the linearity of expectation,

$$\mathbb{E}\left[d_{t+1} \mid d_t\right] \leq d_t + \frac{m}{cn} - \frac{1}{n} \cdot \frac{d_t(c - 2\Delta) + m}{c} = d_t \cdot \left(1 - \frac{c - 2\Delta}{nc}\right),$$

then by the Markov inequality argument,

$$\Pr\left[d_t \geq 1\right] \leq \mathbb{E}\left[d_t\right] \leq \mathbb{E}\left[d_0\right] \cdot \left(1 - \frac{c - 2\Delta}{nc}\right)^t \leq n \cdot \exp\left(-\frac{c - 2\Delta}{nc} \cdot t\right).$$

The proof is immediate by lemma 1.12.10. □

Remark 1.12.36. By definition 1.12.6, the theorem 1.12.34 and theorem 1.12.35 are rapidly mixing.

Remark 1.12.37. The theorem 1.12.34 and theorem 1.12.35 are from [Jer95].

1.12.6 Path Coupling

The motivation of using "path coupling" is that, it starts with coupling between pairs of states differ in small ways, then generalize to arbitrary pairs of states. It may be easier to construct coupling of adjacent states, and any pairs of states can be connected with a path, then we couple all pairs of states along the path, such that the expected distance contracts at each step; by summing along the path, this implies contraction for any pairs of states.

Similar to example 1.12.17, exercise 1.12.20, theorem 1.12.34, and theorem 1.12.35, we conclude by utilizing the expected-distance bound for bounding disagreement probability via a Markov inequality argument, hence bounding the mixing-time with lemma 1.12.10. Eventually, this makes coupling argument easier to be established.

Example 1.12.38 (Markov Chain for Independent Sets)**.** Consider a Markov chain for sampling independent sets over a graph with $\Delta \leq 4$. Let X_t be the independent set at time t. At a step, the chain samples $(u, v) \xleftarrow{R} E$, and

- With probability 1/3, $X_{t+1} = X_t \setminus \{u, v\}$.
- With probability 1/3, let $Y = X_t \setminus \{u\} \cup \{v\}$. If Y is an independent set, $X_{t+1} = Y$; otherwise, $X_{t+1} = X_t$.
- With probability 1/3, let $Y = X_t \setminus \{v\} \cup \{u\}$. If Y is an independent set, $X_{t+1} = Y$; otherwise, $X_{t+1} = X_t$.

Similar to exercise 1.12.16, the stationary distribution of X_t is uniform over all independent sets: By the cutset theorem 1.7.56, the total probability of adjacent states transitioning to the current state is the same as the total probability that the current state transitioning to adjacent states. Since $P_{a,b} = (3|E|)^{-1}$ for any pair of adjacent states (a, b), then uniform distribution is a trivial solution.

We start by coupling a pair of states (X_t, Y_t) that differ by just one vertex, and this coupling can be extended to all pairs of states. Let $Z_t = (X_t \setminus Y_t) \cup (Y_t \setminus X_t)$, and we say a vertex v is bad if $v \in Z$; otherwise v is good. Let $d_t = |Z_t|$, then d_t is the random variable counting bad vertices.

We continue by assuming $d_t = 1$, and WLOG let $X_t = I$, $Y_t = I \cup \{x\}$. We apply a simple coupling by performing the same move on both chains, and show $\mathbb{E}[d_{t+1} \mid d_t] \leq d_t$ when $d_t = 1$, or equivalently, $\mathbb{E}[d_{t+1} - d_t \mid d_t = 1] \leq 0$.

Since a change in d_t is caused only by a move involving x or the neighbors of x, we restrict our attention to the edges sampled adjacent to x, neighbors of x, or neighbors of neighbors of x [14]. Let δ_z be a random variable over $\{-1, 1\}$ for a vertex z between time t and $t+1$: $\delta_z = 1$ when z goes from good to bad, and $\delta_z = -1$ when z goes from bad to good. By linearity of expectation,

$$\mathbb{E}[d_{t+1} - d_t \mid d_t = 1] = \mathbb{E}\left[\sum_{w \in \Gamma^+(\Gamma^+(x))} \delta_w \,\middle|\, d_t = 1\right] = \sum_{w \in \Gamma^+(\Gamma^+(x))} \mathbb{E}[\delta_w \mid d_t = 1].$$

We shall show that when $\Delta \leq 4$, $\mathbb{E}[\delta_w \mid d_t = 1] \leq 0$. For each vertex $y \in \Gamma(x)$:

- If $\left|\Gamma(y) \cap I\right| \geq 2$, only moves on edge (x, y) changes $d_{t+1} - d_t$, either removing y and adding x, or removing $\{x, y\}$. With probability $2 \cdot (3|E|)^{-1}$, if the (x, y) edge is chosen, then $d_{i+1} - d_i$ decreases by 1.
- If $\left|\Gamma(y) \cap I\right| = 0$, for moves on edges with y,
  - For increasing $d_{t+1} - d_t$, for $w \in \Gamma(y) \setminus \{x\}$, removing w and adding x works on X_t but fails on Y_t. Since $\left|\Gamma(y) \setminus \{x\}\right| \leq 3$ by $\Delta \leq 4$, by union bound, $d_{t+1} - d_t$ increases with probability at most $3 \cdot (3|E|)^{-1} = |E|^{-1}$.
  - For decreasing $d_{t+1} - d_t$, any option among 3 possible moves on (x, y) decreases bad vertices. With probability $|E|^{-1}$, if the (x, y) edge is chosen, then $d_{t+1} - d_t$ decreases by 1.

  Hence, for moves on edges with y, $\mathbb{E}[d_{t+1} - d_t \mid d_t = 1] \leq 0$.
- If $\left|\Gamma(y) \cap I\right| = 1$, and let $z \in I \cap \Gamma(y)$, for moves on edges with y,
  - For increasing $d_{t+1} - d_t$, removing z and adding y works on X_t but not Y_t. With probability $(3|E|)^{-1}$, $d_{t+1} - d_t$ increases by 2 if (y, z) edge is chosen and removes z and adds y.
  - For decreasing $d_{t+1} - d_t$, both adding x and removing y, and removing $\{x, y\}$ works. With probability $2 \cdot (3|E|)^{-1}$, if the (x, y) edge is chosen, then $d_{t+1} - d_t$ decreases by 1.

  Hence, for move on edge with y, $\mathbb{E}[d_{t+1} - d_t \mid d_t = 1] = 0$.

[14] Self noting: need to consider the case where an edge is chosen with one end in I, and the other is a neighbor of x.

Since for every pair of edge, we have

$$\sum_{w\in\Gamma^+(\Gamma^+(x))} \mathbb{E}\left[\delta_w \mid d_t = 1, \text{move on edge } (a,b)\right] \le 0,$$

as was shown in the previous discussion of $y \in \Gamma(x)$, then

$$\begin{aligned}\mathbb{E}\left[d_{t+1} - d_t \mid d_t = 1\right] &= \sum_{w\in\Gamma^+(\Gamma^+(x))} \mathbb{E}\left[\delta_w \mid d_t = 1\right] \\ &= \frac{1}{|E|} \sum_{w\in\Gamma^+(\Gamma^+(x))} \sum_{(a,b)\in E} \mathbb{E}\left[\delta_w \mid d_t = 1, \text{move on edge } (a,b)\right] \\ &= \frac{1}{|E|} \sum_{(a,b)\in E} \sum_{w\in\Gamma^+(\Gamma^+(x))} \mathbb{E}\left[\delta_w \mid d_t = 1, \text{move on edge } (a,b)\right] \le 0.\end{aligned}$$

We now generalize the coupling for $d_t > 1$ case. To couple (X_t, Y_t) where $d_t > 1$, we create a chain of states

$$Z_0 = X_t, \ldots, Z_{d_t} = Y_t,$$

where neighboring pair of states (Z_{i-1}, Z_i) has $|Z_{i-1} \setminus Z_i| + |Z_i \setminus Z_{i-1}| = 1$, that each successive Z_i is obtained from Z_{i-1} by either removing an element in $X_t \setminus Y_t$ or adding an element in $Y_t \setminus X_t$.

Our coupling arises as follows. When a move is made in $Z_0 = X_t$, the coupling for $d_t = 1$ case gives a move for the state Z_1. This move in Z_1 can similarly be coupled with a move in state Z_2, and so on, until the move in Z_{d_t-1} yields a move for $Z_{d_t} = Y_t$. Let Z_i' be the state after the move is made from state Z_i, and define

$$\Delta\left(Z_{i-1}', Z_i'\right) = \left|Z_{i-1}' \setminus Z_i'\right| + \left|Z_i' \setminus Z_{i-1}'\right|.$$

Since $\mathbb{E}\left[d_{i+1} - d_i \mid d_i = 1\right] \le 0$, then

$$\mathbb{E}\left[\Delta\left(Z_{i-1}', Z_i'\right)\right] \le 1.$$

Moreover, note that $(Z_0', Z_{d_t}') = (X_{t+1}, Y_{t+1})$, then by the triangle inequality for sets

$$(A \setminus C) \cup (C \setminus A) \subseteq (A \setminus B) \cup (B \setminus C) \cup (B \setminus A) \cup (C \setminus B),$$

as for any element $a \in A \setminus C$, either $a \in B$, then $a \in B \setminus C$; or $a \notin B$, then $a \in A \setminus B$. Then

$$\Delta\left(Z_0', Z_{d_t}'\right) = \left|Z_0' \setminus Z_{d_t}'\right| + \left|Z_{d_t}' \setminus Z_0'\right| \le \sum_{i\in[1,d_t]} \left|Z_{i-1}' \setminus Z_i'\right| + \sum_{i\in[1,d_t]} \left|Z_i' \setminus Z_{i-1}'\right|,$$

and by linearity of expectation,

$$\mathbb{E}\left[d_{t+1} \mid d_t\right] = \mathbb{E}\left[\Delta\left(Z_0', Z_{d_t}'\right)\right] \le \sum_{i\in[1,d_t]} \mathbb{E}\left[\Delta\left(Z_{i-1}', Z_i'\right)\right] \le d_t,$$

and eventually

$$\mathbb{E}\left[d_{t+1}\right] = \mathbb{E}\left[\mathbb{E}\left[d_{t+1} \mid d_t\right]\right] \le \mathbb{E}\left[d_t\right].$$

Suppose $\mathbb{E}\left[d_{t+1} \mid d_t\right] \le \beta d_t$ for $\beta < 1$, the rest is taken over by lemma 1.12.10, given a sufficient number of steps and a Markov inequality argument.

Remark 1.12.39. The example 1.12.38 derives from [BD97].

The example 1.12.38 illustrates the philosophy of path coupling: global distance is controlled by understanding local disagreements. Rather than constructing a coupling for arbitrary pairs of states, we first choose a pre-metric on the state space, whose edges represent elementary discrepancies between configurations.

In example 1.12.38, the natural choice is to connect two states when they differ at a vertex, so that the induced path metric is Hamming distance. Once such a pre-metric is fixed, any pair of states can be joined by a shortest path of discrepancies, and contraction on adjacent pairs propagates to contraction for all pairs. Thus the pre-metric is the structure that makes the path-coupling argument possible: it identifies the right local geometry of the state space.

Definition 1.12.40 (Pre-Metric)**.** A pre-metric on Ω is a positively-weighted connected undirected graph such that all edges are shortest paths. More specifically, if (x, y) are adjacent in the pre-metric, the weight of the edge should be the least weight of any path from x to y in the pre-metric.

Theorem 1.12.41 (Path Coupling Theorem [BD97])**.** *Suppose there exists a coupling defined for all adjacent pair of states in the pre-metric such that for all adjacent* (X, Y)*,*

$$\mathbb{E}\left[d\left(X', Y'\right) \mid X, Y\right] \leq (1-\alpha)\, d\left(X, Y\right),$$

where d is the shortest path metric, then this coupling can be extended to a coupling between all pair of states that also satisfy the contraction inequality.

Exercise 1.12.42 (Exercise 12.7 [MU17])**.** Continuing from example 1.12.38, we have shown there is $\beta < 1$ such that

$$\mathbb{E}\left[d_{t+1} \mid d_t\right] \leq \beta d_t.$$

Let d^* be the maximum distance over all possible pairs of initial states for the coupling.

- Give an upper bound for $\tau(\varepsilon)$ in terms of β and d^*.
- Suppose we have $\mathbb{E}\left[d_{t+1} \mid d_t\right] \leq d_t$, $d_{t+1} \in [d_t - 1, d_t + 1]$, and $\Pr\left[d_t \neq d_{t+1}\right] \geq \gamma$.
  Give an upper bound for $\tau(\varepsilon)$ in terms of β, d^*, and γ.
- Show the mixing time of the graph coloring chain in theorem 1.12.34 and theorem 1.12.35, is $\mathsf{poly}(|V|, \ln(1/\varepsilon))$, even when the number of colors is only 2Δ.
- Show the mixing time of the Markov chain for independent sets in example 1.12.38 is $\mathsf{poly}(|V|, \ln(1/\varepsilon))$.

Proof. Suppose we have the $\beta < 1$, that $\mathbb{E}\left[d_{t+1} \mid d_t\right] \leq \beta d_t$, then

$$\mathbb{E}\left[d_{t+1}\right] = \mathbb{E}\left[\mathbb{E}\left[d_{t+1} \mid d_t\right]\right] \leq \beta \cdot \mathbb{E}\left[d_t\right].$$

Since d^* is the maximum distance over all possible pairs of states for the coupling, then

$$\mathbb{E}\left[d_t\right] \leq \beta^t \cdot d^*.$$

By the same Markov inequality argument, we have

$$\Pr\left[d_t \geq 1\right] \leq \mathbb{E}\left[d_t\right] \leq \beta^t \cdot d^*.$$

Supposing $\varepsilon = \beta^t \cdot d^*$, then

$$\tau\left(\varepsilon\right) \leq \left\lceil \frac{\ln \varepsilon - \ln d^*}{\ln \beta} \right\rceil.$$

If $\Pr\left[d_{t+1} \neq d_t\right] \geq \gamma$, then let $p^+ + p^- = \Pr\left[d_{t+1} \neq d_t\right]$, where $p^+ = \Pr\left[d_{t+1} = d_t + 1\right]$, and $p^- = \Pr\left[d_{t+1} = d_t - 1\right]$. Consider d_t as a Markov chain for random walking over $[0, d^*]$, where $d_0 = d^*$, state 0 is the absorbing state, and the transition probability from state d^* to state $d^* - 1$ is $p^+ + p^- = \Pr\left[d_{t+1} \neq d_t\right]$. Let $h_{i,0}$ be the expected number of steps to be absorbed in 0, then we have the following linear system

$$\begin{aligned}
(p^+ + p^-)h_{1,0} &= 1 + p^+ h_{2,0} + p^- h_{0,0} \\
(p^+ + p^-)h_{2,0} &= 1 + p^+ h_{3,0} + p^- h_{1,0} \\
&\cdots \\
(p^+ + p^-)h_{d^*-1,0} &= 1 + p^+ h_{d^*,0} + p^- h_{d^*-2,0} \\
(p^+ + p^-)h_{d^*,0} &= 1 + p^+ h_{d^*-1,0} + p^- h_{d^*-1,0},
\end{aligned}$$

where $h_{0,0} = 0$.

The chain is lazy, as it stays put with probability at most $1-\gamma$. We continue by deriving the expected number of active moves. Let $q = \Pr[d_{i+1} \neq d_i]$, then conditioned that a move is made,

$$\begin{aligned} q^+ &= \Pr[d_{i+1} = d_i + 1 \mid d_{i+1} \neq d_i] = p^+/q \\ q^- &= \Pr[d_{i+1} = d_i - 1 \mid d_{i+1} \neq d_i] = p^-/q, \end{aligned}$$

and let $g_{i,0}$ be the expected number of active moves, then

$$\begin{aligned} g_{1,0} &= 1 + q^+ g_{2,0} + q^- g_{0,0} \\ g_{2,0} &= 1 + q^+ g_{3,0} + q^- g_{1,0} \\ &\dots \\ g_{d^*-1,0} &= 1 + q^+ g_{d^*,0} + q^- g_{d^*-2,0} \\ g_{d^*,0} &= 1 + g_{d^*-1,0}, \end{aligned}$$

where $g_{0,0} = h_{0,0} = 0$. Since $q^+ + q^- = 1$, and $q^- \geq 1/2$, then consider a Markov chain with $q^+ = q^- = 1/2$, and

$$g_{i,0} = (2d^* - i)i,$$

and therefore $g_{d^*,0} = (d^*)^2$. Since $q^- \geq 1/2$, the real $g_{d^*,0} \leq (d^*)^2$.

Since $\Pr[d_{t+1} \neq d_t] \geq \gamma$, then the expected wait time before an active move is at most γ^{-1}, and therefore

$$h_{d^*,0} \leq \gamma^{-1} (d^*)^2 .$$

Hence, when $L = \left\lceil 2\gamma^{-1} (d^*)^2 \right\rceil$, let T be the random variable for the steps before being absorbed, by Markov inequality

$$\Pr[T \geq L] \leq \frac{\mathbb{E}[T]}{L} \leq \frac{1}{2}.$$

Similar to exercise 1.12.19, by the Markov property, $\Pr[T \geq kL] \leq 2^{-k}$. Then when $\varepsilon = 2^{-k}$,

$$\tau(\varepsilon) \leq \left\lceil \log \frac{1}{\varepsilon} \right\rceil L. \tag{1.24}$$

For path coupling in the Markov chain sampling graph coloring, we first show a pre-metric over all valid coloring, such that any valid pair of adjacent valid coloring differ at the color of a single vertex. Supposing that X_0 and Y_0 are disagreeing only at v, let c_1 and c_2 be the colors of v at X_0 and Y_0, and apply the prior coupling in theorem 1.12.35: if we change $w \in \Gamma(v)$ in X_t to one of the colors in $\{c_1, c_2\}$, we change w in Y_t to the other color; otherwise, apply the same move on both chains.

One thing to beware is, in theorem 1.12.34 and theorem 1.12.35, the distance function d_t is measuring the number of disagreeing vertices, while the current d_t is measuring the minimal number of steps to transition from one valid coloring to the other, as the pre-metric is defined on Ω of valid colorings. Therefore, after a bad move in the coupling, where v is c_1, w is c_2 in X_t, while v is c_2, w is c_1 in Y_t, the d_t is 3 rather than 2 [15].

We continue bounding $\mathbb{E}[d_{t+1} \mid d_t = 1]$. Since v can be corrected with at least $c - \Delta$ colors, then

$$\Pr[d_{t+1} = d_t - 1 \mid d_t = 1] \geq \frac{1}{n} \cdot \frac{c - \Delta}{c}.$$

On the other hand, a bad move happens only on $\Gamma(v)$ with 1 color configuration on both chains, hence

$$\Pr[d_{t+1} = d_t + 2 \mid d_t = 1] \leq \frac{\Delta}{n} \cdot \frac{1}{c}.$$

[15] Self noting: check Daskalakis's slides https://people.csail.mit.edu/costis/6896sp11/lec7s.pdf.

Therefore,

$$\mathbb{E}\left[d_{t+1} \mid d_t = 1\right] \le 1 + 2 \cdot \frac{\Delta}{n} \cdot \frac{1}{c} - \frac{1}{n} \cdot \frac{c - \Delta}{c} = 1 - \frac{c - 3\Delta}{nc} = \beta.$$

Now let the maximal distance be d^*, and construct a path joined by adjacent valid states over Ω with the pre-metric by $Z_0 = X_t$ and $Z_{d_t} = Y_t$. When a move is made on Z_i, the move from coupling is made on Z_{i+1}, and let Z_i' be the state after the move is made on state Z_i. By linearity of expectation,

$$\mathbb{E}\left[d_{t+1} \mid d_t\right] = \mathbb{E}\left[\Delta\left(Z_0', Z_{d_t}'\right)\right] \le \sum_{i \in [1, d_t]} \mathbb{E}\left[\Delta\left(Z_{i-1}', Z_i'\right)\right] \le d_t \cdot \beta,$$

where $\Delta(Z_{i-1}', Z_i')$ is defined as the distance between Z_{i-1}' and Z_i' in the pre-metric, and finally,

$$\mathbb{E}\left[d_t\right] = \mathbb{E}\left[\mathbb{E}\left[d_t \mid d_{t-1}\right]\right] \le \mathbb{E}\left[d_{t-1}\right] \cdot \beta \le d^* \cdot \beta^t,$$

which gives a similar result to theorem 1.12.34 and theorem 1.12.35 but $c \ge 3\Delta + 1$.

Now we use Hamming distance as pre-metric: the states along the path are not all valid states, yet it still serves as a path coupling. Let $\tilde{\Omega}$ be the set for all possible coloring states, and $\Omega \subseteq \tilde{\Omega}$, then any adjacent states (X_0, Y_0) are still differing at the color of a vertex, but they are not necessarily valid states.

A move is made by choosing a vertex v and a color c, if c is not appearing in the neighboring $\Gamma(v)$. One observation is, if $X_0 \in \Omega$, then the chain stays in Ω; otherwise, the chain will turn into Ω at some time step, and it will stay in Ω. Hence, coupling any pair of $(X_0, Y_0) \in \Omega^2$ over the enlarged state space $\tilde{\Omega}$ still preserves the marginal probabilities, while the path may pass through $\tilde{\Omega} \setminus \Omega$ does not hurt the argument, because the path is only used to define the metric for the path coupling theorem.

Applying the same coupling in theorem 1.12.35, conditioned that (X_0, Y_0) are adjacent in the Hamming metric, then on the bad move, d_t increases by 1. Therefore, with the same probability bounds on distance changing,

$$\mathbb{E}\left[d_{t+1} \mid d_t = 1\right] \le 1 - \frac{c - 2\Delta}{nc} = \beta,$$

and by the same path coupling argument, $\mathbb{E}\left[d_t\right] \le \mathbb{E}\left[d_{t-1}\right] \cdot \beta$ gives the same result as theorem 1.12.35 for $c \ge 2\Delta + 1$.

For $c = 2\Delta$, $\mathbb{E}\left[d_{t+1} \mid d_t\right] \le d_t$. We can loosely lower bound by

$$\Pr\left[d_{t+1} \ne d_t\right] \ge \gamma = (nc)^{-1},$$

as there is at least 1 move fixing a disagreeing vertex, then the eq. (1.24) is $\operatorname{poly}\left(|V| = n, \log\left(1/\varepsilon\right)\right)$.

For independent set sampling in example 1.12.38, we can also loosely lower bound by

$$\Pr\left[d_{t+1} \ne d_t\right] \ge \gamma = (3n)^{-1},$$

as a disagreeing vertex v can be removed by choosing an edge (u, v), and removing $\{u, v\}$. By eq. (1.24), the mixing time bound is $\operatorname{poly}\left(|V| = n, \log\left(1/\varepsilon\right)\right)$. □

Exercise 1.12.43 (Exercise 12.16 [MU17]). Consider the Markov chain sampling independent set in a graph $G = (V, E)$. On a step, sample $v \xleftarrow{R} V$ and $b \xleftarrow{R} \{0, 1\}$. If $v \in X_t$, $X_{t+1} = X_t \setminus \{v\}$ on $b = 1$, if $v \notin X_t$ and $X_t \cup \{v\}$ is independent, $X_{t+1} = X_t \cup \{v\}$ on $b = 0$; otherwise, $X_{t+1} = X_t$.

- Show that the stationary distribution of the chain is uniform over all independent sets.
- Consider the chain over cycles and line graphs, where a line graph over n vertices is constructed by connecting vertex $1, \ldots, n$ with an edge, and a cycle graph is the same with an addition of an additional edge $(1, n)$.

  Devise a coupling (X_t, Y_t) for the Markov chain on line graphs and cycle graphs: if $d_t = |X_t \setminus Y_t| + |Y_t \setminus X_t|$, then at each step the coupling is at least as likely to reduce d_t as to increase d_t. Show that it is rapidly mixing.
- For the cases of line graphs and cycle graphs, derive the exact formulas for the number of independent sets.

Proof. The stationary argument is similar to the one in example 1.12.38: for any pair of adjacent state of independent sets, which are differing by a vertex, the transition probability $P_{a,b} = (2n)^{-1}$, then uniform distribution is a trivial solution for stationary distribution.

We use Hamming metric as the pre-metric: though we use the enlarged state space $\tilde{\Omega}$ of all subsets of V, for the initial states $(X_0, Y_0) \in \Omega^2$, the coupling Markov chains still preserve their marginal probabilities, hence the coupling is still valid. This follows the same argument in exercise 1.12.42. Now consider the same coupling by taking on same sampled v and b on both (X_t, Y_t) chains, where (X_t, Y_t) are adjacent states. WLOG let w be the disagreeing vertex, $X_t = I$, and $Y_t = I \cup \{w\}$. We are interested in $v \in \Gamma^+(w)$, as $v \in V \setminus \Gamma^+(w)$ does not change d_t.

- If $v = w$, any b decreases the difference, and $\mathbb{E}[d_{t+1} - d_t \mid d_t = 1, v = w] = -1$.
- If $v \in \Gamma(w)$, then by $\Delta \leq 2$, we know $|\Gamma(v) \cap I| \leq 1$.
  - If $|\Gamma(v) \cap I| = 0$, then the distance is unchanged when $b = 1$, and the distance increases by 1 when $b = 0$. Hence, conditioned that $v \in \Gamma(w)$ and $|\Gamma(v) \cap I| = 0$, $\mathbb{E}[d_{t+1} - d_t \mid d_t = 1] = 1/2$.
  - If $|\Gamma(v) \cap I| = 1$, then no operation changes d_t. Conditioned that $v \in \Gamma(w)$ and $|\Gamma(v) \cap I| = 1$, $\mathbb{E}[d_{t+1} - d_t \mid d_t = 1] = 0$.

Moreover, since $\Delta \leq 2$, then $|\Gamma(w)| \leq 2$. Let X be the random variable for the vertex to choose, then $\Pr[X = v] = n^{-1}$. By linearity of expectation,

$$\mathbb{E}[d_{t+1} - d_t \mid d_t = 1] = \mathbb{E}[\mathbb{E}[d_{t+1} - d_t \mid d_t = 1, X]] \leq 0.$$

With the same trick from exercise 1.12.42, we loosely lower bound by $\Pr[d_{t+1} \neq d_t] \geq (2n)^{-1} = \gamma$, as there is a choice out of $2n$ choices to fix the disagreement at w, then eq. (1.24) is $\text{poly}\left(|V| = n, \log(1/\varepsilon)\right)$.

For the number of independent sets in line graphs and cycle graphs, this is basically dynamic programming. Let L_i be the number of independent sets of line graph with i vertices. By including the first vertex or not, we have

$$L_i = L_{i-1} + L_{i-2}.$$

Let C_i be the number of independent sets of cycle graph with i vertices. If the first vertex is not chosen, then there are L_{i-1} independent sets, while if the first vertex is chosen, then there are L_{i-3} independent sets, hence

$$C_i = L_{i-1} + L_{i-3}.$$

□

Exercise 1.12.44 (Exercise 12.18 [MU17]). Use path coupling to simplify example 1.12.17 and exercise 1.12.20.

Proof. We enhance the result in exercise 1.12.16, by showing that the Markov chain over fixed size independent sets of size $k \leq n/(2\Delta + 2)$ is ergodic. Let $I = X_t \cap Y_t$, $A_t = X_t \setminus I$, and $B_t = Y_t \setminus I$. Consider following 2 easier corner case:

- If $|B_t| = |B_t \setminus \Gamma(A_t)|$, then $B_t \cap \Gamma(X_t) = \varnothing$, and the transition takes $|A_t| = |B_t|$ steps by moving from A_t to B_t.
- If $|B_t| = |B_t \cap \Gamma(A_t)|$, since $|A_t| \leq |X_t| \leq k$, then $|V \setminus \Gamma^+(X_t)| \geq n/2$, and B_t can move to $V \setminus \Gamma^+(X_t)$. The transition can take $2|A_t|$ steps by moving A_t to an independent set in $V \setminus \Gamma^+(X_t)$, then moving back into B_t.

In general case, X_t can be broken into 3 parts: I, $A_t \cap \Gamma(B_t)$, and $A_t \setminus \Gamma(B_t)$. Since $|A_t| \leq k$ and $|V \setminus \Gamma^+(X_t)| \geq n/2$, then there exists a way to move A_t to $V \setminus \Gamma^+(X_t)$ in $|A_t|$ steps, such that $B_t \setminus \Gamma(A_t)$ are all occupied. The remaining steps are transitioning occupied vertices back into $B_t \setminus \Gamma(A_t)$.

The uniform distribution being stationary distribution argument is immediate by the cutset, that the neighboring pair of independent sets transition to each other with transition probability $1/kn$ everywhere, and the chain has odd length loop of going back to itself, hence it is aperiodic.

Let D_t be the pre-metric over all valid k-sized independent sets, which is the least number of steps to transition from X_t to Y_t. Conditioned that $D_t = 1$, let $X_t = I \cup \{u\}$ and $Y_t = I \cup \{v\}$. At a step, sample $w \overset{R}{\leftarrow} X_t$ and $x \overset{R}{\leftarrow} V$, and a transition happens when $w \in X_t$ and $\Gamma(x) \cap X_t = \varnothing$.

If $w \neq u$, $w_x = w_y = w$; otherwise $(w_x, w_y) = (u, v)$. Let $S_1 = \Gamma^+(u)$, $S_2 = \Gamma^+(v)$, and M be the mapping between S_1 and S_2, such that elements in $S_1 \cap S_2$ are mapped to themselves, elements are mapped bijectively between $S_1 \setminus S_2$ and $S_2 \setminus S_1$ as many as possible, bijectively map u and v, with the remaining mapped to itself. Hence, if $x \in S_1 \cup S_2$, $z_x = x$ and $z_y = M(x)$; otherwise, $z_x = z_y = x$.

The transition step is coupled by: Try removing w_x and adding z_x to X_t, and removing w_y and adding z_y to Y_t. To decrease the distance, $(w_x, w_y) = (u, v)$ and $x \notin \Gamma^+(X_t \cup Y_t \setminus \{u, v\})$, then the probability is lower bounded by

$$\Pr[D_{t+1} = D_t - 1 \mid D_t = 1] \geq \frac{1}{k} \cdot \frac{n - (\Delta + 1)(k - 1)}{n}.$$

To increase the distance, $w \in I$ and $z_x \in S_2$, such that the number of moves where at least a chain transitions is upper bounded by $\max(|S_1|, |S_2|) \leq \Delta + 1$ by exercise 1.12.20, hence the probability is upper bounded by

$$\Pr[D_{t+1} = D_t + 1 \mid D_t = 1] \leq \frac{k - 1}{k} \cdot \frac{\Delta + 1}{n}.$$

By linearity of expectation,

$$\mathbb{E}[D_{t+1} \mid D_t = 1] \leq 1 - \frac{1}{k} \cdot \frac{n - (\Delta + 1)(k - 1)}{n} + \frac{k - 1}{k} \cdot \frac{\Delta + 1}{n} = 1 - \frac{n - 2(\Delta + 1)(k - 1)}{kn}.$$

Again by path coupling argument, let

$$Z_0 = X_t, \ldots, Z_{d_t} = Y_t$$

be a chain of states where neighboring states have pre-metric distance 1. The move in Z_0 gives a coupled move for Z_1, and eventually leads to a coupled move for Z_{d_t}. Let Z'_i be the state after the move is made from Z_i, and let

$$\Delta\left(Z'_{i-1}, Z'_i\right) = \left|Z'_{i-1} \setminus Z'_i\right|.$$

By linearity of expectation,

$$\mathbb{E}[D_{t+1} \mid D_t] = \sum_{i \in [1, D_t]} \mathbb{E}\left[\Delta\left(Z'_{i-1}, Z'_i\right)\right] \leq D_t \cdot \left(1 - \frac{n - 2(\Delta + 1)(k - 1)}{kn}\right),$$

and we conclude the proof. □

1.13 Martingales

1.13.1 Martingales

Definition 1.13.1 (Martingale)**.** A sequence of random variables $Z_0, Z_1, \ldots$ is a martingale with respect to another sequence $X_0, X_1, \ldots$ if for all $n \geq 0$, the following conditions hold:

- Z_n is a function of $X_0, \ldots, X_n$.
- $\mathbb{E}[|Z_n|] < \infty$.
- $\mathbb{E}[Z_{n+1} \mid X_0, \ldots, X_n] = Z_n$.

A sequence of random variables $Z_0, Z_1, \ldots$ is called a martingale when it is a martingale with respect to itself, that is

- $\mathbb{E}[|Z_n|] < \infty$.
- $\mathbb{E}[Z_{n+1} \mid Z_0, \ldots, Z_n] = Z_n$.

Example 1.13.2 (Martingale from Fair Games)**.** Consider a gambler who plays a sequence of fair games. Let X_i be the winning on the i^{th} game, that is either $+1$ or -1, and Z_i be the gambler's total winning of the first i games. Since $Z_0 = 0$ and

$$Z_i = Z_0 + \sum_{j \in [1,i]} X_j,$$

then by linearity of expectation,

$$\mathbb{E}[Z_{i+1} \mid X_1, \ldots, X_i] = Z_i + \mathbb{E}[X_{i+1} \mid X_1, \ldots, X_i] = Z_i,$$

where the second equality holds as $\mathbb{E}[X_{i+1} \mid X_1, \ldots, X_i] = 0$ on any input of $X_1, \ldots, X_i$ by fair game.

Thus, $Z_1, \ldots, Z_n$ is a martingale with respect to the sequence $X_1, \ldots, X_n$.

Exercise 1.13.3 (Exercise 13.2 [MU17])**.** Let $X_0, X_1, \ldots$ be a sequence of random variables, and let $S_i = \sum_{j \in [1,i]} X_j$. Show that $S_0, S_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$, then for all $i \neq j$, $\mathbb{E}\left[X_i X_j\right] = 0$.

Proof. If $S_0, S_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$, then by definition 1.13.1,

$$S_i = \mathbb{E}[S_{i+1} \mid X_0, \ldots, X_i] = S_i + \mathbb{E}[X_{i+1} \mid X_0, \ldots, X_i],$$

where the second equality holds by linearity of expectation, then for all $i \geq 0$,

$$\mathbb{E}[X_{i+1} \mid X_0, \ldots, X_i] = 0.$$

Hence, on an $i \geq 1$ and any $j < i$, conditioned that $X_j = x$,

$$\mathbb{E}\left[X_i \mid X_j = x\right] = \mathbb{E}\left[\mathbb{E}\left[X_i \mid X_0, \ldots, X_j = x, \ldots, X_{i-1}\right]\right] = 0.$$

Therefore, we conclude the proof by

$$\begin{aligned}
\mathbb{E}\left[X_i X_j\right] &= \mathbb{E}\left[\mathbb{E}\left[X_i X_j \mid X_j\right]\right] \\
&= \sum_{x \in \Omega} \mathbb{E}\left[X_i X_j \mid X_j = x\right] \Pr\left[X_j = x\right] \\
&= \sum_{x \in \Omega} x\, \mathbb{E}\left[X_i \mid X_j = x\right] \Pr\left[X_j = x\right] = 0.
\end{aligned}$$

□

Exercise 1.13.4 (Exercise 13.1 [MU17])**.** Show that if $Z_0, Z_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$, then it is a martingale with respect to itself.

Proof. By definition 1.2.6 and lemma 1.2.7, we use the tower property

$$\mathbb{E}[\mathbb{E}[Y \mid X] \mid Z] = \mathbb{E}[Y \mid Z]$$

for random variables X, Y, Z such that Z is determined by X, or X carries at least as much information as Z, or Z can be obtained from X from some function.

Since $Z_0, \ldots, Z_n$ is a martingale with respect to $X_0, \ldots, X_n$, then by definition 1.13.1, for all $i \in [0, n]$,

$$Z_i = \mathbb{E}[Z_{i+1} \mid X_0, \ldots, X_i].$$

By the prior fact,

$$\mathbb{E}[\mathbb{E}[Z_{i+1} \mid X_0, \ldots, X_i] \mid Z_0, \ldots, Z_i] = \mathbb{E}[Z_{i+1} \mid Z_0, \ldots, Z_i].$$

Yet by definition 1.13.1,

$$\mathbb{E}[\mathbb{E}[Z_{i+1} \mid X_0, \ldots, X_i] \mid Z_0, \ldots, Z_i] = \mathbb{E}[Z_i \mid Z_0, \ldots, Z_i] = Z_i.$$

Therefore we conclude by $\mathbb{E}[Z_{i+1} \mid Z_0, \ldots, Z_i] = Z_i$. □

Exercise 1.13.5 (Exercise 13.4 [MU17])**.** Let $X_1, X_2, \ldots$ be independent identically distributed random variables with expectation 0 and variance $\sigma^2 < \infty$. Let

$$Z_n = \left(\sum_{i \in [1,n]} X_i \right)^2 - n\sigma^2.$$

Show that $Z_1, Z_2, \ldots$ is a martingale.

Proof. Let $Y_n = \sum_{i \in [1,n]} X_i$, then $Z_n = Y_n^2 - n\sigma^2$. Since $\mathbb{E}[X_i] = 0$, then $\sigma^2 = \mathbb{E}\left[X_i^2\right]$. We show as follows: For $n \geq 1$,

$$\begin{aligned}
\mathbb{E}[Z_{n+1} \mid X_1, \ldots, X_n] &= \mathbb{E}\left[Z_n + X_{n+1}^2 + 2X_{n+1}Y_n - \sigma^2 \mid X_1, \ldots, X_n\right] \\
&= Z_n + \mathbb{E}\left[X_{n+1}^2\right] + 2\,\mathbb{E}[X_{n+1}Y_n] - \sigma^2 \\
&= Z_n + \mathbb{E}\left[X_{n+1}^2\right] + 2\,\mathbb{E}[X_{n+1}]\,\mathbb{E}[Y_n] - \sigma^2 \\
&= Z_n + \mathbb{E}\left[X_{n+1}^2\right] - \sigma^2 = Z_n,
\end{aligned}$$

where the second and the third equality are from the independence, and the last two equalities are from expectation being 0. By exercise 1.13.4, we conclude the proof. □

Definition 1.13.6 (Doob Martingale [Doo40])**.** Let $X_0, \ldots, X_n$ be a sequence of random variables, and Y be a random variable depending on $X_0, \ldots, X_n$ with $\mathbb{E}[|Y|] < \infty$, then for $i \in [0, n]$,

$$Z_i = \mathbb{E}[Y \mid X_0, \ldots, X_i]$$

gives a martingale with respect to $X_0, \ldots, X_n$, as

$$\begin{aligned}
\mathbb{E}[Z_{i+1} \mid X_0, \ldots, X_i] &= \mathbb{E}[\mathbb{E}[Y \mid X_0, \ldots, X_{i+1}] \mid X_0, \ldots, X_i] \\
&= \mathbb{E}[Y \mid X_0, \ldots, X_i] \\
&= Z_i,
\end{aligned}$$

where the second equality holds by lemma 1.2.7 and definition 1.2.6.

Remark 1.13.7. One can view the Doob Martingale as a sequence of refined predictions to Y, where each element Z_i is the expectation of Y when the values of $X_0, \ldots, X_i$ are known. Hence, the refined predictions gradually using more information on the values of the random variables $X_0, \ldots, X_n$.

Exercise 1.13.8 (Exercise 13.3 [MU17])**.** Let $X_0 = 0$ and for $j \geq 0$, X_{j+1} is chosen uniformly over real interval $[X_j, 1]$. Show that, for $k \geq 0$, the sequence $Y_k = 2^k(1 - X_k)$ is a martingale.

Proof. Since Y_{k+1} depends only on Y_k, then

$$\mathbb{E}[Y_{k+1} \mid Y_0, \ldots, Y_k] = \mathbb{E}[Y_{k+1} \mid Y_k].$$

Moreover, since $\mathbb{E}[X_{k+1} \mid X_k] = (1 + X_k)/2$, then

$$\mathbb{E}[Y_{k+1} \mid X_k] = \mathbb{E}\left[2^{k+1}(1 - X_{k+1}) \mid X_k\right] = 2^{k+1} - 2^{k+1}\,\mathbb{E}[X_{k+1} \mid X_k] = 2^k(1 - X_k) = Y_k.$$

By $\mathbb{E}[Y_k \mid Y_k] = Y_k$, we show that

$$\mathbb{E}[\mathbb{E}[Y_{k+1} \mid X_k] \mid Y_k] = \mathbb{E}[Y_{k+1} \mid Y_k] = Y_k,$$

which concludes the proof. □

Lemma 1.13.9. *If $Z_0, Z_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$, then $\mathbb{E}[Z_n] = \mathbb{E}[Z_0]$ for any $n \geq 0$.*

Proof. By definition 1.13.1, for any $n \geq 1$

$$\mathbb{E}[Z_n \mid X_0, \ldots, X_{n-1}] = Z_{n-1}.$$

Given that

$$\mathbb{E}[Z_{n-1} \mid X_0, \ldots, X_{n-2}] = Z_{n-2},$$

we have

$$\begin{aligned} Z_{n-2} &= \mathbb{E}[Z_{n-1} \mid X_0, \ldots, X_{n-2}] \\ &= \mathbb{E}[\mathbb{E}[Z_n \mid X_0, \ldots, X_{n-1}] \mid X_0, \ldots, X_{n-2}] \\ &= \mathbb{E}[Z_n \mid X_0, \ldots, X_{n-2}]. \end{aligned}$$

By induction, we have $Z_0 = \mathbb{E}[Z_n \mid X_0]$, then

$$\mathbb{E}[Z_0] = \mathbb{E}[\mathbb{E}[Z_n \mid X_0]] = \mathbb{E}[Z_n],$$

which concludes the proof. □

Example 1.13.10 (Galton-Watson Process)**.** The Galton-Watson process was used in theorem 1.6.60, but was originally used to statistically investigate the extinction of family names. Let G_t be the random variable for the individuals in the t^{th} generation, and $X_{t,k}$ be the number of offspring of the k^{th} individual in the t^{th} generation. Each individual gives birth to offspring independently, and their numbers of offspring are identically distributed. Hence,

$$G_t = \sum_{k \in [1, G_{t-1}]} X_{t-1,k},$$

and let $\mu = \mathbb{E}[X_{t,k}]$. By linearity of expectation,

$$\mathbb{E}[G_{t+1} \mid G_1, \ldots, G_t] = \mu G_t.$$

Let $M_t = \mu^{-t} \cdot G_t$, then

$$\mathbb{E}[M_{t+1} \mid G_1, \ldots, G_t] = \mathbb{E}\left[\mu^{-(t+1)} \cdot G_{t+1} \mid G_1, \ldots, G_t\right] = \mu^{-t} \cdot G_t = M_t,$$

and we have

$$\mathbb{E}[\mathbb{E}[M_{t+1} \mid G_1, \ldots, G_t] \mid M_1, \ldots, M_t] = \mathbb{E}[M_{t+1} \mid M_1, \ldots, M_t].$$

On the other hand,

$$\mathbb{E}[\mathbb{E}[M_{t+1} \mid G_1, \ldots, G_t] \mid M_1, \ldots, M_t] = \mathbb{E}[M_t \mid M_1, \ldots, M_t] = M_t,$$

which makes $M_1, \ldots$ a martingale.

Example 1.13.11 (Pólya's Urn)**.** Suppose there are black balls and white balls in an urn, that are identical except for the colors. Each time pick a ball, and put it back with another identical ball with same color. For simplicity, we start at round 2, and ends at round n where there are n balls in the urn. Let X_n be the number of black balls in the urn, and $Z_n = X_n/n$ be the ratio of black balls in the urn. Then for all $n \geq 2$,

$$\mathbb{E}[Z_{n+1} \mid Z_2, \ldots, Z_n] = Z_n \cdot \frac{nZ_n + 1}{n+1} + (1 - Z_n) \cdot \frac{nZ_n}{n+1} = Z_n,$$

which makes $Z_2, Z_3, \ldots$ a martingale.

1.13.2 Stopping Times

By lemma 1.13.9, if the number of games is fixed initially, then the expected gain from the fair games is 0, if $Z_0 = 0$. But suppose the number of games is not fixed initially, for example the gambler choose to play a random number of games, by deciding when to quit based on the outcome of the games already played.

Definition 1.13.12 (Stopping Time)**.** A nonnegative integer-valued random variable T is a stopping time for the sequence $Z_0, Z_1, \ldots$ if the probability of the event $T = n$ is independent of the variables $\{Z_{n+j} \mid Z_0, \ldots, Z_n\}_{j \geq 1}$, which are the random variables $Z_{n+1}, Z_{n+2}, \ldots$ conditioned on the values of $Z_0, \ldots, Z_n$.

A stopping time is corresponding to a strategy of when to stop based only on the outcome seen so far. But given that one can stop by strategy such as "winning reaches 10 for the first time", it would be helpful to characterize the conditions on the stopping time T that maintain the property $\mathbb{E}[Z_T] = \mathbb{E}[Z_0] = 0$. The subtle problem is that, T might not be finite in this case, that the gambler will gamble infinite number of times with non-zero probability. The martingale stopping time theorem shows that, under certain conditions, and in particular when the stopping time is bounded or has bounded expectation, the expected value of the martingale at the stopping time is $\mathbb{E}[Z_0]$.

Theorem 1.13.13 (Martingale Stopping Time Theorem)**.** *If $Z_0, Z_1, \ldots$ is a martingale with respect to $X_0, X_1, \ldots$ and if T is a stopping time for $X_0, X_1, \ldots$, then*

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_0]$$

whenever one of the following holds:

- *The Z_i is bounded, such that there exists constant c for all $i \geq 0$, $|Z_i| \leq c$.*
- *T is bounded.*
- *$\mathbb{E}[T] < \infty$, and there is a constant c such that $\mathbb{E}[|Z_{i+1} - Z_i| \mid X_0, \ldots, X_i] \leq c$.*

Exercise 1.13.14 (Exercise 13.5 [MU17])**.** Consider the gambler's ruin example 1.7.51, stopping by either losing ℓ_1 or winning ℓ_2. Let X_n be the winning in the n^{th} game, Y_n be the winning by the end of the n^{th} game, and $Z_n = Y_n^2 - n$. Show that $Z_1, Z_2, \ldots$ is a martingale, and let T be a stopping time, determine $\mathbb{E}[Z_T]$ and $\mathbb{E}[T]$.

Proof. Since $\operatorname{Var}[X_i] = 1$ and $\mathbb{E}[X_i] = 0$, by exercise 1.13.5, $Z_1, Z_2, \ldots$ is a martingale.

We have shown in exercise 1.7.68 that $\mathbb{E}[T] = \ell_1 \ell_2$.

The theorem 1.13.13 applies by

$$|Z_{i+1} - Z_i| = |(Y_{i+1} - Y_i)(Y_{i+1} + Y_i) - 1| \leq |Y_{i+1} - Y_i| \cdot |Y_{i+1} + Y_i| + 1 \leq 2\max(\ell_1, \ell_2) + 1,$$

then $\mathbb{E}[Z_T] = \mathbb{E}[Z_1] = 0$, which means $\mathbb{E}\left[Y_T^2\right] = \mathbb{E}[T]$. Since Y_T^2 can either be ℓ_1^2 or ℓ_2^2, by example 1.7.51 [16],

$$\mathbb{E}\left[Y_T^2\right] = \ell_1^2 \cdot \frac{\ell_2}{\ell_1 + \ell_2} + \ell_2^2 \cdot \frac{\ell_1}{\ell_1 + \ell_2} = \ell_1 \ell_2,$$

which derives $\mathbb{E}[T]$ in a martingale way. □

[16] Now we can show by Y_i is a martingale in example 1.13.2, then $\mathbb{E}[Y_T] = 0$, to derive the winning or losing probabilities.

Exercise 1.13.15 (Exercise 13.6 [MU17]). Consider the gambler's ruin example 1.7.51, stopping by either losing ℓ_1 or winning ℓ_2, and winning probability $p < 1/2$. Let X_n be the winning in the n^{th} game, and Z_n be the winning by the end of the n^{th} game.

- Show that
$$A_n = \left(\frac{1-p}{p}\right)^{Z_n}$$
is a martingale with mean 1.
- Determine the probability that the player wins ℓ_2 before losing ℓ_1.
- Show that $B_n = Z_n - (2p-1)n$ is a martingale with mean 0.
- Let T be the stopping time when the player finishes playing. Determine $\mathbb{E}[Z_T]$ and $\mathbb{E}[T]$.

Proof. We have shown in exercise 1.7.69 that
$$A_n = \left(\frac{1-p}{p}\right)^{Z_n},$$
we have $\mathbb{E}[A_{n+1}] = \mathbb{E}[A_n] = 1$. Since $Z_{n+1} = Z_n + X_{n+1}$,
$$\mathbb{E}[A_{n+1} \mid X_1, \dots, X_n] = p\left(\frac{1-p}{p}\right)^{Z_n+1} + (1-p)\left(\frac{1-p}{p}\right)^{Z_n-1} = \left(\frac{1-p}{p}\right)^{Z_n} = A_n,$$
which proves that $A_1, A_2, \dots$ is a martingale by exercise 1.13.4.

With this $A_1, A_2, \dots$ martingale, then let q be the probability of being absorbed into ℓ_2, and
$$\mathbb{E}[A_T] = q\left(\frac{1-p}{p}\right)^{\ell_2} + (1-q)\left(\frac{1-p}{p}\right)^{-\ell_1} = 1,$$
such that
$$q = \left(1 - \left(\frac{1-p}{p}\right)^{\ell_1}\right) \cdot \left(1 - \left(\frac{1-p}{p}\right)^{\ell_1+\ell_2}\right)^{-1}.$$

Since $\mathbb{E}[X_i] = p - (1-p) = 2p - 1$, then by linearity of expectation,
$$\mathbb{E}[B_{n+1} \mid X_1, \dots, X_n] = B_n + \mathbb{E}\left[X_{n+1} - (2p-1) \mid X_1, \dots, X_n\right] = B_n,$$
the second equality holds by the independence of the games. Hence, $B_1, B_2, \dots$ is a martingale by exercise 1.13.4.

Since A_n is bounded, then by theorem 1.13.13, $\mathbb{E}[A_T] = \mathbb{E}[A_1] = 1$. Therefore,
$$\mathbb{E}[Z_T] = -\ell_1 \cdot (1-q) + \ell_2 \cdot q = (\ell_1 + \ell_2) \cdot q - \ell_1.$$
By $\mathbb{E}[B_T] = 0 = \mathbb{E}[Z_T] - (2p-1)\mathbb{E}[T]$, we have
$$\mathbb{E}[T] = \frac{\mathbb{E}[Z_T]}{2p-1} = \frac{\ell_1 - (\ell_1 + \ell_2) \cdot q}{1-2p}.$$
□

Example 1.13.16 (A Ballot Theorem). Suppose there are 2 candidates running an election, where candidate A obtains a votes, and candidate B obtains $b < a$ votes. The votes come in random order, and can be viewed as sampled from uniformly random permutations. Candidate A always with higher votes has probability $(a-b)/(a+b)$.

Let $n = a + b$, and let S_k be the leading votes of A to B after k votes. Then $S_n = a - b$. For $0 \le k \le n-1$, let
$$X_k = \frac{S_{n-k}}{n-k}.$$

We first show that $X_1, X_2, \ldots, X_n$ is a martingale. Since

$$\mathbb{E}[S_{n-k-1} \mid S_n, \ldots, S_{n-k}] = \mathbb{E}[S_{n-k-1} \mid S_{n-k}] = \frac{n-k+S_{n-k}}{2(n-k)} \cdot (S_{n-k}-1) + \frac{n-k-S_{n-k}}{2(n-k)} \cdot (S_{n-k}+1) = \frac{n-k-1}{n-k} \cdot S_{n-k},$$

where the first equality holds by S_{n-k-1} is only related to S_{n-k} counting backwards, and the second equality holds by the probability of the $(n-k)^{\text{th}}$ vote being A or B, conditioned on S_{n-k}, then

$$\mathbb{E}[X_{k+1} \mid S_n, \ldots, S_{n-k}] = \frac{S_{n-k}}{n-k} = X_k,$$

and

$$\mathbb{E}[X_{k+1} \mid X_1, \ldots, X_k] = \mathbb{E}[\mathbb{E}[X_{k+1} \mid S_n, \ldots, S_{n-k}] \mid X_1, \ldots, X_k] = X_k,$$

which completes the proof of martingale.

Moreover, define a stopping T for the first $X_T = 0$, otherwise $T = n-1$, then since T is bounded, and $X_0, \ldots, X_{n-1}$ is also bounded, by theorem 1.13.13,

$$\mathbb{E}[X_T] = \mathbb{E}[X_0] = \frac{S_n}{n} = \frac{a-b}{a+b}.$$

By the stopping time T, the voting is either A and B are drawing, or A leads throughout the n votes. In the first case, $X_T = 0$ as $S_T = 0$, otherwise, $X_T = 1$ as $S_T = n$. Hence,

$$\mathbb{E}[X_T] = \Pr\left[\text{A leads throughout}\right] \cdot 1 + \Pr\left[\text{A and B are drawing}\right] \cdot 0 = \frac{a-b}{a+b},$$

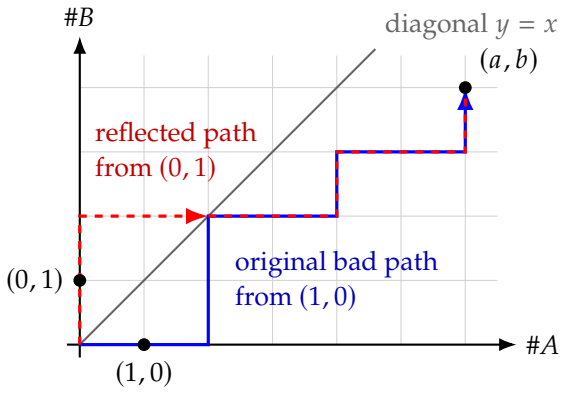
which completes the proof.

Remark 1.13.17. The example 1.13.16 is also called Bertrand's ballot theorem, which can be proved in other ways.

One of the earliest proofs [And87] solved the problem directly, but the "geometric reflection" trick derives from his work. Consider a 2-dimensional lattice, a vote for A moves is a step right, and a vote for B is a step up. Hence, after n moves, the particle moves from $(0,0)$ to (a,b). An observation is, if A is ahead in the count throughout, then the particle is constantly below the line $y = x$ (except for the initial state with no vote).

Moreover, for any permutation that start with a vote for B, then it has to be at the diagonal $y = x$ at least once. The reflection trick is: we can always reflect a trajectory above the diagonal to the symmetry one below the diagonal $y = x$, such that the move is still a bad move, but the move is beginning with a vote for A to $(1,0)$.

Since all the valid trajectories begin with a vote for A, we just exclude all the bad trajectories beginning with $(1,0)$, and all such bad trajectories can be reflected to the trajectories beginning with $(0,1)$.



Therefore, the number of valid trajectories is

$$\binom{a+b-1}{a-1} - \binom{a+b-1}{b-1} = \frac{a-b}{a+b} \cdot \binom{a+b}{a}.$$

When $a = b+1$, this deteriorates to lemma 1.7.67, which are Dycks path and Catalans number C_b.

1.13.3 Wald's Equation

Wald's equation is an important corollary of theorem 1.13.13, which handles the expectation of the sum of independent random variables, where the number of random variables being summed is itself a random variable.

Theorem 1.13.18 (Wald's Equation [Wal47]). *Let $X_1, X_2, \ldots$ be independent, identically distributed random variables with distribution X, and T be a stopping time for this sequence. If T and X have bounded expectation, then*

$$\mathbb{E}\left[\sum_{i\in[1,T]} X_i\right] = \mathbb{E}[T]\cdot\mathbb{E}[X]\,.$$

Proof. We first show a version for the nonnegative random variables. For $i \geq 1$, let

$$Z_i = \sum_{j\in[1,i]} \left(X_j - \mathbb{E}\left[X_j\right]\right),$$

then $Z_1, Z_2, \ldots$ is a martingale with respect to $X_1, X_2, \ldots$, with $\mathbb{E}[Z_i] = 0$ by exercise 1.13.15.

Since $\mathbb{E}[T] < \infty$, and

$$\mathbb{E}[|Z_{i+1} - Z_i| \mid X_1, \ldots, X_i] = \mathbb{E}[|X_{i+1} - \mathbb{E}[X]|] \leq 2\,\mathbb{E}[X]\,, \tag{1.25}$$

then by theorem 1.13.13,

$$\mathbb{E}[Z_T] = \mathbb{E}[Z_1] = 0.$$

Therefore, by linearity of expectation,

$$\mathbb{E}[Z_T] = \mathbb{E}\left[\sum_{i\in[1,T]} (X_i - \mathbb{E}[X])\right] = \mathbb{E}\left[\sum_{i\in[1,T]} X_i\right] - \mathbb{E}[T]\cdot\mathbb{E}[X]\,,$$

which completes the proof.

In general case, where X is any random variable, the main trick is writing $X = X^+ - X^-$, where

$$X^+ = \max(X, 0)\,, \quad X^- = \max(-X, 0)\,.$$

Hence, write $X_i = X_i^+ - X_i^-$ for all $i \geq 1$, and by the prior nonnegative case,

$$\mathbb{E}\left[\sum_{i\in[1,T]} X_i^+\right] = \mathbb{E}[T]\cdot\mathbb{E}[X^+], \mathbb{E}\left[\sum_{i\in[1,T]} X_i^-\right] = \mathbb{E}[T]\cdot\mathbb{E}[X^-]\,.$$

By linearity of expectation,

$$\begin{aligned}\mathbb{E}\left[\sum_{i\in[1,T]} X_i\right] &= \mathbb{E}\left[\sum_{i\in[1,T]} X_i^+ - \sum_{i\in[1,T]} X_i^-\right] = \mathbb{E}\left[\sum_{i\in[1,T]} X_i^+\right] - \mathbb{E}\left[\sum_{i\in[1,T]} X_i^-\right] \\ &= \mathbb{E}[T]\cdot(\mathbb{E}[X^+] - \mathbb{E}[X^-]) \\ &= \mathbb{E}[T]\cdot\mathbb{E}[X]\,,\end{aligned}$$

we conclude the proof for the general case. □

Remark 1.13.19. Since T is a stopping time for $X_1, X_2, \ldots$, then event $\{T \geq i\}$ is dependent only on $X_1, \ldots, X_{i-1}$ by definition 1.13.12. The mutual independence over X_i ensures that X_i is independent of past $X_1, \ldots, X_{i-1}$, hence X_i is independent of $\{T \geq i\}$. The independence is a convenient sufficient condition, that can be used to construct martingale in exercise 1.13.15 and used here, and help simplify the third condition of the theorem 1.13.13 in eq. (1.25).

After theorem 1.13.18, remark 1.13.19, and exercise 1.13.15, for the stopping time in the case of independent random variables, we have an equivalent, yet simpler version of stopping time as follows.

Definition 1.13.20 (Stopping Time for Independent Random Variables)**.** Let $Z_0, Z_1, \ldots$ be a sequence of independent random variables. A nonnegative, integer-valued random variable T is a stopping time for the sequence if the event $\{T = n\}$ is independent of $Z_{n+1}, Z_{n+2}, \ldots$.

Example 1.13.21 (Dice Rolling)**.** Consider a gambling game where a player first rolls a die. If the outcome is X then roll X new standard dice and gain Z, which is the sum of the outcome of the X dice. This is exactly an application of theorem 1.13.18 to derive the expected outcome. By definition 1.13.20 and theorem 1.13.18, let X be the stopping time, and $Y_1, Y_2, \ldots$ be i.i.d. random variables for dice rolling scores with distribution Y. Then by theorem 1.13.18,

$$\mathbb{E}\left[\sum_{i\in[1,X]} Y_i\right] = \mathbb{E}[X]\cdot\mathbb{E}[Y] = \left(\frac{1+n}{2}\right)^2,$$

if the die is a n faced fair die.

Example 1.13.22 (Server Message Transmission)**.** Consider n servers communicating using a shared channel, where time is divided in discrete slots. At each time slot, any server that needs to send a packet can transmit it through the channel. If exactly one packet is sent at that time, the transmission is successful, otherwise none are successful. At a time slot, a server transmits a packet with probability $1/n$.

Let $X_1, X_2, \ldots$ be the independent 0-1 random variables for a successful packet transmission on the i^{th} time slot with distribution X, then

$$\Pr[X = 1] = \binom{n}{1}\cdot\frac{1}{n}\cdot\left(1-\frac{1}{n}\right)^{n-1} = \left(1-\frac{1}{n}\right)^{n-1}.$$

Similar to section 1.5.1, let

$$\varphi_g(n) = (n-1)\ln\left(1-\frac{1}{n}\right),$$

then

$$\varphi_g'(n) = \ln\left(1-\frac{1}{n}\right) + \frac{1}{n}.$$

Let $t = 1 - 1/n$, and $n \geq 2$ means $1/2 \leq t < 1$, then $\varphi_g'(n) < 0$. Therefore, when $n \geq 2$, since

$$\lim_{n\to\infty}\left(1-\frac{1}{n}\right)^{n-1} = e^{-1},$$

then e^{-1} is a lower bound by

$$p = \Pr[X = 1] = \left(1-\frac{1}{n}\right)^{n-1} \geq e^{-1}.$$

Let N be the random variable for the number of packets successfully sent until each server has successfully sent at least 1 packet. By the Coupon's Collector's Problem,

$$\mathbb{E}[N] = n\sum_{i\in[1,n]} i^{-1} = nH(n) = n\ln n + cn.$$

Let Y_i be the number of time slots between the $(i-1)^{\text{th}}$ successful transmission and the i^{th} successful transmission. Then $Y_1, Y_2, \ldots$ are independent and geometrically distributed with success probability p, therefore $\mathbb{E}[Y_i] = 1/p \leq e$. Moreover, whichever server successfully transmitted a packet is independent of the waiting times Y_i, so the Wald argument applies to this sum. By theorem 1.13.18,

$$\mathbb{E}\left[\sum_{i\in[1,N]} Y_i\right] = \mathbb{E}[N]\cdot\mathbb{E}[Y_1] \leq e\,(n\ln n + cn).$$

Exercise 1.13.23 (Exercise 13.11 [MU17]). A parking-lot attendant has mixed up n keys for n cars. The n car owners arrive together. The attendant gives each owner a key according to a permutation chosen uniformly at random from all permutations. If an owner receives his key, he takes it and leaves; otherwise, he returns the key to the attendant. The attendant repeats the process with the remaining keys and car owners. This continues until all owners receive the keys to their cars. Let R be the number of rounds until all car owners receive their keys, and let X_i be the number of owners who receive their car keys in the i^{th} round. Show that

$$Y_i = \sum_{j\in[1,i]} \left(X_j - \mathbb{E}\left[X_j \mid X_1, \ldots, X_{j-1}\right]\right)$$

is a martingale, and use theorem 1.13.13 to derive $\mathbb{E}[R]$.

Proof. To show that $Y_1, Y_2, \ldots$ is a martingale, we have

$$\begin{aligned}\mathbb{E}[Y_i \mid Y_1, \ldots, Y_{i-1}] &= Y_{i-1} + \mathbb{E}[X_i - \mathbb{E}[X_i \mid X_1, \ldots, X_{i-1}] \mid Y_1, \ldots, Y_{i-1}] \\ &= Y_{i-1} + \mathbb{E}[X_i \mid Y_1, \ldots, Y_{i-1}] - \mathbb{E}[X_i \mid Y_1, \ldots, Y_{i-1}] = Y_{i-1}.\end{aligned}$$

Since both X_{i+1} and $\mathbb{E}[X_{i+1} \mid X_1, \ldots, X_i]$ are in $[0, n]$, then by triangle inequality and tower property,

$$\begin{aligned}\mathbb{E}[|Y_{i+1} - Y_i| \mid Y_1, \ldots, Y_i] &= \mathbb{E}[|X_{i+1} - \mathbb{E}[X_{i+1} \mid X_1, \ldots, X_i]| \mid Y_1, \ldots, Y_i] \\ &\le 2\,\mathbb{E}[X_{i+1} \mid Y_1, \ldots, Y_i] \le 2n,\end{aligned}$$

we can apply theorem 1.13.13, that

$$\mathbb{E}[Y_R] = \mathbb{E}[Y_1] = 0.$$

Moreover, by exercise 1.2.4, for $i \ge 1$,

$$\mathbb{E}[X_i \mid X_1, \ldots, X_{i-1}] = 1.$$

Hence, by the stopping time R, we have

$$\mathbb{E}[Y_R] = \mathbb{E}\left[\sum_{i\in[1,R]} X_i - \sum_{i\in[1,R]} \mathbb{E}[X_i \mid X_1, \ldots, X_{i-1}]\right] = \mathbb{E}\left[\sum_{i\in[1,R]} X_i\right] - \mathbb{E}[R] = 0.$$

Since by stopping time R,

$$\sum_{i\in[1,R]} X_i = n,$$

we have $\mathbb{E}[R] = n$. □

Exercise 1.13.24 (Exercise 13.12 [MU17]). Alice and Bob play each other in a chess tournament, where the first player to win four games wins the match. The players are evenly matched, so the probability that each player wins each game is $1/2$, independent of other games. The number of minutes for each game is uniformly distributed over the integers in $[30, 60]$, again independent of other games. What is the expected time they spend playing this match?

Proof. Let $X_1, X_2, \ldots$ be the time used in the i^{th} game with distribution X, then $\mathbb{E}[X] = 45$.

Let T be the stopping time, and WLOG let Alice win, as Bob wins symmetrically. Since Alice wins the final game, then there are $\binom{3}{0}$ ways to finish in 4 games, $\binom{4}{1}$ ways to finish in 5 games, $\binom{5}{2}$ ways to finish in 6 games, $\binom{6}{3}$ ways to finish in 7 games. The expected number of games is

$$\mathbb{E}[T] = \left(\binom{3}{0} + \binom{4}{1} + \binom{5}{2} + \binom{6}{3}\right)^{-1} \cdot \left(\binom{3}{0} \cdot 4 + \binom{4}{1} \cdot 5 + \binom{5}{2} \cdot 6 + \binom{6}{3} \cdot 7\right) = \frac{32}{5}.$$

By theorem 1.13.18,

$$\mathbb{E}\left[\sum_{i\in[1,T]} X_i\right] = \mathbb{E}[X] \cdot \mathbb{E}[T] = 288.$$

□

Exercise 1.13.25 (Exercise 13.13 [MU17]). Consider the following algorithm for sorting n numbers. Start by choosing one of the n numbers uniformly at random, and place it first. Then choose one of the remaining $n-1$ numbers uniformly at random, and place it second. If the second number is smaller than the first, start over again from the beginning. Otherwise, next choose one of the remaining $n-2$ numbers uniformly at random, place it third, and so on. The algorithm starts over from the beginning whenever it finds that the k^{th} item placed is smaller than the $(k-1)^{\text{th}}$ item. Determine the expected number of times the algorithm tries to place a number, assuming that the input consists of n distinct numbers.

Proof. Let $X_1, X_2, \ldots$ be the random variables of the lengths of the incremental sequences drawn from the n distinct numbers, then by the stopping time T, $X_T = n$. Since the success probability of drawing an incremental sequence of length n is $(n!)^{-1}$, then $\mathbb{E}[T] = n!$.

Let $Y_{[1,k]}$ be the random variables for the prefix sequence of length k, then we have

$$\Pr\left[Y_{[1,k]} = \mathbf{y}_{[1,k]}\right] = \frac{(n-k)!}{n!},$$

as each permutation is uniformly distributed. Hence,

$$\Pr[X \geq k] = \binom{n}{k}\frac{(n-k)!}{n!} = \frac{1}{k!},$$

and the expectation of the length of incremental sequence is

$$\mathbb{E}[X] = \sum_{k\in[1,n]} \Pr[X \geq k] = \sum_{k\in[1,n]} \frac{1}{k!} \leq e,$$

the upper bound can be seen in section 1.5.1.

Since X_i are i.i.d., then we can apply theorem 1.13.18, the expected number of draws is

$$\mathbb{E}\left[\sum_{i\in[1,T]} X_i\right] = \mathbb{E}[T]\cdot\mathbb{E}[X] = n!\sum_{k\in[1,n]}\frac{1}{k!}.$$

□

Exercise 1.13.26 (Exercise 13.14 [MU17]). Suppose we are arranging a chain of n dominoes so that, once we are done, we can have them fall sequentially by knocking down the lead domino. Each time one tries to place a domino in the chain, there is some probability that it falls. In this case, one must start over from the very first domino.

- Call each time placing a domino a trial, succeeding with probability p. Use theorem 1.13.18, find the expected number of trials necessary before the arrangement is ready.
- Suppose instead that one can break the arrangement in k components, each of size n/k, in such a way so that once a component is complete, it will not fall for further dominoes. Find the expected number of trials necessary before the arrangement is ready.

Proof. Let $X_1, X_2, \ldots$ be the longest length of a domino arrangement, and T be the stopping time, such that $X_T = n$. Since X_i is geometrically distributed for the number of steps to the first failure, and X_i are i.i.d., then $\mathbb{E}[X_i] = (1-p)^{-1}$. On the other hand, T is also geometrically distributed with success probability p^n, therefore $\mathbb{E}[T] = p^{-n}$.

Hence, by theorem 1.13.18,

$$\mathbb{E}\left[\sum_{i\in[1,T]} X_i\right] = \mathbb{E}[X_1]\cdot\mathbb{E}[T] = \frac{1}{(1-p)p^n}.$$

When there are k segments of sub-arrangements, the expected number of trials to finish a segment is $(1-p)^{-1}p^{-n/k}$, and by linearity of expectation, the expected number of trials to finish all pieces is $k(1-p)^{-1}p^{-n/k}$. □

Exercise 1.13.27 (Exercise 13.15 [MU17]). Use theorem 1.13.18 to derive the following $\mathbb{E}[N]$.

- Let $X_1, X_2, \ldots$ be independent exponential random variables with $\mathbb{E}[X_i] = 1$. Given $k \in \mathbb{R}^+$, define
$$N = \min\{n : \sum_{i\in[1,n]} X_i > k\}.$$
- Let $Y_1, Y_2, \ldots$ be independent uniform random variables in $(0,1)$. Given $k \in (0,1)$, define
$$N = \min\{n : \prod_{i\in[1,n]} Y_i < k\}.$$

Proof. Let $S_m = \sum_{i\in[1,m]} X_i$. Since $\mathbb{E}[X_i] = 1$, by theorem 1.13.18,
$$\mathbb{E}[S_N] = \mathbb{E}[X_1] \cdot \mathbb{E}[N] = \mathbb{E}[N].$$
Moreover, since X_i are exponential random variables, then by the memorylessness property, for any $a \geq 0$,
$$\mathbb{E}[X_i - a \mid X_i > a] = 1.$$
Given that $S_{N-1} \leq k$, and $S_N > k$, we have $X_N > k - S_{N-1}$, then
$$\mathbb{E}[X_N - (k-s) \mid X_N > k-s, S_{N-1} = s] = 1$$
holds for any $0 \leq s \leq k$, therefore $\mathbb{E}[N] = k+1$.

For $Y_1, Y_2, \ldots$, let $Z_i = -\ln Y_i$, then the stopping time is defined as
$$N = \min\{n : \sum_{i\in[1,n]} Z_i > -\ln k\},$$
and $\mathbb{E}[Z_i] = 1$ by
$$\mathbb{E}[Z_i] = \int_0^1 -\ln x \, dx = 1.$$
Let $Q_m = \sum_{i\in[1,m]} Z_i$, by theorem 1.13.18,
$$\mathbb{E}[Q_N] = \mathbb{E}[Z_1] \cdot \mathbb{E}[N] = \mathbb{E}[N].$$
Since $Q_{N-1} \leq -\ln k$, while $Q_N > -\ln k$, we have $Z_N > -\ln k - Q_{N-1}$. Moreover,
$$\Pr[Z_i > a] = \Pr[Y_i < e^{-a}] = e^{-a},$$
as they are the same event. Hence, for all $0 < s \leq -\ln k$,
$$\mathbb{E}[Z_N - (-\ln k - s) \mid Z_N > -\ln k - s, Q_{N-1} = s] = \frac{1}{\Pr[Y_N < k \cdot e^s]} \int_0^{k\cdot e^s} (-\ln x - (-\ln k - s))\, dx = 1,$$
which proves that $\mathbb{E}[Q_N] = \mathbb{E}[N] = -\ln k + 1$.

Another way of showing is, Z_i is an exponential random variable with probability density function $f(x) = e^{-x}$, hence $\mathbb{E}[Z_i] = 1$. Applying the prior result, the threshold is $-\ln k$, and $\mathbb{E}[N] = -\ln k + 1$. □

1.13.4 Azuma-Hoeffding's Inequality

Azuma-Hoeffding says a martingale with bounded increment per step has total gain $X_t - X_0$ concentrating around 0, even if the steps are not independent.

Theorem 1.13.28 (Azuma-Hoeffding's Inequality [Azu67])**.** *Let $X_0, X_1, \ldots$ be a martingale such that*

$$|X_k - X_{k-1}| \leq c_k,$$

then for all $t \geq 1$ and $\varepsilon > 0$,

$$\Pr\left[|X_t - X_0| \geq \varepsilon\right] \leq 2\exp\left(-\frac{\varepsilon^2}{2\sum_{i\in[1,t]} c_i^2}\right).$$

Proof. The proof is similar to theorem 1.4.10, using moment-generating functions.

Start by bounding the upper tail. Let $Y_i = X_i - X_{i-1}$, such that $|Y_i| \leq c_i$, and since $X_0, X_1, \ldots$ is a martingale,

$$\mathbb{E}\left[Y_i \mid X_0, \ldots, X_{i-1}\right] = \mathbb{E}\left[X_i - X_{i-1} \mid X_0, \ldots, X_{i-1}\right] = \mathbb{E}\left[X_i \mid X_0, \ldots, X_{i-1}\right] - X_{i-1} = 0. \tag{1.26}$$

Write

$$Y_i = -c_i \cdot \frac{1 - Y_i/c_i}{2} + c_i \cdot \frac{1 + Y_i/c_i}{2}.$$

Since $\exp(\lambda Y_i)$ is a convex function, by theorem 1.2.5,

$$\begin{aligned}
\exp(\lambda Y_i) &= \exp\left(-c_i\lambda \cdot \frac{1 - Y_i/c_i}{2} + c_i\lambda \cdot \frac{1 + Y_i/c_i}{2}\right) \\
&\leq \frac{1 - Y_i/c_i}{2} \cdot \exp(-c_i\lambda) + \frac{1 + Y_i/c_i}{2} \cdot \exp(c_i\lambda) \\
&= \frac{1}{2}\left(\exp(\lambda c_i) + \exp(-\lambda c_i)\right) + \frac{Y_i}{2c_i}\left(\exp(\lambda c_i) - \exp(-\lambda c_i)\right).
\end{aligned}$$

Now consider

$$\mathbb{E}\left[\exp(\lambda Y_i) \mid X_0, \ldots, X_{i-1}\right] \leq \frac{1}{2}\left(\exp(\lambda c_i) + \exp(-\lambda c_i)\right). \tag{1.27}$$

By section 1.5.1,

$$\frac{1}{2}\left(\exp(\lambda c_i) + \exp(-\lambda c_i)\right) = \sum_{k\geq 0} \frac{(\lambda c_i)^{2k}}{(2k)!}.$$

Since $(2k)! \geq 2^k \cdot k!$,

$$\frac{1}{2}\left(\exp(\lambda c_i) + \exp(-\lambda c_i)\right) = \sum_{k\geq 0} \frac{(\lambda c_i)^{2k}}{(2k)!} \leq \sum_{k\geq 0} \frac{(\lambda c_i)^{2k}}{2^k \cdot k!} = \sum_{k\geq 0} \frac{1}{k!} \cdot \left(\frac{(\lambda c_i)^2}{2}\right)^k = \exp\left(\frac{(\lambda c_i)^2}{2}\right). \tag{1.28}$$

To bound the exponential moment used in Markov's inequality, similar to the Chernoff and Hoeffding bounds,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\lambda\left(X_t - X_0\right)\right)\right] &= \mathbb{E}\left[\exp\left(\lambda \sum_{i\in[1,t]} Y_i\right)\right] = \mathbb{E}\left[\prod_{i\in[1,t]} \exp(\lambda Y_i)\right] = \mathbb{E}\left[\mathbb{E}\left[\prod_{i\in[1,t]} \exp(\lambda Y_i) \,\middle|\, X_1, \ldots, X_{t-1}\right]\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\exp(\lambda Y_t) \cdot \prod_{i\in[1,t-1]} \exp(\lambda Y_i) \,\middle|\, X_1, \ldots, X_{t-1}\right]\right] \\
&= \mathbb{E}\left[\prod_{i\in[1,t-1]} \exp(\lambda Y_i) \cdot \mathbb{E}\left[\exp(\lambda Y_t) \mid X_0, \ldots, X_{t-1}\right]\right] \\
&\leq \exp\left(\frac{\lambda^2 c_t^2}{2}\right) \cdot \mathbb{E}\left[\prod_{i\in[1,t-1]} \exp(\lambda Y_i)\right] \leq \exp\left(\frac{\lambda^2}{2} \sum_{i\in[1,t]} c_i^2\right),
\end{aligned} \tag{1.29}$$

where the last inequality is by induction of eq. (1.27) and eq. (1.28). The rest follows the same method as the Chernoff or Hoeffding bounds,

$$\begin{aligned}\Pr\left[X_t - X_0 \geq \varepsilon\right] &= \Pr\left[\exp\left(\lambda\left(X_t - X_0\right)\right) \geq \exp\left(\lambda\varepsilon\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(\lambda\left(X_t - X_0\right)\right)\right] \cdot \exp\left(-\lambda\varepsilon\right) \\ &\leq \exp\left(\frac{\lambda^2}{2}\sum_{i\in[1,t]} c_i^2 - \lambda\varepsilon\right) \leq \exp\left(-\frac{\varepsilon^2}{2\sum_{i\in[1,t]} c_i^2}\right),\end{aligned} \tag{1.30}$$

where the last inequality is by minimizing at $\lambda = \varepsilon/\sum_{i\in[1,t]} c_i^2$.

To bound the lower tail, let $Z_i = X_{i-1} - X_i$, then $\mathbb{E}\left[Z_i \mid X_0, \ldots, X_{i-1}\right] = 0$ by the same proof in eq. (1.26), and

$$\mathbb{E}\left[\exp\left(\lambda Z_i\right) \mid X_0, \ldots, X_{i-1}\right] \leq \exp\left(\frac{\left(\lambda c_i\right)^2}{2}\right)$$

by eq. (1.27) and eq. (1.28). We conclude the lower bound by eq. (1.29) and eq. (1.30). □

Theorem 1.13.29 (Generalized Azuma-Hoeffding's Inequality [Azu67])**.** *Let $X_0, X_1, \ldots$ be a martingale such that*

$$B_k \leq X_k - X_{k-1} \leq B_k + d_k,$$

for some constant d_k and random variables B_k that may be functions of $X_0, \ldots, X_{k-1}$. Then for all $t \geq 1$ and $\varepsilon > 0$,

$$\Pr\left[\left|X_t - X_0\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2\varepsilon^2}{\sum_{i\in[1,t]} d_i^2}\right).$$

Proof. Start with the upper tail. Let $Y_i = X_i - X_{i-1}$, and $\mathbb{E}\left[Y_i \mid X_0, \ldots, X_{i-1}\right] = 0$ by eq. (1.26). Consider bounding the exponential moment as in eq. (1.29):

$$\mathbb{E}\left[\exp\left(\lambda\left(X_t - X_0\right)\right)\right] = \mathbb{E}\left[\prod_{i\in[1,t-1]} \exp\left(\lambda Y_i\right) \cdot \mathbb{E}\left[\exp\left(\lambda Y_t\right) \mid X_0, \ldots, X_{t-1}\right]\right].$$

By lemma 1.4.8,

$$\mathbb{E}\left[\exp\left(\lambda Y_t\right) \mid X_0, \ldots, X_{t-1}\right] \leq \exp\left(\frac{\lambda^2 d_t^2}{8}\right).$$

By induction,

$$\mathbb{E}\left[\exp\left(\lambda\left(X_t - X_0\right)\right)\right] \leq \exp\left(\frac{\lambda^2}{8}\sum_{i\in[1,t]} d_i^2\right). \tag{1.31}$$

We conclude by

$$\begin{aligned}\Pr\left[X_t - X_0 \geq \varepsilon\right] &= \Pr\left[\exp\left(\lambda\left(X_t - X_0\right)\right) \geq \exp\left(\lambda\varepsilon\right)\right] \\ &\leq \mathbb{E}\left[\exp\left(\lambda\left(X_t - X_0\right)\right)\right] \cdot \exp\left(-\lambda\varepsilon\right) \\ &\leq \exp\left(\frac{\lambda^2}{8}\sum_{i\in[1,t]} d_i^2 - \lambda\varepsilon\right) \leq \exp\left(-\frac{2\varepsilon^2}{\sum_{i\in[1,t]} d_i^2}\right),\end{aligned} \tag{1.32}$$

where the last inequality is by minimizing at $\lambda = 4\varepsilon/\sum_{i\in[1,t]} d_i^2$.

To bound the lower tail, let $Z_i = X_{i-1} - X_i$, then $\mathbb{E}\left[Z_i \mid X_0, \ldots, X_{i-1}\right] = 0$ by the same proof in eq. (1.26), and

$$\mathbb{E}\left[\exp\left(\lambda Z_i\right) \mid X_0, \ldots, X_{i-1}\right] \leq \exp\left(\frac{\lambda^2 d_i^2}{8}\right)$$

by lemma 1.4.8. We conclude the lower bound by eq. (1.31) and eq. (1.32). □

Exercise 1.13.30 (Exercise 13.17 [MU17]). Given a bag with r red balls and g green balls, suppose that we uniformly sample n balls from the bag without replacement. Use Azuma-Hoeffding to show that the number of red balls in the sample concentrates tightly around $nr/(r+g)$.

Proof. Let $X_1, \ldots, X_n$ be 0-1 random variables indicating if the i^{th} draw is a red ball, and let $Y = \sum_{i\in[1,n]} X_i$ be the random variable for the total number of red balls drawn. Then

$$Z_0 = \mathbb{E}[Y], \quad Z_k = \mathbb{E}[Y \mid X_1, \ldots, X_k]$$

is a Doob martingale. Let $Y_k = \sum_{i\in[1,k]} X_i$, such that $Y_n = Y$. Then

$$Z_{k-1} = \mathbb{E}[Y \mid X_1, \ldots, X_{k-1}] = Y_{k-1} + (n-k+1)\cdot\frac{r-Y_{k-1}}{r+g-k+1},$$
$$Z_k = \mathbb{E}[Y \mid X_1, \ldots, X_k] = Y_k + (n-k)\cdot\frac{r-Y_k}{r+g-k}.$$

The difference is

$$\begin{aligned} Z_k - Z_{k-1} &= Y_k + (n-k)\cdot\frac{r-Y_k}{r+g-k} - \left(Y_{k-1} + (n-k+1)\cdot\frac{r-Y_{k-1}}{r+g-k+1}\right) \\ &= \frac{r+g-n}{r+g-k}\left(X_k + \frac{Y_{k-1}-r}{r+g-k+1}\right) = \frac{r+g-n}{r+g-k+1}\left(X_k + \frac{Y_k-r}{r+g-k}\right) \end{aligned}$$

The upper bound is

$$\begin{aligned} Z_k - Z_{k-1} &= \frac{r+g-n}{r+g-k}\left(X_k + \frac{Y_{k-1}-r}{r+g-k+1}\right) \\ &\le X_k + \frac{Y_{k-1}-r}{r+g-k+1} \le X_k \le 1. \end{aligned}$$

The lower bound is

$$\begin{aligned} Z_k - Z_{k-1} &= \frac{r+g-n}{r+g-k+1}\left(X_k + \frac{Y_k-r}{r+g-k}\right) \\ &\ge \frac{r+g-n}{r+g-k+1}\cdot\frac{Y_k-r}{r+g-k} \ge -\frac{r-Y_k}{r+g-k} \ge -1. \end{aligned}$$

Given that $|Z_k - Z_{k-1}| \le 1$, and $\mathbb{E}[Y] = n\cdot r/(r+g)$ by linearity of expectation, we conclude with theorem 1.13.28

$$\Pr[|Y - \mathbb{E}[Y]| \ge \varepsilon] \le 2\exp\left(-\frac{\varepsilon^2}{2n}\right).$$

□

Remark 1.13.31. Interestingly, if we let $X_0, \ldots, X_n$ be the fraction of the remaining red balls in exercise 1.13.30, it is also a martingale. Indeed,

$$\begin{aligned} \mathbb{E}[X_{k+1} \mid X_0, \ldots, X_k] &= X_k \cdot \frac{(r+g-k)X_k - 1}{r+g-k-1} + (1-X_k)\cdot\frac{(r+g-k)X_k}{r+g-k-1} \\ &= \frac{-X_k}{r+g-k-1} + \frac{(r+g-k)X_k}{r+g-k-1} = X_k. \end{aligned}$$

Moreover, $X_0, \ldots, X_n$ is a Doob martingale with respect to itself, and lemma 1.13.9,

$$\mathbb{E}[X_n] = \mathbb{E}[X_0] = \frac{r}{r+g}.$$

The fractional martingale can be viewed as an inverse of example 1.13.11.

McDiarmid's inequality says that, for a function that is not too sensitive, namely changing any one of its inputs will not change its value by much, if the input variables are mutually independent, then the function value depending on the independent random inputs should concentrate around its expectation.

Definition 1.13.32 (Lipschitz Condition [McD89])**.** A function $f(X_1, \ldots, X_n)$ is said to satisfy a Lipschitz condition with bound c if for any i and any set of values $x_1, \ldots, x_n$ and y_i,

$$\left|f(x_1, \ldots, y_i, \ldots, x_n) - f(x_1, \ldots, x_i, \ldots, x_n)\right| \leq c.$$

That is, changing the value of any single coordinate can change the function value by at most c.

Let

$$\begin{aligned} Z_0 &= \mathbb{E}\left[f\left(X_1, \ldots, X_n\right)\right], \\ Z_k &= \mathbb{E}\left[f\left(X_1, \ldots, X_n\right) \mid X_1, \ldots, X_k\right]. \end{aligned}$$

The sequence $Z_0, Z_1, \ldots, Z_n$ is a Doob martingale. Moreover, we claim if $X_1, \ldots, X_n$ are mutually independent, then there exist random variables $B_1, \ldots, B_n$ such that for all $k \in [1, n]$,

$$B_k \leq Z_k - Z_{k-1} \leq B_k + c.$$

We have a counterexample for the dependent case: Let $X_1, \ldots, X_n$ be the random variables drawn from the same fair coin toss Y, and $f(X_1, \ldots, X_n) = \left(\sum_{i\in[1,n]} X_i\right)/n$, with a Lipschitz condition bound $c = 1/n$. Then,

$$Z_0 = \mathbb{E}\left[\frac{1}{n}\sum_{i\in[1,n]} X_i\right] = \frac{1}{2}, \quad Z_1 = \mathbb{E}\left[\frac{1}{n}\sum_{i\in[1,n]} X_i \,\middle|\, X_1\right] = Y.$$

Now $Y - 1/2$ has a range much larger than $1/n$, so the claim of existence of $B_1, \ldots, B_n$ does not hold.

Another counterexample is as follows. Let $X_1, \ldots, X_{k-1}$ be independent random variables drawn from fair coin tosses, and let $X_k, \ldots, X_n$ be random variables equal to the same fair coin toss Y. Let $\ell = n - k + 1$, so

$$Z_{k-1} = \mathbb{E}\left[\frac{1}{n}\sum_{i\in[1,n]} X_i \,\middle|\, X_1, \ldots, X_{k-1}\right] = \frac{1}{n}\sum_{i\in[1,k-1]} X_i + \frac{\ell}{2n}, \quad Z_k = \mathbb{E}\left[\frac{1}{n}\sum_{i\in[1,n]} X_i \,\middle|\, X_1, \ldots, X_k\right] = \frac{1}{n}\sum_{i\in[1,k-1]} X_i + \frac{\ell Y}{n}.$$

The difference $\ell/n \cdot (Y - 1/2)$ has a range $\ell/n \geq 1/n$, so the claim fails.

Theorem 1.13.33 (McDiarmid's Inequality [McD89])**.** *Let f be a function on n variables that satisfies the Lipschitz condition in definition 1.13.32 with bound c. Let $X_1, \ldots, X_n$ be independent random variables that are in the domain of f. Then*

$$\Pr\left[\left|f(X_1, \ldots, X_n) - \mathbb{E}\left[f(X_1, \ldots, X_n)\right]\right| \geq \varepsilon\right] \leq 2\exp\left(-\frac{2\varepsilon^2}{nc^2}\right).$$

Proof. Let S_k denote $X_1, \ldots, X_k$, simplify $f(X_1, \ldots, X_n)$ as $f(\overline{X})$, and let $f_k(\overline{X}, x) = f(X_1, \ldots, X_{k-1}, x, X_{k+1}, \ldots, X_n)$.

Consider the upper bound and the lower bound of $Z_k - Z_{k-1}$, where $Z_k - Z_{k-1}$ is upper bounded by

$$\sup_x \mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}, X_k = x\right] - \mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}\right] = \sup_x \left((Z_k \mid X_k = x) - Z_{k-1}\right),$$

and $Z_k - Z_{k-1}$ is lower bounded by B_k, defined as [17]

$$\inf_x \mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}, X_k = x\right] - \mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}\right] = \inf_x \left((Z_k \mid X_k = x) - Z_{k-1}\right).$$

[17] If $\overline{X}$ here have support (the possible value taken) being finite number of values, we can use max and min to substitute sup and inf.

Then consider the range between the upper and lower bound of $Z_k - Z_{k-1}$:

$$\begin{aligned}\sup_x(Z_k \mid X_k = x) - \inf_x(Z_k \mid X_k = x) &= \sup_x \mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}, X_k = x\right] - \inf_x \mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}, X_k = x\right] \\ &= \sup_{x,y}\left(\mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}, X_k = x\right] - \mathbb{E}\left[f(\overline{X}) \,\middle|\, S_{k-1}, X_k = y\right]\right) \\ &= \sup_{x,y} \mathbb{E}\left[f_k(\overline{X}, x) - f_k(\overline{X}, y) \,\middle|\, S_{k-1}\right],\end{aligned} \tag{1.33}$$

where the last equality holds because X_k is independent of $\overline{X} \setminus X_k$; this follows from conditional probability,

$$\begin{aligned}&\Pr\left[\overline{X} = (z_1, \ldots, z_{k-1}, x, \ldots, z_n) \,\middle|\, S_{k-1} = (z_1, \ldots, z_{k-1}), X_k = x\right] = \\ &\quad \Pr\left[\overline{X} \setminus X_k = (z_1, \ldots, z_{k-1}, z_{k+1}, \ldots, z_n) \,\middle|\, S_{k-1} = (z_1, \ldots, z_{k-1})\right].\end{aligned}$$

Since f satisfies the Lipschitz condition with bound c, and $X_1, \ldots, X_n$ are independent, eq. (1.33) is at most c. We conclude the proof with theorem 1.13.29. □

Remark 1.13.34. McDiarmid is an instance of Azuma-Hoeffding with a Doob martingale from independent random variables and a function satisfying the Lipschitz condition. The condition to apply Azuma-Hoeffding is to have bounded increments, which are guaranteed by independence and Lipschitz condition bound c. Independence says that revealing X_k does not alter the conditional probability distribution of the unrevealed random variables, as they do not depend on X_k. Therefore, the effect of revealing X_k can be compared as if only the k^{th} coordinate changes, such that the martingale difference is bounded by the same c.

Exercise 1.13.35 (Exercise 13.8 [MU17])**.** In the bin-packing problem, we are given items with sizes $a_1, \ldots, a_n$ with $a_i \in [0, 1]$. The goal is to pack them into minimum number of bins, with each bin being able to hold any collection of items whose total sizes sum to at most 1. Suppose that each of the a_i is chosen independently according to some distribution, let P be the number of bins required in the best packing of the resulting items. Prove that

$$\Pr\left[|P - \mathbb{E}[P]| \geq \varepsilon\right] \leq 2\exp\left(-2\varepsilon^2/n\right).$$

Proof. Let $f(a_1, \ldots, a_n)$ be the minimum number of bins needed to pack n items. Then the Lipschitz condition bound of f is at most 1, since a bin has capacity at most 1. Starting with $f(a_1, \ldots, a_k = 0, \ldots, a_n)$ bins, and changing to $a_k = 1$, the worst one can do is to put a_k in a separate bin and preserve the arrangement for the other items. Hence

$$f(a_1, \ldots, a_k = 1, \ldots, a_n) \leq f(a_1, \ldots, a_k = 0, \ldots, a_n) + 1.$$

We then conclude the proof with theorem 1.13.33 with $c = 1$. □

Exercise 1.13.36 (Exercise 13.9 [MU17])**.** Consider an n-cube with $|V| = 2^n$. Let $S \subseteq V$ with $|S| > 0$, and $x \xleftarrow{R} V$. Let $D(x, S)$ be the minimum number of coordinates in which x and y differ over all points $y \in S$. Give a bound on

$$\Pr\left[|D(x, S) - \mathbb{E}[D(x, S)]| > \varepsilon\right].$$

Proof. Let $\overline{X} = (X_1, \ldots, X_n)$ be independent fair coin tosses, $f(\overline{X}) = D(\overline{X}, S)$, and $f_k(\overline{x}, y) = f(x_1, \ldots, x_{k-1}, y, \ldots, x_n)$. Let d be the Hamming distance from $\overline{x} \xleftarrow{R} \{0,1\}^n$ to $\overline{w} \in S$, which is $f(\overline{x})$, and we have the following:

- If $\overline{w}$ is still the closest element in S, then either $w_k = y$ and $f_k(\overline{x}, y) = d - 1$, or $w_k \neq y$ and $f_k(\overline{x}, y) = d + 1$.
- $f_k(\overline{x}, y) = d$, where $y \neq w_k$, but there is an element $\overline{w}' \in S$ that disagrees in $d + 1$ coordinates with $\overline{x}$, and $w'_k = y$.

Alternatively, changing one coordinate either increases or decreases the distance to any element in S by 1. Thus, the Lipschitz condition bound $c = 1$. We conclude with theorem 1.13.33 that

$$\Pr\left[|D(x, S) - \mathbb{E}[D(x, S)]| > \varepsilon\right] < 2\exp\left(-2\varepsilon^2/n\right).$$

□

Exercise 1.13.37 (Exercise 13.10 [MU17])**.** Consider generalizing Hoeffding's bound to independent random variables ranging in $[0,1]$, and derive a tail bound for $S_n = \sum_{i\in[1,n]} X_i$.

Proof. Consider a Doob martingale over

$$Z_0 = \mathbb{E}[S_n], \quad Z_k = \mathbb{E}[S_n \mid X_1, \ldots, X_k],$$

and S_n has Lipschitz condition bound $c = 1$. Immediately by theorem 1.13.33,

$$\Pr[|S_n - \mathbb{E}[S_n]| \geq \varepsilon] \leq 2\exp\left(-2\varepsilon^2/n\right).$$

□

Example 1.13.38 (DNA Pattern Matching)**.** Let $\overline{X} = (X_1, \ldots, X_n)$ be a sequence of characters sampled independently and uniformly at random over an alphabet Σ, and let $\mathbf{b} \in \Sigma^k$ be a fixed pattern string of k characters. Let F be the number of occurrences of $\mathbf{b}$ in the random string $\overline{X}$. By linearity of expectation, $\mathbb{E}[F] = (n-k+1)\cdot s^{-k}$. Let

$$Z_0 = \mathbb{E}[F], \quad Z_i = \mathbb{E}[F \mid X_1, \ldots, X_i]$$

be a Doob martingale. Using $|Z_i - Z_{i-1}| \leq k$, theorem 1.13.28 gives

$$\Pr[|F - \mathbb{E}[F]| \geq \varepsilon] \leq 2\exp\left(-\frac{\varepsilon^2}{2nk^2}\right).$$

Taking a closer look, $|Z_i - Z_{i-1}| \leq k$ is rather wasteful, as F as a function over $\overline{X}$ has Lipschitz condition bound $c = k$. Thus, by theorem 1.13.33 or theorem 1.13.29,

$$\Pr[|F - \mathbb{E}[F]| \geq \varepsilon] \leq 2\exp\left(-\frac{2\varepsilon^2}{nk^2}\right).$$

Exercise 1.13.39 (Exercise 13.16 [MU17])**.** A subsequence of a string is any string that can be obtained by deleting characters. Consider $\mathbf{x}, \mathbf{y} \overset{R}{\leftarrow} \{0,1\}^n$, and the longest common subsequence (LCS) of $\mathbf{x}$ and $\mathbf{y}$. Show that the expected length of the LCS is in $(c_1 n, c_2 n)$ where $c_1 > 1/2$ and $c_2 < 1$ when n is sufficiently large [18]. Use theorem 1.13.33 to show that the length of the LCS is sharply concentrated around its mean.

Proof. Let L_ℓ be the length of the LCS of two binary strings of length ℓ. Then we have

$$L_\ell + 1 > \mathbb{E}[L_{\ell+1} \mid L_\ell] > L_\ell + \frac{1}{2},$$

as $x_{\ell+1} = y_{\ell+1}$ with probability $1/2$, and $L_{\ell+1} = L_\ell + 1$ with non-zero probability when $x_{\ell+1} \neq y_{\ell+1}$.

Let $Z_i = (X_i, Y_i)$ for the i^{th} bit in $\mathbf{x}$ and $\mathbf{y}$. The Doob martingale is constructed by

$$Z_0 = \mathbb{E}[L_n], \quad Z_k = \mathbb{E}[L_n \mid Z_1, \ldots, Z_k].$$

The Lipschitz condition bound is 1, as any common subsequence's length increases or decreases by 1:

- Either the previous LCS remains, no matter what new assignment to the k^{th} coordinate.
- Or the LCS length is unchanged, but altering the k^{th} bits improves another common subsequence.

Hence, we conclude with theorem 1.13.33 by

$$\Pr[|L_n - \mathbb{E}[L_n]| \geq \varepsilon] \leq 2\exp\left(-2\varepsilon^2/n\right).$$

□

[18] There are interesting open problems on pushing the bounds closer.

Example 1.13.40 (Balls and Bins)**.** Recall in m balls and n bins model, we let $\{X_i\}_{i\in[1,m]}$ be the random variables for the bin into which the i^{th} ball falls, and let F be the number of empty bins after m bins are thrown. Then

$$Z_0 = \mathbb{E}[F], \quad Z_k = \mathbb{E}[F \mid X_1, \ldots, X_k]$$

is a Doob martingale. Moreover, the Lipschitz condition bound is 1, as changing which bin the i^{th} ball lands increases or decreases the number of empty bins after m balls at most by 1. We conclude with theorem 1.13.33 that

$$\Pr[|F - \mathbb{E}[F]| \geq \varepsilon] \leq 2\exp\left(-2\varepsilon^2/m\right),$$

and the expected number of empty bins is $\mathbb{E}[F] = n(1-1/n)^m$.

Exercise 1.13.41 (Exercise 13.18 [MU17])**.** We have shown before in Bloom filter (e.g., exercise 1.5.29) that the fractions of entries that are 0 in a Bloom filter is concentrated around $p' = (1-1/n)^{km}$, where m is the number of data items, k is the number of hash functions, and n is the size of the Bloom filter in bits. Derive a similar concentration result using a martingale inequality.

Proof. This is a direct adaptation of example 1.13.40, using a km balls and n bins model. Let $\{X_i\}_{i\in[1,km]}$ be the 0-1 random variables for which bit the hash hits, and Y be the number of 0 bits after m items, then $\mathbb{E}[Y] = p'n$, and

$$Z_0 = \mathbb{E}[Y], \quad Z_k = \mathbb{E}[Y \mid X_1, \ldots, X_k]$$

is a Doob martingale, and Y is a function over $\{X_i\}_{i\in[1,km]}$ with Lipschitz condition bound $c = 1$. By theorem 1.13.33,

$$\Pr[|Y - \mathbb{E}[Y]| \geq \varepsilon] \leq 2\exp\left(-\frac{2\varepsilon^2}{km}\right).$$

□

Exercise 1.13.42 (Exercise 13.20 [MU17])**.** We improve the bound in example 1.13.40 and reuse all defined notations. Let A_i denote the number of bins that are empty after the i^{th} ball is thrown.

- Show that $Z_{i-1} = A_{i-1}(1-1/n)^{m-i+1}$.
- Show that, if the i^{th} ball lands in a bin that is empty, $Z_i = (A_{i-1}-1)(1-1/n)^{m-i}$.
- Show that, if the i^{th} ball lands in a bin that is not empty, $Z_i = A_{i-1}(1-1/n)^{m-i}$.
- Show that theorem 1.13.29 applies with $d_i = (1-1/n)^{m-i}$, and

$$\Pr[|F - \mathbb{E}[F]| \geq \varepsilon] \leq 2\exp\left(-\frac{2\varepsilon^2(2n-1)}{n^2 - \mathbb{E}[F]^2}\right).$$

Proof. The probability of a bin remaining empty after $m - i + 1$ balls is $(1-1/n)^{m-i+1}$. By linearity of expectation,

$$Z_{i-1} = \mathbb{E}[F \mid X_1, \ldots, X_{i-1}] = A_{i-1}(1-1/n)^{m-i+1}.$$

Whether the i^{th} ball lands into an empty bin or not determines the number of empty bins onwards after the i^{th} ball. If the i^{th} ball lands into an empty bin, $A_i = A_{i-1} - 1$; otherwise, $A_i = A_{i-1}$. Still by linearity of expectation,

$$Z_i = \mathbb{E}[F \mid X_1, \ldots, X_i] = A_i(1-1/n)^{m-i}.$$

Let $B_i = (A_{i-1}/n - 1)\cdot(1-1/n)^{m-i}$, then $d_i = (1-1/n)^{m-i}$ such that $B_i \leq Z_i - Z_{i-1} \leq B_i + d_i$. By theorem 1.13.29,

$$\sum_{i\in[1,m]} d_i^2 = \sum_{j\in[0,m-1]} \left(1-\frac{1}{n}\right)^{2j} = \left(1-\left(1-\frac{1}{n}\right)^{2m}\right)\cdot\left(1-\left(1-\frac{1}{n}\right)^{2}\right)^{-1} = \frac{n^2 - n^2(1-1/n)^{2m}}{2n-1} = \frac{n^2 - \mathbb{E}[F]^2}{2n-1},$$

which proves

$$\Pr[|F - \mathbb{E}[F]| \geq \varepsilon] \leq 2\exp\left(-\frac{2\varepsilon^2(2n-1)}{n^2 - \mathbb{E}[F]^2}\right).$$

□

Remark 1.13.43. The key step is not McDiarmid itself, but showing the bounded increment in the martingale; it can either be directly shown, or by independence and Lipschitz condition bound.

Example 1.13.44 (Chromatic Number [SS87])**.** Given a random graph G in $G_{n,p}$, the chromatic number $\chi(G)$ is the minimum number of colors needed to color all vertices of the graph, so no adjacent vertices have the same color.

Let $G_i \subseteq G$ be the random subgraph induced by the set of vertices $[1, i]$, so G_i depends on G_{i-1}. Then

$$Z_0 = \mathbb{E}[\chi(G)], \quad Z_k = \mathbb{E}[\chi(G) \mid G_1, \dots, G_k]$$

forms a Doob martingale.

Now consider bounding $Z_k - Z_{k-1}$. For $H = G - v_k$, formed by removing the k^{th} vertex,

$$\chi(H) \leq \chi(G) \leq \chi(H) + 1, \tag{1.34}$$

where the first inequality holds because H has one less vertex than G, and the second inequality holds because adding v_k to H introduces at most one more new color.

Hence, consider G and G' differing only on the edges between v_k and $V \backslash \{v_k\}$. Then $\chi(G), \chi(G') \in [\chi(H), \chi(H)+1]$. By eq. (1.34), it is immediate that

$$\mathbb{E}[\chi(H) \mid G_1, \dots, G_{k-1}] \leq \mathbb{E}[\chi(G) \mid G_1, \dots, G_{k-1}] \leq \mathbb{E}[\chi(H) \mid G_1, \dots, G_{k-1}] + 1.$$

Moreover, since for any G and G' differing on the edges to v_k, $\chi(G), \chi(G') \in [\chi(H), \chi(H) + 1]$, for any G_k,

$$\mathbb{E}[\chi(H) \mid G_1, \dots, G_{k-1}] \leq \mathbb{E}[\chi(G) \mid G_1, \dots, G_k] \leq \mathbb{E}[\chi(H) \mid G_1, \dots, G_{k-1}] + 1.$$

Thus $d_i = 1$ for theorem 1.13.29, and we conclude that

$$\Pr\left[|\chi(G) - \mathbb{E}[\chi(G)]| \geq \varepsilon\sqrt{n}\right] \leq 2\exp\left(-2\varepsilon^2\right).$$

Exercise 1.13.45 (Exercise 13.19 [MU17])**.** Consider a random graph G in $G_{n,N}$ where $N = cn$ for $c > 0$. Let X be the number of isolated vertices, namely vertices of degree 0. Determine $\mathbb{E}[X]$ and show

$$\Pr\left[|X - \mathbb{E}[X]| \geq 2\varepsilon\sqrt{cn}\right] \leq 2\exp\left(-\varepsilon^2/2\right).$$

Proof. Let X_i be the 0-1 random variable for the i^{th} vertex being unconnected to any of N edges, and $\binom{n}{2}$ is the total number of possible edges. Then there are

$$\binom{\binom{n}{2} - (n-1)}{N} = \binom{\binom{n-1}{2}}{N}$$

graphs isolating the i^{th} vertex, and

$$\Pr[X_i = 1] = \binom{\binom{n-1}{2}}{N}\binom{\binom{n}{2}}{N}^{-1}.$$

By linearity of expectation,

$$\mathbb{E}\left[X = \sum_{i \in [1,n]} X_i\right] = n\binom{\binom{n-1}{2}}{N}\binom{\binom{n}{2}}{N}^{-1}.$$

Let $\{E_i\}_{i \in [1,N]}$ be random variables in $[1, \binom{n}{2}]$ for the location of the edge, and

$$Z_0 = \mathbb{E}[X], \quad Z_k = \mathbb{E}[X \mid E_1, \dots, E_k]$$

is a Doob martingale. An edge can decrease Z_{k-1} by 2 by joining 2 isolated vertices, so $|Z_k - Z_{k-1}| \leq c = 2$, and we conclude with theorem 1.13.28

$$\Pr\left[|X - \mathbb{E}[X]| \geq 2\varepsilon\sqrt{cn}\right] \leq 2\exp\left(-\frac{4\varepsilon^2 N}{2\sum_{i \in [1,N]} c^2}\right) = 2\exp\left(-\varepsilon^2/2\right).$$

An alternative way to further bound $Z_k - Z_{k-1}$ follows. Let I_k be the number of isolated vertices removed by the E_k, A_k be the number of isolated vertices after the E_k, such that $A_k = A_{k-1} - I_k$. A vertex remains isolated with probability

$$\Pr[X_i = 1 \mid E_1, \ldots, E_k] = \binom{\binom{n-1}{2} - k}{N-k} \binom{\binom{n}{2} - k}{N-k}^{-1} = q_k.$$

Moreover, we have the following iteration

$$q_{k-1} = \frac{\binom{n-1}{2} - (k-1)}{\binom{n}{2} - (k-1)} \cdot q_k.$$

By linearity of expectation, $Z_k = A_k \cdot q_k$. Then

$$Z_k - Z_{k-1} = (A_{k-1} - I_k) \cdot q_k - A_{k-1} \cdot \frac{\binom{n-1}{2} - (k-1)}{\binom{n}{2} - (k-1)} \cdot q_k = q_k \cdot \left(\frac{n-1}{\binom{n}{2} - (k-1)} \cdot A_{k-1} - I_k \right). \tag{1.35}$$

Notice that

$$\mathbb{E}[I_k \mid E_1, \ldots, E_{k-1}] = \frac{n-1}{\binom{n}{2} - (k-1)} \cdot A_{k-1},$$

and I_k conditioned on $E_1, \ldots, E_{k-1}$ is in $[0, 2]$. Hence eq. (1.35) is of interval length $d_k = 2q_k \leq 2$, and we conclude with theorem 1.13.29

$$\Pr\left[|X - \mathbb{E}[X]| \geq 2\varepsilon\sqrt{cn}\right] \leq 2\exp\left(-\frac{2 \cdot 4\varepsilon^2 N}{\sum_{k \in [1,N]} d_k^2}\right) = 2\exp\left(-2\varepsilon^2\right).$$

A few notes from prior attempts follow. We have

$$\begin{aligned}\mathbb{E}[X \mid E_1, \ldots, E_{k-1}] = {} & \mathbb{E}[X \mid E_1, \ldots, E_{k-1}, I_k = 2] \cdot \Pr[I_k = 2 \mid E_1, \ldots, E_{k-1}] \\ & + \mathbb{E}[X \mid E_1, \ldots, E_{k-1}, I_k = 1] \cdot \Pr[I_k = 1 \mid E_1, \ldots, E_{k-1}] \\ & + \mathbb{E}[X \mid E_1, \ldots, E_{k-1}, I_k = 0] \cdot \Pr[I_k = 0 \mid E_1, \ldots, E_{k-1}].\end{aligned}$$

Let Y be the number of isolated vertices for a graph in $G_{n,N-1}$. We have

$$\mathbb{E}[X \mid E_1, \ldots, E_{k-1}] \leq \mathbb{E}[Y \mid E_1, \ldots, E_{k-1}],$$

since the N^{th} edge can still remove isolated vertices by

$$\mathbb{E}[X \mid E_1, \ldots, E_{N-1}] \leq \mathbb{E}[Y \mid E_1, \ldots, E_{N-1}].$$

Let B_k be the number of edges removing 2 isolated vertices, C_k be the number of edges removing 1 isolated vertex, and D_k be the number of edges removing 0 isolated vertex. For $G_{n,N}$ and $G_{n,N-1}$, the invariant is

$$B_k + C_k + D_k + k = \binom{n}{2}.$$

Then, we have

$$\mathbb{E}[X \mid E_1, \ldots, E_{k-1}, I_k = 2] \geq \mathbb{E}[Y \mid E_1, \ldots, E_{k-1}] - 2,$$

as $D_k \geq D_{k-1}$, $B_k \leq B_{k-1}$, and $B_{k-1} + C_{k-1} + D_{k-1} = B_k + C_k + D_k + 1$ in this case. Similarly,

$$\mathbb{E}[X \mid E_1, \ldots, E_{k-1}, I_k = 1] \geq \mathbb{E}[Y \mid E_1, \ldots, E_{k-1}] - 1,$$

as $D_k \geq D_{k-1}$, $B_k \leq B_{k-1}$, and $B_{k-1} + C_{k-1} + D_{k-1} = B_k + C_k + D_k + 1$. Then, we have

$$\mathbb{E}[X \mid E_1, \ldots, E_{k-1}, I_k = 0] \leq \mathbb{E}[Y \mid E_1, \ldots, E_{k-1}],$$

where $B_k = B_{k-1}$, $C_k = C_{k-1}$, and $D_k = D_{k-1} - 1$. X and Y have $N - k$ edges to go, but X has one fewer edge not decreasing the number of isolated vertices, making it more likely to decrease the number of isolated vertices. □

1.14 Sample Complexity, VC Dimension, and Rademacher Complexity

TODO: ...

1.15 Pairwise Independence and Universal Hash Function

1.15.1 Pairwise Independence

1.15.2 Chebyshev's Inequality for Pairwise Independent Random Variables

1.15.3 Universal Families of Hash Functions

1.15.4 Application in Finding Heavy Hitters in Data Streams

TODO: *after balls and bins, Poisson, and maybe Probabilistic methods.*

1.16 Power Laws and Related Distributions

TODO: ...

1.17 Balanced Allocations and Cuckoo Hashing

TODO: *should be after balls bins?*

2 Pseudorandomness

TODO: *Vadhan* [*Vad*12] *Pseudorandomness*

3 Probabilitic Method

TODO: *Alon Spencer* [*AS*16] *Probabilistic Method*

4 Various Papers

4.1 Leftover Hash Lemma and HILL Entropy

TODO: *after entropy in* [MU17], *learn LHL* [ILL89] *and HILL entropy* [HILL99]

4.2 Discrete Gaussian Leftover Hash Lemma

TODO: *should learn from* [*AGHS*13]

4.3 Schwartz-Zippel by Dana Moshkovitz

The geometric viewpoint in [Mos10] is to foliate $\mathbb{F}_q^n$ by parallel lines, and count their intersections with the zero set of a polynomial. The key is to choose a direction for which no whole line lies in the zero set.

The existence of the direction uses only the following elementary fact.

Lemma 4.3.1 (Nonzero evaluation [Mos10])**.** *Let $\mathbb{F}_q$ be a finite field with q elements, and $f \in \mathbb{F}_q[X_1, \ldots, X_n]$ be nonzero with individual degrees less than q. Then f has a nonzero evaluation in $\mathbb{F}_q^n$.*

Proof. Fixing coordinates successively while keeping the remaining polynomial nonzero is always possible: if every substitution $X_i = a$ for $a \in \mathbb{F}_q$ over a nonzero remaining polynomial P, made P identically zero, then

$$\prod_{a \in \mathbb{F}_q} (X_i - a) = X_i^q - X_i \mid P,$$

contradicting $\deg_{X_i} P < q$. After all coordinates are fixed, the resulting value is nonzero. □

We now restate Schwartz-Zippel [Sch80, Zip79] in Moshkovitz's viewpoint [Mos10].

Lemma 4.3.2 (Schwartz-Zippel [Sch80, Zip79])**.** *Let $\mathbb{F}_q$ be a finite field with q elements, and $f \in \mathbb{F}_q[X_1, \ldots, X_n]$ be nonzero of total degree d, with $n \geq 1$. From uniformly random $\mathbf{x} \xleftarrow{R} \mathbb{F}_q^n$,*

$$\Pr\left[f(\mathbf{x}) = 0\right] \leq \frac{d}{q}.$$

Equivalently, as for the zero set of f over $\mathbb{F}_q^n$,

$$\left|\{\mathbf{x} \in \mathbb{F}_q^n : f(\mathbf{x}) = 0\}\right| \leq dq^{n-1}.$$

Proof. The cases $d = 0$ and $d \geq q$ are immediate, so assume $1 \leq d < q$.

Write $f = g + h$, where nonzero g consists of all the degree-d monomials, and thus $\deg h < d$. By lemma 4.3.1, g has a nonzero evaluation over $\mathbb{F}_q^n$, and by g being all degree-d monomials, $g(\mathbf{0}) = 0$. Hence, there exists $\mathbf{y} \in \mathbb{F}_q^n \setminus \{\mathbf{0}\}$ such that $g(\mathbf{y}) \neq 0$.

For any basepoint $\mathbf{x}$, substitution gives $p(t) = f(\mathbf{x} + t\mathbf{y})$ that $\deg_t p = d$, whose degree-d term has coefficient $g(\mathbf{y})$, and has at most d roots.

There are q^{n-1} distinct parallel lines $\mathbf{x} + t\mathbf{y}$, as $\left|\mathbb{F}_q^n / \mathrm{Span}\,(\mathbf{y})\right| = q^{n-1}$, and thus at most dq^{n-1} zero evaluations, as each line contributes at most d roots. □

The same counting pattern appears in the Ligero argument below: choose a direction along which every parallel line contains few exceptional points, then count across the partition. There the points are received words close to a code; here they are zeros of f.

4.4 Ligero Interleaved Linear Code Testing

For a code $C \subseteq \Sigma^n$ and a vector $\mathbf{v} \in \Sigma^n$,

- denote by $d(\mathbf{v}, C)$ the minimum distance of $\mathbf{v}$ from C, namely the number of positions in which $\mathbf{v}$ differs from the closest codeword in C,
- denote by $d(V, C)$ the minimum distance between a vector set V and a code C, that

$$d(V, C) = \min_{\mathbf{v} \in V} d(\mathbf{v}, C),$$

- denote by $\Delta(\mathbf{v}, C)$ the set of positions in which $\mathbf{v}$ differs from such a closest codeword,
- denote by $\Delta(V, C)$ the set of positions in which some $\mathbf{v} \in V$ differs from their closest codeword in C, that

$$\Delta(V, C) = \bigcup_{\mathbf{v} \in V} \Delta(\mathbf{v}, C).$$

Definition 4.4.1 (Interleaved Code)**.** Let $L \subset \mathbb{F}^n$ be an $[n, k, d]$ linear code over $\mathbb{F}$. We let L^m denote the $[n, mk, d]$ (interleaved) code over $\mathbb{F}^m$ whose codewords are all $\mathbf{U} \in \mathbb{F}^{m \times n}$, such that each row $\mathbf{u}_i$ of $\mathbf{U}$ satisfies $\mathbf{u}_i \in L$, and let $\mathbf{u}_{*,j}$ denote the j^{th} symbol (column) of $\mathbf{U}$.

A randomized algorithm [AHIV17] to test the membership of $\mathbf{U}$ in L^m follows, where $\mathbf{U}$ is stored in an oracle O:

- $\mathcal{V}$ samples $\mathbf{r} \xleftarrow{R} \mathbb{F}^m$ and sends over $\mathbf{r}$ to $\mathcal{P}$.
- $\mathcal{P}$ sends back $\mathbf{w} \leftarrow \mathbf{r}^{\mathsf{T}}\mathbf{U}$ to $\mathcal{V}$.
- $\mathcal{V}$ samples a set of positions $Q \xleftarrow{R} [1, n]^\ell$, and queries O for symbols $\mathbf{u}_{*,i}$ for all $i \in Q$.
- $\mathcal{V}$ checks if $\mathbf{w}$ is consistent with $\{\mathbf{u}_{*,i}\}_{i \in Q}$ by $\langle \mathbf{r}, \mathbf{u}_{*,i} \rangle = w_i$.

Lemma 4.4.2. *Let e be a positive integer such that $e < d/4$. Suppose $d(\mathbf{U}^*, L^m) > e$, then for a random $\mathbf{w}^*$ in* RowSpan$(\mathbf{U}^*)$,

$$\Pr\left[d\left(\mathbf{w}^*, L\right) \leq e\right] \leq \frac{e+1}{|\mathbb{F}|}.$$

Proof. We introduce the following lemma, stating that if there is an element in the rowspan of $\mathbf{U}^*$ has large distance, then such element is almost everywhere in the rowspan space.

Lemma 4.4.3. *Suppose $\mathbf{u}, \mathbf{v} \in$ RowSpan $(\mathbf{U}^*)$ with $d\left(\mathbf{u}, L\right) > 2e$, then over the line $f(\alpha) = \alpha \cdot \mathbf{u} + \mathbf{v}$ where $\alpha \in \mathbb{F}$,*

$$\Pr\left[d\left(\alpha \cdot \mathbf{u} + \mathbf{v}, L\right) \leq e\right] \leq |\mathbb{F}|^{-1}.$$

Proof. On the line $f(\alpha) = \alpha \cdot \mathbf{u} + \mathbf{v}$, suppose there are 2 distinct elements $f(\alpha_0)$ and $f(\alpha_1)$, such that

$$d\left(\alpha_0 \cdot \mathbf{u} + \mathbf{v}, L\right) \leq e \text{ and } d\left(\alpha_1 \cdot \mathbf{u} + \mathbf{v}, L\right) \leq e.$$

By triangle inequality,

$$d\left((\alpha_0 - \alpha_1) \cdot \mathbf{u}, L\right) \leq d\left(\alpha_0 \cdot \mathbf{u} + \mathbf{v}, L\right) + d\left(\alpha_1 \cdot \mathbf{u} + \mathbf{v}, L\right) \leq 2e,$$

this contradicts against the assumption of $d(\mathbf{u}, L) > 2e$. Hence, at most 1 element on $f(\alpha)$ has distance at most e. □

Corollary 4.4.4. *Suppose $\mathbf{u} \in$ RowSpan $(\mathbf{U}^*)$ has $d\left(\mathbf{u}, L\right) > 2e$, then for a random $\mathbf{w}^*$ in* RowSpan $(\mathbf{U}^*)$,

$$\Pr\left[d\left(\mathbf{w}^*, L\right) \leq e\right] \leq |\mathbb{F}|^{-1}.$$

Proof. We foliate RowSpan $(\mathbf{U}^*)$ with $|\mathbb{F}|^{m-1}$ parallel lines similar to [Mos10], where each line is defined by $\alpha \cdot \mathbf{u} + \mathbf{v}$. By lemma 4.4.3, each line has at most 1 element of distance at most e, and thus at most $|\mathbb{F}|^{m-1}$ such elements, which completes the proof. □

On the other hand, if every element in RowSpan $(\mathbf{U}^*)$ has distance to L at most $2e$, fixing a differed position j, we argue that there are at most 1 element over a line, with distance to L is at most e, and the element is not differed at j.

Lemma 4.4.5. *Suppose* $\mathbf{u}, \mathbf{v} \in \text{RowSpan}(\mathbf{U}^*)$ *with* $j \in \Delta(\mathbf{u}, L)$, *then*

$$\Pr\left[d(\alpha \cdot \mathbf{u} + \mathbf{v}, L) \leq e \bigwedge j \notin \Delta(\alpha \cdot \mathbf{u} + \mathbf{v}, L)\right] \leq |\mathbb{F}|^{-1}.$$

Proof. If there are 2 distinct elements on $f(\alpha)$ both have the j^{th} symbol matched, and both distances to L are at most e, then by triangle inequality,

$$\begin{aligned} d\left(\text{Span}\left(f(\alpha_0), f(\alpha_1)\right), L\right) &= \min_{a,b} d\left(a \cdot f(\alpha_0) + b \cdot f(\alpha_1), L\right) \\ &\leq d(\alpha_0 \cdot \mathbf{u} + \mathbf{v}, L) + d(\alpha_1 \cdot \mathbf{u} + \mathbf{v}, L) \\ &\leq 2e < d/2, \end{aligned}$$

all elements in the Span $\left(f(\alpha_0), f(\alpha_1)\right)$ are in the unique decoding radius, hence all elements have j^{th} symbol matched. Yet this contradicts $\mathbf{u} \in \text{Span}\left(f(\alpha_0), f(\alpha_1)\right)$, that $j \in \Delta(\mathbf{u}, L)$. Hence, there are at most 1 such element on the $f(\alpha)$. □

The following corollary is immediate by the space foliation idea in [Mos10].

Corollary 4.4.6. *For a random* $\mathbf{w}^*$ *in* RowSpan $(\mathbf{U}^*)$, *and* $j \in \Delta(\mathbf{U}^*, L^m)$,

$$\Pr\left[d(\mathbf{w}^*, L) \leq e \bigwedge j \notin \Delta(\mathbf{w}^*, L)\right] \leq |\mathbb{F}|^{-1}.$$

Since $d(\mathbf{U}^*, L^m) > e$, we union bound over at least $e+1$ differed positions, upper bounded by $(e+1) \cdot |\mathbb{F}|^{-1}$. □

An improved result of distance preserving of random linear combination over linear code is also in [AHIV17].

Lemma 4.4.7. *Let* e *be a positive integer such that* $e < d/3$. *Suppose* $d(\mathbf{U}^*, L^m) > e$, *then for a random* $\mathbf{w}^*$ *in* RowSpan$(\mathbf{U}^*)$,

$$\Pr\left[d(\mathbf{w}^*, L) \leq e\right] \leq \frac{e+1}{|\mathbb{F}|}.$$

Proof. We start by showing there exists elements in RowSpan $(\mathbf{U}^*)$ such that the distance to L greater than e.

Lemma 4.4.8. *Suppose* $d(\mathbf{U}^*, L^m) > e$, *and* $|\mathbb{F}| > e$, *then there exists* $\mathbf{w}^* \in \text{RowSpan}(\mathbf{U}^*)$ *such that* $d(\mathbf{w}^*, L) > e$.

Proof. Suppose all $\mathbf{w}^* \in \text{RowSpan}(\mathbf{U}^*)$ have $d(\mathbf{w}^*, L) \leq e$, then let $\mathbf{u}^* \in \text{RowSpan}(\mathbf{U}^*)$ maximizes the distance to L,

$$\max_{\mathbf{v}^* \in \text{RowSpan}(\mathbf{U}^*)} d(\mathbf{v}^*, L) = d(\mathbf{u}^*, L) \leq e.$$

There must exist another element $\mathbf{x}^*$ in the rowspan, such that

$$\Delta(\mathbf{x}^*, L) \setminus \Delta(\mathbf{u}^*, L) \neq \emptyset,$$

otherwise $d(\mathbf{u}^*, L) = d(\mathbf{U}^*, L^m) > e$, contradicting to the prior assumption of $d(\mathbf{u}^*, L) \leq e$.

We also want to know, if all $\mathbf{w}^* \in \text{RowSpan}(\mathbf{U}^*)$ are of distance at most e to L, does the nearest codewords follow the same linear combination as the $\mathbf{w}^*$'s (or we cannot analyze by the union of differed positions).

Construct a line $f(\alpha) = \alpha \cdot \mathbf{u}^* + \mathbf{x}^*$, and let $\mathbf{u}, \mathbf{x}, \mathbf{y}$ be the codewords within e distance to $\mathbf{u}^*, \mathbf{x}^*, f(\alpha)$, then

$$d(\alpha \cdot \mathbf{u}^* + \mathbf{x}^*, \alpha \cdot \mathbf{u} + \mathbf{x}) \leq d(\mathbf{u}^*, \mathbf{u}) + d(\mathbf{x}^*, \mathbf{x}) \leq 2e,$$

the first inequality holds by the triangle inequality, and the second holds by the assumption that any element in the rowspan is of distance to L at most e. Then the distance between 2 valid codewords is

$$d(\alpha \cdot \mathbf{u}^* + \mathbf{x}^*, \mathbf{y}) \leq d\left(\alpha \cdot \mathbf{u}^* + \mathbf{x}^*, f(\alpha)\right) + d\left(f(\alpha), \mathbf{y}\right) \leq 2e + e < d,$$

which leads to a contradiction, as any 2 valid codewords are at least of d distance. Hence, $\mathbf{y} = \alpha \cdot \mathbf{u} + \mathbf{x}$.

Assuming all $\mathbf{w}^* \in \text{RowSpan}(\mathbf{U}^*)$ have $d(\mathbf{w}^*, L) \leq e$, now we know the linear combination of $\mathbf{w}^*$'s is the same as the linear combination of the nearest $\mathbf{w}$ codewords. Let

$$\Delta_{\mathbf{u}^*,\mathbf{x}^*} = \Delta(\mathbf{u}^*, L) \cap \Delta(\mathbf{x}^*, L).$$

Then the probability for any position in $\Delta_{\mathbf{u}^*,\mathbf{x}^*}$ being cancelled on a random α is at most $|\Delta_{\mathbf{u}^*,\mathbf{x}^*}| \cdot |\mathbb{F}|^{-1}$ by union bound. Hence, by

$$|\Delta_{\mathbf{u}^*,\mathbf{x}^*}| < e < |\mathbb{F}|,$$

first inequality holds by $d(\mathbf{u}^*, L) < e$, then the probability that no position in $\Delta_{\mathbf{u}^*,\mathbf{x}^*}$ being cancelled is positive, which proves the existence of $\mathbf{w}^*$ that $d(\mathbf{w}^*, L) > e$. □

Since there exists elements in the rowspan such that the distance to L is greater than e, in order to strengthen the result from lemma 4.4.2, it suffices to show that in an affine subspace of $\mathbb{F}^n$, either all points are of distances within e to L, or almost all are not. This reduces to showing the same in 1-dimensional space.

Lemma 4.4.9. *For any $\mathbf{u}, \mathbf{v} \in \mathbb{F}^n$ constructing a line $\alpha \cdot \mathbf{u} + \mathbf{v}$,*

- *either every points have $d(\alpha \cdot \mathbf{u} + \mathbf{v}, L) \leq e$,*
- *or at most e points have $d(\alpha \cdot \mathbf{u} + \mathbf{v}, L) \leq e$.*

Proof. We show this with 2 vectors of Hamming weight at most e, and see if any element in the 1-dimensional affine subspace can be within distance e to any valid non-zero codeword. Let $\mathbf{u}^*, \mathbf{v}^*$ be 2 vectors of Hamming weights at most e, then they define a line $f(\alpha) = \alpha \cdot \mathbf{u}^* + \mathbf{v}^*$. The lemma can be proved in 2 cases:

- $\left|\text{supp}(\mathbf{u}^*) \cup \text{supp}(\mathbf{v}^*)\right| \leq e$. Then any element on the line are within distance e to $\mathbf{0}$.
- $\left|\text{supp}(\mathbf{u}^*) \cup \text{supp}(\mathbf{v}^*)\right| > e$. The intersection $\Delta_{\mathbf{u}^*,\mathbf{v}^*}$ has at most $e-1$ elements. By a similar argument of differed position cancelling in lemma 4.4.8, there are at most $e-1$ points on the line with overlapped disagreeing positions cancelled, hence there are at most $e-1$ points within distance e to $\mathbf{0}$.

  Suppose there is a non-zero codeword $\mathbf{c}$ such that an $\alpha \cdot \mathbf{u}^* + \mathbf{v}^*$ is within distance e to $\mathbf{c}$, but by triangle inequality,

  $$d(\mathbf{0}, \mathbf{c}) \leq d(\mathbf{0}, \alpha \cdot \mathbf{u}^* + \mathbf{v}^*) + d(\alpha \cdot \mathbf{u}^* + \mathbf{v}^*, \mathbf{c}) \leq d(\mathbf{0}, \mathbf{v}^*) + d(\mathbf{0}, \mathbf{u}^*) + d(\alpha \cdot \mathbf{u}^* + \mathbf{v}^*, \mathbf{c}) \leq 3e < d,$$

  which is a contradiction. Hence, there is only $\mathbf{0}$ that is within e distance to elements on the line.

□

By lemma 4.4.8 and lemma 4.4.9, the proof for lemma 4.4.7 is immediate. □

We conclude with the soundness analysis, when $\mathbf{U}^n$ is far from L^m.

Theorem 4.4.10. *Let $e < d/3$ be a positive integer. Suppose $d(\mathbf{U}^*, L^m) > e$, then for any malicious $\mathcal{P}$ strategy, $\mathcal{V}$ accepts with probability at most*

$$\left(1 - \frac{e+1}{n}\right)^t + \frac{e+1}{|\mathbb{F}|}.$$

Proof. Let $\mathbf{w}^{*\mathsf{T}} = \mathbf{r}^{\mathsf{T}}\mathbf{U}^*$, the probability upper bound follows:

$$\begin{aligned}
\Pr\left[\mathcal{V} \text{ accepts}\right] &\leq \Pr\left[\mathcal{V} \text{ accepts} \mid d(\mathbf{w}^*, L) > e\right] + \Pr\left[d(\mathbf{w}^*, L) \leq e\right] \\
&\leq \binom{n-e-1}{t} \cdot \binom{n}{t}^{-1} + \frac{e+1}{|\mathbb{F}|} \\
&\leq \left(1 - \frac{e+1}{n}\right)^t + \frac{e+1}{|\mathbb{F}|},
\end{aligned}$$

the first inequality holds by exercise 1.4.14 trick, while the second holds by lemma 4.4.7. □

4.5 Interleaved Linear Code Testing with Logarithmic Randomness

TODO: *Diamond Posen* [*DP*24]

4.6 Basefold Random Linear Foldable Code and the Distance Proof

Lemma 4.6.1 (Lemma 2 [CF24]). *Fix any generator matrix $\mathbf{G}_{d-1}$ where the encoding of any nonzero message $\mathbf{m} \in \mathbb{F}^{k_{d-1}}$ has fewer than t_{d-1} zeroes. For any subset $S \subseteq [1, n_{d-1}]$,*

- *if $|S| < t_{d-1}$, then $|m_d(S)| < |\mathbb{F}|^{2(t_{d-1}-|S|)}$,*
- *if $|S| \geq t_{d-1}$, then $|m_d(S)| = 1$.*

Proof. When $|S| \geq t_{d-1}$, only $\mathbf{m} = \mathbf{0}$ can make all alphabets over S being 0. Hence, $\dim\ker\left(G^{\mathsf{T}}_{d-1,S}\right) = 0$, $|m_d(S)| = 1$.

Let $T \subseteq [1, n_{d-1}] \setminus S$ of size $t_{d-1} - |S|$. We first have

$$\begin{aligned}\operatorname{rank}(G^{\mathsf{T}}_{d-1,T\cup S}) &\leq \operatorname{rank}\left(G^{\mathsf{T}}_{d-1,S}\right) + \operatorname{rank}\left(G^{\mathsf{T}}_{d-1,T}\right)\\ &\leq \operatorname{rank}\left(G^{\mathsf{T}}_{d-1,S}\right) + (t_{d-1} - |S|).\end{aligned}$$

By Rank-Nullity,

$$k_{d-1} = \operatorname{rank}\left(G^{\mathsf{T}}_{d-1,T\cup S}\right) + \dim\ker\left(G^{\mathsf{T}}_{d-1,T\cup S}\right),$$

since the linear code alphabets over $T \cup S$ is $\mathbf{0}$ iff $\mathbf{m} = \mathbf{0}$ [19], then

$$\dim\ker\left(\mathbf{G}^{\mathsf{T}}_{d-1,T\cup S}\right) = 0,$$

hence $\operatorname{rank}\left(G^{\mathsf{T}}_{d-1,T\cup S}\right) = k_{d-1}$, and $\operatorname{rank}\left(G^{\mathsf{T}}_{d-1,S}\right) \geq k_{d-1} - (t_{d-1} - |S|)$. Again by Rank-Nullity,

$$k_{d-1} = \operatorname{rank}\left(G^{\mathsf{T}}_{d-1,S}\right) + \dim\ker\left(G^{\mathsf{T}}_{d-1,S}\right),$$

and hence,

$$\dim\ker\left(G^{\mathsf{T}}_{d-1,S}\right) \leq t_{d-1} - |S|,$$

which completes the proof, that $|m_d(S)| \leq |F|^{2(t_{d-1}-|S|)}$. □

TODO: *basefold stuffs* [CF24].

[19] or by $t_{d-1} \geq k_{d-1}$ as $n_{d-1} - t_{d-1} + 1$ is the distance of the $(d-1)^{\text{th}}$ level linear code, which is no more than $n_{d-1} - k_{d-1} + 1$ by MDS code.

4.7 Bulletproof

TODO: ... [$BBB^{+}18$]

4.8 Sublinear Lattice ZKP for Arithmetic Circuits

The work [BBC+18] achieves sublinear communication complexity in the lattice based ZK argument system for linear relations. More specifically, they can prove knowledge of short pre-images $\{\mathbf{s}_i\}_{i\in[1,\ell]}$ over $\mathbb{Z}_q^v$ for a linear relation

$$\mathbf{AS} \equiv \mathbf{T} \mod q,$$

where $A \in \mathbb{Z}_q^{r\times v}$ and $\mathbf{T} \in \mathbb{Z}_q^{r\times \ell}$ are public, and $\mathbf{S} = [\mathbf{s}_1 \mid \ldots \mid \mathbf{s}_\ell]$. The trick is to use a challenge matrix $\mathbf{C} \xleftarrow{R} \{0,1\}^{\ell\times(\lambda+2)}$, such that $\mathbf{SC}$ is still full of short coefficients, but the number of relations shrink from ℓ to $\lambda+2$. Applying the same technique in [Lyu09, Lyu12], we achieve zero knowledge.

Ring notation and norms. Let $d = d(\lambda)$ be a power of two, and let $R = \mathbb{Z}[X]/(X^d+1)$, $R_q = R/qR$, and $R_p = R/pR$. For $f \in R_q$, we always take the centered coefficient representative before measuring its size. If this representative is $f = \sum_{i=0}^{d-1} f_i X^i$, then f_i denotes the coefficient of X^i. Define

$$\|f\|_\infty = \max_{i\in[0,d-1]} |f_i|, \qquad \|f\|_2 = \left(\sum_{i=0}^{d-1} f_i^2\right)^{1/2}.$$

For $\mathbf{z} = (z_1, \ldots, z_m) \in R^m$, define

$$\|\mathbf{z}\|_\infty = \max_{j\in[m]} \|z_j\|_\infty, \qquad \|\mathbf{z}\|_2 = \left(\sum_{j=1}^{m} \|z_j\|_2^2\right)^{1/2}.$$

Commitment Scheme. For a commitment scheme $\Pi_{\mathsf{Com}} = (\mathsf{Setup}, \mathsf{Com})$, let $\mathcal{M}_{\mathsf{Com}}$ denote the message space, and $\mathcal{R}_{\mathsf{Com}}$ denote the randomness space. The setup algorithm outputs a commitment key $\mathsf{ck} \leftarrow \mathsf{Setup}(1^\lambda, 1^\ell)$, and the commitment algorithm takes $\mathbf{m} \in \mathcal{M}_{\mathsf{Com}}$ and $\mathbf{r} \in \mathcal{R}_{\mathsf{Com}}$ as inputs, and outputs $\mathbf{c} \leftarrow \mathsf{Com}_{\mathsf{ck}}(\mathbf{m};\mathbf{r})$. An opening of $\mathbf{c}$ is a pair $(\mathbf{m},\mathbf{r})$, and verification recomputes $\mathsf{Com}_{\mathsf{ck}}(\mathbf{m};\mathbf{r})$.

Definition 4.8.1 (Binding). $\Pi_{\mathsf{Com}} = (\mathsf{Setup}, \mathsf{Com})$ is computationally binding if for every efficient adversary $\mathcal{A}$,

$$\Pr\left[\mathsf{Com}_{\mathsf{ck}}(\mathbf{m}_0;\mathbf{r}_0) = \mathsf{Com}_{\mathsf{ck}}(\mathbf{m}_1;\mathbf{r}_1) \wedge \mathbf{m}_0 \neq \mathbf{m}_1 \;\middle|\; \begin{array}{r} \mathsf{ck} \leftarrow \mathsf{Setup}(1^\lambda, 1^\ell), \\ (\mathbf{m}_0,\mathbf{r}_0,\mathbf{m}_1,\mathbf{r}_1) \leftarrow \mathcal{A}(1^\lambda, \mathsf{ck}) \end{array}\right] \leq \mathsf{negl}(\lambda),$$

If the same condition holds for unbounded adversaries, then $\Pi_{\mathsf{Com}} = (\mathsf{Setup}, \mathsf{Com})$ is statistically binding.

Definition 4.8.2 (Hiding). $\Pi_{\mathsf{Com}} = (\mathsf{Setup}, \mathsf{Com})$ is computationally hiding if for every efficient adversary $(\mathcal{A}_0, \mathcal{A}_1)$,

$$\left|\Pr\left[b = b' \;\middle|\; \begin{array}{r} \mathsf{ck} \leftarrow \mathsf{Setup}(1^\lambda, 1^\ell), \\ (\mathbf{m}_0, \mathbf{m}_1, \mathsf{st}) \leftarrow \mathcal{A}_0(1^\lambda, \mathsf{ck}), \\ b \xleftarrow{R} \{0,1\}, \\ \mathbf{r} \leftarrow \mathcal{R}_{\mathsf{Com}}, \\ \mathbf{c}_b \leftarrow \mathsf{Com}_{\mathsf{ck}}(\mathbf{m}_b;\mathbf{r}), \\ b' \leftarrow \mathcal{A}_1(\mathsf{st}, \mathbf{c}_b) \end{array}\right] - \frac{1}{2}\right| \leq \mathsf{negl}(\lambda).$$

If the same condition holds for unbounded adversaries, then $\Pi_{\mathsf{Com}} = (\mathsf{Setup}, \mathsf{Com})$ is statistically hiding.

Module-SIS. The binding property of Ajtai commitments is from the hardness of module-SIS assumption.

Definition 4.8.3 (Module-SIS [LS15]). Let $q = q(\lambda), m = m(\lambda), n = n(\lambda), \beta = \beta(\lambda)$, and $d = d(\lambda)$. We say $\mathsf{MSIS}_{n,m,d,q,\beta}$ holds if for any efficient adversary $\mathcal{A}$ outputting $\mathbf{z} \in R^m$, the following holds

$$\Pr\left[\mathbf{Az} \equiv \mathbf{0} \mod q \wedge 0 < \|\mathbf{z}\|_\infty \leq \beta \mid \mathbf{A} \xleftarrow{R} R_q^{n\times m}, \mathbf{z} \leftarrow \mathcal{A}(\mathbf{A})\right] \leq \mathsf{negl}(\lambda).$$

When $n = 1$, the assumption specializes to the Ring-SIS assumption [PR06, LM06].

Ajtai Commitment. We now recall the Ajtai commitment instantiated from module-SIS assumption.

Definition 4.8.4 (Ajtai Commitment [Ajt96])**.** Over the above ring, we define the following parameters:

- ℓ: the message dimension,
- k: the number of output ring elements.
- $\mathcal{M}_{\mathsf{Ajtai}}$: the message space R_p^{ℓ}, where elements of R_p are lifted to R_q by their centered coefficient representatives.
- $\mathcal{R}_{\mathsf{Ajtai}}$: the randomness space $R_q^{2k\log_p q}$.
- $\mathcal{D}_{\mathsf{Ajtai}}$: the subgaussian distribution $\mathcal{D}_{\sigma}^{2k\log_p q}$ over the randomness space.
- B_{Ajtai}: the binding space defined as follows,

$$B_{\mathsf{Ajtai}} = \left\{ \mathbf{s} = \left[\mathbf{r}^{\mathsf{T}} \mid \mathbf{m}^{\mathsf{T}}\right]^{\mathsf{T}} \in R^{\ell+2k\log_p q} \ : \ \|\mathbf{s}\|_{\infty} \le B \right\}.$$

We construct $\Pi_{\mathsf{Ajtai}} = (\mathsf{Setup}, \mathsf{Com})$ over R_p as follows:

- $\mathsf{Setup}(1^{\lambda}, 1^{\ell})$: Sample $\mathbf{A}_1 \xleftarrow{\text{R}} R_q^{k\times 2k\log_p q}$ and $\mathbf{A}_2 \xleftarrow{\text{R}} R_q^{k\times \ell}$, and output $\mathsf{ck} = \mathbf{A} = [\mathbf{A}_1 \mid \mathbf{A}_2]$.
- $\mathsf{Com}_{\mathsf{ck}}(\mathbf{m}; \mathbf{r})$: On input the message $\mathbf{m} \in R_p^{\ell}$ and $\mathbf{r} \in R_q^{2k\log_p q}$, output $\mathbf{c} = \mathbf{A}_1\mathbf{r} + \mathbf{A}_2\mathbf{m} \in R_q^{k}$.

An opening is $(\mathbf{m}, \mathbf{r})$, and verification recomputes the commitment.

Lemma 4.8.5. *Ajtai commitment is computationally binding and statistically hiding.*

Proof. Hiding follows from the Leftover Hash Lemma: $(\mathbf{A}_1, \mathbf{A}_1\mathbf{r})$ is statistically close to uniformly random, and hence the commitment scheme is statistically hiding. Let $n' = k, m' = \ell + 2k\log_p q$, and $2B \le \beta$, so long as $\mathsf{MSIS}_{n',m',d,q,\beta}$ is hard, the commitment scheme is computationally binding. □

Amortized Linear Relation Proof. We now describe the amortized proof for many linear relations over R_q. The public statement consists of matrices $\mathbf{A} \in R_q^{r\times v}$ and $\mathbf{T} \in R_q^{r\times \ell}$, and the witness is $\mathbf{S} = [\mathbf{s}_1 \mid \cdots \mid \mathbf{s}_{\ell}] \in R^{v\times \ell}$, where each $\mathbf{s}_i$ has ℓ_{∞}-norm upper bounded by a bound B_{lin}, satisfying $\mathbf{AS} \equiv \mathbf{T} \mod q$.

By linearity of the Ajtai commitment definition 4.8.4, the commitments to the compressed columns $\mathbf{SC}$ can be obtained by applying the same linear combinations to the commitments of the columns of $\mathbf{S}$.

We begin by describing the "compression lemma" that help establish amortized proof.

Lemma 4.8.6. *Let $\mathbb{G}$ be an abelian group, and $\mathbf{x} = [x_1, \ldots, x_{\ell}]^{\mathsf{T}} \in \mathbb{G}^{\ell}$ be nonzero. For $\mathbf{c} \xleftarrow{R} \{0,1\}^{\ell}$, $\Pr[\langle \mathbf{c}, \mathbf{x}\rangle = 0] \le 1/2$. Consequently, for $\mathbf{C} \xleftarrow{R} \{0,1\}^{\ell\times t}$, $\Pr\left[\mathbf{x}^{\mathsf{T}}\mathbf{C} = \mathbf{0}^{\mathsf{T}}\right] \le 2^{-t}$.*

Proof. Pick i with $x_i \neq 0$, and fix all c_j for $j \neq i$. The two possible sums differ by x_i, so at most one of them is zero. The conditional probability is at most $1/2$, and averaging over the fixed bits gives the claim. □

Immediate by lemma 4.8.6, if $\mathbf{AS'} \not\equiv \mathbf{T} \mod q$, for $\mathbf{C} \xleftarrow{\text{R}} \{0,1\}^{\ell\times t}$, $\Pr\left[\mathbf{AS'C} \equiv \mathbf{TC} \mod q\right] \le 2^{-t}$.

Protocol Construction. The prover commits to each $\mathbf{s}_i$ using the Ajtai commitment definition 4.8.4.

Let C be the challenge set, that is $\{0,1\}$ when $R = \mathbb{Z}$, otherwise $\{0\} \cup \{\pm X^j\}_{j<d}$ when $R = \mathbb{Z}[X]/(X^d + 1)$.

Commit. For $i \in [1, \ell]$, the prover samples $\mathbf{r}_i \leftarrow \mathcal{D}_{\mathsf{Ajtai}}$ and sends $\mathbf{c}_i = \mathsf{Com}_{\mathsf{ck}}(\mathbf{s}_i; \mathbf{r}_i)$. Let $\mathbf{C}_{\mathsf{com}} = [\mathbf{c}_1 \mid \ldots \mid \mathbf{c}_{\ell}]$.

Challenge. The verifier samples $\mathbf{C} \xleftarrow{\text{R}} C^{\ell\times(\lambda+2)}$.

Response. The prover computes $\mathbf{Y} \leftarrow \mathbf{SC}$ and the compressed randomness $\mathbf{R} = [\mathbf{r}_1 \mid \ldots \mid \mathbf{r}_{\ell}]\mathbf{C}$, and sends over $(\mathbf{Y}, \mathbf{R})$.

Verification. The verifier checks $(\mathbf{Y}, \mathbf{R})$ are consistent with the compressed commitments by $\mathbf{C}_{\mathsf{Com}}\mathbf{C} = \mathsf{Com}_{\mathsf{ck}}(\mathbf{Y}; \mathbf{R})$.

Completeness is immediate.

Special Knowledge Soundness. We now show the (special) knowledge soundness of an accepting prover.

Here, v is the witness dimension for each relation, $t = \lambda + 2$, and ℓ is the number of relations being batched. Fix a statement $(\mathbf{A}, \mathbf{T})$ and a first message $\mathbf{C}_{\mathsf{Com}}$.

Theorem 4.8.7. *If $\mathcal{P}^*$ responds to a random challenge $\mathbf{C} \xleftarrow{R} C^{\ell \times t}$, and convinces the verifier with probability $\varepsilon \geq 2^{-\lambda}$, then the extractor should recover a matrix $\mathbf{S} \in R^{v \times \ell}$ such that $\mathbf{AS} \equiv c\mathbf{T} \mod q$[20], together with extracted randomness, open the commitments in $\mathbf{C}_{\mathsf{Com}}$, and the extracted columns are short. The extractor runs in expected time $\mathsf{poly}(\lambda, 1/\varepsilon)$.*

We start by describing how to extract a column of the witness $(\mathbf{s}_i, \mathbf{r}_i)$, which can be generalized to all columns. For an extractor $\mathcal{E}^{\mathcal{P}^*}$, it rewinds the prover to its first message $\mathbf{C}_{\mathsf{Com}}$ (so as to pin down the prover randomness), together with all rows of the challenge $\mathbf{C} \in C^{\ell \times t}$. It then resamples only the i^{th} row of $\mathbf{C}$, yielding $\mathbf{C}'$.

We now introduce a helpful lemma via probabilistic method related to the extraction process.

Lemma 4.8.8 (Heavy Row Lemma [Dam10])**.** *Let $\mathbf{H} \in \{0,1\}^{n \times m}$ has density at least ε, namely at least εnm entries are 1. We say that a row $a \in [1, n]$ is heavy, if there are more than $\varepsilon m/2$ entries are 1, namely the density in this row is at least $\varepsilon/2$. Then a uniformly random 1-entry of $\mathbf{H}$ lies in a heavy row with probability at least $1/2$.*

Proof. The total number of 1-entries in $\mathbf{H}$ is at least εmn, while the number of 1-entries in non-heavy rows is at most $\varepsilon mn/2$. Hence, the probability of an 1-entry lying in a light row is at most $1/2$, and an 1-entry lying in a heavy row is at least $1/2$. □

Therefore, consider building up a matrix $\mathbf{H}$ over $\{0,1\}$ as follows. The rows are indexed by the prover randomness and by all challenge rows except the i^{th} row. The columns are indexed by the possible i^{th} row challenges $\boldsymbol{\alpha}_i \in C^t$. The entry is 1 if $\mathcal{P}^*$ gives an accepting response on the corresponding full challenge matrix, and 0 otherwise.

The density of $\mathbf{H}$ is at least ε as the prover convinces the verifier with probability at least ε. Hence, by lemma 4.8.8, conditioned on a random challenge is accepting, with probability at least $1/2$ the corresponding row of $\mathbf{H}$ is heavy. Thus, after the extractor finds one accepting transcript in a heavy row, resampling only the i^{th} row in $\mathbf{C}$ gives another accepting transcript with probability at least $\varepsilon/2$.

Immediately, this gives a randomized algorithm for the extractor to extract a column of witness $(\mathbf{s}_i, \mathbf{r}_i)$ in expected polynomial time. First, the extractor samples accepting $(\mathbf{C}, \mathbf{Y}, \mathbf{R})$ on the first $\mathbf{C}_{\mathsf{Com}}$. Then, fix the prover randomness and $\mathbf{C}$ except the i^{th} row. Resample the i^{th} row independently up to $N = O(\lambda/\varepsilon)$ times. If one obtains an accepting $(\mathbf{C}', \mathbf{Y}', \mathbf{R}')$ whose i^{th} row differs from that of $\mathbf{C}$, continue with the extraction below. Otherwise, abort and restart.

The runtime analysis is similar to the expectation of a geometrically distributed random variable. Similar to the relaxation in lemma 1.5.36, we have

$$\begin{aligned}\Pr[\text{abort}] &= \Pr\left[\text{abort} \mid i^{\text{th}} \text{ row of } \mathbf{C} \text{ is heavy for } \mathbf{H}\right] \cdot \Pr\left[i^{\text{th}} \text{ row of } \mathbf{C} \text{ is heavy for } \mathbf{H}\right] \\ &\quad + \Pr\left[\text{abort} \mid i^{\text{th}} \text{ row of } \mathbf{C} \text{ is light for } \mathbf{H}\right] \cdot \Pr\left[i^{\text{th}} \text{ row of } \mathbf{C} \text{ is light for } \mathbf{H}\right] \\ &\leq \Pr\left[\text{abort} \mid i^{\text{th}} \text{ row of } \mathbf{C} \text{ is heavy for } \mathbf{H}\right] + \Pr\left[i^{\text{th}} \text{ row of } \mathbf{C} \text{ is light for } \mathbf{H}\right].\end{aligned}$$

The second term is at most $1/2$ by the lemma 4.8.8. The first term is upper bounded as follows. Since the i^{th} row of $\mathbf{C}$ is heavy on $\mathbf{H}$, the probability of choosing a rejecting challenge is at most $1 - \varepsilon/2$, also excluding the first accepting $\mathbf{C}$, then the rejecting probability is at most $1 - \varepsilon/2 + |C|^{-t}$. Repeating N times, the first term is at most $(1 - \varepsilon/2 + |C|^{-t})^N$. Thus, the aborting probability is at most

$$\frac{1}{2} + \left(1 - \frac{\varepsilon}{2} + \frac{1}{|C|^t}\right)^N.$$

If $\varepsilon > 2^{-\lambda}$, then $1 - \varepsilon/2 + |C|^{-t} > 1 - \varepsilon/4 > \exp(-\varepsilon/4)$, which leads to the final upper bound being $1/2 + 2^{-\lambda}$. Since each run takes $O(N)$ time, then extractor runs in expected $\mathsf{poly}(\lambda, 1/\varepsilon)$.

[20] When $R = \mathbb{Z}$, $c = 1$; otherwise when R is a polynomial ring, $c = 2$.

Suppose the extractor obtains two accepting transcripts $(\mathbf{C}, \mathbf{Y}, \mathbf{R})$ and $(\mathbf{C}', \mathbf{Y}', \mathbf{R}')$ whose challenges differ only on the i^{th} row, and differ nontrivially in that row. Write $\boldsymbol{\alpha}_i, \boldsymbol{\alpha}'_i \in C^t$ for the two i^{th} rows and let

$$\Delta\mathbf{C} = \mathbf{C} - \mathbf{C}', \qquad \Delta\mathbf{Y} = \mathbf{Y} - \mathbf{Y}', \qquad \Delta\mathbf{R} = \mathbf{R} - \mathbf{R}',$$

then the commitment openings satisfy

$$\mathbf{C}_{\mathsf{Com}} \cdot \Delta\mathbf{C} = \mathsf{Com}_{\mathsf{ck}}(\Delta\mathbf{Y}; \Delta\mathbf{R}).$$

Choose a coordinate j such that $\alpha = (\boldsymbol{\alpha}_i - \boldsymbol{\alpha}'_i)_j \in C - C$, then the j^{th} column of the first equation gives

$$\mathsf{Com}_{\mathsf{ck}}(\Delta\mathbf{y}_j; \Delta\mathbf{r}_j) \equiv \mathbf{A}_1 \cdot \Delta\mathbf{r}_j + \mathbf{A}_2 \cdot \Delta\mathbf{y}_j \equiv \alpha\mathbf{c}_i \mod q,$$

so $(\Delta\mathbf{y}_j, \Delta\mathbf{r}_j)$ is an opening of the scaled commitment $\alpha\mathbf{c}_i$.

By the challenge-set property, there exists a short $g \in R$ such that $g\alpha = c$. Hence

$$\mathsf{Com}_{\mathsf{ck}}(g \cdot \Delta\mathbf{y}_j; g \cdot \Delta\mathbf{r}_j) \equiv c\mathbf{c}_i \mod q.$$

The extractor therefore obtains the i^{th} scaled opening, and repeating the same row-local procedure for every $i \in [1, \ell]$ yields openings of all columns of $c\mathbf{C}_{\mathsf{Com}}$. Shortness follows from the response norm checks in accepting transcripts; the exact checks and constants are accounted for in the norm-growth analysis.

4.9 A Non-PCP Approach to Succinct Quantum-Safe ZK

The high-level idea of [BLNS20] is twofold. On one hand, it upgrades the amortized linear-relation proof for Ajtai commitments from [BBC+18] into a *levelled* SIS commitment, making the commitment object recursively compressing. On the other hand, it gives a lattice analogue of the Bulletproofs folding argument [BBB+18]. Concretely, we have:

- levelled SIS commitments, with communication

$$O\left(N^{1/(d+1)} \cdot \left(d^3\lambda \log^2 N + d\lambda^2\right)\right)$$

  for d levels, and hence $\mathsf{polylog}(N)$ communication on $d+1 = \log N$;
- a lattice analogue of the Bulletproofs folding argument [BBB+18], with smaller polylogarithmic proof size but much larger extraction slack in knowledge soundness.

N is the number of committed secret values. Recall the note section 4.8 on [BBC+18] that the one-level construction has $\tilde{O}(\sqrt{N})$ communication after balancing the number of commitments and the opening proof size after amortization. We begin by showing how [BLNS20] reduces the communication via a two-level commitment instantiation. As in the previous lattice argument note, we suppress the zero-knowledge layer: the commitment is hiding, and the proof is made zero-knowledge by Gaussian masking with rejection sampling in the Fiat-Shamir-with-aborts style.

Two-Level Commitment Map. The levelling idea starts by committing to commitments. On message $\mathbf{S} \in \mathbb{Z}^{m_1 m_2 \times m_3}$, commitment keys $\mathbf{A}_1 \in \mathbb{Z}_{q_1}^{n \times n m_1}$, $\mathbf{A}_2 \in \mathbb{Z}_{q_2}^{n \times m_2}$, and $q_1 > q_2$. The top commitment is

$$\mathbf{T} = \mathbf{A}_1 \cdot \left((\mathbf{I}_{m_1} \otimes \mathbf{A}_2) \cdot \mathbf{S} \mod q_2\right) \mod q_1.$$

This is SIS-commitment-like at each level: the lower level compresses blocks of $\mathbf{S}$, and the upper level compresses the lower-level commitments. At this point this is just the commitment map: an honest opening is a matrix $\mathbf{S}$ whose image under the above map is $\mathbf{T}$, and $\mathbf{S}$ is short.

General Levelled Commitments. For d levels, let $N = \prod_{i \in [1,d+1]} m_i$, $m_0 = 1$, and $q_1 > \cdots > q_d$. The bottom matrix is $\mathbf{A}_d \in \mathbb{Z}_{q_d}^{n \times m_d}$, and for $i < d$, $\mathbf{A}_i \in \mathbb{Z}_{q_i}^{n \times n m_i}$. Define $M_{i,j} = m_i \cdots m_j$ and $M_i = M_{1,i}$, and define recursively

$$F_{i,j}(\mathbf{S}) = \begin{cases} \mathbf{A}_i \mathbf{S} \mod q_i, & i = j, \\ F_{i,j-1}\left(\left(\mathbf{I}_{M_{i,j-1}} \otimes \mathbf{A}_j\right) \cdot \mathbf{S} \mod q_j\right), & i < j. \end{cases}$$

The statement is $\mathbf{T} \leftarrow F_{1,d}(\mathbf{S})$, where d is the number of levels.

Folding Operations. Similar to the challenges in section 4.8, the challenges are matrices over the challenge set C the same as the one in section 4.8, with shape $\mathbf{C}_1 \xleftarrow{R} C^{m_{d+1} \times \lambda}$, and $\mathbf{C}_i \xleftarrow{R} C^{m_{i-1}\lambda \times \lambda}$ for $i \in [2, d]$.

For a matrix $\mathbf{Z}$ whose rows are split into m_i consecutive blocks $\mathbf{Z}_1, \ldots, \mathbf{Z}_{m_i}$ of equal size, define the block transpose

$$\mathsf{BT}_i(\mathbf{Z}) = \left[\mathbf{Z}_1 \mid \ldots \mid \mathbf{Z}_{m_i}\right].$$

We also use the recursive folding map Fold_i from [BLNS20]. The base case is

$$\mathsf{Fold}_0(\mathbf{U}_1; \mathbf{C}_1) = \mathbf{U}_1 \mathbf{C}_1.$$

For $i > 0$, split the input tuple $(\mathbf{U}_1, \ldots, \mathbf{U}_{M_i})$ into m_i consecutive subtuples, such that for $j \in [1, m_i]$,

$$\bar{\mathbf{U}}_j = \left(\mathbf{U}_{(j-1)M_{i-1}+1}, \ldots, \mathbf{U}_{jM_{i-1}}\right).$$

Then

$$\mathsf{Fold}_i(\mathbf{U}_1,\dots,\mathbf{U}_{M_i};\mathbf{C}_1,\dots,\mathbf{C}_{i+1}) = \mathsf{BT}_i\begin{pmatrix}\mathsf{Fold}_{i-1}(\bar{\mathbf{U}}_1;\mathbf{C}_1,\dots,\mathbf{C}_i)\\ \vdots \\ \mathsf{Fold}_{i-1}(\bar{\mathbf{U}}_{m_i};\mathbf{C}_1,\dots,\mathbf{C}_i)\end{pmatrix}\mathbf{C}_{i+1}.$$

Lemma 4.9.1 (Fold properties [BLNS20]). *Fix i and $\{\mathbf{U}_j\}_{j\in[1,M_i]}$ with compatible dimensions. For challenges $\{\mathbf{C}_j\}_{j\in[1,i+1]}$ with the shapes above:*

- *there exist matrices $\{\mathbf{D}_i\}_{i\in[1,M_t]}$ with $\|\mathbf{D}_i\|_\infty \le \lambda^i$ such that*

$$\mathsf{Fold}_i(\mathbf{U}_1,\dots,\mathbf{U}_{M_i};\mathbf{C}_1,\dots,\mathbf{C}_{i+1}) = \sum_{t\in[1,M_i]}\mathbf{U}_t\mathbf{D}_t;$$

- *for every compatible matrix $\mathbf{B}$,*

$$\mathbf{B}\cdot\mathsf{Fold}_i(\mathbf{U}_1,\dots,\mathbf{U}_{M_i};\mathbf{C}_1,\dots,\mathbf{C}_{i+1}) = \mathsf{Fold}_i(\mathbf{BU}_1,\dots,\mathbf{BU}_{M_i};\mathbf{C}_1,\dots,\mathbf{C}_{i+1});$$

- *if each $\mathbf{U}_t$ is vertically split as $\mathbf{U}_t = [\mathbf{U}_{t,1}^\mathsf{T} \mid \cdots \mid \mathbf{U}_{t,r}^\mathsf{T}]^\mathsf{T}$, then*

$$\mathsf{Fold}_i(\mathbf{U}_1,\dots,\mathbf{U}_{M_i};\mathbf{C}_1,\dots,\mathbf{C}_{i+1}) = \begin{bmatrix}\mathsf{Fold}_i(\mathbf{U}_{1,1},\dots,\mathbf{U}_{M_i,1};\mathbf{C}_1,\dots,\mathbf{C}_{i+1})\\ \vdots \\ \mathsf{Fold}_i(\mathbf{U}_{1,r},\dots,\mathbf{U}_{M_i,r};\mathbf{C}_1,\dots,\mathbf{C}_{i+1})\end{bmatrix}.$$

Proof. We prove all three parts by induction on i. For $i = 0$, the definition gives $\mathsf{Fold}_0(\mathbf{U}_1;\mathbf{C}_1) = \mathbf{U}_1\mathbf{C}_1$. Thus the first part holds with $\mathbf{D}_1 = \mathbf{C}_1$, and the second and third parts are immediate.

Assume the claims hold for $i-1$. Split $(\mathbf{U}_1,\dots,\mathbf{U}_{M_i})$ into subtuples $\bar{\mathbf{U}}_1,\dots,\bar{\mathbf{U}}_{m_i}$ as was defined Fold_i, and write

$$\mathbf{W}_j = \mathsf{Fold}_{i-1}(\bar{\mathbf{U}}_j;\mathbf{C}_1,\dots,\mathbf{C}_i).$$

Also split the rows of $\mathbf{C}_{i+1}$ into m_i blocks $\mathbf{C}_{i+1}^{(1)},\dots,\mathbf{C}_{i+1}^{(m_i)}$, each of shape $\lambda\times\lambda$. Then

$$\mathsf{Fold}_i(\mathbf{U}_1,\dots,\mathbf{U}_{M_i};\mathbf{C}_1,\dots,\mathbf{C}_{i+1}) = \sum_{j\in[1,m_i]}\mathbf{W}_j\mathbf{C}_{i+1}^{(j)}.$$

By induction, each $\mathbf{W}_j$ is a bounded linear combination of the matrices in $\bar{\mathbf{U}}_j$. Multiplying by $\mathbf{C}_{i+1}^{(j)}$ gives coefficient matrices whose ℓ_∞ norm grows by at most a factor λ, since challenge entries have coefficient norm at most 1. Hence the new coefficients $\mathbf{D}_t$ satisfy $\|\mathbf{D}_t\|_\infty \le \lambda^i$, proving the first part.

For the second part, using the induction on each subtuple,

$$\mathbf{BW}_j = \mathsf{Fold}_{i-1}(\mathbf{B}\bar{\mathbf{U}}_j;\mathbf{C}_1,\dots,\mathbf{C}_i).$$

Therefore multiplying Fold_i by $\mathbf{B}$ on the left is the same as replacing every input $\mathbf{U}_t$ by $\mathbf{BU}_t$ before the recursive fold.

For the third part, since each $\mathbf{U}_t$ splits into r vertical blocks, then by induction, each $\mathbf{W}_j$ splits into the corresponding folded block. Horizontal concatenation followed by multiplication by $\mathbf{C}_{i+1}$ preserves this vertical block decomposition, so the a^{th} vertical output block is

$$\mathsf{Fold}_i(\mathbf{U}_{1,a},\dots,\mathbf{U}_{M_i,a};\mathbf{C}_1,\dots,\mathbf{C}_{i+1}).$$

This proves the stacking identity and completes the induction. □

To describe the prover's messages that are folded, we write $\mathbf{S}$ as $\mathbf{S}_{i_1,\dots,i_{d-1}}$ for $(i_1,\dots,i_{d-1}) \in [1,m_1]\times\dots\times[1,m_{d-1}]$, as a leaf of $m_d \times m_{d+1}$ elements. In the levelled commitment perspective, the intermediate commitments are viewed as internal nodes of the tree, and for $k \in [1, d-2]$, define intermediate vertices recursively by

$$\mathbf{S}_{i_1,\dots,i_k} = \left(\mathbf{I}_{m_{k+1}} \otimes \mathbf{A}_{k+2}\right) \begin{bmatrix} \mathbf{S}_{i_1,\dots,i_k,1} \\ \vdots \\ \mathbf{S}_{i_1,\dots,i_k,m_{k+1}} \end{bmatrix} \mod q_{k+2},$$

of shape $nm_{k+1} \times m_{d+1}$ elements. Finally,

$$\mathbf{S}_\varnothing = (\mathbf{I}_{m_1} \otimes \mathbf{A}_2) \begin{bmatrix} \mathbf{S}_1 \\ \vdots \\ \mathbf{S}_{m_1} \end{bmatrix} \mod q_2,$$

with shape $nm_1 \times m_{d+1}$ elements, so $\mathbf{A}_1\mathbf{S}_\varnothing \equiv \mathbf{T} \mod q_1$. For the i^{th} response, the prover uses the tuple

$$\mathbf{V}_i = \begin{cases} \mathbf{S}_\varnothing, & i = 1, \\ (\mathbf{S}_{\boldsymbol{\alpha}})_{\boldsymbol{\alpha}\in[1,m_1]\times\dots\times[1,m_{i-1}]}, & i > 1, \end{cases}$$

where $\mathbf{V}_i$ is ordered with the first coordinate varying fastest. This is the protocol order used by Fold_{i-1}; the recursive definition of each intermediate vertex still stacks its children under a fixed prefix.

Amortized Levelled Proof. We now show the 0-1 compression invariant $\mathbf{ASC} \equiv \mathbf{TC} \mod q$ from [BBC+18] in the levelled setting of [BLNS20]. The point of lemma 4.9.1 is twofold here: the vertical decomposition lets us treat a folded intermediate commitment as the stack of folded child commitments, while the left multiplication property lets us pull out the next SIS matrix $\mathbf{A}_{i+1}$. Then BT_i only rearranges this vertical stack into the horizontal sibling blocks expected by the next challenge. With this abstraction, we show by induction over the levels of the commitment tree.

- At the root, $\mathbf{A}_1\mathbf{S}_\varnothing \equiv \mathbf{T} \mod q_1$, and therefore we have $\mathbf{A}_1\mathbf{Z}_1 \equiv \mathbf{TC}_1 \mod q_1$.
- For the inductive step, write $\mathbf{V}_{i+1} = (\mathbf{V}_{i,1},\dots,\mathbf{V}_{i,m_i})$, where $\mathbf{V}_{i,j}$ is the tuple of the j^{th} children of the vertices in $\mathbf{V}_i$. By the definition of intermediate commitments, each vertex in $\mathbf{V}_i$ is the vertical stack of the $\mathbf{A}_{i+1}$-commitments to its children. Using the vertical decomposition and left multiplication properties of lemma 4.9.1, we have

$$\mathbf{Z}_i \equiv \begin{bmatrix} \mathbf{A}_{i+1}\mathsf{Fold}_{i-1}(\mathbf{V}_{i,1};\mathbf{C}_1,\dots,\mathbf{C}_i) \\ \vdots \\ \mathbf{A}_{i+1}\mathsf{Fold}_{i-1}(\mathbf{V}_{i,m_i};\mathbf{C}_1,\dots,\mathbf{C}_i) \end{bmatrix} \mod q_{i+1}.$$

  Applying BT_i and the next challenge gives

$$\mathsf{BT}_i(\mathbf{Z}_i)\mathbf{C}_{i+1} \equiv \mathbf{A}_{i+1}\mathsf{Fold}_i(\mathbf{V}_{i+1};\mathbf{C}_1,\dots,\mathbf{C}_{i+1}) = \mathbf{A}_{i+1}\mathbf{Z}_{i+1} \mod q_{i+1}.$$

The block transpose only exposes the m_i sibling blocks of $\mathbf{Z}_i$ in the order needed for the next challenge matrix.

We now show an amortized opening protocol for the levelled commitment, which is public-coin, so all consistency and norm checks can be deferred to the end of the verification.

Commit. The prover computes the levelled commitment $\mathbf{T} \leftarrow F_{1,d}(\mathbf{S})$ and sends $\mathbf{T}$.

i^{th} Challenge. The verifier samples the random challenge matrix $\mathbf{C}_i$ with the shape specified above.

i^{th} Response. After receiving $\mathbf{C}_i$, the prover sends $\mathbf{Z}_i \leftarrow \mathsf{Fold}_{i-1}(\mathbf{V}_i;\mathbf{C}_1,\dots,\mathbf{C}_i)$, of shape $nm_i \times \lambda$ elements.

Verify. After receiving all responses $\mathbf{Z}_1,\dots,\mathbf{Z}_d$, the verifier checks the response are short, and the consistency by

$$\mathbf{A}_1\mathbf{Z}_1 \equiv \mathbf{TC}_1 \mod q_1,$$
$$\mathbf{A}_{i+1}\mathbf{Z}_{i+1} \equiv \mathsf{BT}_i(\mathbf{Z}_i)\mathbf{C}_{i+1} \mod q_{i+1},$$

where $i \in [1, d-1]$.

Knowledge soundness and relaxed extraction. The opening argument does not extract an exact opening to the original commitment map, and the source of this relaxation is already illustrated in the one-level vanilla construction of [BBC+18]: special soundness subtracts two accepting responses, so the target is not scaled in the plain SIS setting, but the extracted preimage has a larger norm. If each accepted response has norm bounded by B, then the extracted difference is naturally bounded by $2B$.

In the extractor of [BBC+18], the extracted opening is represented by bounded integers. However, in a levelled commitment, the extracted preimage is not just an opening witness; it is also meant to be an internal value produced by a lower-level commitment. The same internal value is seen in two ways: the upper-layer consumes an integer representative, as SIS is strong when the integer bound is still small, while the lower-layer only sees the corresponding residue class modulo the lower modulus. The gap between these two views is a multiple of the lower modulus, and these multiples are exactly the carry matrices in $\widetilde{F}$.

For two levels, this definition specializes to a relaxed opening $(\bar{\mathbf{S}}, \mathbf{R})$ satisfying

$$\mathbf{A}_1 \cdot \left((\mathbf{I}_{m_1} \otimes \mathbf{A}_2)\bar{\mathbf{S}} \mod q_2 + q_2\mathbf{R}\right) \equiv \mathbf{T} \mod q_1.$$

The extra $q_2\mathbf{R}$ term records the missing multiples of q_2 when the lower-level computation is viewed as an integer input to the top commitment. This relaxation is still meaningful when parameters force the extracted $(\bar{\mathbf{S}}, \mathbf{R})$ to stay within SIS-binding bounds.

Although the commitment is tree shaped, extraction is better viewed depth-wise. All vertices at a same level are folded into one response, and the relaxed opening records the consistency between neighboring layers.

We now define the relaxed version of the levelled commitment map, in parallel with the exact map $F_{i,j}$:

$$\widetilde{F}_{i,j}(\mathbf{S}; \mathbf{R}_i, \ldots, \mathbf{R}_{j-1}) = \begin{cases} \mathbf{A}_i\mathbf{S} \mod q_i, & i = j, \\ \widetilde{F}_{i,j-1}\left(\left(\mathbf{I}_{M_{i,j-1}} \otimes \mathbf{A}_j\right) \cdot \mathbf{S} \mod q_j + q_j\mathbf{R}_{j-1}; \mathbf{R}_i, \ldots, \mathbf{R}_{j-2}\right), & i < j. \end{cases}$$

A bounded relaxed opening to $\mathbf{T}$ is a tuple $(\bar{\mathbf{S}}, \mathbf{R}_1, \ldots, \mathbf{R}_{d-1})$ such that every column of $\bar{\mathbf{S}}$ has norm at most B_S, every column of $\mathbf{R}_i$ has ℓ_∞ norm at most B_R, and

$$\widetilde{F}_{1,d}(\bar{\mathbf{S}}; \mathbf{R}_1, \ldots, \mathbf{R}_{d-1}) \equiv \mathbf{T} \mod q_1.$$

An exact opening to $\mathbf{T}$ is the case where all carry matrices $\mathbf{R}$ are zeroed out.

Similar to the theorem 4.8.7 in [BBC+18], we define the knowledge soundness for [BLNS20].

Theorem 4.9.2. *If a prover convinces a verifier with probability $\varepsilon > 2^{-\lambda+1} \cdot (4dN)^{2d}$, there is an extractor outputs a bounded relaxed opening, where $B_S = 2^d B$ and $B_{R_k} = 2^k(M_{k-1}m_{d+1}\lambda^{k-1} + 2)$ for $k \in [1, d-1]$. It runs in expected $\mathsf{poly}(\lambda, 1/\varepsilon)$ time.*

Noticing

$$\mathbf{A}_{i+1}\mathbf{Z}_{i+1} \equiv \mathsf{BT}_i(\mathbf{Z}_i)\mathbf{C}_{i+1} \mod q_{i+1}$$

is exactly compressing $m_i \times \lambda$ commitments to λ ones as in [BBC+18], we can use the same row-local rewinding idea as theorem 4.8.7.

Fix the α^{th} column of $\mathbf{T}$. The desired column-wise relaxed opening has the form

$$\widetilde{F}_{1,d}(\bar{\mathbf{s}}_\alpha; \mathbf{r}_1, \ldots, \mathbf{r}_{d-1}) \equiv \mathbf{t}_\alpha \mod q_1,$$

where $\mathbf{r}_i$ is the α^{th} column of $\mathbf{R}_i$ for $i \in [1, d-1]$. Iterating α over $[1, m_{d+1}]$ extracts the full set of relaxed opening.

For now, ignore the runtime of sampling challenges for accepting transcripts, and suppose the extractor obtains two accepting transcripts whose first challenges $\mathbf{C}_1$ and $\mathbf{C}'_1$ differ only in the α^{th} row. The verification equation gives

$$\mathbf{A}_1\mathbf{Z}_1 \equiv \mathbf{T}\mathbf{C}_1 \mod q_1,$$
$$\mathbf{A}_1\mathbf{Z}'_1 \equiv \mathbf{T}\mathbf{C}'_1 \mod q_1.$$

Choose a coordinate r where the two α^{th} challenge rows differ, and orient the subtraction such that $(\mathbf{C}_1 - \mathbf{C}'_1)_{\alpha,r} = 1$. Taking the r^{th} column after subtracting the two accepting equations gives

$$\mathbf{A}_1(\mathbf{z}_{1,r} - \mathbf{z}'_{1,r}) \equiv \mathbf{t}_\alpha \mod q_1.$$

Thus the extractor obtains a bounded preimage of $\mathbf{t}_\alpha$ under $\mathbf{A}_1$.

In a levelled commitment, this vector is only the upper-layer input; the tree construction is then used to recursively explain this extracted column by lower-level preimages.

For the next transition, write the row blocks of $\mathbf{Z}_1$ as $\mathbf{Z}_{1,1}, \ldots, \mathbf{Z}_{1,m_1}$, so that

$$\mathsf{BT}_1(\mathbf{Z}_1) = [\mathbf{Z}_{1,1} \mid \cdots \mid \mathbf{Z}_{1,m_1}].$$

The r^{th} column of $\mathbf{Z}_1$ is then spread across the columns $\{r_j = r + (j-1)\lambda\}_{j\in[1,m_1]}$ of $\mathsf{BT}_1(\mathbf{Z}_1)$, and the verification is

$$\mathbf{A}_2\mathbf{Z}_2 \equiv \mathsf{BT}_1(\mathbf{Z}_1)\mathbf{C}_2 \mod q_2.$$

To extract the j^{th} block of the r^{th} column of $\mathbf{Z}_1$, the extractor therefore rewinds the r_j^{th} row of $\mathbf{C}_2$. If two accepting second challenges $\mathbf{C}_2$ and $\mathbf{C}'_2$ differ only in this row, choose a coordinate s and orient such that $(\mathbf{C}_2 - \mathbf{C}'_2)_{r_j,s} = 1$. Taking the s^{th} column after subtracting the two level-two equations gives

$$\mathbf{A}_2(\mathbf{z}_{2,s} - \mathbf{z}'_{2,s}) \equiv \mathsf{BT}_1(\mathbf{Z}_1)_{r_j} \mod q_2.$$

The right hand side is the j^{th} block of the extracted r^{th} column of $\mathbf{Z}_1$, which is the r_j^{th} column of $\mathsf{BT}_1(\mathbf{Z}_1)$. Iterating over $j \in [1, m_1]$ explains all blocks of that column; these j's become the child slots in the transcript tree.

The heavy-row argument lemma 4.8.8 from [Dam10] is still the runtime core of the extractor.

Lemma 4.9.3 (Scaled Heavy Row Lemma)**.** *Let $\mathbf{H} \in \{0,1\}^{n\times m}$ have density at least ε. Fix $\delta \le \varepsilon$, and call a row δ-heavy if its row-density is at least δ. If a 1-entry of $\mathbf{H}$ is sampled uniformly, then its row is δ-heavy with probability at least $1 - \delta/\varepsilon$. Moreover, conditioned on a δ-heavy row, resampling a column gives another 1-entry with probability at least δ. If the resampled column is required to differ from the original one, this probability is at least $\delta - 1/m$.*

Proof. There are at most δmn 1-entries in light rows, so at least $(\varepsilon - \delta)mn$ 1-entries in heavy rows, and the probability lower bound $1 - \delta/\varepsilon$ holds. The rest follows the same argument in lemma 4.8.8. □

The 0-1 matrices are used in the runtime analysis, serving the same challenge-indexing purpose as in [BBC+18], but now used in the extraction tree. Fix a partial transcript prefix $\boldsymbol{\pi}_i = (\mathbf{C}_1, \mathbf{Z}_1, \ldots, \mathbf{C}_i, \mathbf{Z}_i)$. First define

$$\mathbf{H}_i^{(\boldsymbol{\pi}_i)}\left[\chi, \mathbf{C}'_{i+1}\right]\left[\mathbf{C}'_{i+2}, \ldots, \mathbf{C}'_d\right] \in \{0,1\}.$$

The row is indexed by the prover randomness χ and the current challenge $\mathbf{C}'_{i+1}$. The column is indexed by the future challenge choice $\mathbf{C}'_{i+2}, \ldots, \mathbf{C}'_d$. The entry is 1 iff these choices give a full accepting transcript extending $\boldsymbol{\pi}_i$, such that

$$\mathbf{H}_{i-1}^{\boldsymbol{\pi}_{i-1}}[\chi, \mathbf{C}_i]\left[\mathbf{C}'_{i+1}, \ldots, \mathbf{C}'_d\right] = \mathbf{H}_i^{(\boldsymbol{\pi}_i)}\left[\chi, \mathbf{C}'_{i+1}\right]\left[\mathbf{C}'_{i+2}, \ldots, \mathbf{C}'_d\right].$$

To apply the heavy-row argument to one row of the next challenge, reindex this same 0-1 matrix. Let ρ be a row index in $\mathbf{C}'_{i+1}$. Write $\mathbf{c}'^{\mathsf{T}}_{i+1,\rho}$ for the ρ^{th} row of $\mathbf{C}'_{i+1}$, and $\hat{\mathbf{C}}'_{i+1,\rho}$ for the rows obtained from $\mathbf{C}'_{i+1}$ by deleting $\mathbf{c}'^{\mathsf{T}}_{i+1,\rho}$. Define

$$H_i'^{(\boldsymbol{\pi}_i,\rho)}\left[\chi, \hat{\mathbf{C}}'_{i+1,\rho}\right]\left[\mathbf{c}'^{\mathsf{T}}_{i+1,\rho}, \mathbf{C}'_{i+2}, \ldots, \mathbf{C}'_d\right] = H_i^{(\boldsymbol{\pi}_i)}\left[\chi, \mathbf{C}'_{i+1}\right]\left[\mathbf{C}'_{i+2}, \ldots, \mathbf{C}'_d\right].$$

Thus the rows of $H_i'^{(\boldsymbol{\pi}_i,\rho)}$ fix the prover randomness and the non-interesting rows of $\mathbf{C}'_{i+1}$, while its columns are indexed by the ρ^{th} row of $\mathbf{C}'_{i+1}$ together with all later challenges. At the root, $\boldsymbol{\pi}_0 = \emptyset$ and the first rewinding uses the α^{th} row of $\mathbf{C}'_1$. Beyond the root, BT determines how the current extraction target is seen at the next level: it becomes several target rows in $\mathbf{C}'_{i+1}$ spaced by λ, as was previously discussed.

The original one-level use sets $\delta = \varepsilon/2$, and the probability is at least $1/2$ in lemma 4.8.8. For the tree extractor, the same lemma is used with a larger slack K: set the heavy threshold to $\delta = \varepsilon/K$, and an accepting entry lands in a heavy row with probability at least $1 - 1/K$. After landing in such a row, resampling only the target row(s) in the challenge being rewound gives another accepting challenge with probability at least $\varepsilon/K - |C|^{-\lambda}$.

The extractor in [BLNS20] builds a tree of partial transcripts, and the leaves are accepting full transcripts. A state at depth i is written as

$$\mathsf{TC}_i([j_1, \ldots, j_i], \mathbf{C}_i, \mathbf{Z}_i, t).$$

Here $[j_1, \ldots, j_i]$ is the address of the vertex in the tree, where each j_k records which child slot was taken:

- The block of the current column (because of BT) being followed,
- One of the two rewounded children, left or right.

The arguments $\mathbf{C}_i, \mathbf{Z}_i, t$ are the current challenge, the current response, and the column currently being extracted. Reading the challenges and responses along the addressed path gives the prefix $\boldsymbol{\pi}_i = (\mathbf{C}_1, \mathbf{Z}_1, \ldots, \mathbf{C}_i, \mathbf{Z}_i)$ used when rewinding the prover.

The first step is the same operation with empty prefix and target row α in $\mathbf{C}_1$ (namely the α^{th} column of $\mathbf{T}$). Sample two first challenges that differ only in this row, run the prover on both, and choose a coordinate t where the two α^{th} rows differ, oriented so the difference is 1. The two resulting children are depth-1 states storing their own $\mathbf{C}_1, \mathbf{Z}_1$, and both carry the index t. Their edge already gives

$$\mathbf{A}_1(\mathbf{z}_{1,t}^{(0)} - \mathbf{z}_{1,t}^{(1)}) \equiv \mathbf{t}_\alpha \mod q_1.$$

Fix a non-root, non-leaf state $\mathsf{TC}_i([j_1, \ldots, j_i], \mathbf{C}_i, \mathbf{Z}_i, t)$. The first task is to determine which rows of the next challenge have to be rewound. The t^{th} column of $\mathbf{Z}_i$ is spread by BT_i across the rows

$$\rho_\ell = t + (\ell - 1)\lambda, \qquad \ell \in [1, m_i],$$

of the next challenge $\mathbf{C}_{i+1}$. For each row ρ_ℓ, the algorithm creates a pair of child candidates by sampling two challenges $\widetilde{\mathbf{C}}_{i+1}$ and $\widetilde{\mathbf{C}}'_{i+1}$ that are equivalent except for the ρ_ℓ^{th} row, and rewinding the prover from the prefix $\boldsymbol{\pi}_i$ to obtain the two responses $\widetilde{\mathbf{Z}}_{i+1}$ and $\widetilde{\mathbf{Z}}'_{i+1}$.

For $i < d - 1$, these candidates are the next partial-transcript children. For $i = d - 1$, they must be accepting leaf transcripts instead.

- For the first child, the extractor keeps resampling the ρ_ℓ^{th} row of $\widetilde{\mathbf{C}}_{i+1}$ and reruns the prover until the resulting transcript is accepting; it aborts after $\lambda K/\varepsilon$ failed attempts.
- For the second child, it rewinds to the same prefix and repeats the same search for $\widetilde{\mathbf{C}}'_{i+1}$; it aborts after $2\lambda K/\varepsilon$ failed attempts.

Choose a coordinate $s_\ell \in [1, \lambda]$ where the two ρ_ℓ^{th} challenge rows differ,

$$\left(\widetilde{\mathbf{C}}_{i+1} - \widetilde{\mathbf{C}}'_{i+1}\right)_{\rho_\ell, s_\ell} \neq 0,$$

and orient the pair so the difference is 1. The two responses become the $(\ell, 0)$ and $(\ell, 1)$ children, and both use s_ℓ as the next column to be extracted. The edge created by this local rewind is

$$\mathbf{A}_{i+1}\left(\widetilde{\mathbf{z}}_{i+1,s_\ell} - \widetilde{\mathbf{z}}'_{i+1,s_\ell}\right) \equiv \mathsf{BT}_i(\mathbf{Z}_i)_{\rho_\ell} \mod q_{i+1}.$$

Thus a depth i state with $i \geq 1$ creates $2m_i$ children, and level $i > 0$ has $2^i M_{i-1}$ vertices, as in the paper.

Now we can bound the abort probability in the same order as the paper. Let κ denote the per-step heavy-row fraction parameter, and let the retry parameter in the tree construction be $K = \kappa^{2d}$. Fix a level-$(d-1)$ vertex V, and write

$$(V_0, V_1, \ldots, V_{d-2}, V)$$

for the path from the root V_0 to V. Let $r_1, \ldots, r_\gamma$ be the indices for which V_{r_j} is a right child. Let $\mathsf{abort}(V)$ be the event that the tree construction aborts while trying to complete this fixed leaf vertex V. Write $h_i(\eta)$ for the event that the row selected by this path is η-heavy in $\mathbf{H}_i$, and write $h'_i(\eta)$ for the corresponding event in the reindexed matrix $\mathbf{H}'_i$. The aborting probability for left and right children is derived differently: for a left child it comes directly from $\mathbf{H}_i$, while for a right child it first comes through the reindexed matrix $\mathbf{H}'_i$ and then through $\mathbf{H}_i$.

The first step is the only boundary case. If V_1 is a right child, then the first challenge was obtained by resampling one target row while keeping the other rows fixed, so the proof first goes through $\mathbf{H}'_0$:

$$\Pr[\mathsf{abort}(V)] \le \Pr\left[\mathsf{abort}(V) \mid h'_0(\varepsilon/\kappa)\right] + \Pr\left[\neg h'_0(\varepsilon/\kappa)\right] \le \Pr\left[\mathsf{abort}(V) \mid h'_0(\varepsilon/\kappa)\right] + 1/\kappa.$$

Here $\mathbf{H}'_0$ moves the target row of $\mathbf{C}'_1$ from the row index into the column index: a row of $\mathbf{H}'_0$ is fixed by the non-target rows of $\mathbf{C}'_1$, and is the concatenation of the ordinary $\mathbf{H}_0$ rows obtained by varying that target row. Thus the heavy-row lemma applies to $\mathbf{H}'_0$ after this row-column rearrangement. Conditioned on $h'_0(\varepsilon/\kappa)$, the selected row of $\mathbf{H}'_0$ has density at least ε/κ. Now split this selected row according to the value of the target row. Each piece is one ordinary row of $\mathbf{H}_0$, so another application of the heavy-row lemma gives

$$\Pr\left[\neg h_0(\varepsilon/\kappa^2) \mid h'_0(\varepsilon/\kappa)\right] \le 1/\kappa,$$

thus

$$\Pr\left[\mathsf{abort}(V) \mid h'_0(\varepsilon/\kappa)\right] \le \Pr\left[\mathsf{abort}(V) \mid h'_0(\varepsilon/\kappa) \wedge h_0(\varepsilon/\kappa^2)\right] + 1/\kappa.$$

If V_1 is a left child, then the first challenge is sampled as a full challenge, and the bound starts directly from $\mathbf{H}_0$:

$$\Pr[\mathsf{abort}(V)] \le \Pr[\mathsf{abort}(V) \mid h_0(\varepsilon/\kappa)] + \Pr[\neg h_0(\varepsilon/\kappa)] \le \Pr[\mathsf{abort}(V) \mid h_0(\varepsilon/\kappa)] + 1/\kappa.$$

After this boundary step, the proof continues in the same way through $h_1, \ldots, h_{d-2}$, inserting h'_{r_j-1} immediately before h_{r_j-1} for every right child V_{r_j} on the path. Let E be the conjunction of these heavy-row events, such that the threshold divided by a factor of κ at each peeling step. Then

$$\Pr[\neg E] \le (d-1+\gamma)/\kappa \le 2(d-1)/\kappa.$$

On E, the event $h_{d-2}(\varepsilon/\kappa^{d-1+\gamma})$ holds. Concretely, for the prefix $\boldsymbol{\pi}_{d-2}$ on the path to V, the row $\mathbf{H}_{d-2}^{(\boldsymbol{\pi}_{d-2})}[\chi, \mathbf{C}_{d-1}]$ has density at least $\varepsilon/\kappa^{d-1+\gamma}$. The columns of this row are indexed by the remaining final challenge $\mathbf{C}_d$, so this is the success probability for one attempt in the first final-challenge search. Since $\gamma \le d-1$ and $K = \kappa^{2d}$, this local success probability is lower-bounded by $\varepsilon/\kappa^{d-1+\gamma} \ge \varepsilon/K$. With $\lambda K/\varepsilon$ attempts, the first leaf abort probability less than $e^{-\lambda}$.

The paired leaf search needs one more heavy-row step. Write ρ for the target row of the final challenge determined by this leaf. After the first final challenge is fixed, the extractor resamples only this row for the second final challenge, so the analysis passes to the row

$$\mathbf{H}'^{(\boldsymbol{\pi}_{d-1},\rho)}_{d-1}[\chi, \hat{\mathbf{C}}'_{d,\rho}]$$

whose columns are indexed by the resampled target row. Conditioned on $h'_{d-1}(\varepsilon/\kappa^{d+\gamma})$, this row has density at least $\varepsilon/\kappa^{d+\gamma}$. This event fails with probability at most $1/\kappa$, and, when it holds, the successful resampling probability is at least

$$\varepsilon/\kappa^{d+\gamma} - |C|^{-\lambda},$$

where the subtraction excludes choosing the same target row again. By the lower bound on ε in theorem 4.9.2, this is still at least $\varepsilon/(2K)$, and the $2\lambda K/\varepsilon$ attempts make the second leaf abort probability less than $e^{-\lambda}$.

Combining the failure of E, the extra possible failure of $h'_{d-1}(\varepsilon/\kappa^{d+\gamma})$, and the two leaf searches, for a fixed leaf vertex we get

$$\Pr[\mathsf{abort}(V)] \le 2e^{-\lambda} + 2d/\kappa.$$

The paper sets $\kappa = 4dN$. Union bounding over all vertices at level $d-1$, which are $2^{d-1}M_{d-2}$ of them, given that d is a constant, and $2^{d-1}M_{d-2} < N$, then the aborting probability is at most

$$\frac{2N}{e^{\lambda}} + \frac{2dN}{\kappa} = \frac{2N}{e^{\lambda}} + \frac{1}{2}.$$

Thus one run of TreeConstruct succeeds with constant probability, up to the negligible $2N/e^{\lambda}$ term. For a fixed target column α, running $O(\lambda)$ independent copies gives a successful tree except with negligible probability. The extractor needs this for all m_{d+1} columns of $\mathbf{T}$, so it repeats the same procedure for every $\alpha \in [1, m_{d+1}]$. Since $m_{d+1} \le N$, $N = \mathsf{poly}(\lambda)$, d is constant, and $K = (4dN)^{2d} = \mathsf{poly}(\lambda)$, the total expected running time remains $\mathsf{poly}(\lambda, 1/\varepsilon)$.

Once the tree is built, extraction is just reading off neighboring children. For a vertex whose current target is the t^{th} column, the two children created for the row ρ_ℓ give

$$\mathbf{A}_{i+1}\left(\widetilde{\mathbf{z}}_{i+1,s_\ell} - \widetilde{\mathbf{z}}'_{i+1,s_\ell}\right) \equiv \mathsf{BT}_i(\mathbf{Z}_i)_{\rho_\ell} \mod q_{i+1}.$$

Collecting these sibling differences over $\ell \in [1, m_i]$ gives the preimage block for the parent target column after the block transition. Repeating down the tree and concatenating the extracted blocks level by level produces the relaxed opening $(\bar{\mathbf{s}}, \mathbf{r}_1, \ldots, \mathbf{r}_{d-1})$ for the chosen column of $\mathbf{T}$; Iterating over $[1, m_{d+1}]$ columns gives the full relaxed opening.

The binding intuition is then about the relaxed relation extracted from the argument. For two levels, write the lifted inputs seen by the top matrix $\mathbf{A}_1$ as

$$\bar{\mathbf{S}}_\varnothing = (\mathbf{I}_{m_1} \otimes \mathbf{A}_2)\bar{\mathbf{S}} \mod q_2 + q_2\mathbf{R},$$
$$\bar{\mathbf{S}}'_\varnothing = (\mathbf{I}_{m_1} \otimes \mathbf{A}_2)\bar{\mathbf{S}}' \mod q_2 + q_2\mathbf{R}'.$$

Both relaxed openings to the same $\mathbf{T}$ satisfy

$$\mathbf{A}_1\bar{\mathbf{S}}_\varnothing \equiv \mathbf{A}_1\bar{\mathbf{S}}'_\varnothing \equiv \mathbf{T} \mod q_1.$$

Suppose the two bounded relaxed openings $(\bar{\mathbf{S}}, \mathbf{R})$ and $(\bar{\mathbf{S}}', \mathbf{R}')$ are distinct, and both open to the same $\mathbf{T}$. If $\mathbf{R} \ne \mathbf{R}'$, then the two integer representations of the upper-layer inputs are different, and hence

$$\mathbf{A}_1(\bar{\mathbf{S}}_\varnothing - \bar{\mathbf{S}}'_\varnothing) \equiv \mathbf{0} \mod q_1$$

gives a short SIS relation for $\mathbf{A}_1$. If $\mathbf{R} = \mathbf{R}'$, then $\bar{\mathbf{S}} \ne \bar{\mathbf{S}}'$. In this case, either the lower-layer images of $\mathbf{I}_{m_1} \otimes \mathbf{A}_2$ differ modulo q_2, which again makes the lifted top inputs different and gives SIS for $\mathbf{A}_1$, or their equality modulo q_2 gives a short nonzero SIS solution for $\mathbf{A}_2$. The separated moduli and norm bounds are chosen so that the differences stay inside the SIS-binding range.

Parameters and slack. For constant d, the parameter balancing in the paper gives communication

$$O\left(N^{1/(d+1)} \cdot \left(d^3\lambda \log^2 N + d\lambda^2\right)\right).$$

If $d + 1 = \log N$, this becomes

$$O(\lambda \log^5 N + \lambda^2 \log N).$$

The extractor recovers a relaxed opening: a matrix $\bar{\mathbf{S}}$ and matrices $\mathbf{R}_1, \ldots, \mathbf{R}_{d-1}$ satisfying the relaxed recursive commitment equation. The communication saving therefore comes with knowledge slack, with parameters set so that the slack remains below the relevant SIS bounds.

Bulletproofs-style lattice folding. The second construction asks whether the Bulletproofs [BBB+18] folding trick can be made lattice-compatible. The starting statement is the RSIS equation

$$\mathbf{As} = \mathbf{t}, \qquad \mathbf{A} = [\mathbf{A}_1 \mid \mathbf{A}_2] \in R^{1\times k}, \qquad \mathbf{s} = \begin{bmatrix}\mathbf{s}_1\\ \mathbf{s}_2\end{bmatrix} \in R^k,$$

over $R = \mathbb{Z}[X]/(X^n + 1)$. Here the number of integer secrets is $N = kn$. In one folding round, the prover sends

$$\mathbf{l} = \mathbf{A}_1\mathbf{s}_2, \qquad \mathbf{r} = \mathbf{A}_2\mathbf{s}_1.$$

The verifier samples a monomial challenge $c \xleftarrow{R} \{X^i : i \in \mathbb{Z}_{2n}\}$, and the prover responds with

$$\mathbf{z} = \mathbf{s}_1 + c\mathbf{s}_2.$$

Then

$$(c\mathbf{A}_1 + \mathbf{A}_2)\mathbf{z} = c^2\mathbf{l} + c\mathbf{t} + \mathbf{r},$$

so the verifier can replace the old statement $(\mathbf{A}, \mathbf{t})$ by

$$\mathbf{B} = c\mathbf{A}_1 + \mathbf{A}_2, \qquad \mathbf{t}' = c^2\mathbf{l} + c\mathbf{t} + \mathbf{r},$$

with witness $\mathbf{z}$ of half the length. The verifier also checks that $\mathbf{z}$ remains short.

Repeating for d rounds reduces the witness length by 2^d; setting $d = \log k$ leaves a single ring element and gives logarithmic-length communication.

For one round, extraction rewinds to three accepting transcripts with distinct challenges c_1, c_2, c_3 gives

$$(c_j\mathbf{A}_1 + \mathbf{A}_2)\mathbf{z}_j = c_j^2\mathbf{l} + c_j\mathbf{t} + \mathbf{r}, \qquad j \in [1, 3],$$

for the same $\mathbf{l}, \mathbf{r}, \mathbf{t}$. Choose coefficients $\lambda_1, \lambda_2, \lambda_3 \in R$ by solving the Vandermonde system

$$\begin{bmatrix} c_1^2 & c_2^2 & c_3^2 \\ c_1 & c_2 & c_3 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix}\lambda_1\\ \lambda_2\\ \lambda_3\end{bmatrix} = \begin{bmatrix}0\\ 1\\ 0\end{bmatrix}.$$

Then the $\mathbf{l}$ and $\mathbf{r}$ terms cancel, and formally

$$\mathbf{A}_1 \sum_{j=1}^{3} \lambda_j c_j \mathbf{z}_j + \mathbf{A}_2 \sum_{j=1}^{3} \lambda_j \mathbf{z}_j = \mathbf{A} \sum_{j=1}^{3} \lambda_j \begin{bmatrix} c_j\mathbf{z}_j \\ \mathbf{z}_j \end{bmatrix} = \mathbf{t}.$$

The challenge set is chosen to consist of monomials X^i. For pairwise distinct monomials, the paper uses the fact that $2/(X^i - X^j)$ has coefficients in $\{-1, 0, 1\}$, and obtains $\|8\lambda_j\|_\infty \le 2n^2$. After clearing this denominator, one obtains a one-round relaxed preimage

$$\bar{\mathbf{z}} = 8\sum_{j=1}^{3} \lambda_j \begin{bmatrix} c_j\mathbf{z}_j \\ \mathbf{z}_j \end{bmatrix}, \qquad \mathbf{A}\bar{\mathbf{z}} = 8\mathbf{t},$$

with $\|\bar{\mathbf{z}}\|_\infty \le O(12n^3 p)$ when each accepted response is bounded by $2p$.

This is the sense in which the Bulletproofs-style extraction is approximate: the extractor does not recover a short preimage of the original target $\mathbf{t}$, but a larger preimage of the scaled target $8\mathbf{t}$. Recursing through d folds multiplies both effects. Extraction gives a relaxed solution to

$$\mathbf{A}\bar{\mathbf{z}} = 8^d\mathbf{t}$$

with

$$\|\bar{\mathbf{z}}\|_\infty = O(n^{3d}12^d p)$$

when $\|\mathbf{s}\|_\infty \le p$. Thus logarithmic folding buys a much smaller transcript, but pays a large slack: the target is scaled by 8^d and the extracted witness norm grows by roughly $(12n^3)^d$. With repetition to reduce soundness error, the proof size is

$$O\left(\frac{\lambda N \log(2^d p)}{2^d \log n} + \lambda d^2 n\right).$$

Takeaway. Both constructions are non-PCP routes to succinct, plausibly quantum-safe ZK, but neither extracts an exact original witness. Levelled commitments have larger proofs but smaller, more controllable slack. Bulletproofs folding has the better proof-size expression when slack is cheap, but the relaxed equation $\mathbf{A}\bar{\mathbf{z}} = 8^d\mathbf{t}$ and coefficient growth can force much larger parameters in applications where the extracted witness feeds into other cryptographic components. The paper's main open direction is to keep the algebraic folding/compression while reducing coefficient growth enough to become competitive with generic PCP/Merkle-tree based post-quantum arguments in concrete size.

4.10 Short Invertible Polynomial Ring Elements

This note combines two related papers. The first work [LS18] explains how the factorization pattern of cyclotomic rings modulo q affects the availability of short invertible elements, which is for the challenge spaces in Fiat-Shamir with Abort sigma protocol. The second work [BLS19] uses the fully splitting case to turn a coordinate-wise shortness statement into an algebraic identity via the NTT.

Cyclotomic Polynomial Ring Splitting. Let $R = \mathbb{Z}[X]/(\Phi_\ell(X))$ and $R_q = \mathbb{Z}_q[X]/(\Phi_\ell(X))$. The factorization of Φ_ℓ modulo q controls how much CRT structure is available in R_q. If Φ_ℓ splits into a few irreducible factors, one obtains a partially splitting ring. This is useful for fast multiplication and for constructing large challenge sets whose nonzero differences remain invertible [LS18].

For an integer $m > 1$, write

$$\varphi(m) = m \cdot \prod_{\substack{p \text{ prime} \\ p \mid m}} \frac{p-1}{p}, \qquad \delta(m) = \prod_{\substack{p \text{ prime} \\ p \mid m}} p,$$

such that $m = \varphi(m)/\varphi(\delta(m)) \times \delta(m)$, and

$$\tau(m) = \begin{cases} m, & m \text{ is odd}, \\ m/2, & m \text{ is even}. \end{cases}$$

For $a \in \mathbb{Z}_m^*$,

$$\mathsf{ord}_m(a) = \min\{k > 0 : a^k \equiv 1 \mod m\}.$$

The paper also writes $s_1(m)$ for the largest singular value of the Vandermonde matrix associated with the primitive m-th roots of unity. We only need φ and ord for the splitting theorem below; s_1 appears in the later bounds showing that small elements are invertible.

The cyclotomic polynomial Φ_m is fully splitting modulo q if it factors into linear terms over $\mathbb{Z}_q$:

$$\Phi_m(X) \equiv \prod_{j \in [1, \varphi(m)]} (X - r_j) \mod q,$$

where the $r_j \in \mathbb{Z}_q^*$ are exactly the primitive m^{th} roots of unity modulo q, such that $\mathsf{ord}_q(r_j) = \varphi(m)$. Since $\mathbb{Z}_q^*$ is cyclic of order $q-1$, this happens exactly when $\mathbb{Z}_q$ contains primitive m^{th} roots of unity, namely

$$m \mid q - 1 \qquad \Leftrightarrow \qquad q \equiv 1 \mod m.$$

In this case the CRT decomposes

$$\mathbb{Z}_q[X]/(\Phi_m(X)) \cong \prod_{j \in [1, \varphi(m)]} \mathbb{Z}_q[X]/(X - r_j) \cong \mathbb{Z}_q^{\varphi(m)}.$$

The following theorem describes the characteristics of cyclotomic polynomial ring splitting, and more specifically, in the partial splitting case.

Theorem 4.10.1 (Cyclotomic Partial Splitting [LS18])**.** *Let*

$$m = \prod_i p_i^{e_i}, \qquad z = \prod_i p_i^{f_i},$$

where $e_i \geq 1$ and $1 \leq f_i \leq e_i$. If q is prime, $q \equiv 1 \mod z$, and $\mathsf{ord}_m(q) = m/z$, then for r_j the primitive z^{th} roots of unity,

$$\Phi_m(X) \equiv \prod_{j \in [1, \varphi(z)]} (X^{m/z} - r_j) \mod q$$

and each factor $X^{m/z} - r_j$ is irreducible over $\mathbb{Z}_q$.

The condition $q \equiv 1 \mod z$ means, there is a size z multiplicative subgroup in $\mathbb{Z}_q^*$, such that r_j are z^{th} primitive roots of unity, while $\mathsf{ord}_m(q) = m/z$ means, there is no elements in $\mathbb{Z}_q^*$ that r_j can split to in the directions defined by the prime factors of m/z.

It is better to understand $\mathsf{ord}_m(q)$ in the Frobenius endomorphism aspect of view.

Definition 4.10.2 (Frobenius endomorphism and orbit)**.** Let $\overline{\mathbb{Z}_q}$ be an ambient field containing $\mathbb{Z}_q$ and all the roots of $\Phi_m(X)$. We use it only so that we can discuss roots of Φ_m even when they do not lie in $\mathbb{Z}_q$. The polynomials are still defined over $\mathbb{Z}_q$, meaning their coefficients lie in $\mathbb{Z}_q$; we only evaluate them over $\overline{\mathbb{Z}}_q$. The q-Frobenius map is

$$\mathsf{Endo}_q(\alpha) = \alpha^q.$$

It fixes $\mathbb{Z}_q$ pointwise, since $a^q = a$ for every $a \in \mathbb{Z}_q$. For $\alpha \in \overline{\mathbb{Z}_q}$, its Frobenius orbit is defined as $\alpha, \alpha^q, \alpha^{q^2}, \ldots$ and the orbit length is the smallest $t > 0$ such that $\alpha^{q^t} = \alpha$.

If α is a primitive m^{th} root of unity, then the condition $\alpha^{q^t} = \alpha$ for orbit to end is equivalent to $q^t \equiv 1 \mod m$ over the exponent. Hence the Frobenius orbit length of a primitive m^{th} root is $\mathsf{ord}_m(q)$.

We then construct the minimal polynomial from a Frobenius orbit, which is irreducible over $\mathbb{Z}_q$.

Lemma 4.10.3 (Minimal polynomial from Frobenius orbit)**.** *Let $\alpha \in \overline{\mathbb{Z}_q}$ have Frobenius orbit length t, then*

$$\mu_\alpha(X) = \prod_{i \in [0,t-1]} \left(X - \alpha^{q^i}\right)$$

is the minimal polynomial for α over $\mathbb{Z}_q$. In particular, $\mu_\alpha(X)$ is irreducible over $\mathbb{Z}_q$ and has coefficients in $\mathbb{Z}_q$, and $\deg(\mu_\alpha) = t$.

Proof. This product lies in $\mathbb{Z}_q[X]$ as the q-Frobenius map cyclically permutes the roots in the product, which means

$$\prod_{i \in [0,t-1]} \left(X - \left(\alpha^{q^i}\right)^q\right) = \prod_{i \in [1,t]} \left(X - \alpha^{q^i}\right) = \mu_\alpha(X),$$

and all the binomial terms are over Z_q satisfying $b = b^q$, then each coefficient $c^q = c$, therefore each $c \in \mathbb{Z}_q$.

Let $g(X) \in \mathbb{Z}_q[X]$ satisfy $g(\alpha) = 0$. By prior discussion, $g(\alpha^q) = 0$ as each coefficient over $\mathbb{Z}_q$ has $c = c^q$. Hence, $g(X)$ vanishes over the Frobenius orbit, so $\mu_\alpha | g$, and μ_g is the minimal polynomial for α over $\mathbb{Z}_q$.

A minimal polynomials over a field are irreducible, or otherwise contradicting the minimality of the degree. □

If ζ is a m^{th} root of unity, $\zeta^{m/z}$ is a z^{th} root of unity. The primitiveness is preserved, say ζ is a primitive m^{th} root of unity, then $\zeta^{m/z}$ is a primitive z^{th} root of unity.

With $\mathsf{ord}_m(q) = m/z$, for a primitive m^{th} root of unity in $\overline{\mathbb{Z}}_q$, the Frobenius orbit has length m/z. By $q \equiv 1 \mod z$, we know the z^{th} roots of unity are over $\mathbb{Z}_q$. Hence, for each primitive m^{th} root of unity ζ, we have

$$\left(\zeta^{q^i}\right)^{m/z} = \left(\zeta^{m/z}\right)^{q^i} = r^{q^i} = r,$$

where $\zeta^{m/z} = r$, and r is a primitive z^{th} root of unity. Therefore, the m/z long Frobenius orbit vanishes over $X^{m/z} - r_j$ that is irreducible over $\mathbb{Z}_q$ for all $j \in [1, \varphi(z)]$.

The choice of z controls the splitting pattern: the ring has $\varphi(z)$ CRT slots, each tied to an extension field of degree m/z over $\mathbb{Z}_q$. The fully splitting case is when $z = m$, where $\mathsf{ord}_m(q) = 1$, namely $q \equiv 1 \mod m$, then

$$\Phi_m(X) \equiv \prod_{j \in [1,\varphi(m)]} (X - r_j) \mod q.$$

By CRT decomposition, evaluation at these roots gives the NTT isomorphism. Let $n = \varphi(m)$, then

$$R_q \cong \mathbb{Z}_q^n, \qquad s \mapsto \widehat{s} = (s(r_1), \ldots, s(r_n)),$$

and multiplication in R_q becomes coordinate-wise multiplication after applying the NTT. In the power-of-two case $R = \mathbb{Z}[X]/(X^n + 1) = \mathbb{Z}[X]/(\Phi_{2n}(X))$, this condition becomes $q \equiv 1 \mod 2n$, so $X^n + 1$ has all of its roots in $\mathbb{Z}_q$.

Splitting and Short Invertability. Let $R_m = \mathbb{Z}[X]/(\Phi_m(X))$ and let $\omega_1, \ldots, \omega_{\varphi(m)}$ be the primitive m^{th} roots of unity defined over complex field. The short-invertibility question in [LS18] is whether a nonzero short element of $R_{m,q}$ can vanish in one of the CRT coordinates induced by the above splitting. Suppose

$$\Phi_m(X) \equiv \prod_j f_j(X) \mod q,$$

where the f_j are irreducible over $\mathbb{Z}_q$, then

$$R_{m,q} \cong \prod_j \mathbb{Z}_q[X]/(f_j(X)),$$

and $y \in R_{m,q}$ is invertible when $y \not\equiv 0 \mod f_j(X)$ for every irreducible f_j. Hence, non-invertibility means that y lands in the kernel of at least one CRT projection, and [LS18] views such kernels as ideal lattices under the canonical embedding. For $\mathbf{y} = \sum_{i\in[0,\varphi(m)-1]} y_i X^i \in R_m$, the canonical embedding is the complex evaluation vector

$$\sigma(\mathbf{y}) = \left(\mathbf{y}(\omega_1), \ldots, \mathbf{y}(\omega_{\varphi(m)})\right) \in \mathbb{C}^{\varphi(m)}, \tag{4.1}$$

and the embedding norm is

$$\|\mathbf{y}\|_e = \|\sigma(\mathbf{y})\|_2 . \tag{4.2}$$

Let V_m be the associated Vandermonde matrix

$$\mathbf{V}_m = \begin{pmatrix} 1 & \omega_1 & \cdots & \omega_1^{\varphi(m)-1} \\ \vdots & \vdots & & \vdots \\ 1 & \omega_{\varphi(m)} & \cdots & \omega_{\varphi(m)}^{\varphi(m)-1} \end{pmatrix}.$$

The parameter $s_1(m)$ defined as

$$s_1(m) = \max_{\mathbf{u}\in\mathbb{C}^{\varphi(m)}\setminus\{0\}} \frac{\|V_m\mathbf{u}\|_2}{\|\mathbf{u}\|_2}$$

is the largest singular value of V_m, which is used to compare coefficient size with embedding size:

$$\|\mathbf{y}\|_e = \|\mathbf{V}_m\mathbf{y}\|_2 \le s_1(m)\,\|\mathbf{y}\|_2 , \tag{4.3}$$

where $\mathbf{y} = (y_0, \ldots, y_{\varphi(m)-1})$ is the coefficient vector of $\mathbf{y}$.

We now construct the main result of [LS18] from the lower bound of embedding norm of an ideal lattice.

Lemma 4.10.4 (Embedding lower bound for ideal lattices [PR07, LS18])**.** *Let Λ be an ideal lattice in R_m, for every nonzero $\mathbf{y} \in \Lambda$ satisfies*

$$\|\mathbf{y}\|_e \ge \sqrt{\varphi(m)} \cdot \det(\Lambda)^{1/\varphi(m)}.$$

Applying eq. (4.3), we get

$$\|\mathbf{y}\|_2 \ge \lambda_1(\Lambda) \ge \frac{\sqrt{\varphi(m)}}{s_1(m)} \cdot \det(\Lambda)^{1/\varphi(m)}.$$

Consider $\Lambda = \{\mathbf{x} \in R_m : \mathbf{x} \equiv 0 \mod (f_i, q)\}$, and the next result is immediate by lemma 4.10.4. First, Λ itself is an additive subgroup of $\mathbb{Z}^{\varphi(m)}$. Second, for any $a \in \Lambda$ and $b \in R_m$, we have $ab \in \Lambda$. Hence, Λ is an ideal lattice.

Lemma 4.10.5 (Polynomial ring ℓ_2 lower bound [LS18])**.** *Suppose $\Phi_m(X)$ splits modulo q into d distinct irreducible factors of degree $\varphi(m)/d$. If $\mathbf{y} \in R_m$ is nonzero and $\mathbf{y} \equiv 0 \mod (f_j(X), q)$ for one irreducible factor f_j, then*

$$\|\mathbf{y}\|_2 \ge \frac{\sqrt{\varphi(m)}}{s_1(m)} q^{1/d}.$$

Since $\det(\Lambda) = \left|\mathbb{Z}^{\varphi(m)}/\Lambda\right| = q^{\varphi(m)/d}$, then lemma 4.10.5 is immediate.

If a nonzero $\mathbf{y}$ vanishes in one CRT coordinate, then $\mathbf{y}$ lies in Λ, and lemma 4.10.5 forces $\|\mathbf{y}\|_2$ to be large. Thus any nonzero $\mathbf{y}$ whose coefficient norm is below this lower bound cannot vanish in any CRT coordinate.

Theorem 4.10.6 (Short invertibility in partially splitting rings [LS18])**.** *Under the partial-splitting setup*

$$\Phi_m(X) \equiv \prod_{j\in[1,\varphi(z)]} (X^{m/z} - r_j) \mod q,$$

any nonzero $\mathbf{y} \in R_{m,q}$ *satisfying either*

$$\|\mathbf{y}\|_\infty < \frac{1}{s_1(z)} q^{1/\varphi(z)}$$

or

$$\|\mathbf{y}\|_2 < \frac{\sqrt{\varphi(m)}}{s_1(m)} q^{1/\varphi(z)}$$

is invertible in $R_{m,q}$.

Informally, short nonzero elements do not wrap around to zero in any NTT coordinate.

4.11 Lattice-Based Exact Argument of Knowledge

From [LS18] to [BLS19]. The role of splitting is different in [LS18] and [BLS19]. In [LS18], the NTT coordinates are used to argue that short nonzero elements cannot vanish in any coordinate, and hence are invertible. In [BLS19], we instead use the fully splitting case to expose the witness coordinates themselves: after the NTT, the ternary condition on $\widehat{\mathbf{s}}$ is equivalent to the single ring identity $\mathbf{s}(\mathbf{s}-\mathbf{1})(\mathbf{s}-\mathbf{2}) = \mathbf{0}$. The remaining issue is no longer ring splitting, but how to enforce this identity together with $\mathbf{A}\widehat{\mathbf{s}} = \overline{u}$ inside an exact proof of knowledge. In our case, $R_q = \mathbb{Z}_q[X]/(\Phi_\ell)$ is a power-of-2 fully splitting cyclotomic ring by $q \equiv 1 \mod \ell$, $\Phi_\ell = X^n + 1$, and $n = \varphi(\ell)$, so $\ell = 2n$.

The NTT Trick for Shortness. The starting exact relation is to prove knowledge of a short vector $\widehat{\mathbf{s}} \in \mathbb{Z}_q^n$ such that

$$\mathbf{A}\widehat{\mathbf{s}} = \overline{u} \mod q.$$

We focus on $\widehat{\mathbf{s}} \in \{0,1,2\}^n$. For $\mathbf{s} \in R_q$ whose NTT representation is $\widehat{\mathbf{s}}$, the ternary constraint is equivalent to

$$\widehat{\mathbf{s}} \in \{0,1,2\}^n \quad \Leftrightarrow \quad \widehat{\mathbf{s}} \circ (\widehat{\mathbf{s}} - \widehat{\mathbf{1}}) \circ (\widehat{\mathbf{s}} - \widehat{\mathbf{2}}) = \widehat{\mathbf{0}} \quad \Leftrightarrow \quad \mathbf{s}(\mathbf{s}-\mathbf{1})(\mathbf{s}-\mathbf{2}) = \mathbf{0} \in R_q, \tag{4.4}$$

and the "shortness" check is turned into an algebraic identity over a fully splitting ring. This is how exact shortness is proved without falling back to a coordinate-wise combinatorial argument.

First Attempt From Prior Approximate Arguments of Knowledge. We start from the Commit-and-Open Sigma-Protocol, which originated from [Lyu12], and was explicitly constructed in [BDL+18] for the approximate argument of knowledge for the commitment opening.

Since we cannot challenge $\mathbf{A}\hat{\mathbf{s}}$ with short invertible ring elements, as $\hat{\mathbf{s}} \in \mathbb{Z}_q^n$ is incompatible with the challenge applied through the polynomial product. Going down the path, for the knowledge soundness of the Sigma-Protocol, the challenge has to be from $\mathbb{Z}_q$ for a sufficiently large challenge set. Yet, given that the new challenge set is $\mathbb{Z}_q$, the norm of $c \cdot \hat{\mathbf{s}}$ can be arbitrary, and the masking message $\overline{r} \in \mathbb{Z}_q^n$ has to be uniformly random. Again, this violates the commitment binding in the Commit-and-Open Sigma-Protocol, where $\mathbf{A}$ as the Ajtai commitment has to bind $\overline{r}$ to pin down the masking against the random oracle.

Looking closely, in previous approximate arguments of knowledge [BBC+18, BDL+18], $\mathbf{A}$ simultaneously takes over the role of the statement, and the binding commitment that binds the masking message and the last opening. The dual roles enforce that the masking and opening have to be short, or otherwise Ajtai will not bind.

In [BLS19], the role of binding commitment is lifted from $\mathbf{A}$, and moved to a lattice-based commitment scheme that binds arbitrary messages, such that the canonical representations $\mathbf{s}$ and $\mathbf{r} = \mathsf{NTT}^{-1}(\overline{r})$ are committed and pinned to the random oracle. In this way, $\mathbf{A}$ is treated as more of a linear system of the relation to be proved, rather than a commitment scheme that instantiates a Commit-and-Open Sigma-Protocol. The lifting of the binding role allows the challenge c to be uniformly random over $\mathbb{Z}_q$, and the masking $\overline{r}$ for $c \cdot \hat{\mathbf{s}}$ to be uniformly random over $\mathbb{Z}_q^n$.

Why [BDL+18] Commitment. Ajtai commitment [Ajt96] has been great as a compressing commitment, when the message is short. The binding and extractability hold by the SIS hardness assumption. However in our scenario, the NTT representation is short for the message to be committed, but not the canonical form of the message itself. Hence, we need a commitment that lets us manipulate arbitrary ring messages algebraically and homomorphically, while preserving hiding, binding, and extractability.

Now consider a commitment satisfying the requirements above, though it is not compressing like Ajtai commitment. The vanilla BDLOP commitment [BDL+18] is constructed as follows. For $\beta > 0$, write

$$S_\beta = \{\mathbf{a} \in R : \|\mathbf{a}\|_\infty \le \beta\}. \tag{4.5}$$

The public parameter is a pair of matrices:

$$\mathbf{A}_1 = \begin{bmatrix} \mathbf{I}_n & \mathbf{A}_1' \end{bmatrix} \in R_q^{n \times k}, \qquad \mathbf{A}_1' \xleftarrow{\text{R}} R_q^{n \times (k-n)},$$
$$\mathbf{A}_2 = \begin{bmatrix} \mathbf{0}^{\ell_{\text{Com}} \times n} & \mathbf{I}_{\ell_{\text{Com}}} & \mathbf{A}_2' \end{bmatrix} \in R_q^{\ell_{\text{Com}} \times k}, \quad \mathbf{A}_2' \xleftarrow{\text{R}} R_q^{\ell_{\text{Com}} \times (k-n-\ell_{\text{Com}})}.$$

Equivalently, the commitment map is defined by the vertical stack

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \in R_q^{(n+\ell_{\mathsf{Com}})\times k}.$$

On message $\overline{\mathbf{m}} \in R_q^{\ell_{\mathsf{Com}}}$, sample short randomness $\bar{\mathbf{r}} \xleftarrow{\mathrm{R}} S_\beta^k$ and output

$$\bar{\mathbf{c}} = \begin{bmatrix} \bar{\mathbf{c}}_1 \\ \bar{\mathbf{c}}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \bar{\mathbf{r}} + \begin{bmatrix} \mathbf{0}^n \\ \overline{\mathbf{m}} \end{bmatrix}.$$

Thus the message is added only to the lower coordinates, while both coordinates are still linear functions of the same short randomness $\bar{\mathbf{r}}$. The relaxed opening form used in the proof system allows an extracted scalar $\mathbf{f}$ from the challenge set:

$$\mathbf{f} \cdot \bar{\mathbf{c}} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \end{bmatrix} \bar{\mathbf{r}} + \mathbf{f} \cdot \begin{bmatrix} \mathbf{0}^n \\ \overline{\mathbf{m}} \end{bmatrix}, \tag{4.6}$$

where the honest opening has $\mathbf{f} = \mathbf{1}$ and $\bar{\mathbf{r}}$ remains short. This identity-block form is what makes opening proofs compact: the prover can prove knowledge of a short $\bar{\mathbf{r}}$ for the top part and recover the message from the lower part.

Definition 4.11.1 ($\mathsf{SKS}^2_{n,k,\gamma}$)**.** Given $\mathbf{A}' \xleftarrow{\mathrm{R}} R_q^{n\times(k-n)}$, find a nonzero vector $\overline{\mathbf{y}} = (\mathbf{y}_1, \ldots, \mathbf{y}_k)^\mathsf{T} \in R^k$ such that

$$\begin{bmatrix} \mathbf{I}_n & \mathbf{A}' \end{bmatrix} \overline{\mathbf{y}} = \mathbf{0} \in R_q^n, \qquad \|\mathbf{y}_i\|_2 \le \gamma.$$

This is exactly the Module-SIS hardness assumption where $\mathbf{A}$ is in Hermite Normal Form.

Definition 4.11.2 ($\mathsf{DKS}^\infty_{n,k,\beta}$)**.** Given $\mathbf{A}' \xleftarrow{\mathrm{R}} R_q^{n\times(k-n)}$, distinguish

$$\left(\mathbf{A}', \begin{bmatrix} \mathbf{I}_n & \mathbf{A}' \end{bmatrix} \overline{\mathbf{y}}\right), \qquad \overline{\mathbf{y}} \xleftarrow{\mathrm{R}} S_\beta^k,$$

from $(\mathbf{A}', \overline{\mathbf{u}})$ for uniform $\overline{\mathbf{u}} \xleftarrow{\mathrm{R}} R_q^n$. This is equivalent to the Module-LWE hardness assumption, when the number of samples is limited, namely for a limited number of MLWE samples sharing the same secret.

The hiding reduction is to $\mathsf{DKS}^\infty_{n+\ell_{\mathsf{Com}},k,\beta}$: if $\mathbf{A}\bar{\mathbf{r}}$ is computationally (or statistically) indistinguishable from uniform, then adding $(\mathbf{0}, \overline{\mathbf{m}})$ does not reveal $\overline{\mathbf{m}}$. The binding reduction is to $\mathsf{SKS}^2_{n,k,\gamma}$ for the top block $\mathbf{A}_1$: two different short relaxed openings give a nonzero vector of the form $\mathbf{f}'\bar{\mathbf{r}} - \mathbf{f}\bar{\mathbf{r}}'$ with $\mathbf{A}_1(\mathbf{f}'\bar{\mathbf{r}} - \mathbf{f}\bar{\mathbf{r}}') = \mathbf{0}$.

BDLOP Specialized in [BLS19]. Now back to the [BLS19]. The challenge set C is defined by

$$C = \{\mathbf{0}\} \cup \{X^i\}_{i\in[0,\ell-1]} = \{\mathbf{0}\} \cup \{\pm X^i\}_{i\in[0,n-1]}, \tag{4.7}$$

such that $|C| = \ell + 1 = 2n + 1$. The challenge-difference set is $\overline{C} := (C - C) \setminus \{\mathbf{0}\}$, and every element in $\overline{C}$ is invertible: Since R_q is fully splitting, then

$$R_q \cong \prod_{j\in\mathbb{Z}_\ell^*} \mathbb{Z}_q[X]/\left(X - \xi^j\right),$$

where ξ is the ℓ^{th} primitive root of unity. Each $X^\alpha - X^\beta \in \overline{C}$ has $X^\alpha - X^\beta \equiv \xi^{\alpha j} - \xi^{\beta j} \mod \left(X - \xi^j\right)$ for all $j \in \mathbb{Z}_\ell^*$, and $\xi^{\alpha j} - \xi^{\beta j}$ is zeroed out iff $\alpha \equiv \beta \mod \ell$. Hence, all of the CRT slots are nonzero, and they are invertible. Differences involving $\mathbf{0}$ are of form $\pm X^\alpha$, which are units in R_q and are invertible.

Moreover, for every $\bar{\mathbf{f}} = \mathbf{f}_1 - \mathbf{f}_2 \in \overline{C}$ and $\mathbf{a} \in R$, multiplication by $\bar{\mathbf{f}}$ increases the coefficient ℓ_2-norm by at most a factor of 2:

$$\left\|\bar{\mathbf{f}}\mathbf{a}\right\|_2 \le \|\mathbf{f}_1\mathbf{a}\|_2 + \|\mathbf{f}_2\mathbf{a}\|_2 \le 2\|\mathbf{a}\|_2. \tag{4.8}$$

One thing to note: the ℓ_2-norm in [BLS19] is the canonical embedding ℓ_2-norm, while the ℓ_2-norm is the coefficient ℓ_2-norm. The canonical embedding ℓ_2-norm is the ℓ_2-norm for the eq. (4.1), which was shown in [LS18]. In our R power-of-two cyclotomic ring, we have

$$\begin{aligned}\|\mathbf{a}\|_e^2 &= \sum_{j\in\mathbb{Z}_\ell^*} \left|\mathbf{a}\left(\omega^j\right)\right|^2 = \sum_{j\in\mathbb{Z}_\ell^*} \mathbf{a}\left(\omega^j\right)\overline{\mathbf{a}\left(\omega^j\right)} = \sum_{j\in\mathbb{Z}_\ell^*} \mathbf{a}\left(\omega^j\right)\mathbf{a}\left(\omega^{-j}\right)\\ &= n\|\mathbf{a}\|_2^2 + \sum_{i\neq k\in[0,n-1]} a_i a_k \sum_{j\in\mathbb{Z}_\ell^*} \omega^{j(i-k)} = n\|\mathbf{a}\|_2^2,\end{aligned}$$

where ω here is the primitive ℓ^{th} complex root of unity, namely $\omega = \exp(2\pi i/\ell)$, as was written in eq. (4.2).

In the [BLS19] specialization for [BDL+18], RLWE and RSIS are needed for hiding and binding.

Definition 4.11.3 ($\mathsf{RSIS}_{k,B}$)**.** An efficient algorithm $\mathcal{A}$ has advantage ε in solving $\mathsf{RSIS}_{k,B}$ if

$$\Pr\left[\|\bar{\mathbf{s}}\|_2 \le B \ \wedge\ \left[\mathbf{1}, \bar{\mathbf{a}}^\mathsf{T}\right]\bar{\mathbf{s}} = \mathbf{0} \ \wedge\ \bar{\mathbf{s}} \neq \bar{\mathbf{0}} \,\middle|\, \bar{\mathbf{a}} \xleftarrow{\text{R}} R_q^k,\ \bar{\mathbf{s}} \leftarrow \mathcal{A}(\bar{\mathbf{a}})\right] \ge \varepsilon,$$

where $\bar{\mathbf{s}} \in R^{k+1}$ and the zero-equation is evaluated in R_q, and $\|\bar{\mathbf{s}}\|_2$ is the coefficient ℓ_2-norm.

Fix a distribution χ over $\mathbb{Z}$. We write χ^n for the distribution over R, where a sample is obtained by sampling n coefficients independently from χ, and χ^{kn} denotes the corresponding distribution over R^k. In the instantiation of [BLS19], the distribution χ is over $\{-1, 0, 1\}$, outputting ± 1 with probability $5/16$ each, and 0 with probability $3/8$.

Definition 4.11.4 (RLWE_m)**.** Let $m \ge 1$. An efficient algorithm $\mathcal{A}$ has advantage ε in solving RLWE_m if

$$\left|\Pr\left[b = 1 \,\middle|\, \begin{array}{l}\bar{\mathbf{a}} \xleftarrow{\text{R}} R_q^m,\\ \mathbf{s} \leftarrow \chi^n,\\ \bar{\mathbf{e}} \leftarrow \chi^{mn},\\ b \leftarrow \mathcal{A}(\bar{\mathbf{a}}, \bar{\mathbf{a}}\mathbf{s} + \bar{\mathbf{e}})\end{array}\right] - \Pr\left[b = 1 \,\middle|\, \begin{array}{l}\bar{\mathbf{a}} \xleftarrow{\text{R}} R_q^m,\\ \bar{\mathbf{u}} \xleftarrow{\text{R}} R_q^m,\\ b \leftarrow \mathcal{A}(\bar{\mathbf{a}}, \bar{\mathbf{u}})\end{array}\right]\right| \ge \varepsilon.$$

The RLWE_m problem is hard if every efficient algorithm has only negligible advantage.

The BDLOP commitment is specialized to commit to $\overline{\mathbf{m}} = (\mathbf{m}_2, \mathbf{m}_3, \mathbf{m}_4, \mathbf{m}_5)^\mathsf{T} \in R_q^4$, and the public parameter is a matrix $\mathbf{B} \in R_q^{5\times 6}$ of the form

$$\mathbf{B} = \begin{bmatrix}\mathbf{1} & \mathbf{b}_{1,2} & \mathbf{b}_{1,3} & \mathbf{b}_{1,4} & \mathbf{b}_{1,5} & \mathbf{b}_{1,6}\\ \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{b}_{2,6}\\ \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{0} & \mathbf{b}_{3,6}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{0} & \mathbf{b}_{4,6}\\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{1} & \mathbf{b}_{5,6}\end{bmatrix}, \qquad \mathbf{b}_{i,j} \xleftarrow{\text{R}} R_q.$$

This is the same identity-block idea as vanilla BDLOP, but with one top coordinate and four message coordinates. To commit to $\overline{\mathbf{m}}$, sample short randomness $\bar{\mathbf{r}} \leftarrow \chi^{6n}$ from the error distribution and output

$$\bar{\mathbf{t}} = \mathbf{B}\bar{\mathbf{r}} + \begin{bmatrix}\mathbf{0}\\ \overline{\mathbf{m}}\end{bmatrix} \in R_q^5.$$

For $\sigma > 0$, let $\rho_\sigma(\overline{\mathbf{u}}) := \exp\left(-\|\overline{\mathbf{u}}\|_2^2/(2\sigma^2)\right)$, and $\rho_\sigma(S) := \sum_{\overline{\mathbf{x}}\in S}\rho_\sigma(\overline{\mathbf{x}})$. The discrete Gaussian centered at $\overline{\mathbf{v}} \in R^k$ is

$$D_{\overline{\mathbf{v}},\sigma}^{kn}(\overline{\mathbf{z}}) = \rho_\sigma(\overline{\mathbf{z}} - \overline{\mathbf{v}})\,\rho_\sigma^{-1}\left(R^k\right). \tag{4.9}$$

When $\overline{\mathbf{v}} = \overline{\mathbf{0}}$, we write D_σ^{kn}.

The relaxed opening relation from extractor is

$$\mathbf{f} \cdot \bar{\mathbf{t}} = \mathbf{B}\bar{\mathbf{r}} + \mathbf{f} \cdot \begin{bmatrix} \mathbf{0} \\ \overline{\mathbf{m}} \end{bmatrix},$$

where $\mathbf{f}$ is either from the challenge-difference set $\overline{C}$, or $\mathbf{f} = \mathbf{1}$ for an honest opening. $\bar{\mathbf{r}}$ needs to be short, written as $\|\bar{\mathbf{r}}\|_2 \leq 2B$, with $B = \sigma\sqrt{12n}$.

In [BLS19] the hiding is reduced to the RLWE_5 hardness, and the binding is reduced to the $\mathsf{RSIS}_{5,8B}$ hardness, where $8B$ derives from the ℓ_2-norm of $\mathbf{f}_1\bar{\mathbf{r}}_0 - \mathbf{f}_0\bar{\mathbf{r}}_1$ in eq. (4.10), recall eq. (4.8) for factor 2 and $\|\bar{\mathbf{r}}\|_2 \leq 2B$.

Short Invertible Elements Alone Won't Cut It. A short invertible element itself cannot give knowledge with relaxed norm bounds, not even the exact knowledge. Its role is more of, so long as the elements in the challenge-difference $\overline{C}$ are invertible, we can reduce binding to the hardness assumptions like SIS.

After the first definition of binding by definition 4.8.1, we define the property again as follows.

Definition 4.11.5 (Binding)**.** A commitment scheme is ε-binding if for all efficient $\mathcal{A}$,

$$\Pr\left[\mathbf{m}_0 \neq \mathbf{m}_1 \wedge \mathsf{Open}_{\mathsf{pp}}(\mathbf{m}_0, \mathbf{c}, \mathbf{r}_0) = 1 \wedge \mathsf{Open}_{\mathsf{pp}}(\mathbf{m}_1, \mathbf{c}, \mathbf{r}_1) = 1 \;\middle|\; \begin{array}{r} \mathsf{pp} \leftarrow \mathsf{Setup}(1^\lambda), \\ (\mathbf{m}_0, \mathbf{r}_0, \mathbf{m}_1, \mathbf{r}_1, \mathbf{c}) \leftarrow \mathcal{A}(1^\lambda, \mathsf{pp}) \end{array}\right] < \varepsilon.$$

For [BDL+18], the $(\mathbf{f}, \overline{\mathbf{m}}, \bar{\mathbf{r}})$ is a relaxed opening proof. Supposing we have a pair of relaxed openings satisfying

$$\mathbf{f}_0 \cdot \bar{\mathbf{t}} = \mathbf{B}\bar{\mathbf{r}}_0 + \mathbf{f}_0 \cdot \begin{bmatrix} \mathbf{0} \\ \overline{\mathbf{m}}_0 \end{bmatrix},$$
$$\mathbf{f}_1 \cdot \bar{\mathbf{t}} = \mathbf{B}\bar{\mathbf{r}}_1 + \mathbf{f}_1 \cdot \begin{bmatrix} \mathbf{0} \\ \overline{\mathbf{m}}_1 \end{bmatrix}$$

with $\mathbf{m}_0 \neq \mathbf{m}_1$, then cancelling out $\bar{\mathbf{t}}$ term, we have

$$\mathbf{B}(\mathbf{f}_1 \cdot \bar{\mathbf{r}}_0 - \mathbf{f}_0 \cdot \bar{\mathbf{r}}_1) + \mathbf{f}_0 \cdot \mathbf{f}_1 \cdot \begin{bmatrix} \mathbf{0} \\ \overline{\mathbf{m}}_0 - \overline{\mathbf{m}}_1 \end{bmatrix} = \bar{\mathbf{0}}, \tag{4.10}$$

separating by the vertical stack

$$\mathbf{A}_1(\mathbf{f}_1 \cdot \bar{\mathbf{r}}_0 - \mathbf{f}_0 \cdot \bar{\mathbf{r}}_1) = \mathbf{0},$$
$$\mathbf{A}_2(\mathbf{f}_1 \cdot \bar{\mathbf{r}}_0 - \mathbf{f}_0 \cdot \bar{\mathbf{r}}_1) = \mathbf{f}_0 \cdot \mathbf{f}_1 \cdot (\overline{\mathbf{m}}_1 - \overline{\mathbf{m}}_0).$$

It is that $\mathbf{f}_0 \cdot \mathbf{f}_1$ being invertible, ensuring $\mathbf{f}_0 \cdot \mathbf{f}_1 \cdot (\overline{\mathbf{m}}_0 - \overline{\mathbf{m}}_1) \neq \bar{\mathbf{0}}$, or otherwise the product with a non-invertible ring element a may be 0. Given that $\mathbf{f}_1 \cdot \bar{\mathbf{r}}_0 - \mathbf{f}_0 \cdot \bar{\mathbf{r}}_1 \neq \bar{\mathbf{0}}$, reducing to the SIS hardness of $\mathbf{A}_1(\mathbf{f}_1 \cdot \bar{\mathbf{r}}_0 - \mathbf{f}_0 \cdot \bar{\mathbf{r}}_1)$.

For [Ajt96], recall definition 4.8.4, write $\mathbf{A} = [\mathbf{A}_1 \mid \mathbf{A}_2]$ as the commitment key, $\mathbf{s} = \left[\mathbf{r}^\mathsf{T} \mid \mathbf{m}^\mathsf{T}\right]^\mathsf{T}$ as the message and randomness, and $\mathbf{c} \leftarrow \mathbf{As} = \mathbf{A}_1\mathbf{r} + \mathbf{A}_2\mathbf{m}$ as the Ajtai commitment. Deriving from the same Sigma-Protocol in [Lyu12], a relaxed opening proof $(\mathbf{f}, \overline{\mathbf{m}}, \bar{\mathbf{r}})$ from extraction must satisfy

$$\mathbf{f} \cdot \bar{\mathbf{c}} = \mathbf{A}_1\bar{\mathbf{r}} + \mathbf{A}_2\overline{\mathbf{m}},$$

where the honest opening has $\mathbf{f} = \mathbf{1}$.

If the same commitment is opened with $\mathbf{s}' = \left[\mathbf{r}'^\mathsf{T} \mid \mathbf{m}'^\mathsf{T}\right]^\mathsf{T}$, then for $\mathbf{s} = \left[\bar{\mathbf{r}}^\mathsf{T} \mid \overline{\mathbf{m}}^\mathsf{T}\right]^\mathsf{T}$ with $\mathbf{s} \cdot \mathbf{f}^{-1} \neq \mathbf{s}'$, we immediately reduce binding to SIS hardness assumption; or otherwise without invertibility the existence of an SIS solution when $\mathbf{f} \cdot \mathbf{s}' - \mathbf{s} = \bar{\mathbf{0}}$ is hard to reason (say an entry of $\mathbf{s}$ is 0 while the same entry in $\mathbf{s}'$ is non-zero but annihilated). Again, the preservation of non-zeroness by invertible elements helps with the hardness reduction.

To sum up, the short invertible elements in the relaxed opening ensure the binding reduction, but do not directly give knowledge soundness.

What if A is Weak. For [BLS19], the relation to prove is $\mathbf{A}\hat{\mathbf{s}} \equiv \overline{u} \mod q$, where each entry in $\hat{\mathbf{s}}$ is within $\{0, 1, 2\}$. In the prior Commit-and-Open Sigma-Protocol attempt based on previous approximate arguments of knowledge, we lifted the role of the binding commitment from $\mathbf{A}$. My ill-contrived question is now: if $\mathbf{A}$ is not a strong Ajtai instance, can we still have zero knowledge for arbitrary $\mathbf{A}$, unlike in previous works such as [BBC+18, BLNS20, BDL+18]?

I started by considering an extreme case where $\mathbf{A}$ is an identity matrix. Then $\overline{u} = \hat{\mathbf{s}}$, and the random masking $\overline{r}$ is sent in the clear. An honest-but-curious verifier can remove $\overline{r}$ from $\overline{z}$, which is the third move of the Sigma-Protocol proposed by modifying the original opening protocol in [BDL+18] (ignoring the BDLOP commitment for now), and derive the witness $\hat{\mathbf{s}}$.

On the other hand, such an accepting transcript is totally simulatable! Still ignoring the BDLOP part,

- an accepting transcript contains the first message $\mathbf{A}\overline{r}$ of the masking through the $\mathbf{A}$ linear system, the second challenge c, which is uniformly random over $\mathbb{Z}_q$, and the third response $c \cdot \hat{\mathbf{s}} + \overline{r}$;
- a simulated transcript first samples the third response $\overline{z}$ uniformly at random and then samples the challenge $c' \xleftarrow{R} \mathbb{Z}_q$; the first masking through $\mathbf{A}$ can be simulated by $\mathbf{A}\overline{z} - c' \cdot \overline{u}$.

Both the second and third messages are uniformly distributed. The first masking $\mathbf{A}\overline{r}$ in a real accepting transcript is uniform over $\mathrm{im}(\mathbf{A})$ (uniformly random over the whole codomain if $\mathbf{A}$ has full row rank), which is determined by the distribution of $\overline{r}$, while in the simulated transcript, the distribution of $\mathbf{A}\overline{z} - c' \cdot \overline{u}$ is determined by c' and $\overline{z}$, which are both uniformly random, and $\overline{u} \in \mathrm{im}(\mathbf{A})$ as HVZK is defined for a valid statement. Then $\mathbf{A}\overline{z} - c' \cdot \overline{u} = \mathbf{A}\left(\overline{z} - c' \cdot \overline{s}'\right)$, where $\mathbf{A}\overline{s}' \equiv \overline{u} \mod q$, so the first masking is also simulatable as part of the joint transcript distribution. Generalizing from $\mathbf{A} = \mathbf{I}_n$, simulatability holds regardless of $\mathbf{A}$.

So what exactly does HVZK really say? Literally, HVZK says that an honestly generated accepting transcript, where the prover holds the witness and the statement is public, is computationally or statistically close to a simulated transcript generated using only the statement. Now let T_{real} be the honestly generated accepting transcript, T_{sim} be the simulated transcript, $\mathcal{A}$ be an efficient algorithm attempting to recover the witness from the transcript, and

$$P_{\mathsf{real}} = \Pr\left[R\left(x, \mathcal{A}(x, T_{\mathsf{real}})\right) = 1\right], \quad P_{\mathsf{sim}} = \Pr\left[R\left(x, \mathcal{A}(x, T_{\mathsf{sim}})\right) = 1\right].$$

$R\left(x, \mathcal{A}\left(x, T\right)\right)$ can be computed efficiently because membership in the NP relation R can be checked efficiently and $\mathcal{A}$ is efficient. If HVZK is computational, then let $R(x, \mathcal{A}(x, \cdot))$ be an efficient distinguisher, hence

$$|P_{\mathsf{real}} - P_{\mathsf{sim}}| \leq \mathsf{negl}(\lambda). \tag{4.11}$$

If HVZK is statistical, then eq. (4.11) follows from the statistical distance being negligible. Let $\mathcal{B}$ be an algorithm that samples T_{sim} from the simulator and runs $\mathcal{A}(x, T_{\mathsf{sim}})$. Since the simulator is efficient, $\mathcal{B}$ is efficient. Then by eq. (4.11),

$$\Pr\left[R(x, \mathcal{B}(x)) = 1\right] - \mathsf{negl}(\lambda) \leq P_{\mathsf{real}} \leq \Pr\left[R(x, \mathcal{B}(x)) = 1\right] + \mathsf{negl}(\lambda),$$

or equivalently

$$P_{\mathsf{real}} - \mathsf{negl}(\lambda) \leq \Pr\left[R(x, \mathcal{B}(x)) = 1\right] \leq P_{\mathsf{real}} + \mathsf{negl}(\lambda).$$

In this way, HVZK can be interpreted as, it does not give one more advantage in finding the witness for the statement, with or without an honestly generated accepting transcript. But ZK is irrelevant to $\mathbf{A}$ being a hard instance or not. If the problem is easy, then the problem is in P, and there is no witness to hide.

Thus, any efficient witness-recovery advantage obtained from the honest verifier's view could already be obtained from the public statement alone. If the relation is hard, no such non-negligible advantage can exist, and the *witness hiding* against an honest verifier follows from HVZK and the hardness of the relation [21]. See also [Dam10].

In our particular setting, if $(\mathbf{A}, \overline{u})$ is sampled from a hard ISIS instance distribution, then finding $\overline{r}$ from $\mathbf{A}\overline{r}$ in the transcript for short $\hat{\mathbf{s}}'$ is at least as hard as solving the ISIS instance; namely, efficient recovery of the masking value $\overline{r}$ would give an efficient solver for the ISIS relation.

Henceforth, we let the statement $(\mathbf{A}, \overline{u})$ be sampled from a hard ISIS instance distribution for witness privacy.

[21] This is an honest-verifier analogue of the witness-hiding interpretation introduced by [FS90]. Their formal definition is distributional and applies to potentially malicious verifiers, whereas the present conclusion follows only for the honest verifier from HVZK.

After the First Sigma-Protocol. Recall the construction of the first Sigma-Protocol attempt without BDLOP:

- The prover sends out $\overline{w} \leftarrow \mathbf{A}\overline{r}$, where $\overline{r}$ is uniformly random.
- The verifier challenges the prover by sending $c \xleftarrow{R} \mathbb{Z}_q$.
- The prover replies with $c \cdot \hat{\mathbf{s}} + \overline{r}$.

It is definitely flawed, a malicious prover who does not know $\mathbf{s}$ can send out the first masking $\overline{w}'$ at random, and on verifier replies with $c \xleftarrow{R} \mathbb{Z}_q$, find an arbitrary image $\overline{z}'$ such that $\mathbf{A}\overline{z}' = \overline{w}' + c \cdot \overline{u}$, where $\mathbf{A}\hat{\mathbf{s}} \equiv \overline{u} \mod q$ in statement.

Starting from the approximate argument similar to [BDL+18, Lyu12]. For HVZK, namely the accepting transcript is simulatable, the first message for the masking randomness should be simulatable in the joint distribution of the transcript, or otherwise supposing the accepting transcript whose first message being raw masking randomness is efficiently simulatable, then such efficient simulator can be used to solve the statement, which is a hard ISIS instance, and it does not make sense. By contradiction, the first message to be sent out should be "hiding".

On the other hand, one could have just sent over the raw masking randomness to the RO, and let the RO decide which challenge to send back. But due to the goal of reaching ZK, one has to use a (cryptographic) object in between the masking randomness and the RO. Intuitively, one should pin the masking randomness against the RO, such that the challenge is decided and fixed after the RO absorbs the first message, and the represented masking randomness.

Since the RO itself cannot guarantee what value the first message represents, we need the "binding" property, which says it is computationally/statistically hard to find different preimages that lead to the same message. With such property, an efficient knowledge extractor can extract (relaxed) knowledge of the witness from a pair of forked transcripts from the first message, or otherwise a malicious prover can provide "branch-dependent" decompositions, say $z_0 = y_0 + c_0 \cdot s_0$ and $z_1 = y_1 + c_1 \cdot s_1$, so rewinding does not give valid witness. Binding removes the possibility to equivocate, and makes it possible for extraction from rewinding.

With the property of binding and hiding, we introduce the notion of commitment.

In the vanilla approximate arguments such as [BDL+18, Lyu12], stopping at the third move is sufficient. It has to be either breaking the SIS hardness assumption, with probability at most $\mathsf{negl}(\lambda)$, or the malicious prover prepares a forged commitment $\overline{t}$ to the masking randomness, and a forged response short $\overline{z}$, such that the challenge c happens to satisfy $\mathbf{A}\overline{z} = \overline{t} + c \cdot \overline{u}$, with probability at most $|C|^{-1}$. That sums up to be the knowledge soundness, and it suffices to stop at the third move as the binding property allows for rewinding for knowledge extraction.

For the current protocol [BLS19], the same argument no longer applies. The masking randomness is not pinned to the RO, even if $\mathbf{A}$ stands for an Ajtai commitment, it will not bind for a message with unbounded length. Moreover, on the third response, in the previous protocols [BDL+18, Lyu12], it suffices to use binding of short messages to stop there, while in [BLS19], the norm is arbitrary, so we will resort to proving the linear relation $\overline{r} + c \cdot \hat{\mathbf{s}} = \hat{\mathbf{z}}$, let alone proving the shortness of $\hat{\mathbf{s}}$.

Quickly recap what we have here, so after the first Sigma-Protocol, we transformed the problem into proving the linear relation $\hat{\mathbf{z}} = c \cdot \hat{\mathbf{s}} + \overline{r}$, where $\hat{\mathbf{z}}$ is in the third move prover response, while $\overline{r}$ is the masking randomness, and $\hat{\mathbf{s}}$ is the witness. Finally, we also need to prove the shortness of $\hat{\mathbf{s}}$ in the statement.

Proving Linear Relation over Elements in [BDL+18]. Deriving from the protocol for linear relation in [BDL+18], we consider an adaptation here in the RSIS and RLWE settings to prove the linear relation of the committed elements.

Let $\mathbf{B} \in R_q^{k \times (k+1)}$ be a BDLOP commitment public parameter, and $\overline{\mathbf{m}} \in R_q^{k-1}$ be the message, then

$$\overline{\mathbf{c}} = \mathbf{B}\overline{\mathbf{r}} + \begin{bmatrix} \mathbf{0} \\ \overline{\mathbf{m}} \end{bmatrix} \in R_q^k$$

is the commitment to $\overline{\mathbf{m}}$, where $\overline{\mathbf{r}} \leftarrow \chi^{(k+1)n}$.

Let $f(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1})$ be a linear function of $\{\mathbf{x}_i\}_{i \in [1,k-1]}$ whose total degree is at most 1, $\overline{\mathbf{b}}_i$ be the i^{th} row of $\mathbf{B}$, and

$\overline{\mathbf{b}}_f = f\left(\overline{\mathbf{b}}_2, \ldots, \overline{\mathbf{b}}_k\right)$. Then Π_{lin} is defined for statement

$$f(\mathbf{m}_1, \ldots, \mathbf{m}_{k-1}) = \mathbf{w}.$$

- The prover samples $\overline{\mathbf{r}}' \leftarrow D_\sigma^{(k+1)n}$, and sends over

$$\overline{\mathbf{w}} = \begin{bmatrix} \overline{\mathbf{b}}_1^\mathsf{T} \\ \overline{\mathbf{b}}_f^\mathsf{T} \end{bmatrix} \overline{\mathbf{r}}'.$$

- The verifier samples $\mathbf{f} \xleftarrow{R} C$, where C is the short challenge set, and sends over $\mathbf{f}$.
- The prover runs reject sampling, and outputs $\overline{\mathbf{z}} \leftarrow \overline{\mathbf{r}}' + \mathbf{f} \cdot \overline{\mathbf{r}}$.
- The verifier checks $\overline{\mathbf{z}}$ being short, and the following equality holds

$$\overline{\mathbf{w}} + \mathbf{f} \cdot \begin{bmatrix} \mathbf{c}_1 \\ f(\mathbf{c}_2, \ldots, \mathbf{c}_k) \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{b}}_1^\mathsf{T} \\ \overline{\mathbf{b}}_f^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}} + \mathbf{f} \cdot \begin{bmatrix} \mathbf{0} \\ \mathbf{w} \end{bmatrix}.$$

Moreover, the linear relation scheme Π_{lin} can be extended to multiple linear functions.

Piece Things Together Without Shortness Check. We now piece Π_{lin} and the first Sigma-Protocol attempt together for the construction of the argument system just to prove the reduced statement $\hat{\mathbf{z}} = \overline{r} + c \cdot \hat{\mathbf{s}}$ without the norm check.

Here we let $\hat{\mathbf{y}} = \mathsf{NTT}(\mathbf{y})$ be the masking randomness, which used to be represented by $\overline{r}$. We use $\mathbf{B} \in R_q^{3\times 4}$ for the BDLOP commitment public parameter, such that we commit to $\mathbf{s}$ and $\mathbf{y}$.

- The prover sends out $\overline{w} \leftarrow \mathbf{A}\hat{\mathbf{y}}$, where $\hat{\mathbf{y}}$ is uniformly random. Moreover, the prover samples $\overline{\mathbf{r}} \leftarrow \chi^{4n}$, commits

$$\overline{\mathbf{c}} \leftarrow \mathbf{B}\overline{\mathbf{r}} + \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \\ \mathbf{s} \end{bmatrix},$$

  and sends $\overline{\mathbf{c}}$ over to the verifier.
- The verifier challenges the prover by sending $c \xleftarrow{R} \mathbb{Z}_q$.
- The prover replies with $\mathbf{z} \leftarrow c \cdot \mathbf{s} + \mathbf{y}$, and starts Π_{lin} by:
  - The prover samples $\overline{\mathbf{y}}' \leftarrow D_\sigma^{4n}$, let $f(\mathbf{y}, \mathbf{s}) = c \cdot \mathbf{s} + \mathbf{y}$, and sends over

$$\overline{\mathbf{w}}' = \begin{bmatrix} \overline{\mathbf{b}}_1^\mathsf{T} \\ \overline{\mathbf{b}}_f^\mathsf{T} \end{bmatrix} \overline{\mathbf{y}}'.$$

  - The verifier samples $\mathbf{f} \xleftarrow{R} C$, where C is the short challenge set, and sends over $\mathbf{f}$.
  - The prover runs reject sampling, and outputs $\overline{\mathbf{z}}' \leftarrow \overline{\mathbf{y}}' + \mathbf{f} \cdot \overline{\mathbf{r}}$.
- The verifier checks that $\left\|\overline{\mathbf{z}}'\right\|_2 \le B$ and that the following equality holds:

$$\overline{\mathbf{w}}' + \mathbf{f} \cdot \begin{bmatrix} \mathbf{c}_1 \\ f(\mathbf{c}_2, \mathbf{c}_3) \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{b}}_1^\mathsf{T} \\ \overline{\mathbf{b}}_f^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}}' + \mathbf{f} \cdot \begin{bmatrix} \mathbf{0} \\ \mathbf{z} \end{bmatrix},$$

  and $\overline{w} + c \cdot \overline{u} = \mathbf{A}\hat{\mathbf{z}}$ holds.

Shortness Check Over the Sigma-Protocol Response. To adapt the previous NTT shortness check in eq. (4.4) for $\hat{\mathbf{z}}$, we have

$$(\mathbf{z} - 2c)(\mathbf{z} - c)\mathbf{z} = \mathbf{y}^3 + 3(\mathbf{s} - 1)\mathbf{y}^2 \cdot c + (3\mathbf{s}^2 - 6\mathbf{s} + 2)\mathbf{y} \cdot c^2 + (\mathbf{s} - 1)(\mathbf{s} - 2)\mathbf{s} \cdot c^3.$$

To prove the degree-3 evaluation over c, the trivial way is to use Π_{lin} to prove the commitments to $\mathbf{y}^3$, $3(\mathbf{s}-1)\mathbf{y}^2$, and $(3\mathbf{s}^2 - 6\mathbf{s} + 2)\mathbf{y}$ satisfy the linear relation over coefficients $1, c, c^2$, then the public parameter should be $R_q^{6\times 7}$.

One optimization in [BLS19] is to reuse the commitment to $\mathbf{s}$, such that the linear relation to be done is over $1, c$, while c^2 should be multiplying 0, and

$$\begin{aligned}(\mathbf{z} - 2c)(\mathbf{z} - c)\mathbf{s} &= \mathbf{s}\mathbf{y}^2 + \mathbf{y}\mathbf{s}(2\mathbf{s} - 3) \cdot c + (\mathbf{s} - 2)(\mathbf{s} - 1)\mathbf{s} \cdot c^2 \\ &= \mathbf{y}(2\mathbf{s} - 3) \cdot \mathbf{z} - \mathbf{y}^2(\mathbf{s} - 3) + (\mathbf{s} - 2)(\mathbf{s} - 1)\mathbf{s} \cdot c^2.\end{aligned}$$

Either way is fine, but this allows for committing to only 4 elements, and the public parameter is decreased to $R_q^{5\times 6}$.

Final Construction. We now present the final construction, and we use $\mathbf{B} \in R_q^{5\times 6}$ for the BDLOP public parameter, such that we commit to $\mathbf{y}$, $\mathbf{s}$, $\mathbf{y}(2\mathbf{s} - 3)$, and $\mathbf{y}^2(\mathbf{s} - 3)$.

- The prover sends out $\overline{w} \leftarrow \mathbf{A}\hat{\mathbf{y}}$, where $\hat{\mathbf{y}}$ is uniformly at random. Moreover, the prover samples $\overline{\mathbf{r}} \leftarrow \chi^{6n}$, commits

$$\overline{\mathbf{c}} \leftarrow \mathbf{B}\overline{\mathbf{r}} + \begin{bmatrix} \mathbf{0} \\ \mathbf{y} \\ \mathbf{s} \\ \mathbf{y}(2\mathbf{s} - 3) \\ \mathbf{y}^2(\mathbf{s} - 3) \end{bmatrix},$$

  and sends $\overline{\mathbf{c}}$ over to the verifier.
- The verifier challenges the prover by sending $c \xleftarrow{R} \mathbb{Z}_q$.
- The prover replies with $\mathbf{z} \leftarrow c \cdot \mathbf{s} + \mathbf{y}$, and starts Π_{lin} by:
  - The prover samples $\overline{\mathbf{y}}' \leftarrow D_\sigma^{6n}$, let

$$\begin{aligned}f_0(\mathbf{y}, \mathbf{s}, \mathbf{m}_4, \mathbf{m}_5) &= c \cdot \mathbf{s} + \mathbf{y}, \\ f_1(\mathbf{y}, \mathbf{s}, \mathbf{m}_4, \mathbf{m}_5) &= (\mathbf{z} - 2c)(\mathbf{z} - c) \cdot \mathbf{s} - \mathbf{m}_4 \cdot \mathbf{z} + \mathbf{m}_5,\end{aligned}$$

    and sends over

$$\overline{\mathbf{w}}' = \begin{bmatrix} \overline{\mathbf{b}}_1^\mathsf{T} \\ \overline{\mathbf{b}}_{f_0}^\mathsf{T} \\ \overline{\mathbf{b}}_{f_1}^\mathsf{T} \end{bmatrix} \overline{\mathbf{y}}'.$$

  - The verifier samples $\mathbf{f} \xleftarrow{R} C$, where C is the short challenge set, and sends over $\mathbf{f}$.
  - The prover runs reject sampling, and outputs $\overline{\mathbf{z}}' \leftarrow \overline{\mathbf{y}}' + \mathbf{f} \cdot \overline{\mathbf{r}}$.
- The verifier checks $\overline{\mathbf{z}}'$ being short, the following equality holds

$$\overline{\mathbf{w}}' + \mathbf{f} \cdot \begin{bmatrix} \mathbf{c}_1 \\ f_0(\mathbf{c}_2, \ldots, \mathbf{c}_5) \\ f_1(\mathbf{c}_2, \ldots, \mathbf{c}_5) \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{b}}_1^\mathsf{T} \\ \overline{\mathbf{b}}_{f_0}^\mathsf{T} \\ \overline{\mathbf{b}}_{f_1}^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}}' + \mathbf{f} \cdot \begin{bmatrix} \mathbf{0} \\ \mathbf{z} \\ \mathbf{0} \end{bmatrix},$$

  and $\overline{w} + c \cdot \overline{u} = \mathbf{A}\hat{\mathbf{z}}$ holds.

Knowledge Extraction. We let ε be the convincing probability of the prover, so the prover generates an accepting transcript in expected $1/\varepsilon$ steps. The knowledge extraction process is similar to that in [BLNS20], where one needs distinct challenges $c_1, c_2, c_3 \in \mathbb{Z}_q$ because there is a degree-2 polynomial over c in the shortness check, and the last step is an adaptation of the BDLOP proof of a linear relation, for each c_i, by [BDL+18], we need 2 distinct $\mathbf{f}_{i,1}, \mathbf{f}_{i,2} \in C$ for a relaxed opening, and therefore in total we need 6 accepting transcripts forked in 3 by 2.

We begin by bounding the expected runtime of the extractor, where the "heavy-row" argument [Dam10] is used in a similar way as [BBC+18, BLNS20]. If the probability of convincing is ε, consider all transcripts indexed by the first challenge $c \in \mathbb{Z}_q$ and the second challenge $\mathbf{f} \in C$, in a 0-1 matrix.

Start with the first pair of accepting transcripts decided by c_1. A "heavy row" indexed by challenge c is a row with entries being 1 of fraction at least $\varepsilon/2$, then by the "heavy-row" argument in [Dam10], the first accepting transcript is on a heavy row with probability at least $1/2$. Conditioned that the first accepting transcript is on a heavy row, the second accepting transcript is obtained with probability at least $\varepsilon/2 - |C|^{-1}$, as $\mathbf{f}_{1,1} \neq \mathbf{f}_{1,2}$. Hence, the expected runtime for the first pair is

$$\frac{1}{\varepsilon} + \frac{2}{\varepsilon - 2|C|^{-1}}.$$

The second pair of accepting transcripts is similar: The first of them is obtained with probability at least $\varepsilon - 1/q$ as $c_2 \neq c_1$, then let the heavy row with entries being 1 of fraction at least $\varepsilon/2 - 1/(2q)$, then the first of them lands on a heavy row with probability at least $1/2$, and conditioned that, the second of them is obtained with probability at least $\varepsilon/2 - 1/(2q) - |C|^{-1}$. Hence, the expected runtime for the second pair is

$$\frac{1}{\varepsilon - q^{-1}} + \frac{2}{\varepsilon - q^{-1} - 2|C|^{-1}}.$$

Hence, conditioned on all three pairs being on heavy rows, occurring with probability at least $1/8$, the expected runtime is at most

$$T = \frac{9}{\varepsilon - 2q^{-1} - 2|C|^{-1}},$$

and we cap the runtime of the extractor at $2T$. By Markov's inequality, finding 3 pairs of accepting transcripts in $2T$ steps succeeds with probability at least $1/16$ (conditioned that all are on heavy rows, the probability is at least $1/2$ by Markov's inequality, then $1/16$ is immediate). Restarting after failure thus gives expected runtime at most $32T$ [22].

For $i \in [1,3]$, let $\mathbf{f}'_i = \mathbf{f}_{i,1} - \mathbf{f}_{i,2}$, and $\overline{\mathbf{z}}''_i = \overline{\mathbf{z}}'_{i,1} - \overline{\mathbf{z}}'_{i,2}$ be the relaxed-opening randomness obtained by taking the difference of the two masked responses. The difference $\mathbf{f}'_i$ is invertible over R_q by $\overline{C}$, as $\mathbf{f}_{i,1}, \mathbf{f}_{i,2} \in C$ are distinct. The two transcripts have the same c_i, $\mathbf{z}_i$, and three masking terms. Subtracting their three verification equations cancels the common masking terms and gives a branch-specific pair of linear functions. Namely, let

$$\begin{aligned} g_i &:= (\mathbf{z}_i - c_i)(\mathbf{z}_i - 2c_i), \\ f_{0,i}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5) &:= \mathbf{x}_2 + c_i \mathbf{x}_3, \\ f_{1,i}(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5) &:= g_i \mathbf{x}_3 - \mathbf{z}_i \mathbf{x}_4 + \mathbf{x}_5, \end{aligned}$$

with the corresponding rows

$$\begin{aligned} \overline{\mathbf{b}}^{\mathsf{T}}_{f_{0,i}} &:= \overline{\mathbf{b}}^{\mathsf{T}}_2 + c_i \overline{\mathbf{b}}^{\mathsf{T}}_3, \\ \overline{\mathbf{b}}^{\mathsf{T}}_{f_{1,i}} &:= g_i \overline{\mathbf{b}}^{\mathsf{T}}_3 - \mathbf{z}_i \overline{\mathbf{b}}^{\mathsf{T}}_4 + \overline{\mathbf{b}}^{\mathsf{T}}_5. \end{aligned}$$

The subtraction is then written in the same matrix form as the masked BDLOP opening:

$$\mathbf{f}'_i \begin{bmatrix} \mathbf{c}_1 \\ f_{0,i}(\mathbf{c}_2, \ldots, \mathbf{c}_5) \\ f_{1,i}(\mathbf{c}_2, \ldots, \mathbf{c}_5) \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{b}}^{\mathsf{T}}_1 \\ \overline{\mathbf{b}}^{\mathsf{T}}_{f_{0,i}} \\ \overline{\mathbf{b}}^{\mathsf{T}}_{f_{1,i}} \end{bmatrix} \overline{\mathbf{z}}''_i + \mathbf{f}'_i \begin{bmatrix} \mathbf{0} \\ \mathbf{z}_i \\ \mathbf{0} \end{bmatrix}.$$

[22] [BLS19] states $16T$, apparently omitting the factor 2 from the timeout.

Define the extracted messages by

$$\mathbf{y}_i^* = \mathbf{c}_2 - (\mathbf{f}_i')^{-1}\overline{\mathbf{b}}_2^{\mathsf{T}}\overline{\mathbf{z}}_i'', \qquad \mathbf{s}_i^* = \mathbf{c}_3 - (\mathbf{f}_i')^{-1}\overline{\mathbf{b}}_3^{\mathsf{T}}\overline{\mathbf{z}}_i'',$$
$$\mathbf{m}_{4,i}^* = \mathbf{c}_4 - (\mathbf{f}_i')^{-1}\overline{\mathbf{b}}_4^{\mathsf{T}}\overline{\mathbf{z}}_i'', \qquad \mathbf{m}_{5,i}^* = \mathbf{c}_5 - (\mathbf{f}_i')^{-1}\overline{\mathbf{b}}_5^{\mathsf{T}}\overline{\mathbf{z}}_i''.$$

The first row of this equation and these definitions give the relaxed opening

$$\mathbf{f}_i'\overline{\mathbf{c}} = \mathbf{B}\overline{\mathbf{z}}_i'' + \mathbf{f}_i' \begin{bmatrix} \mathbf{0} \\ \mathbf{y}_i^* \\ \mathbf{s}_i^* \\ \mathbf{m}_{4,i}^* \\ \mathbf{m}_{5,i}^* \end{bmatrix}, \qquad \left\|\overline{\mathbf{z}}_i''\right\|_2 \le 2B.$$

If two outer branches yield different messages, the two relaxed openings break binding and give an $\mathsf{RSIS}_{5,8B}$ solution. Otherwise, all three branches open to fixed messages $\mathbf{y}^*, \mathbf{s}^*, \mathbf{m}_4^*, \mathbf{m}_5^*$ that are independent of c_i.

The second row of the matrix equation then gives

$$\mathbf{z}_i = f_{0,i}(\mathbf{y}^*, \mathbf{s}^*, \mathbf{m}_4^*, \mathbf{m}_5^*) = \mathbf{y}^* + c_i\mathbf{s}^* \qquad \text{for } i \in [1,3], \tag{4.12}$$

while the third row gives

$$f_{1,i}(\mathbf{y}^*, \mathbf{s}^*, \mathbf{m}_4^*, \mathbf{m}_5^*) = g_i\mathbf{s}^* - \mathbf{z}_i\mathbf{m}_4^* + \mathbf{m}_5^* = 0.$$

Substituting eq. (4.12) into the definition of g_i gives

$$g_i = \big(\mathbf{y}^* + c_i(\mathbf{s}^* - 1)\big)\big(\mathbf{y}^* + c_i(\mathbf{s}^* - 2)\big),$$

substituting this expression for g_i, together with the expression for $\mathbf{z}_i$ in eq. (4.12), into the third-row equation yields a degree-2 polynomial in c_i whose leading coefficient is

$$\mathbf{s}^*(\mathbf{s}^* - 1)(\mathbf{s}^* - 2).$$

The polynomial vanishes at the three distinct points c_1, c_2, c_3. The corresponding Vandermonde matrix is invertible over $\mathbb{Z}_q$, hence also over R_q, and therefore

$$\mathbf{s}^*(\mathbf{s}^* - 1)(\mathbf{s}^* - 2) = 0.$$

Applying the NTT shows that $\widehat{\mathbf{s}}^* \in \{0,1,2\}^n$.

Finally, subtracting the outer verification equations for two distinct challenges gives

$$\mathbf{A}(\widehat{\mathbf{z}}_1 - \widehat{\mathbf{z}}_2) = (c_1 - c_2)\overline{u}.$$

By eq. (4.12), $\widehat{\mathbf{z}}_1 - \widehat{\mathbf{z}}_2 = (c_1 - c_2)\widehat{\mathbf{s}}^*$. Since $c_1 - c_2$ is invertible in $\mathbb{Z}_q$, we conclude that

$$\mathbf{A}\widehat{\mathbf{s}}^* = \overline{u}.$$

Therefore, the extractor outputs an exact witness $\widehat{\mathbf{s}}^* \in \{0,1,2\}^n$, or it produces an $\mathsf{RSIS}_{5,8B}$ solution.

Knowledge Soundness. Following [BG93, Dam10], the knowledge error κ can be viewed as an upper bound on the acceptance probability that does not yet force extractable knowledge. Informally, it is the best convincing probability that the analysis permits for a prover that does not know a correct witness. If the prover goes beyond this threshold, the extractor can take a witness out of its behavior. This interpretation does not assert that a prover without a witness can actually attain κ: the bound need not be tight, and a particular prover may do strictly worse.

For comparison, in a flat protocol with a uniform challenge set of size N, if any r distinct accepting branches from the same prefix suffice for extraction, then at most $r-1$ branches may remain non-extractable, giving $\kappa = (r-1)/N$.

Thus two-branch extraction gives $1/N$, while three-branch extraction gives $2/N$. This is only the flat intuition: the extractor above requires a nested 3-by-2 challenge tree.

For the current protocol, let

$$\kappa := \frac{2}{q} + \frac{2}{|C|}.$$

When obtaining the third pair of transcripts, excluding the two previously used outer challenges loses at most $2/q$ of the acceptance probability. On a heavy row, the acceptance probability for the second challenge is therefore at least one half of $\varepsilon - 2/q$. Excluding the previously used second challenge then leaves probability at least

$$\frac{\varepsilon}{2} - \frac{1}{q} - \frac{1}{|C|} = \frac{1}{2}(\varepsilon - \kappa)$$

for obtaining the required inner fork. Hence $\varepsilon > \kappa$ makes every denominator in the extraction above positive, and the heavy-row argument of [Dam10] finds the nested challenge tree in expected time $\mathsf{poly}(\lambda)/(\varepsilon - \kappa)$.

The extractor then outputs either an exact witness $\widehat{\mathbf{s}}^* \in \{0,1,2\}^n$ satisfying $\mathbf{A}\widehat{\mathbf{s}}^* = \overline{u}$, or an $\mathsf{RSIS}_{5,8B}$ solution. Under the $\mathsf{RSIS}_{5,8B}$ hardness assumption, the latter event is negligible. This realizes the Bellare-Goldreich interpretation: acceptance beyond κ by a non-negligible gap yields extractable knowledge, with the inverse gap governing the expected extraction time.

Rejection Sampling and Honest-Prover Bounds. All norms in this paragraph are coefficient ℓ_2-norms. Let $\overline{\mathbf{v}} := \mathbf{f}\overline{\mathbf{r}}$, where $\overline{\mathbf{r}} \leftarrow \chi^{6n}$ and $\mathbf{f} \xleftarrow{R} C$. Since $\mathbf{f} \in C$ defined in eq. (4.7), it does not increase the coefficient norm. Moreover, $\|\overline{\mathbf{r}}\|_2^2$ is a sum of $6n$ independent Bernoulli trials with success probability $5/8$. Hence, for $0 < \delta \le 1$,

$$\Pr\left[\|\overline{\mathbf{v}}\|_2 \le T_{\text{norm}}\right] \ge 1 - \exp\left(-\frac{5\delta^2 n}{4}\right), \qquad T_{\text{norm}} := \sqrt{(1+\delta)\frac{15n}{4}}$$

by Chernoff's bound. Choose δ so that the failure probability is at most 2^{-101}, and choose $\sigma \ge 5T_{\text{norm}}$.

For $\overline{\mathbf{y}}' \leftarrow D_\sigma^{6n}$, the uncorrected response $\overline{\mathbf{z}}' = \overline{\mathbf{y}}' + \overline{\mathbf{v}}$ is distributed according to $D_{\overline{\mathbf{v}},\sigma}^{6n}$, centered at $\overline{\mathbf{v}}$ rather than $\overline{\mathbf{0}}$. The prover accepts this response with probability

$$\min\left\{\frac{1}{12} \cdot \frac{\rho_\sigma(\overline{\mathbf{z}}')}{\rho_\sigma(\overline{\mathbf{z}}' - \overline{\mathbf{v}})}, 1\right\}.$$

By the rejection-sampling lemma of [Lyu12, BLS19], acceptance occurs with probability at least $1/12 - 2^{-104}$. Conditioned on acceptance, the joint distribution of $(\overline{\mathbf{v}}, \overline{\mathbf{z}}')$ is within statistical distance 2^{-100} of independently sampling $\overline{\mathbf{v}}$ as above and $\overline{\mathbf{z}}' \leftarrow D_\sigma^{6n}$. Thus the accepted response is statistically close to a zero-centered Gaussian independent of the shift $\overline{\mathbf{v}} = \mathbf{f}\overline{\mathbf{r}}$.

Finally, by [Ban93], over $\overline{\mathbf{z}}' \leftarrow D_\sigma^{6n}$, the Gaussian tail bound gives

$$\Pr\left[\left\|\overline{\mathbf{z}}'\right\|_2 \le \sigma\sqrt{12n}\right] \ge 1 - (2/e)^{3n/2}.$$

We therefore set $B := \sigma\sqrt{12n}$. An accepted response fails the verifier's norm check with probability at most

$$2^{-100} + (2/e)^{3n/2}.$$

Upon rejection, the prover aborts and restarts with fresh randomness, so an honest execution succeeds with probability approximately $1/12$ per attempt.

4.12 Practical Product Proof from Lattice Commitments

Polynomial Ring Elements Don't Split Into Many Factors. One contribution of [ALS20], building on [LS18], is the observation that polynomial ring elements cannot split into many factors for a short invertible challenge set. More specifically, consider the particular setting of a power-of-two ring from theorem 4.10.1 and theorem 4.10.6.

Lemma 4.12.1 (Power-of-two Subgroup Order [LS18])**.** *Let $4 \le k \le n$ be both power of two, and $a \equiv 1 + k \mod 2k$, then $\mathsf{ord}_n(a) = n/k$, namely the cyclic subgroup $\langle a \rangle \subset \mathbb{Z}_n^*$ generated by a has order n/k.*

Proof. The $k = n$ case is trivial, as $a \equiv 1 + n \mod 2n$. Henceforth assume $k < n$.

Since $a \equiv 1 + k \mod 2k$, we write $a = 1 + a_1 k$ where a_1 is odd. As a result,

$$a^2 = 1 + 2ka_1 + k^2 a_1^2 = 1 + (2k)a_1\,(1 + ka_1/2) = 1 + (2k)a_2,$$

Since $k \ge 4$ is a power of two, $1 + ka_1/2$ is odd, and hence so is a_2. By $a^2 \equiv 1 + 2k \mod 4k$, we have

$$a^{n/k} \equiv 1 + n \mod 2n,$$

and moreover, $a^{n/(2k)} \equiv 1 + n/2 \mod n$. Therefore, $\langle a \rangle \subset \mathbb{Z}_n^*$ has order n/k. □

Then theorem 4.10.1 can be updated as follows.

Corollary 4.12.2. *Let $m = 2d$ and $z = 2\ell$ both be powers of two, with $4 \le z \le m$. If q is prime with $q \equiv 1 + z \mod 2z$, then for distinct primitive z^{th} roots of unity r_j,*

$$\Phi_m(X) = X^d + 1 \equiv \prod_{j \in [1,\ell]} \left(X^{d/\ell} - r_j\right) \mod q$$

and each factor $X^{d/\ell} - r_j$ is irreducible over $\mathbb{Z}_q$.

We also update theorem 4.10.6 as follows.

Corollary 4.12.3. *Following corollary 4.12.2, any $\mathbf{y} \in R_{m,q}$ satisfying*

$$0 < \|\mathbf{y}\|_\infty < \ell^{-1/2} \cdot q^{1/\ell} = \frac{1}{s_1(z)} q^{1/\varphi(z)}$$

is invertible in $R_{m,q}$.

Proof. By [ACX21, LPR13], $s_1(z) \le \sqrt{\tau(z)} = \sqrt{\ell}$, with equality here. The rest follows from $\varphi(z) = \ell$. □

For pairwise difference of distinct challenges being invertible (to reduce binding to the openings to SIS hardness), we need $q^{1/\ell} \cdot \ell^{-1/2} > 2$, and eventually $q > \left(2\sqrt{\ell}\right)^\ell$. The following table shows that the polynomial ring cannot split into many irreducible factors without making the modulus too large for computational efficiency.

Number of factors ℓ	Required modulus q
4	$q > 2^8$
8	$q > 2^{20}$
16	$q > 2^{48}$
32	$q > 2^{112}$
64	$q > 2^{256}$

Table 4.12.1: Required modulus size for different numbers of irreducible factors.

This gives rise to the question behind [ALS20]: what if the challenge set is not globally invertible, but the probability that a pairwise difference is noninvertible is low, such as $\mathrm{poly}(q^{-1})$? The goal is a good challenge set even when the ring splits into a large number of irreducible factors.

Prime Splittings and Automorphisms. To describe the splitting in corollary 4.12.2 better, we point out that R_q has a group of automorphisms $\mathsf{Aut}(R_q)$ that is isomorphic to $\mathbb{Z}_{2d}^*$. We write $\mathsf{Aut}(R_q)$ for the cyclotomic automorphism subgroup consisting of the maps $f(X) \mapsto f(X^i)$ for $i \in \mathbb{Z}_{2d}^*$. This group is isomorphic to $\mathbb{Z}_{2d}^*$:

$$i \to \sigma_i \ : \ \mathbb{Z}_{2d}^* \to \mathsf{Aut}(R_q),$$

where

$$\begin{aligned} f(X) &= \sum_{j\in[0,d-1]} a_j X^j, \\ \sigma_i(f(X)) &= \sum_{j\in[0,d-1]} a_j \sigma_i(X^j) = \sum_{j\in[0,d-1]} a_j X^{ij} = f(X^i). \end{aligned} \tag{4.13}$$

The group $\mathsf{Aut}(R_q)$ acts transitively on prime ideals $(X^{d/\ell} - \xi)$ in R_q, and every σ_i factors through field isomorphisms

$$\mathbb{Z}_q[X]/\left(X^{d/\ell} - \xi\right) \cong R_q/\left(X^{d/\ell} - \xi\right) \to \mathbb{Z}_q[X]/\left(X^{d/\ell} - \xi^{i^{-1}}\right) \cong R_q/\left(\sigma_i\left(X^{d/\ell} - \xi\right)\right).$$

Concretely, by automorphism, we have

$$\sigma_i\left(\langle X^{d/\ell} - \xi\rangle\right) = \langle X^{id/\ell} - \xi\rangle = \langle X^{d/\ell} - \xi^{i^{-1}}\rangle$$

in R_q, then for $\mathbf{f} \in R_q$,

$$\sigma_i\left(\mathbf{f} \mod \left(X^{d/\ell} - \xi\right)\right) = \sigma_i(\mathbf{f}) \mod \left(X^{d/\ell} - \xi^{i^{-1}}\right).$$

The multiplicative subgroup $\langle 2\ell + 1\rangle \subseteq \mathbb{Z}_{2d}^*$ has order d/ℓ by corollary 4.12.2, and any $i \in \langle 2\ell + 1\rangle$ stabilizes the prime ideal $\left(X^{d/\ell} - \xi\right)$, namely

$$\sigma_i\left(X^{d/\ell} - \xi\right) = X^{d/\ell} - \xi,$$

as ξ is a primitive $2\ell^{\text{th}}$ root of unity, then $i \in \langle 2\ell + 1\rangle \subset \mathbb{Z}_{2d}^*$ has $\xi^{i^{-1}} = \xi$. The quotient group $\mathbb{Z}_{2d}^*/\langle 2\ell + 1\rangle \cong \mathbb{Z}_{2\ell}^*$ has order ℓ, and since elements in $\langle 2\ell + 1\rangle$ stabilize the prime ideal, we can index the ℓ prime ideals by

$$\left(X^d + 1\right) = \prod_{i\in\mathbb{Z}_{2d}^*/\langle 2\ell+1\rangle} \left(X^{d/\ell} - \xi^i\right). \tag{4.14}$$

Moreover, for $k \mid \ell$ with $k < \ell$, $\langle 2\ell/k + 1\rangle \subset \mathbb{Z}_{2d}^*$ has order kd/ℓ, and thus $\langle 2\ell/k + 1\rangle/\langle 2\ell + 1\rangle$ has order k. Then

$$\left(X^{kd/\ell} - \xi^k\right) = \prod_{i\in\langle 2\ell/k+1\rangle/\langle 2\ell+1\rangle} \left(X^{d/\ell} - \xi^i\right), \tag{4.15}$$

and therefore

$$\left(X^d + 1\right) = \prod_{i\in\mathbb{Z}_{2d}^*/\langle 2\ell/k+1\rangle} \left(X^{kd/\ell} - \xi^{ik}\right) = \prod_{i\in\mathbb{Z}_{2d}^*/\langle 2\ell/k+1\rangle} \prod_{j\in\langle 2\ell/k+1\rangle/\langle 2\ell+1\rangle} \left(X^{d/\ell} - \xi^{ij}\right).$$

Another way of showing it is, let $\sigma = \sigma_{2\ell/k+1}$, since $\{(2\ell/k + 1)^i \mod 2d\}_{i\in[0,k-1]}$ gives $\langle 2\ell/k + 1\rangle/\langle 2\ell + 1\rangle$, then

$$\left(X^{kd/\ell} - \xi^k\right) = \prod_{i\in[0,k-1]} \sigma^i\left(X^{d/\ell} - \xi\right). \tag{4.16}$$

By $\mathbb{Z}_{2d}^*/\langle 2\ell/k + 1\rangle \cong \mathbb{Z}_{2\ell/k}^*$, then

$$\left(X^d + 1\right) = \prod_{i\in\mathbb{Z}_{2\ell/k}^*} \prod_{j\in[0,k-1]} \sigma^j\left(X^{d/\ell} - \xi^i\right),$$

where the prime ideals are indexed by $\mathbb{Z}_{2\ell/k}^* \times [0, k-1]$.

Distribution in the NTT Coefficients. Consider the distribution for the coefficients in $R_q = \mathbb{Z}_q[X]/(X^d+1)$, where each coefficient is over $\{-1,0,1\}$, 0 has probability p, and ± 1 has probability $(1-p)/2$ each. Such distribution C is the distribution for the challenge set.

Lemma 4.12.4 (Distributional Invariance across NTT Factors [ALS20]). *Let $\mathbf{x} \leftarrow C$ where the coefficients are i.i.d. Then $\mathbf{x} \mod \left(X^{d/\ell} - \xi^i\right)$ is identically distributed as $\mathbf{x} \mod \left(X^{d/\ell} - \xi^j\right)$ for all $i, j \in \mathbb{Z}_{2\ell}^*$.*

Proof. First, for any automorphism $\sigma \in \mathsf{Aut}(R_q)$, if $\mathbf{x} \leftarrow C$, then $\mathbf{x}$ is identically distributed as $\sigma(\mathbf{x})$, as the coefficients of $\sigma(\mathbf{x})$ are also i.i.d. over $\{-1,0,1\}$ with the same distribution as C.

Suppose $(X^{d/\ell} - \xi^i)$ is a prime ideal in $\mathbb{Z}_q[X]/\left(X^d+1\right)$, then as previously discussed,

$$R_q/\left(X^{d/\ell} - \xi^i\right) \cong R_q/\left(X^{d/\ell} - \xi^j\right)$$

as ξ has an order 2ℓ, and each NTT component is isomorphic to each other under an automorphism in $\mathsf{Aut}(R_q)$, then

$$\sigma\left(\mathbf{x} \mod \left(X^{d/\ell} - \xi^i\right)\right) = \sigma(\mathbf{x}) \mod \left(X^{d/\ell} - \xi^j\right)$$

if $\sigma \in \mathsf{Aut}(R_q)$ induces the isomorphism. Thus, the prime-ideal case for distributional invariance across NTT factors is concluded.

If $(X^{d/\ell} - \xi^i)$ is not irreducible, then neither is $(X^{d/\ell} - \xi^j)$, and both split into the same number of irreducible factors. By the existence of an automorphism inducing an isomorphism from one irreducible factor to the other, we conclude the proof. □

Let $\alpha = d/\ell$. We reorder the coefficients of $\mathbf{c} \leftarrow C$ as follows:

$$\begin{array}{llll} \mathbf{c} = c_0 & + c_\alpha X^\alpha & + \cdots & + c_{d-\alpha} X^{d-\alpha} \\ \quad c_1 X & + c_{\alpha+1} X^{\alpha+1} & + \cdots & + c_{d-\alpha+1} X^{d-\alpha+1} \\ & & \cdots & \\ \quad c_{\alpha-1} X^{\alpha-1} & + c_{2\alpha-1} X^{2\alpha-1} & + \cdots & + c_{d-1} X^{d-1}. \end{array}$$

Hence, the d/ℓ coefficients of $\mathbf{c} \mod (X^\alpha - \xi)$ are i.i.d., which can be shown as

$$\mathbf{c} \equiv \begin{bmatrix} 1 & X & \cdots & X^{\alpha-1} \end{bmatrix} \begin{bmatrix} c_0 & + & c_\alpha \xi & + \cdots & + & c_{d-\alpha}\xi^{\ell-1} \\ c_1 & + & c_{\alpha+1}\xi & + \cdots & + & c_{d-\alpha+1}\xi^{\ell-1} \\ & & & \cdots & & \\ c_{\alpha-1} & + & c_{2\alpha-1}\xi & + \cdots & + & c_{d-1}\xi^{\ell-1} \end{bmatrix} \mod (X^\alpha - \xi), \tag{4.17}$$

where each row, namely each coefficient of $\mathbf{c} \mod (X^\alpha - \xi)$, is represented by the random variable

$$Y = \sum_{i \in [0,\ell-1]} \xi^i X_i,$$

and the X_i are i.i.d. according to the $\{-1,0,1\}$ distribution. We now present the lemma for the distribution of Y.

Lemma 4.12.5 (Projected Challenge Anti-Concentration [ALS20]). *For all $x \in \mathbb{Z}_q$,*

$$\Pr[Y = x] \le \frac{1}{q} + \frac{1}{q} \sum_{j \in \mathbb{Z}_q^*} \prod_{k \in [0,\ell-1]} \left| p + (1-p)\cos\left(2\pi j \xi^k / q\right) \right|.$$

Proof. Let $P : \mathbb{Z}_q \to [0,1]$ be the probability mass function. It is a convolution of the following distributions:

$$\mu_k(0) = p, \quad \mu_k(\xi^k) = \mu_k(-\xi^k) = (1-p)/2$$

for $k \in [0, \ell-1]$, as $Y = \sum_{i\in[0,\ell-1]} \xi^i X_i$. Hence, by Fourier analysis,

$$\hat{\mu}_k(j) = \sum_{\gamma\in\mathbb{Z}_q} \mu_k(\gamma) \exp\left(-2\pi i j\gamma/q\right) = p + (1-p)\cos\left(2\pi j\xi^k/q\right).$$

By convolution,

$$\hat{P}(j) = \prod_{k\in[0,\ell-1]} \hat{\mu}_k(j) = \prod_{k\in[0,\ell-1]} \left(p + (1-p)\cos\left(2\pi j\xi^k/q\right)\right).$$

By inverse Fourier transform,

$$\begin{aligned} P(x) &= \frac{1}{q}\sum_{j\in\mathbb{Z}_q} \hat{P}(j)\exp\left(2\pi \mathrm{i} jx/q\right) = \frac{1}{q} + \frac{1}{q}\sum_{j\in\mathbb{Z}_q^*} \hat{P}(j)\exp\left(2\pi \mathrm{i} jx/q\right) \\ &\le \frac{1}{q} + \frac{1}{q}\sum_{j\in\mathbb{Z}_q^*} \left|\hat{P}(j)\right| = \frac{1}{q} + \frac{1}{q}\sum_{j\in\mathbb{Z}_q^*}\prod_{k\in[0,\ell-1]} \left|p + (1-p)\cos\left(2\pi j\xi^k/q\right)\right|. \end{aligned}$$

□

By taking advantage of symmetry, we can reduce the number of terms needed to compute the probability upper bound by a factor of 2ℓ.

Lemma 4.12.6. *For all $x \in \mathbb{Z}_q$,*

$$\Pr[Y = x] \le \frac{1}{q} + \frac{2\ell}{q}\sum_{j\in\mathbb{Z}_q^*/\langle\xi\rangle}\prod_{k\in[0,\ell-1]} \left|p + (1-p)\cos\left(2\pi j\xi^k/q\right)\right|. \tag{4.18}$$

Proof. Since $\langle\xi\rangle \subset \mathbb{Z}_q^*$ has order 2ℓ, then $\{1, \xi, \ldots, \xi^{\ell-1}\} = \langle\xi\rangle/\pm 1$. Supposing $g \in \mathbb{Z}_q^*$ has order $q-1$ that generates $\mathbb{Z}_q^*$, and $\xi = g^{(q-1)/(2\ell)}$, then the cosets of $\langle\xi\rangle$ can be represented by $\mathbb{Z}_q^*/\langle\xi\rangle = \{1, \ldots, g^{(q-1)/(2\ell)-1}\}$. Hence, multiplying any ξ^m to $\{j, j\xi, \ldots, j\xi^{\ell-1}\}$ permutes these elements up to sign. Since cosine is even, then $\hat{P}(j) = \hat{P}(j\xi^m)$. □

Each row can be represented in Horner's polynomial evaluation form

$$x_0 + \xi\left(x_1 + \xi\left(x_2 + \ldots + \xi\left(x_{\ell-2} + \xi x_{\ell-1}\right)\right)\right),$$

which can be transformed into a Markov chain: $Z_0 = 0$, $Z_{i+1} = \xi Z_i + X_i$ where X_i follows the $\{-1, 0, 1\}$ distribution.

Module-SIS and Module-LWE. We recall the assumptions underlying the binding and hiding properties of the commitment scheme in [ALS20].

Definition 4.12.7 ($\mathsf{MSIS}_{n,m,B}$)**.** Let $n, m \ge 1$ and $0 < B < q$. An efficient $\mathcal{A}$ has advantage ε in solving $\mathsf{MSIS}_{n,m,B}$ if

$$\Pr\left[0 < \|\bar{\mathbf{x}}\|_2 \le B \ \wedge\ \mathbf{A}\bar{\mathbf{x}} = \bar{\mathbf{0}} \,\middle|\, \mathbf{A} \xleftarrow{\text{R}} R_q^{n\times m}, \bar{\mathbf{x}} \leftarrow \mathcal{A}(\mathbf{A})\right] \ge \varepsilon,$$

where $\bar{\mathbf{x}} \in R_q^m$, the equation is over R_q, and the norm is the coefficient ℓ_2-norm of the whole vector, using centered representatives.

Fix a coefficient distribution χ over $\mathbb{Z}$. As before, χ^{kd} denotes independent sampling of all coefficients in R^k.

Definition 4.12.8 ($\mathsf{MLWE}_{n,m,\chi}$)**.** Let $n, m \ge 1$. A PPT algorithm $\mathcal{A}$ has advantage ε in solving $\mathsf{MLWE}_{n,m,\chi}$ if

$$\left|\Pr\left[\mathcal{A}(\mathbf{A}, \mathbf{A}\bar{\mathbf{s}} + \bar{\mathbf{e}}) = 1 \,\middle|\, \begin{array}{l}\mathbf{A} \xleftarrow{\text{R}} R_q^{m\times n},\\ \bar{\mathbf{s}} \leftarrow \chi^{nd},\\ \bar{\mathbf{e}} \leftarrow \chi^{md}\end{array}\right] - \Pr\left[\mathcal{A}(\mathbf{A}, \bar{\mathbf{u}}) = 1 \,\middle|\, \begin{array}{l}\mathbf{A} \xleftarrow{\text{R}} R_q^{m\times n},\\ \bar{\mathbf{u}} \xleftarrow{\text{R}} R_q^m\end{array}\right]\right| \ge \varepsilon.$$

All samples within each experiment are independent, and the arithmetic is over R_q.

Following [ALS20], we omit m and write $\mathsf{MSIS}_{n,B}$ and $\mathsf{MLWE}_{n,\chi}$. The BDLOP commitment matrix follows, where μ and λ denote the module ranks for MSIS and MLWE, respectively.

For L committed ring elements, put $N = \mu + \lambda + L$. The block structure used in [BLS19] generalizes to

$$\mathbf{B} = \begin{bmatrix} \mathbf{I}_\mu & \mathbf{U} \\ \mathbf{0}^{L\times\mu} & \mathbf{I}_L \end{bmatrix} \begin{bmatrix} \mathbf{I}_\mu & \mathbf{0}^{\mu\times L} & \mathbf{V} \\ \mathbf{0}^{L\times\mu} & \mathbf{I}_L & \mathbf{W} \end{bmatrix} = \begin{bmatrix} \mathbf{I}_\mu & \mathbf{U} & \mathbf{V}+\mathbf{UW} \\ \mathbf{0}^{L\times\mu} & \mathbf{I}_L & \mathbf{W} \end{bmatrix} \in R_q^{(\mu+L)\times N}, \tag{4.19}$$

where $\mathbf{U} \xleftarrow{R} R_q^{\mu\times L}$, $\mathbf{V} \xleftarrow{R} R_q^{\mu\times\lambda}$, and $\mathbf{W} \xleftarrow{R} R_q^{L\times\lambda}$. The instance in [BLS19] has $\mu = \lambda = 1$ and $L = 4$, and for the instance in [ALS20], the binding and hiding relies on $\mathsf{MSIS}_{\mu,8\kappa\beta}$ and $\mathsf{MLWE}_{\lambda,\chi}$.

In [ALS20], the public matrix is sampled uniformly with the same dimensions by $\mathbf{B} \xleftarrow{R} R_q^{(\mu+L)\times N}$ (for readability). For the hiding property, supposing the first $\mu + L$ columns form an invertible $\mathbf{C} \in R_q^{(\mu+L)\times(\mu+L)}$, such that

$$\mathbf{B} = \begin{bmatrix} \mathbf{C} & \mathbf{D} \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{I}_{\mu+L} & \mathbf{C}^{-1}\mathbf{D} \end{bmatrix} = \mathbf{CE},$$

then multiplying $\mathbf{E}$ with $\bar{\mathbf{r}} \leftarrow \chi^{Nd}$ yields a MLWE sample $\bar{\mathbf{g}}$, and $\bar{\mathbf{g}}$ multiplying with invertible $\mathbf{C}$ makes the result computationally indistinguishable to a uniform sample, so as to mask the message $\bar{\mathbf{m}}$. The formal argument is the knapsack-MLWE duality, see [EZS+19, Appendix C]. The binding is still reduced to $\mathsf{MSIS}_{\mu,8\kappa\beta}$ in the same reduction.

Another way of seeing is

$$\mathbf{B}' = \mathbf{R} \begin{bmatrix} \mathbf{I}_\mu & \mathbf{U} \\ \mathbf{0}^{L\times\mu} & \mathbf{I}_L \end{bmatrix} \begin{bmatrix} \mathbf{I}_\mu & \mathbf{0}^{\mu\times L} & \mathbf{V} \\ \mathbf{0}^{L\times\mu} & \mathbf{I}_L & \mathbf{W} \end{bmatrix} = \mathbf{R} \begin{bmatrix} \mathbf{I}_\mu & \mathbf{U} & \mathbf{V}+\mathbf{UW} \\ \mathbf{0}^{L\times\mu} & \mathbf{I}_L & \mathbf{W} \end{bmatrix},$$

where $\mathbf{R}$ is uniform over invertible $R_q^{(\mu+L)\times(\mu+L)}$, then hiding and binding are still reduced to $\mathsf{MLWE}_{\lambda,\chi}$ and $\mathsf{MSIS}_{\mu,8\kappa\beta}$.

To commit to $\bar{\mathbf{m}} = (\mathbf{m}_1, \ldots, \mathbf{m}_L)^\mathsf{T} \in R_q^L$, sample $\bar{\mathbf{r}} \leftarrow \chi^{Nd}$ and output

$$\begin{bmatrix} \bar{\mathbf{t}}_0 \\ \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_L \end{bmatrix} = \mathbf{B}\bar{\mathbf{r}} + \begin{bmatrix} \bar{\mathbf{0}} \\ \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_L \end{bmatrix}, \qquad \bar{\mathbf{t}}_0 = \mathbf{B}_0\bar{\mathbf{r}}, \quad \mathbf{t}_i = \langle \bar{\mathbf{b}}_i, \bar{\mathbf{r}} \rangle + \mathbf{m}_i.$$

Challenge Set without Difference Invertibility. An immediate application of the result on the NTT slot coefficient distribution is a distribution over C under which the coefficients are sampled i.i.d. over $\{0, 1\}$. [23] Hence, a difference of independently sampled distinct elements is non-invertible with probability at most $\ell p^{d/\ell}$ by union bound.

It is helpful to start from the BDLOP construction, as [ALS20] improves on [BDL+18] with a challenge set without the property of *globally invertible pairwise differences*. Regarding the commitment

$$\begin{bmatrix} \bar{\mathbf{t}}_0 \\ \mathbf{t}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_0 \\ \bar{\mathbf{b}}_1^\mathsf{T} \end{bmatrix} \bar{\mathbf{r}} + \begin{bmatrix} \bar{\mathbf{0}} \\ \mathbf{m} \end{bmatrix},$$

in the proof of opening, the prover commits to a short masking vector $\bar{\mathbf{y}}$ from a discrete Gaussian distribution, sends $\bar{\mathbf{w}} = \mathbf{B}_0\bar{\mathbf{y}}$, the verifier challenges with $\mathbf{c} \leftarrow C$, the prover replies with $\bar{\mathbf{z}} = \bar{\mathbf{y}} + \mathbf{c} \cdot \bar{\mathbf{r}}$, and rejection sampling is used to make $\bar{\mathbf{z}}$ statistically close to a discrete Gaussian sample centered at $\mathbf{0}$. The verifier checks $\|\bar{\mathbf{z}}\|_2 \le \beta$ and $\mathbf{B}_0\bar{\mathbf{z}} = \bar{\mathbf{w}} + \mathbf{c} \cdot \bar{\mathbf{t}}_0$.

If C has this property, the best [BDL+18] can extract from the proof of opening is the relaxed opening eq. (4.6).

- On one hand, one cannot guarantee that $\mathbf{f}^{-1} \cdot \bar{\mathbf{r}}$ is short after cancelling out $\mathbf{f}$ from $\bar{\mathbf{r}}$, which can be arbitrarily long.
- On the other hand, the binding property should prevent a single commitment from admitting relaxed openings to distinct messages; otherwise, the opening proof would be meaningless.

The binding reduction to the SIS is achieved by the pairwise difference being invertible, see eq. (4.10).

[23] A bit flawed construction, elaborate later.

Hence, the best [BDL+18] can do is the relaxed opening, and uniquely interpolating $\overline{\mathbf{r}}^*$ and $\overline{\mathbf{y}}^*$ out of a pair of forked accepting transcripts. But the scheme cannot guarantee the shortness of $\overline{\mathbf{r}}^*$ and $\overline{\mathbf{y}}^*$.

In [ALS20], the challenge set is not pairwise difference invertible, so an extractor can obtain a "relaxed opening" from a pair of forked accepting transcripts, but binding is not immediately reducible to SIS, and we cannot uniquely interpolate the prover's replies $\overline{\mathbf{z}}$ and $\overline{\mathbf{z}}'$ to obtain $\overline{\mathbf{y}}^*$ and $\overline{\mathbf{r}}^*$, from only a pair of forked accepting transcripts

$$\overline{\mathbf{z}} = \overline{\mathbf{y}}^* + \mathbf{c} \cdot \overline{\mathbf{r}}^*, \quad \overline{\mathbf{z}}' = \overline{\mathbf{y}}^* + \mathbf{c}' \cdot \overline{\mathbf{r}}^*. \tag{4.20}$$

But we can restore the interpolation for $\overline{\mathbf{y}}^*$ and $\overline{\mathbf{r}}^*$, by piecing together local interpolations over each prime ideal splitting R_q. More specifically, though $\Delta\mathbf{c} = \mathbf{c} - \mathbf{c}'$ is not invertible, not all of the CRT slots are zeroed out. Suppose that

$$R_q = \mathbb{Z}_q[X]/\left(X^d + 1\right) \cong \mathbb{Z}_q[X]/(\varphi_1) \times \cdots \times \mathbb{Z}_q[X]/(\varphi_\ell)\,,$$

where $\varphi_1, \ldots, \varphi_\ell$ are irreducible polynomial factors, and supposing $\Delta\mathbf{c} \not\equiv 0 \mod \varphi_i$, then interpolation over the i^{th} CRT slot gives

$$\overline{\mathbf{z}}_i \equiv \overline{\mathbf{y}}_i^* + \mathbf{c}_i \cdot \overline{\mathbf{r}}_i^* \mod \varphi_i, \quad \overline{\mathbf{z}}_i' \equiv \overline{\mathbf{y}}_i^* + \mathbf{c}_i' \cdot \overline{\mathbf{r}}_i^* \mod \varphi_i,$$

and by piecing together the interpolated $\overline{\mathbf{r}}_i^*$ and $\overline{\mathbf{y}}_i^*$ over each CRT slot $i \in [1, \ell]$, we have a pair $(\overline{\mathbf{r}}^*, \overline{\mathbf{y}}^*)$.

Moreover, we will show in the following, for this pair of global interpolations $(\overline{\mathbf{y}}^*, \overline{\mathbf{r}}^*)$ pieced out CRT slot-wise, on every accepting transcript prefixed by $\overline{\mathbf{w}}$, the prover must reply in the degree-1 affine form in eq. (4.20).

Lemma 4.12.9. *If we have, for $i \in [1, \ell]$, a pair of accepting transcripts $(\overline{\mathbf{w}}, \mathbf{c}_i, \overline{\mathbf{z}}_i)$ and $(\overline{\mathbf{w}}, \mathbf{c}_i', \overline{\mathbf{z}}_i')$ for the same commitment, with $\Delta\mathbf{c}_i := \mathbf{c}_i - \mathbf{c}_i' \not\equiv 0 \mod \varphi_i$, then every accepting transcript for that commitment with the same $\overline{\mathbf{w}}$ must have that $\overline{\mathbf{z}} = \overline{\mathbf{y}}^* + \mathbf{c} \cdot \overline{\mathbf{r}}^*$ where $(\overline{\mathbf{y}}^*, \overline{\mathbf{r}}^*)$ are fixed independently of $\mathbf{c}$, or otherwise we have a $\mathsf{MSIS}_{\mu,8\kappa\beta}$ solution for $\mathbf{B}_0$, where κ is a bound for the ℓ_1-norm of the challenges. Moreover, $\mathbf{B}_0\overline{\mathbf{y}}^* = \overline{\mathbf{w}}$ and $\mathbf{B}_0\overline{\mathbf{r}}^* = \bar{\mathbf{t}}_0$.*

Proof. Let $\overline{\mathbf{y}}^{*'}$ be such that $\overline{\mathbf{z}} = \overline{\mathbf{y}}^{*'} + \mathbf{c} \cdot \overline{\mathbf{r}}^*$. Then fix an arbitrary $i \in [1, \ell]$,

$$\mathbf{B}_0\left(\overline{\mathbf{z}}_i - \overline{\mathbf{z}}_i'\right) = \Delta\mathbf{c}_i \cdot \bar{\mathbf{t}}_0, \quad \mathbf{B}_0\left(\overline{\mathbf{z}} - \overline{\mathbf{z}}_i\right) = (\mathbf{c} - \mathbf{c}_i)\,\bar{\mathbf{t}}_0.$$

Cross-multiplying $(\mathbf{c} - \mathbf{c}_i)$ and $\Delta\mathbf{c}_i$, then subtracting them, should yield either a $\mathsf{MSIS}_{\mu,8\kappa\beta}$ solution for $\mathbf{B}_0$, where the norm bound follows from

$$\|\Delta\mathbf{c}_i\|_1 \cdot \|\overline{\mathbf{z}} - \overline{\mathbf{z}}_i\|_2 + \|\mathbf{c} - \mathbf{c}_i\|_1 \cdot \left\|\overline{\mathbf{z}}_i - \overline{\mathbf{z}}_i'\right\|_2 \leq (2\kappa) \cdot (2\beta) + (2\kappa) \cdot (2\beta) = 8\kappa\beta,$$

or it yields

$$(\mathbf{c} - \mathbf{c}_i) \cdot \left(\overline{\mathbf{z}}_i - \overline{\mathbf{z}}_i'\right) = \Delta\mathbf{c}_i \cdot (\overline{\mathbf{z}} - \overline{\mathbf{z}}_i)\,.$$

Since $\Delta\mathbf{c}_i \not\equiv 0 \mod \varphi_i$, then

$$(\mathbf{c} - \mathbf{c}_i) \cdot \Delta\mathbf{c}_i \cdot \overline{\mathbf{r}}^* \equiv \Delta\mathbf{c}_i \cdot \left(\overline{\mathbf{y}}^{*'} - \overline{\mathbf{y}}^* + (\mathbf{c} - \mathbf{c}_i) \cdot \overline{\mathbf{r}}^*\right) \mod \varphi_i,$$

namely $\Delta\mathbf{c}_i \cdot \left(\overline{\mathbf{y}}^{*'} - \overline{\mathbf{y}}^*\right) \equiv 0 \mod \varphi_i$. Hence, $\overline{\mathbf{y}}^{*'} \equiv \overline{\mathbf{y}}^* \mod \varphi_i$. If it holds for all $i \in [1, \ell]$, then $\overline{\mathbf{y}}^{*'} = \overline{\mathbf{y}}^*$.

The statement for $\mathbf{B}_0\overline{\mathbf{y}}^* = \overline{\mathbf{w}}$ and $\mathbf{B}_0\overline{\mathbf{r}}^* = \bar{\mathbf{t}}_0$ is immediate by CRT slot-wise local interpolation. □

Lemma 4.12.9 is helpful, as it explains that the prover has to reply in the degree-1 form w.r.t. the challenge, and it gives rise to the following augmentation of the relaxed opening, allowing for binding reduction to SIS even if we don't rely on the pairwise difference invertibility property.

Definition 4.12.10 (Weak Opening)**.** A weak opening for the commitment $\bar{\mathbf{t}}^\mathsf{T} = \left[\bar{\mathbf{t}}_0^\mathsf{T} \mid \mathbf{t}_1\right]$ consists of $(\Delta\mathbf{c}_i)_{i\in[1,\ell]} \in R_q^\ell$, $\overline{\mathbf{r}}^* \in R_q^N$ and $\mathbf{m}^* \in R_q$ such that for each $\Delta\mathbf{c}_i \in R_q$, $\Delta\mathbf{c}_i \not\equiv 0 \mod (\varphi_i)$, $\|\Delta\mathbf{c}_i\|_1 \leq 2\kappa$, $\|\Delta\mathbf{c}_i \cdot \overline{\mathbf{r}}^*\|_2 \leq 2\beta$, and

$$\mathbf{B}_0\overline{\mathbf{r}}^* = \bar{\mathbf{t}}_0, \quad \langle\overline{\mathbf{b}}_1, \overline{\mathbf{r}}^*\rangle + \mathbf{m}^* = \mathbf{t}_1.$$

The key point is to supply ℓ pairwise differences of challenges, such that each CRT slot is nonzero, and eventually this allows for binding reduction to SIS, which shows as follows.

Lemma 4.12.11. *The commitment scheme is binding w.r.t. the weak openings if* $\mathsf{MSIS}_{\mu,8\kappa\beta}$ *is hard, namely, valid weak openings*

$$(\{\Delta\mathbf{c}_i\}, \overline{\mathbf{r}}^*, \mathbf{m}^*), \quad \left(\{\Delta\mathbf{c}'_i\}, \overline{\mathbf{r}}^{*'}, \mathbf{m}^{*'}\right)$$

with $\mathbf{m}^* \neq \mathbf{m}^{*'}$ *give an* $\mathsf{MSIS}_{\mu,8\kappa\beta}$ *solution for* $\mathbf{B}_0$.

Proof. If $\mathbf{m}^* \neq \mathbf{m}^{*'}$, then $\overline{\mathbf{r}}^* \neq \overline{\mathbf{r}}^{*'}$, and there must be at least one $i \in [1, \ell]$ such that $\overline{\mathbf{r}}^* - \overline{\mathbf{r}}^{*'} \not\equiv \overline{0} \mod (\varphi_i)$. Then

$$\Delta\mathbf{c}_i \cdot \left(\Delta\mathbf{c}'_i \cdot \overline{\mathbf{r}}^*\right) - \Delta\mathbf{c}'_i \cdot \left(\Delta\mathbf{c}_i \cdot \overline{\mathbf{r}}^{*'}\right) = \Delta\mathbf{c}_i \cdot \Delta\mathbf{c}'_i \cdot \left(\overline{\mathbf{r}}^* - \overline{\mathbf{r}}^{*'}\right) \not\equiv \overline{0} \mod \varphi_i, \tag{4.21}$$

which gives a $\mathsf{MSIS}_{\mu,8\kappa\beta}$ solution for $\mathbf{B}_0$. The $8\kappa\beta$ derives from $\|\Delta\mathbf{c}_i\|_1 \leq 2\kappa$ and $\|\Delta\mathbf{c}_i \cdot \overline{\mathbf{r}}^*\|_2 \leq 2\beta$ in definition 4.12.10. □

Looking back, we can reverse engineer the motivation from [BDL+18] as follows. Starting from a challenge set that is not globally pairwise difference invertible, then to obtain an augmentation of the relaxed opening, we should collect ℓ pairs of forked accepting transcripts starting from $\overline{\mathbf{w}}$ the commitment to the masking vector, such that each CRT slot is once nonzero, and we piece out a global interpolation over R_q by local interpolations over $R_q/(\varphi_i)$. What is different from [BDL+18] in [ALS20] is that we need to interpolate many times, while in [BDL+18] one interpolation is sufficient, as C in [BDL+18] is globally pairwise difference invertible. To ensure that the linearity of the prover reply $\mathbf{z}$ against the short challenge set elements is still preserved, the lemma 4.12.9 shows that the affine form is respected for all accepting transcripts beginning with $\overline{\mathbf{w}}$ for the commitment. A good trick in lemma 4.12.9 is, on a reply that does not have the affine form, move all the eccentric parts down to the degree 0 constant $\overline{\mathbf{y}}^{*'}$, breaking SIS.

With lemma 4.12.9, the form of definition 4.12.10 makes sense, as lemma 4.12.9 rules out the non-linearity in each "relaxed opening" in the weak opening, so $\overline{\mathbf{y}}^*$ is always cancelled out. Finally, the lemma 4.12.11 concludes that the weak opening binds against the commitment by SIS hardness.

All of this effort is necessary because C is not pairwise difference invertible, so extraction cannot be done by one pair of forked accepting transcripts, but rather interpolations over multiple pairs. The weak opening enables binding reduction to SIS by providing $\Delta\mathbf{c}_i$ whose i^{th} CRT slot is non-zero, and lemma 4.12.9 rules out all other forms of extracted openings by showing the uniqueness of the line.

The analysis in [ALS20] has some interesting reasoning for a larger challenge set: Suppose the challenge set C is globally pairwise difference invertible, and R_q is full splitting. Fix a CRT slot, there are q distinct values on that CRT slot, and for the challenge set C, by the pigeonhole principle, there are at most q challenges in C, or otherwise there exists collision on the CRT slot, then the pairwise difference is not always invertible. Hence, the soundness is at best $1/q$. Even when R_q splits into a few low-degree factors, we cannot quantify the exact size of such globally pairwise difference invertible C.

Now I should explain why section 4.12 does not work out of the box. Indeed, if $\mathbf{c}, \mathbf{c}' \xleftarrow{\text{R}} \{0,1\}^d$, then the coefficients of $\mathbf{c} - \mathbf{c}'$ are i.i.d. with $P(\pm 1) = 1/4$ and $P(0) = 1/2$. Yet it does not mean the anti-concentration bound p in eq. (4.18) can be immediately applied here, whose coefficient 0 probability is $1/2$, such that the knowledge soundness is $p^{d/\ell}$. Looking closely, let X, X' be the random variables for the j^{th} coefficient in the i^{th} CRT slot of $\mathbf{c}, \mathbf{c}'$, then

$$\sum_{x \in \mathbb{Z}_q} \Pr[X = x \wedge X' = x] = \sum_{x \in \mathbb{Z}_q} \Pr[X = x]^2 = \Pr[X = X'] = p.$$

The last equality holds from eq. (4.18), where the inequality is equality when $x = 0$ for the distribution of $X - X'$. Hence

$$\max_{x \in \mathbb{Z}_q} \Pr[X = x]^2 \leq \sum_{x \in \mathbb{Z}_q} \Pr[X = x]^2 = p,$$

and hence the knowledge soundness induced by the challenge set over $\{0,1\}^d$ is no longer $p^{d/\ell}$ but rather $p^{d/(2\ell)}$, as the maximum point probability of a coefficient in the i^{th} CRT slot of $\mathbf{c}$ is at most the square root of the anti-concentration bound formed for the $\{-1,0,1\}$ symmetric distribution.

BDLOP with Ring Automorphism. Now we construct a BDLOP in the ring automorphism settings. By changing the challenge set distribution, say [LS18], to the distribution defined previously over $\{-1,0,1\}^d$, where distribution for each coefficient is i.i.d., $P(\pm 1) = 1/4$ and $P(0) = 1/2$, then let p be the maximum probability in eq. (4.18), then the knowledge soundness should be $p^{d/\ell}$. More specifically, if the prover convinces the verifier with probability $\varepsilon > p^{d/\ell}$, we obtain the first accepting transcript on challenge $\mathbf{c}$ in expected $1/\varepsilon$ time, then for another challenge $\mathbf{c}'$,

$$\begin{aligned}\varepsilon = \Pr[\mathcal{P} \text{ succeeds}] &= \Pr\left[\mathcal{P} \text{ succeeds} \mid \mathbf{c}' \not\equiv \mathbf{c} \mod (\varphi_i)\right] \cdot \Pr\left[\mathbf{c}' \not\equiv \mathbf{c} \mod (\varphi_i)\right] \\ &+ \Pr\left[\mathcal{P} \text{ succeeds} \mid \mathbf{c}' \equiv \mathbf{c} \mod (\varphi_i)\right] \cdot \Pr\left[\mathbf{c}' \equiv \mathbf{c} \mod (\varphi_i)\right] \\ &\leq \Pr\left[\mathcal{P} \text{ succeeds} \mid \mathbf{c}' \not\equiv \mathbf{c} \mod (\varphi_i)\right] + \Pr\left[\mathbf{c}' \equiv \mathbf{c} \mod (\varphi_i)\right],\end{aligned}$$

the convincing probability on $\mathbf{c}' \not\equiv \mathbf{c} \mod \varphi_i$ is at least $\varepsilon - p^{d/\ell}$, and the extractor runs in expected polynomial time if $\varepsilon - p^{d/\ell}$ is $\mathsf{poly}^{-1}(\lambda)$.

If $p^{d/\ell}$ is not negligible, we can run k copies of the protocol in parallel, and reduce the soundness down to $p^{kd/\ell}$. To achieve this in a 3-move Sigma protocol, they challenge with $\mathbf{c}$ and its automorphisms. I suppose the observation stems from: for a $\mathbf{c}$ sampled over $\{-1,0,1\}^d$, a coefficient in a coarse ideal $X^{kd/\ell} - \xi^{jk}$ takes on a value with probability at most p by eq. (4.18) [24]. Recall that $\sigma = \sigma_{2\ell/k+1}$ stabilizes the ideal $(X^{kd/\ell} - \xi^{jk})$, such that

$$\left(X^{kd/\ell} - \xi^{jk}\right) = \prod_{i \in [0,k-1]} \sigma^i \left(X^{d/\ell} - \xi^j\right) = \prod_{i \in \langle 2\ell/k+1 \rangle / \langle 2\ell+1 \rangle} \left(X^{d/\ell} - \xi^{ij}\right) \tag{4.22}$$

by eqs. (4.15) and (4.16), such that if $\mathbf{c} - \mathbf{c}' \not\equiv 0 \mod (X^{kd/\ell} - \xi^{jk})$, then any of $\sigma^i(\mathbf{c} - \mathbf{c}') \not\equiv 0 \mod (X^{kd/\ell} - \xi^{jk})$, and any of the CRT slot under the coarse CRT factor is at least once non-zero in the k automorphisms.

The construction is then immediate. Sample k random masking elements and commit to them, then the verifier replies with the challenge $\mathbf{c}$, and the prover replies with $\bar{\mathbf{z}}_i = \bar{\mathbf{y}}_i + \sigma^i(\mathbf{c}) \cdot \bar{\mathbf{r}}$ under rejection sampling. If the prover runs in unit time and convinces the verifier with probability $\varepsilon > p^{kd/\ell}$, then the extraction takes

$$\frac{1}{\varepsilon} + \frac{\ell}{k} \cdot \frac{1}{\varepsilon - p^{kd/\ell}}$$

expected time, as sampling

- The first accepting transcript with challenge $\mathbf{c}$ takes expected $1/\varepsilon$ time.
- An accepting transcript with challenge $\mathbf{c}_j$ with $\mathbf{c} - \mathbf{c}_j \not\equiv 0 \mod (X^{kd/\ell} - \xi^{jk})$ takes expected $(\varepsilon - p^{kd/\ell})^{-1}$ time.

Strawman Product Proof. We begin with a rudimentary product proof using random masking elements. For

$$\begin{bmatrix} \bar{\mathbf{t}}_0 \\ \mathbf{t}_1 \\ \mathbf{t}_2 \\ \mathbf{t}_3 \end{bmatrix} = \mathbf{B}\bar{\mathbf{r}} + \begin{bmatrix} \bar{\mathbf{0}} \\ \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix}, \tag{4.23}$$

the relation to be proved is $\mathbf{m}_1 \cdot \mathbf{m}_2 = \mathbf{m}_3$. The strawman solution in [ALS20] is by

$$\begin{bmatrix} \bar{\mathbf{t}}'_0 \\ \mathbf{t}'_1 \\ \mathbf{t}'_2 \\ \mathbf{t}'_3 \\ \mathbf{t}'_4 \\ \mathbf{t}'_5 \end{bmatrix} = \mathbf{B}'\bar{\mathbf{r}}' + \begin{bmatrix} \bar{\mathbf{0}} \\ \mathbf{a}_1 \\ \mathbf{a}_2 \\ \mathbf{a}_3 \\ \mathbf{a}_1\mathbf{m}_2 + \mathbf{a}_2\mathbf{m}_1 - \mathbf{a}_3 \\ \mathbf{a}_1\mathbf{a}_2 \end{bmatrix}, \tag{4.24}$$

[24] Here the p is not defined for $X^{d/\ell} - \xi^j$ but for $X^{kd/\ell} - \xi^{kj}$.

that is committing to random masking polynomials $\mathbf{a}_i$ and garbage terms. The invariant is the masked relation

$$(\mathbf{a}_1 + \mathbf{x}\mathbf{m}_1)(\mathbf{a}_2 + \mathbf{x}\mathbf{m}_2) - \mathbf{x}(\mathbf{a}_3 + \mathbf{x}\mathbf{m}_3) = \mathbf{a}_1\mathbf{a}_2 + \mathbf{x}(\mathbf{m}_1\mathbf{a}_2 + \mathbf{m}_2\mathbf{a}_1 - \mathbf{a}_3) + \mathbf{x}^2(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3) \tag{4.25}$$

after a challenge $\mathbf{x}$ from the verifier. More specifically, for the masked openings

$$\mathbf{f}_i = \mathbf{a}_i + \mathbf{x}\mathbf{m}_i,$$

we can prove that $\mathbf{t}'_i + \mathbf{x}\mathbf{t}_i$ commits to $\mathbf{f}_i$. The verifier can then check eq. (4.25) by seeing if $\mathbf{t}'_5 + \mathbf{x}\mathbf{t}'_4$ commits to $\mathbf{f}_1\mathbf{f}_2 - \mathbf{x}\mathbf{f}_3$

$$\mathbf{f}_1\mathbf{f}_2 - \mathbf{x}\mathbf{f}_3 = \mathbf{a}_1\mathbf{a}_2 + \mathbf{x}(\mathbf{m}_1\mathbf{a}_2 + \mathbf{m}_2\mathbf{a}_1 - \mathbf{a}_3) + \mathbf{x}^2(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3),$$

such that the quadratic relation is checked linearly over committed elements, and the desired product relation is the vanishing of its degree-2 coefficient. Essentially, the masking technique is used in [BLS19, YAZ+19], where a product proof appears inside a shortness proof.

We now put things together conceptually as follows. Suppose the prover knows an opening $(\mathbf{m}_1, \mathbf{m}_2, \mathbf{m}_3, \bar{\mathbf{r}})$ for the commitment in eq. (4.23):

- The prover commits to the masked terms and garbage terms in eq. (4.24).
- The verifier replies with a uniformly random challenge $\mathbf{x}$.
- The prover replies with masked openings $\mathbf{f}_i = \mathbf{a}_i + \mathbf{x}\mathbf{m}_i$.

  Moreover, it commits to the masking randomness $\bar{\mathbf{y}}$ and $\bar{\mathbf{y}}'$ by

  $$\bar{\mathbf{w}} = \mathbf{B}_0\bar{\mathbf{y}}, \quad \bar{\mathbf{w}}' = \mathbf{B}'_0\bar{\mathbf{y}}',$$

  and sends over $\mathbf{v}_i = \mathbf{x} \cdot \langle \bar{\mathbf{b}}_i, \bar{\mathbf{y}} \rangle + \langle \bar{\mathbf{b}}'_i, \bar{\mathbf{y}}' \rangle$, and $\mathbf{u} = \mathbf{x} \cdot \langle \bar{\mathbf{b}}'_4, \bar{\mathbf{y}}' \rangle + \langle \bar{\mathbf{b}}'_5, \bar{\mathbf{y}}' \rangle$ [25].
- The verifier replies with a short challenge $\mathbf{c}$.
- The prover runs rejection sampling, and replies

  $$\bar{\mathbf{z}} = \bar{\mathbf{y}} + \mathbf{c} \cdot \bar{\mathbf{r}}, \quad \bar{\mathbf{z}}' = \bar{\mathbf{y}}' + \mathbf{c} \cdot \bar{\mathbf{r}}'.$$

- The verifier checks $\bar{\mathbf{z}}$ and $\bar{\mathbf{z}}'$ are short,

  $$\begin{aligned} \mathbf{B}_0\bar{\mathbf{z}} &= \bar{\mathbf{w}} + \mathbf{c} \cdot \bar{\mathbf{t}}_0, \\ \mathbf{B}'_0\bar{\mathbf{z}}' &= \bar{\mathbf{w}}' + \mathbf{c} \cdot \bar{\mathbf{t}}'_0, \\ \mathbf{v}_i + \mathbf{c}\left(\mathbf{t}'_i + \mathbf{x}\mathbf{t}_i\right) &= \mathbf{x} \cdot \langle \bar{\mathbf{b}}_i, \bar{\mathbf{z}} \rangle + \langle \bar{\mathbf{b}}'_i, \bar{\mathbf{z}}' \rangle + \mathbf{c}\mathbf{f}_i, \end{aligned}$$

  and checks the quadratic masked relation eq. (4.25) by ensuring $\gamma = \mathbf{x}\mathbf{t}'_4 + \mathbf{t}'_5 - (\mathbf{f}_1\mathbf{f}_2 - \mathbf{x}\mathbf{f}_3)$ commits to 0 by

  $$\mathbf{u} + \mathbf{c}\gamma = \mathbf{x} \cdot \langle \bar{\mathbf{b}}'_4, \bar{\mathbf{z}}' \rangle + \langle \bar{\mathbf{b}}'_5, \bar{\mathbf{z}}' \rangle.$$

Regardless of the proof size, there is a catch regarding the knowledge soundness: $\mathbf{x}$ should be pairwise difference invertible, which falls back the prior discussion of the challenge set construction [26], and reusing the same masks for different $\mathbf{x}$ breaks zero knowledge.

[25] Recall the proving linear relation over elements in BDLOP abstracted in note for [ALS20].

[26] Suppose the polynomial ring splits completely. The knowledge soundness will be at best $1/q$ by the pigeonhole principle.

Computationally Indistinguishable Masked Opening. An observation to the prior product proof from masked opening is, the values $\mathbf{f}_i = \mathbf{a}_i + \mathbf{x}\mathbf{m}_i$ provide unconditional statistical hiding. We have

$$\mathbf{f}'_i = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{z}} \rangle - \mathbf{c}\mathbf{t}_i = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{y}} \rangle - \mathbf{c}\mathbf{m}_i,$$

such that $\langle \overline{\mathbf{b}}_i, \overline{\mathbf{y}} \rangle$ is the computationally indistinguishable mask, and the masked relation in eq. (4.25) converts to

$$\mathbf{f}'_1\mathbf{f}'_2 + \mathbf{c}\mathbf{f}'_3 = \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}} \rangle \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle + \mathbf{c}\left(\langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}} \rangle - \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}} \rangle \mathbf{m}_2 - \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle \mathbf{m}_1 \right) + \mathbf{c}^2 \left(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3 \right). \tag{4.26}$$

With one garbage term $\langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}} \rangle - \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}} \rangle \mathbf{m}_2 - \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle \mathbf{m}_1$ to commit to (and sending out claim $\langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}} \rangle \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle + \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}} \rangle$), the new construction can be more efficient. Hence, the masked openings are inherent in the Σ-protocol.

The protocol so far works as follows:

- The prover commits to the masking randomness $\overline{\mathbf{w}} = \mathbf{B}_0 \overline{\mathbf{y}}$, the message and the garbage term

$$\begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_1 \\ \mathbf{t}_2 \\ \mathbf{t}_3 \\ \mathbf{t}_4 \end{bmatrix} = \mathbf{B}\overline{\mathbf{r}} + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \\ \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}} \rangle - \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}} \rangle \mathbf{m}_2 - \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle \mathbf{m}_1 \end{bmatrix},$$

  and derives the claim $\mathbf{v} = \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}} \rangle \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle + \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}} \rangle$.
- The verifier challenges with $\mathbf{c}$.
- The prover runs rejection sampling and replies with $\overline{\mathbf{z}} \leftarrow \overline{\mathbf{y}} + \mathbf{c}\overline{\mathbf{r}}$.
- The verifier checks that $\overline{\mathbf{z}}$ is short and that

$$\begin{aligned} \overline{\mathbf{w}} + \mathbf{c}\overline{\mathbf{t}}_0 &= \mathbf{B}_0 \overline{\mathbf{z}}, \\ \mathbf{f}_1\mathbf{f}_2 + \mathbf{c}\mathbf{f}_3 + \mathbf{f}_4 &= \mathbf{v}, \end{aligned}$$

  where $\mathbf{f}_i = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{z}} \rangle - \mathbf{c}\mathbf{t}_i$ is the new computationally indistinguishable masked opening.

Soundness Amplification from Automorphism Repetition, Recap. After the strawman construction of the product proof, we pointed out the knowledge soundness issue from uniformly random $\mathbf{x}$, and the same issue remains in the computationally indistinguishable masked opening, which is challenged by $\mathbf{c}$.

The prior structured parallel repetition using automorphisms becomes handy, and it is a good time to recap here. Before the "automorphism repetition", the malicious prover can cheat by, say, if $\overline{\mathbf{r}}$ is correct except at the i^{th} CRT slot, then there must exist $\overline{\boldsymbol{\delta}}$ and $\overline{\Delta} = \mathbf{B}_0 \overline{\boldsymbol{\delta}}$ satisfying

$$\overline{\boldsymbol{\delta}} \not\equiv \overline{\mathbf{0}} \mod \left(X^{d/\ell} - \xi^i \right), \quad \overline{\boldsymbol{\delta}} \equiv \overline{\mathbf{0}} \mod \left(X^{d/\ell} - \xi^j \right)$$

for $j \neq i$, and $\overline{\Delta}$ is non-zero only at the i^{th} CRT slot, such that $\mathbf{B}_0 \overline{\mathbf{r}}_0 = \overline{\mathbf{t}}_0$ for $\overline{\mathbf{r}} = \overline{\mathbf{r}}_0 + \overline{\boldsymbol{\delta}}$, and hence $\mathbf{B}_0 \overline{\mathbf{r}} = \overline{\mathbf{t}}_0 + \overline{\Delta}$.

Supposing the prover guessed the challenge's i^{th} CRT lane is $\mathbf{s} \in \mathbb{Z}_q[X]/\left(X^{d/\ell} - \xi^i\right)$, let $\mathbf{c}_s$ be the lifted R_q element of $\mathbf{s}$ such that it is only non-zero at the i^{th} CRT lane, and the forging follows:

- The prover samples $\overline{\mathbf{y}}$ the masking randomness, and sends out $\overline{\mathbf{w}} \leftarrow \mathbf{B}_0 \overline{\mathbf{y}} + \mathbf{c}_s \overline{\Delta}$.
- The verifier challenges with an element $\mathbf{c}$ such that $\mathbf{c} \equiv \mathbf{c}_s \mod \left(X^{d/\ell} - \xi^i\right)$.
- The prover runs rejection sampling and replies with $\overline{\mathbf{z}} \leftarrow \overline{\mathbf{y}} + \mathbf{c}\overline{\mathbf{r}}$.
- The forged proof still goes through by

$$\overline{\mathbf{w}} + \mathbf{c}\overline{\mathbf{t}}_0 = \mathbf{B}_0 \overline{\mathbf{y}} + \mathbf{c}_s \overline{\Delta} + \mathbf{c}\left(\mathbf{B}_0 \overline{\mathbf{r}} - \overline{\Delta} \right) = \mathbf{B}_0 \left(\overline{\mathbf{y}} + \mathbf{c}\overline{\mathbf{r}} \right) + \overline{\Delta}\left(\mathbf{c}_s - \mathbf{c} \right) = \mathbf{B}_0 \overline{\mathbf{z}}.$$

The trick is to compensate by the right amount of $\overline{\Delta}$ if the prover guesses the relevant part of $\mathbf{c}$ correctly, as $\bar{\mathbf{t}}_0$ has a deficit of $\overline{\Delta}$ relative to $\mathbf{B}_0\bar{\mathbf{r}}$.

Moving to the "automorphism repetition", for a CRT slot determined by $\left(X^{d/\ell} - \xi^j\right)$, instead of being tested by a single challenge $\mathbf{c} \mod \left(X^{d/\ell} - \xi^j\right)$, it is tested by k challenges deriving from

$$\mathbf{c} \mod \left(X^{kd/\ell} - \xi^{kj}\right) \cong \left(\mathbf{c} \mod \left(X^{d/\ell} - \xi^j\right), \sigma(\mathbf{c}) \mod \left(X^{d/\ell} - \xi^j\right), \ldots, \sigma^{k-1}(\mathbf{c}) \mod \left(X^{d/\ell} - \xi^j\right)\right). \quad (4.27)$$

Suppose any one of the small CRT slots under $\left(X^{kd/\ell} - \xi^{kj}\right)$ [27] is incorrect in $\bar{\mathbf{r}}$, the malicious prover needs to guess $\mathbf{c} \mod \left(X^{kd/\ell} - \xi^{kj}\right)$, and forge across the k repetitions at the corresponding slots with the same algorithm. The guessing probability is at most the knowledge-soundness bound $p^{kd/\ell}$.

Automorphism Repetition in Product Proof. We can view the proof of $\mathbf{m}_1\mathbf{m}_2 = \mathbf{m}_3$ as proving

$$\mathbf{m}_1\mathbf{m}_2 \equiv \mathbf{m}_3 \mod \left(X^{d/\ell} - \xi^j\right)$$

for each $j \in \mathbb{Z}_{2d}^*/\langle 2\ell + 1\rangle$, simultaneously for ℓ CRT slots. With the same motivation, the vanilla product proof testing eq. (4.26) only suffices for non-negligible soundness, as a CRT slot of the message is tested with a single CRT slot of the challenge, instead of being tested k times by eq. (4.27).

To build up the intuition, consider the vanilla product proof for eq. (4.26), and suppose $\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3$ is nonzero only at the j^{th} CRT slot. For simplicity, we write $\mathbf{a}_i = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{y}}\rangle$, and the malicious prover guesses by adding an element $\mathbf{s}$ to the claim $\mathbf{v}$, such that $\mathbf{s}$ is only nonzero at the j^{th} CRT slot, then

$$\mathbf{f}_1\mathbf{f}_2 + \mathbf{c}\mathbf{f}_3 + \mathbf{f}_4 - \mathbf{v} = (\mathbf{a}_1\mathbf{a}_2 + \mathbf{a}_4 - \mathbf{v}) + \mathbf{c}\,(\mathbf{a}_3 - \mathbf{a}_1\mathbf{m}_2 - \mathbf{a}_2\mathbf{m}_1 - \mathbf{m}_4) + \mathbf{c}^2\,(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3) \equiv \mathbf{0} \mod \left(X^{d/\ell} - \xi^j\right) \quad (4.28)$$

has at most 2 roots for the j^{th} CRT slot, and thus the guessing probability is at most $2p^{d/\ell}$ by a union bound.

Now the genuine question is how to implement "automorphism repetition" for the product proof using eq. (4.26). Consider a product proof utilizing the "automorphism parallelism" that uses the challenge $\mathbf{c}$ and its automorphisms but separates the claims and commitments to the garbage terms as follows:

- The prover commits to the masking randomness $\overline{\mathbf{w}}_i = \mathbf{B}_0\overline{\mathbf{y}}_i$ for $i \in [0, k-1]$, the message commitment

$$\begin{bmatrix} \bar{\mathbf{t}}_0 \\ \mathbf{t}_1 \\ \mathbf{t}_2 \\ \mathbf{t}_3 \end{bmatrix} = \mathbf{B}\bar{\mathbf{r}} + \begin{bmatrix} \bar{\mathbf{0}} \\ \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix},$$

  and the garbage terms and the claims for $i \in [0, k-1]$ by

$$\mathbf{t}_{4,i} = \langle \overline{\mathbf{b}}_4, \bar{\mathbf{r}}\rangle + \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}}_i\rangle - \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}}_i\rangle\mathbf{m}_2 - \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}}_i\rangle\mathbf{m}_1,$$
$$\mathbf{v}_i = \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}}_i\rangle + \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}}_i\rangle\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}}_i\rangle.$$

- The verifier challenges with $\mathbf{c}$.
- The prover runs rejection sampling and replies with $\overline{\mathbf{z}}_i \leftarrow \overline{\mathbf{y}}_i + \sigma^i(\mathbf{c}) \cdot \bar{\mathbf{r}}$.
- The verifier checks that $\overline{\mathbf{z}}_i$ is short and that

$$\overline{\mathbf{w}}_i + \sigma^i(\mathbf{c}) \cdot \bar{\mathbf{t}}_0 = \mathbf{B}_0\overline{\mathbf{z}}_i,$$
$$\mathbf{f}_{1,i}\mathbf{f}_{2,i} + \sigma^i(\mathbf{c}) \cdot \mathbf{f}_{3,i} + \mathbf{f}_{4,i} = \mathbf{v}_i,$$

  where, for $a \in [1, 3]$,

$$\mathbf{f}_{a,j} = \langle \overline{\mathbf{b}}_a, \overline{\mathbf{z}}_j\rangle - \sigma^j(\mathbf{c}) \cdot \mathbf{t}_a, \quad \mathbf{f}_{4,j} = \langle \overline{\mathbf{b}}_4, \overline{\mathbf{z}}_j\rangle - \sigma^j(\mathbf{c}) \cdot \mathbf{t}_{4,j}.$$

[27] See eq. (4.22).

Again, let $\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3$ be nonzero only at one of the fine CRT slots under $\left(X^{kd/\ell} - \xi^{kj}\right)$. By eq. (4.28), for one forgery of the claim $\mathbf{v}$, the quadratic constraint admits at most 2 roots modulo the fine CRT slot, then by eq. (4.27), forging the claims $\mathbf{v}_0, \ldots, \mathbf{v}_{k-1}$ for the nonzero CRT slot gives at most 2^k roots modulo $\left(X^{kd/\ell} - \xi^{kj}\right)$, namely, in each repetition, a forgery on the CRT slot admits 2 roots for $\sigma^i(\mathbf{c}) \mod \left(X^{d/\ell} - \xi^j\right)$. Hence, at most 2^k values of $\mathbf{c} \mod \left(X^{kd/\ell} - \xi^{kj}\right)$ can make the k CRT slots across k repetitions vanish simultaneously.

Note that in the degree-2 setting, the malicious prover can take advantage of the higher degree: it does not have to guess the exact $\mathbf{c} \mod \left(X^{kd/\ell} - \xi^{kj}\right)$, but instead can guess among 2^k values. The resulting guessing probability is at most $2^k p^{kd/\ell}$.

Automorphism and Communication Complexity. The repetition over automorphisms provides a concrete benefit. In fact, for the prior construction utilizing the automorphism repetition, where $\{\mathbf{t}_{4,i}\}_{i\in[0,k-1]}$ commits to the garbage terms, and $\{\mathbf{v}_i\}_{i\in[0,k-1]}$ are the claims, using a set of independent challenges still suffices, and the independence makes it easier to reason about the soundness.

This is surely better than the strawman construction regarding the concrete proof size, yet the structural nature of the automorphism is absent. Recalling the automorphism in eq. (4.13), consider an inverse σ_i^{-1} such that

$$\sigma_i^{-1}\left(\sigma_i(\mathbf{f}_1) \cdot \sigma_i(\mathbf{f}_2)\right) = \sigma_i^{-1}\left(\sigma_i(\mathbf{f}_1 \cdot \mathbf{f}_2)\right) = \mathbf{f}_1 \cdot \mathbf{f}_2,$$

then for eq. (4.28) and eq. (4.26), we can "inverse rotate" from

$$\mathbf{f}_{1,i}\mathbf{f}_{2,i} + \sigma^i(\mathbf{c}) \cdot \mathbf{f}_{3,i} = \mathbf{a}_{1,i}\mathbf{a}_{2,i} + \sigma^i(\mathbf{c})\left(\mathbf{a}_{3,i} - \mathbf{a}_{1,i}\mathbf{m}_2 - \mathbf{a}_{2,i}\mathbf{m}_1\right) + \sigma^i(\mathbf{c})^2\left(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3\right)$$

to

$$\sigma^{-i}\left(\mathbf{f}_{1,i}\mathbf{f}_{2,i}\right) + \mathbf{c} \cdot \sigma^{-i}\left(\mathbf{f}_{3,i}\right) = \sigma^{-i}\left(\mathbf{a}_{1,i}\mathbf{a}_{2,i}\right) + \mathbf{c}\sigma^{-i}\left(\mathbf{a}_{3,i} - \mathbf{a}_{1,i}\mathbf{m}_2 - \mathbf{a}_{2,i}\mathbf{m}_1\right) + \mathbf{c}^2\sigma^{-i}\left(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3\right),$$

such that for $i \in [0, k-1]$, the $\mathbf{c}$ automorphisms are aligned back to the original $\mathbf{c}$. This structure allows us to combine k claims and garbage commitments into one through a random linear combination.

Let $\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_{k-1}$ be sampled uniformly over R_q. Then the garbage-term commitment and the claim are

$$\mathbf{t}_4 = \langle \overline{\mathbf{b}}_4, \overline{\mathbf{r}} \rangle + \sum_{i\in[0,k-1]} \boldsymbol{\alpha}_i \sigma^{-i}\left(\mathbf{a}_{3,i} - \mathbf{a}_{1,i}\mathbf{m}_2 - \mathbf{a}_{2,i}\mathbf{m}_1\right), \quad \mathbf{v} = \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}}_0 \rangle + \sum_{i\in[0,k-1]} \boldsymbol{\alpha}_i \sigma^{-i}\left(\mathbf{a}_{1,i}\mathbf{a}_{2,i}\right), \tag{4.29}$$

for the combined statement

$$\sum_{i\in[0,k-1]} \boldsymbol{\alpha}_i \cdot \sigma^{-i}\left(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3\right) = \mathbf{0}. \tag{4.30}$$

The final protocol follows:

- The prover commits to the masking randomness $\overline{\mathbf{w}}_i = \mathbf{B}_0\overline{\mathbf{y}}_i$ for $i \in [0, k-1]$, the message commitment

$$\begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_1 \\ \mathbf{t}_2 \\ \mathbf{t}_3 \end{bmatrix} = \mathbf{B}\overline{\mathbf{r}} + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{m}_1 \\ \mathbf{m}_2 \\ \mathbf{m}_3 \end{bmatrix}.$$

- The verifier replies with the uniform randomness $\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_{k-1}$.
- The prover derives $\mathbf{t}_4$ and $\mathbf{v}$ in eq. (4.29) and replies.
- The verifier challenges with $\mathbf{c}$.
- The prover runs rejection sampling and replies with $\overline{\mathbf{z}}_i \leftarrow \overline{\mathbf{y}}_i + \sigma^i(\mathbf{c}) \cdot \overline{\mathbf{r}}$.
- The verifier checks that $\overline{\mathbf{z}}_i$ is short, that $\overline{\mathbf{w}}_i + \sigma^i(\mathbf{c}) \cdot \overline{\mathbf{t}}_0 = \mathbf{B}_0\overline{\mathbf{z}}_i$, and

$$\mathbf{f}_{4,0} + \sum_{i\in[0,k-1]} \boldsymbol{\alpha}_i \sigma^{-i}\left(\mathbf{f}_{1,i}\mathbf{f}_{2,i} + \sigma^i(\mathbf{c}) \cdot \mathbf{f}_{3,i}\right) = \mathbf{f}_{4,0} + \sum_{i\in[0,k-1]} \boldsymbol{\alpha}_i \sigma^{-i}\left(\mathbf{f}_{1,i}\mathbf{f}_{2,i}\right) + \mathbf{c} \sum_{i\in[0,k-1]} \boldsymbol{\alpha}_i \sigma^{-i}\left(\mathbf{f}_{3,i}\right) = \mathbf{v},$$

  where $\mathbf{f}_{i,j} = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{z}}_j \rangle - \sigma^j(\mathbf{c}) \cdot \mathbf{t}_i$.

Now supposing a CRT slot is nonzero for $\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3$, then the automorphisms

$$\sigma^{-i}\left(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3\right) \mod \sigma^{-i}\left(X^{d/\ell} - \xi^j\right)$$

are all nonzero. For the combined statement eq. (4.30), we analyze the number of CRT slots under

$$\sum_{i\in[0,k-1]} \boldsymbol{\alpha}_i \cdot \sigma^{-i}\left(\mathbf{m}_1\mathbf{m}_2 - \mathbf{m}_3\right) \mod \left(X^{kd/\ell} - \xi^{kj}\right)$$

are nonzero, and analyze the guessing probability of a malicious prover.

- If the k CRT slots are all nonzero, then it follows exactly from my version of the product proof without utilizing automorphism well, namely the guessing probability is at most $2^k p^{kd/\ell}$, as $\mathbf{c} \mod \left(X^{kd/\ell} - \xi^{kj}\right)$ takes on at most 2^k values.
- If i of the k CRT slots are zero, then on a set of i slots being zeroed out by the randomness, $\mathbf{c} \mod \left(X^{kd/\ell} - \xi^{kj}\right)$ takes on at most $q^{id/\ell}2^{k-i}$ values.

Hence, the guessing probability is at most

$$\sum_{i\in[0,k]} \binom{k}{i} \left(\frac{1}{q^{d/\ell}}\right)^i \left(1 - \frac{1}{q^{d/\ell}}\right)^{k-i} 2^{k-i} q^{id/\ell} p^{kd/\ell} \le 3^k p^{kd/\ell}. \tag{4.31}$$

Moreover, to batch multiple product relations, it suffices to have

$$\sum_j \sum_{i\in[0,k-1]} \boldsymbol{\alpha}_{i,j} \cdot \sigma^{-i}\left(\mathbf{m}_{1,j}\mathbf{m}_{2,j} - \mathbf{m}_{3,j}\right) = \mathbf{0},$$

namely, to take a random linear combination with $\{\{\boldsymbol{\alpha}_{i,j}\}_{i\in[0,k-1]}\}_j$ over R_q.

The observation is, if there exists nonzero $\mathbf{m}_{1,j}\mathbf{m}_{2,j} - \mathbf{m}_{3,j}$, under the chosen coarse slot containing nonzero error, for each of the k CRT slots, random linear combination over independent $\boldsymbol{\alpha}_{i,j} \mod \varphi$ vanishes with probability at most $q^{-d/\ell}$ by Schwartz-Zippel, and the randomness of each slot is independent, so the soundness in eq. (4.31) does not change.

4.13 Practical Exact Proofs from Lattices

We now apply the product proof in [ALS20] from the computationally indistinguishable masked opening to [BLS19]. A baseline protocol for the exact proof follows:

- The prover sends out $\overline{w} \leftarrow A\hat{\mathbf{y}}$ and $\overline{\mathbf{w}} \leftarrow \mathbf{B}_0\overline{\mathbf{y}}$, where $\mathbf{y} \xleftarrow{R} R_q$ and $\overline{\mathbf{y}} \leftarrow D_\sigma^{Nd}$, and samples $\overline{\mathbf{r}} \leftarrow \chi^{Nd}$ to commit

$$\begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_1 \\ \mathbf{t}_2 \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_1^{\mathsf{T}} \\ \overline{\mathbf{b}}_2^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{r}} + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{y} \\ \mathbf{s} \end{bmatrix}.$$

- The verifier challenges by sending $c \xleftarrow{R} \mathbb{Z}_q$.
- The prover replies with $\mathbf{z} \leftarrow c \cdot \mathbf{s} + \mathbf{y}$, and starts Π_{lin} and the shortness check in [ALS20] by:
  - The prover derives $\overline{\mathbf{b}}_{f_0} := \overline{\mathbf{b}}_1 + c \cdot \overline{\mathbf{b}}_2$, replies $\mathbf{w}' \leftarrow \langle \overline{\mathbf{b}}_{f_0}, \overline{\mathbf{y}} \rangle$, then it commits

$$\begin{bmatrix} \mathbf{t}_3 \\ \mathbf{t}_4 \end{bmatrix} \leftarrow \begin{bmatrix} \overline{\mathbf{b}}_3^{\mathsf{T}} \\ \overline{\mathbf{b}}_4^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{r}} + \begin{bmatrix} \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}} \rangle - 3\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^2 \mathbf{s} \\ \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle \left(3\mathbf{s}^2 - \mathbf{1}\right) \end{bmatrix}, \tag{4.32}$$

    and replies $\mathbf{t}_3, \mathbf{t}_4$, and the claim $\mathbf{v} \leftarrow \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^3 + \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}} \rangle$.
  - The verifier samples $\mathbf{c}$ over the $\{-1, 0, 1\}$ challenge set C, with each coefficient has $P(\pm 1) = 1/4$, $P(0) = 1/2$.
  - The prover runs rejection sampling, and outputs $\overline{\mathbf{z}}' \leftarrow \overline{\mathbf{y}} + \mathbf{c} \cdot \overline{\mathbf{r}}$.
- The verifier checks by ensuring $\overline{\mathbf{z}}'$ being short, $A\hat{\mathbf{z}} = \overline{w} + c \cdot \overline{u}$,

$$\begin{bmatrix} \overline{\mathbf{w}} \\ \mathbf{w}' \end{bmatrix} + \mathbf{c} \cdot \begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_1 + c \cdot \mathbf{t}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_{f_0}^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{z}}' + \mathbf{c} \cdot \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{z} \end{bmatrix},$$

  and the shortness check from the product proof in [ALS20] by

$$\mathbf{f}_2 \left(\mathbf{f}_2 - \mathbf{c}\right) \left(\mathbf{f}_2 + \mathbf{c}\right) + \mathbf{f}_3 + \mathbf{c}\mathbf{f}_4 = \mathbf{v}, \tag{4.33}$$

  where $\mathbf{f}_i = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{z}}' \rangle - \mathbf{c}\mathbf{t}_i$ for $i \in [2, 4]$.

The 2 garbage terms defined in eq. (4.32) derive from the degree-3 shortness check over $\{-1, 0, 1\}$ from the product proof over the masked opening, similar to the one in [ALS20]:

$$\begin{aligned} \mathbf{f}_2 \left(\mathbf{f}_2 - \mathbf{c}\right) \left(\mathbf{f}_2 + \mathbf{c}\right) &= \left(\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle - \mathbf{c}\mathbf{s}\right) \left(\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle - \mathbf{c}\left(\mathbf{s} - \mathbf{1}\right)\right) \left(\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle - \mathbf{c}\left(\mathbf{s} + \mathbf{1}\right)\right) \\ &= \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^3 - 3\mathbf{s}\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^2 \mathbf{c} + \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle \left(3\mathbf{s}^2 - \mathbf{1}\right) \mathbf{c}^2 - \left(\mathbf{s}^3 - \mathbf{s}\right) \mathbf{c}^3. \end{aligned}$$

Hence, when $\mathbf{f}_2$ is the masked opening for $\mathbf{s}$, we let the garbage terms defined in eq. (4.32), such that

$$\begin{aligned} \mathbf{f}_4 &= \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}} \rangle - \mathbf{c}\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle \left(3\mathbf{s}^2 - \mathbf{1}\right), \\ \mathbf{f}_3 &= \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}} \rangle - \mathbf{c}\left(\langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}} \rangle - 3\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^2 \mathbf{s}\right), \end{aligned}$$

so $\mathbf{f}_3$ can cancel out the computationally indistinguishable masking $\langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}} \rangle$ in $\mathbf{f}_4$, and the claim $\mathbf{v}$ is

$$\mathbf{f}_2 \left(\mathbf{f}_2 - \mathbf{c}\right) \left(\mathbf{f}_2 + \mathbf{c}\right) + \mathbf{f}_3 + \mathbf{c}\mathbf{f}_4 = \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^3 + \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}} \rangle - \left(\mathbf{s}^3 - \mathbf{s}\right) \mathbf{c}^3 = \mathbf{v} - \left(\mathbf{s}^3 - \mathbf{s}\right) \mathbf{c}^3,$$

where a satisfying $\mathbf{s}$ should be vanishing in $\mathbf{s}^3 - \mathbf{s}$.

The current baseline protocol deviates from [BLS19], as c was needed for both the random linear combination to embed $\mathbf{s}$ to $\mathbf{z}$ (later proved by the linear relation proof), and the shortness check from the product proof, now c only serves the first purpose.

Now we move forward and amplify the soundness for the baseline protocol via automorphism repetition, such that the BDLOP commits to k uniform maskings, 1 message, and 2 garbage terms for shortness check. The masked openings for the garbage terms, and the combined claim, are formed as

$$\begin{aligned}
\mathbf{f}_{k+3} &= \langle \overline{\mathbf{b}}_{k+3}, \overline{\mathbf{y}}_1 \rangle - \mathbf{c} \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_{j+1} \sigma^{-j} \left(\langle \overline{\mathbf{b}}_{k+1}, \overline{\mathbf{y}}_{j+1} \rangle \left(3\mathbf{s}^2 - \mathbf{1} \right) \right) = \langle \overline{\mathbf{b}}_{k+3}, \overline{\mathbf{y}}_1 \rangle - \mathbf{cm}_{k+3}, \\
\mathbf{f}_{k+2} &= \langle \overline{\mathbf{b}}_{k+2}, \overline{\mathbf{y}}_1 \rangle - \mathbf{c} \left(\langle \overline{\mathbf{b}}_{k+3}, \overline{\mathbf{y}}_1 \rangle - 3 \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_{j+1} \sigma^{-j} \left(\langle \overline{\mathbf{b}}_{k+1}, \overline{\mathbf{y}}_{j+1} \rangle^2 \mathbf{s} \right) \right) = \langle \overline{\mathbf{b}}_{k+2}, \overline{\mathbf{y}}_1 \rangle - \mathbf{cm}_{k+2}, \\
\mathbf{v} &= \langle \overline{\mathbf{b}}_{k+2}, \overline{\mathbf{y}}_1 \rangle + \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_{j+1} \sigma^{-j} \left(\langle \overline{\mathbf{b}}_{k+1}, \overline{\mathbf{y}}_{j+1} \rangle^3 \right),
\end{aligned} \tag{4.34}$$

such that the random linear combination yields the statement and the invariant for combined claim

$$\begin{aligned}
\mathbf{0} &= \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_{j+1} \sigma^{-j} \left(\mathbf{s}^3 - \mathbf{s} \right), \\
\mathbf{v} &= \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_{j+1} \sigma^{-j} \left(\mathbf{f}_{k+1,j+1} \left(\mathbf{f}_{k+1,j+1} - \sigma^j(\mathbf{c}) \right) \left(\mathbf{f}_{k+1,j+1} + \sigma^j(\mathbf{c}) \right) \right) + \mathbf{f}_{k+2} + \mathbf{c} \mathbf{f}_{k+3}.
\end{aligned} \tag{4.35}$$

The amplified baseline protocol follows:

- The prover sends out $\overline{w}_i \leftarrow A\hat{\mathbf{y}}_i$ and $\overline{\mathbf{w}}_i \leftarrow \mathbf{B}_0 \overline{\mathbf{y}}_i$, where $\mathbf{y}_i \xleftarrow{R} R_q$ and $\overline{\mathbf{y}}_i \leftarrow D_\sigma^{Nd}$, and samples $\overline{\mathbf{r}} \leftarrow \chi^{Nd}$ to commit

$$\begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_1 \\ \vdots \\ \mathbf{t}_k \\ \mathbf{t}_{k+1} \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_1^\mathsf{T} \\ \vdots \\ \overline{\mathbf{b}}_k^\mathsf{T} \\ \overline{\mathbf{b}}_{k+1}^\mathsf{T} \end{bmatrix} \overline{\mathbf{r}} + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_k \\ \mathbf{s} \end{bmatrix}.$$

- The verifier challenges by sending $c_1, \ldots, c_k \xleftarrow{R} \mathbb{Z}_q$ and $\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_k \xleftarrow{R} R_q$.
- The prover replies with $\mathbf{z}_i \leftarrow c_i \cdot \mathbf{s} + \mathbf{y}_i$ for $i \in [1, k]$, and starts Π_{lin} and the shortness check in [ALS20] by:
  - The prover derives $\overline{\mathbf{b}}_{f_i} := \overline{\mathbf{b}}_i + c_i \cdot \overline{\mathbf{b}}_{k+1}$, replies $\mathbf{w}'_i \leftarrow \langle \overline{\mathbf{b}}_{f_i}, \overline{\mathbf{y}}_i \rangle$, then it commits

$$\begin{bmatrix} \mathbf{t}_{k+2} \\ \mathbf{t}_{k+3} \end{bmatrix} \leftarrow \begin{bmatrix} \overline{\mathbf{b}}_{k+2}^\mathsf{T} \\ \overline{\mathbf{b}}_{k+3}^\mathsf{T} \end{bmatrix} \overline{\mathbf{r}} + \begin{bmatrix} \mathbf{m}_{k+2} \\ \mathbf{m}_{k+3} \end{bmatrix},$$

    and replies with $\mathbf{t}_{k+2}, \mathbf{t}_{k+3}$, and the claim $\mathbf{v}$ in eq. (4.34),
  - The verifier samples $\mathbf{c}$ over the $\{-1, 0, 1\}$ challenge set C, with each coefficient has $P(\pm 1) = 1/4$, $P(0) = 1/2$.
  - The prover runs rejection sampling, and outputs $\overline{\mathbf{z}}'_i \leftarrow \overline{\mathbf{y}}_i + \sigma^{i-1}(\mathbf{c}) \cdot \overline{\mathbf{r}}$ for $i \in [1, k]$.
- The verifier checks by ensuring $\overline{\mathbf{z}}'_i$ being short, $A\hat{\mathbf{z}}_i = \overline{w}_i + c_i \cdot \overline{u}$ for $i \in [1, k]$,

$$\begin{bmatrix} \overline{\mathbf{w}}_i \\ \mathbf{w}'_i \end{bmatrix} + \sigma^{i-1}(\mathbf{c}) \cdot \begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_i + c_i \cdot \mathbf{t}_{k+1} \end{bmatrix} = \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_{f_i}^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}}'_i + \sigma^{i-1}(\mathbf{c}) \cdot \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{z}_i \end{bmatrix},$$

  and the shortness check from the product proof in [ALS20] by checking the second equality eq. (4.35), where

$$\mathbf{f}_{k+1,j+1} = \langle \overline{\mathbf{b}}_{k+1}, \overline{\mathbf{z}}'_{j+1} \rangle - \sigma^j(\mathbf{c}) \, \mathbf{t}_{k+1},$$

  for $j \in [0, k-1]$, and $\mathbf{f}_j = \langle \overline{\mathbf{b}}_j, \overline{\mathbf{z}}'_1 \rangle - \mathbf{c}\mathbf{t}_j$ for $j \in \{k+2, k+3\}$.

But now the baseline protocol augmented with automorphism repetition is more like the prior attempt to amplify soundness for product proof with automorphism repetition. The shortness proof itself needs 2 garbage terms, while the linearity test for each repetition needs k separate masking terms. A question arises that how can BDLOP commits to only a constant number of messages, while we can still amplify the soundness.

Linear Relation Testing via NTT Averaging. We consider an interactive protocol testing the matrix multiplication statement $A\hat{\mathbf{s}} = \overline{u}$ without $\mathbf{y} \xleftarrow{R} R_q$: Consider a linear prover, which is bound to a witness chosen before the challenge, and must return the verifier-specified linear evaluation of the witness, then

- The verifier challenges by $\overline{\gamma} \xleftarrow{R} \mathbb{Z}_q^d$.
- The prover replies with $z \leftarrow \overline{\gamma}^{\mathsf{T}} A\hat{\mathbf{s}}$.
- The verifier checks $z = \langle \overline{\gamma}, \overline{u} \rangle$.

By Schwartz-Zippel, if $A\hat{\mathbf{s}} - \overline{u} \neq \overline{0}$, over the randomness of $\overline{\gamma}$, the probability of $\overline{\gamma}^{\mathsf{T}} (A\hat{\mathbf{s}} - \overline{u}) = 0$ is q^{-1}.

Put it in the polynomial ring computation over the NTT terms, we first consider the lemma in [ENS20].

Lemma 4.13.1 (NTT Averaging Lemma). *Let R_q be defined as eq. (4.14), $\mathbf{f} \in R_q$, and $\hat{\mathbf{f}}_i = \mathbf{f} \mod \left(X^{d/\ell} - \xi^i\right)$ for $i \in \mathbb{Z}_{2\ell}^*$, then when we lift $\hat{\mathbf{f}}_i$ to R_q, we have*

$$\frac{1}{\ell} \sum_{i \in \mathbb{Z}_{2\ell}^*} \hat{\mathbf{f}}_i = \sum_{j \in [0, d/\ell - 1]} f_j X^j.$$

Proof. Recall the eq. (4.17), then noticing

$$\sum_{i \in [0, \ell-1]} \xi^{(2i+1)v} = \xi^v \frac{\xi^{2\ell v} - 1}{\xi^{2v} - 1} = 0$$

for any $v \in [1, \ell - 1]$, then all the ξ^j in eq. (4.17) with degree $j > 0$ are filtered out in the averaging summation of $\hat{\mathbf{f}}_i$, while the constant coefficients $\{f_i\}_{i \in [0, d/\ell - 1]}$ are scaled up by ℓ. □

Consider $\mathsf{NTT}^{-1}(\overline{\gamma}^{\mathsf{T}} A) \in R_q$, then by lemma 4.13.1, $d^{-1} \langle \overline{\gamma}^{\mathsf{T}} A, \hat{\mathbf{s}} \rangle$ is in the constant coefficient of $\mathsf{NTT}^{-1}(\overline{\gamma}^{\mathsf{T}} A) \cdot \mathbf{s} \in R_q$.

- The prover sends out $\overline{\mathbf{w}} \leftarrow \mathbf{B}_0 \overline{\mathbf{y}}$, where $\overline{\mathbf{y}} \leftarrow D_\sigma^{Nd}$, samples $\mathbf{g} \xleftarrow{R} \{R_q : g_0 = 0\}$ and $\overline{\mathbf{r}} \leftarrow \chi^{Nd}$, to commit

$$\begin{bmatrix} \overline{\mathbf{t}}_0 \\ \overline{\mathbf{t}}_1 \\ \overline{\mathbf{t}}_2 \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_1^{\mathsf{T}} \\ \overline{\mathbf{b}}_2^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{r}} + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{g} \\ \mathbf{s} \end{bmatrix}.$$

- The verifier challenges by sending $\overline{\gamma} \xleftarrow{R} \mathbb{Z}_q^d$.
- The prover replies with $\mathbf{z} \leftarrow \mathsf{NTT}^{-1}(d\overline{\gamma}^{\mathsf{T}} A) \cdot \mathbf{s} + \mathbf{g}$, and starts Π_{lin} and the shortness check in [ALS20] by:
  - The prover derives $\overline{\mathbf{b}}_{f_0} := \overline{\mathbf{b}}_1 + \mathsf{NTT}^{-1}(d\overline{\gamma}^{\mathsf{T}} A) \cdot \overline{\mathbf{b}}_2$, replies $\mathbf{w}' \leftarrow \langle \overline{\mathbf{b}}_{f_0}, \overline{\mathbf{y}} \rangle$, then it commits

$$\begin{bmatrix} \mathbf{t}_3 \\ \mathbf{t}_4 \end{bmatrix} \leftarrow \begin{bmatrix} \overline{\mathbf{b}}_3^{\mathsf{T}} \\ \overline{\mathbf{b}}_4^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{r}} + \begin{bmatrix} \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}} \rangle - 3 \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^2 \mathbf{s} \\ \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle \left(3\mathbf{s}^2 - \mathbf{1}\right) \end{bmatrix},$$

    and replies $\mathbf{t}_3, \mathbf{t}_4$, and the claim $\mathbf{v} \leftarrow \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}} \rangle^3 + \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}} \rangle$.
  - The verifier samples $\mathbf{c}$ over the $\{-1, 0, 1\}$ challenge set C, with each coefficient has $P(\pm 1) = 1/4$, $P(0) = 1/2$.

– The prover runs rejection sampling, and outputs $\overline{\mathbf{z}}' \leftarrow \overline{\mathbf{y}} + \mathbf{c} \cdot \overline{\mathbf{r}}$.

- The verifier checks by ensuring $\overline{\mathbf{z}}'$ being short, $z_0 = \langle \overline{\gamma}, \overline{u} \rangle$,

$$\begin{bmatrix} \overline{\mathbf{w}} \\ \mathbf{w}' \end{bmatrix} + \mathbf{c} \cdot \begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_1 + \mathsf{NTT}^{-1}(d\overline{\gamma}^{\mathsf{T}} A) \cdot \mathbf{t}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_{f_0}^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{z}}' + \mathbf{c} \cdot \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{z} \end{bmatrix},$$

and the shortness check from the product proof in [ALS20] by eq. (4.33).

Coefficient Filtering via Automorphisms. The prior construction from NTT averaging suffices as a single repetition, but for soundness amplification by automorphism repetition, the masking term over $\{R_q : g_0 = 0\}$ makes it admissible for only a single repetition.

We now want a way to filter out coefficients over R_q via linear operations and automorphisms, or otherwise the new linearity test is no better than the [BLS19] one.

We start by an observation that can filter out all the odd powered coefficients. Since

$$\sigma_{d+1}(X) = X^{d+1} \equiv -X \mod (X^d + 1),$$

then

$$\mathbf{f} + \sigma_{d+1}(\mathbf{f}) = 2\left(f_0 + f_2 X^2 + \cdots + f_{d-2} X^{d-2}\right).$$

Then, in $\log k$ rounds, the automorphisms $\sigma_{d+1}, \sigma_{d/2+1}, \ldots, \sigma_{2d/k+1}$ can yield

$$k\left(f_0 + f_k X^k + \cdots + f_{d-k} X^{d-k}\right).$$

Moreover, after r round for $r \in [1, \log_2 k]$, all the automorphisms used are indexed by $\{\sigma_{i \cdot 2d/2^r+1}\}_{i \in [0, 2^r-1]}$.

Lemma 4.13.2 (Automorphism Filtering Lemma)**.** *For $k \leq d$ and $k \mid d$, then for any $\mathbf{f} \in R_q$,*

$$F_k(\mathbf{f}) = \sum_{i \in [0,k-1]} \sigma_{i \cdot 2d/k+1}(\mathbf{f}) = k\left(f_0 + f_k X^k + \cdots + f_{d-k} X^{d-k}\right).$$

Moreover, when $k < d$,

$$F_k(\mathbf{f}) = \sum_{i \in [0,k-1]} \sigma^i_{2d/k+1}(\mathbf{f}). \tag{4.36}$$

Now with lemma 4.13.2, each repetition can be mapped to coefficients of degree in $i \in [0, k-1]$, by multiplying terms X^i, then the masking term can be relaxed to $\mathbf{g} \xleftarrow{\mathrm{R}} \{R_q : g_i = 0, i \in [0, k-1]\}$, the responses masked by $\mathbf{g}$ is,

$$\mathbf{z} \leftarrow \mathbf{g} + \sum_{i \in [0,k-1]} k^{-1} X^i F_k\left(\mathsf{NTT}^{-1}\left(d\overline{\gamma}_i^{\mathsf{T}} A\right) \cdot \mathbf{s}\right), \tag{4.37}$$

which can be derived from the commitment

$$\mathbf{t} = \mathbf{t}_1 + \sum_{i \in [0,k-1]} k^{-1} X^i F_k\left(\mathsf{NTT}^{-1}\left(d\overline{\gamma}_i^{\mathsf{T}} A\right) \cdot \mathbf{t}_2\right) = \langle \overline{\mathbf{b}}_1, \overline{\mathbf{r}} \rangle + \sum_{i \in [0,k-1]} k^{-1} X^i F_k\left(\mathsf{NTT}^{-1}\left(d\overline{\gamma}_i^{\mathsf{T}} A\right) \cdot \langle \overline{\mathbf{b}}_2, \overline{\mathbf{r}} \rangle\right) + \mathbf{z}, \tag{4.38}$$

where $\mathbf{t}_1$ commits to $\mathbf{g}$, and $\mathbf{t}_2$ commits to $\mathbf{s}$.

Yet, there are automorphism terms in eq. (4.38), and the prior vanilla Π_{lin} no longer suffices. We start by stepping back and thinking up a way to prove linear relation over 2 BDLOP commitments from different commitment keys (assuming they are of same size). Supposing they commit to $\mathbf{m}_0$ and $\mathbf{m}_1$, by sampling $\overline{\mathbf{r}}_0, \overline{\mathbf{r}}_1 \leftarrow \chi^{Nd}$, and

$$\begin{bmatrix} \overline{\mathbf{t}}_0 \\ \mathbf{t}_1 \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_1^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{r}}_0 + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{m}_0 \end{bmatrix}, \quad \begin{bmatrix} \overline{\mathbf{t}}_2 \\ \mathbf{t}_3 \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{B}_2 \\ \overline{\mathbf{b}}_3^{\mathsf{T}} \end{bmatrix} \overline{\mathbf{r}}_1 + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{m}_1 \end{bmatrix},$$

and $\mathbf{m}_2 = \mathbf{m}_0 + \mathbf{m}_1$ is the public statement. The protocol works as follows:

- The prover samples $\overline{\mathbf{y}}_0, \overline{\mathbf{y}}_1 \leftarrow D_\sigma^{Nd}$, and outputs $\overline{\mathbf{w}}_0 \leftarrow \mathbf{B}_0\overline{\mathbf{y}}_0$, $\overline{\mathbf{w}}_1 \leftarrow \mathbf{B}_2\overline{\mathbf{y}}_1$, and $\mathbf{w}_2 \leftarrow \langle\overline{\mathbf{b}}_1, \overline{\mathbf{y}}_0\rangle + \langle\overline{\mathbf{b}}_3, \overline{\mathbf{y}}_1\rangle$.
- The verifier challenges with $\mathbf{c}$ over the challenge set C, where each coefficient has $P(\pm 1) = 1/4$, $P(0) = 1/2$.
- The prover runs rejection sampling, and replies $\overline{\mathbf{z}}'_0 \leftarrow \overline{\mathbf{y}}_0 + \mathbf{c} \cdot \bar{\mathbf{r}}_0$ and $\overline{\mathbf{z}}'_1 \leftarrow \overline{\mathbf{y}}_1 + \mathbf{c} \cdot \bar{\mathbf{r}}_1$.
- The verifier checks $\overline{\mathbf{z}}'_0$ and $\overline{\mathbf{z}}'_1$ are short, and

$$\begin{bmatrix} \mathbf{B}_0 \\ \mathbf{0} \\ \overline{\mathbf{b}}_1^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}}'_0 + \begin{bmatrix} \mathbf{0} \\ \mathbf{B}_2 \\ \overline{\mathbf{b}}_3^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}}'_1 + \mathbf{c} \cdot \begin{bmatrix} \overline{\mathbf{0}} \\ \overline{\mathbf{0}} \\ \mathbf{m}_2 \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{w}}_0 \\ \overline{\mathbf{w}}_1 \\ \mathbf{w}_2 \end{bmatrix} + \mathbf{c} \cdot \begin{bmatrix} \bar{\mathbf{t}}_0 \\ \bar{\mathbf{t}}_2 \\ \mathbf{t}_1 + \mathbf{t}_3 \end{bmatrix}$$

The soundness is at most p by eq. (4.18). Automorphism repetition can be applied to have soundness to at most p^k.

With this protocol, we can actually prove the public statement of automorphism filtering $F_k(\mathbf{s})$ for committed $\mathbf{s}$. The intuition was to view $\sigma(\mathbf{B}_0)$ and $\sigma(\overline{\mathbf{b}}_1)$ as different commitment keys, where $\sigma = \sigma_{2d/k+1}$ for some $k < d$ and $k|d$, and the commitment is formed from $\bar{\mathbf{r}} \leftarrow \chi^{Nd}$ by

$$\begin{bmatrix} \bar{\mathbf{t}}_0 \\ \mathbf{t}_1 \end{bmatrix} = \begin{bmatrix} \mathbf{B}_0 \\ \overline{\mathbf{b}}_1^\mathsf{T} \end{bmatrix} \bar{\mathbf{r}} + \begin{bmatrix} \overline{\mathbf{0}} \\ \mathbf{s} \end{bmatrix}.$$

- The prover samples $\overline{\mathbf{y}}_i \leftarrow D_\sigma^{Nd}$, outputs $\overline{\mathbf{w}}_i \leftarrow \mathbf{B}_0\overline{\mathbf{y}}_i$ for $i \in [0, k-1]$, and

$$\mathbf{w}_k \leftarrow \sum_{i\in[0,k-1]} \sigma^{-i}\left(\langle\overline{\mathbf{b}}_1, \overline{\mathbf{y}}_i\rangle\right). \tag{4.39}$$

- The verifier challenges with $\mathbf{c}$ over the challenge set C, where each coefficient has $P(\pm 1) = 1/4$, $P(0) = 1/2$.
- The prover runs rejection sampling, and replies $\overline{\mathbf{z}}'_i \leftarrow \overline{\mathbf{y}}_i + \sigma^i(\mathbf{c}) \cdot \bar{\mathbf{r}}$ for $i \in [0, k-1]$.
- The verifier checks $\overline{\mathbf{z}}'_i$ are short for $i \in [0, k-1]$, and

$$\begin{bmatrix} \mathbf{B}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \overline{\mathbf{b}}_1^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}}'_0 + \begin{bmatrix} \mathbf{0} \\ \sigma^{-1}(\mathbf{B}_0) \\ \vdots \\ \mathbf{0} \\ \sigma^{-1}(\overline{\mathbf{b}}_1^\mathsf{T}) \end{bmatrix} \sigma^{-1}(\overline{\mathbf{z}}'_1) + \cdots + \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \sigma^{-(k-1)}(\mathbf{B}_0) \\ \sigma^{-(k-1)}(\overline{\mathbf{b}}_1^\mathsf{T}) \end{bmatrix} \sigma^{-(k-1)}(\overline{\mathbf{z}}'_{k-1}) + \mathbf{c} \cdot \begin{bmatrix} \overline{\mathbf{0}} \\ \vdots \\ \vdots \\ \overline{\mathbf{0}} \\ F_k(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{w}}_0 \\ \sigma^{-1}(\overline{\mathbf{w}}_1) \\ \vdots \\ \sigma^{-(k-1)}(\overline{\mathbf{w}}_{k-1}) \\ \mathbf{w}_k \end{bmatrix} + \mathbf{c} \cdot \begin{bmatrix} \bar{\mathbf{t}}_0 \\ \sigma^{-1}(\bar{\mathbf{t}}_0) \\ \vdots \\ \sigma^{-(k-1)}(\bar{\mathbf{t}}_0) \\ F_k(\mathbf{t}_1) \end{bmatrix}. \tag{4.40}$$

The soundness is at most p by eq. (4.18). Now take a closer look: The prior protocol uses $\mathbf{c}$ to test the automorphism filtering via the linear relation test over distinct BDLOP commitment keys in eq. (4.40). For automorphism repetition, think concretely from $\overline{\mathbf{z}}'_1 \leftarrow \overline{\mathbf{y}}_1 + \sigma(\mathbf{c}) \cdot \bar{\mathbf{r}}$, then the next $\mathbf{w}_{k,1}$ to send before $\mathbf{c}$ is

$$\mathbf{w}_{k,1} \leftarrow \sum_{i\in[0,k-1]} \sigma^{-i}\left(\langle\overline{\mathbf{b}}_1, \overline{\mathbf{y}}_{i+1 \bmod k}\rangle\right),$$

and eq. (4.40) becomes

$$\begin{bmatrix} \mathbf{B}_0 \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \overline{\mathbf{b}}_1^\mathsf{T} \end{bmatrix} \overline{\mathbf{z}}'_1 + \begin{bmatrix} \mathbf{0} \\ \sigma^{-1}(\mathbf{B}_0) \\ \vdots \\ \mathbf{0} \\ \sigma^{-1}(\overline{\mathbf{b}}_1^\mathsf{T}) \end{bmatrix} \sigma^{-1}(\overline{\mathbf{z}}'_2) + \cdots + \begin{bmatrix} \mathbf{0} \\ \vdots \\ \mathbf{0} \\ \sigma^{-(k-1)}(\mathbf{B}_0) \\ \sigma^{-(k-1)}(\overline{\mathbf{b}}_1^\mathsf{T}) \end{bmatrix} \sigma^{-(k-1)}(\overline{\mathbf{z}}'_0) + \sigma(\mathbf{c}) \cdot \begin{bmatrix} \overline{\mathbf{0}} \\ \vdots \\ \vdots \\ \overline{\mathbf{0}} \\ F_k(\mathbf{s}) \end{bmatrix} = \begin{bmatrix} \overline{\mathbf{w}}_1 \\ \sigma^{-1}(\overline{\mathbf{w}}_2) \\ \vdots \\ \sigma^{-(k-1)}(\overline{\mathbf{w}}_0) \\ \mathbf{w}_{k,1} \end{bmatrix} + \sigma(\mathbf{c}) \cdot \begin{bmatrix} \bar{\mathbf{t}}_0 \\ \sigma^{-1}(\bar{\mathbf{t}}_0) \\ \vdots \\ \sigma^{-(k-1)}(\bar{\mathbf{t}}_0) \\ F_k(\mathbf{t}_1) \end{bmatrix}.$$

By induction,

$$\mathbf{w}_{k,j} \leftarrow \sum_{i\in[0,k-1]} \sigma^{-i}\left(\langle\overline{\mathbf{b}}_1, \overline{\mathbf{y}}_{i+j \bmod k}\rangle\right) = \sigma^j(\mathbf{w}_k). \tag{4.41}$$

Supposing a slot under the slots grouped by $\left(X^k - \xi^k\right)$ is wrong, then the adversary needs to guess $\mathbf{c}$ mod $\left(X^k - \xi^k\right)$ to convince the verifier, with the same strategy for the affine forging in weak opening in the note for [ALS20], by fixing the pre-challenge $\overline{\mathbf{w}}$ for the relevant error term, yielding the soundness of at most p^k, where p is eq. (4.18) for the guessing probability of each coefficient in coarse $\left(X^k - \xi^k\right)$.

The rest is to prove the automorphism filtering of each $\mathsf{NTT}^{-1}(d\overline{\gamma}_j^{\mathsf{T}} A)\mathbf{s}$ for $j \in [0, k-1]$ by stacking them up over X^j, normalizing by k^{-1}, masking by $\mathbf{g}$, and eventually eq. (4.37). To carry on the automorphism filtering test derived from the linearity test, each $\mathbf{w}_{k,j}$ derived from eq. (4.39) is indexed by $j \in [0, k-1]$ for each $\overline{\gamma}_j$ by

$$\mathbf{w}_{k,j} \leftarrow \sum_{i \in [0,k-1]} \sigma^{-i} \left(\langle \mathsf{NTT}^{-1}(d\overline{\gamma}_j^{\mathsf{T}} A) \cdot \overline{\mathbf{b}}_2, \overline{\mathbf{y}}_i \rangle \right).$$

Correspondingly, the verifier can derive

$$\mathbf{u}_{k,j} \leftarrow \sum_{i \in [0,k-1]} \sigma^{-i} \left(\langle \mathsf{NTT}^{-1}(d\overline{\gamma}_j^{\mathsf{T}} A) \cdot \overline{\mathbf{b}}_2, \overline{\mathbf{z}}'_i \rangle \right),$$

and on the third round prover response, by eq. (4.41), for each $i \in [0, k-1]$ automorphic challenges $\sigma^i(\mathbf{c})$,

$$\mathbf{w}_{f_i} \leftarrow \langle \overline{\mathbf{b}}_1, \overline{\mathbf{y}}_i \rangle + \sum_{j \in [0,k-1]} k^{-1} X^j \sigma^i \left(\mathbf{w}_{k,j} \right), \tag{4.42}$$

and the verifier can derive correspondingly

$$\mathbf{u}_{f_i} \leftarrow \langle \overline{\mathbf{b}}_1, \overline{\mathbf{z}}'_i \rangle + \sum_{j \in [0,k-1]} k^{-1} X^j \sigma^i \left(\mathbf{u}_{k,j} \right), \tag{4.43}$$

such that for each $i \in [0, k-1]$ automorphic challenges $\sigma^i(\mathbf{c})$,

$$\mathbf{w}_{f_i} + \sigma^i(\mathbf{c}) \cdot \underbrace{\left(\mathbf{t}_1 + \sum_{j \in [0,k-1]} k^{-1} X^j F_k \left(\mathsf{NTT}^{-1} \left(d\overline{\gamma}_j^{\mathsf{T}} A \right) \cdot \mathbf{t}_2 \right) \right)}_{\mathbf{t}} = \mathbf{u}_{f_i} + \sigma^i(\mathbf{c}) \cdot \underbrace{\left(\mathbf{g} + \sum_{j \in [0,k-1]} k^{-1} X^j F_k \left(\mathsf{NTT}^{-1} \left(d\overline{\gamma}_j^{\mathsf{T}} A \right) \cdot \mathbf{s} \right) \right)}_{\mathbf{z}}, \tag{4.44}$$

where $\mathbf{t}, \mathbf{z}$ were in eq. (4.38) and eq. (4.37).

Again, now the masked openings for the garbage terms, and the combined claim, are formed as

$$\begin{aligned}
\mathbf{f}_4 &= \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}}_0 \rangle - \mathbf{c} \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_j \sigma^{-j} \left(\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}}_j \rangle \left(3\mathbf{s}^2 - \mathbf{1} \right) \right) = \langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}}_0 \rangle - \mathbf{c}\mathbf{m}_4, \\
\mathbf{f}_3 &= \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}}_0 \rangle - \mathbf{c} \left(\langle \overline{\mathbf{b}}_4, \overline{\mathbf{y}}_0 \rangle - 3 \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_j \sigma^{-j} \left(\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}}_j \rangle^2 \mathbf{s} \right) \right) = \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}}_0 \rangle - \mathbf{c}\mathbf{m}_3, \\
\mathbf{v} &= \langle \overline{\mathbf{b}}_3, \overline{\mathbf{y}}_0 \rangle + \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_j \sigma^{-j} \left(\langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}}_j \rangle^3 \right),
\end{aligned} \tag{4.45}$$

such that the random linear combination yields the statement and the invariant for combined claim

$$\begin{aligned}
\mathbf{0} &= \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_j \sigma^{-j} \left(\mathbf{s}^3 - \mathbf{s} \right), \\
\mathbf{v} &= \sum_{j \in [0,k-1]} \boldsymbol{\alpha}_j \sigma^{-j} \left(\mathbf{f}_{2,j} \left(\mathbf{f}_{2,j} - \sigma^j(\mathbf{c}) \right) \left(\mathbf{f}_{2,j} + \sigma^j(\mathbf{c}) \right) \right) + \mathbf{f}_3 + \mathbf{c}\mathbf{f}_4,
\end{aligned} \tag{4.46}$$

where $\mathbf{f}_i = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{z}}'_0 \rangle - \mathbf{c}\mathbf{t}_i = \langle \overline{\mathbf{b}}_i, \overline{\mathbf{y}}_0 \rangle - \mathbf{c}\mathbf{m}_i$ for $i \in \{3, 4\}$, each $\mathbf{f}_{2,j}$ for $j \in [0, k-1]$ is

$$\mathbf{f}_{2,j} = \langle \overline{\mathbf{b}}_2, \overline{\mathbf{z}}'_j \rangle - \sigma^j(\mathbf{c})\, \mathbf{t}_2 = \langle \overline{\mathbf{b}}_2, \overline{\mathbf{y}}_j \rangle - \sigma^j(\mathbf{c})\, \mathbf{s},$$

and the second equality holds if the prover is honest.

The protocol for a single matrix multiplication claim follows:

- The prover sends $\overline{\mathbf{w}}_i \leftarrow \mathbf{B}_0\overline{\mathbf{y}}_i$, where $\overline{\mathbf{y}}_i \leftarrow D_\sigma^{Nd}$, samples $\mathbf{g} \xleftarrow{R} \{R_q : g_j = 0, j \in [0, k-1]\}$ and $\overline{\mathbf{r}} \leftarrow \chi^{Nd}$ to commit

$$\begin{bmatrix}\overline{\mathbf{t}}_0\\ \mathbf{t}_1\\ \mathbf{t}_2\end{bmatrix} \leftarrow \begin{bmatrix}\mathbf{B}_0\\ \overline{\mathbf{b}}_1^{\mathsf{T}}\\ \overline{\mathbf{b}}_2^{\mathsf{T}}\end{bmatrix}\overline{\mathbf{r}} + \begin{bmatrix}\overline{\mathbf{0}}\\ \mathbf{g}\\ \mathbf{s}\end{bmatrix}.$$

- The verifier challenges by sending $\overline{\gamma}_0, \ldots, \overline{\gamma}_{k-1} \xleftarrow{R} \mathbb{Z}_q^d$ and $\boldsymbol{\alpha}_0, \ldots, \boldsymbol{\alpha}_{k-1} \xleftarrow{R} R_q$.
- The prover replies $\mathbf{z}$ in eq. (4.37), and starts Π_{lin} and the shortness check in [ALS20] by:
  - The prover replies each $\mathbf{w}_{f_i}$ for $i \in [0, k-1]$ in eq. (4.42), then it commits

$$\begin{bmatrix}\mathbf{t}_3\\ \mathbf{t}_4\end{bmatrix} \leftarrow \begin{bmatrix}\overline{\mathbf{b}}_3^{\mathsf{T}}\\ \overline{\mathbf{b}}_4^{\mathsf{T}}\end{bmatrix}\overline{\mathbf{r}} + \begin{bmatrix}\mathbf{m}_3\\ \mathbf{m}_4\end{bmatrix},$$

    and replies with $\mathbf{t}_3, \mathbf{t}_4$, and the claim $\mathbf{v}$ in eq. (4.45),
  - The verifier samples $\mathbf{c}$ over the $\{-1, 0, 1\}$ challenge set C, with each coefficient has $P(\pm 1) = 1/4$, $P(0) = 1/2$.
  - The prover runs rejection sampling, and outputs $\overline{\mathbf{z}}_i' \leftarrow \overline{\mathbf{y}}_i + \sigma^i(\mathbf{c}) \cdot \overline{\mathbf{r}}$ for $i \in [0, k-1]$.
- The verifier checks $\overline{\mathbf{z}}_i'$ being short, $z_j = \langle \overline{\gamma}_j, \overline{u} \rangle$ for $j \in [0, k-1]$.

  It then derives $\mathbf{u}_{f_i}$ for $i \in [0, k-1]$ in eq. (4.43), $\mathbf{t}$ in eq. (4.38), checks

$$\begin{aligned}\overline{\mathbf{w}}_i + \sigma^i(\mathbf{c}) \cdot \overline{\mathbf{t}}_0 &= \mathbf{B}_0\overline{\mathbf{z}}_i',\\ \mathbf{w}_{f_i} + \sigma^i(\mathbf{c}) \cdot \mathbf{t} &= \mathbf{u}_{f_i} + \sigma^i(\mathbf{c}) \cdot \mathbf{z}\end{aligned}$$

  for $i \in [0, k-1]$, which is eq. (4.44) and eq. (4.40),

  Finally, it checks the shortness of $\hat{\mathbf{s}}$ by the second equality eq. (4.46).

Cost Comparison Against [BLS19]. We compare the published noninteractive constructions of [BLS19, ENS20] by counting transmitted elements and their coefficient widths.

Let n be the total witness length, $A \in \mathbb{Z}_q^{m \times n}$, d be the R_q degree, and $b_q = \lceil \log_2 q \rceil$. We retain N for the opening-vector width and μ for the top commitment rank. Counts exclude the public statement and commitment key. The following are the papers' historical parameter choices for approximately 128-bit security.

Parameter	BLS19	ENS20
Matrix dimensions $m \times n$	1024×2048	1024×2048
Coefficient width b_q	32	32
Ring degree d	2048	128
Witness polynomials n/d	1	16
Repetitions	$t = 4, t' = 3$	$k = 4$
Commitment ranks (λ, μ)	$(1, 1)$	$(10, 9)$

In particular, ENS separates the witness length from the ring degree: Their example uses 16 witness polynomials.

BLS: transmitted objects. In the optimized Fiat–Shamir protocol of [BLS19, Algorithms 2–3 and Section 3.4], each of the t commitment vectors has five ring elements, with one shared across repetitions. There are t masked witness replies and tt' short opening vectors of width 6.

Component	Count	Bits per coefficient or choice
Commitments	$1 + 4t$ ring elements	b_q
Masked witness replies	t ring elements	b_q
Short opening replies	$6tt'$ ring elements	b_z
Scalar challenges	t scalars	b_q
Monomial challenges	tt' choices	$\lceil \log_2 \ell \rceil$

Here $d = n$ and $\ell = 2d$ for the power-of-two benchmark. The proof size in bits is therefore

$$S_{\mathrm{BLS}} = (1 + 5t)db_q + 6tt'db_z + t\big(b_q + t'\lceil \log_2 \ell \rceil\big). \tag{4.47}$$

For $\sigma = 919.13$, the short-coefficient width is $b_z = \lceil \log_2(12\sigma) \rceil = 14$. Substituting the benchmark parameters gives

Component	Calculation in bytes	Size
Full part	$21 \cdot 2048 \cdot 32/8 = 172{,}032$	168 KiB
Short part	$72 \cdot 2048 \cdot 14/8 = 258{,}048$	252 KiB
Challenges	$4(32 + 3 \cdot 12)/8 = 34$	34 bytes
Total	430,114	420.0332 KiB

ENS: transmitted objects. For n/d witness polynomials, the committed message count and opening width are

$$L = n/d + 3, \qquad N = \lambda + \mu + L.$$

The additional 3 messages are the $\mathbf{g}$ linear-test mask, and the two shortness garbage terms.

Besides the $\mu + L$ commitment elements, ENS transmits one masked linear-test polynomial, called h in the paper. For our single-polynomial notation it is the publicly centered response

$$h = \mathbf{z} - \sum_{j=0}^{k-1} X^j \langle \overline{\gamma}_j, \overline{u} \rangle,$$

so its first k coefficients are known to be zero. After reconstructing the first messages, the remaining short replies contain kN ring elements. With $b_z = \lceil \log_2(2\delta_1) \rceil$, [ENS20, Section 4.2, Equation (11)] gives the basic bit count

$$\begin{aligned} S_{\mathrm{ENS}}^{\mathrm{basic}} &= (\mu + n/d + 4)db_q + k(\lambda + \mu + n/d + 3)db_z \\ &= [n + (\mu + 4)d]b_q + k[n + (\lambda + \mu + 3)d]b_z. \end{aligned} \tag{4.48}$$

This leading count omits the 256-bit challenge seed and the saving from the k known coefficients of h.

For bookkeeping, insert $\delta_1 = 2^{18}$ from [ENS20, Appendix B.1], so $b_z = 19$, $L = 19$, and $N = 38$:

Component	Calculation in bytes	Size
Full part	$29 \cdot 128 \cdot 32/8 = 14{,}848$	14.5 KiB
Short part	$152 \cdot 128 \cdot 19/8 = 46{,}208$	45.125 KiB
Basic total	61,056	59.625 KiB

Although ENS has more short polynomials (152 versus 72), it has 19456 short coefficients, compared to BLS's 147456. ENS and BLS contain 29 and 21 full ring elements, respectively, or 3712 and 43008 coefficients.

4.14 Lattice Proofs for Exact Euclidean Norm Bounds and Quadratic Relations

One of the motivations of [LNP22] is that, the prior works [BLS19, ENS20] for $A\hat{\mathbf{s}} = \overline{\mathbf{u}}$ restrict each entry of $\hat{\mathbf{s}}$ being over $\{-1, 0, 1\}$, and thus committing the $\mathbf{m} = \mathsf{NTT}^{-1}(\hat{\mathbf{s}})$, which can be arbitrarily long, can be expensive with [BDL+18]. If we want to work with $\hat{\mathbf{s}}$ with larger ℓ_∞-norm, then more garbage terms are needed in the commitment.

The question they asked was: Is NTT packing all that necessary, as both concrete efficiency and communication complexity can be improved if no NTT packing is involved. In [LNP22], the encoding over CRT slots [BLS19, ALS20, ENS20] is not used, messages are committed in coefficient form. In this case, R_q is not necessarily fully splitting.

Challenge Set. There are many ways to construct challenge sets for the R_q:

- In [LS18], the corollary 4.12.3 gives a short challenge set, such that all nonzero pairwise differences are invertible, but by corollaries 4.12.2 and 4.12.3, the modulus grows with the number of splitting factors and the norm bound of the set of short elements, and some computation is in table 4.12.1;
- In [BLS19], the eq. (4.7) suffices;
- In [ALS20], the anti-concentration lemma 4.12.5 gives the max point probability of a CRT slot, and the probability for pairwise difference being non-invertible is upper bounded by a union bound. The extractor can extract CRT slot by slot, so it is not required to be globally pairwise difference invertible.

In [LNP22], they found another set of elements in R_q that is invertible.

Lemma 4.14.1. *By corollary 4.12.2, let $p \equiv 5 \pmod 8$ be a prime, then any nonzero $\mathbf{c} \in R_p$ such that $\sigma_{-1}(\mathbf{c}) = \mathbf{c}$ is invertible.*

Proof. By corollary 4.12.2, if $p \equiv 5 \bmod 8$ is a prime, then

$$R_p = \mathbb{Z}_p[X]/(X^d + 1) \cong \mathbb{Z}_p[X]/(X^{d/2} - r) \times \mathbb{Z}_p[X]/(X^{d/2} + r),$$

r being the 4^{th} primitive root of unity, splitting 2 irreducible factors. Since $\sigma_{-1}(\mathbf{c}) = \mathbf{c}$, then satisfying elements are

$$\mathbf{c} = c_0 + \sum_{i \in [1, d/2-1]} c_i \left(X^i - X^{d-i}\right) = c_0 + \sum_{i \in [1, d/2-1]} c_i X^i - \sum_{i \in [1, d/2-1]} c_i X^{d-i},$$

as any $X^i - X^{d-i}$ satisfies $\sigma_{-1}(X^i - X^{d-i}) = X^i - X^{d-i}$. Hence, for each factor $(X^{d/2} \pm r)$, we have

$$\mathbf{c} \equiv c_0 + \sum_{i \in [1, d/2-1]} \left(c_i \pm r c_{d/2-i}\right) X^i \mod \left(X^{d/2} \pm r\right).$$

We claim at least one of $c_i - r c_{d/2-i}$ and $c_{d/2-i} - r c_i$ is nonzero, given that at least one of $c_i, c_{d/2-i}$ is nonzero: Supposing $c_i = r c_{d/2-i}$, then $(1 - r^2) c_{d/2-i} \neq 0$. The case for $c_i + r c_{d/2-i}$ and $c_{d/2-i} + r c_i$ is the same idea.

Hence, a nonzero $\mathbf{c}$ is nonzero at both factors if $\sigma_{-1}(\mathbf{c}) = \mathbf{c}$. □

The lemma 4.14.1 can be applied to the R_q, where $q = \prod q_i$, and $q_1 < \cdots < q_k$, with all $q_i \equiv 5 \pmod 8$ are primes, then $R_q = \prod R_{q_i}$, and thus each R_{q_i} splits into 2 irreducible factors. Hence, any nonzero $\mathbf{c} \in R_q$ satisfying $\sigma_{-1}(\mathbf{c}) = \mathbf{c}$ and $0 < \|\mathbf{c}\|_\infty < q_1$, $\mathbf{c}$ is invertible.

The short invertible set of elements from lemma 4.14.1 shares a similar vibe to [LS18]. The similarity is generally "If it is just too short, it will not wrap around in any of the CRT slots".

- The current one finds a pattern of invertibility $\sigma_{-1}(\mathbf{c}) = \mathbf{c}$ that applies to any R_{q_i} splitting into 2 irreducible factors in the above setting, then any nonzero $\mathbf{c} \in R_q$ such that $\sigma_{-1}(\mathbf{c}) = \mathbf{c}$, that is shorter than q_1 in ℓ_∞-norm is invertible, as the coefficient representations within $[-q_1 + 1, q_1 - 1]$ can remain the same for all $[-q_i + 1, q_i - 1]$ where $q_i \geq q_1$.
- The [LS18] one is, considering $R_q \cong \prod \mathbb{Z}_q[X]/(f_i(X))$, then for the ideal lattice $\Lambda = \{\mathbf{x} \in R_m : \mathbf{x} \equiv 0 \mod (f_i, q)\}$, if nonzero $\mathbf{x} \in R_q$ is shorter than $\lambda_1(\Lambda)$ in ℓ_2-norm, then $\mathbf{x}$ cannot belong to the CRT kernel lattice Λ.

Now we move on and consider the norm growth when multiplying with challenge set elements, namely we want to bound $\|\mathbf{c}\bar{\mathbf{r}}\|_2$. The main idea of the next lemma is to upper bound the operator norm of $\mathsf{rot}(\mathbf{c}) \in \mathbb{Z}^{d\times d}$, such that

$$\|\mathbf{c}\bar{\mathbf{r}}\| \leq \left(\max_{\|\mathbf{x}\|_2=1} \|\mathbf{c}\mathbf{x}\|_2\right) \|\bar{\mathbf{r}}\|_2 = \|\mathsf{rot}(\mathbf{c})\|_2 \|\bar{\mathbf{r}}\|_2 .$$

Lemma 4.14.2. *Let $\bar{\mathbf{r}} \in R^\ell$ and $\mathbf{c} \in R$, then for any power-of-two κ, we have*

$$\|\mathbf{c}\bar{\mathbf{r}}\|_2 \leq \|\sigma_{-1}(\mathbf{c}^\kappa)\,\mathbf{c}^\kappa\|_1^{1/(2\kappa)} \cdot \|\bar{\mathbf{r}}\|_2 .$$

Proof. We first write $\mathbf{C} = \mathsf{rot}(\mathbf{c}) \in \mathbb{Z}^{d\times d}$, then

$$\|\mathbf{C}\|_2^2 = \max_{\|\overline{x}\|_2=1} \|\mathbf{C}\overline{x}\|_2^2 = \max_{\|\overline{x}\|_2=1} \overline{x}^\mathsf{T}\mathbf{C}^\mathsf{T}\mathbf{C}\overline{x}, \tag{4.49}$$

where the last equality holds by squaring up and summing up all entries in $\mathbf{C}\overline{x}$, then by eq. (4.49), $\mathbf{C}^\mathsf{T}\mathbf{C}$ is positive semidefinite. By the spectral theorem, $\mathbf{C}^\mathsf{T}\mathbf{C}$ has an orthonormal eigenbasis, that eigenvectors $\{\overline{v}_i\}_{i\in[1,d]}$ are mutually perpendicular and of ℓ_2-norm being 1. Hence, for $\overline{x}$ with $\|\overline{x}\|_2 = 1$,

$$\mathbf{C}^\mathsf{T}\mathbf{C}\overline{x} = \mathbf{C}^\mathsf{T}\mathbf{C} \sum_{i\in[1,d]} \alpha_i \overline{v}_i = \sum_{i\in[1,d]} \alpha_i \lambda_i \overline{v}_i, \tag{4.50}$$

and by the $\|\overline{x}\|_2 = 1$,

$$1 = \left\| \sum_{i\in[1,d]} \alpha_i \overline{v}_i \right\|_2^2 = \sum_{i\in[1,d]} \alpha_i^2 + \sum_{i\neq j\in[1,d]} \alpha_i\alpha_j \langle \overline{v}_i, \overline{v}_j\rangle = \sum_{i\in[1,d]} \alpha_i^2. \tag{4.51}$$

Therefore, conditioned that eq. (4.51), we have

$$\left\|\mathbf{C}^\mathsf{T}\mathbf{C}\right\|_2 = \max_{\|\overline{x}\|=1} \left\|\mathbf{C}^\mathsf{T}\mathbf{C}\overline{x}\right\|_2 = \max_{\{\alpha_i\}_{i\in[1,d]}} \left\| \sum_{i\in[1,d]} \alpha_i\lambda_i\overline{v}_i \right\|_2 = \max_{\{\alpha_i\}_{i\in[1,d]}} \left(\sum_{i\in[1,d]} \alpha_i^2\lambda_i^2 \right)^{1/2} = \max_{i\in[1,d]} |\lambda_i| ,$$

where the third equality holds by eq. (4.51). Moreover, continuing from eq. (4.50), we finish eq. (4.49) by

$$\|\mathbf{C}\|_2^2 = \max_{\{\alpha_i\}_{i\in[1,d]}} \sum_{i\in[1,d]} \alpha_i^2\lambda_i = \max_{i\in[1,d]} \lambda_i = \left\|\mathbf{C}^\mathsf{T}\mathbf{C}\right\|_2 ,$$

where the last equality holds by the positive semidefiniteness of $\mathbf{C}^\mathsf{T}\mathbf{C}$, such that each λ_i is nonnegative. Continuing on, since $\mathbf{A} = \mathbf{C}^\mathsf{T}\mathbf{C}$ is positive semidefinite, then $\mathbf{A}^\mathsf{T}\mathbf{A} = \mathbf{A}^2$ as $\mathbf{A}$ is diagonal symmetric, then

$$\|\mathbf{C}\|_2^{2\kappa} = \left\|\mathbf{C}^\mathsf{T}\mathbf{C}\right\|_2^\kappa = \left\|\left(\mathbf{C}^\mathsf{T}\mathbf{C}\right)^\kappa\right\|_2 .$$

For $\mathbf{u}, \mathbf{v} \in R$, another observation for $\mathbf{u}\mathbf{v} = \mathsf{rot}(\mathbf{u})\overline{v}$ is,

$$\mathsf{rot}(\mathbf{u})\overline{v} = \sum_{i\in[0,d-1]} u_i \mathsf{rot}(X^i)\overline{v},$$

and therefore by triangle inequality,

$$\|\mathsf{rot}(\mathbf{u})\overline{v}\|_2 \leq \sum_{i\in[0,d-1]} |u_i| \left\|\mathsf{rot}(X^i)\overline{v}\right\|_2 = \|\mathbf{u}\|_1 \cdot \|\mathbf{v}\|_2 . \tag{4.52}$$

Hence, for $\mathbf{v}$ with $\|\mathbf{v}\|_2 = 1$ and coefficients over $\mathbb{R}$, by operator norm eq. (4.49), $\|\mathsf{rot}(\mathbf{u})\|_2 \leq \|\mathbf{u}\|_1$. The rest is immediate by the fact that $\mathbf{C}^\mathsf{T} = \mathsf{rot}(\sigma_{-1}(\mathbf{c}))$, then

$$\left(\mathbf{C}^\mathsf{T}\mathbf{C}\right)^2 = \mathsf{rot}\left(\sigma_{-1}\left(\sigma_{-1}(\mathbf{c})\mathbf{c}\right)\right) \mathsf{rot}\left(\sigma_{-1}(\mathbf{c})\mathbf{c}\right) = \mathsf{rot}\left(\sigma_{-1}(\mathbf{c})^2\mathbf{c}^2\right) = \mathsf{rot}\left(\sigma_{-1}(\mathbf{c}^2)\mathbf{c}^2\right) ,$$

and by induction,

$$\|\mathbf{C}\|_2^{2\kappa} = \left\|\mathbf{C}^\mathsf{T}\mathbf{C}\right\|_2^\kappa = \left\|\left(\mathbf{C}^\mathsf{T}\mathbf{C}\right)^\kappa\right\|_2 = \|\mathsf{rot}\left(\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa\right)\|_2 \leq \|\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa\|_1 .$$

□

Moreover, by eq. (4.52), we have $\|\mathbf{uv}\|_1 \le \|\mathbf{u}\|_1 \|\mathbf{v}\|_1$, then $\|\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa\|_1^{1/(2\kappa)}$ is monotonically nonincreasing by

$$\left\|\sigma_{-1}(\mathbf{c}^{2\kappa})\mathbf{c}^{2\kappa}\right\|_1^{1/(4\kappa)} = \left\|(\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa)^2\right\|_1^{1/(4\kappa)} \le \|\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa\|_1^{1/(2\kappa)}.$$

On the other hand, by Cauchy-Schwartz, $\|\mathbf{v}\|_1 \le \sqrt{d}\,\|\mathbf{v}\|_2$, then

$$\|\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa\|_1^{1/(2\kappa)} \le d^{1/(4\kappa)}\,\|\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa\|_2^{1/(2\kappa)} \le d^{1/(4\kappa)}\left\|\left(\mathbf{C}^\mathsf{T}\mathbf{C}\right)^\kappa\right\|_2^{1/(2\kappa)} = d^{1/(4\kappa)}\,\|\mathbf{C}\|_2,$$

where the second inequality holds by letting the rotational matrix multiplying against $\overline{e}_0 = [1, 0, \ldots, 0]^\mathsf{T}$.

Therefore,

$$\|\mathbf{C}\|_2 \le \left\|\sigma_{-1}(\mathbf{c}^{2\kappa})\mathbf{c}^{2\kappa}\right\|_1^{1/(4\kappa)} \le \|\sigma_{-1}(\mathbf{c}^\kappa)\mathbf{c}^\kappa\|_1^{1/(2\kappa)} \le d^{1/(4\kappa)}\,\|\mathbf{C}\|_2,$$

and the upper bound by lemma 4.14.2 is monotonically nonincreasing, and is approaching the operator norm $\|\mathbf{C}\|_2$.

Given lemma 4.14.2 as an estimation for the ℓ_2 operator norm of $\mathbf{c} \in R$, we construct the challenge set C defined over the elements specified by lemma 4.14.1, that for a power-of-two k,

$$C := \{\mathbf{c} \in S_\kappa^\sigma \;:\; \left\|\sigma_{-1}(\mathbf{c}^k)\mathbf{c}^k\right\|_1^{1/(2k)} \le \eta\}, \tag{4.53}$$

where $S_\kappa^\sigma = \{\mathbf{c} \in S_\kappa \;:\; \sigma(\mathbf{c}) = \mathbf{c}\}$ for $\sigma \in \mathrm{Aut}(R_q)$, and S_κ is defined in eq. (4.5), that $\kappa \le (q_1 - 1)/2$. Eventually, the goal is to make the ℓ_∞-norm of challenges to be bounded by $(q_1 - 1)/2$, and estimate $\|\mathrm{rot}(\mathbf{c})\|_2 \le \eta$, such that $\|\mathbf{c}\overline{\mathbf{r}}\|_2 \le \eta\,\|\overline{\mathbf{r}}\|_2$.

Rejection Sampling and Parameters. We haven't get to the extended MLWE from [LNS21], so we use the vanilla rejection sampling from [Lyu12] throughout this note. In [LNP22], the masking randomness for [BDL+18] and [Ajt96] is sampled over a Gaussian distribution, and the next lemma help describe the concentration property of a sample.

Lemma 4.14.3 (Gaussian Tail Bound [Ban93]). *Let $\overline{\mathbf{z}} \leftarrow D_s^{md}$, then for $t > 1$,*

$$\Pr\left[\|\overline{\mathbf{z}}\|_2 > t \cdot s\sqrt{md}\right] < \left(t \exp\left(\frac{1 - t^2}{2}\right)\right)^{md}.$$

In [LNP22], the lemma 4.14.3 is parameterized with $t = \sqrt{2}$, then

$$\Pr\left[\|\overline{\mathbf{z}}\|_2 > s\sqrt{2md}\right] < (2/e)^{md/2}.$$

We continue on describing the rejection sampling from the same algorithm in [Lyu12], but parameterized differently from [BLS19]. The key idea is using the outputting probability threshold by [Lyu12] for $\overline{\mathbf{z}} \leftarrow \overline{\mathbf{y}} + \overline{\mathbf{v}}$ where $\overline{\mathbf{y}} \leftarrow D_s^{md}$

$$\min\left(\frac{1}{M} \cdot \frac{D_s^{md}(\overline{\mathbf{z}})}{D_{\overline{\mathbf{v}},s}^{md}(\overline{\mathbf{z}})}, 1\right) = \min\left(\frac{1}{M} \cdot \frac{\rho_s(\overline{\mathbf{z}})}{\rho_s(\overline{\mathbf{z}} - \overline{\mathbf{v}})}, 1\right), \tag{4.54}$$

where ρ_s is defined in eq. (4.9), and since we are working over Gaussian distributions, the equality holds. Moreover, over $\overline{\mathbf{z}} \leftarrow D_s^{md}$,

$$\Pr\left[D_s^{md}(\overline{\mathbf{z}}) \le M \cdot D_{\overline{\mathbf{v}},s}^{md}(\overline{\mathbf{z}})\right] \ge 1 - 2^{-128}, \tag{4.55}$$

then the threshold eq. (4.54) uses the left term with overwhelming probability: The M is chosen by $s := \gamma\,\|\overline{\mathbf{v}}\|_2$, then

$$\frac{\rho_s(\overline{\mathbf{z}})}{\rho_s(\overline{\mathbf{z}} - \overline{\mathbf{v}})} = \exp\left(\frac{-2\langle\overline{\mathbf{z}}, \overline{\mathbf{v}}\rangle + \|\overline{\mathbf{v}}\|_2^2}{2s^2}\right) \le \exp\left(\frac{28s\,\|\overline{\mathbf{v}}\|_2 + \|\overline{\mathbf{v}}\|_2^2}{2s^2}\right) = \exp\left(\frac{14}{\gamma} + \frac{1}{2\gamma^2}\right) = M, \tag{4.56}$$

(if $\|\overline{\mathbf{v}}\|_2 \le T$, then $s := \gamma T$), and the second inequality holds with overwhelming probability by Lemma 4.3 of [Lyu12]:

$$\Pr\left[|\langle\overline{\mathbf{z}}, \overline{\mathbf{v}}\rangle| \ge 14s\,\|\overline{\mathbf{v}}\|_2\right] \le 2\exp(-98) \le 2^{-128},$$

thus eq. (4.55) holds, the accepting probability is at least $(1 - 2^{-128})/M$, and conditioned on the sample's acceptance, the statistical distance of the real distribution is within 2^{-128} against the simulated distribution.

ABDLOP Commitment Scheme. Consider combining Ajtai [Ajt96] and BDLOP [BDL+18] together, then we have ABDLOP commitment [LNP22]. Recall eq. (4.19), where the commitment key is over $R_q^{(\mu+L)\times N}$:

- Binding is reduced to $\mathsf{MSIS}_{\mu,N,B}$.
- Hiding is reduced to $\mathsf{MLWE}_{\lambda,\mu+L,\chi}$, where commitment randomness is drawn from distribution χ^{Nd}.

Onwards, we use m_1 to denote the length of the Ajtai message, and m_2 to denote the length of the BDLOP randomness. The combination is done by stacking an Ajtai commitment to the SIS part in the BDLOP commitment:

$$\begin{bmatrix} \bar{\mathbf{t}}_A \\ \bar{\mathbf{t}}_B \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{0} \end{bmatrix} \bar{\mathbf{s}}_1 + \begin{bmatrix} \mathbf{A}_2 \\ \mathbf{B} \end{bmatrix} \bar{\mathbf{s}}_2 + \begin{bmatrix} \bar{\mathbf{0}} \\ \overline{\mathbf{m}} \end{bmatrix} \in R_q^{\mu+L},$$

such that binding is reduced to $\mathsf{MSIS}_{\mu,m_1+m_2,B}$, while hiding is reduced to $\mathsf{MLWE}_{\lambda,\mu+L,\chi}$.

Assume $\|\bar{\mathbf{s}}_1\|_2 \le \alpha$ and $\|\bar{\mathbf{s}}_2\|_\infty \le \nu$. By $\|\mathrm{rot}(\mathbf{c})\|_2 \le \eta$ from lemma 4.14.2, the ℓ_2-norm bound for the centering shifts are $T_1 := \eta\alpha$ and $T_2 := \eta\nu\sqrt{m_2 d}$. Choose $\gamma_1, \gamma_2 > 0$, and thus $s_i := \gamma_i T_i$ and $M_i := \exp\left(14/\gamma_i + 1/(2\gamma_i^2)\right)$ in eq. (4.56).

Combining the proof of opening for Ajtai [Ajt96] and [BDL+18], a three move ABDLOP proof of opening follows:

- The prover samples $\bar{\mathbf{y}}_1 \leftarrow D_{s_1}^{m_1 d}$ and $\bar{\mathbf{y}}_2 \leftarrow D_{s_2}^{m_2 d}$, and sends over $\overline{\mathbf{w}} \leftarrow \mathbf{A}_1\bar{\mathbf{y}}_1 + \mathbf{A}_2\bar{\mathbf{y}}_2$.
- The verifier challenges by $\mathbf{c} \leftarrow C$ defined in eq. (4.53).
- The prover runs rejection sampling, and replies $\bar{\mathbf{z}}_1 \leftarrow \bar{\mathbf{y}}_1 + \mathbf{c}\cdot\bar{\mathbf{s}}_1$ and $\bar{\mathbf{z}}_1 \leftarrow \bar{\mathbf{y}}_2 + \mathbf{c}\cdot\bar{\mathbf{s}}_2$.
- The verifier checks shortness $\|\bar{\mathbf{z}}_1\|_2 \le s_1\sqrt{2m_1 d}$, $\|\bar{\mathbf{z}}_2\|_2 \le s_2\sqrt{2m_2 d}$, and

$$\overline{\mathbf{w}} + \mathbf{c}\cdot\bar{\mathbf{t}}_A = \mathbf{A}_1\bar{\mathbf{z}}_1 + \mathbf{A}_2\bar{\mathbf{z}}_2.$$

For MSIS hardness reduction for binding, the concrete parameter instantiated is $B = 8\eta\sqrt{s_1^2(2m_1 d) + s_2^2(2m_2 d)}$. Similar to eqs. (4.10) and (4.21), supposing we have a pair of accepting $(\overline{\mathbf{w}}, \bar{\mathbf{v}}, \mathbf{c}, \bar{\mathbf{z}}_1, \bar{\mathbf{z}}_2)$ and $(\overline{\mathbf{w}}, \bar{\mathbf{v}}, \mathbf{c}', \bar{\mathbf{z}}_1', \bar{\mathbf{z}}_2')$, then let $\Delta\mathbf{c} := \mathbf{c} - \mathbf{c}'$, $\Delta\bar{\mathbf{z}}_1 := \bar{\mathbf{z}}_1 - \bar{\mathbf{z}}_1'$, and $\Delta\bar{\mathbf{z}}_1 := \bar{\mathbf{z}}_2 - \bar{\mathbf{z}}_2'$, an extracted relaxed opening has $(\Delta\mathbf{c}, \Delta\bar{\mathbf{z}}_1, \Delta\bar{\mathbf{z}}_2, \overline{\mathbf{m}})$, such that

$$\Delta\mathbf{c}\begin{bmatrix} \bar{\mathbf{t}}_A \\ \bar{\mathbf{t}}_B \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{0} \end{bmatrix} \Delta\bar{\mathbf{z}}_1 + \begin{bmatrix} \mathbf{A}_2 \\ \mathbf{B} \end{bmatrix} \Delta\bar{\mathbf{z}}_2 + \Delta\mathbf{c}\begin{bmatrix} \bar{\mathbf{0}} \\ \overline{\mathbf{m}} \end{bmatrix}, \tag{4.57}$$

where the term $\Delta\mathbf{c}\cdot\overline{\mathbf{m}} = \Delta\mathbf{c}\cdot\bar{\mathbf{t}}_B - \mathbf{B}\Delta\bar{\mathbf{z}}_2$. Given relaxed openings $(\Delta\mathbf{c}, \Delta\bar{\mathbf{z}}_1, \Delta\bar{\mathbf{z}}_2, \overline{\mathbf{m}})$ and $(\Delta\mathbf{c}', \Delta\bar{\mathbf{z}}_1', \Delta\bar{\mathbf{z}}_2', \overline{\mathbf{m}}')$, that at least one of $\Delta\mathbf{c}'\Delta\bar{\mathbf{z}}_1 \ne \Delta\mathbf{c}\Delta\bar{\mathbf{z}}_1'$ and $\overline{\mathbf{m}} \ne \overline{\mathbf{m}}'$ is true, then

$$\bar{\mathbf{0}} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{0} & \mathbf{B} \end{bmatrix} \left(\Delta\mathbf{c}' \cdot \begin{bmatrix} \Delta\bar{\mathbf{z}}_1 \\ \Delta\bar{\mathbf{z}}_2 \end{bmatrix} - \Delta\mathbf{c} \cdot \begin{bmatrix} \Delta\bar{\mathbf{z}}_1' \\ \Delta\bar{\mathbf{z}}_2' \end{bmatrix} \right) + \Delta\mathbf{c}\Delta\mathbf{c}' \begin{bmatrix} \bar{\mathbf{0}} \\ \overline{\mathbf{m}} - \overline{\mathbf{m}}' \end{bmatrix},$$

which gives a solution to the MSIS instance $[\mathbf{A}_1 \mid \mathbf{A}_2] \in R_q^{\mu\times(m_1+m_2)}$ in either case. Since for $\bar{\boldsymbol{\beta}} := \left[\Delta\bar{\mathbf{z}}_1^\mathsf{T} \mid \Delta\bar{\mathbf{z}}_2^\mathsf{T}\right]^\mathsf{T}$,

$$\left\|\bar{\boldsymbol{\beta}}\right\|_2 \le 2\sqrt{s_1^2(2m_1 d) + s_2^2(2m_2 d)},$$

and by lemma 4.14.2, $\|\mathrm{rot}(\mathbf{c})\|_2 \le \eta$ for a $\mathbf{c} \in C$, thus $\|\mathrm{rot}(\Delta\mathbf{c})\|_2 \le 2\eta$, which concludes the proof for B derivation.

Now consider a linear relation proof over R_q. Consider instantiating from just Ajtai commitment for $\mathbf{R}\bar{\mathbf{s}}_1 = \bar{\mathbf{u}}$.

- The prover samples $\bar{\mathbf{y}}_1 \leftarrow D_{s_1}^{m_1 d}$ and sends over $\overline{\mathbf{w}} \leftarrow \mathbf{A}_1\bar{\mathbf{y}}_1$ and $\bar{\mathbf{v}} \leftarrow \mathbf{R}\bar{\mathbf{y}}_1$.
- The verifier challenges by $\mathbf{c} \leftarrow C$ defined in eq. (4.53).
- The prover runs rejection sampling, and replies $\bar{\mathbf{z}}_1 \leftarrow \bar{\mathbf{y}}_1 + \mathbf{c}\cdot\bar{\mathbf{s}}_1$.
- The verifier checks shortness $\|\bar{\mathbf{z}}_1\|_2 \le s_1\sqrt{2m_1 d}$ and

$$\overline{\mathbf{w}} + \mathbf{c}\cdot\bar{\mathbf{t}} = \mathbf{A}_1\bar{\mathbf{z}}_1,$$
$$\bar{\mathbf{v}} + \mathbf{c}\cdot\bar{\mathbf{u}} = \mathbf{R}\bar{\mathbf{z}}_1.$$

Continuing on, consider the BDLOP commitment for $\mathbf{R}\overline{\mathbf{m}} = \overline{\mathbf{u}}$, which is basically Π_{lin}.

- The prover samples $\overline{\mathbf{y}}_2 \leftarrow D_{s_2}^{m_2 d}$ and sends over $\overline{\mathbf{w}} \leftarrow \mathbf{A}_2\overline{\mathbf{y}}_2$ and $\overline{\mathbf{v}} \leftarrow \mathbf{R}\mathbf{B}\overline{\mathbf{y}}_2$.
- The verifier challenges by $\mathbf{c} \leftarrow C$ defined in eq. (4.53).
- The prover runs rejection sampling, and replies $\overline{\mathbf{z}}_2 \leftarrow \overline{\mathbf{y}}_2 + \mathbf{c} \cdot \overline{\mathbf{s}}_2$.
- The verifier checks shortness $\|\overline{\mathbf{z}}_2\|_2 \le s_2\sqrt{2m_2 d}$ and

$$\begin{bmatrix}\overline{\mathbf{w}} \\ \overline{\mathbf{v}}\end{bmatrix} + \mathbf{c} \cdot \begin{bmatrix}\bar{\mathbf{t}}_A \\ \mathbf{R}\bar{\mathbf{t}}_B\end{bmatrix} = \begin{bmatrix}\mathbf{A}_2 \\ \mathbf{R}\mathbf{B}\end{bmatrix}\overline{\mathbf{z}}_2 + \mathbf{c} \cdot \begin{bmatrix}\overline{\mathbf{0}} \\ \mathbf{u}\end{bmatrix}.$$

Hence, we can somehow merge them together for the statement $\mathbf{R}_1\overline{\mathbf{s}}_1 + \mathbf{R}_2\overline{\mathbf{m}} = \overline{\mathbf{u}}$ by

- The prover samples $\overline{\mathbf{y}}_1 \leftarrow D_{s_1}^{m_1 d}$, $\overline{\mathbf{y}}_2 \leftarrow D_{s_2}^{m_2 d}$ and sends over $\overline{\mathbf{w}} \leftarrow \mathbf{A}_1\overline{\mathbf{y}}_1 + \mathbf{A}_2\overline{\mathbf{y}}_2$ and $\overline{\mathbf{v}} \leftarrow \mathbf{R}_2\mathbf{B}\overline{\mathbf{y}}_2 - \mathbf{R}_1\overline{\mathbf{y}}_1$.
- The verifier challenges by $\mathbf{c} \leftarrow C$ defined in eq. (4.53).
- The prover runs rejection sampling, and replies $\overline{\mathbf{z}}_1 \leftarrow \overline{\mathbf{y}}_1 + \mathbf{c} \cdot \overline{\mathbf{s}}_1$ and $\overline{\mathbf{z}}_2 \leftarrow \overline{\mathbf{y}}_2 + \mathbf{c} \cdot \overline{\mathbf{s}}_2$.
- The verifier checks shortness $\|\overline{\mathbf{z}}_1\|_2 \le s_1\sqrt{2m_1 d}$, $\|\overline{\mathbf{z}}_2\|_2 \le s_2\sqrt{2m_2 d}$, and

$$\begin{bmatrix}\overline{\mathbf{w}} \\ \overline{\mathbf{v}}\end{bmatrix} + \mathbf{c} \cdot \begin{bmatrix}\bar{\mathbf{t}}_A \\ \mathbf{R}_2\bar{\mathbf{t}}_B\end{bmatrix} + \begin{bmatrix}\overline{\mathbf{0}} \\ \mathbf{R}_1\overline{\mathbf{z}}_1\end{bmatrix} = \begin{bmatrix}\mathbf{A}_1 \\ \mathbf{0}\end{bmatrix}\overline{\mathbf{z}}_1 + \begin{bmatrix}\mathbf{A}_2 \\ \mathbf{R}_2\mathbf{B}\end{bmatrix}\overline{\mathbf{z}}_2 + \mathbf{c} \cdot \begin{bmatrix}\overline{\mathbf{0}} \\ \mathbf{u}\end{bmatrix}.$$

Hence, a relaxed opening $(\Delta\mathbf{c}, \Delta\overline{\mathbf{z}}_1, \Delta\overline{\mathbf{z}}_2, \overline{\mathbf{m}})$ satisfies eq. (4.57), $\|\Delta\overline{\mathbf{z}}_1\|_2 \le 2s_1\sqrt{2m_1 d}$, $\|\Delta\overline{\mathbf{z}}_2\|_2 \le 2s_2\sqrt{2m_2 d}$, and

$$\mathbf{R}_1\Delta\overline{\mathbf{z}}_1 + \mathbf{R}_2\left(\Delta\mathbf{c} \cdot \overline{\mathbf{m}}\right) = \Delta\mathbf{c} \cdot \overline{\mathbf{u}}.$$

Now from the linear relation proof over R_q, consider the linear relation proof over $\mathbb{Z}_q$. Consider the inner product performed over the coefficients of polynomial ring via automorphism σ_{-1}, that the constant coefficient of $\sigma_{-1}(\mathbf{a})\mathbf{b}$ is the inner product $\langle \overline{a}, \overline{b} \rangle$ of coefficients in $\mathbf{a} = \sum a_i X^i$ and $\mathbf{b} = \sum b_i X^i$, allowing for encoding the Z_q linear relation over the coefficients: $\langle \sigma_{-1}(\overline{\mathbf{r}}_1), \overline{\mathbf{s}}_1 \rangle + \langle \sigma_{-1}(\overline{\mathbf{r}}_2), \overline{\mathbf{m}} \rangle = \mathbf{u}$. Augmenting the BDLOP message for maskings like [ENS20] by

$$\begin{bmatrix}\bar{\mathbf{t}}_A \\ \bar{\mathbf{t}}_B \\ \bar{\mathbf{t}}_g\end{bmatrix} = \begin{bmatrix}\mathbf{A}_1 \\ \mathbf{0} \\ \mathbf{0}\end{bmatrix}\overline{\mathbf{s}}_1 + \begin{bmatrix}\mathbf{A}_2 \\ \mathbf{B} \\ \mathbf{B}_g\end{bmatrix}\overline{\mathbf{s}}_2 + \begin{bmatrix}\overline{\mathbf{0}} \\ \overline{\mathbf{m}} \\ \overline{\mathbf{g}}\end{bmatrix} \in R_q^{\mu+L+\rho},$$

where $\overline{\mathbf{g}} \in R_q^{\rho}$ are the masking terms. Start by proving with the Π_{lin} for R_q in ABDLOP, with a single repetition:

- The prover commits $\mathbf{g} \xleftarrow{\mathrm{R}} \{R_q : g_0 = 0\}$ by $\mathbf{t}_g \leftarrow \langle \overline{\mathbf{b}}_g, \overline{\mathbf{s}}_2 \rangle + \mathbf{g}$, and sends $\mathbf{t}_g$ over.
- The verifier challenges by sampling $\gamma \xleftarrow{\mathrm{R}} \mathbb{Z}_q$.
- The prover replies $\mathbf{u}' \leftarrow \mathbf{g} + \gamma \cdot \mathbf{u}$, and starts the Π_{lin} with ABDLOP for $\langle \gamma\sigma_{-1}(\overline{\mathbf{r}}_1), \overline{\mathbf{s}}_1 \rangle + \langle \gamma\sigma_{-1}(\overline{\mathbf{r}}_2), \overline{\mathbf{m}} \rangle + \mathbf{g} = \mathbf{u}'$ by
  - The prover samples $\overline{\mathbf{y}}_1 \leftarrow D_{s_1}^{m_1 d}$, $\overline{\mathbf{y}}_2 \leftarrow D_{s_2}^{m_2 d}$ and replies

$$\begin{aligned}\overline{\mathbf{w}} &\leftarrow \mathbf{A}_1\overline{\mathbf{y}}_1 + \mathbf{A}_2\overline{\mathbf{y}}_2, \\ \mathbf{v} &\leftarrow \left(\gamma\sigma_{-1}(\overline{\mathbf{r}}_2)^{\mathsf{T}}\mathbf{B} + \overline{\mathbf{b}}_g^{\mathsf{T}}\right)\overline{\mathbf{y}}_2 - \gamma\sigma_{-1}(\overline{\mathbf{r}}_1)^{\mathsf{T}}\overline{\mathbf{y}}_1.\end{aligned}$$

  - The verifier challenges by $\mathbf{c} \leftarrow C$ defined in eq. (4.53).
  - The prover runs rejection sampling, and replies $\overline{\mathbf{z}}_1 \leftarrow \overline{\mathbf{y}}_1 + \mathbf{c} \cdot \overline{\mathbf{s}}_1$ and $\overline{\mathbf{z}}_2 \leftarrow \overline{\mathbf{y}}_2 + \mathbf{c} \cdot \overline{\mathbf{s}}_2$.
- The verifier checks shortness $\|\overline{\mathbf{z}}_1\|_2 \le s_1\sqrt{2m_1 d}$, $\|\overline{\mathbf{z}}_2\|_2 \le s_2\sqrt{2m_2 d}$,

$$\begin{bmatrix}\overline{\mathbf{w}} \\ \mathbf{v}\end{bmatrix} + \mathbf{c} \cdot \begin{bmatrix}\bar{\mathbf{t}}_A \\ \gamma\sigma_{-1}(\overline{\mathbf{r}}_2)^{\mathsf{T}}\bar{\mathbf{t}}_B + \mathbf{t}_g\end{bmatrix} + \begin{bmatrix}\overline{\mathbf{0}} \\ \gamma\sigma_{-1}(\overline{\mathbf{r}}_1)^{\mathsf{T}}\overline{\mathbf{z}}_1\end{bmatrix} = \begin{bmatrix}\mathbf{A}_1 \\ \mathbf{0}\end{bmatrix}\overline{\mathbf{z}}_1 + \begin{bmatrix}\mathbf{A}_2 \\ \gamma\sigma_{-1}(\overline{\mathbf{r}}_2)^{\mathsf{T}}\mathbf{B} + \overline{\mathbf{b}}_g^{\mathsf{T}}\end{bmatrix}\overline{\mathbf{z}}_2 + \mathbf{c} \cdot \begin{bmatrix}\overline{\mathbf{0}} \\ \mathbf{u}'\end{bmatrix},$$

  and the constant term satisfies $u'_0 = \gamma u_0$.

For multiple linear relations, apply random linear combination over $\mathbb{Z}_q$ under the same mask $\mathbf{g}$, with soundness q_1^{-1}, say any CRT slot under the split by $\mathbb{Z}_q = \prod \mathbb{Z}_{q_i}$ is wrong, the uniformly random challenges that are independently sampled can fix up that slot with probability at most q_1^{-1} by Schwartz-Zippel. The random linear combination over multiple linear relations is similar to the style in [ALS20] for multiple product relations.

When repetitions are greater than once, a straightforward way is to make ρ exactly the number of repetitions. But a closer observation would be, given that R_q at least has 2 automorphisms, we can use the automorphism filtering from [ENS20], such that a masking can hold 2 repetitions by $\mathbf{g} \xleftarrow{R} \{R_q : g_0 = g_1 = 0\}$.

We show by 1 linear relation with 2 γ challenges setting:

- The prover commits $\mathbf{g} \xleftarrow{R} \{R_q : g_0 = g_1 = 0\}$ by $\mathbf{t}_g \leftarrow \langle \overline{\mathbf{b}}_g, \overline{\mathbf{s}}_2 \rangle + \mathbf{g}$, and sends $\mathbf{t}_g$ over.
- The verifier challenges by sampling $\gamma_0, \gamma_1 \xleftarrow{R} \mathbb{Z}_q$.
- The prover replies

$$\mathbf{u}' \leftarrow \mathbf{g} + \frac{\gamma_0 + \gamma_1 X}{2} \cdot F_2(\mathbf{u}),$$

and starts the Π_{lin} with ABDLOP for

$$\mathbf{g} + \frac{\gamma_0 + \gamma_1 X}{2} \cdot F_2(\langle \sigma_{-1}(\overline{\mathbf{r}}_1), \overline{\mathbf{s}}_1 \rangle + \langle \sigma_{-1}(\overline{\mathbf{r}}_2), \overline{\mathbf{m}} \rangle),$$

where F_2 is defined in eq. (4.36), by

  - The prover samples $\overline{\mathbf{y}}_{1,0}, \overline{\mathbf{y}}_{1,1} \leftarrow D_{s_1}^{m_1 d}$, $\overline{\mathbf{y}}_{2,0}, \overline{\mathbf{y}}_{2,1} \leftarrow D_{s_2}^{m_2 d}$, computes

$$\boldsymbol{\alpha} \leftarrow 2^{-1} \sum_{i \in [0,1]} \sigma_{d+1}^{-i} \left(\sigma_{-1}(\overline{\mathbf{r}}_2)^\intercal \mathbf{B} \overline{\mathbf{y}}_{2,i} - \sigma_{-1}(\overline{\mathbf{r}}_1)^\intercal \overline{\mathbf{y}}_{1,i} \right),$$

    and replies

$$\begin{aligned} \overline{\mathbf{w}}_i &\leftarrow \mathbf{A}_1 \overline{\mathbf{y}}_{1,i} + \mathbf{A}_2 \overline{\mathbf{y}}_{2,i}, \\ \mathbf{v}_i &\leftarrow (\gamma_0 + \gamma_1 X) \sigma_{d+1}^i(\boldsymbol{\alpha}) + \overline{\mathbf{b}}_g^\intercal \overline{\mathbf{y}}_{2,i}, \end{aligned}$$

    where σ_{d+1} is due to lemma 4.14.1 on R_q splitting.
  - The verifier challenges by $\mathbf{c} \leftarrow C$ defined in eq. (4.53).
  - The prover runs rejection sampling, and replies $\overline{\mathbf{z}}_{1,i} \leftarrow \overline{\mathbf{y}}_{1,i} + \sigma_{d+1}^i(\mathbf{c}) \cdot \overline{\mathbf{s}}_1$ and $\overline{\mathbf{z}}_{2,i} \leftarrow \overline{\mathbf{y}}_{2,i} + \sigma_{d+1}^i(\mathbf{c}) \cdot \overline{\mathbf{s}}_2$.
- The verifier checks shortness $\|\overline{\mathbf{z}}_{1,i}\|_2 \le s_1 \sqrt{2 m_1 d}$, $\|\overline{\mathbf{z}}_{2,i}\|_2 \le s_2 \sqrt{2 m_2 d}$, automorphism filtering checks for Π_{lin}:

$$\begin{aligned} &\begin{bmatrix} \mathbf{A}_1 \\ \mathbf{0} \\ -\frac{\gamma_0 + \gamma_1 X}{2} \sigma_{-1}(\overline{\mathbf{r}}_1)^\intercal \end{bmatrix} \overline{\mathbf{z}}_{1,i} + \begin{bmatrix} \mathbf{A}_2 \\ \mathbf{0} \\ \frac{\gamma_0 + \gamma_1 X}{2} \sigma_{-1}(\overline{\mathbf{r}}_2)^\intercal \mathbf{B} + \overline{\mathbf{b}}_g^\intercal \end{bmatrix} \overline{\mathbf{z}}_{2,i} \\ &\quad + \begin{bmatrix} \mathbf{0} \\ \sigma_{d+1}^{-1}(\mathbf{A}_1) \\ -\frac{\gamma_0 + \gamma_1 X}{2} \sigma_{d+1}^{-1}\left(\sigma_{-1}(\overline{\mathbf{r}}_1)^\intercal\right) \end{bmatrix} \sigma_{d+1}^{-1}(\overline{\mathbf{z}}_{1,1-i}) + \begin{bmatrix} \mathbf{0} \\ \sigma_{d+1}^{-1}(\mathbf{A}_2) \\ \frac{\gamma_0 + \gamma_1 X}{2} \sigma_{d+1}^{-1}\left(\sigma_{-1}(\overline{\mathbf{r}}_2)^\intercal \mathbf{B}\right) \end{bmatrix} \sigma_{d+1}^{-1}(\overline{\mathbf{z}}_{2,1-i}) + \sigma_{d+1}^i(\mathbf{c}) \cdot \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{u}' \end{bmatrix} \\ &\qquad = \begin{bmatrix} \overline{\mathbf{w}}_i \\ \sigma_{d+1}^{-1}(\overline{\mathbf{w}}_{1-i}) \\ \mathbf{v}_i \end{bmatrix} + \sigma_{d+1}^i(\mathbf{c}) \begin{bmatrix} \overline{\mathbf{t}}_A \\ \sigma_{d+1}^{-1}(\overline{\mathbf{t}}_A) \\ \frac{\gamma_0 + \gamma_1 X}{2} F_2\left(\sigma_{-1}(\overline{\mathbf{r}}_2)^\intercal \overline{\mathbf{t}}_B\right) + \mathbf{t}_g \end{bmatrix} \end{aligned}$$

holds for each $i \in [0, 1]$, and the constant term satisfies $u_i' = \gamma_i u_0$.

Quadratic Relation Proof. Now also consider for quadratic relation proof over ABDLOP, extending from [ALS20]. Considering the quadratic relation proof over BDLOP by the computationally indistinguishable masking in eq. (4.26), the quadratic relation can be extended to arbitrary pair of subset sums of the BDLOP messages by

$$\begin{aligned}\bar{\mathbf{f}}^\mathsf{T}\mathbf{R}\bar{\mathbf{f}} &= \left(\mathbf{c}\overline{\mathbf{m}} - \mathbf{B}\overline{\mathbf{y}}_2\right)^\mathsf{T}\mathbf{R}\left(\mathbf{c}\overline{\mathbf{m}} - \mathbf{B}\overline{\mathbf{y}}_2\right) \\ &= \mathbf{c}^2\overline{\mathbf{m}}^\mathsf{T}\mathbf{R}\overline{\mathbf{m}} - \mathbf{c}\left(\overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\overline{\mathbf{m}} + \overline{\mathbf{m}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}}\right) + \overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}},\end{aligned} \tag{4.58}$$

where $\bar{\mathbf{f}} = \mathbf{c}\overline{\mathbf{m}} - \mathbf{B}\overline{\mathbf{y}}_2$ the computationally indistinguishible masked opening, and $\overline{\mathbf{y}} = \mathbf{B}\overline{\mathbf{y}}_2$. Continuing on, consider the linear terms and the constant terms, we have the relation

$$\overline{\mathbf{m}}^\mathsf{T}\mathbf{R}\overline{\mathbf{m}} + \bar{\mathbf{r}}_1^\mathsf{T}\overline{\mathbf{m}} + \mathbf{r}_0 = \mathbf{0}, \tag{4.59}$$

the corresponding masked opening should give

$$\begin{aligned}\bar{\mathbf{f}}^\mathsf{T}\mathbf{R}\bar{\mathbf{f}} + \mathbf{c}\bar{\mathbf{r}}_1^\mathsf{T}\bar{\mathbf{f}} + \mathbf{c}^2\mathbf{r}_0 &= \left(\mathbf{c}\overline{\mathbf{m}} - \mathbf{B}\overline{\mathbf{y}}_2\right)^\mathsf{T}\mathbf{R}\left(\mathbf{c}\overline{\mathbf{m}} - \mathbf{B}\overline{\mathbf{y}}_2\right) + \mathbf{c}\bar{\mathbf{r}}_1^\mathsf{T}\left(\mathbf{c}\overline{\mathbf{m}} - \mathbf{B}\overline{\mathbf{y}}_2\right) + \mathbf{c}^2\mathbf{r}_0 \\ &= \mathbf{c}^2\left(\overline{\mathbf{m}}^\mathsf{T}\mathbf{R}\overline{\mathbf{m}} + \bar{\mathbf{r}}_1^\mathsf{T}\overline{\mathbf{m}} + \mathbf{r}_0\right) - \mathbf{c}\left(\overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\overline{\mathbf{m}} + \overline{\mathbf{m}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}} + \bar{\mathbf{r}}_1^\mathsf{T}\overline{\mathbf{y}}\right) + \overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}},\end{aligned} \tag{4.60}$$

such that the degree-1 term $\mathbf{g}_0 = -\left(\overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\overline{\mathbf{m}} + \overline{\mathbf{m}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}} + \bar{\mathbf{r}}_1^\mathsf{T}\overline{\mathbf{y}}\right)$ is the additional garbage term to commit.

Now we consider the Ajtai part. Instead of the computationally indistinguishable opening $\bar{\mathbf{f}} = \mathbf{c}\overline{\mathbf{m}} - \overline{\mathbf{y}}$, we have statistically indistinguishable opening $\bar{\mathbf{z}} = \mathbf{y}_1 + \mathbf{c}\bar{\mathbf{s}}_1$. Hence, the computation structure in eqs. (4.58) and (4.60) remains if we combine the masks by

$$\overline{\mathbf{y}} = \begin{bmatrix}-\overline{\mathbf{y}}_1 \\ \mathbf{B}\overline{\mathbf{y}}_2\end{bmatrix}, \quad \bar{\mathbf{s}} = \begin{bmatrix}\bar{\mathbf{s}}_1 \\ \mathbf{m}\end{bmatrix}, \tag{4.61}$$

and $\bar{\mathbf{f}} = \mathbf{c}\cdot\bar{\mathbf{s}} - \overline{\mathbf{y}}$, then eq. (4.60) becomes

$$\bar{\mathbf{f}}^\mathsf{T}\mathbf{R}\bar{\mathbf{f}} + \mathbf{c}\bar{\mathbf{r}}_1^\mathsf{T}\bar{\mathbf{f}} + \mathbf{c}^2\mathbf{r}_0 = \mathbf{c}^2\left(\bar{\mathbf{s}}^\mathsf{T}\mathbf{R}\bar{\mathbf{s}} + \bar{\mathbf{r}}_1^\mathsf{T}\bar{\mathbf{s}} + \mathbf{r}_0\right) - \mathbf{c}\left(\overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\bar{\mathbf{s}} + \bar{\mathbf{s}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}} + \bar{\mathbf{r}}_1^\mathsf{T}\overline{\mathbf{y}}\right) + \overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}},$$

such that the degree-1 term $\mathbf{g}_0 = -\left(\overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\bar{\mathbf{s}} + \bar{\mathbf{s}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}} + \bar{\mathbf{r}}_1^\mathsf{T}\overline{\mathbf{y}}\right)$ is the additional garbage term to commit.

Hence, a prototype for quadratic relation over ABDLOP commitment scheme follows:

- The prover samples $\overline{\mathbf{y}}_1 \leftarrow D_{s_1}^{m_1 d}$, $\overline{\mathbf{y}}_2 \leftarrow D_{s_2}^{m_2 d}$, commits to $\mathbf{g}_0$ by $\mathbf{t}_g \leftarrow \bar{\mathbf{b}}_g^\mathsf{T}\bar{\mathbf{s}}_2 + \mathbf{g}_0$, and replies

$$\begin{aligned}\overline{\mathbf{w}} &\leftarrow \mathbf{A}_1\overline{\mathbf{y}}_1 + \mathbf{A}_2\overline{\mathbf{y}}_2, \\ \mathbf{v} &\leftarrow \overline{\mathbf{y}}^\mathsf{T}\mathbf{R}\overline{\mathbf{y}} + \bar{\mathbf{b}}_g^\mathsf{T}\overline{\mathbf{y}}_2.\end{aligned}$$

- The verifier challenges by $\mathbf{c} \leftarrow C$ defined in eq. (4.53).
- The prover runs rejection sampling, and replies $\bar{\mathbf{z}}_1 \leftarrow \overline{\mathbf{y}}_1 + \mathbf{c}\cdot\bar{\mathbf{s}}_1$ and $\bar{\mathbf{z}}_2 \leftarrow \overline{\mathbf{y}}_2 + \mathbf{c}\cdot\bar{\mathbf{s}}_2$.
- The verifier checks shortness $\|\bar{\mathbf{z}}_1\|_2 \le s_1\sqrt{2m_1 d}$, $\|\bar{\mathbf{z}}_2\|_2 \le s_2\sqrt{2m_2 d}$, derives $\mathbf{f}_g \leftarrow \bar{\mathbf{b}}_g^\mathsf{T}\bar{\mathbf{z}}_2 - \mathbf{c}\mathbf{t}_g$ and

$$\bar{\mathbf{f}} \leftarrow \begin{bmatrix}\bar{\mathbf{z}}_1 \\ \mathbf{c}\bar{\mathbf{t}}_B - \mathbf{B}\bar{\mathbf{z}}_2\end{bmatrix},$$

and checks by

$$\begin{aligned}\mathbf{A}_1\bar{\mathbf{z}}_1 + \mathbf{A}_2\bar{\mathbf{z}}_2 &= \overline{\mathbf{w}} + \mathbf{c}\bar{\mathbf{t}}_A \\ \bar{\mathbf{f}}^\mathsf{T}\mathbf{R}\bar{\mathbf{f}} + \mathbf{c}\bar{\mathbf{r}}_1^\mathsf{T}\bar{\mathbf{f}} + \mathbf{c}^2\mathbf{r}_0 + \mathbf{f}_g &= \mathbf{v}.\end{aligned}$$

Addtionally, noticing the challenges from C has $\sigma(\mathbf{c}) = \mathbf{c}$, then the quadratic relation can be extended over all $\bar{\mathbf{s}}$'s automorphisms, that the message got expanded to

$$\bar{\mathbf{s}} = \begin{bmatrix} \{\sigma^i(\bar{\mathbf{s}}_1)\}_{i\in[0,k-1]} \\ \{\sigma^i(\overline{\mathbf{m}})\}_{i\in[0,k-1]} \end{bmatrix}, \tag{4.62}$$

as each opening has $\sigma^j(\bar{\mathbf{z}}_i) = \sigma^j(\bar{\mathbf{y}}_i) + \mathbf{c} \cdot \sigma^j(\bar{\mathbf{s}}_i)$ by self-automorphism $\mathbf{c} = \sigma(\mathbf{c})$, which is similar to the technique for automorphism filtering in eq. (4.40). Then eq. (4.61) can be changed to

$$\bar{\mathbf{y}} = \begin{bmatrix} \{\sigma^i(-\bar{\mathbf{y}}_1)\}_{i\in[0,k-1]} \\ \{\sigma^i(\mathbf{B}\bar{\mathbf{y}}_2)\}_{i\in[0,k-1]} \end{bmatrix}, \quad \bar{\mathbf{s}} = \begin{bmatrix} \{\sigma^i(\bar{\mathbf{s}}_1)\}_{i\in[0,k-1]} \\ \{\sigma^i(\overline{\mathbf{m}})\}_{i\in[0,k-1]} \end{bmatrix}. \tag{4.63}$$

If the prover is honest, the convincing probabilty is approximately

$$\frac{1}{M_1 M_2} = \exp\left(\frac{14}{\gamma_1} + \frac{14}{\gamma_2} + \frac{1}{2\gamma_1^2} + \frac{1}{2\gamma_2^2}\right)^{-1}$$

by eq. (4.55), as we are just using [Lyu12] rejection sampling.

The knowledge soundness analysis is similar to the heavy row argument lemma 4.8.8, but the extractor behavior is described and developed in [ACK21]. We begin by the expected draws for negative hypergeometric distribution.

Lemma 4.14.4. *Given a bin with N balls of which M are marked. Then the attempts to draw uniformly at random from the bin without replacement until $k \le M$ marked balls are drawn is distributed according to the negative hypergeometric distribution. The expected number of draws is $k(N+1)/(M+1)$.*

Proof. The sequence of draws without replacement can be viewed as a permutation of N balls, and we are interested in the number of draws up to the first k marked ones. For an unmarked ball, it can be in one of the $M+1$ positions separated by M marked ones, and there are k of them before the k^{th} marked one. By the linearity of expectation, the expected number of draws is

$$k + (N - M) \cdot \frac{k}{M+1} = \frac{k(N+1)}{M+1}.$$

□

An immediate application is, let $\mathbf{H} \in \{0,1\}^{R\times N}$ be the binary matrix for R rows indexed by the prover's randomness, and the N columns indexed by verifier challenges, then let the i^{th} row has accepting dencity ε_i, supposing the first draw over the row is accepting, then drawing for the next 2 acceptance is expected to be $2/\varepsilon_i$ by lemma 4.14.4. Supposing the extractor aborts if the first draw is rejected, or otherwise it samples uniformly at random the remaining entries of the row without replacement until three acceptances or the row is exhausted, and supposing there are at least 3 acceptance on the row, then

$$\mathbb{E}\left[Q \mid i^{\text{th}} \text{ row}\right] = 1 \cdot (1 - \varepsilon_i) + (1 + 2/\varepsilon_i) \cdot \varepsilon_i = 3,$$

and supposing there are less than 3 acceptence on the row, then

$$\mathbb{E}\left[Q \mid i^{\text{th}} \text{ row}\right] = 1 \cdot (1 - \varepsilon_i) + N \cdot \varepsilon_i \le 3.$$

Averaging over the prover randomness, we thus have

$$\mathbb{E}[Q] = \mathbb{E}\left[\mathbb{E}\left[Q \mid i^{\text{th}} \text{ row}\right]\right] \le 3, \tag{4.64}$$

where the success probability is

$$\sum_{k\in[3,N]} \frac{k}{N} \cdot \delta_k = \sum_{k\in[1,N]} \frac{k}{N} \cdot \delta_k - \frac{\delta_1}{N} - \frac{2\delta_2}{N} = \varepsilon - \frac{\delta_1 + 2\delta_2}{N}, \tag{4.65}$$

where δ_k is the fraction of rows that have k acceptances, ε is the fraction of acceptance in $\mathbf{H}$, and eq. (4.65) is at least $\varepsilon - 3/N$ by a coarse analysis, and $\varepsilon - 2/N$ by a finer one. Hence, if checks on an entry of $\mathbf{H}$ takes T, and $\varepsilon > 2/N$, then extracting three accepting transcripts by restarting over fresh prover randomness, is at most expected $3T/(\varepsilon - 2/N)$.

Now, in expected $3T/(\varepsilon - 2/N)$ time, we have 3 accepting transcripts forked from the same prover randomenss, hence the same first prover message, by $(\overline{\mathbf{w}}, \mathbf{v}, \mathbf{t}_g, \mathbf{c}_i, \overline{\mathbf{z}}_{1,i}, \overline{\mathbf{z}}_{2,i})$ for $i \in [0, 2]$.

Let $\Delta\mathbf{c} \leftarrow \mathbf{c}_0 - \mathbf{c}_1$, $\Delta\overline{\mathbf{z}}_1 \leftarrow \overline{\mathbf{z}}_{1,0} - \overline{\mathbf{z}}_{1,1}$, $\Delta\overline{\mathbf{z}}_2 \leftarrow \overline{\mathbf{z}}_{2,0} - \overline{\mathbf{z}}_{2,1}$, and $\bar{\mathbf{s}}_1 \leftarrow \Delta\overline{\mathbf{z}}_1 \cdot (\Delta\mathbf{c})^{-1}$, $\bar{\mathbf{s}}_2 \leftarrow \Delta\overline{\mathbf{z}}_2 \cdot (\Delta\mathbf{c})^{-1}$. Further, we have $\overline{\mathbf{m}} \leftarrow \bar{\mathbf{t}}_B - \mathbf{B}\bar{\mathbf{s}}_2$ and $\mathbf{g}_0 \leftarrow \mathbf{t}_g - \overline{\mathbf{b}}_g^{\mathsf{T}}\bar{\mathbf{s}}_2$, then

$$\begin{bmatrix} \bar{\mathbf{t}}_A \\ \bar{\mathbf{t}}_B \\ \bar{\mathbf{t}}_g \end{bmatrix} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \bar{\mathbf{s}}_1 + \begin{bmatrix} \mathbf{A}_2 \\ \mathbf{B} \\ \overline{\mathbf{b}}_g^{\mathsf{T}} \end{bmatrix} \bar{\mathbf{s}}_2 + \begin{bmatrix} \bar{\mathbf{0}} \\ \overline{\mathbf{m}} \\ \mathbf{g}_0 \end{bmatrix},$$

and $\|\Delta\mathbf{c}\|_\infty \le 2\kappa$ by eq. (4.53), $\|\Delta\overline{\mathbf{z}}_i\|_2 \le 2s_i\sqrt{2m_i d}$ for $i \in [1, 2]$.

Continuing on, we can have $\overline{\mathbf{y}}_1 \leftarrow \overline{\mathbf{z}}_{1,0} - \mathbf{c}_0 \cdot \bar{\mathbf{s}}_1$, $\overline{\mathbf{y}}_2 \leftarrow \overline{\mathbf{z}}_{2,0} - \mathbf{c}_0 \cdot \bar{\mathbf{s}}_2$, and under the same interpolation,

$$\begin{aligned} \overline{\mathbf{z}}_{1,0} - \mathbf{c}_0 \cdot \bar{\mathbf{s}}_1 &= \overline{\mathbf{z}}_{1,1} - \mathbf{c}_1 \cdot \bar{\mathbf{s}}_1, \\ \overline{\mathbf{z}}_{2,0} - \mathbf{c}_0 \cdot \bar{\mathbf{s}}_2 &= \overline{\mathbf{z}}_{2,1} - \mathbf{c}_1 \cdot \bar{\mathbf{s}}_2. \end{aligned}$$

Applying the derived $\bar{\mathbf{s}}_1$, $\bar{\mathbf{s}}_2$, and $\mathbf{c}_2$, against $\overline{\mathbf{z}}_{1,2}$ and $\overline{\mathbf{z}}_{2,2}$, we have either $\overline{\mathbf{y}}_1$ and $\overline{\mathbf{y}}_2$ being the masking randomness, or we find a $\mathsf{MSIS}_{n,m_1+m_2,B}$ solution for $[\mathbf{A}_1\mathbf{A}_2]$, where B was defined near eq. (4.57), that for $i \in [1, 2]$,

$$\overline{\mathbf{u}}_i \leftarrow \Delta\mathbf{c}\left(\overline{\mathbf{z}}_{i,2} - \mathbf{c}_2\bar{\mathbf{s}}_i - \overline{\mathbf{y}}_i\right) = \Delta\mathbf{c}\left(\overline{\mathbf{z}}_{i,2} - \overline{\mathbf{z}}_{i,0}\right) + (\mathbf{c}_0 - \mathbf{c}_2)\,\Delta\overline{\mathbf{z}}_i$$

solves $[\mathbf{A}_1\mathbf{A}_2]$ by $\mathbf{A}_1\overline{\mathbf{u}}_1 + \mathbf{A}_2\overline{\mathbf{u}}_2 = \bar{\mathbf{0}}$.

Onwards assuming the MSIS is respected, then derive eq. (4.63), $\mathbf{g}_0' = -\left(\overline{\mathbf{y}}^{\mathsf{T}}\mathbf{R}\bar{\mathbf{s}} + \bar{\mathbf{s}}^{\mathsf{T}}\mathbf{R}\overline{\mathbf{y}} + \bar{\mathbf{r}}_1^{\mathsf{T}}\overline{\mathbf{y}}\right)$, such that

$$\begin{bmatrix} 1 & \mathbf{c}_0 & \mathbf{c}_0^2 \\ 1 & \mathbf{c}_1 & \mathbf{c}_1^2 \\ 1 & \mathbf{c}_2 & \mathbf{c}_2^2 \end{bmatrix} \begin{bmatrix} \overline{\mathbf{y}}^{\mathsf{T}}\mathbf{R}\overline{\mathbf{y}} + \overline{\mathbf{b}}_g^{\mathsf{T}}\overline{\mathbf{y}}_2 - \mathbf{v} \\ \mathbf{g}_0' - \mathbf{g}_0 \\ \bar{\mathbf{s}}^{\mathsf{T}}\mathbf{R}\bar{\mathbf{s}} + \bar{\mathbf{r}}_1^{\mathsf{T}}\bar{\mathbf{s}} + \mathbf{r}_0 \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

must hold for the extracted variables. Since pairwise challenge differences are invertible, the Vandermonde matrix is invertible; hence every coefficient is zero, including the claimed quadratic relation.

Many Quadratic Relations Proof. To instantiate the proof for multiple quadratc relations, we sample uniformly random $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N \xleftarrow{\text{R}} R_q$ and apply linear combination over quadratic relations defined by eq. (4.59).

For soundness analysis, it is similar to the linear combination for multiple linear relations. Recall the soundness error q_1^{-1} derived by $\gamma_1, \ldots, \gamma_\ell \xleftarrow{\text{R}} \mathbb{Z}_q$, it was due to $\mathbb{Z}_q \cong \prod \mathbb{Z}_{q_i}$, then the random linear combination over a CRT slot being fixed to 0 has probability at most q_1^{-1}. Now consider random linear combination over $\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_N \xleftarrow{\text{R}} R_q$, over some fixed $\mathbf{e}_1, \ldots, \mathbf{e}_N \in R_q$ that are not all zero, then by

$$R_q \cong \prod \left(\mathbb{F}_{q_i^{d/2}} \times \mathbb{F}_{q_i^{d/2}}\right),$$

supposing $q_i \equiv 5 \mod 8$ makes $R_{q_i} \cong \mathbb{F}_{q_i^{d/2}} \times \mathbb{F}_{q_i^{d/2}}$, then fixing a nonzero CRT slot has probability at most $q_1^{-d/2}$.

Hence, starting from the extractor that extracts for a single quadratic relation, the knowledge soundness here is $2/N + q_1^{-d/2}$, where the additional term is from some relation is not satisfying.

Consider using the extractor for one quadratic relation. After the first extraction, we have an opening satisfying

$$\bar{\mathbf{s}}^{\mathsf{T}}\left(\sum \boldsymbol{\mu}_i\mathbf{R}_i\right)\bar{\mathbf{s}} + \left(\sum \boldsymbol{\mu}_i\bar{\mathbf{r}}_{1,i}^{\mathsf{T}}\right)\bar{\mathbf{s}} + \sum \boldsymbol{\mu}_i\mathbf{r}_{0,i} = \mathbf{0}.$$

Yet, we cannot rule out the case that some subordinate quadratic relation is not satisfying.

Again, we design an extractor for multiple quadratic relations, utilizing the extractor for one quadratic relation, under the trick by [ACK21] utilizing lemma 4.14.4. For a fixed opening violating some underlying quadratic relations, it may succeed with probability at most $q_1^{-d/2}$ over random linear combination by prior discussion, without breaking MSIS.

Here, the opening is fixed independently of the sampled $\overline{\boldsymbol{\mu}}$ for random linear combination, and a malicious prover capable of producing different openings for the same commitment may instead, choose an opening according to $\overline{\boldsymbol{\mu}}$. Hence, in this case, $q_1^{-d/2}$ does not directly bound the probability of obtaining some invalid opening in one extraction.

We therefore seek a second successful extraction for the same commitment, forked from independently sampled $\overline{\boldsymbol{\mu}}$. If the two extracted relaxed openings differ, the binding reduction yields an MSIS solution; otherwise, the same openings may still violate some underlying quadratic relations. Hence, a second extraction tells malicious prover being lucky from the case of finding an MSIS solution.

The new extractor invokes the extractor for one quadratic relation, if the first extraction failed, abort immediately; otherwise, restart extraction over independently sampled $\overline{\boldsymbol{\mu}}$ until a success. Let p be the probability of a successful extraction for one quadratic relation, E be the event that the first extraction is successful, and X be the number of invocations for the extractor for one quadratic relation, then the expected number of invocations is

$$\mathbb{E}[X] = \mathbb{E}[X \mid E]\Pr[E] + \mathbb{E}[X \mid \neg E]\Pr[\neg E] = \left(1 + \frac{1}{p}\right) \cdot p + 1 \cdot (1 - p) = 2.$$

Hence, by eq. (4.64), the expected runtime of the extractor for one quadratic relation is at most $3T$, then the expected runtime of the extractor for multiple quadratic relations is at most $6T$.

For the success probability of such extractor for multiple quadratic relations, there are three disjoint events:

- E_0: Both openings are the same, and all underlying quadratic relations are satisfied.
- E_1: Both openings are the same, but some underlying quadratic relations are violated.
- E_2: The openings are distinct, or an MSIS solution on either extraction.

A prover capable of breaking binding may choose an invalid opening by $\overline{\boldsymbol{\mu}}$ in the first extraction, yet return the same opening in the second extraction. Such an execution does not give an MSIS solution and belongs to E_1. We therefore need to bound $\Pr[E_1]$ even for such provers. By the construction of the extractor for multiple quadratic relations,

$$\Pr[E_0 \vee E_1 \vee E_2] = p \geq \varepsilon - \frac{2}{N}, \tag{4.66}$$

where ε is the prover's convincing probability, and eq. (4.66) holds by, on the first successful extration, the second extraction keeps rebooting up to another successful extraction. Hence,

$$\Pr[E_0 \vee E_2] = p - \Pr[E_1] \geq \varepsilon - \frac{2}{N} - \Pr[E_1].$$

Let X_0, X_1 be the random variables for the output of the first and second extraction, and h_x be the probability of an extraction gives relaxed opening x, then $h_\perp = 1 - p \leq 1 - \varepsilon + 2/N$, and

$$\Pr[X_1 = x \mid X_0 = x] = \sum_{r \geq 0} (1-p)^r h_x = \frac{h_x}{p}.$$

Therefore, summing overall openings that violates some underlynig quadratic relations, we have

$$\Pr[E_1] = \sum_x \Pr[X_1 = x \mid X_0 = x]\Pr[X_0 = x] = \sum_x \frac{h_x}{p} \cdot h_x \leq \frac{q_1^{-d/2}}{p} \sum_x h_x \leq q_1^{-d/2},$$

the third inequality holds by $h_x \leq q_1^{-d/2}$ for a fixed invalid opening x, and the last inequality holds by the successful extraction probability, and therefore

$$\Pr[E_0 \vee E_2] \geq \varepsilon - \frac{2}{N} - q_1^{-d/2}.$$

Vanishing Constant Coefficient Proof. We denote the protocol for proving multiple quadratic relations by Π^2_{many}. We consider proving multiple quadratic relations, whose evaluations

$$\bar{\mathbf{s}}^{\mathsf{T}}\mathbf{R}_i\bar{\mathbf{s}} + \bar{\mathbf{r}}_{1,i}^{\mathsf{T}}\bar{\mathbf{s}} + \mathbf{r}_{0,i} \tag{4.67}$$

over $\bar{\mathbf{s}}$ defined in eq. (4.62), for $i \in [1, M]$, have constant coefficients being 0. Write $F_i(\bar{\mathbf{s}})$ for eq. (4.67), and extend the BDLOP commitment key by$\mathbf{B}_g \in R_q^{\lambda \times m_2}$. The protocol proceeds as follows:

- The prover samples $\bar{\mathbf{g}} \xleftarrow{R} \{\mathbf{a} \in R_q : a_0 = 0\}^\lambda$, commits $\bar{\mathbf{t}}_g \leftarrow \mathbf{B}_g\bar{\mathbf{s}}_2 + \bar{\mathbf{g}}$, and sends $\bar{\mathbf{t}}_g$ to the verifier
- The verifier replies $\boldsymbol{\Gamma} \xleftarrow{R} \mathbb{Z}_q^{\lambda \times M}$.
- The prover computes

$$\begin{bmatrix}\mathbf{h}_1\\ \vdots \\ \mathbf{h}_\lambda\end{bmatrix} \leftarrow \begin{bmatrix}\mathbf{g}_1\\ \vdots \\ \mathbf{g}_\lambda\end{bmatrix} + \boldsymbol{\Gamma}\begin{bmatrix}F_1(\bar{\mathbf{s}})\\ \vdots \\ F_M(\bar{\mathbf{s}})\end{bmatrix},$$

  and sends $\mathbf{h}_1, \ldots, \mathbf{h}_\lambda$.
- Invoke Π^2_{many} to prove

$$\mathbf{g}_j + \sum_{i=1}^{M} \gamma_{j,i}F_i(\bar{\mathbf{s}}) - \mathbf{h}_j = \mathbf{0}$$

  for $j \in [1, \lambda]$.
- The verifier accepts if Π^2_{many} accepts and $h_{j,0} = 0$ for $j \in [1, \lambda]$.

Additional quadratic relations over R_q over the original committed witness can be included in the same invocation of Π^2_{many}, which combines them with the above well-formedness equations.

Similar to the analysis for the multiple quadratic relation proof, the knowledge soundness here is $2/N + q_1^{-d/2} + q_1^{-\lambda}$. The extractor rewinds at the $\boldsymbol{\Gamma}$ challenge, and invokes the extractor for multiple quadratic relations. By the [ACK21] trick, the expected number of invocations to the extractor for multiple quadratic relations is 2, and thus the expected runtime is at most $12T$. With the same analysis, splitting into 3 disjoint events, the probability of 2 openings being distinct is at most $q_1^{-\lambda}$, completing the soundness analysis.

Proving Norm Bounds. We begin by a lemmas from [BL17, LNS21] as a motivation. Recall the binomial distribution from fair coin, where Bin_κ is a distribution from

$$\sum_{i \in [1,\kappa]} (X_i - Y_i),$$

where X_i and Y_i are fair coins over $\{0, 1\}$. We begin with the prototype in [BL17].

Lemma 4.14.5 (ℓ_∞ Lower-Tail Bound for Binary Compression [BL17]). *For all $\bar{s} \in \mathbb{Z}_p^k$, $\bar{c}' \in \{0,1\}^k$, and $y \in \mathbb{Z}_p$, over the randomness of $\bar{c} \xleftarrow{R} \{0,1\}^k$,*

$$\Pr\left[\left\|\langle \bar{s}, \bar{c} - \bar{c}'\rangle + y\right\|_\infty < \frac{1}{2}\left\|\bar{s}\right\|_\infty\right] \le \frac{1}{2}.$$

Proof. Let $F(\bar{c}) := \langle \bar{s}, \bar{c} - \bar{c}'\rangle + y$, and $\|\bar{s}\|_\infty = |s_i|$, then let $\bar{c}''$ has c_i bit flipped, and

$$F(\bar{c}) - F(\bar{c}'') = (2c_i - 1)s_i,$$

and hence $\left\|F(\bar{c}) - F(\bar{c}'')\right\|_\infty = \|\bar{s}\|_\infty$. By triangle inequality, for all $\bar{c} \in \{0,1\}^k$,

$$\|\bar{s}\|_\infty \le \|F(\bar{c})\|_\infty + \left\|F(\bar{c}'')\right\|_\infty,$$

such that at least one of $\|F(\bar{c})\|_\infty$ and $\left\|F(\bar{c}'')\right\|_\infty$ is at least $\|\bar{s}\|_\infty/2$, completing the proof. □

With lemma 4.14.5, the following result is immediate.

Lemma 4.14.6 (ℓ_∞ Lower-Tail Bound for Binomial Compression [BL17, LNS21]). *For all $\overline{s} \in \mathbb{Z}_p^k$ and $y \in \mathbb{Z}_p$, over the randomness of $\overline{c} \xleftarrow{R} \mathsf{Bin}_1^k$,*

$$\Pr\left[\left\|\langle \overline{s}, \overline{c}\rangle + y\right\|_\infty < \frac{1}{2}\left\|\overline{s}\right\|_\infty\right] \leq \frac{1}{2}.$$

Proof. By lemma 4.14.5, we have the claim for the fixed $\overline{c}'$ version of claim. By $\overline{c}' \xleftarrow{R} \{0,1\}^k$,

$$\Pr\left[\left\|\langle \overline{s}, \overline{c}''\rangle + y\right\|_\infty < \frac{1}{2}\left\|\overline{s}\right\|_\infty\right] = \mathbb{E}\left[\Pr\left[\left\|\langle \overline{s}, \overline{c} - \overline{c}'\rangle + y\right\|_\infty < \frac{1}{2}\left\|\overline{s}\right\|_\infty \middle| \overline{c}'\right]\right] \leq \frac{1}{2},$$

where $\overline{c}'' \leftarrow \mathsf{Bin}_1^k$, and the inequality holds by lemma 4.14.5, as any probability in the expectation is at most $1/2$. □

The range by lemma 4.14.6 gives a probability bound for the ℓ_∞-norm in $[\|\overline{s}\|_\infty/2, k\|\overline{s}\|_\infty]$, and by lemma 1.4.8, we can show without y that

$$\Pr\left[|\langle \overline{c}, \overline{s}\rangle| \geq C\sqrt{k}\left\|\overline{s}\right\|_\infty\right] \leq 2\exp\left(-C^2/2\right),$$

and given a concretely sufficiently large k, we can use a large constant C. The form of $y + \langle \overline{s}, \overline{c}\rangle$ is useful for the sigma protocol, yet the range of the ℓ_∞-norm has a factor of $O(\sqrt{k})$, which is too large.

Johnson-Lindenstrauss Lemma [JL84], conceptually shows the existence of a linear map $\mathbf{V} \in \mathbb{R}^{\ell \times m}$ sampled over continuous Gaussian distribution, such that for k distinct points in m-dimensional vector space be linearly embedded into ℓ-dimensional space, where $\ell = O(\log k)$, and their distances over ℓ_2-norm are preserved approximately.

Lemma 4.14.7 (Johnson–Lindenstrauss [JL84]). *Let $S \subseteq \mathbb{R}^m$ with $|S| = k \geq 2$, and $0 < \varepsilon < 1$. There exists $\ell = O(\varepsilon^{-2}\log k)$ and a linear map $\mathbf{V} \in \mathbb{R}^{\ell \times m}$ such that, for all $\overline{x}, \overline{y} \in S$,*

$$(1-\varepsilon)\left\|\overline{x} - \overline{y}\right\|_2^2 \leq \left\|\mathbf{V}\overline{x} - \mathbf{V}\overline{y}\right\|_2^2 \leq (1+\varepsilon)\left\|\overline{x} - \overline{y}\right\|_2^2.$$

In [JL84], the lemma 4.14.7 shows the existence of such linear map, and in [IM98] we have a randomized algorithm over Gaussian matrix, such that with high probability, the linear map satisfies the conditions in lemma 4.14.7.

Lemma 4.14.8 (Gaussian Randomized Johnson–Lindenstrauss [IM98]). *Let $S \subseteq \mathbb{R}^m$ with $|S| = k \geq 2$, and let $\varepsilon, \delta \in (0,1)$. There exists a constant $C > 0$ such that the following holds.*

Choose a positive integer

$$\ell \geq C\varepsilon^{-2}\log(k/\delta),$$

and sample $\mathbf{V} \in \mathbb{R}^{\ell \times m}$ by independently drawing each entry from $\mathcal{N}(0, 1/\ell)$.

Then, with probability at least $1-\delta$ over $\mathbf{V}$, simultaneously for all $\overline{x}, \overline{y} \in S$,

$$(1-\varepsilon)\left\|\overline{x} - \overline{y}\right\|_2^2 \leq \left\|\mathbf{V}\overline{x} - \mathbf{V}\overline{y}\right\|_2^2 \leq (1+\varepsilon)\left\|\overline{x} - \overline{y}\right\|_2^2.$$

Achlioptas [Ach03] shows that the same asymptotic dimension bound and distance-preservation guarantee can be obtained using independent random signs.

Lemma 4.14.9 (Sign Randomized Johnson–Lindenstrauss [Ach03]). *The conclusion of lemma 4.14.8 remains valid for entries $V_{i,j} = X_{i,j}/\sqrt{\ell}$, where the $X_{i,j}$ are sampled independently and uniformly over ± 1.*

In our scenario, we don't need to handle k points in this case, but we do need to preserve the pairwise difference in Euclidean space. More specifically, we want the distance between a fixed vector and zero preserved approximately. The fixed vector is the witness in the Sigma protocol. We therefore study how closely its projected norm follows its original norm, with high probability.

Gentry et al. [GHL22] use an integer projection from $\mathbf{R} \leftarrow \mathsf{Bin}_1^{256\times m}$ to reduce the dimension of the vector whose shortness is to be proved. To obtain concrete norm estimates, they compare this projection with a Gaussian matrix $\mathbf{G} \in \mathbb{R}^{256\times m}$ whose entries are sampled independently from $\mathcal{N}(0, 1/2)$, matching the variance of Bin_1.

For a fixed nonzero vector $\overline{w} \in \mathbb{Z}^m$,

$$\frac{2\,\|\mathbf{G}\overline{w}\|_2^2}{\|\overline{w}\|_2^2} \sim \chi^2_{256}.$$

To see this, consider the output of row i:

$$Y_i := (\mathbf{G}\overline{w})_i = \sum_{j=1}^{m} G_{i,j} w_j.$$

A fixed linear combination of independent Gaussian variables is Gaussian. Since $\mathbb{E}\left[G_{i,j}\right] = 0$ and $\mathrm{Var}\left[G_{i,j}\right] = 1/2$,

$$Y_i \sim \mathcal{N}\left(0, \frac{1}{2}\sum_{j\in[1,m]} w_j^2\right) = \mathcal{N}\left(0, \frac{\|\overline{w}\|_2^2}{2}\right).$$

Moreover, the Y_i are independent, since they use independently sampled rows of $\mathbf{G}$.

Thus, normalizing each output by its standard deviation gives independent standard normal variables

$$Z_i := \frac{\sqrt{2}}{\|\overline{w}\|_2} Y_i \sim \mathcal{N}(0,1).$$

In particular,

$$\|\mathbf{G}\overline{w}\|_2^2 = \sum_{i\in[1,256]} Y_i^2 = \frac{\|\overline{w}\|_2^2}{2} \sum_{i\in[1,256]} Z_i^2.$$

The factor $\|\overline{w}\|_2^2$ therefore cancels:

$$\frac{2\,\|\mathbf{G}\overline{w}\|_2^2}{\|\overline{w}\|_2^2} = \sum_{i=1}^{256} Z_i^2 \sim \chi^2_{256}.$$

Here, the 256 degrees of freedom count the independent standard normal variables being squared.

Let $X \sim \chi^2_{256}$. Numerically evaluating its tails gives

$$\Pr[X < 60] \approx 2^{-130.98} < 2^{-128}, \qquad \Pr[X > 674] \approx 2^{-128.30} < 2^{-128}.$$

Consequently, the normalization above yields the Gaussian bounds in [GHL22, Fact 3.1]:

$$\begin{aligned} \Pr_{\mathbf{G}}\left[\|\mathbf{G}\overline{w}\|_2^2 < 30\,\|\overline{w}\|_2^2\right] &= \Pr[X < 60] < 2^{-128}, \\ \Pr_{\mathbf{G}}\left[\|\mathbf{G}\overline{w}\|_2^2 > 337\,\|\overline{w}\|_2^2\right] &= \Pr[X > 674] < 2^{-128}. \end{aligned} \tag{4.68}$$

The constants 30 and 337 arise by dividing the chi-square thresholds 60 and 674 by 2.

Gentry et al. [GHL22, Corollary 3.2] then heuristically apply these same bounds to the projection over binomial distribution $\mathbf{R} \leftarrow \mathsf{Bin}_1^{256\times m}$, whose entries have the same variance as those of $\mathbf{G}$. The Gaussian bounds in eq. (4.68) are exact probability statements; their transfer to the binomial projection is heuristic.

More generally, replacing the Gaussian variance $1/2$ by $\kappa/2$ gives

$$\frac{2\,\|\mathbf{G}\overline{w}\|_2^2}{\kappa\,\|\overline{w}\|_2^2} \sim \chi^2_{256}.$$

Under the same Gaussian substitution heuristic, for $\mathbf{R} \leftarrow \mathsf{Bin}_{\kappa}^{256\times m}$ and $X \sim \chi^2_{256}$, LNP [LNP22, Lemma 2.8] obtains

$$\Pr_{\mathbf{R}}\left[\|\mathbf{R}\overline{w}\|_2^2 < 13\kappa\,\|\overline{w}\|_2^2\right] \approx \Pr[X < 26] \le 2^{-256},$$
$$\Pr_{\mathbf{R}}\left[\|\mathbf{R}\overline{w}\|_2^2 > 337\kappa\,\|\overline{w}\|_2^2\right] \approx \Pr[X > 674] \le 2^{-128}.$$

Here, $\approx$ denotes the heuristic comparison. The smaller lower-tail threshold gives the stronger 2^{-256} estimate needed in the subsequent argument.

Lemma 4.14.10 (ℓ_2 Lower-Tail Bound for Binomial Compression [LNP22]). *Fix $m, p \in \mathbb{Z}^+$ and a bound $0 < b \le p/(41m)$. Let $\overline{w} \in \mathbb{Z}_p^m$ satisfy $\|\overline{w}\|_2 \ge b$, and $\overline{y} \in \mathbb{Z}_p^{256}$ is arbitrary. Norms of residue vectors use centered representatives.*

For $\mathbf{R} \leftarrow \mathsf{Bin}_1^{256\times m}$ with independent entries, over the randomness of $\mathbf{R}$,

$$\Pr\left[\left\|\mathbf{R}\overline{w} + \overline{y} \bmod p\right\|_2 < 4b\right] < 2^{-128}.$$

Proof. We begin by separating out the case where $\|\overline{w}\|_\infty \ge p/(4m)$. By the given bound condition,

$$4b \le \frac{4p}{41m} \le \frac{p}{8m}.$$

By lemma 4.14.6, for any row $i \in [1, 256]$, over the randomness of i^{th} row $\overline{r}_i^{\mathsf{T}}$,

$$\Pr\left[\left|\langle \overline{r}_i, \overline{w}\rangle + y_i\right| < \frac{1}{2}\|\overline{w}\|_\infty\right] \le \frac{1}{2}.$$

All rows must satisfy such event to have the ℓ_2-norm of the projected vector be at most $4b$, of at most 2^{-256} probability.

Now consider the complementary case $\|\overline{w}\|_\infty < p/(4m)$. Consider all the computation over $\mathbb{Z}$ rather than $\mathbb{Z}_p$, and let $X_i = \langle \overline{r}_i, \overline{w}\rangle$, then we have $|X_i| \le \|\overline{w}\|_1 < p/4$, and thus each $X_i + y_i$ has $\left|X_i + y_i\right| < 3p/4$.

With a centered representation, a long $\left|X_i + y_i\right| > p/2$ has wrapped representation $p/4 < \left|X_i + y_i \bmod p\right| < p/2$, yet we have $4b \le p/(8m) < p/4$, and thus with $\|\overline{w}\|_\infty < p/(4m)$,

$$\Pr\left[\left\|\mathbf{R}\overline{w} + \overline{y} \bmod p\right\|_2 < 4b\right] = \Pr\left[\left\|\mathbf{R}\overline{w} + \overline{y}\right\|_2 < 4b\right],$$

as the settings with rows of large ℓ_∞-norm wrapping into $(p/4, p/2)$ are not counted anyways.

Continuing on, we use the MGF trick in Chernoff's bound, by

$$\begin{aligned}\Pr\left[\left\|\mathbf{R}\overline{w} + \overline{y}\right\|_2 < 4b\right] &= \Pr\left[\sum_{i\in[1,256]}\left(X_i + y_i\right)^2 < 16b^2\right] \le \exp\left(16sb^2\right)\mathbb{E}\left[\prod_{i\in[1,256]}\exp\left(-s\left(X_i + y_i\right)^2\right)\right] \\ &= \exp\left(16sb^2\right)\prod_{i\in[1,256]}\mathbb{E}\left[\exp\left(-s\left(X_i + y_i\right)^2\right)\right]\end{aligned}$$

for any $s \ge 0$. We observe that, for a binomial random variable $B \sim \mathsf{Bin}_1$,

$$\mathbb{E}\left[\exp\left(iuB\right)\right] = \frac{1}{4}\left(\exp\left(-iu\right) + \exp\left(iu\right)\right) + \frac{1}{2} = \frac{1}{2}\left(\cos\left(u\right) + 1\right) = \cos^2\left(u/2\right).$$

With this result, we can introduce the linear shift y such that

$$\begin{aligned}\mathbb{E}\left[\cos\left(u\left(X_i + y_i\right)\right)\right] &= \frac{1}{2}\left(\mathbb{E}\left[\exp\left(ui\left(X_i + y_i\right)\right)\right] + \mathbb{E}\left[\exp\left(-ui\left(X_i + y_i\right)\right)\right]\right) = \cos\left(uy_i\right)\prod_{j\in[1,m]}\cos^2\left(uw_j/2\right) \\ &\le \prod_{j\in[1,m]}\cos^2\left(uw_j/2\right) = \mathbb{E}\left[\cos\left(uX_i\right)\right].\end{aligned}$$

Another fact we are using is, let $G \sim \mathcal{N}(0,1)$ and any $t \in \mathbb{R}$, we have $\mathbb{E}[\cos(tG)] = \exp(-t^2/2)$. Taking independent G of X_i, we substitute $u = \sqrt{2s}G$ into the preceding inequality and average over G. Since cosine is bounded, we can exchange the order of expectations and obtain

$$\begin{aligned}\mathop{\mathbb{E}}_{X_i}\left[\exp\left(-s(X_i+y_i)^2\right)\right] &= \mathop{\mathbb{E}}_{G}\mathop{\mathbb{E}}_{X_i}\left[\cos\left(\sqrt{2s}G(X_i+y_i)\right)\right] \\ &\le \mathop{\mathbb{E}}_{G}\mathop{\mathbb{E}}_{X_i}\left[\cos\left(\sqrt{2s}GX_i\right)\right] = \mathop{\mathbb{E}}_{X_i}\left[\exp\left(-sX_i^2\right)\right].\end{aligned}$$

Hence, the shift y_i is removed from the upper bound.

Substituting this into the Chernoff bound, and using that the X_i have the same distribution, gives

$$\Pr\left[\left\|\mathbf{R}\overline{w}+\overline{y}\right\|_2 < 4b\right] \le \exp\left(16sb^2\right)\prod_{i\in[1,256]}\mathbb{E}\left[\exp\left(-sX_i^2\right)\right] \tag{4.69}$$

for every $s \ge 0$. It remains to bound the expectation for a single row and then choose s.

The next goal is to upper bound $\mathbb{E}\left[\exp\left(-sX_i^2\right)\right]$ such that the X_i can be taken out, and it only matters with $\|\overline{w}\|_2$. We began by the Taylor expansion for $\exp(-sa)$ over a, that is

$$\exp(-sa) \le 1 - sa + \frac{s^2a^2}{2}. \tag{4.70}$$

By the linearity of expectation, we have

$$\mathbb{E}\left[\exp\left(-sX_i^2\right)\right] \le 1 - s\,\mathbb{E}\left[X_i^2\right] + \frac{s^2}{2}\mathbb{E}\left[X_i^4\right].$$

As was discussed in remark 1.6.39, we can derive the second and fourth moment of X_i by

$$\begin{aligned}\mathbb{E}\left[X_i^2\right] &= \frac{1}{2}\|\overline{w}\|_2^2, \\ \mathbb{E}\left[X_i^4\right] &= \frac{1}{2}\sum_{j\in[1,m]} w_j^4 + 3\cdot\frac{1}{4}\sum_{j\ne k} w_j^2w_k^2 = \frac{3}{4}\|\overline{w}\|_2^4 - \frac{1}{4}\sum_{j\in[1,m]} w_j^4 \le \frac{3}{4}\|\overline{w}\|_2^4.\end{aligned} \tag{4.71}$$

By $b \le \|\overline{w}\|_2$ in lemma, we have $\exp\left(16sb^2\right) \le \exp\left(16s\|\overline{w}\|_2^2\right)$. Writing

$$t = s\|\overline{w}\|_2^2 \ge 0, \tag{4.72}$$

the loose upper bound is 2^{-52}.

To derive a tigher bound from eq. (4.70), we consider coefficients $s \ge A \ge 0, B \ge 0$ such that

$$\exp(-sa) \le 1 - Aa + Ba^2 = F(a),$$

holds for all $a \ge 0$, and the degree-2 series expansion upper bounds $\mathbb{E}\left[\exp(-sa)\right]$ tigher.

Averaging both sides by $a = X_i^2$,

$$\begin{aligned}\mathbb{E}\left[\exp\left(-sX_i^2\right)\right] &\le 1 - A\,\mathbb{E}\left[X_i^2\right] + B\,\mathbb{E}\left[X_i^4\right] \le 1 - A\frac{\|\overline{w}\|_2^2}{2} + B\frac{3\|\overline{w}\|_2^4}{4} \\ &= 1 - \frac{A}{3}\cdot\frac{3\|\overline{w}\|_2^2}{2} + \frac{B}{3}\left(\frac{3\|\overline{w}\|_2^2}{2}\right)^2 = \frac{2}{3} + \frac{1}{3}F\left(\frac{3\|\overline{w}\|_2^2}{2}\right).\end{aligned}$$

By earlier requirement, $F(a) \ge \exp(-sa)$, then the upper bound is closest when $F(c) = \exp(-sc)$ when $c = 3\|\overline{w}\|_2^2/2$, and it is they are not intersecting at c. Hence,

$$\begin{cases}1 - Ac + Bc^2 = \exp(-sc), \\ -A + 2Bc = -s\exp(-sc).\end{cases} \tag{4.73}$$

We therefore obtain

$$B = \frac{1-(1+sc)\exp(-sc)}{c^2}, \qquad A = \frac{2-(2+sc)\exp(-sc)}{c}.$$

We first verify $s - A \geq 0$. Put $x = sc \geq 0$. Then

$$s - A = \frac{x - 2 + (x+2)\exp(-x)}{c}. \tag{4.74}$$

The numerator vanishes at $x = 0$, and its derivative is

$$1 - (1+x)\exp(-x) \geq 0,$$

since $\exp(x) \geq 1 + x$. Thus $A \leq s$.

To verify the upper bound throughout the domain, let

$$H(a) = F(a) - \exp(-sa).$$

Our construction gives

$$H(0) = H(c) = H'(c) = 0.$$

If $s = 0$, then $A = B = 0$ and H is identically zero. Otherwise, $H'''(a) = s^3\exp(-sa) > 0$, so H' is strictly convex.

Since $H(0) = H(c)$, Rolle's theorem gives some $r \in (0, c)$ with $H'(r) = 0$. Together with $H'(c) = 0$, strict convexity implies that H' is negative between r and c, and positive outside this interval. Therefore, H increases from $H(0) = 0$ to $H(r)$, decreases back to $H(c) = 0$, and increases thereafter. Consequently, $H(a) \geq 0$ for all $a \geq 0$.

Again let $t = s\,\|\overline{w}\|_2^2$, and we are at 2^{-96} for an upper bound.

To further improve the bound, we use a degree-4 polynomial and moments up to order 8. For $t = s\,\|\overline{w}\|_2^2 \geq 0$,

$$F_4(a) = 1 - A_4 a + B_4 a^2 - C_4 a^3 + D_4 a^4, \tag{4.75}$$

where

$$\begin{aligned}
A_4 &= \frac{1}{\|\overline{w}\|_2^2}\left(\frac{5}{2} - \left(\frac{64}{27} + \frac{16t}{9}\right)\exp(-t) - \left(\frac{7}{54} + \frac{t}{9}\right)\exp(-4t)\right),\\
B_4 &= \frac{1}{\|\overline{w}\|_2^4}\left(\frac{33}{16} - \left(\frac{16}{9} + \frac{8t}{3}\right)\exp(-t) - \left(\frac{41}{144} + \frac{t}{4}\right)\exp(-4t)\right),\\
C_4 &= \frac{1}{\|\overline{w}\|_2^6}\left(\frac{5}{8} - \left(\frac{4}{9} + t\right)\exp(-t) - \left(\frac{13}{72} + \frac{t}{6}\right)\exp(-4t)\right),\\
D_4 &= \frac{1}{\|\overline{w}\|_2^8}\left(\frac{1}{16} - \left(\frac{1}{27} + \frac{t}{9}\right)\exp(-t) - \left(\frac{11}{432} + \frac{t}{36}\right)\exp(-4t)\right).
\end{aligned}$$

For all $a \geq 0$, we have

$$\exp(-sa) \leq F_4(a).$$

Averaging both sides with $a = X_i^2$ and using the moments up to order 8,

$$\mathbb{E}\left[\exp\left(-sX_i^2\right)\right] \leq \mathbb{E}\left[F_4(X_i^2)\right] \leq \frac{137}{256} + \left(\frac{191}{432} + \frac{5t}{144}\right)\exp(-t) + \left(\frac{157}{6912} - \frac{t}{576}\right)\exp(-4t).$$

Using $b \leq \|\overline{w}\|_2$, we upper bound by 2^{-144}. A construction with touching points derived from the moments, including its coefficient derivation and binomial correction analysis, is given in appendix A. This is mainly AI's work, I cannot pull it off myself. □

TODO: *Stuffs for* [LNP22].

TODO: *Write a short note on rigorous replacements for the projection bounds in* [LNP22, *Lemmas* 2.8–2.9]. *Use the exact representation of* Bin_κ *as a sum of independent signs scaled by* 1/2 *and* [Ach03, *Eq.* (13)] *to write up the upper-tail bound replacing* 337 *by* 343, *with error below* 2^{-128}. *Combine it with the shifted modular lower-tail bound in lemma* 4.14.10 *and its moment analysis in appendix A, then track the resulting completeness and soundness parameters in Section* 5 *of* [LNP22].

4.15 Practical Sublinear Proofs for R1CS from Lattices

TODO: *stuff from* [*NS*22].

4.16 LaBRADOR

TODO: ... [*BS*23]

4.17 Greyhound

TODO: ... [*NS*24]

4.18 Lattice-Based Succinct Arguments for NP with Polylog Verification

TODO: ... [*BCS*23]

4.19 Greco

So the motivation of Greco [Bot24] is to prove the validity of some RLWE ciphertext, more specifically BFV ciphertext [Bra12, FV12] whose ciphertext is improved with CRT [HPS19, KPZ21]. The main idea of Greco is from [dPLS19].

Throughout, coefficients in $\mathbb{Z}_q$ are lifted to their centered integer representation, unless stated otherwise.

Start with BFV ciphertext [Bra12, FV12] from RLWE over R_q^2, where $R_q = \mathbb{Z}_q[X]/(X^n+1)$. We have

$$(\mathsf{ct}_0, \mathsf{ct}_1) = (\mathbf{a}\cdot\mathbf{s} + \mathbf{e} + k, -\mathbf{a}),$$

where $\mathbf{a} \xleftarrow{R} R_q$, $\mathbf{s} \xleftarrow{R} \{-1,0,1\}^n$, $\mathbf{e} \leftarrow \chi^n$, all the computation are carried over R_q, and

$$k = \left\lfloor \frac{q}{t} \left\lfloor m \right\rceil_t \right\rceil = \frac{q}{t}\left\lfloor m \right\rceil_t - \frac{1}{t}\left\lfloor mq \right\rceil_t \in \left[-\frac{q}{2}, \frac{q}{2}\right) \cap \mathbb{Z},$$

where $\lfloor\cdot\rceil$ means rounding to the nearest integer, $\lfloor z \rceil_t$ means centered remainder of z modulo t, and t is the plaintext modulus. With the requirement that q and t are coprime, then

$$k \equiv -t^{-1}\left\lfloor mq \right\rceil_t \mod q,$$

which was described in [KPZ21]. With a composite ciphertext modulus $q = \prod q_i$, we have CRT splitting such that

$$(\mathsf{ct}_0, \mathsf{ct}_1) \cong \left((\mathsf{ct}_{0,0}, \mathsf{ct}_{1,0}) \mod q_0, \ldots, (\mathsf{ct}_{0,\ell-1}, \mathsf{ct}_{1,\ell-1}) \mod q_{\ell-1}\right),$$

and thus for the plaintext encoding,

$$k \equiv k_{0,i} k_1 \mod q_i,$$

where $k_{0,i} \equiv -t^{-1} \mod q_i$, and $k_1 = \left\lfloor mq \right\rceil_t$.

Greco [Bot24] extends from [dPLS19] in the CRT modulus splitting setting. The gain is not taking advantage of the parallelism in BFV, but subtly allowing for the circuit arithmetic be simulated over $\mathbb{Z}_p$ as if it is over $\mathbb{Z}$, where p is the order of elliptic curve commitment group, used by for example [KZG10].

Lifting all the polynomial arithmetics from R_{q_i} over $\mathbb{Z}[X]$, then we have

$$\mathbf{a}_i \cdot \mathbf{s} + \mathbf{e} + k_{0,i}k_1 + \mathbf{r}_{1,i}q_i + \mathbf{r}_{2,i}(X^n+1) = \mathsf{ct}_{0,i} \tag{4.76}$$

for all $i \in [0, \ell-1]$, such that for all $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{Z}[X]$,

$$\boldsymbol{\alpha} q_i + \boldsymbol{\beta}(X^n+1) \equiv \mathbf{0} \mod \left(q_i, (X^n+1)\right). \tag{4.77}$$

Given a sufficiently large modulus p such that the ℓ_∞-norm does not wrap around over $\mathbb{Z}_p$, namely the summation of the ℓ_∞-norms of both sides of eq. (4.76) is strictly less than p, we can simulate eq. (4.76) over $\mathbb{Z}_p[X]$.

Too Many Range Checks. We now show how to decrease the number range checks in [Bot24]. Over $\mathbb{Z}[X]$, let

$$\mathbf{d}_i = \mathsf{ct}_{0,i} - (\mathbf{a}_i \cdot \mathbf{s} + \mathbf{e} + k_{0,i}k_1) \in \mathbb{Z}[X], \tag{4.78}$$

which must be of form $\boldsymbol{\alpha}_i(X^n+1) + \boldsymbol{\beta}_i$, as X^n+1 is monic, division over $\mathbb{Z}[X]$ gives unique quotient and remainder $\boldsymbol{\alpha}_i, \boldsymbol{\beta}_i \in \mathbb{Z}[X]$, $\deg \boldsymbol{\alpha}_i \le n-2$, and $\deg \boldsymbol{\beta}_i \le n-1$. For accepting relations, namely vaild ciphertexts, we have

$$\mathbf{d}_i \equiv \mathbf{0} \mod \left(q_i, (X^n+1)\right),$$

then by eq. (4.77), $\boldsymbol{\beta}_i = q_i\boldsymbol{\gamma}_i$ for some $\boldsymbol{\gamma}_i \in \mathbb{Z}[X]$ where $\deg \boldsymbol{\gamma}_i \le n-1$. Therefore, $\mathbf{d}_i = \boldsymbol{\alpha}_i(X^n+1) + q_i\boldsymbol{\gamma}_i$.

Greco let $\mathbf{r}_{2,i}$ be center reprepresented over $\mathbb{Z}_{q_i}[X]$. We thus write $\boldsymbol{\alpha}_i = \left\lfloor \boldsymbol{\alpha}_i \right\rceil_{q_i} + q_i\mathbf{u}_{2,i}$ for some $\mathbf{u}_{2,i} \in \mathbb{Z}[X]$, then

$$\mathbf{d}_i = \boldsymbol{\alpha}_i(X^n+1) + q_i\boldsymbol{\gamma}_i = \left(\left\lfloor \boldsymbol{\alpha}_i \right\rceil_{q_i} + q_i\mathbf{u}_{2,i}\right)(X^n+1) + q_i\boldsymbol{\gamma}_i$$

$$= \lfloor \boldsymbol{\alpha}_i \rceil_{q_i} (X^n + 1) + q_i \left(\mathbf{u}_{2,i} (X^n + 1) + \boldsymbol{\gamma}_i \right).$$

Essentially, the identity exhibits a witness pair with $\mathbf{r}_{2,i}$ centered modulo q_i, and the compensating multiple of $X^n + 1$ in $\mathbf{r}_{1,i}$. Therefore, their $\deg \mathbf{r}_{1,i}$ is upper bounded by $2n - 2$.

It means for the $\mathbf{r}_{1,i}$ and $\mathbf{r}_{2,i}$, there are $3n-2$ elements range checked per q_i by Greco, which is the same as [dPLS19]. As of the ranges,

- For $\mathbf{r}_{2,i}$ with centered representation modulo q_i, we have

$$\|\mathbf{r}_{2,i}\|_\infty \leq \frac{q_i - 1}{2}. \tag{4.79}$$

- For $\mathbf{r}_{1,i}$, we consider the following over $\mathbb{Z}[X]$:

$$\mathbf{r}_{1,i} = q_i^{-1} \left(\mathsf{ct}_{0,i} - (\mathbf{a}_i \cdot \mathbf{s} + \mathbf{e} + k_{0,i} k_1) - \mathbf{r}_{2,i} (X^n + 1) \right),$$

and let $\mathbf{e}$ be B-bounded. The $\mathsf{ct}_{0,i} - (\mathbf{a}_i \cdot \mathbf{s} + \mathbf{e} + k_{0,i} k_1)$ has ℓ_∞-norm bounded by

$$D_i = \frac{q_i - 1}{2} + n \frac{q_i - 1}{2} + B + |k_{0,i}| \cdot \frac{t - 1}{2}, \tag{4.80}$$

as $k_{0,i}$ can be publicly derived, $\mathbf{a}_i$ and k_1 are centered represented, $\mathbf{e}$ is not long enough to wrap back, and

$$\|\mathbf{r}_{2,i} (X^n + 1)\|_\infty = \|\mathbf{r}_{2,i}\|_\infty \leq \frac{q_i - 1}{2}$$

as $\deg \mathbf{r}_{2,i} \leq n - 2$. Hence,

$$\|\mathbf{r}_{1,i}\|_\infty \leq q_i^{-1} \left(\frac{q_i - 1}{2} + n \frac{q_i - 1}{2} + B + |k_{0,i}| \cdot \frac{t - 1}{2} + \frac{q_i - 1}{2} \right) = q_i^{-1} \left((n + 2) \cdot \frac{q_i - 1}{2} + B + |k_{0,i}| \cdot \frac{t - 1}{2} \right). \tag{4.81}$$

To ensure that p is sufficiently large for the relation over $\mathbb{Z}[X]$, such that the polynomial relation is simulatable over $\mathbb{Z}_p[X]$, we require the summation of the ℓ_∞-norms of both sides of the equation to be not exceeding p by

$$p > 2 \left((n + 2) \cdot \frac{q_i - 1}{2} + B + |k_{0,i}| \cdot \frac{t - 1}{2} \right). \tag{4.82}$$

To sum up, we have bounds for $\mathbf{r}_{1,i}$ and $\mathbf{r}_{2,i}$ in eqs. (4.79) and (4.81) from eq. (4.82) for [Bot24, dPLS19].

It is worth noticing the following fact about polynomial product over $\mathbb{Z}[X]$ and modulo $X^n + 1$.

Lemma 4.19.1. *Let* $\mathbf{f}, \mathbf{g} \in \mathbb{Z}[X]$ *have* $\deg \mathbf{f} \leq n - 1$ *and* $\deg \mathbf{g} \leq n - 1$, *then*

$$\|\mathbf{fg}\|_\infty \leq n \|\mathbf{f}\|_\infty \|\mathbf{g}\|_\infty,$$
$$\|\mathbf{fg} \mod (X^n + 1)\|_\infty \leq n \|\mathbf{f}\|_\infty \|\mathbf{g}\|_\infty.$$

Proof. For $\mathbf{fg}$ product over $\mathbb{Z}[X]$,

$$\|\mathbf{fg}\|_\infty \leq \sum_{j \in [0, n-1]} |f_j| \left\| X^j \mathbf{g} \right\|_\infty = \sum_{j \in [0, n-1]} |f_j| \|\mathbf{g}\|_\infty \leq n \|\mathbf{f}\|_\infty \|\mathbf{g}\|_\infty.$$

For $\mathbf{fg} \mod (X^n + 1)$ product over $\mathbb{Z}[X]$,

$$\|\mathbf{fg} \mod (X^n + 1)\|_\infty \leq \sum_{j \in [0, n-1]} |f_j| \left\| X^j \mathbf{g} \mod (X^n + 1) \right\|_\infty = \sum_{j \in [0, n-1]} |f_j| \|\mathbf{g}\|_\infty \leq n \|\mathbf{f}\|_\infty \|\mathbf{g}\|_\infty.$$

□

We use independently supplied witnesses $\mathbf{r}_{1,i}, \mathbf{r}_{2,i}$ with $\deg \mathbf{r}_{1,i} \leq n-1$ and $\deg \mathbf{r}_{2,i} \leq n-2$. It suffices to check the full polynomial relation eq. (4.76) over $\mathbb{Z}[X]$ by

- checking $q_i \mathbf{r}_{1,i} \equiv \mathbf{d}_i \mod (X^n+1)$,
- and checking $\mathbf{r}_{2,i} = \boldsymbol{\alpha}_i$, where $\boldsymbol{\alpha}_i$ is decomposed from $\mathbf{d}_i$.

By lemma 4.19.1, for eq. (4.78), $\|\mathbf{d}_i\|_\infty$ and $\|\mathbf{d}_i \mod (X^n+1)\|_\infty$ share the same ℓ_∞-norm bound eq. (4.80), as all other terms have degree at most $n-1$, while lemma 4.19.1 applies to $\mathbf{a}_i \cdot \mathbf{s}$ over $\mathbb{Z}[X]$. Hence,

$$\|\mathbf{r}_{1,i}\|_\infty \leq q_i^{-1} \left((n+1) \cdot \frac{q_i - 1}{2} + B + |k_{0,i}| \cdot \frac{t-1}{2} \right) = q_i^{-1} D_i$$

by eq. (4.80), as $\mathbf{r}_{2,i}(X^n+1)$ is filtered out. Then over $\mathbb{Z}[X]/(X^n+1)$, if we relation is simulatable over $\mathbb{Z}_p[X]/(X^n+1)$, we need

$$p > 2D_i = 2 \left((n+1) \cdot \frac{q_i - 1}{2} + B + |k_{0,i}| \cdot \frac{t-1}{2} \right).$$

We pause by a philosophical question: Why do we need range checks here? We are essentially checking a $\mathbb{Z}[X]$ relation over $\mathbb{Z}_p[X]$, and thus all the computation over $\mathbb{Z}$ is within $[-(p-1)/2, (p-1)/2]$ [28].

- For $\mathbf{r}_{1,i}$ case, it carries a scalar q_i, and thus it needs to be range checked [29].
- But for $\mathbf{r}_{2,i}$, multiplying with a X^n+1, which is monic (highest degree coefficient being 1), and does not scale up the ℓ_∞-norm for the coefficients of terms with degree at least n. Hence, for a sufficiently large p by

  $$p > 2D_i > 2n \|\mathbf{a}_i\|_\infty \|\mathbf{s}\|_\infty,$$

  the equality between $\mathbf{r}_{2,i}$ and $\boldsymbol{\alpha}_i$ over $\mathbb{Z}_p[X]$ can be lifted to $\mathbb{Z}[X]$ in centered representation without range check.

Therefore, there are only n range checks need to be performed over $\mathbf{r}_{1,i}$, bounded by $q_i^{-1} D_i$, assuming that $p > 2D_i$.

Degree-3 Trick Like [BLS19]. Inducing a lookup-based range check for entries in $\mathbf{s}$ is a bit costly. Say $\mathbf{s} \in \{-1, 0, 1\}^n$, then each entries can be made into

$$(s_i - 1)\, s_i\, (s_i + 1) = 0,$$

which is not exactly the same, but in the same spirit as [BLS19].

TODO: ... [*Bot*24]

[28] Greco [Bot24, dPLS19] range checks both $\mathbf{r}_{1,i}$ and $\mathbf{r}_{2,i}$ to instantiate the p lower bound, but it confuses me about the point of range checks.
[29] Think of an counterexample where $p = 101$, $q_i = 7$, and $r = 29$ not range checked, for invalid $d = 1$.

References

[Ach03] Dimitris Achlioptas. Database-friendly random projections: Johnson–Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66(4):671–687, 2003.

[ACK21] Thomas Attema, Ronald Cramer, and Lisa Kohl. A compressed Σ-protocol theory for lattices. In *CRYPTO*, pages 549–579, 2021.

[ACX21] Thomas Attema, Ronald Cramer, and Chaoping Xing. A note on short invertible ring elements and applications to cyclotomic and trinomial number fields. *Mathematical Cryptology*, 1(1):45–70, 2021.

[AGHS13] Shweta Agrawal, Craig Gentry, Shai Halevi, and Amit Sahai. Discrete gaussian leftover hash lemma over infinite domains. In *ASIACRYPT*, pages 97–116, 2013.

[AHIV17] Scott Ames, Carmit Hazay, Yuval Ishai, and Muthuramakrishnan Venkitasubramaniam. Ligero: Lightweight sublinear arguments without a trusted setup. In *ACM CCS*, pages 2087–2104, 2017.

[Ajt96] Miklós Ajtai. Generating hard instances of lattice problems. In *STOC*, pages 99–108, 1996.

[ALS20] Thomas Attema, Vadim Lyubashevsky, and Gregor Seiler. Practical product proofs for lattice commitments. In *CRYPTO*, pages 470–499, 2020.

[And87] Désiré André. Solution directe du problème résolu par m. bertrand. *Comptes Rendus de l'Académie des Sciences, Paris*, 105:436–437, 1887.

[AS16] Noga Alon and Joel H. Spencer. *The Probabilistic Method*. Wiley, 4th edition, 2016.

[Azu67] Kazuoki Azuma. Weighted sums of certain dependent random variables. *Tohoku Mathematical Journal*, 19(3):357–367, 1967.

[Ban93] Wojciech Banaszczyk. New bounds in some transference theorems in the geometry of numbers. *Mathematische Annalen*, 296:625–635, 1993.

[BBB+18] Benedikt Bünz, Jonathan Bootle, Dan Boneh, Andrew Poelstra, Pieter Wuille, and Gregory Maxwell. Bulletproofs: Short proofs for confidential transactions and more. In *IEEE S&P*, pages 315–334, 2018.

[BBC+18] Carsten Baum, Jonathan Bootle, Andrea Cerulli, Rafaël del Pino, Jens Groth, and Vadim Lyubashevsky. Sub-linear lattice-based zero-knowledge arguments for arithmetic circuits. In *CRYPTO*, pages 669–699, 2018.

[BCS23] Jonathan Bootle, Alessandro Chiesa, and Katerina Sotiraki. Lattice-based succinct arguments for NP with polylogarithmic-time verification. In *CRYPTO*, pages 227–251, 2023.

[BD97] Russ Bubley and Martin E. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *FOCS*, pages 223–231, 1997.

[BDL+18] Carsten Baum, Ivan Damgård, Vadim Lyubashevsky, Sabine Oechsner, and Chris Peikert. More efficient commitments from structured lattice assumptions. In *SCN*, pages 368–385, 2018.

[Bec91] József Beck. An algorithmic approach to the lovász local lemma. i. *Random Struct. Algorithms*, 2(4):343–365, 1991.

[BG93] Mihir Bellare and Oded Goldreich. On defining proofs of knowledge. In *CRYPTO*, pages 390–420, 1993.

[BL17] Carsten Baum and Vadim Lyubashevsky. Simple amortized proofs of shortness for linear relations over polynomial rings. Cryptology ePrint Archive, Paper 2017/759, 2017.

[BLNS20] Jonathan Bootle, Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. A Non-PCP approach to succinct Quantum-Safe Zero-Knowledge. In *CRYPTO*, pages 441–469, 2020.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. Self-testing/correcting with applications to numerical problems. *Journal of Computer and System Sciences*, 47(3):549–595, 1993.

[BLS19] Jonathan Bootle, Vadim Lyubashevsky, and Gregor Seiler. Algebraic techniques for short(er) exact lattice-based zero-knowledge proofs. In *CRYPTO*, pages 176–202, 2019.

[Bot24] Enrico Bottazzi. Greco: Fast zero-knowledge proofs for valid FHE RLWE ciphertexts formation. Cryptology ePrint Archive, Paper 2024/594, 2024.

[Bra12] Zvika Brakerski. Fully homomorphic encryption without modulus switching from classical gapsvp. In *CRYPTO*, pages 868–886, 2012.

[BS23] Ward Beullens and Gregor Seiler. LaBRADOR: Compact proofs for R1CS from Module-SIS. In *CRYPTO*, pages 518–548, 2023.

[CF24] Hadas Zeilberger Binyi Chen and Ben Fisch. Basefold: Efficient multilinear polynomial commitment schemes from foldable codes. In *CRYPTO*, pages 138–169, 2024.

[Cho94] Kwok Pui Choi. On the medians of gamma distributions and an equation of ramanujan. *Proceedings of the American Mathematical Society*, 121(1):245–251, 1994.

[Dam10] Ivan Damgård. On Σ-protocols. Lecture notes, Cryptologic Protocol Theory, 2010. CPT 2010, v.2.

[Dob40] Wolfgang Doblin. Éléments d'une théorie générale des chaînes simples constantes de Markoff. *Annales scientifiques de l'École Normale Supérieure*, 57:61–111, 1940.

[Doo40] Joseph L. Doob. Regularity properties of certain families of chance variables. *Transactions of the American Mathematical Society*, 47(3):455–486, 1940.

[DP24] Benjamin E. Diamond and Jim Posen. Proximity testing with logarithmic randomness. *IACR Communications in Cryptology*, 1(1), 2024.

[dPLS19] Rafaël del Pino, Vadim Lyubashevsky, and Gregor Seiler. Short discrete log proofs for FHE and ring-LWE ciphertexts. In *PKC*, pages 344–373, 2019.

[EL75] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. In András Hajnal, Richard Rado, and Vera T. Sós, editors, *Infinite and Finite Sets*, volume 10 of *Colloquia Mathematica Societatis János Bolyai*, pages 609–627. North-Holland, Amsterdam, 1975.

[Eli72] Peter Elias. The efficient construction of an unbiased random sequence. *The Annals of Mathematical Statistics*, 43(3):865–870, 1972.

[ENS20] Muhammed F. Esgin, Ngoc Khanh Nguyen, and Gregor Seiler. Practical exact proofs from lattices: New techniques to exploit fully-splitting rings. In *ASIACRYPT*, pages 259–288, 2020.

[Ete81] Nasrollah Etemadi. An elementary proof of the strong law of large numbers. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 55(1):119–122, 1981.

[EZS+19] Muhammed F. Esgin, Raymond K. Zhao, Ron Steinfeld, Joseph K. Liu, and Dongxi Liu. MatRiCT: Efficient, scalable and post-quantum blockchain confidential transactions protocol. In *ACM CCS*, pages 567–584, 2019.

[Fel68a] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 1. John Wiley & Sons, 1968.

[Fel68b] William Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley & Sons, 1968.

[FS90] Uriel Feige and Adi Shamir. Witness indistinguishable and witness hiding protocols. In *STOC*, pages 416–426, 1990.

[FV12] Junfeng Fan and Frederik Vercauteren. Somewhat practical fully homomorphic encryption. Cryptology ePrint Archive, Paper 2012/144, 2012.

[GHL22] Craig Gentry, Shai Halevi, and Vadim Lyubashevsky. Practical non-interactive publicly verifiable secret sharing with thousands of parties. In *EUROCRYPT*, pages 458–487, 2022.

[HILL99] Johan Håstad, Russell Impagliazzo, Leonid A. Levin, and Michael Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28(4):1364–1396, 1999.

[Hoe63] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963.

[HPS19] Shai Halevi, Yuriy Polyakov, and Victor Shoup. An improved RNS variant of the BFV homomorphic encryption scheme. In *Topics in Cryptology – CT-RSA 2019*, pages 83–105, 2019.

[ILL89] Russell Impagliazzo, Leonid A. Levin, and Michael Luby. Pseudo-random generation from one-way functions. In *STOC*, pages 12–24, 1989.

[IM98] Piotr Indyk and Rajeev Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *STOC*, pages 604–613, 1998.

[Jer95] Mark Jerrum. A very simple algorithm for estimating the number of k-colorings of a low-degree graph. *Random Structures & Algorithms*, 7(2):157–165, 1995.

[JL84] William B. Johnson and Joram Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In *Conference in Modern Analysis and Probability*, volume 26 of *Contemporary Mathematics*, pages 189–206. American Mathematical Society, 1984.

[JS89] Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM Journal on Computing*, 18(6):1149–1178, 1989.

[JVV86] Mark R. Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.*, 43:169–188, 1986.

[KL51] Solomon Kullback and Richard A. Leibler. On information and sufficiency. *The Annals of Mathematical Statistics*, 22(1):79–86, 1951.

[KPZ21] Andrey Kim, Yuriy Polyakov, and Vincent Zucca. Revisiting homomorphic encryption schemes for finite fields. In *ASIACRYPT*, pages 608–639, 2021.

[Kra49] Leon G. Kraft. A device for quantizing, grouping, and coding amplitude-modulated pulses. M.s. thesis, Massachusetts Institute of Technology, Cambridge, MA, 1949.

[KZG10] Aniket Kate, Gregory M. Zaverucha, and Ian Goldberg. Constant-size commitments to polynomials and their applications. In *ASIACRYPT*, pages 177–194, 2010.

[LM06] Vadim Lyubashevsky and Daniele Micciancio. Generalized compact knapsacks are collision resistant. In *ICALP*, pages 144–155, 2006.

[LMR94] F. Thomson Leighton, Bruce M. Maggs, and Satish B. Rao. Packet routing and job-shop scheduling in O(congestion + dilation) steps. *Combinatorica*, 14(2):167–186, 1994.

[LNP22] Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Maxime Plançon. Lattice-based zero-knowledge proofs and applications: Shorter, simpler, and more general. In *CRYPTO*, pages 71–101, 2022.

[LNS21] Vadim Lyubashevsky, Ngoc Khanh Nguyen, and Gregor Seiler. Shorter lattice-based zero-knowledge proofs via one-time commitments. In *PKC*, pages 215–241, 2021.

[Lov79] László Lovász. On the shannon capacity of a graph. *IEEE Transactions on Information Theory*, 25(1):1–7, 1979.

[LP17] David A. Levin and Yuval Peres. *Markov Chains and Mixing Times*. American Mathematical Society, Providence, RI, 2 edition, 2017. With contributions by Elizabeth L. Wilmer.

[LPR13] Vadim Lyubashevsky, Chris Peikert, and Oded Regev. A toolkit for Ring-LWE cryptography. In *EUROCRYPT*, pages 35–54, 2013.

[LS15] Adeline Langlois and Damien Stehlé. Worst-case to average-case reductions for module lattices. *Des. Codes Cryptogr.*, 75(3):565–599, 2015.

[LS18] Vadim Lyubashevsky and Gregor Seiler. Short, invertible elements in partially splitting cyclotomic rings and applications to lattice-based zero-knowledge proofs. In *EUROCRYPT*, pages 204–224, 2018.

[Lub66] David Lubell. A short proof of sperner's lemma. *Journal of Combinatorial Theory*, 1:299, 1966.

[Lyu09] Vadim Lyubashevsky. Fiat–Shamir with aborts: Applications to lattice and factoring-based signatures. In *ASIACRYPT*, pages 598–616, 2009.

[Lyu12] Vadim Lyubashevsky. Lattice signatures without trapdoors. In *EUROCRYPT*, pages 738–755, 2012.

[Mat88] Peter C. Matthews. Covering problems for brownian motion on spheres. *Annals of Probability*, 16(1):189–199, 1988.

[McD89] Colin McDiarmid. On the method of bounded differences. In *Surveys in Combinatorics, 1989*, volume 141 of *London Mathematical Society Lecture Note Series*, pages 148–188. Springer, 1989.

[McM56] Brockway McMillan. Two inequalities implied by unique decipherability. *IRE Transactions on Information Theory*, 2(4):115–116, 1956.

[Mos09] Robin A. Moser. A constructive proof of the lovász local lemma. In *STOC*, pages 343–350, 2009.

[Mos10] Dana Moshkovitz. An alternative proof of the schwartz–zippel lemma. *Electron. Colloquium Comput. Complex.*, 2010:96, 2010.

[Mos12] Robin A. Moser. *Exact Algorithms for Constraint Satisfaction Problems*. PhD thesis, ETH Zurich, 2012.

[MT10] Robin A. Moser and Gábor Tardos. A constructive proof of the general lovász local lemma. *J. ACM*, 57(2), 2010.

[MU17] Michael Mitzenmacher and Eli Upfal. *Probability and Computing: Randomization and Probabilistic Techniques in Algorithms and Data Analysis*. Cambridge University Press, Cambridge, UK, 2nd edition, 2017.

[NS22] Ngoc Khanh Nguyen and Gregor Seiler. Practical sublinear proofs for R1CS from lattices. In *CRYPTO*, pages 133–162, 2022.

[NS24] Ngoc Khanh Nguyen and Gregor Seiler. Greyhound: Fast polynomial commitments from lattices. In *CRYPTO*, pages 243–275, 2024.

[Pap91] Christos H. Papadimitriou. On selecting a satisfying truth assignment. In *FOCS*, pages 163–169, 1991.

[Per92] Yuval Peres. Iterating Von Neumann's procedure for extracting random bits. *The Annals of Statistics*, 20(1):590–597, 1992.

[Pós76] László Pósa. Hamiltonian circuits in random graphs. *Discrete Mathematics*, 14(4):359–364, 1976.

[PR06] Chris Peikert and Alon Rosen. Efficient collision-resistant hashing from worst-case assumptions on cyclic lattices. In *TCC*, pages 145–166, 2006.

[PR07] Chris Peikert and Alon Rosen. Lattices that admit logarithmic worst-case to average-case connection factors. In *STOC*, pages 478–487, 2007.

[Rag88] Prabhakar Raghavan. Probabilistic construction of deterministic algorithms: Approximating packing integer programs. *J. Comput. Syst. Sci.*, 37(2):130–143, 1988.

[Sch80] Jacob T. Schwartz. Fast probabilistic algorithms for verification of polynomial identities. *J. ACM*, 27(4), 1980.

[Sha48a] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, 1948.

[Sha48b] Claude E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(4):623–656, 1948.

[Sha56] Claude E. Shannon. The zero error capacity of a noisy channel. *IRE Transactions on Information Theory*, IT-2(3):8–19, 1956.

[Spe28] Emanuel Sperner. Ein satz über untermengen einer endlichen menge. *Mathematische Zeitschrift*, 27:544–548, 1928.

[Spe77] Joel Spencer. Asymptotic lower bounds for ramsey functions. *Discrete Mathematics*, 20(1):69–76, 1977.

[SS87] Eli Shamir and Joel Spencer. Sharp concentration of the chromatic number on random graphs $G_{n,p}$. *Combinatorica*, 7:121–129, 1987.

[Top07] Flemming Topsøe. Some bounds for the logarithmic function. In Yeol Je Cho, Jong Kyu Kim, and Sever S. Dragomir, editors, *Inequality Theory and Applications*, volume 4, pages 137–151. Nova Science Publishers, New York, 2007.

[Tse63] Michael Lvovitch Tsetlin. Finite automata and models of simple forms of behaviour. *Russian Mathematical Surveys*, 18(4):1–27, 1963.

[Vad12] Salil P. Vadhan. *Pseudorandomness*. Now Publishers Inc., Hanover, MA, 2012.

[VB81] Leslie G. Valiant and Gordon J. Brebner. Universal schemes for parallel communication. In *STOC*, pages 263–277, 1981.

[von51] John von Neumann. Various techniques used in connection with random digits. In A. S. Householder, G. E. Forsythe, and H. H. Germond, editors, *Monte Carlo Method*, volume 12 of *National Bureau of Standards Applied Mathematics Series*, pages 36–38. U.S. Government Printing Office, Washington, DC, 1951. Summary written by George E. Forsythe.

[Wal47] Abraham Wald. *Sequential Analysis*. John Wiley & Sons, 1947.

[YAZ+19] Rupeng Yang, Man Ho Au, Zhenfei Zhang, Qiuliang Xu, Zuoxia Yu, and William Whyte. Efficient lattice-based zero-knowledge arguments with standard soundness: Construction and applications. In *CRYPTO*, pages 147–175, 2019.

[Zip79] Richard Zippel. Probabilistic algorithms for sparse polynomials. In *EUROSAM*, 1979.

A Degree-Four Bound for Binomial Compression

Using the setup of lemma 4.14.10 and the variables X_i from its proof, we derive a degree-4 polynomial upper bound for $\mathbb{E}[\exp(-sX_i^2)]$. We retain the parameter t from eq. (4.72).

The degree-2 construction leading to eq. (4.73) extends to two positive touching points. We derive these points from the moments and check the binomial corrections separately.

Exact moments up to order eight. The entries r_{ij} have vanishing odd moments, and every positive even moment is $1/2$. To compute the moments of the sum, let $S_j = \sum_{k=1}^{j} r_{ik} w_k$. For each positive integer r, independence and the binomial expansion give

$$\mathbb{E}[S_j^{2r}] = \mathbb{E}[S_{j-1}^{2r}] + \frac{1}{2}\sum_{k=1}^{r} \binom{2r}{2k} w_j^{2k}\, \mathbb{E}[S_{j-1}^{2r-2k}].$$

Here $S_0 = 0$, with zeroth moment equal to 1. For example, the surviving terms of the sixth moment have exponent patterns (6), (4, 2), and (2, 2, 2), so

$$\mathbb{E}[X_i^6] = \frac{1}{2}\|\overline{w}\|_6^6 + \frac{15}{4}\sum_{j\neq k} w_j^4 w_k^2 + \frac{45}{4}\sum_{j<k<\ell} w_j^2 w_k^2 w_\ell^2.$$

The second and fourth moments are given in eq. (4.71). Collecting the terms above, and applying the same recurrence through order eight, gives the remaining exact identities

$$\begin{aligned}
\mathbb{E}[X_i^6] &= \frac{15}{8}\|\overline{w}\|_2^6 - \frac{15}{8}\|\overline{w}\|_2^2\|\overline{w}\|_4^4 + \frac{1}{2}\|\overline{w}\|_6^6, \\
\mathbb{E}[X_i^8] &= \frac{105}{16}\|\overline{w}\|_2^8 - \frac{105}{8}\|\overline{w}\|_2^4\|\overline{w}\|_4^4 \\
&\quad + 7\|\overline{w}\|_2^2\|\overline{w}\|_6^6 + \frac{35}{16}\|\overline{w}\|_4^8 - \frac{17}{8}\|\overline{w}\|_8^8.
\end{aligned} \tag{A.1}$$

We retain these identities rather than replacing every moment by an upper bound: the coefficients of the polynomial alternate in sign.

Deriving the two touching points from the moments. We use the polynomial form in eq. (4.75), with coefficients determined below. The terms in eqs. (4.71) and (A.1) depending only on $\|\overline{w}\|_2$ are

$$m_1 = \frac{\|\overline{w}\|_2^2}{2}, \quad m_2 = \frac{3\|\overline{w}\|_2^4}{4}, \quad m_3 = \frac{15\|\overline{w}\|_2^6}{8}, \quad m_4 = \frac{105\|\overline{w}\|_2^8}{16}.$$

These are not the full binomial moments. We first seek positive weights and points $0 < a_- < a_+$ such that

$$1 - A_4 m_1 + B_4 m_2 - C_4 m_3 + D_4 m_4 = \lambda_0 F_4(0) + \lambda_- F_4(a_-) + \lambda_+ F_4(a_+). \tag{A.2}$$

Matching coefficients requires

$$\lambda_0 + \lambda_- + \lambda_+ = 1, \qquad \lambda_- a_-^r + \lambda_+ a_+^r = m_r \quad (r = 1, 2, 3, 4).$$

Each of the two points satisfies $a_\pm^2 = (a_- + a_+)a_\pm - a_- a_+$. Multiplying by further powers and taking the same weighted sums eliminates the weights:

$$\begin{cases} m_3 = (a_- + a_+)m_2 - a_- a_+ m_1, \\ m_4 = (a_- + a_+)m_3 - a_- a_+ m_2. \end{cases}$$

Substitution of the four values of m_r gives

$$a_- + a_+ = 5\|\overline{w}\|_2^2, \qquad a_- a_+ = \frac{15}{4}\|\overline{w}\|_2^4.$$

Therefore the touching points are

$$a_- = \frac{5-\sqrt{10}}{2}\,\|\overline{w}\|_2^2\,, \qquad a_+ = \frac{5+\sqrt{10}}{2}\,\|\overline{w}\|_2^2\,. \tag{A.3}$$

The first two moment equations then give

$$\lambda_- = \frac{a_+ m_1 - m_2}{a_-(a_+ - a_-)}, \qquad \lambda_+ = \frac{m_2 - a_- m_1}{a_+(a_+ - a_-)}.$$

Simplifying, all three weights are positive:

$$\lambda_0 = \frac{8}{15}, \qquad \lambda_- = \frac{7+2\sqrt{10}}{30}, \qquad \lambda_+ = \frac{7-2\sqrt{10}}{30}.$$

For a polynomial satisfying $F_4(a) \geq \exp(-sa)$, the right-hand side of eq. (A.2) is smallest when the polynomial equals the exponential at a_- and a_+. Applying the same tangency condition as in eq. (4.73), we impose

$$F_4(a_\pm) = \exp(-sa_\pm), \qquad F_4'(a_\pm) = -s\exp(-sa_\pm).$$

These four equations determine the four coefficients. The construction below also verifies that their solution is an upper bound on the whole nonnegative axis.

Deriving the coefficients. First construct a cubic H that matches the two values and two slopes. Write $d = a_+ - a_- > 0$ and $E_\pm = \exp(-sa_\pm)$, and start at the left point:

$$H(a) = E_- - sE_-(a-a_-) + k_2(a-a_-)^2 + h_3(a-a_-)^3.$$

The two conditions at a_- already hold. The conditions at a_+ are

$$\begin{cases} k_2 d^2 + h_3 d^3 = E_+ - E_- + sdE_-, \\ 2k_2 d + 3h_3 d^2 = s(E_- - E_+). \end{cases}$$

Solving this two-by-two system gives

$$h_3 = \frac{2(E_- - E_+) - sd(E_- + E_+)}{d^3}, \qquad k_2 = \frac{(2sd-3)E_- + (sd+3)E_+}{d^2}.$$

Expanding $H(a) = h_0 + h_1 a + h_2 a^2 + h_3 a^3$, we obtain

$$\begin{aligned} h_2 &= k_2 - 3a_- h_3, \\ h_1 &= -sE_- - 2a_- k_2 + 3a_-^2 h_3, \\ h_0 &= (1+sa_-)E_- + a_-^2 k_2 - a_-^3 h_3. \end{aligned}$$

The quartic term $(a-a_-)^2(a-a_+)^2$ and its derivative both vanish at the touching points. We can therefore add a multiple of this term without changing any of the four conditions. Choose that multiple to make the value at 0 equal to 1:

$$F_4(a) = H(a) + D_4(a-a_-)^2(a-a_+)^2, \qquad D_4 = \frac{1-h_0}{a_-^2 a_+^2}.$$

Using the sum and product of the touching points,

$$\begin{aligned} (a-a_-)^2(a-a_+)^2 = a^4 - 10\,\|\overline{w}\|_2^2\,a^3 + \frac{65}{2}\,\|\overline{w}\|_2^4\,a^2 \\ - \frac{75}{2}\,\|\overline{w}\|_2^6\,a + \frac{225}{16}\,\|\overline{w}\|_2^8\,. \end{aligned}$$

Consequently the coefficients in the original variable a are

$$\begin{aligned} D_4 &= \frac{16(1-h_0)}{225\,\|\overline{w}\|_2^8}, \\ C_4 &= 10\,\|\overline{w}\|_2^2\,D_4 - h_3, \\ B_4 &= \frac{65}{2}\,\|\overline{w}\|_2^4\,D_4 + h_2, \\ A_4 &= \frac{75}{2}\,\|\overline{w}\|_2^6\,D_4 - h_1. \end{aligned} \tag{A.4}$$

All these coefficients depend on the same free parameter s through $E_\pm = \exp(-sa_\pm)$; no numerical value of s has been fixed.

The polynomial upper bounds the exponential. The interpolation remainder has one factor for matching the value at 0 and two factors for matching the value and slope at each positive point. For each $a \geq 0$ away from the touching points, the Hermite remainder is

$$\exp(-sa) - F_4(a) = -\frac{s^5\exp(-s\xi)}{5!}a(a-a_-)^2(a-a_+)^2 \leq 0,$$

for some ξ in the interval containing $0, a, a_-, a_+$. At the touching points equality holds by construction. Thus $\exp(-sa) \leq F_4(a)$ for every $a \geq 0$ and $s \geq 0$.

We will also need three coefficient inequalities. First, applying the cubic remainder to H at 0 gives

$$1 - H(0) = \frac{s^4\exp(-s\xi_0)}{4!}a_-^2a_+^2 \geq 0,$$

so $D_4 \geq 0$. The numerator in eq. (4.74) was shown to be nonnegative. Applying this with $x = sd \geq 0$, the cubic coefficient satisfies

$$h_3 = -\frac{E_-}{d^3}\left(x - 2 + (x+2)\exp(-x)\right) \leq 0.$$

Therefore

$$C_4 \geq 10\,\|\overline{w}\|_2^2\,D_4. \tag{A.5}$$

Finally, subtract the endpoint derivatives of the cubic:

$$H'(a_+) - H'(a_-) = d\left(2h_2 + 15\,\|\overline{w}\|_2^2\,h_3\right) = s(E_- - E_+).$$

Substituting eq. (A.4) gives

$$-2B_4 + 15\,\|\overline{w}\|_2^2\,C_4 - 85\,\|\overline{w}\|_2^4\,D_4 = -\frac{s(E_- - E_+)}{d} \leq 0. \tag{A.6}$$

Each inequality holds throughout $s \geq 0$, including $s = 0$ when F_4 is identically 1.

The binomial moment correction is nonpositive. Substitution of the exact moments gives

$$\mathbb{E}[F_4(X_i^2)] = 1 - A_4m_1 + B_4m_2 - C_4m_3 + D_4m_4 + \Delta,$$

where

$$\begin{aligned} \Delta = &\left(-\frac{B_4}{4} + \frac{15C_4\,\|\overline{w}\|_2^2}{8} - \frac{105D_4\,\|\overline{w}\|_2^4}{8}\right)\|\overline{w}\|_4^4 \\ &+ \left(7D_4\,\|\overline{w}\|_2^2 - \frac{C_4}{2}\right)\|\overline{w}\|_6^6 \\ &+ D_4\left(\frac{35\,\|\overline{w}\|_4^8}{16} - \frac{17\,\|\overline{w}\|_8^8}{8}\right). \end{aligned}$$

The coefficient inequalities proved above imply

$$-\frac{B_4}{4}+\frac{15C_4\|\overline{w}\|_2^2}{8}-\frac{105D_4\|\overline{w}\|_2^4}{8}\le-\frac{5}{2}D_4\|\overline{w}\|_2^4,$$
$$7D_4\|\overline{w}\|_2^2-\frac{C_4}{2}\le 2D_4\|\overline{w}\|_2^2.$$

Hence

$$\Delta\le D_4\Big(-\frac{5}{2}\|\overline{w}\|_2^4\|\overline{w}\|_4^4+2\|\overline{w}\|_2^2\|\overline{w}\|_6^6 + \frac{35}{16}\|\overline{w}\|_4^8-\frac{17}{8}\|\overline{w}\|_8^8\Big).$$

We use two inequalities between the coefficient sums. Term by term,

$$\|\overline{w}\|_2^4 w_j^4+w_j^8-2\|\overline{w}\|_2^2 w_j^6=w_j^4\left(\|\overline{w}\|_2^2-w_j^2\right)^2\ge 0,$$

and consequently $2\|\overline{w}\|_2^2\|\overline{w}\|_6^6\le\|\overline{w}\|_2^4\|\overline{w}\|_4^4+\|\overline{w}\|_8^8$. Also the weights $p_j=w_j^2/\|\overline{w}\|_2^2$ sum to 1. Jensen's inequality for the cube function on nonnegative inputs gives

$$\frac{\|\overline{w}\|_8^8}{\|\overline{w}\|_2^2}=\sum_j p_j(w_j^2)^3\ge\left(\sum_j p_j w_j^2\right)^3=\left(\frac{\|\overline{w}\|_4^4}{\|\overline{w}\|_2^2}\right)^3.$$

Thus $\|\overline{w}\|_8^8\ge\|\overline{w}\|_4^{12}/\|\overline{w}\|_2^4$. Since $D_4\ge 0$, these two inequalities preserve the upper-bound direction. Writing $\rho=\|\overline{w}\|_4^4/\|\overline{w}\|_2^4$, we obtain

$$\begin{aligned}\Delta&\le D_4\left(-\frac{3}{2}\|\overline{w}\|_2^4\|\overline{w}\|_4^4+\frac{35}{16}\|\overline{w}\|_4^8-\frac{9}{8}\|\overline{w}\|_8^8\right)\\&\le D_4\|\overline{w}\|_2^4\|\overline{w}\|_4^4\left(-\frac{3}{2}+\frac{35\rho}{16}-\frac{9\rho^2}{8}\right)\\&=-D_4\|\overline{w}\|_2^4\|\overline{w}\|_4^4\left(\frac{9}{8}\left(\rho-\frac{35}{36}\right)^2+\frac{503}{1152}\right)\le 0.\end{aligned}$$

This proves the comparison with the moment terms depending only on $\|\overline{w}\|_2$, uniformly in $\overline{w}$ and throughout $s\ge 0$.

The expectation bound and its numerical consequence. Since $F_4(0)=1$ and $F_4(a_\pm)=\exp(-sa_\pm)$, eq. (A.2) now yields

$$\begin{aligned}\mathbb{E}[\exp(-sX_i^2)]\le\mathbb{E}[F_4(X_i^2)]\le L(t),&\\ L(t):=\frac{8}{15}+\frac{7+2\sqrt{10}}{30}\exp\left(-\frac{5-\sqrt{10}}{2}t\right)&\\ +\frac{7-2\sqrt{10}}{30}\exp\left(-\frac{5+\sqrt{10}}{2}t\right).&\end{aligned}\tag{A.7}$$

This holds for every $t\ge 0$. The moment matching minimizes the expression $1-A_4m_1+B_4m_2-C_4m_3+D_4m_4$ among degree-4 polynomial upper bounds. The separate argument for $\Delta\le 0$ is what permits its use for the actual binomial variables.

Substituting eq. (A.7) into eq. (4.69), and using $b\le\|\overline{w}\|_2$, bounds the small-coordinate probability by

$$\Phi(t):=\exp(16t)L(t)^{256}.$$

We minimize this expression including its exponential prefactor. Put $u = (5 - \sqrt{10})/2$ and $v = (5 + \sqrt{10})/2$, so $L(t) = \lambda_0 + \lambda_- \exp(-ut) + \lambda_+ \exp(-vt)$. Since $L(t) > 0$, the sign of $\Phi'(t)$ is the sign of

$$G(t) = L(t) + 16L'(t) = \lambda_0 + \lambda_-(1 - 16u)\exp(-ut) + \lambda_+(1 - 16v)\exp(-vt).$$

Both $u, v > 1/16$, so

$$G'(t) = \lambda_- u(16u - 1)\exp(-ut) + \lambda_+ v(16v - 1)\exp(-vt) > 0.$$

Moreover $G(0) = 1 - 16/2 = -7$ and $\lim_{t\to\infty} G(t) = 8/15 > 0$. Thus G has exactly one zero, which is the unique global minimum of Φ on the nonnegative axis. Numerically,

$$t_* \approx 2.649584348025418, \qquad \log_2 \Phi(t_*) \approx -144.987751949687.$$

The rational choice $t = 53/20$ gives essentially the same value:

$$\log_2 \Phi(53/20) \approx -144.987750241968.$$

For a certified inequality, rational interval bounds for $\sqrt{10}$ and the exponential give

$$\Phi(53/20) < \frac{101}{200}\, 2^{-144} < 2^{-144}.$$

The numerical minimum remains slightly above 2^{-145}, and satisfies the target in lemma 4.14.10.

YUP, FUCK THIS SHIT.
I AIN'T EVER GONNA PLAY THIS SHIT GAME AGAIN.
.....