

A SAMPLING LOVÁSZ LOCAL LEMMA FOR LARGE DOMAIN SIZES

AUTHOR(S)

ABSTRACT. We present polynomial-time algorithms for approximate counting and sampling solutions to constraint satisfaction problems (CSPs) with atomic constraints within the local lemma regime:

$$pD^{2+o_q(1)} \lesssim 1.$$

When the domain size q of each variable becomes sufficiently large, this almost matches the known lower bound $pD^2 \gtrsim 1$ for approximate counting and sampling solutions to atomic CSPs [BGG⁺19, GGW22], thus establishing an almost tight sampling Lovász local lemma for large domain sizes.

CONTENTS

1. Introduction	2
2. Preliminaries and notations	6
3. Constraint-wise coupling with exponential decay of correlation	9
4. Linear programming from constraint-wise coupling	15
5. Counting and sampling via linear programming	22
6. Conclusion and open problems	25
References	25
Appendix A. Basic properties of the linear program	28
Appendix B. Probabilistic properties of the linear program	29

1. INTRODUCTION

Constraint satisfaction problems (CSPs) are ubiquitous in Computer Science, and their solution spaces have been a subject of great research interest. A CSP is represented by a collection of constraints defined on a set of variables, where a solution is an assignment of variables such that all constraints are satisfied. A cornerstone tool in studying CSP solution spaces is the *Lovász local lemma* [EL75], which establishes the following sufficient condition for the existence of a CSP solution by interpreting the space of assignments as a product probability space and the violation of each constraint as a bad event:

$$(1) \quad ep(D + 1) \leq 1,$$

where p stands for the maximum *violation probability* of each constraint, and D stands for the *dependency degree*, given by the maximum number of other constraints that a constraint can share variables with. This condition (1) was later shown to be essentially tight [She85]. Subsequent work on the *algorithmic Lovász local lemma* seeks to constructively find a CSP solution by efficient algorithms. This has led to a long line of research [Bec91, Alo91, MR99, CS00, Sri08, Mos09, MT10], culminating in algorithms for efficiently finding a CSP solution up to the condition in (1). Together, these contributions establish a sharp threshold for the existence/construction of CSP solutions.

On the other hand, a considerable amount of work has been focused on the *counting/sampling Lovász local lemma* [BGG⁺19, HSZ19, Moi19, GLLZ19, FGYZ21a, FHY21, JPV21a, JPV21b, HSW21, GGW22, QWZ22, FGW22, HWY22, HWY23a, QW24], which aims to characterize a local lemma type regime under which the problem of (approximately) counting or (almost uniformly) sampling CSP solutions is tractable. Hardness results in [BGG⁺19, GGW22] have shown that this counting/sampling variant of LLL requires a strictly stronger condition $pD^2 \lesssim 1$, where \lesssim hides lower-order factors and constants. This holds true even when restricted to some canonical sub-classes of CSPs, such as k -CNFs or hypergraph colorings. Regarding upper bounds, the current state-of-the-art [HWY23a] shows that counting/sampling CSP solutions is efficiently solvable under the condition $pD^5 \lesssim 1$. However, the correct threshold for the counting/sampling LLL is not yet clear. The following question is fundamental to our understanding of the critical phenomenon for counting and sampling CSP solutions:

Is $pD^2 \lesssim 1$ the correct threshold for the counting/sampling Lovász local lemma?

Despite numerous works on the topic and successive improvements on the algorithmic threshold, current techniques have encountered barriers towards closing such gaps due to their reliance on a “freezing” paradigm. Originally introduced by Beck [Bec91] to handle the non-self-reducibility of the local lemma regime, the “freezing” paradigm ensures that a local lemma type condition is invariantly satisfied under arbitrary pinnings produced in the process of incrementally constructing a satisfying CSP solution, which imposes slackness in the local lemma regime. As a result, the best upper bound obtained using this paradigm was $pD^4 \lesssim 1$ [Sri08, Alo91, MR99], where the extra D^3 factor came from the use of a structure called $\{2, 3\}$ -tree.

In the realm of the counting and sampling Lovász local lemma, Beck’s technique is highly prevalent. Moitra’s seminal work [Moi19] on counting k -SAT solutions introduced the method of “mark/unmark”, where a local lemma type condition is preserved under arbitrary pinning of marked variables. This “mark/unmark” method also introduced slackness in the local lemma regime, and can be viewed as a static version of the “freezing” paradigm. Subsequently, this static version of “freezing” was later generalized to non-static settings [GLLZ19, JPV21b, HWY22, HWY23a, HWY23b], and also enabled the development of several Markov chain Monte Carlo approaches [FGYZ21a, FHY21, JPV21a, HSW21, GGGHP22, CMM23] for fast sampling. To this day, the current best regime $pD^5 \lesssim 1$ for sampling general CSPs still relies on the idea of “freezing”. The slackness in this case arises from the use of a structure called generalized $\{2, 3\}$ -trees [HWY22, HWY23a], which is similar to the source of slackness in [Sri08].

These advancements suggest a barrier in the current approach to counting and sampling LLL. Novel techniques are required to establish a tight counting and sampling Lovász local lemma.

1.1. Our results. In this work, we establish the tractability of approximate counting and sampling solutions to atomic constraint satisfaction problems within an improved local lemma regime. As the domain size of each variable increases, this condition approaches $pD^2 \lesssim 1$, providing a positive answer the major open question on counting and sampling LLL, particularly for large domain sizes.

A CSP instance (or a CSP formula) is denoted by $\Phi = (V, \mathcal{Q}, C)$. Here, V is a set of $n = |V|$ variables; each variable $v \in V$ is endowed with a finite domain Q_v , altogether $\mathcal{Q} \triangleq \bigotimes_{v \in V} Q_v$; and C is a set of constraints, where each constraint $c \in C$ is a function $c : \bigotimes_{v \in \text{vbl}(c)} Q_v \rightarrow \{\text{True}, \text{False}\}$ defined on a subset of variables, denoted by $\text{vbl}(c) \subseteq V$. A CSP formula $\Phi = (V, \mathcal{Q}, C)$ is called *atomic* if each constraint $c \in C$ is violated by exactly one assignment in the domain of c , i.e., $|c^{-1}(\text{False})| = 1$. The atomicity of constraints is typical for many classical constraint satisfaction problems, including k -CNF and hypergraph colorings¹, and is a natural assumption in previous studies of LLL [AI14, HV15, Kol16, HS17a, HS17b, HS19, AIS19, Har21, FHY21, JPV21a, HSW21].

Given a CSP formula $\Phi = (V, \mathcal{Q}, C)$, we use \mathcal{P} to denote the uniform (product) distribution over all possible assignments in \mathcal{Q} . The following parameters Φ can be defined:

- *width* $k \triangleq \max_{c \in C} |\text{vbl}(c)|$;
- *minimum domain size* $q_{\min} \triangleq \min_{v \in V} |Q_v|$;
- *maximum domain size* $q_{\max} \triangleq \max_{v \in V} |Q_v|$;
- *dependency degree* $D \triangleq \max_{c \in C} |\{c' \in C \setminus \{c\} \mid \text{vbl}(c) \cap \text{vbl}(c') \neq \emptyset\}|$;
- *maximum violation probability* $p \triangleq \max_{c \in C} \mathcal{P}[-c]$.

The following condition characterizes the local lemma regime achieved in this paper.

Condition 1. $\Phi = (V, \mathcal{Q}, C)$ is an atomic CSP formula satisfying

$$(2) \quad (8e)^3 \cdot p \cdot (D + 1)^{2+\zeta} \leq 1,$$

where

$$\zeta = \frac{2 \ln(2 - 1/q_{\min})}{\ln q_{\min} - \ln(2 - 1/q_{\min})}.$$

Our main results are the following algorithms for counting/sampling LLL under Condition 1.

Theorem 1.1 (counting LLL). *Assume Condition 1 for $\Phi = (V, \mathcal{Q}, C)$. There exists a deterministic algorithm that, given any Φ and any $\varepsilon \in (0, 1)$, outputs an estimate \hat{Z} such that*

$$(1 - \varepsilon)Z_{\Phi} \leq \hat{Z} \leq (1 + \varepsilon)Z_{\Phi}$$

within time $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q_{\max})}\right)$, where Z_{Φ} represents the number of solutions to Φ .

Theorem 1.2 (sampling LLL). *Assume Condition 1 for $\Phi = (V, \mathcal{Q}, C)$. There exists an algorithm that, given any Φ and any $\varepsilon \in (0, 1)$, outputs a random assignment $X \in \mathcal{Q}$ distributed as $\hat{\mu}$ such that*

$$d_{\text{TV}}(\hat{\mu}, \mu_{\Phi}) \leq \varepsilon$$

within time $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q_{\max})}\right)$, where μ_{Φ} represents the uniform distribution over all solutions to Φ .

Remark 1.3 (local lemma regime). The $\zeta = \zeta(q)$ in Condition 1 decreases monotonically as q increases. In particular, as q tends to infinity, ζ approaches 0, leading Condition 1 to approach $pD^2 \lesssim 1$, which matches the $pD^2 \gtrsim 1$ lower bound [BGG⁺19, GGW22]. Conversely, with $q = 2$ (the Boolean domain), ζ achieves its maximum value of $\log_{8/3}(9/4) \approx 2.82$. In this case, Condition 1 becomes $pD^{4.82} \lesssim 1$, improving upon the previous best upper bound of $pD^5 \lesssim 1$ for counting/sampling LLL [HWY23a].

Remark 1.4 (non-uniform width). Condition 1 does not involve the width k , making it applicable to CSP formulas with non-uniform widths. This aspect is particularly desirable from the LLL perspective. Previously, such generality was only attained in [FHY21] under a local lemma type condition $pD^{350} \lesssim 1$.

Remark 1.5 (time complexity). The $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q)}\right)$ time complexities of Theorems 1.1 and 1.2 align with previous results for deterministic counting LLL [Moi19, GLLZ19, JPV21b, HWY23a, FGW⁺23]. In fact, such complexity bounds with k, D in the exponents appear to be intrinsic to deterministic approximate counting to this day, which relies on exhaustive enumerations of local structures.

We then apply our results to two typical classes of atomic CSPs: hypergraph q -colorings and k -CNF.

¹For hypergraph q -colorings, each hyperedge may correspond to q atomic constraints, each of which forbids the edge to be colored monochromatically with one particular color.

1.1.1. *Application to hypergraph colorings.* Let $H = (V, \mathcal{E})$ be a k -uniform hypergraph, where $|e| = k$ for all $e \in \mathcal{E}$. A proper hypergraph q -coloring $X \in [q]^V$ assigns one of the q colors to each $v \in V$, ensuring no edge is monochromatic. We use Δ to denote the maximum degree of the hypergraph, i.e., each vertex belongs to at most Δ hyperedges. Hypergraph colorings are foundational combinatorial objects that have been a key to the discovery of the Lovász Local lemma [EL75]. The local lemma was initially devised to show that a proper q -coloring exists if $q \geq C\Delta^{\frac{1}{k}}$, where C is a sufficiently large constant.

For the problem of approximate counting/sampling proper hypergraph q -colorings, the standard local Markov chains are efficient for sampling when $q \geq C\Delta$ for some constant C [BDK06, BDK08], which is also necessary for the irreducibility of local Markov chains [FM11]. For *simple* hypergraphs, where any two distinct hyperedges share at most one vertex, it has been shown that local Markov chains rapidly mix when $q \geq \max\{C_k \log n, 500k^3\Delta^{\frac{1}{k-3}}\}$ [FM11, FA17], for hypergraphs with n vertices. The first result for efficiently approximate counting/sampling hypergraph colorings under LLL condition is [GLLZ19]. By adapting Moitra’s technique to a non-static “freezing”, they obtained polynomial-time algorithms in the local lemma regime $q \geq 357\Delta^{\frac{14}{k-14}}$ for $k \geq 28$. This regime has been improved in several subsequent works [FHY21, JPV21b, JPV21a, HSW21], resulting in the current best upper bound $q \geq 310\Delta^{\frac{3}{k-3}}$ for $k \geq 24$ in [JPV21a, HSW21]. On the lower bound side, it was proved in [GGW22] that no efficient algorithm exists for the problem when $q \gtrsim \Delta^{\frac{2}{k}}$ unless $\mathbf{NP}=\mathbf{RP}$.

Our results produce the following corollary on counting/sampling proper hypergraph q -colorings, closing the gap between the current upper and lower bounds for the problem.

Corollary 1.6 (counting/sampling hypergraph q -colorings). *For any $\zeta > 0$, there exists a finite q_0 such that given any k -uniform hypergraph $H = (V, \mathcal{E})$ on n vertices with maximum degree Δ and $q \geq q_0$ colors, if $k \geq 8$ and*

$$q \geq 70\Delta^{\frac{2+\zeta}{k-2-\zeta}},$$

then for any $\varepsilon \in (0, 1)$:

- (approximate counting) the total number of proper q -colorings of H can be estimated within relative error $1 \pm \varepsilon$ deterministically within time $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q)}\right)$;
- (almost uniform sampling) an almost uniform random q -coloring $X \in [q]^V$ can be generated within time $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q)}\right)$, such that the distribution of X is ε -close (in total variation distance) to the uniform distribution over all proper q -colorings of H .

The $q_0 = q_0(\zeta)$ in Corollary 1.6 grows as $q_0 = \exp(O(1/\zeta))$ in ζ .

1.1.2. *Application to k -CNF.* A standard setting for LLL is the k -CNF (conjunctive normal form) with uniform width k and maximum variable-degree d . A CNF is called a (k, d) -formula if each clause has exactly k distinct variables and each variable is contained in at most d distinct clauses.

For approximate counting/sampling k -CNF solutions, the breakthrough of Moitra [Moi19] was the first to give an efficient algorithm in the LLL regime $k \geq 60 \log k + 60 \log d + 300$. This bound was then improved in a series of works [FGYZ21a, JPV21b, JPV21a, HSW21, HWY23a], with the current best bound given in [HWY23a] as $k \geq 5 \log d + 5 \log k + O(1)$. The lower bound proved in [BGG⁺19] says that unless $\mathbf{NP}=\mathbf{RP}$, $k \gtrsim 2 \log d + 2 \log k$ is necessary for efficient algorithms to exist.

Our results improve the current best bound for approximate counting/sampling k -CNF solutions.

Corollary 1.7 (counting/sampling k -CNF solutions). *Given any (k, d) -formula Φ on n variables, if*

$$k \geq 4.82 \log k + 4.82 \log d + 5,$$

then for any $\varepsilon \in (0, 1)$:

- (approximate counting) the total number of satisfying assignments for Φ can be estimated within relative error $1 \pm \varepsilon$ deterministically within time $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q)}\right)$;
- (almost uniform sampling) an almost uniform random satisfying assignment $X \in \{\text{True}, \text{False}\}^V$ can be generated within time $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q)}\right)$, such that the distribution of X is ε -close (in total variation distance) to the uniform distribution over all satisfying assignments for Φ .

1.2. **Technique overview.** We then outline our general approach for the main results (Theorems 1.1 and 1.2) and provide a brief overview of the techniques.

Our high-level approach falls into the following two-step framework:

- (1) In the first step, we construct a coupling between the joint distributions defined by two CSP instances that differ locally. We manage to show that the discrepancy in this coupling decays at an exponential rate in the local lemma regime characterized by Condition 1.
- (2) Next, we leverage the coupling constructed in the first step to devise a linear program. By mimicking the transitions of the coupling procedure, this linear program can efficiently bootstrap the marginal ratios as long as the original coupling procedure converges efficiently.

This LP-based framework for counting LLL was initiated in the seminal work of Moitra [Moi19] and was subsequently generalized and refined in [GLLZ19, GGGY20, JPV21b].

Despite this high-level framework, our method differs fundamentally from all previous works and circumvents the barrier encountered by the previous approaches.

Our method completely dispenses with the “freezing” paradigm relied upon in previous works to establish correlation decay properties in the local lemma regime. In addition, it also eliminates the need for ensuring a “factorization property” that after properly pinning the marked variables, the residual formula breaks down into logarithmic-sized connected components, which was required in previous works to achieve counting/sampling using marginal estimators through an exhaustive enumeration.²

Specifically, we introduce a novel constraint-wise coupling, whose exponential decay of correlation is due to an average-case percolation-style analysis, thus avoiding to preserve a local-lemma-type condition with respect to a worst-case pinning. Also, we propose a “constraint-wise self-reducibility” for local lemma regime to replace the “factorization property” to obtain counting/sampling algorithms, providing a new perspective for sampling/counting LLL. Next, we provide a more in-depth discussion.

1.2.1. *Constraint-wise coupling with exponential decay of correlation.* The correlation decay properties have been a key to approximate counting and sampling in many classic and contemporary works [Dob70, GMP05, Wei06, BGK⁺07, SST12, LLY13, LY13, YZ13, GKM15, PR17, LSS19, ALO20, CLV20, CGŠV21, FGYZ21b]. For sampling CSP solutions, the correlation decay properties have also been used explicitly or implicitly in various previous algorithmic results [HSZ19, Moi19, BGG⁺19, GLLZ19, GGGY20, FGYZ21a, FHY21, JPV21a, JPV21b, HSW21, QWZ22, FGW22, HWY22, GGGHP22, HWY23a, FGW⁺23, CMM23, HWY23b].

Fix the variable set V with domain \mathcal{Q} , and for any set of constraints C use μ_C to denote the uniform distribution over solutions to $\Phi = (V, \mathcal{Q}, C)$. Our work establishes a correlation decay property within the local lemma regime of Condition 1. Such a correlation decay property is captured by the decay of discrepancy in a novel constraint-wise coupling between μ_C and the distribution with one fewer constraint $\mu_{C \setminus \{c_0\}}$ (formally stated in Theorem 3.1).

We expose the idea of this coupling. Decompose $\mu_{C \setminus \{c_0\}}$ using the law of total probability:

$$\mu_{C \setminus \{c_0\}} = \mu_{C \setminus \{c_0\}}(c_0) \cdot \mu_C + \mu_{C \setminus \{c_0\}}(\neg c_0) \cdot \mu_{C \setminus \{c_0\}}(\cdot \mid \neg c_0).$$

Inspired from this, we can establish a coupling between μ_C and $\mu_{C \setminus \{c_0\}}$ as follows:

- with probability $\mu_{C \setminus \{c_0\}}(c_0)$, establish a coupling between between identical distributions μ_C and μ_C , which can already be perfectly coupled;
- with probability $\mu_{C \setminus \{c_0\}}(\neg c_0)$, establish a coupling between between $\mu_{C \setminus \{c_0\}}(\cdot \mid \neg c_0)$ and μ_C .

Here, in the latter case, a negated constraint $\neg c_0$ was introduced into the distribution. For atomic CSP, this will force the assignment on $\text{vbl}(c_0)$ to be the unique violating assignment $\pi = \text{False}(c_0)$.

To address this issue, we further decompose μ_C as follows:

$$\mu_C = \sum_{\rho \in \mathcal{Q}_{\text{vbl}(c_0)}} \mu_C(\rho) \cdot \mu_C(\cdot \mid \rho),$$

which inspire us to achieve the coupling in this case as: first, sample $\rho \sim \mu_{C, \text{vbl}(c_0)}$ according to the marginal distribution $\mu_{C, \text{vbl}(c_0)}$ induced by μ_C on $\text{vbl}(c_0)$, and then, establish the coupling between the conditional distributions $\mu_{C \setminus \{c_0\}}(\cdot \mid \pi)$ and $\mu_C(\cdot \mid \rho)$. In this way, we eliminated the negated constraint

²See the last paragraph of Section 1.2 in [JPV21b] for a discussion on the specific barriers of this paradigm.

$\neg c_0$ from the distribution, albeit potentially introducing a discrepancy in the assignment over $\text{vbl}(c_0)$. It is important to note that the coupling $\mu_{C \setminus \{c_0\}}(\cdot \mid \pi)$ and $\mu_C(\cdot \mid \rho)$, after some simplification, is equivalent to the coupling of two distributions specified by two sets of constraints on the same variable set, thereby can be resolved recursively.

The recursively constructed coupling in our approach may remind us of the recursive coupling technique introduced by Goldberg, Martin, and Mike to prove strong spatial mixing [GMP05]. However, there are significant differences between our approach and theirs. Firstly, our coupling proceeds in a constraint-wise manner. Secondly, the main distinction arises from the fact that the LLL regime is not self-reducible. In their work [GMP05], self-reducibility plays a crucial role in the recursive coupling for establishing strong spatial mixing. The one-step worst-case contraction bounded there allows for the application of path coupling [BD97]. However, in local lemma regimes, self-reducibility is not guaranteed, and the LLL condition degrades after pinning too many variables. Consequently, we must keep track of the entire evolution of the coupling process and apply an average-case percolation-style analysis. An intriguing finding is that the random choices involved in our constraint-wise coupling procedure can be explicitly identified using the principle of deferred decision. This simplifies the analysis considerably and may be of independent interest.

1.2.2. Linear programs from constraint-wise coupling. We will transform the exponentially decayed coupling process constructed above into a linear program that mimics the transition probabilities of this coupling procedure. This linear program will be used to bootstrap the marginal ratios between the total numbers of CSP solutions to $\Phi = (V, Q, C)$ and $(V, Q, C \setminus \{c_0\})$.

Since our coupling proceeds in a constraint-wise manner and its analysis is considerably non-local, it is significantly different from previous works within the same general framework. Therefore, several new ideas are required for designing the linear program as well as the final counting/sampling algorithms:

- Unlike in all previous LP-based methods for counting LLL [Moi19, GLLZ19, GGGY20, JPV21b], a key property known as “local uniformity,” which used to be enforced by freezing, is no longer guaranteed in our setting. Instead of relying on “local uniformity,” which introduces slackness in the local lemma regimes, we utilize a 2-tree structure that provides certificates for overflow in the coupling procedure, as well as a new class of constraints in the linear program. This 2-tree structure, critical to our algorithm, closely aligns with similar structures that have appeared in the lower bounds [BGG⁺19, GGW22], shedding light on why our algorithms approach the critical threshold.
- In the process of converting a marginal estimator to an approximate counting or sampling algorithm within local lemma regimes where variable-wise self-reducibility does not hold, previous works relied on the “factorization” to complete partial assignments. However, this approach introduces slackness in local lemma regimes. Our finding shows that such additional slackness is unnecessary for efficient counting/sampling LLL. Instead, we utilize “constraint-wise self-reducibility” for local lemma regimes and show that the followings are efficient:
 - estimate the constraint-wise marginal probability $\mu_C(c)$;
 - incrementally update some sample $X \sim \mu_{C \setminus \{c\}}$ to $Y \sim \mu_C$.

2. PRELIMINARIES AND NOTATIONS

2.1. CSP defined by atomic constraints.

2.1.1. Basic concepts for CSPs. A CSP is described by a collection of constraints defined on a set of variables. Formally, an instance of a constraint satisfaction problem, called a *CSP formula*, is denoted by $\Phi = (V, Q(= \bigotimes_{v \in V} Q_v), C)$. Here, V is a set of $n = |V|$ random variables, where each random variable $v \in V$ is endowed with a finite domain Q_v of size $q_v \triangleq |Q_v| \geq 2$; and C gives a collection of local constraints, such that each $c \in C$ is a constraint function $c : \bigotimes_{v \in \text{vbl}(c)} Q_v \rightarrow \{\text{True}, \text{False}\}$ defined on a subset of variables, denoted by $\text{vbl}(c) \subseteq V$. An assignment $\sigma \in Q$ is called *satisfying* for Φ if

$$\Phi(\sigma) \triangleq \bigwedge_{c \in C} c(\sigma_{\text{vbl}(c)}) = \text{True}.$$

In the context of LLL, each constraint c can be interpreted as a bad event A_c , which happens when the assignment on $\text{vbl}(c)$ violates c . For any subset of constraints $\mathcal{E} \subseteq C$, denote $\text{vbl}(\mathcal{E}) \triangleq \bigcup_{c \in \mathcal{E}} \text{vbl}(c)$. For any subset of variables $\Lambda \subseteq V$, denote $\mathcal{Q}_\Lambda \triangleq \bigotimes_{v \in \Lambda} \mathcal{Q}_v$. Also, we say Φ is *satisfiable* if at least one satisfying assignment to Φ exists.

We say some constraint $c \in C$ is defined by *atomic bad events*, or simply, *atomic*, if it is violated by exactly one configuration in $\mathcal{Q}_{\text{vbl}(c)}$. For atomic constraints c , we use $\text{False}(c)$ to denote the only violating configuration of c in $\mathcal{Q}_{\text{vbl}(c)}$. Moreover, when all constraints in C are atomic, we say Φ is atomic.

2.1.2. Notations for (partial) assignments. For a partial assignment $\sigma \in \mathcal{Q}_\Lambda$ specified over a subset of variables $\Lambda \subseteq V$, we use $\Lambda(\sigma) = \Lambda$ to denote the set of assigned variables in σ . For any partial assignment σ and any $S \subseteq \Lambda(\sigma)$, we use σ_S to denote $\bigotimes_{v \in S} \sigma(v)$. We further write $\sigma_v = \sigma_{\{v\}}$ for $v \in V$.

For any two assignments σ, τ such that $\Lambda(\sigma) \cap \Lambda(\tau) = \emptyset$, we define $\sigma \wedge \tau \in \mathcal{Q}_{\Lambda(\sigma) \cup \Lambda(\tau)}$ as the concatenation of σ and τ such that for any $v \in \Lambda(\sigma) \cup \Lambda(\tau)$,

$$(\sigma \wedge \tau)(v) = \begin{cases} \sigma(v) & v \in \Lambda(\sigma), \\ \tau(v) & v \in \Lambda(\tau). \end{cases}$$

We will use \emptyset to specifically denote an empty assignment, distinguishing from the empty set \emptyset .

2.1.3. Notations for events and probability measures. We then specify some notations for events and probability measures related to the CSP.

Definition 2.1 (simple notations for events). For the simplicity of notations, we will use:

- any constraint $c \in C$ to denote the event that this constraint is satisfied;
- any subset of constraints $\mathcal{E} \subseteq C$ to denote the event that all constraints in \mathcal{E} are satisfied;
- any partial assignment σ to denote the event that the assignment on $\Lambda(\sigma)$ is precisely σ .

Note that under this definition, the notation $\sigma \wedge \tau$ as a concatenation of assignment is consistent with the same notation where σ and τ are considered as events, and $\sigma \wedge \tau$ is the logical and of the two events.

We use \mathcal{P} to denote the uniform product distribution over the space \mathcal{Q} . For any subset of variables $\Lambda \subseteq V$. We use $\mu = \mu_\Phi$ denote the distribution over all satisfying assignments of Φ induced by \mathcal{P} , i.e.

$$\mu_\Phi \triangleq \mathcal{P}(\cdot | C).$$

μ_Φ is well-defined only when Φ is satisfiable.

When the variable set V and the domain \mathcal{Q} is clear, for some set of constraints \mathcal{E} defined over V , and some assignment σ defined over $\Lambda(\sigma)$, we stipulate the following notations for (conditional) distributions:

$$\mu_{\mathcal{E}} \triangleq \mathcal{P}(\cdot | \mathcal{E}), \quad \mu_{\mathcal{E}}^\sigma \triangleq \mathcal{P}(\cdot | \mathcal{E} \wedge \sigma).$$

For some probability distribution μ and some subset of variables $\Lambda \subseteq V$, we use μ_Λ to denote the marginal distribution induced by μ on Λ . We use commas to separate multiple subscripts, for example, we use $\mu_{\mathcal{E}, \Lambda}^\sigma$ to denote the marginal distribution induced by $\mu_{\mathcal{E}}^\sigma$ on Λ .

2.1.4. Pinned formula and pinned constraints. For a subset of variables $\Lambda \subseteq V$ and a partial assignment $\sigma \in \mathcal{Q}_\Lambda$ specified on Λ , the pinned formula $\Phi = (V, \mathcal{Q}, C)$ under σ , denoted by $\Phi^\sigma = (V^\sigma, \mathcal{Q}^\sigma, C^\sigma)$, is a new CSP formula such that $V^\sigma = V \setminus \Lambda(\sigma)$, $\mathcal{Q}^\sigma = \mathcal{Q}_{V \setminus \Lambda(\sigma)}$ and the C^σ is obtained from C by:

- (1) replacing each original constraint $c \in C$ with the corresponding pinned constraint c^σ , where $\text{vbl}(c^\sigma) = \text{vbl}(c) \setminus \Lambda(\sigma)$ and $c^\sigma(\tau) = c(\tau \wedge \sigma_{\Lambda(\sigma) \cap \text{vbl}(c)})$ for any $\tau \in \mathcal{Q}_{\text{vbl}(c^\sigma)}$;
- (2) removing all the resulting constraints that have already been satisfied.

Whenever a pinning σ is applied to a CSP formula $\Phi = (V, \mathcal{Q}, C)$, we always assume that σ does not violate any constraint in C . Under such assumption, Φ^σ is always well-defined and $\mu_{\Phi^\sigma} = \mu_{V \setminus \Lambda(\sigma)}^\sigma$. Moreover, if Φ is atomic, then so is the pinned formula Φ^σ . We use C^* to denote the set of all possible constraints obtained from pinning some constraint in C with some non-violating σ , including the unpinned constraints in C . Finally, for each (possibly) pinned constraint $c \in C^*$, we use c^O to denote its original unpinned constraint in C .

2.2. Lovász Local lemma. The celebrated Lovász local lemma gives a sufficient criterion for a CSP solution to exist:

Theorem 2.2 ([EL75]). *Given a CSP formula $\Phi = (V, \mathcal{Q}, C)$, if the following holds*

$$(3) \quad \exists x \in (0, 1)^C \quad \text{s.t.} \quad \forall c \in C : \quad \mathcal{P}[-c] \leq x(c) \prod_{\substack{c' \in C \setminus \{c\} \\ \text{vbl}(c) \cap \text{vbl}(c') \neq \emptyset}} (1 - x(c')),$$

then

$$\mathcal{P}[C] \geq \prod_{c \in C} (1 - x(c)) > 0.$$

When the condition in (3) is satisfied, the probability of any event in the uniform distribution μ over all satisfying assignments can be well approximated by the probability of the event in the product distribution.

Theorem 2.3 ([HSS11, Theorem 2.1]). *Given a CSP formula $\Phi = (V, \mathcal{Q}, C)$, if (3) holds, then for any event \mathcal{A} that is determined by the assignment on a subset of variables $\text{vbl}(\mathcal{A}) \subseteq V$,*

$$\mathcal{P}[\mathcal{A} \mid C] \leq \mathcal{P}[\mathcal{A}] \prod_{\substack{c \in C \\ \text{vbl}(c) \cap \text{vbl}(\mathcal{A}) \neq \emptyset}} (1 - x(c))^{-1}.$$

2.3. Dependency graph and 2-trees. The dependency graph is a key notion for the Lovász local lemma.

Definition 2.4 (dependency graph). Let $\Phi = (V, \mathcal{Q}, C)$ be a CSP formula and let $\mathcal{E} \subseteq C$ be a subset of constraints.

- $G_\Phi(\mathcal{E})$ denotes the dependency graph induced by \mathcal{E} , which is a graph with vertex set \mathcal{E} such that there is an edge between distinct $c, c' \in \mathcal{E}$ if and only if $\text{vbl}(c) \cap \text{vbl}(c') \neq \emptyset$.
- $G_\Phi^2(\mathcal{E})$ denotes the square graph of $G_\Phi(\mathcal{E})$, in which there is an edge between distinct $c, c' \in \mathcal{E}$ if and only if $\text{dist}_{G_\Phi}(c, c') \leq 2$.

The notion of 2-tree is an important combinatorial structure introduced by Alon [Alo91], which has played key roles in algorithmic and sampling LLL.

Definition 2.5 (2-tree). Let $G = (V, E)$ be a graph and $\text{dist}_G(\cdot, \cdot)$ denote the shortest path distance in G . A 2-tree in G is a subset of vertices $T \subseteq V$ such that:

- for any $u, v \in T$, $\text{dist}_G(u, v) \geq 2$;
- T is connected if an edge is added between each $u, v \in T$ such that $\text{dist}_G(u, v) = 2$.

The following lemma bounds the number of 2-trees of a certain size in bounded-degree graphs.

Lemma 2.6 ([FGYZ21a, Corollary 5.7]). *Let $G = (V, E)$ be a graph with maximum degree d and $v \in V$ be a vertex. Then the number of 2-trees in G of size ℓ containing v is at most $\frac{(ed^2)^{\ell-1}}{2}$.*

2.4. Total variation distance and coupling. Let μ and ν be two probability distributions over the same (finite) state space Ω . Their total variation distance is defined by

$$d_{\text{TV}}(\mu, \nu) = \frac{1}{2} \sum_{x \in \Omega} |\mu(x) - \nu(x)| = \max_{A \subseteq \Omega} (\mu(A) - \nu(A)).$$

A coupling \mathcal{D} of two distributions μ and ν is a joint distribution over $\Omega \times \Omega$ whose projection on the first (or second) coordinate is μ (or ν). The well-known coupling lemma is often used to bound total variation distances, given as follows.

Lemma 2.7 ([LP17, Proposition 4.7]). *Let $\mathcal{D} : (X, Y)$ be any coupling of μ and ν , then*

$$d_{\text{TV}}(\mu, \nu) \leq \Pr_{\mathcal{D}} [X \neq Y].$$

3. CONSTRAINT-WISE COUPLING WITH EXPONENTIAL DECAY OF CORRELATION

We prove a correlation decay property for the uniform distribution μ over all satisfying assignments for the atomic CSP formula Φ satisfying Condition 1. This correlation decay property is captured by the decay of discrepancy in a coupling between μ and the distribution with one fewer constraint.

Theorem 3.1. *Let the CSP formula $\Phi = (V, Q, C)$ satisfy Condition 1. Let $c_0 \in C$ be an arbitrary constraint. There exists a coupling (X, Y) of μ_C and $\mu_{C \setminus \{c_0\}}$, such that for any integer $K \geq 1$,*

$$\Pr [d_{\text{Ham}}(X, Y) \geq k \cdot (D + 1) \cdot K] \leq 2^{-K},$$

where $d_{\text{Ham}}(X, Y) \triangleq \sum_{v \in V} I[X(v) \neq Y(v)]$ denotes the Hamming distance between X and Y .

Theorem 3.1 says that in the local lemma regime characterized by Condition 1, the total influence of any particular constraint decays exponentially. This is the first time that such an exponential decay of correlation has been found in a local lemma regime as Condition 1. This is achieved by a new constraint-wise coupling process. The construction and analysis of this novel process completely bypass the use of Beck's freezing approach, which introduced slackness and was used in almost every previous work.

The rest of this section is devoted to proving Theorem 3.1.

3.1. A constraint-wise coupling. First, we present the construction of the coupling in Theorem 3.1. Fix an instance of atomic CSP formula $\Phi = (V, Q, C)$. The coupling is defined by a recursive procedure $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$, which takes as input two sets $\mathcal{E}, \mathcal{F} \subseteq C^*$ of pinned constraints, along with two partial assignments $\sigma, \tau \in Q_\Lambda$ specified on some subset $\Lambda \subseteq V$ of variables, with the promise that both \mathcal{E}^σ and \mathcal{F}^τ are satisfiable. The procedure tries to produce a random pair of assignments $(X, Y) \in Q \times Q$ distributed according to a coupling of the joint distributions $\mu_{\mathcal{E}}^\sigma$ and $\mu_{\mathcal{F}}^\tau$.

This recursive procedure $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ is given in Algorithm 1. An arbitrary total ordering is assumed on C^* , which contains all original constraints in C and all their possible pinnings.

Algorithm 1: $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$

Instance: atomic CSP formula $\Phi = (V, Q, C)$;

Input : two subsets of pinned constraints $\mathcal{E}, \mathcal{F} \subseteq C^*$, and two partial assignments $\sigma, \tau \in Q_\Lambda$ specified on the same subset $\Lambda \subseteq V$ of variables;

Output : a pair of assignments $(X, Y) \in Q \times Q$;

```

1 if  $\mathcal{E}^\sigma = \mathcal{F}^\tau$  then
2   | let  $(X, Y)$  be drawn according to the coupling of  $\mu_{\mathcal{E}}^\sigma$  and  $\mu_{\mathcal{F}}^\tau$  that always satisfies  $X_{V \setminus \Lambda} = Y_{V \setminus \Lambda}$ ;
3   | return  $(X, Y)$ ;
4 if  $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$  then
5   | choose the smallest  $c \in \mathcal{F}^\tau \setminus \mathcal{E}^\sigma$ ;
6   | with probability  $\mu_{\mathcal{E}}^\sigma(c)$  do
7   |   | return  $\text{Couple}(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$ ;
8   | else
9   |   | let  $\pi = \text{False}(c)$  and draw a random  $\rho \sim \mu_{\mathcal{F}, \text{vbl}(c)}^\tau$ ;
10  |   | return  $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ ;
11 else
12  | choose the smallest  $c \in \mathcal{E}^\sigma \setminus \mathcal{F}^\tau$ ;
13  | with probability  $\mu_{\mathcal{F}}^\tau(c)$  do
14  |   | return  $\text{Couple}(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau)$ ;
15  | else
16  |   | draw a random  $\pi \sim \mu_{\mathcal{E}, \text{vbl}(c)}^\sigma$  and let  $\rho = \text{False}(c)$ ;
17  |   | return  $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ ;

```

Remark 3.2 (well-definedness of Algorithm 1). It is easy to see that for atomic CSP, if $\mathcal{E}^\sigma = \mathcal{F}^\tau$, then the joint distributions $\mu_{\mathcal{E}}^\sigma$ and $\mu_{\mathcal{F}}^\tau$ are identically defined over $V \setminus \Lambda$. Thus the coupling in Line 2 must exist.

Also, note that the constraints c chosen in Line 5 and Line 12 (which are used later) are pinned constraints, and are chosen according to the total ordering assumed over all pinned constraints in C^* .

Idea of the coupling. Algorithm 1 formalizes the ideas outlined in Section 1.2. The coupling procedure maintains a pair of “sub-instances with boundaries” (\mathcal{E}, σ) and (\mathcal{F}, τ) , which initially are $(C \setminus \{c_0\}, \emptyset)$ and (C, \emptyset) . The procedure tries to couple the joint distributions $\mu_{\mathcal{E}}^\sigma$ and $\mu_{\mathcal{F}}^\tau$. When the reduced sets of constraints $\mathcal{E}^\sigma = \mathcal{F}^\tau$, we have that $\mu_{\mathcal{E}}^\sigma$ and $\mu_{\mathcal{F}}^\tau$ are identically distributed over the set $V \setminus \Lambda$ of unassigned variables, and hence we can perfectly couple the unassigned variables (Lines 1-3). Otherwise $\mathcal{E}^\sigma \neq \mathcal{F}^\tau$, there must exist at least one (pinned) constraint $c \in \mathcal{E}^\sigma \Delta \mathcal{F}^\tau$, say $c \in \mathcal{F}^\tau \setminus \mathcal{E}^\sigma$ (Lines 4-10). The joint distribution $\mu_{\mathcal{E}}^\sigma$ can be decomposed based on the satisfaction and violation of c as:

$$\mu_{\mathcal{E}}^\sigma = \mu_{\mathcal{E}}^\sigma(c) \cdot \mu_{\mathcal{E}}^\sigma(\cdot \mid c) + \mu_{\mathcal{E}}^\sigma(\neg c) \cdot \mu_{\mathcal{E} \cup \{c\}}^\sigma(\cdot \mid \neg c).$$

The coupling algorithm then proceeds as follows according to the above decomposition:

- With probability $\mu_{\mathcal{E}}^\sigma(c)$, try to couple $\mu_{\mathcal{E}}^\sigma(\cdot \mid c)$ and $\mu_{\mathcal{F}}^\tau$ (Lines 6-7), which is solved recursively by the same coupling procedure on the new pair of sub-instances $(\mathcal{E} \cup \{c\}, \sigma)$ and (\mathcal{F}, τ) .
- With probability $\mu_{\mathcal{E}}^\sigma(\neg c)$, try to couple $\mu_{\mathcal{E}}^\sigma(\cdot \mid \neg c)$ and $\mu_{\mathcal{F}}^\tau$ (Lines 8-10). Because c is a (pinned) atomic constraint, $\mu_{\mathcal{E}}^\sigma(\cdot \mid \neg c)$ fixes the assignment on $\text{vbl}(c)$ to be $\pi = \text{False}(c)$. On the other hand, the joint distribution $\mu_{\mathcal{F}}^\tau$ can be decomposed based on the assignment on $\text{vbl}(c)$ as:

$$\mu_{\mathcal{F}}^\tau = \sum_{\rho \in \mathcal{Q}_{\text{vbl}(c)}} \mu_{\mathcal{F}}^\tau(\rho) \cdot \mu_{\mathcal{F}}^\tau(\cdot \mid \rho).$$

Then the coupling between $\mu_{\mathcal{E}}^\sigma(\cdot \mid \neg c)$ and $\mu_{\mathcal{F}}^\tau$ can be solved recursively on the new pair of sub-instances $(\mathcal{E}, \sigma \wedge \pi)$ and $(\mathcal{F}, \tau \wedge \rho)$, where $\pi = \text{False}(c)$ and $\rho \sim \mu_{\mathcal{F}, \text{vbl}(c)}^\tau$.

The mirrored case with $c \in \mathcal{E}^\sigma \setminus \mathcal{F}^\tau$ is processed symmetrically (Lines 11-17).

Algorithm 1 *per se* does not give an efficient procedure for computing the output (X, Y) , since it relies on calculating some nontrivial marginal probabilities. But it is useful for exhibiting the correlation decay property in the local lemma regime, which in turn is useful for the algorithmic implications.

Correctness of the coupling. Necessarily, the recursive procedure satisfies the invariant condition that the two joint distributions $\mu_{\mathcal{E}}^\sigma$ and $\mu_{\mathcal{F}}^\tau$ are coupled correctly.

Lemma 3.3 (soundness of coupling). *Assume Condition 1. For any constraint $c_0 \in C$, the procedure $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$ terminates with probability 1 and returns a coupling of μ_C and $\mu_{C \setminus \{c_0\}}$.*

Proof. We first prove that the coupling procedure is well-defined. It suffices to show by structural induction in the top-down order of recursion that for each recursive call $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$:

$$(4) \quad \Lambda(\sigma) = \Lambda(\tau), \quad \mathcal{P}[\mathcal{E} \wedge \sigma] > 0, \quad \mathcal{P}[\mathcal{F} \wedge \tau] > 0.$$

The base case is $(\mathcal{E}, \mathcal{F}, \sigma, \tau) = (C \setminus \{c_0\}, C, \emptyset, \emptyset)$. Then (4) holds by Condition 1 and Theorem 2.2.

For the induction step, assume the current call is $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$, then by the induction hypothesis, both \mathcal{E}^σ and \mathcal{F}^τ are well-defined and $\Lambda(\sigma) = \Lambda(\tau)$. We then only prove the case when $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$. The case when $\mathcal{F}^\tau \subseteq \mathcal{E}^\sigma$ follows analogously.

Let c be the smallest constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$. Note that by the induction hypothesis, $\mu_{\mathcal{E}}^\sigma(c)$ is well-defined. Also, when $\mu_{\mathcal{E}}^\sigma(c) > 0$ we have

$$\mathcal{P}[(\mathcal{E} \cup \{c\}) \wedge \sigma] = \mathcal{P}[\mathcal{E} \wedge \sigma] \cdot \mu_{\mathcal{E}}^\sigma(c) > 0,$$

proving (4) for the direct recursive call $\text{Couple}(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$ on Line 7.

For any direct recursive call $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ such that $\pi = \text{False}(c)$ and $\rho \in \mathcal{Q}_{\text{vbl}(c)}$ on Line 10, $\Lambda(\sigma \wedge \pi) = \Lambda(\tau \wedge \rho)$ trivially holds. Also, it must hold that $\mu_{\mathcal{E}}^\sigma(\neg c) > 0, \mu_{\mathcal{F}}^\tau(\rho) > 0$ by Lines 8-Line 9 of Algorithm 1, meaning π implies $\neg c$. Therefore, we have

$$\mathcal{P}[\mathcal{E} \wedge \sigma \wedge \pi] = \mathcal{P}[\mathcal{E} \wedge \pi \wedge \neg c] = \mu_{\mathcal{E}}^\sigma(\neg c) \cdot \mathcal{P}[\mathcal{E} \wedge \sigma] > 0,$$

$$\mathcal{P}[\mathcal{F} \wedge \tau \wedge \rho] = \mu_{\mathcal{F}}^\tau(\rho) \cdot \mathcal{P}[\mathcal{F} \wedge \tau] > 0,$$

finishing the proof of the induction step. This proves that the procedure is well-defined.

Knowing the procedure is well-defined, notice that Algorithm 1 terminates when $\mathcal{E}^\sigma = \mathcal{F}^\tau$, and in each recursive step, either the size of $\mathcal{E}^\sigma \Delta \mathcal{F}^\tau$ reduces by one, or the number of unassigned variables in σ and τ reduces by at least one. Due to the finiteness of the number of (pinned) constraints and variables, the procedure $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$ terminates eventually.

At last, we apply a structural induction in the bottom-up order of recursion, to prove that each possible recursive call of $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ produces a correct coupling of $\mu_{\mathcal{E}}^\sigma$ and $\mu_{\mathcal{F}}^\tau$.

The base case is when $\mathcal{E}^\sigma = \mathcal{F}^\tau$, and the above claim holds by Lines 1-3 of Algorithm 1.

For the induction step, by symmetry, we only prove for the case when $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$. Let c be the smallest constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$. From Lines 5-10 of Algorithm 1, we have the following two cases:

- With probability $\mu_{\mathcal{E}}^\sigma(c)$, $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ returns the output of $\text{Couple}(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$. By the induction hypothesis, the random pair of assignments (X_1, Y_1) returned by $\text{Couple}(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$ at Line 7 follow the marginal distributions

$$X_1 \sim \mu_{\mathcal{E} \cup \{c\}}^\sigma = \mu_{\mathcal{E}}^\sigma(\cdot \mid c) \quad \text{and} \quad Y_1 \sim \mu_{\mathcal{F}}^\tau.$$

- With probability $\mu_{\mathcal{E}}^\sigma(\neg c)$, $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ returns the output of $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ for $\pi = \text{False}(c)$ and $\rho \sim \mu_{\mathcal{F}, \text{vbl}(c)}^\tau$. By the induction hypothesis, the random pair of assignments (X_2, Y_2) returned by $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ at Line 10 follow the marginal distributions:

$$X_2 \sim \mu_{\mathcal{E}}^\sigma(\cdot \mid \pi) = \mu_{\mathcal{E}}^\sigma(\cdot \mid \neg c) \quad \text{and} \quad Y_2 \sim \sum_{\rho \in \mathcal{Q}_{\text{vbl}(c)}} \mu_{\mathcal{F}}^\tau(\rho) \cdot \mu_{\mathcal{F}}^\tau(\cdot \mid \rho) = \mu_{\mathcal{F}}^\tau.$$

Hence, the pair of assignments (X, Y) returned at Line 10 follow the marginal distributions:

$$\begin{aligned} X &\sim \mu_{\mathcal{E}}^\sigma(c) \cdot \mu_{\mathcal{E}}^\sigma(\cdot \mid c) + \mu_{\mathcal{E}}^\sigma(\neg c) \cdot \mu_{\mathcal{E}}^\sigma(\cdot \mid \neg c) &= \mu_{\mathcal{E}}^\sigma, \\ Y &\sim \mu_{\mathcal{E}}^\sigma(c) \cdot \mu_{\mathcal{F}}^\tau + \mu_{\mathcal{E}}^\sigma(\neg c) \cdot \mu_{\mathcal{F}}^\tau &= \mu_{\mathcal{F}}^\tau. \end{aligned}$$

This finishes the last case of the induction step and the proof of the lemma. \square

3.2. 2-tree witness for discrepancy. We need to bound the decay of correlation stated in Theorem 3.1. Throughout, we fix the input atomic CSP formula $\Phi = (V, Q, C)$ and an arbitrary constraint $c_0 \subseteq C$ assumed in Theorem 3.1. Our goal is to bound the discrepancy of the coupling measured in Hamming distance between the random pair (X, Y) produced by $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$. To achieve this goal, we introduce a notion of witness for the discrepancy. Inspired by the 2-tree structure [Alo91], the witness is defined as a subset of independent constraints accessed by Algorithm 1.

A technical aspect worth paying attention to is that Algorithm 1 deals with pinned constraints $c \in C^*$; however, independence between the original unpinned constraints $c^O \in C$ should be ensured to properly bound the discrepancy. This is formalized through the following definition.

Definition 3.4 (witness for discrepancy). Given a run of Algorithm 1 from $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$, the *witness set* $B = B_{c_0} \subseteq C$ is a set of (unpinned) constraints, constructed as follows:

- Initially, $B = \emptyset$.
- Whenever $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ is recursively called at Line 10 or Line 17, add the original unpinned constraint $c^O \in C$ into B if it does not share variables with any constraints already in B . Formally, let $B \leftarrow B \bowtie c$, where the *join operator* $B \bowtie c$ is defined as:

$$(5) \quad B \bowtie c \triangleq \begin{cases} B & \exists c' \in B \text{ s.t. } \text{vbl}(c^O) \cap \text{vbl}(c') \neq \emptyset; \\ B \cup \{c^O\} & \text{otherwise.} \end{cases}$$

The set B constructed in the definition above gives a witness for the discrepancy in the coupling produced by Algorithm 1, in the sense which we will see later that the hamming distance $d_{\text{Ham}}(X, Y)$ is always upper bounded by $k \cdot (D + 1) \cdot |B|$ for the output (X, Y) of $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$.

Recall the definitions of dependency graph (Definition 2.4) and 2-tree (Definition 2.5), we have the following observation.

Lemma 3.5. *During the process in Definition 3.4, the set B is always a 2-tree in the dependency graph G_Φ .*

Proof. Note that by Definition 3.4 and (5), we directly have the set B always disjoint. It remains to show that $G_{\Phi}^2(B)$ is always connected. We then prove this claim by a forward induction.

The induction basis is when the current parameter for Couple is $(\mathcal{E}, \mathcal{F}, \sigma, \tau) = (C \setminus \{c_0\}, C, \emptyset, \emptyset)$, and the claim holds by convention as we initialized B as \emptyset .

For the induction step, assume the current parameter $(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ and the current witness set B . By the induction hypothesis, $G_{\Phi}^2(B)$ is connected. We then prove the hypothesis for all possible direct recursive calls. If $\mathcal{E}^{\sigma} = \mathcal{F}^{\tau}$, no further recursion will be called and we are already done. Otherwise $\mathcal{E}^{\sigma} \neq \mathcal{F}^{\tau}$, we then only prove the case when $\mathcal{F}^{\tau} \not\subseteq \mathcal{E}^{\sigma}$. The case when $\mathcal{F}^{\tau} \subseteq \mathcal{E}^{\sigma}$ follows analogously. Let c be the smallest constraint in $\mathcal{F}^{\tau} \setminus \mathcal{E}^{\sigma}$, we have two cases:

- We recurse to Couple($\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau$) with B remains unchanged, then the claim trivially holds.
- We recurse to Couple($\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho$) for $\pi = \text{False}(c)$ and some $\rho \in \mathcal{Q}_{\text{vbl}(c)}$ and we update $B \leftarrow B \bowtie c$. If $B \bowtie c = B$, then the claim trivially holds by the induction hypothesis, otherwise assume $B \bowtie c = B \cup \{c^O\}$, we then have two additional cases:
 - c^O is exactly c_0 , in this case it must hold that $(\mathcal{E}, \mathcal{F}, \sigma, \tau) = (C \setminus \{c_0\}, C, \emptyset, \emptyset)$, since otherwise after the first step c_0 is satisfied in both \mathcal{E}^{σ} and \mathcal{F}^{τ} , and any pinned constraint of it can never have been chosen by the algorithm. In this case, $B \bowtie c$ is simply $\{c_0^O\}$ and therefore the claim holds.
 - Otherwise, it must satisfy that $\text{vbl}(c) \neq \text{vbl}(c^O)$, since this is the only way that c can be added into $\mathcal{E}^{\sigma} \Delta \mathcal{F}^{\tau}$. Take any variable $v \in \text{vbl}(c^O) \setminus \text{vbl}(c)$, and let $c' \in C^*$ be the pinned constraint containing v chosen by the procedure when v is assigned. We must have $(c')^O \notin B$ since otherwise we have $B \bowtie c^O = B$ by (5). Hence, there must exist another constraint $c^* \in B$ such that $\text{vbl}(c') \cap \text{vbl}(c^*) \neq \emptyset$ by the time we update $B \leftarrow B \bowtie c'$ according to (5). Hence c^* and c^O are adjacent in G_{Φ}^2 and $G_{\Phi}^2(B \bowtie c) = G_{\Phi}^2(B \cup \{c^O\})$ is also connected by the induction hypothesis.

□

To establish Theorem 3.1, it is intuitive to explore a truncated version of Algorithm 1, limited to $|B| < K$ for an integer $K \geq 1$, and then bound the probability of truncation. This will be formalized by the random process introduced in Definition 3.6. Here, we monitor the sequence of input arguments across all recursive calls of Couple, along with the associated witness set B , truncating it once its size surpasses K . This formal approach elucidates the structure of the recursive execution of Algorithm 1, thereby enhancing our understanding and paving the way for subsequent algorithmic insights.

Definition 3.6 (random process simulating K -truncated Algorithm 1). Let $\mathfrak{X} \sim \mu_{C \setminus \{c_0\}}$ and $\mathfrak{Y} \sim \mu_C$ be drawn independently beforehand. Define the random process $P^{\text{cp}} = P_K^{\text{cp}} = \{(\mathcal{E}_t, \mathcal{F}_t, \sigma_t, \tau_t, B_t)\}_{t \geq 0}$ starting from the initial state $(\mathcal{E}_0, \mathcal{F}_0, \sigma_0, \tau_0, B_0) = (C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$ as follows:

- (1) If $|B| = K$ or $\mathcal{E}^{\sigma} = \mathcal{F}^{\tau}$, the process stops and $(\mathcal{E}_t, \mathcal{F}_t, \sigma_t, \tau_t, B_t)$ is the outcome of the process.
- (2) Otherwise, suppose $\mathcal{F}_t^{\tau} \not\subseteq \mathcal{E}_t^{\sigma}$. Let c be the smallest pinned constraint in $\mathcal{F}_t^{\tau} \setminus \mathcal{E}_t^{\sigma}$.

$$(\mathcal{E}_{t+1}, \mathcal{F}_{t+1}, \sigma_{t+1}, \tau_{t+1}, B_{t+1}) \leftarrow \begin{cases} (\mathcal{E}_t \cup \{c\}, \mathcal{F}_t, \sigma_t, \tau_t, B_t) & c \text{ is satisfied by } \mathfrak{X}_{\text{vbl}(c)}; \\ (\mathcal{E}_t, \mathcal{F}_t, \sigma_t \wedge \mathfrak{X}_{\text{vbl}(c)}, \tau_t \wedge \mathfrak{Y}_{\text{vbl}(c)}, B_t \bowtie c) & \text{otherwise.} \end{cases}$$

- (3) Otherwise $\mathcal{F}_t^{\tau} \subseteq \mathcal{E}_t^{\sigma}$. Let c be the smallest pinned constraint in $\mathcal{E}_t^{\sigma} \setminus \mathcal{F}_t^{\tau}$.

$$(\mathcal{E}_{t+1}, \mathcal{F}_{t+1}, \sigma_{t+1}, \tau_{t+1}, B_{t+1}) \leftarrow \begin{cases} (\mathcal{E}_t, \mathcal{F}_t \cup \{c\}, \sigma_t, \tau_t, B_t) & c \text{ is satisfied by } \mathfrak{Y}_{\text{vbl}(c)}; \\ (\mathcal{E}_t, \mathcal{F}_t, \sigma_t \wedge \mathfrak{X}_{\text{vbl}(c)}, \tau_t \wedge \mathfrak{Y}_{\text{vbl}(c)}, B_t \bowtie c) & \text{otherwise.} \end{cases}$$

Let $\mu^{\text{cp}} = \mu_K^{\text{cp}}$ denote the distribution of the outcome $(\mathcal{E}_{\infty}, \mathcal{F}_{\infty}, \sigma_{\infty}, \tau_{\infty}, B_{\infty})$ of this process, and let $\mathcal{L}^{\text{cp}} = \text{supp}(\mu^{\text{cp}})$ be its support. Let $\mathcal{L}_K^{\text{cp}}$ be the set of “truncated” outcomes, i.e.:

$$\mathcal{L}_K^{\text{cp}} = \{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \mid \mu^{\text{cp}}((\mathcal{E}, \mathcal{F}, \sigma, \tau, B)) > 0 \wedge |B| = K\}.$$

At last, let \mathcal{V}^{cp} denote the set of all possible $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ with $\Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}}] > 0$.

Remark 3.7. Note that for atomic CSP Φ satisfying Condition 1, due to Theorem 2.2 and total probability, the random process P^{cp} and the distribution μ^{cp} in Definition 3.6 are well-defined.

Remark 3.8 (explicitly identified randomness). In Definition 3.6, all randomness used by the process P^{cp} is identified with the two pre-generated random assignments $\mathfrak{X} \sim \mu_{C \setminus \{c_0\}}$ and $\mathfrak{Y} \sim \mu_C$. However, in Algorithm 1, random choices (in Lines 6, 9, 13 and 16) are generated at the moment they are used. According to the principle of deferred decision, this distinction does not affect the definition of P^{cp} , and the process faithfully simulates Algorithm 1. Nevertheless, such explicit identification of randomness is crucial for our analyses of the coupling and its algorithmic implications.

A key observation is that the correlation decay of the coupling is bounded by the probability of truncated outcomes of the process constructed in Definition 3.6.

Lemma 3.9. *Assume Condition 1. For the output (X, Y) of $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$,*

$$\Pr [d_{\text{Ham}}(X, Y) \geq k \cdot (D + 1) \cdot K] \leq \mu^{\text{cp}} [\mathcal{L}_K^{\text{cp}}].$$

Proof. Note that by Definition 3.4, each time we recursively call $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ at Line 10 or Line 17 of Algorithm 1, we update $B \leftarrow B \bowtie c$, at most k variables is additionally assigned to σ/τ and any pinned constraint of c^O will never been chosen by the procedure in future steps. Note that each constraint in C shares constraint with at most D other constraints, hence by (5), after $D + 1$ such operations the size of B increases by at least one. Also, by Line 3 of Algorithm 1 we have $d_{\text{Ham}}(X, Y)$ is upper bounded by the total number of variables assigned in σ/τ during $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$. Therefore, we have

$$d_{\text{Ham}}(X, Y) \geq k \cdot (D + 1) \cdot K \implies |B| \geq K.$$

We claim that the process in Definition 3.6 can be coupled with Algorithm 1 initialized by $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$ so that each time we recursively call $\text{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ with witness set B , we will also move to the same tuple $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in Definition 3.6. Under the coupling assumed by the claim, it follows that

$$|B| \geq K \implies (\mathcal{E}^{\text{cp}}, \mathcal{F}^{\text{cp}}, \sigma^{\text{cp}}, \tau^{\text{cp}}, B^{\text{cp}}) \in \mathcal{L}_K^{\text{cp}},$$

and the lemma immediately follows.

We strengthen the claim that conditioning on we are currently at some tuple $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in Definition 3.6, it follows that

$$(6) \quad \mathfrak{X} \sim \mu_{\mathcal{E}}^{\sigma}, \quad \mathfrak{Y} \sim \mu_{\mathcal{F}}^{\tau}.$$

and prove the strengthened claim using a structural induction in the top-down order. The base case is for the initial call $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$ with witness set $B = \emptyset$. Note that in Definition 3.6 we also initially have $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) = (C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$, and

$$\mathfrak{X} \sim \mu_{C \setminus \{c_0\}} = \mu_{\mathcal{E}}^{\sigma}, \quad \mathfrak{Y} \sim \mu_C = \mu_{\mathcal{F}}^{\tau}.$$

For the induction step, assume we are currently executing $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$ with witness set B . By the induction hypothesis, we are also at the same tuple $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in Definition 3.6, so that (6) holds.

Now it remains to compare Algorithm 1 with Definition 3.6 and notice that under (6), the transition probabilities to the next tuples are the same in both. Therefore, the next tuple in both can still be perfectly coupled. Also, it can be verified that (6) also holds for each possible branch in Definition 3.6. This finishes the proof of the claim and the lemma. \square

The following tail bound for the measure of truncated outcomes is crucial for proving Theorem 3.1.

Lemma 3.10. *Assume Condition 1. For any integer $K \geq 1$,*

$$\mu^{\text{cp}} [\mathcal{L}_K^{\text{cp}}] \leq 2^{-K}.$$

Theorem 3.1 then follows directly from combining Lemmas 3.3, 3.9 and 3.10. All we need to do now is prove Lemma 3.10.

3.3. Refutation of large witnesses. By the above arguments, it suffices to bound the probability $\mu^{\text{cp}}[\mathcal{L}_K^{\text{cp}}]$. Consider the random outcome generated by the process in Definition 3.6:

$$(\mathcal{E}^{\text{cp}}, \mathcal{F}^{\text{cp}}, \sigma^{\text{cp}}, \tau^{\text{cp}}, B^{\text{cp}}) \sim \mu^{\text{cp}}.$$

It then suffices to show that the witness set B^{cp} is very unlikely to become too large. We first establish the following tail bound for any particular large enough subset of B^{cp} .

Lemma 3.11. *Assume Condition 1. For any subset of disjoint constraints $T \subseteq C$,*

$$\Pr[T \subseteq B^{\text{cp}}] \leq p^{\frac{2|T|}{2+\zeta}} \cdot (1 - ep)^{2(D+1)|T|}.$$

Proof. We claim that for each $c \in T$, the event $c \in B^{\text{cp}}$ implies the following event

$$(7) \quad \mathcal{A}_c : \forall v \in \text{vbl}(c), \quad \mathfrak{X}(v) = \text{False}(c)(v) \vee \mathfrak{Y}(v) = \text{False}(c)(v).$$

To prove the claim, by contradiction we suppose that there exists some $v \in \text{vbl}(c)$ such that both $\mathfrak{X}(v) \neq \text{False}(c)(v)$ and $\mathfrak{Y}(v) \neq \text{False}(c)(v)$. Let c' be the pinned constraint chosen by the procedure when c is added into B^{cp} at some step in Definition 3.6, then we have $c = (c')^O$. It must hold that $v \in \Lambda(c')$, as otherwise c' is both satisfied in \mathcal{E} and \mathcal{F} conditioning on the assignment of \mathfrak{X} and \mathfrak{Y} on $\text{vbl}(c) \setminus \Lambda(c')$, and could not have been chosen by the procedure. However, $v \in \Lambda(c')$ is also not possible as otherwise by the time c' is chosen in Definition 3.6, c' will be satisfied by both $\mathfrak{X}_{\text{vbl}(c')}$ and $\mathfrak{Y}_{\text{vbl}(c')}$ and will not be added into B^{cp} , a contradiction. The above claim is proved.

Thus we have

$$(8) \quad \begin{aligned} \Pr[T \subseteq B^{\text{cp}}] &\leq \Pr\left[\bigwedge_{c \in T} \mathcal{A}_c\right] \\ &\stackrel{\text{(by disjointness between } c \in T)}{\leq} \prod_{c \in T} \Pr[\mathcal{A}_c] \\ &\leq \prod_{c \in T} \left(\prod_{v \in \text{vbl}(c)} \left(\frac{2}{|Q_v|} - \frac{1}{|Q_v|^2} \right) \cdot (1 - ep)^{-2(D+1)} \right). \end{aligned}$$

Here, the last inequality is by interpreting the probability space for generating $\mathfrak{X} \sim \mu_C$ and $\mathfrak{Y} \sim \mu_{C \setminus \{c_0\}}$ as the product space over two copies of the distribution \mathcal{P} , conditioning on that all constraints in C are satisfied in the first copy, and all constraints in $C \setminus \{c_0\}$ are satisfied in the second copy. Note that this can be viewed as an LLL distribution with dependency degree at most D and violation probability of each bad event at most p . Also, each event \mathcal{A}_c is mutually dependent with all but $2(D+1)$ bad events. Therefore, setting $x(c) = ep$ for each bad event c and applying Theorem 2.3 gives to the above last inequality.

Note that since C is a set of atomic constraints, we have for each $c \in C$,

$$\prod_{v \in \text{vbl}(c)} \frac{1}{|Q_v|} \leq p,$$

and for any $v \in \text{vbl}(c)$,

$$\frac{\ln(2/|Q_v| - 1/|Q_v|^2)}{\ln(1/|Q_v|)} = 2 - \frac{\ln(2/|Q_v| - 1)}{\ln(1/|Q_v|)} \geq 2 - \frac{\ln(2/q_{\min} - 1)}{\ln q_{\min}} = \frac{2}{2 + \zeta},$$

where in above we use that $\frac{\ln(2x-1)}{\ln x}$ is a monotonically decreasing function for all $x > 1$.

Hence, combining with (8) we have

$$\Pr[T \subseteq B^{\text{cp}}] \leq \prod_{c \in T} \left(p^{\frac{2}{2+\zeta}} \cdot (1 - ep)^{-2(D+1)} \right) = p^{\frac{2|T|}{2+\zeta}} \cdot (1 - ep)^{-2(D+1)|T|}. \quad \square$$

With the help of Lemma 3.11, we are ready to prove Lemma 3.10, completing the proof of Theorem 3.1.

Proof of Lemma 3.10. Let $\mathbb{T}_K^{c_0}$ be the set of 2-trees in G_Φ of size K containing c_0 . Since $K \geq 1$ we have all $(\mathcal{E}^{\text{cp}}, \mathcal{F}^{\text{cp}}, \sigma^{\text{cp}}, \tau^{\text{cp}}, B^{\text{cp}}) \in \mathcal{V}_K^{\text{cp}}$ satisfy that $c_0 \in B^{\text{cp}}$ by Condition 1 and Definition 3.6. Then we have

$$\begin{aligned}
& \mu^{\text{cp}}[\mathcal{L}_K^{\text{cp}}] \\
\text{(by Definition 3.4)} & \leq \Pr[|B^{\text{cp}}| \geq K] \\
\text{(by Lemma 3.5)} & \leq \sum_{T \in \mathbb{T}_K^{c_0}} \Pr[T \subseteq B^{\text{cp}}] \\
\text{(by Lemma 3.11)} & \leq \sum_{T \in \mathbb{T}_K^{c_0}} p^{\frac{2K}{2+\zeta}} \cdot (1-ep)^{-2(D+1)K} \\
\text{(by Lemma 2.6)} & \leq \frac{(eD^2)^{K-1}}{2} \cdot p^{\frac{2L}{2+\zeta}} \cdot (1-ep)^{-2(D+1)K} \\
& \leq \left(eD^2 \cdot p^{\frac{2}{2+\zeta}} \cdot (1-ep)^{-2(D+1)} \right)^K \\
\text{(by Condition 1)} & \leq 2^{-K}.
\end{aligned}$$

□

4. LINEAR PROGRAMMING FROM CONSTRAINT-WISE COUPLING

In this section, we will present a linear program that is the central component for the counting and sampling algorithms within the local lemma regime, as stated in Theorem 1.1 and Theorem 1.2.

As in previous works for LP-based counting LLL [Moi19, GLLZ19, JPV21b], we will set up a linear program to mimic the transition probabilities in the coupling procedure (Definition 3.6) with the goal of bootstrapping the constraint-wise marginal ratios. However, our construction of the coupling differs vastly from previous works. Additionally, our analysis of the coupling is near-critical in reaching the optimal threshold, making the design of the linear program challenging and requiring significant novelty. Nevertheless, we have managed to distill critical structures from the analysis in Section 3 that are sufficient for constructing the LP.

4.1. Marginal probabilities from coupling procedure. Fix an atomic CSP formula $\Phi = (V, Q, C)$, an arbitrary constraint $c_0 \in C$, and an integer $K \geq 1$. Recall the random process $P^{\text{cp}} = P_K^{\text{cp}} = \{(\mathcal{E}_t, \mathcal{F}_t, \sigma_t, \tau_t, B_t)\}_{t \geq 0}$ and the distribution $\mu^{\text{cp}} = \mu_K^{\text{cp}}$ of its final outcome, both introduced in Definition 3.6. The following defines a family of probabilities arising from a natural one-sided sampling process induced from P^{cp} . These probabilities correspond to the variables of the linear program that will be introduced later.

Definition 4.1 (one-sided sampler and marginal probabilities). Let X (resp. Y) be drawn as:

- draw $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \sim \mu^{\text{cp}}$;
- draw $X \sim \mu_{\mathcal{E}}^{\sigma}$ (and resp. $Y \sim \mu_{\mathcal{F}}^{\tau}$).

For any evaluation of $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$, any $\mathbf{x} \in Q^{\mathcal{E} \wedge \sigma} \triangleq \{\pi \in Q \mid \pi \text{ satisfies } \mathcal{E} \wedge \sigma\}$ and $\mathbf{y} \in Q^{\mathcal{F} \wedge \tau}$, define:

$$\begin{aligned}
p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X & \triangleq \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid X = \mathbf{x}], \\
p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y & \triangleq \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid Y = \mathbf{y}].
\end{aligned}$$

Note that in above definition, $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X = \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid X = \mathbf{x}]$ for any $\mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}$ (and similarly $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y = \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid Y = \mathbf{y}]$ for any $\mathbf{y} \in Q^{\mathcal{F} \wedge \tau}$). This is actually well-defined. And furthermore, these probabilities are in fact marginal probabilities of some partial assignments in some joint distributions. These are formally justified by the following proposition.

Proposition 4.2. *Assume Condition 1. Fix any evaluation of $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$. It holds that*

$$\begin{aligned}
\forall \mathbf{x}, \mathbf{x}' \in Q^{\mathcal{E} \wedge \sigma} : & \quad \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid X = \mathbf{x}] = \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid X = \mathbf{x}'], \\
\forall \mathbf{y}, \mathbf{y}' \in Q^{\mathcal{F} \wedge \tau} : & \quad \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid Y = \mathbf{y}] = \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid Y = \mathbf{y}'].
\end{aligned}$$

Moreover, it holds that

$$(9) \quad p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X = \mu_C(\mathcal{F} \wedge \tau) \quad \text{and} \quad p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma).$$

Proof. Following the same routine as the proof of Lemma 3.9 (more precisely, following (6)), we can couple the one-sided sampler in Definition 4.1 with the process P^{cp} so that the sample X (resp. Y) generated by the one-sided sampler is identified with the random source $\mathfrak{X} \sim \mu_{C \setminus \{c_0\}}$ (resp. $\mathfrak{Y} \sim \mu_C$) used in the construction of P^{cp} in Definition 3.6.

We then only prove the first equality in (9), and the other follows analogously. Fix some $\mathbf{x} \in \mathcal{Q}$, then under the coupling between X and \mathfrak{X} it immediately follows that

$$\Pr[X = \mathbf{x}] = \mu_{C \setminus \{c_0\}}(\mathbf{x}).$$

We claim that for any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\text{cp}}$,

$$(10) \quad \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \wedge X = \mathbf{x}] = \begin{cases} 0 & \mathbf{x} \notin \mathcal{Q}^{\mathcal{E} \wedge \sigma}; \\ \mu_{C \setminus \{c_0\}}(\mathbf{x}) \cdot \mu_C(\mathcal{F} \wedge \tau) & \mathbf{x} \in \mathcal{Q}^{\mathcal{E} \wedge \sigma}, \end{cases}$$

and the lemma follows from the law of conditional probability.

We then prove (10). Note that the first case is immediate by Definition 4.1.

Following the proof of Lemma 3.9, we have that for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\text{cp}}$,

$$(11) \quad \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}}] = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \mu_C(\mathcal{F} \wedge \tau).$$

Therefore, for any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\text{cp}}$ and any $\mathbf{x} \in \mathcal{Q}^{\mathcal{E} \wedge \sigma}$,

$$\begin{aligned} & \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \wedge X = \mathbf{x}] \\ &= \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}}] \cdot \Pr[X = \mathbf{x} \mid (\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}}] \\ \text{(by Definition 4.1)} &= \Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}}] \cdot \mu_{\mathcal{E}}^{\sigma}(\mathbf{x}) \\ \text{(by (11))} &= \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \mu_C(\mathcal{F} \wedge \tau) \cdot \mu_{\mathcal{E}}^{\sigma}(\mathbf{x}) \\ \text{(by the chain rule)} &= \mu_{C \setminus \{c_0\}}(\mathbf{x}) \cdot \mu_C(\mathcal{F} \wedge \tau), \end{aligned}$$

finishing the proof of (10). Here, the last equality is additionally by that $\mathcal{E} \wedge \sigma$ implies $C \setminus \{c_0\}$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\text{cp}}$, an argument easily proven through an induction on Definition 3.6. \square

The following lists some basic properties of $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X$ and $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y$.

Proposition 4.3. *Assume Condition 1. The followings hold for the $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X$ and $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y$:*

- (1) $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X, p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \in [0, 1]$. In particular, $p_{(C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)}^X = p_{(C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)}^Y = 1$.
- (2) For any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\text{cp}} \setminus \mathcal{L}^{\text{cp}}$, where \mathcal{V}^{cp} and \mathcal{L}^{cp} are defined in Definition 3.6,
 - (a) if $\mathcal{F}^{\tau} \not\subseteq \mathcal{E}^{\sigma}$, letting $c \in \mathcal{F}^{\tau} \setminus \mathcal{E}^{\sigma}$ be the smallest and $\pi = \text{False}(c)$,

$$\begin{aligned} p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &= p_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}^X = \sum_{\substack{\rho \in \mathcal{Q}_{\text{vbl}(c)} \\ (\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\text{cp}}}} p_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X; \\ p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &= p_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}^Y + p_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^Y, \\ &\text{for all } \rho \in \mathcal{Q}_{\text{vbl}(c)} \text{ and all } (\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\text{cp}}. \end{aligned}$$

- (b) otherwise $\mathcal{F}^{\tau} \subseteq \mathcal{E}^{\sigma}$, letting $c \in \mathcal{E}^{\sigma} \setminus \mathcal{F}^{\tau}$ be the smallest and $\rho = \text{False}(c)$,

$$\begin{aligned} p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &= p_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}^X + p_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X, \\ &\text{for all } \pi \in \mathcal{Q}_{\text{vbl}(c)} \text{ and all } (\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\text{cp}}; \\ p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &= p_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}^Y = \sum_{\substack{\pi \in \mathcal{Q}_{\text{vbl}(c)} \\ (\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\text{cp}}}} p_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^Y. \end{aligned}$$

(3) For any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\text{cp}}$,

$$p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \cdot \frac{|Q^{\mathcal{E} \wedge \sigma}|}{|Q^{C \setminus \{c_0\}}|} = p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \cdot \frac{|Q^{\mathcal{F} \wedge \tau}|}{|Q^C|}.$$

These properties follow directly from Definition 4.1 or from Proposition 4.2, and are easy to verify.

The following bounds for $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X$ and $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y$ with 2-trees follow from a similar argument as in our analysis of the correlation decay property (e.g. the proof of Lemma 3.11).

Proposition 4.4. *Assume Condition 1. Let $\mathbb{T}_K^{c_0}$ be the set of 2-trees in G_Φ of size K containing c_0 . Recall that \mathcal{P} is the product distribution. For each $c \in C$, let \mathcal{A}_c be the event defined as (7), that is,*

$$\mathcal{A}_c : \forall v \in \text{vbl}(c), \quad \mathfrak{X}(v) = \text{False}(c)(v) \vee \mathfrak{Y}(v) = \text{False}(c)(v).$$

Then for any $T \in \mathbb{T}_K^{c_0}$ and $\mathbf{x} \in Q^{C \setminus \{c_0\}}$,

$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\text{cp}}: \\ B=T \wedge \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \leq (1 - \epsilon p)^{(D+1)K} \Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right],$$

and for any $T \in \mathbb{T}_K^{c_0}$ and $\mathbf{y} \in Q^C$,

$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\text{cp}}: \\ B=T \wedge \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \leq (1 - \epsilon p)^{(D+1)K} \Pr_{\mathfrak{X} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{Y} = \mathbf{y} \right].$$

Proof. We only prove the first inequality, and the second one follows analogously.

Let $(\mathcal{E}^{\text{cp}}, \mathcal{F}^{\text{cp}}, \sigma^{\text{cp}}, \tau^{\text{cp}}, B^{\text{cp}}) \sim \mu^{\text{cp}}$ denote the random outcome of the process P^{cp} constructed in Definition 3.6. Following the same routine as the proof of Lemma 3.9 (more precisely, following (6)), we can couple the one-sided sampler in Definition 4.1 with the process P^{cp} so that the sample X (resp. Y) generated by the one-sided sampler is identified with the random source $\mathfrak{X} \sim \mu_{C \setminus \{c_0\}}$ (resp. $\mathfrak{Y} \sim \mu_C$) used in the construction of P^{cp} in Definition 3.6. Fix any $\mathbf{x} \in Q^{C \setminus \{c_0\}}$. Then we have

$$\begin{aligned} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\text{cp}}: \\ B=T \wedge \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &= \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\text{cp}}: \\ B=T \wedge \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{cp}} \mid X = \mathbf{x}] \\ &= \Pr [B^{\text{cp}} = T \mid X = \mathbf{x}] \\ \text{(by the coupling between } X \text{ and } \mathfrak{X}) &= \Pr_{\substack{\mathfrak{X} \sim \mu_{C \setminus \{c_0\}} \\ \mathfrak{Y} \sim \mu_C}} [B^{\text{cp}} = T \mid \mathfrak{X} = \mathbf{x}] \\ &\leq \Pr_{\substack{\mathfrak{X} \sim \mu_{C \setminus \{c_0\}} \\ \mathfrak{Y} \sim \mu_C}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right] \\ \text{(by Theorem 2.3)} &\leq (1 - \epsilon p)^{(D+1)K} \Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right]. \end{aligned}$$

Here, the first inequality follows the same argument in the proof of Lemma 3.11 that $B^{\text{cp}} = T$ implies the event $\bigwedge_{c \in T} \mathcal{A}_c$. \square

Proposition 4.4 utilizes the independence between the random assignments $\mathfrak{X}, \mathfrak{Y}$, which are generated as the random source used by the process P^{cp} constructed in Definition 3.6 for simulating Algorithm 1. Such explicit identification of randomness helped in proving Lemma 3.11 and also will play a key role in constructing the linear program that will be introduced later.

4.2. Setting up the linear program. Next, we set up the linear program that mimics the marginal probabilities $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X$ and $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y$ introduced in Definition 4.1, in order to bootstrap the ratio between the numbers of assignments satisfying C and those satisfying $C \setminus \{c_0\}$.

4.2.1. *The coupling tree.* We define the following notion of the recursion tree for the coupling procedure $\text{Couple}(C \setminus \{c_0\}, C, \emptyset, \emptyset)$, truncated whenever the witness set B has size $|B| \geq K$.

Definition 4.5 (*K-truncated coupling tree*). The *K-truncated coupling tree* $\mathcal{T} = \mathcal{T}_K(\Phi, c_0)$ is a finite rooted tree, with each tree node corresponding to a $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$, which is the tuple appeared in Definition 3.6. The tree \mathcal{T} is inductively constructed as follows:

- (1) The root of \mathcal{T} corresponds to $(C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$, and has depth 0.
- (2) For $i = 0, 1, \dots$: for all existing tree nodes $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$ of depth i in the current \mathcal{T} ,
 - (a) if σ violates \mathcal{E} or τ violates \mathcal{F} or $\mathcal{E}^\sigma = \mathcal{F}^\tau$ or $|B| = K$, then do nothing and hence leave $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ as a leaf node in \mathcal{T} ;
 - (b) else if $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$, then pick the smallest $c \in \mathcal{F}^\tau \setminus \mathcal{E}^\sigma$, add $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$ as a child of $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in \mathcal{T} , and for $\pi = \text{False}(c)$ and each $\rho \in Q_{\text{vbl}(c)}$, add $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)$ as a child of $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in \mathcal{T} ;
 - (c) else, pick the smallest $c \in \mathcal{E}^\sigma \setminus \mathcal{F}^\tau$, add $(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)$ as a child of $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in \mathcal{T} , and for each $\pi \in Q_{\text{vbl}(c)}$ and $\rho = \text{False}(c)$, add $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)$ as a child of $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in \mathcal{T} .

Let \mathcal{L} be the set of leaf nodes in \mathcal{T} . We further define:

- $\mathcal{L}_{\text{coup}} \triangleq \{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L} \mid \mathcal{E}^\sigma = \mathcal{F}^\tau\}$ as the set of “coupled” leaf nodes in \mathcal{T} ;
- $\mathcal{L}_{\text{trun}} \triangleq \{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L} \mid |B| = K\}$ as the set of “truncated” leaf nodes in \mathcal{T} ;
- $\mathcal{L}_{\text{valid}} \triangleq \mathcal{L}_{\text{trun}} \cup \mathcal{L}_{\text{coup}}$ as the set of “valid” leaf nodes in \mathcal{T} ;
- $\mathcal{L}_{\text{invld}} \triangleq \mathcal{L} \setminus \mathcal{L}_{\text{valid}}$ as the set of “invalid” leaf nodes in \mathcal{T} .

Remark 4.6 (well-definedness of coupling tree). For each internal node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in the coupling tree \mathcal{T} , the event $\mathcal{E} \wedge \sigma \wedge \mathcal{F} \wedge \tau$ is partitioned in some way into mutually disjoint events, e.g.:

$$(\mathcal{E} \cup \{c\}) \wedge \sigma \wedge \mathcal{F} \wedge \tau, \quad \mathcal{E} \wedge (\sigma \wedge \text{False}(c)) \wedge \mathcal{F} \wedge (\tau \wedge \rho), \quad \forall \rho \in Q_{\text{vbl}(c)},$$

each of which corresponds to a child. Therefore, all nodes in \mathcal{T} are distinct and \mathcal{T} is well-defined.

Note that every leaf node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}$ satisfies exactly one of the followings:

- $\mathcal{E}^\sigma = \mathcal{F}^\tau$;
- $|B| = K$;
- σ violates \mathcal{E} or τ violates \mathcal{F} .

Therefore, the classification of leaf nodes in Definition 4.5 is also well defined.

Remark 4.7 (difference between the coupling tree and the original coupling process). Note that the coupling tree \mathcal{T} may include more nodes than the tuples $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ that can possibly be generated by the process in Definition 3.6. This is because some branch in the coupling tree \mathcal{T} may have 0 probability in the random process. Some properties are unaffected by such distinction and still hold for the coupling tree, for instance, Lemma 3.5, while others need to be reproved in this stronger setting.

4.2.2. *The linear program.* We now introduce the linear program. The LP is constructed on an instance of *K-truncated coupling tree* introduced in Definition 4.5, where each tree node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ is associated with a pair of variables, corresponding to $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X$ and $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y$. The constraints of the LP are derived from the properties listed in Propositions 4.3 and 4.4.

Definition 4.8 (linear program induced by the coupling). Let $\mathcal{T} = \mathcal{T}_K(\Phi, c_0)$ be the *K-truncated coupling tree*. Let $0 \leq r_- \leq r_+$ be two parameters. The following linear program (for checking feasibility) is defined on the variables $\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X$ and $\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y$ for all $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$:

I. *Range constraints:*

$$\begin{aligned} \hat{p}_{(C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)}^X &= \hat{p}_{(C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)}^Y = 1; \\ \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X, \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &\in [0, 1], \quad \forall (\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T}). \end{aligned}$$

II. *Non-leaf constraints:* For each non-leaf node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T}) \setminus \mathcal{L}$,

(a) if $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$, letting c be the smallest constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$ and $\pi = \text{False}(c)$,

$$\begin{aligned}\hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &= \hat{P}_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}^X = \sum_{\rho \in \mathcal{Q}_{\text{vbl}(c)}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X; \\ \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &= \hat{P}_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}^Y + \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^Y, \quad \forall \rho \in \mathcal{Q}_{\text{vbl}(c)}.\end{aligned}$$

(b) otherwise $\mathcal{F}^\tau \subseteq \mathcal{E}^\sigma$, letting c be the smallest constraint in $\mathcal{E}^\sigma \setminus \mathcal{F}^\tau$ and $\rho = \text{False}(c)$,

$$\begin{aligned}\hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &= \hat{P}_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}^X + \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X, \quad \forall \pi \in \mathcal{Q}_{\text{vbl}(c)}; \\ \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &= \hat{P}_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}^Y = \sum_{\pi \in \mathcal{Q}_{\text{vbl}(c)}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^Y.\end{aligned}$$

III. *Leaf constraints*: For each leaf node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}$,

(a) if it is a coupled leaf $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}$,

$$r_- \cdot \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \leq \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \leq r_+ \cdot \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y.$$

(b) if it is an invalid leaf $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{invld}}$,

$$\begin{aligned}\hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &= 0, \text{ if } \sigma \text{ violates } \mathcal{E}; \\ \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &= 0, \text{ if } \tau \text{ violates } \mathcal{F}.\end{aligned}$$

IV. *Overflow constraints (via 2-tree)*: Let $\mathbb{T}_K^{c_0}$ be the set of 2-trees in G_Φ of size K containing c_0 .

$$\begin{aligned}\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}: \\ B = T \wedge \mathbf{x} \in \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &\leq (1 - \epsilon p)^{(D+1)K} \Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right]; \quad \forall T \in \mathbb{T}_K^{c_0}, \mathbf{x} \in \mathcal{Q}; \\ \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}: \\ B = T \wedge \mathbf{y} \in \mathcal{Q}^{\mathcal{F} \setminus C \wedge \tau}}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &\leq (1 - \epsilon p)^{(D+1)K} \Pr_{\mathfrak{X} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{Y} = \mathbf{y} \right], \quad \forall T \in \mathbb{T}_K^{c_0}, \mathbf{y} \in \mathcal{Q},\end{aligned}$$

where recall that \mathcal{P} is the product distribution and \mathcal{A}_c denotes the event defined in (7), i.e.

$$\mathcal{A}_c : \forall v \in \text{vbl}(c), \quad \mathfrak{X}(v) = \text{False}(c)(v) \vee \mathfrak{Y}(v) = \text{False}(c)(v).$$

Due to the independence between the constraints within a 2-tree, the probabilities involving \mathcal{A}_c in the overflow constraints (Item IV.) of Definition 4.8, can be calculated explicitly as:

$$\begin{aligned}\Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right] &= \prod_{c \in T} \prod_{v \in \text{vbl}(c)} |Q_v|^{-I[x_v \neq \text{False}(c)_v]}, \\ \Pr_{\mathfrak{X} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{Y} = \mathbf{y} \right] &= \prod_{c \in T} \prod_{v \in \text{vbl}(c)} |Q_v|^{-I[y_v \neq \text{False}(c)_v]}.\end{aligned}$$

The overflow constraints (Item IV.) of Definition 4.8 are derived from Proposition 4.4. These constraints encapsulate a critical ‘‘independence property’’ pivotal to our analysis of the correlation decay. They stand as the primary novelty in our linear program design, distinguishing it from previous LP-based approaches [Moi19, GGGY20, JPV21b]. Note, however, that the overflow constraints slightly differ from Proposition 4.4, as the domain of \mathbf{x}/\mathbf{y} have been extended and the conditions under the summation have changed. These differences serve technical purposes for the linear program to be efficient.

While it is natural to question the number of overflow constraints described in Definition 4.8, we will address this concern later (in Proposition 4.9). Specifically, we will show that there are indeed $\exp(K \cdot \text{poly}(k, \log q_{\max}, D))$ distinct overflow constraints, and moreover, it takes approximately that amount of time to both write down the linear program and verify its feasibility.

4.3. Analysis of the linear program. We now present several properties for the above linear program. First, we show that the feasibility of the linear program can be checked efficiently.

Proposition 4.9. *For any $0 \leq r_- \leq r_+$, the feasibility of the linear program in Definition 4.8 can be checked within $\exp(K \cdot \text{poly}(k, \log q_{\max}, D))$ time.*

Proposition 4.9 holds because there are at most $\exp(K \cdot \text{poly}(k, \log q_{\max}, D))$ constraints that need to be checked. Its formal proof is deferred to Appendix A.

Next, we show the soundness of this linear program: The true values of the marginal probabilities satisfy all the linear constraints.

Lemma 4.10. *Assume Condition 1. The LP in Definition 4.8 is feasible for $r_- = r_+ = \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|}$.*

Proof. For each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, we let

$$(12) \quad \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X = \mu_C(\mathcal{F} \wedge \tau), \quad \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma),$$

and show that this is a feasible solution for $r_- = r_+ = \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|}$. We remark that following Lemma 3.3, this choice of the solution in (12) coincides with the real values of $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X, p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y$ defined in Definition 4.1, but extends to all nodes in $V(\mathcal{T})$ instead of just \mathcal{V}^{cp} .

Item I. directly follows from Item 1 of Proposition 4.3. Also, Item II. directly follows from Item 2 of Proposition 4.3.

Item III.b directly follows from (12). To prove Item III.a, note that for $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}$, $\mathcal{E}^\sigma = \mathcal{F}^\tau$ and hence $|Q^{\mathcal{E} \wedge \sigma}| = |Q^{\mathcal{F} \wedge \tau}|$. Therefore,

- if $|Q^{\mathcal{E} \wedge \sigma}| = |Q^{\mathcal{F} \wedge \tau}| > 0$, then Item III.a follows directly from Item 3 of Proposition 4.3;
- otherwise $|Q^{\mathcal{E} \wedge \sigma}| = |Q^{\mathcal{F} \wedge \tau}| = 0$, then $\mu_C(\mathcal{F} \wedge \tau) = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) = 0$ and Item III.a still holds.

The proof of Item IV. falls into a similar vibe as the proof for Proposition 4.4. However, since now the domain for x and y has changed, we need to define another procedure similar to Definition 3.6 for explicitly identifying randomness to interpret the left-hand side in the inequality of Item IV.. The formal proof of Item IV. is deferred to Appendix B. \square

At last, we show that the feasibility of the linear program implies that r_- and r_+ provide respective lower and upper bound approximations for $\frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|}$ with bounded multiplicative error. With this, we can apply a binary search to approximate the true value of $\frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|}$.

Lemma 4.11. *Assume Condition 1. If the LP in Definition 4.8 is feasible for parameters $0 \leq r_- \leq r_+$, then*

$$\left(1 - 2 \cdot 2^{-K}\right) r_- \leq \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|} \leq \left(1 + 2 \cdot 2^{-K}\right) r_+.$$

Proof. First, we claim the following property for the feasible solutions.

Claim 4.12.

$$\begin{aligned} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{valid}} \\ \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &= 1, \quad \text{for all } \mathbf{x} \in Q^{C \setminus \{c_0\}}, \\ \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{valid}} \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &= 1, \quad \text{for all } \mathbf{y} \in Q^C. \end{aligned}$$

This claim can be proved by routinely verifying Definition 4.8. The proof is deferred to Appendix A. By summing the equations in Claim 4.12 over all $c \in Q^{C \setminus \{c_0\}}$ and respectively all $c \in Q^C$, we obtain:

$$(13) \quad \begin{aligned} |Q^{C \setminus \{c_0\}}| &= \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{valid}} \\ \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X, \\ |Q^C| &= \sum_{\mathbf{y} \in Q^C} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{valid}} \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y. \end{aligned}$$

Claim 4.13. *Assume Condition 1. The followings hold.*

$$\begin{aligned} \frac{1}{|Q^{C \setminus \{c_0\}}|} \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}:} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X &\leq 2^{-K}, \\ \frac{1}{|Q^C|} \sum_{\mathbf{y} \in Q^C} \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}:} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y &\leq 2^{-K}. \end{aligned}$$

This claim can be proved by a similar argument to the analysis of correlation decay in Section 3. The proof is deferred to Appendix B.

We then have

$$\begin{aligned} \text{(by (13))} \quad |Q^{C \setminus \{c_0\}}| &= \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{valid}}:} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\ &= \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}:} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\ &\quad + \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}:} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\ \text{(by Claim 4.13)} \quad &\leq \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}} |Q^{\mathcal{E} \wedge \sigma}| \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X + 2^{-K} \cdot |Q^{C \setminus \{c_0\}}|. \end{aligned}$$

Therefore, we have

$$\begin{aligned} |Q^{C \setminus \{c_0\}}| &\in \left[\hat{z}^X, \left(1 + 2 \cdot 2^{-K}\right) \hat{z}^X \right], \\ \text{where } \hat{z}^X &\triangleq \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}} |Q^{\mathcal{E} \wedge \sigma}| \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X; \\ |Q^C| &\in \left[\hat{z}^Y, \left(1 + 2 \cdot 2^{-K}\right) \hat{z}^Y \right], \\ \text{where } \hat{z}^Y &\triangleq \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}} |Q^{\mathcal{F} \wedge \tau}| \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y. \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|} &\leq \left(1 + 2 \cdot 2^{-K}\right) \frac{\sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}} |Q^{\mathcal{E} \wedge \sigma}| \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X}{\sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}} |Q^{\mathcal{F} \wedge \tau}| \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y} \\ &\leq \left(1 + 2 \cdot 2^{-K}\right) r_+, \end{aligned}$$

where the last inequality is due to the leaf constraints of the LP (Item III. of Definition 4.8). Symmetrically, we have

$$\begin{aligned} \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|} &\geq \left(1 - 2 \cdot 2^{-K}\right) \frac{\sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}} |Q^{\mathcal{E} \wedge \sigma}| \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X}{\sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}} |Q^{\mathcal{F} \wedge \tau}| \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y} \\ &\geq \left(1 - 2 \cdot 2^{-K}\right) r_-. \end{aligned}$$

Together, we have

$$\left(1 - 2 \cdot 2^{-K}\right) r_- \leq \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|} \leq \left(1 + 2 \cdot 2^{-K}\right) r_+. \quad \square$$

5. COUNTING AND SAMPLING VIA LINEAR PROGRAMMING

In this section, we utilize the linear program introduced in Section 4 to prove Theorems 1.1 and 1.2.

5.1. Constraint-wise self-reducibility in the local lemma regime. To obtain the algorithms for approximate counting and sampling required in Theorems 1.1 and 1.2, we will first provide a *marginal estimator* and a *dynamic sampler*. Within the local lemma regime characterized by Condition 1, the marginal estimator can efficiently estimate the ratio between the numbers of satisfying assignments for two sets of constraints differing in exactly one constraint; and the dynamic sampler can dynamically update a sample for the uniform satisfying assignment when a new constraint is added. Such a marginal estimator and a dynamic sampler are stated in the following two theorems.

Theorem 5.1 (marginal estimator). *There exists a deterministic algorithm such that for any $\Phi = (V, \mathcal{Q}, C)$ satisfying Condition 1, any $c_0 \in C$ and $\varepsilon \in (0, 1)$, within time $O\left(\left(\frac{1}{\varepsilon}\right)^{\text{poly}(k, D, \log q_{\max})}\right)$ the algorithm outputs an estimate \hat{r} satisfying*

$$(1 - \varepsilon) \frac{Z_\Phi}{Z_{\Phi^{c_0}}} \leq \hat{r} \leq (1 + \varepsilon) \frac{Z_\Phi}{Z_{\Phi^{c_0}}},$$

where Z_Φ and $Z_{\Phi^{c_0}}$ respectively denote the numbers of solutions to Φ and $\Phi^{c_0} = (V, \mathcal{Q}, C \setminus \{c_0\})$.

Theorem 5.2 (dynamic sampler). *There exists a randomized algorithm such that for any $\Phi = (V, \mathcal{Q}, C)$ satisfying Condition 1, any $c_0 \in C$ and $\varepsilon \in (0, 1)$, given access to a sample $\sigma \sim \mu_{\Phi^{c_0}}$ of the uniform distribution $\mu_{\Phi^{c_0}}$ over all solutions to $\Phi^{c_0} = (V, \mathcal{Q}, C \setminus \{c_0\})$, within time $O\left(\left(\frac{1}{\varepsilon}\right)^{\text{poly}(k, D, \log q_{\max})}\right)$ the algorithm updates the σ to an assignment $\tau \in \mathcal{Q}$ satisfying*

$$d_{\text{TV}}(\tau, \mu_\Phi) \leq \varepsilon.$$

Through a constraint-wise self-reduction that maintains the local lemma regime, Theorem 1.1 is implied by Theorem 5.1 and Theorem 1.2 is implied by Theorem 5.2.

Proof of Theorem 1.1. Let $C = \{c_1, c_2, \dots, c_m\}$. Note that we have $m \leq n(D+1)$. For each $0 \leq i \leq m$, we let $C_i = \{c_1, c_2, \dots, c_i\}$, $\Phi_i = (V, \mathcal{Q}, C_i)$ and Z_{Φ_i} denote the number of solutions to Φ_i . Note that $\Phi_m = \Phi$. We can then decompose Z_{Φ_m} in the following way:

$$(14) \quad Z_{\Phi_m} = Z_{\Phi_0} \cdot \frac{Z_{\Phi_m}}{Z_{\Phi_0}} = Z_{\Phi_0} \cdot \prod_{i=1}^m \frac{Z_{\Phi_i}}{Z_{\Phi_{i-1}}}.$$

Note that $Z_{\Phi_0} = \prod_{v \in V} |Q_v|$ is easily computable in $\text{poly}(n, q_{\max})$ time. We can then apply Theorem 5.1

to approximate each $\frac{Z_{\Phi_i}}{Z_{\Phi_{i-1}}}$ up to multiplicative error $\frac{\varepsilon}{4m}$ and multiply them together, by that Φ_i satisfies Condition 1 for each $0 \leq i \leq m$. In this way we obtain an estimate up to error ε within time $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q_{\max})}\right)$, as desired. \square

Proof of Theorem 1.2. Let $C = \{c_1, c_2, \dots, c_m\}$. Note that we have $m \leq n(D+1)$. For each $0 \leq i \leq m$, we let $C_i = \{c_1, c_2, \dots, c_i\}$ and $\Phi_i = (V, \mathcal{Q}, C_i)$. We build a sequence of assignments $\sigma_0, \sigma_1, \dots, \sigma_m \in \mathcal{Q}$ as follows:

- Sample $\sigma_0 \sim \mathcal{P}$.
- For each $1 \leq i \leq m$, use the dynamic sampler in Theorem 5.2 to obtain σ_i from σ_{i-1} with input formula (V, \mathcal{Q}, C_i) , input constraint c_i , input parameter $\frac{\varepsilon}{m}$, and input sample σ_{i-1} . We output an arbitrary final assignment instead if σ_i isn't a satisfying solution to Φ_i at some step $0 \leq i \leq m$.

By Theorem 5.2 and induction, it easily follows from maximally coupling σ_i with a uniform solution of Φ_i and applying the coupling lemma (Lemma 2.7) that

$$\forall 0 \leq i \leq m, d_{\text{TV}}(\sigma_i, \mu_{C_i}) \leq \frac{i}{m} \cdot \varepsilon.$$

Also, the overall running time is bounded by $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k, D, \log q_{\max})}\right)$. Therefore, the theorem is proved. \square

5.2. **Construction of the marginal estimator/sampler.** It remains to prove Theorems 5.1 and 5.2. Theorem 5.1 is proved by directly applying the linear program constructed and analyzed in Section 4.

Proof of Theorem 5.1. We set the truncation threshold as

$$K = 1 + \log \varepsilon^{-1}.$$

Applying a binary search on the parameters r_- and r_+ , one can obtain an estimate of $\frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|}$ within multiplicative error ε by Lemma 4.11. The running time bound follows from Proposition 4.9. \square

We then prove Theorem 5.2. The dynamic sampler claimed in Theorem 5.2 is constructed as follows.

Definition 5.3 (dynamic sampler). The algorithm takes as input a CSP $\Phi = (V, Q, C)$, a constraint $c_0 \in C$ and $\varepsilon \in (0, 1)$, and is given access to an assignment $\sigma^{\text{in}} \in Q$ which satisfies $\Phi^{c_0} = (V, Q, C \setminus \{c_0\})$. The algorithm proceeds as follows to update σ^{in} to a new assignment $\sigma^{\text{out}} \in Q$:

- Set the truncation threshold as

$$K = 1 + \log \left(\frac{\varepsilon}{4} \right)^{-1}.$$

Construct the LP in Definition 4.8 on the K -truncated coupling tree $\mathcal{T} = \mathcal{T}_K(\Phi, c_0)$. Apply a binary search to narrow down the interval $[r_-, r_+]$ so that $r_- \geq \frac{4+\varepsilon}{4+2\varepsilon} r_+$ and the LP is still feasible with parameters r_- and r_+ . By Lemma 4.11, it is guaranteed that

$$\left(1 - \frac{\varepsilon}{4}\right) r_- \leq \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|} \leq \left(1 + \frac{\varepsilon}{4}\right) r_+.$$

Let $\left\{ \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X, \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \right\}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})}$ be the corresponding LP feasible solution.

- Use the following random process to generate a root-to-leaf path in \mathcal{T} . The process starts from the root node $(C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$. In each step, at a non-leaf node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$, it does:
 - (1) if $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$, denoted by c be the smallest pinned constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$,
 - move to $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$ if c is satisfied by $\sigma_{\text{vbl}(c)}^{\text{in}}$;
 - otherwise, move to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \sigma_{\text{vbl}(c)}^{\text{in}}, \tau \wedge \rho, B \bowtie c)$ for each $\rho \in Q_{\text{vbl}(c)}$ w.p.

$$\frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \sigma_{\text{vbl}(c)}^{\text{in}}, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X};$$

- (2) otherwise $\mathcal{F}^\tau \subseteq \mathcal{E}^\sigma$, denoted by c be the smallest pinned constraint in $\mathcal{E}^\sigma \setminus \mathcal{F}^\tau$,
 - move to $(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)$ with probability

$$\frac{\hat{p}_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X};$$

- move to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \sigma_{\text{vbl}(c)}^{\text{in}}, \tau \wedge \rho, B \bowtie c)$ for $\rho = \text{False}(c)$ with probability

$$\frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \sigma_{\text{vbl}(c)}^{\text{in}}, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X}.$$

In the end, the random process arrives at a random leaf $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}$. Let P^{samp} denote the root-to-leaf path generated by this process.

- If the sampled leaf node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \notin \mathcal{L}_{\text{coup}}$, then let $\sigma^{\text{out}} \in Q$ be an arbitrary unsatisfying assignment of Φ ; otherwise, update σ^{in} by changing the values of the assigned variables to the corresponding values as in τ , i.e. let $\sigma^{\text{out}} \leftarrow \tau \wedge \sigma_{V \setminus \Lambda(\tau)}^{\text{in}}$.

We then finish the section by proving the following lemma, which directly implies Theorem 5.2.

Lemma 5.4. *Assume Condition 1 and $\sigma^{\text{in}} \sim \mu_{C \setminus \{c_0\}}$. Then the σ^{out} produced by the dynamic sampler in Definition 5.3 satisfies*

$$d_{\text{TV}}(\sigma^{\text{out}}, \mu_C) \leq \varepsilon.$$

Proof. We need the following lemma, whose proof is deferred to Appendix B.

Lemma 5.5. *Assume Condition 1. Let $\sigma^{\text{in}} \sim \mu_{C \setminus \{c_0\}}$. For any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, we have*

$$\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{samp}}] = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X.$$

Moreover, conditioning on $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{samp}}$, it follows that

$$\sigma^{\text{in}} \sim \mu_{\mathcal{E}}^{\sigma},$$

for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ such that $\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{samp}}] > 0$.

Then for each $\mathbf{y} \in Q^C$, we have

$$\begin{aligned} \Pr [\sigma^{\text{out}} = \mathbf{y}] &= \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \cdot \frac{1}{|Q^{\mathcal{E} \wedge \sigma}|} \\ \text{(by Item III.a of Definition 4.8)} &\geq r_- \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \cdot \frac{1}{|Q^{\mathcal{E} \wedge \sigma}|} \\ (\star) &= r_- \cdot \frac{1}{|Q^C \setminus \{c_0\}|} \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y. \end{aligned}$$

Here, the first equality is by that $\mathcal{E}^{\sigma} = \mathcal{F}^{\tau}$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}$, and hence each solution in $Q^{\mathcal{F} \wedge \tau}$ is generated with equal probability $\frac{1}{|Q^{\mathcal{F} \wedge \tau}|} = \frac{1}{|Q^{\mathcal{E} \wedge \sigma}|}$ by Lemma 5.5. The \star equality is by the chain rule and that $\mathcal{E} \wedge \sigma \implies C \setminus \{c_0\}$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, following the argument in Remark 4.6.

Consider a distribution ν defined over Q such that for each $\mathbf{y} \in Q^C$, we let

$$(15) \quad \nu[\mathbf{y}] = \frac{1}{|Q^C|} \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y,$$

and that distribution of ν over unsatisfying assignments of Φ is carefully designed in a way that

$$(16) \quad d_{\text{TV}}(\nu, \sigma^{\text{out}}) = \left| \frac{1}{|Q^C|} - r_- \cdot \frac{1}{|Q^C \setminus \{c_0\}|} \right| \sum_{\mathbf{y} \in Q^C} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y.$$

By $(1 - \frac{\varepsilon}{4}) r_- \cdot \frac{|Q^C \setminus \{c_0\}|}{|Q^C|} \leq (1 + \frac{\varepsilon}{4}) r_+$ and $r_- \geq \frac{4+\varepsilon}{4+2\varepsilon} r_+$ ensured by the process, we then have

$$(17) \quad \left(1 - \frac{\varepsilon}{2}\right) \cdot \frac{|Q^C \setminus \{c_0\}|}{|Q^C|} \leq r_- \leq \left(1 + \frac{\varepsilon}{2}\right) \cdot \frac{|Q^C \setminus \{c_0\}|}{|Q^C|}.$$

Combining (16) and (17), we obtain

$$(18) \quad d_{\text{TV}}(\nu, \sigma^{\text{out}}) \leq \frac{\varepsilon}{2} \cdot \frac{1}{|Q^C|} \sum_{\mathbf{y} \in Q^C} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \leq \frac{\varepsilon}{2}.$$

Now that by (15) and Claim 4.12, for each $\mathbf{y} \in Q^C$ we have $\nu[\mathbf{y}] \leq \frac{1}{|Q^C|} = \mu_C(\mathbf{y})$. Hence,

$$\begin{aligned} d_{\text{TV}}(\mu_C, \nu) &= \sum_{\mathbf{y} \in Q^C} \left(\frac{1}{|Q^C|} - \frac{1}{|Q^C|} \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \right) \\ \text{(by Claim 4.12)} &= \frac{1}{|Q^C|} \sum_{\mathbf{y} \in Q^C} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}: \\ \mathbf{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \end{aligned}$$

$$\text{(by Claim 4.13)} \quad \leq \frac{\varepsilon}{2}.$$

Combining with (18) and using the triangle inequality for the total variation distance, we obtain

$$d_{\text{TV}}(\mu_C, \sigma^{\text{out}}) \leq d_{\text{TV}}(\mu_C, \nu) + d_{\text{TV}}(\nu, \sigma^{\text{out}}) = \varepsilon,$$

finishing the proof of the lemma. \square

6. CONCLUSION AND OPEN PROBLEMS

In this paper, we present polynomial-time algorithms for approximate counting/almost uniform sampling atomic constraint satisfaction solutions in the regime of $pD^{2+o(1)} \lesssim 1$ (Condition 1), where the $o(1)$ factor approaches zero as the minimum domain size grows. This regime almost matches the lower bound $pD^2 \lesssim 1$ given by [BGG⁺19, GGW22]. Even for the worst-case of Boolean domains, our regime $pD^{4.82} \lesssim 1$ still improves over the previous best regime of $pD^5 \lesssim 1$ [HWY23a].

At the heart of our approach is a novel constraint-wise coupling for CSPs. This coupling, along with its analyses, sharply captures the critical phenomenon for counting/sampling LLL and may be of independent interest, helping pave the way for future improvements.

Here, we outline several future research directions that our findings may serve as a basis for:

- An obvious open problem is establishing the optimal counting/sampling LLL for small domain sizes q , especially for counting/sampling k -CNF, where the domain size $q = 2$.
- Another open direction is to generalize the current approach to general CSPs, possibly with non-atomic constraints.
- Last but not the least, an open aspect is to devise *fast* algorithms for sampling LLL that runs in polynomial time when k and D are allowed to be unbounded or in near-linear time when k and D are bounded. Such fast algorithms should avoid relying on exhaustive enumerations of local structures, unlike the deterministic approximate counting algorithms to this day.

REFERENCES

- [AI14] Dimitris Achlioptas and Fotis Iliopoulos. Random walks that find perfect objects and the Lovász local lemma. In *FOCS*, pages 494–503. IEEE, 2014.
- [AIS19] Dimitris Achlioptas, Fotis Iliopoulos, and Alistair Sinclair. Beyond the Lovász local lemma: point to set correlations and their algorithmic applications. In *FOCS*, pages 725–744. IEEE, 2019.
- [Alo91] Noga Alon. A parallel algorithmic version of the local lemma. In *FOCS*, pages 586–593. IEEE, 1991.
- [ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *FOCS*, pages 1319–1330. IEEE, 2020.
- [BD97] Russ Bubley and Martin Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *FOCS*, page 223. IEEE, 1997.
- [BDK06] Magnus Bordewich, Martin Dyer, and Marek Karpinski. Stopping times, metrics and approximate counting, 2006.
- [BDK08] Magnus Bordewich, Martin Dyer, and Marek Karpinski. Path coupling using stopping times and counting independent sets and colorings in hypergraphs. *ALGORITHMS*, page 2008, 2008.
- [Bec91] József Beck. An algorithmic approach to the Lovász local lemma. *Random Struct. Algorithms*, 2(4):343–365, 1991.
- [BGG⁺19] Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Daniel Štefankovič. Approximation via correlation decay when strong spatial mixing fails. *SIAM J. Comput.*, 48(2):279–349, 2019.
- [BGK⁺07] Mohsen Bayati, David Gamarnik, Dimitriy Katz, Chandra Nair, and Prasad Tetali. Simple deterministic approximation algorithms for counting matchings. In *STOC*, pages 122–127, 2007.

- [CGŠV21] Zongchen Chen, Andreas Galanis, Daniel Štefankovič, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *SODA*, page 1548–1557. SIAM, 2021.
- [CLV20] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of glauber dynamics up to uniqueness via contraction. In *FOCS*, pages 1307–1318. IEEE, 2020.
- [CMM23] Zongchen Chen, Nitya Mani, and Ankur Moitra. From algorithms to connectivity and back: Finding a giant component in random k -SAT. In *SODA*, pages 3437–3470. SIAM, 2023.
- [CS00] Artur Czumaj and Christian Scheideler. Coloring nonuniform hypergraphs: a new algorithmic approach to the general Lovász local lemma. *Random Struct. Algorithms*, 17(3-4):213–237, 2000.
- [Dob70] R. L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theory of Probability & Its Applications*, 15(3):458–486, 1970.
- [EL75] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and finite sets, volume 10 of Colloquia Mathematica Societatis János Bolyai*, pages 609–628, 1975.
- [FA17] Alan Frieze and Michael Anastos. Randomly coloring simple hypergraphs with fewer colors. *Information Processing Letters*, 126:39–42, 2017.
- [FGW22] Weiming Feng, Heng Guo, and Jiaheng Wang. Improved bounds for randomly colouring simple hypergraphs. In *RANDOM*, pages 25:1–25:17. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.
- [FGW⁺23] Weiming Feng, Heng Guo, Chunyang Wang, Jiaheng Wang, and Yitong Yin. Towards derandomising markov chain monte carlo. In *FOCS*, pages 1963–1990. IEEE, 2023.
- [FGYZ21a] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Fast sampling and counting k -SAT solutions in the local lemma regime. *J. ACM*, 68(6):Art. 40, 42, 2021.
- [FGYZ21b] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the Boolean Domain. In *SODA*, pages 1558–1577. SIAM, 2021.
- [FHY21] Weiming Feng, Kun He, and Yitong Yin. Sampling constraint satisfaction solutions in the local lemma regime. In *STOC*, pages 1565–1578. ACM, 2021.
- [FM11] Alan Frieze and Páll Melsted. Randomly coloring simple hypergraphs. *Information Processing Letters*, 111(17):848–853, 2011.
- [GGGHP22] Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Andrés Herrera-Poyatos. Fast sampling of satisfying assignments from random k -sat. *arXiv*, abs/2206.15308, 2022.
- [GGGY20] Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Kuan Yang. Counting solutions to random CNF formulas. In *ICALP*, volume 168 of *LIPICs*, pages 53:1–53:14, 2020.
- [GGW22] Andreas Galanis, Heng Guo, and Jiaheng Wang. Inapproximability of counting hypergraph colourings. *ACM Trans. Comput. Theory*, 2022.
- [GKM15] David Gamarnik, Dmitriy Katz, and Sidhant Misra. Strong spatial mixing of list coloring of graphs. *Random Structures & Algorithms*, 46(4):599–613, 2015.
- [GLLZ19] Heng Guo, Chao Liao, Pinyan Lu, and Chihao Zhang. Counting hypergraph colorings in the local lemma regime. *SIAM J. Comput.*, 48(4):1397–1424, 2019.
- [GMP05] Leslie Ann Goldberg, Russell Martin, and Mike Paterson. Strong spatial mixing with fewer colors for lattice graphs. *SIAM J. Comput.*, 35(2):486–517, 2005.
- [Har21] David G. Harris. Oblivious resampling oracles and parallel algorithms for the lopsided lovász local lemma. *ACM Trans. Algorithms*, 17(1), 2021.
- [HS17a] David G. Harris and Aravind Srinivasan. Algorithmic and enumerative aspects of the moser-tardos distribution. *ACM Trans. Algorithms*, 13(3), 2017.
- [HS17b] David G. Harris and Aravind Srinivasan. A constructive Lovász local lemma for permutations. *Theory Comput.*, 13:Paper No. 17, 41, 2017.
- [HS19] David G. Harris and Aravind Srinivasan. The Moser-Tardos framework with partial resampling. *J. ACM*, 66(5):Art. 36, 45, 2019.
- [HSS11] Bernhard Haeupler, Barna Saha, and Aravind Srinivasan. New constructive aspects of the lovász local lemma. *J. ACM*, 58(6):28:1–28:28, 2011. (Conference version in *FOCS’10*).
- [HSW21] Kun He, Xiaoming Sun, and Kewen Wu. Perfect sampling for (atomic) Lovász local lemma. *arXiv*, abs/2107.03932, 2021.

- [HSZ19] Jonathan Hermon, Allan Sly, and Yumeng Zhang. Rapid mixing of hypergraph independent sets. *Random Struct. Algorithms*, 54(4):730–767, 2019.
- [HV15] Nicholas J. A. Harvey and Jan Vondrák. An algorithmic proof of the Lovász local lemma via resampling oracles. In *FOCS*, pages 1327–1345. IEEE, 2015.
- [HWY22] Kun He, Chunyang Wang, and Yitong Yin. Sampling lovász local lemma for general constraint satisfaction solutions in near-linear time. In *FOCS*, pages 147–158. IEEE, 2022.
- [HWY23a] Kun He, Chunyang Wang, and Yitong Yin. Deterministic counting lovász local lemma beyond linear programming. In *SODA*, pages 3388–3425. SIAM, 2023.
- [HWY23b] Kun He, Kewen Wu, and Kuan Yang. Improved bounds for sampling solutions of random CNF formulas. In *SODA*, pages 3330–3361. SIAM, 2023.
- [JPV21a] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. On the sampling Lovász local lemma for atomic constraint satisfaction problems. *arXiv*, abs/2102.08342, 2021.
- [JPV21b] Vishesh Jain, Huy Tuan Pham, and Thuy Duong Vuong. Towards the sampling lovász local lemma. In *FOCS*, pages 173–183. IEEE, 2021.
- [Kha80] L.G. Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980.
- [Kol16] Vladimir Kolmogorov. Commutativity in the algorithmic Lovász local lemma. In *FOCS*, pages 780–787. IEEE, 2016.
- [LLY13] Liang Li, Pinyan Lu, and Yitong Yin. Correlation decay up to uniqueness in spin systems. In *SODA*, page 67–84. SIAM, 2013.
- [LP17] David A. Levin and Yuval Peres. *Markov chains and mixing times*. American Mathematical Soc., 2017.
- [LSS19] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. Correlation decay and partition function zeros: Algorithms and phase transitions. *SIAM J. Comput.*, pages 200–252, 2019.
- [LY13] Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *RANDOM*, pages 639–654. Springer, 2013.
- [Moi19] Ankur Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. *J. ACM*, 66(2):10:1–10:25, 2019. (Conference version in *STOC*’17).
- [Mos09] Robin A. Moser. A constructive proof of the Lovász local lemma. In *STOC*, pages 343–350. ACM, 2009.
- [MR99] Michael Molloy and Bruce Reed. Further algorithmic aspects of the local lemma. In *STOC ’98 (Dallas, TX)*, pages 524–529. ACM, 1999.
- [MT10] Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász local lemma. *J. ACM*, 57(2):11, 2010.
- [PR17] Viresh Patel and Guus Regts. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM J. Comput.*, 46(6):1893–1919, 2017.
- [QW24] Guoliang Qiu and Jiaheng Wang. Inapproximability of counting independent sets in linear hypergraphs. *Information Processing Letters*, 184:106448, 2024.
- [QWZ22] Guoliang Qiu, Yanheng Wang, and Chihao Zhang. A perfect sampler for hypergraph independent sets. In *ICALP*, pages 103:1–103:16. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.
- [She85] J. B. Shearer. On a problem of Spencer. *Combinatorica*, 5(3):241–245, 1985.
- [Sri08] Aravind Srinivasan. Improved algorithmic versions of the Lovász local lemma. In *SODA*, pages 611–620. SIAM, 2008.
- [SST12] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. In *SODA*, pages 941–953. SIAM, 2012.
- [Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149. ACM, 2006.
- [YZ13] Yitong Yin and Chihao Zhang. Approximate counting via correlation decay on planar graphs. In *SODA*, pages 47–66. SIAM, 2013.

In this section, we will prove two basic properties of the linear program constructed in Section 4, specifically Proposition 4.9 and Claim 4.12.

Recall that we have fixed a CSP $\Phi = (V, Q, C)$ and a constraint $c_0 \in C$. Also, recall the notations in Definition 4.1, the K -truncated coupling tree $\mathcal{T} = \mathcal{T}_K(\Phi, c_0)$ in Definition 4.5, and the linear program in Definition 4.8.

A.1. Proof of Proposition 4.9. Before proving Proposition 4.9, we need to first prove the following lemma, which bounds the depth of the coupling tree \mathcal{T} .

Lemma A.1. \mathcal{T} has depth at most $KD(D+1) + 1$.

Proof. Suppose to the contrary that there exists a node $(\mathcal{E}', \mathcal{F}', \sigma', \tau', B') \in V(\mathcal{T})$ of depth $KD(D+1) + 2$. We track the size of $\mathcal{E}^\sigma \Delta \mathcal{F}^\tau$ when going down the path between the root and $(\mathcal{E}', \mathcal{F}', \sigma', \tau', B')$ and denote it by t . Initially, by Condition 1, we have $t = 1$. Note that each time we either add some c into \mathcal{E} or \mathcal{F} , and let t decrease by one; or we assign some values to σ and τ on $\text{vbl}(c)$, and let t increase by at most $D - 1$ (note that at most D new elements can be added into $\mathcal{E}^\sigma \Delta \mathcal{F}^\tau$ and that c is removed from $\mathcal{E}^\sigma \Delta \mathcal{F}^\tau$). We denote the number of times the latter operation being executed by i .

By $(\mathcal{E}', \mathcal{F}', \sigma', \tau', B')$ is of depth $KD(D+1) + 2$, the above step is repeated for $KD(D+1) + 2$ times. Note that at $(\mathcal{E}', \mathcal{F}', \sigma', \tau', B')$ we have $t \geq 0$, then

$$(D-1) \cdot i + 1 - (KD(D+1) + 2 - i) \geq 0,$$

and hence $i > K(D+1)$. Note that by Definition 4.5 and (5), we have $|B'| > K$, contradicting the truncation condition at Item 2a of Definition 4.5. Therefore, the lemma is proved. \square

We are now ready to prove Proposition 4.9.

Proof of Proposition 4.9. Note that the branching number of the coupling tree is at most $(q_{\max})^{2k}$. Hence by Lemma A.1, the size of the coupling tree is bounded by $\exp(K \cdot \text{poly}(k, \log q_{\max}, D))$.

Therefore, the total number of constraints of Items I. through III. in Definition 4.8 is also bounded by $\exp(K \cdot \text{poly}(k, \log q, D))$. For Item IV. of Definition 4.8, we only show how to efficiently check the feasibility for the first inequality, and the other one follows analogously. We note that after fixing some $T \in \mathbb{T}_K^{c_0}$, for all leaf nodes $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}$ satisfying $B = T$, the following properties hold:

- for any $v \in \Lambda(\sigma)$, there exist $c \in C, c' \in T$ such that $v \in \text{vbl}(c)$ and $\text{vbl}(c) \cap \text{vbl}(c') \neq \emptyset$;
- for any $c \in \mathcal{E} \setminus (C \setminus \{c_0\})$, we have $\Lambda(\sigma) \cap \text{vbl}(c^O) \neq \emptyset$.

Here, the first property can be shown using proof by contradiction: Suppose that such $c \in C, c' \in T$ cannot be found. by the time v is assigned in σ/τ through the root-to-leaf path in \mathcal{T} , let c^* be the (pinned) constraint chosen at the time, then $(c^*)^O$ doesn't share variable with any constraint in B at the time according to the assumption, and thus $(c^*)^O$ must be added into B by Definition 4.8 and (5). Then we have a contradiction by letting $c = c' = c^*$. The second property follows from that each (pinned) constraint c chosen by the procedure and added into \mathcal{E} in Definition 4.8 lies in $\mathcal{E}^\sigma \Delta \mathcal{F}^\tau$ at the time, and c can added into $\mathcal{E}^\sigma \Delta \mathcal{F}^\tau$ only by either

- c is exactly c_0 ;
- or c is pinned by σ/τ , i.e. $\text{vbl}(c^O) \neq \text{vbl}(c)$.

By the above properties, the feasibility of the first equation can be directly verified by enumerating the assignment of \mathbf{x} on variables of constraints within distance 2 of T in G_Φ . This says it suffices to only check $(q_{\max})^{Kk(D+1)^2}$ (partial) assignments of \mathbf{x} to verify the first inequality in Item IV. of Definition 4.8. Combining with the bound on the number of 2-trees in Lemma 2.6, we have the number of constraints in Item IV. of Definition 4.8 can be bounded by $\exp(K \cdot \text{poly}(k, \log q_{\max}, D))$.

Hence, we have both the number of variables and constraints of the linear program defined in Definition 4.8 is at most $\exp(K \cdot \text{poly}(k, \log q_{\max}, D))$. Then the lemma follows from the standard running time guarantees for linear programming (e.g., [Kha80]). \square

A.2. **Proof of Claim 4.12.** The proof of Claim 4.12 comes from verifying several basic definitions of the linear program.

Proof of Claim 4.12. We only prove the first equality of Claim 4.12, then the second equality follows analogously. By Item II. of Definition 4.8 and an induction on the increasing depth of \mathcal{T} , we have

$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}: \\ \mathbf{x} \in \mathcal{Q}^{\mathcal{E} \wedge \sigma}}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^{\mathbf{x}} = 1, \quad \text{for all } \mathbf{x} \in \mathcal{Q}^{C \setminus \{c_0\}}.$$

Note that for any $x \in \mathcal{Q}^{C \setminus \{c_0\}}$, any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{invld}}$ such that $\mathbf{x} \in \mathcal{Q}^{\mathcal{E} \wedge \sigma}$, it must hold that τ violates \mathcal{F} . Then, the lemma holds immediately by applying Item III.b of Definition 4.8. \square

APPENDIX B. PROBABILISTIC PROPERTIES OF THE LINEAR PROGRAM

In this section, we prove several important probabilistic properties related to the linear program. Specifically, we will finish the proof of Lemma 4.10 regarding Item IV. of Definition 4.8, also prove Claim 4.13 and Lemma 5.5. The key ideas of the proofs will heavily resemble the proofs of the correlation decay property in Section 3.

Recall that we have fixed a CSP $\Phi = (V, \mathcal{Q}, C)$ and a constraint $c_0 \in C$. Also, recall the notations in Definition 4.1, the K -truncated coupling tree $\mathcal{T} = \mathcal{T}_K(\Phi, c_0)$ in Definition 4.5, and the linear program in Definition 4.8.

B.1. **Remaining proof of Lemma 4.10.** Recall that for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, we let

$$(19) \quad \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^{\mathbf{x}} = \mu_C(\mathcal{F} \wedge \tau), \quad \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^{\mathbf{y}} = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma).$$

We then finish the proof of Lemma 4.10 by showing the above choice satisfies Item IV. of Definition 4.8. We only prove the first inequality, and the second inequality follows analogously.

We need the definition of a random process, which is almost the same process as Definition 3.6, except that now \mathfrak{X} follows the distribution of \mathcal{P} instead of $\mu_{C \setminus \{c_0\}}$. Also, the random process now can be interpreted as generating a root-to-leaf path in \mathcal{T} . Here in this random process, it is possible to reach those nodes $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{invld}}$ due to the different distribution of \mathfrak{X} .

Definition B.1 (random process associated with the true marginal probabilities). Let $\mathfrak{X}^{\text{mg}} \sim \mathcal{P}$ and $\mathfrak{Y}^{\text{mg}} \sim \mu_C$ be drawn independently beforehand. Define the random process $P^{\text{mg}} = P_K^{\text{mg}} = \{(\mathcal{E}_t, \mathcal{F}_t, \sigma_t, \tau_t, B_t)\}_{t \geq 0}$ starting from the initial state $(\mathcal{E}_0, \mathcal{F}_0, \sigma_0, \tau_0, B_0) = (C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$ as follows:

- (1) If σ violates \mathcal{E} or τ violates \mathcal{F} or $\mathcal{E}^\sigma = \mathcal{F}^\tau$ or $|B| = K$, the process stops and $(\mathcal{E}_t, \mathcal{F}_t, \sigma_t, \tau_t, B_t)$ is the outcome of the process.
- (2) Otherwise, suppose $\mathcal{F}_t^{\tau_t} \not\subseteq \mathcal{E}_t^{\sigma_t}$. Let c be the smallest pinned constraint in $\mathcal{F}_t^{\tau_t} \setminus \mathcal{E}_t^{\sigma_t}$.

$$(\mathcal{E}_{t+1}, \mathcal{F}_{t+1}, \sigma_{t+1}, \tau_{t+1}, B_{t+1}) \leftarrow \begin{cases} (\mathcal{E}_t \cup \{c\}, \mathcal{F}_t, \sigma_t, \tau_t, B_t) & c \text{ is satisfied by } \mathfrak{X}_{\text{vbl}(c)}^{\text{mg}}; \\ \left(\mathcal{E}_t, \mathcal{F}_t, \sigma_t \wedge \mathfrak{X}_{\text{vbl}(c)}^{\text{mg}}, \tau_t \wedge \mathfrak{Y}_{\text{vbl}(c)}^{\text{mg}}, B_t \bowtie c \right) & \text{otherwise.} \end{cases}$$

- (3) Otherwise $\mathcal{F}_t^\tau \subseteq \mathcal{E}_t^\sigma$. Let c be the smallest pinned constraint in $\mathcal{E}_t^\sigma \setminus \mathcal{F}_t^\tau$.

$$(\mathcal{E}_{t+1}, \mathcal{F}_{t+1}, \sigma_{t+1}, \tau_{t+1}, B_{t+1}) \leftarrow \begin{cases} (\mathcal{E}_t, \mathcal{F}_t \cup \{c\}, \sigma_t, \tau_t, B_t) & c \text{ is satisfied by } \mathfrak{Y}_{\text{vbl}(c)}^{\text{mg}}; \\ \left(\mathcal{E}_t, \mathcal{F}_t, \sigma_t \wedge \mathfrak{X}_{\text{vbl}(c)}^{\text{mg}}, \tau_t \wedge \mathfrak{Y}_{\text{vbl}(c)}^{\text{mg}}, B_t \bowtie c \right) & \text{otherwise.} \end{cases}$$

Let μ^{mg} denote the distribution of the outcome $(\mathcal{E}_\infty, \mathcal{F}_\infty, \sigma_\infty, \tau_\infty, B_\infty)$ of this process.

By comparing with the definition of $\mathcal{T} = \mathcal{T}_K(\Phi, c_0)$ in Definition 4.5. We have all tuples possibly visited by Definition B.4 are in $V(\mathcal{T})$.

The following lemma holds for the process in Definition B.1.

Lemma B.2. Assume Condition 1. For each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$,

$$(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \iff \mathfrak{X}^{\text{mg}} \in \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma} \wedge \mathfrak{Y}^{\text{mg}} \in \mathcal{Q}^{\mathcal{F} \wedge \tau}.$$

The following corollary is then direct by combining Lemma B.2 with the law of conditional probability.

Corollary B.3. Assume Condition 1. For each $\mathbf{x} \in \mathcal{Q}$ and each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$,

$$\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \mid \mathfrak{X}^{\text{mg}} = \mathbf{x}] = \begin{cases} 0 & \mathbf{x} \notin \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}; \\ \mu_C(\mathcal{F} \wedge \tau) & \mathbf{x} \in \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}. \end{cases}$$

Proof. For each $\mathbf{x} \in \mathcal{Q}$ and each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}$ we have

$$(20) \quad \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \mid \mathfrak{X}^{\text{mg}} = \mathbf{x}] = \frac{\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \wedge \mathfrak{X}^{\text{mg}} = \mathbf{x}]}{\Pr [\mathfrak{X}^{\text{mg}} = \mathbf{x}]} \\ \text{(by Definition B.1)} = |\mathcal{Q}| \cdot \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \wedge \mathfrak{X}^{\text{mg}} = \mathbf{x}].$$

By Lemma B.2, if $\mathbf{x} \notin \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}$, then the above term clearly equals 0. We then assume $\mathbf{x} \in \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}$. Note that by Definition B.1 and Lemma B.2, conditioning on the event $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}}$, the distribution of \mathfrak{X}^{mg} is uniform over all assignments in $\mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}$. Therefore,

$$\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \wedge \mathfrak{X}^{\text{mg}} = \mathbf{x}] = \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}}] \cdot \frac{1}{|\mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}|} \\ \text{(by Lemma B.2)} = \frac{|\mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}|}{|\mathcal{Q}|} \cdot \frac{|\mathcal{Q}^{\mathcal{F} \wedge \tau}|}{|\mathcal{Q}^C|} \cdot \frac{1}{|\mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}|} \\ (\star) = \frac{1}{|\mathcal{Q}|} \cdot \mu_C(\mathcal{F} \wedge \tau),$$

combining with (20) proves the corollary. Here, the \star equality is by that $\mathcal{F} \wedge \tau \implies C$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, following the argument in Remark 4.6. \square

We are now ready to finish the proof of Lemma 4.10.

Remaining proof of Lemma 4.10. Let $(\mathcal{E}^{\text{mg}}, \mathcal{F}^{\text{mg}}, \sigma^{\text{mg}}, \tau^{\text{mg}}, B^{\text{mg}}) \sim \mu^{\text{mg}}$ denote the random outcome of the process P^{mg} constructed in Definition B.1.

Recall that $\mathbb{T}_K^{c_0}$ is the set of 2-trees in G_Φ of size K containing c_0 . By Corollary B.3, we have for each $\mathbf{x} \in \mathcal{Q}$ and each $T \in \mathbb{T}_K^{c_0}$,

$$(21) \quad \Pr [B^{\text{mg}} = T \mid \mathfrak{X}^{\text{mg}} = \mathbf{x}] = \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}} \\ B=T}} \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \mid \mathfrak{X}^{\text{mg}} = \mathbf{x}] \\ \text{(by Corollary B.3 and (19))} = \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}} \\ B=T \wedge \mathbf{x} \in \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\})}}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^{\mathfrak{X}},$$

which is exactly the expression in the left-hand side of the first inequality in Item IV. of Definition 4.8.

Fix some $T \in \mathbb{T}_K^{c_0}$. Now it remains to bound $\Pr [B^{\text{mg}} = T \mid \mathfrak{X}^{\text{mg}} = \mathbf{x}]$. Here, we leverage the analysis in Lemma 3.11. Following the argument in Lemma 3.11, we have that for each $c \in T$, $c \in B^{\text{mg}}$ implies

$$\mathcal{A}_c^{\text{mg}} : \forall v \in \text{vbl}(c), \quad \mathfrak{X}^{\text{mg}}(v) = \text{False}(c)(v) \vee \mathfrak{Y}^{\text{mg}}(v) = \text{False}(c)(v).$$

Therefore, we have

$$\Pr [B^{\text{mg}} = T \mid \mathfrak{X}^{\text{mg}} = \mathbf{x}] \leq \Pr \left[\bigwedge_{c \in T} \mathcal{A}_c^{\text{mg}} \mid \mathfrak{X}^{\text{mg}} = \mathbf{x} \right] \\ \text{(by Definition B.1)} = \Pr_{\mathfrak{Y} \sim \mu_C} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right] \\ \text{(by Theorem 2.3)} \leq (1 - \epsilon p)^{K(D+1)} \Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right].$$

Combining with (21), Lemma 4.10 is proved. \square

We finish this subsection by proving Lemma B.2.

Proof of Lemma B.2. By Φ satisfy Condition 1, \mathfrak{Y}^{mg} and the procedure in Definition B.1 are well-defined. We then prove by structural induction in the top-down order that for any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ such that $\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}}] > 0$,

$$(22) \quad (\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \iff \mathfrak{X}^{\text{mg}} \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma} \wedge \mathfrak{Y}^{\text{mg}} \in Q^{\mathcal{F} \wedge C \wedge \tau},$$

then Lemma B.2 directly holds.

The base case is when $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) = (C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$. In this case, $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}}$ and $Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma} = Q^{\mathcal{F} \wedge C \wedge \tau} = Q$. So (22) holds, proving the base case.

For the induction step, we assume for the current $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ that $\mathcal{E}^\sigma \neq \mathcal{F}^\tau$, $|B| < K$, σ doesn't violate \mathcal{E} and τ doesn't violate \mathcal{F} . We then only prove the case for $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$. The case when $\mathcal{F}^\tau \subseteq \mathcal{E}^\sigma$ follows analogously.

let c be the smallest constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$. By Definition B.1, we have the following two cases:

- The next tuple is $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$, then we have

$$\begin{aligned} & (\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \\ \text{(by Definition 3.6)} \iff & (\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \wedge \mathfrak{X}_{\text{vbl}(c_0)}^{\text{mg}} \text{ satisfies } c \\ \text{(by I.H.)} \iff & \mathfrak{X}^{\text{mg}} \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma} \wedge \mathfrak{Y}^{\text{mg}} \in Q^{\mathcal{F} \wedge \tau} \wedge \mathfrak{X}_{\text{vbl}(c_0)}^{\text{mg}} \text{ satisfies } c \\ \iff & \mathfrak{X}^{\text{mg}} \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \cup \{c\} \wedge \sigma} \wedge \mathfrak{Y}^{\text{mg}} \in Q^{\mathcal{F} \wedge \tau} \\ (\star) \iff & \mathfrak{X}^{\text{mg}} \in Q^{(\mathcal{E} \cup \{c\}) \setminus (C \setminus \{c_0\}) \wedge \sigma} \wedge \mathfrak{Y}^{\text{mg}} \in Q^{\mathcal{F} \wedge \tau}. \end{aligned}$$

Here, the \star implication is by that each $c \in \mathcal{F}^\tau \setminus \mathcal{E}^\sigma$ chosen by the random process in Definition B.1 must satisfy either

- is exactly c_0 ;
- or is pinned by σ/τ , i.e. $\text{vbl}(c) \neq \text{vbl}(c^0)$,

and hence cannot be in $C \setminus \{c_0\}$.

- The next tuple is $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)$ for $\pi = \text{False}(c)$ and some $\rho \in Q_{\text{vbl}(c)}$, then we have

$$\begin{aligned} & (\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in P^{\text{mg}} \\ \text{(by Definition 3.6)} \iff & (\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \wedge \mathfrak{X}^{\text{mg}} \text{ violates } c \wedge \mathfrak{X}_{\text{vbl}(c)}^{\text{mg}} = \pi \wedge \mathfrak{Y}_{\text{vbl}(c)}^{\text{mg}} = \rho \\ \text{(by } \pi = \text{False}(c)) \iff & (\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{mg}} \wedge \mathfrak{X}_{\text{vbl}(c)}^{\text{mg}} = \pi \wedge \mathfrak{Y}_{\text{vbl}(c)}^{\text{mg}} = \rho \\ \text{(by I.H.)} \iff & \mathfrak{X}^{\text{mg}} \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma} \wedge \mathfrak{Y}^{\text{mg}} \in Q^{\mathcal{F} \wedge \tau} \wedge \mathfrak{X}_{\text{vbl}(c)}^{\text{mg}} = \pi \wedge \mathfrak{Y}_{\text{vbl}(c)}^{\text{mg}} = \rho \\ \iff & \mathfrak{X}^{\text{mg}} \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma \wedge \pi} \wedge \mathfrak{Y}^{\text{mg}} \in Q^{\mathcal{F} \wedge \tau \wedge \rho}, \end{aligned}$$

finishing the induction steps and the proof of the lemma. □

B.2. Proof of Claim 4.13 and Lemma 5.5. In this subsection, we simultaneously prove Claim 4.13 and Lemma 5.5. Similar as the introduction of Definition 3.6 in Section 3 and Definition B.1 in the previous subsection, we need to define the following random process for explicitly identified randomness:

Definition B.4 (two random processes associated with the linear program). We consider the following way of generating root-to-leaf paths of \mathcal{T} . Initiate a random assignment $\mathfrak{X}^{\text{lp}} \in Q$ distributed as either

- $\mathfrak{X}^{\text{lp}} \sim \mathcal{P}$ (type 1 initialization)
- or $\mathfrak{X}^{\text{lp}} \sim \mu_{C \setminus \{c_0\}}$ (type 2 initialization)

The process starts from the root node $(C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$ of \mathcal{T} . In each step, at a non-leaf node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$, it does:

- (1) if $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$, denoted by c be the smallest pinned constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$,
 - move to $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$ if c is satisfied by $\mathfrak{X}_{\text{vbl}(c)}^{\text{lp}}$;

- otherwise, move to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \mathfrak{x}_{\text{vbl}(c)}^{\text{lp}}, \tau \wedge \rho, B \bowtie c)$ for each $\rho \in \mathcal{Q}_{\text{vbl}(c)}$ w.p.

$$\frac{\hat{P}_{\mathcal{E}, \mathcal{F}, \sigma \wedge \mathfrak{x}_{\text{vbl}(c)}^{\text{lp}}, \tau \wedge \rho, B \bowtie c}^X}{\hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X};$$

(2) otherwise $\mathcal{F}^\tau \subseteq \mathcal{E}^\sigma$, denoted by c be the smallest pinned constraint in $\mathcal{E}^\sigma \setminus \mathcal{F}^\tau$,

- move to $(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)$ with probability

$$\frac{\hat{P}_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}^X}{\hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X};$$

- move to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \mathfrak{x}_{\text{vbl}(c)}^{\text{lp}}, \tau \wedge \rho, B \bowtie c)$ for $\rho = \text{False}(c)$ with probability

$$\frac{\hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \mathfrak{x}_{\text{vbl}(c)}^{\text{lp}}, \tau \wedge \rho, B \bowtie c)}^X}{\hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X}.$$

By Item II. of Definition 4.8, it can be verified that for both types of initialization of \mathfrak{x}^{lp} , the above process generates a probability distribution over root-to-leaf paths of \mathcal{T} . For the two processes with \mathfrak{x}^{lp} initialized as type 1/type 2, we refer to the \mathfrak{x}^{lp} as $\mathfrak{x}^{\text{lp1}}/\mathfrak{x}^{\text{lp2}}$, respectively. We let the root-to-leaf path generated be $P^{\text{lp1}}/P^{\text{lp2}}$, and the distribution on the leaf node induced by the process be $\mu^{\text{lp1}}/\mu^{\text{lp2}}$, respectively.

We then present probability bounds for Definition B.4 with different initialization, concluded by the two following lemmas.

Lemma B.5. Assume Condition 1. For each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, it holds that

$$\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}}] = \mathcal{P} [(\mathcal{E} \setminus (C \setminus \{c_0\})) \wedge \sigma] \cdot \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X.$$

Moreover, conditioning on $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}}$, it follows that

$$\mathfrak{x}^{\text{lp1}} \sim \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma,$$

for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ such that $\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}}] > 0$.

Lemma B.6. Assume Condition 1. For each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, it holds that

$$\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp2}}] = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X.$$

Moreover, conditioning on $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp2}}$, it follows that

$$\mathfrak{x}^{\text{lp2}} \sim \mu_{\mathcal{E}}^\sigma,$$

for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ such that $\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp2}}] > 0$.

By Lemma B.6 and comparing Definition B.4 with Definition 5.3, Lemma 5.5 is immediately proved. Also, we have the following corollary by combining Lemma B.5 with the law of conditional probability.

Corollary B.7. Assume Condition 1. For each $\mathbf{x} \in \mathcal{Q}$ and each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$,

$$\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}} \mid \mathfrak{x}^{\text{lp1}} = \mathbf{x}] = \begin{cases} 0 & \mathbf{x} \notin \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}; \\ \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X & \mathbf{x} \in \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}. \end{cases}$$

Proof. For each $\mathbf{x} \in \mathcal{Q}$ and each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$ we have

$$(23) \quad \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}} \mid \mathfrak{x}^{\text{lp1}} = \mathbf{x}] = \frac{\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}} \wedge \mathfrak{x}^{\text{lp1}} = \mathbf{x}]}{\Pr [\mathfrak{x}^{\text{lp1}} = \mathbf{x}]} \\ \text{(by Definition B.4)} = |\mathcal{Q}| \cdot \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}} \wedge \mathfrak{x}^{\text{lp1}} = \mathbf{x}].$$

Note that by Lemma B.5, if $\mathbf{x} \notin Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}$, then the above term clearly equals to 0. We then assume $\mathbf{x} \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}$. Note that by Lemma B.5, conditioning on the event $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$, the distribution of σ is uniform over all assignments in $Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}$. Therefore,

$$\begin{aligned} \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}} \wedge \mathfrak{X}^{\text{lp1}} = \mathbf{x}] &= \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}}] \cdot \frac{1}{|Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}|} \\ \text{(by Lemma B.5)} &= \frac{|Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}|}{|Q|} \cdot \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \cdot \frac{1}{|Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}|} \\ &= \frac{1}{|Q|} \cdot \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X, \end{aligned}$$

combining with (23) proves the corollary. \square

The proofs of Lemmas B.5 and B.6 follow a direct induction on the coupling tree. We provide the proofs of Lemmas B.5 and B.6 at the end of this subsection for completeness.

We are now ready to prove Claim 4.13.

Proof of Claim 4.13. We only prove the first inequality, and the second inequality follows analogously.

Note that we have

$$\begin{aligned} &\frac{1}{|Q^{C \setminus \{c_0\}}|} \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}} \\ \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\ (24) \quad (\star) &= \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}} \\ \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} \frac{\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp2}}]}{|Q^{\mathcal{E} \wedge \sigma}|} \\ (\blacktriangle) &= \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}} \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp2}}], \end{aligned}$$

where the \star equality is by Lemma B.6 and that $\mathcal{E} \wedge \sigma \implies C \setminus \{c_0\}$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, following the argument in Remark 4.6. The \blacktriangle equality is still by that $\mathcal{E} \wedge \sigma \implies C \setminus \{c_0\}$ and exchanging the order of summation.

Let $(\mathcal{E}^{\text{lp1}}, \mathcal{F}^{\text{lp1}}, \sigma^{\text{lp1}}, \tau^{\text{lp1}}, B^{\text{lp1}}) \sim \mu^{\text{lp1}}$ and $(\mathcal{E}^{\text{lp2}}, \mathcal{F}^{\text{lp2}}, \sigma^{\text{lp2}}, \tau^{\text{lp2}}, B^{\text{lp2}}) \sim \mu^{\text{lp2}}$ denote the random leaf of $V(\mathcal{T})$ generated by the two processes in Definition B.4, respectively.

Let $\mathbb{T}_K^{c_0}$ be the set of 2-trees in G_Φ of size K containing c_0 . Note that following the proof of Lemma 3.5 we have $B \subseteq \mathbb{T}_K^{c_0}$ for all $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}$. We then have

$$\begin{aligned} &\frac{1}{|Q^{C \setminus \{c_0\}}|} \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}} \\ \mathbf{x} \in Q^{\mathcal{E} \wedge \sigma}}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\ \text{(by (24))} &= \sum_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}} \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp2}}] \\ &= \sum_{T \in \mathbb{T}_K^{c_0}} \Pr [B^{\text{lp2}} = T] \\ (25) &= \sum_{T \in \mathbb{T}_K^{c_0}} \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \Pr [\mathfrak{X}^{\text{lp2}} = \mathbf{x}] \cdot \Pr [B^{\text{lp2}} = T \mid \mathfrak{X}^{\text{lp2}} = \mathbf{x}] \\ (\star) &= \sum_{T \in \mathbb{T}_K^{c_0}} \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \Pr [\mathfrak{X}^{\text{lp2}} = \mathbf{x}] \cdot \Pr [B^{\text{lp1}} = T \mid \mathfrak{X}^{\text{lp1}} = \mathbf{x}] \\ \text{(by Corollary B.7)} &= \sum_{T \in \mathbb{T}_K^{c_0}} \sum_{\mathbf{x} \in Q^{C \setminus \{c_0\}}} \Pr [\mathfrak{X}^{\text{lp2}} = \mathbf{x}] \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}} \\ \mathbf{x} \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma \wedge B=T}}} \hat{P}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X. \end{aligned}$$

Here, the \star equality is by noting that we can perfectly couple the random processes in Definition B.4 with two types of initialization when $\mathfrak{X}^{\text{lp1}}$ and $\mathfrak{X}^{\text{lp2}}$ are the same value in $\mathcal{Q}^{C \setminus \{c_0\}}$.

Recall \mathcal{A}_c denotes the event defined in (7), i.e.

$$\mathcal{A}_c : \forall v \in \text{vbl}(c), \quad \mathfrak{X}(v) = \text{False}(c)(v) \vee \mathfrak{Y}(v) = \text{False}(c)(v).$$

We further note that

$$\begin{aligned} & \sum_{T \in \mathbb{T}_K^{c_0}} \sum_{\mathbf{x} \in \mathcal{Q}^{C \setminus \{c_0\}}} \Pr[\mathfrak{X}^{\text{lp2}} = \mathbf{x}] \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}: \\ \mathbf{x} \in \mathcal{Q}^{\mathcal{E} \setminus (C \setminus \{c_0\})} \wedge \sigma \wedge B = T}} \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\ (\star) \quad & \leq (1 - ep)^{(D+1)K} \sum_{T \in \mathbb{T}_K^{c_0}} \sum_{\mathbf{x} \in \mathcal{Q}^{C \setminus \{c_0\}}} \Pr[\mathfrak{X}^{\text{lp2}} = \mathbf{x}] \Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X} = \mathbf{x} \right] \\ (\blacktriangle) \quad & = (1 - ep)^{(D+1)K} \sum_{T \in \mathbb{T}_K^{c_0}} \sum_{\mathbf{x} \in \mathcal{Q}_{\text{vbl}(T)}} \Pr[\mathfrak{X}_{\text{vbl}(T)}^{\text{lp2}} = \mathbf{x}] \cdot \Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X}_{\text{vbl}(T)} = \mathbf{x} \right] \\ (\text{by Definition B.4}) \quad & \leq (1 - ep)^{2(D+1)K} \sum_{T \in \mathbb{T}_K^{c_0}} \sum_{\mathbf{x} \in \mathcal{Q}_{\text{vbl}(T)}} \Pr_{\mathfrak{X} \sim \mathcal{P}}[\mathfrak{X}_{\text{vbl}(T)} = \mathbf{x}] \cdot \Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid \mathfrak{X}_{\text{vbl}(T)} = \mathbf{x} \right] \\ & = (1 - ep)^{2(D+1)K} \sum_{T \in \mathbb{T}_K^{c_0}} \Pr_{\mathfrak{X}, \mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \right] \\ (\blacksquare) \quad & \leq (1 - ep)^{2(D+1)K} \sum_{T \in \mathbb{T}_K^{c_0}} p^{\frac{2}{2+\zeta}} \\ (\text{by Lemma 2.6}) \quad & \leq \left(eD^2 \cdot p^{\frac{2}{2+\zeta}} \cdot (1 - ep)^{-2(D+1)} \right)^K \\ (\text{by Condition 1}) \quad & \leq 2^{-K}, \end{aligned}$$

which combining with (25) finishes the proof of the lemma.

Here, the \star inequality is by Item IV. of Definition 4.8. The \blacktriangle equality is by the definition of \mathcal{A}_c that the expression

$$\Pr_{\mathfrak{Y} \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \right]$$

only depends on $\mathfrak{X}_{\text{vbl}(T)}$. The \blacksquare inequality is by following the argument in the proof of Lemma 3.11. \square

We finish this subsection and the overall section by proving Lemmas B.5 and B.6.

Proof of Lemma B.5. We prove the lemma by a structural induction in a top-down order.

The base case is when at the root node, i.e., $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) = (C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$. Note that in this case we have $\Pr[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{P}^{\text{lp1}}] = 1$ and

$$\mathcal{P}[(\mathcal{E} \setminus (C \setminus \{c_0\})) \wedge \sigma] = 1, \quad \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X = 1,$$

where the second equality is by Item 2 of Definition 4.8. Also, it follows from Definition B.4 that

$$\mathfrak{X}^{\text{lp1}} \sim \mathcal{P} = \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma.$$

The base case is proved.

For the induction step, we only prove the case when $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$. The case when $\mathcal{F}^\tau \subseteq \mathcal{E}^\sigma$ follows analogously. Let c be the smallest constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$. From Definition B.4, we have the following cases:

- c is satisfied by $\mathfrak{X}^{\text{lp1}}$. In this case, we go to $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$. By the induction hypothesis, we have $\mathfrak{X}^{\text{lp1}} \sim \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma$ and therefore this event (c is satisfied by $\mathfrak{X}^{\text{lp1}}$) happens with probability

$\mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma(c)$. Then

$$\begin{aligned}
& \Pr [(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}}] \\
& \text{(by Definition B.4)} = \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}}] \cdot \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma(c) \\
& \text{(by I.H.)} = \mathcal{P} [(\mathcal{E} \setminus (C \setminus \{c_0\})) \wedge \sigma] \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \cdot \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma(c) \\
& \text{(by the chain rule)} = \mathcal{P} [((\mathcal{E} \setminus (C \setminus \{c_0\})) \cup \{c\}) \wedge \sigma] \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\
& \quad (\star) = \mathcal{P} [((\mathcal{E} \cup \{c\}) \setminus (C \setminus \{c_0\})) \wedge \sigma] \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\
& \text{(by Item II. of Definition 4.8)} = \mathcal{P} [((\mathcal{E} \cup \{c\}) \setminus (C \setminus \{c_0\})) \wedge \sigma] \cdot \hat{p}_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}^X.
\end{aligned}$$

Here, the \star equality is by that each $c \in \mathcal{F}^\tau \setminus \mathcal{E}^\sigma$ chosen by the random process in Definition B.1 must satisfy either

- is exactly c_0 ;
- or is pinned by σ , i.e., $\text{vbl}(c) \neq \text{vbl}(c^O)$,

and hence cannot be in $C \setminus \{c_0\}$. Also, conditioning on going to $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$, it follows that

$$\mathfrak{X}^{\text{lp1}} \sim \mu_{\mathcal{E} \setminus (C \setminus \{c_0\}) \cup \{c\}}^\sigma = \mu_{(\mathcal{E} \cup \{c\}) \setminus (C \setminus \{c_0\})}^\sigma,$$

finishing the proof of this case.

- c is violated by $\mathfrak{X}^{\text{lp1}}$. Let $\pi = \mathfrak{X}_{\text{vbl}(c)}^{\text{lp1}}$, then we have $\pi = \text{False}(c)$. By the induction hypothesis, we have $\mathfrak{X}^{\text{lp1}} \sim \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma$ and this event (c is violated by $\mathfrak{X}^{\text{lp1}}$) happens with probability

$$\mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma(\neg c) = \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma(\pi).$$

In this case, for each $\rho \in \mathcal{Q}_{\text{vbl}(c)}$ we go to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)$ with probability $\frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X}$. Hence for each $\rho \in \mathcal{Q}_{\text{vbl}(c)}$,

$$\begin{aligned}
& \Pr [(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in P^{\text{lp1}}] \\
& \quad (\star) = \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp1}}] \cdot \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma(\pi) \cdot \frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X} \\
& \text{(by I.H.)} = \mathcal{P} [(\mathcal{E} \setminus (C \setminus \{c_0\})) \wedge \sigma] \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \cdot \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^\sigma(\pi) \cdot \frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X} \\
& \quad (\blacktriangle) = \mathcal{P} [(\mathcal{E} \setminus (C \setminus \{c_0\})) \wedge (\sigma \wedge \pi)] \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X,
\end{aligned}$$

where the \star equality is by Definition B.4 and the \blacktriangle equality is by the chain rule. Also, conditioning on going to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)$, it follows that

$$\mathfrak{X}^{\text{lp1}} \sim \mu_{\mathcal{E} \setminus (C \setminus \{c_0\})}^{\sigma \wedge \pi},$$

finishing the proof of this case and the lemma. □

Proof of Lemma B.6. We prove the lemma by a structural induction in a top-down order.

The base case is when at the root node, i.e., $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) = (C \setminus \{c_0\}, C, \emptyset, \emptyset, \emptyset)$. Note that in this case we have $\Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp2}}] = 1$ and

$$\mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) = 1, \quad \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X = 1,$$

where the second equality is by Item 2 of Definition 4.8. Also, it follows from Definition B.4 that

$$\mathfrak{X}^{\text{lp2}} \sim \mu_{C \setminus \{c_0\}} = \mu_{\mathcal{E}}^\sigma.$$

The base case is proved.

For the induction step, we only prove the case when $\mathcal{F}^\tau \not\subseteq \mathcal{E}^\sigma$. The case when $\mathcal{F}^\tau \subseteq \mathcal{E}^\sigma$ follows analogously. Let c be the smallest constraint in $\mathcal{F}^\tau \setminus \mathcal{E}^\sigma$. From Definition B.4, we have the following cases:

- c is satisfied by $\mathfrak{X}^{\text{lp}2}$. In this case, we go to $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$. By the induction hypothesis, we have $\mathfrak{X}^{\text{lp}2} \sim \mu_{\mathcal{E}}^\sigma$ and this event (c is satisfied by $\mathfrak{X}^{\text{lp}2}$) happens with probability $\mu_{\mathcal{E}}^\sigma(c)$. Then

$$\begin{aligned} & \Pr [(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp}2}] \\ \text{(by Definition B.4)} &= \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp}2}] \cdot \mu_{\mathcal{E}}^\sigma(c) \\ \text{(by I.H.)} &= \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \cdot \mu_{\mathcal{E}}^\sigma(c) \\ (\star) &= \mu_{C \setminus \{c_0\}}((\mathcal{E} \cup \{c\}) \wedge \sigma) \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \\ \text{(by Item II. of Definition 4.8)} &= \mu_{C \setminus \{c_0\}}((\mathcal{E} \cup \{c\}) \wedge \sigma) \cdot \hat{p}_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}^X, \end{aligned}$$

where the \star equality is by the chain rule and that $\mathcal{E} \wedge \sigma \implies C \setminus \{c_0\}$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, following the argument in Remark 4.6. Also, conditioning on going to $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)$, it follows that

$$\mathfrak{X}^{\text{lp}2} \sim \mu_{\mathcal{E} \cup \{c\}}^\sigma,$$

finishing the proof of this case.

- c is violated by $\mathfrak{X}^{\text{lp}2}$. Let $\pi = \mathfrak{X}_{\text{vbl}(c)}^{\text{lp}2}$. By the induction hypothesis, we have $\mathfrak{X}^{\text{lp}2} \sim \mu_{\mathcal{E}}^\sigma$ and this event (c is violated by $\mathfrak{X}^{\text{lp}2}$) happens with probability

$$\mu_{\mathcal{E}}^\sigma(\neg c) = \mu_{\mathcal{E}}^\sigma(\pi).$$

In this case, for each $\rho \in \mathcal{Q}_{\text{vbl}(c)}$ we go to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)$ with probability $\frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X}$. Hence, for each $\rho \in \mathcal{Q}_{\text{vbl}(c)}$,

$$\begin{aligned} & \Pr [(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in P^{\text{lp}2}] \\ \text{(by Definition B.4)} &= \Pr [(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\text{lp}2}] \cdot \mu_{\mathcal{E}}^\sigma(\pi) \cdot \frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X} \\ \text{(by I.H.)} &= \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \cdot \mu_{\mathcal{E}}^\sigma(\pi) \cdot \frac{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X}{\hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X} \\ (\star) &= \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge (\sigma \wedge \pi)) \cdot \hat{p}_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}^X, \end{aligned}$$

where the \star equality is by the chain rule and that $\mathcal{E} \wedge \sigma \implies C \setminus \{c_0\}$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$, following the argument in Remark 4.6. Also, conditioning on going to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)$, it follows that

$$\mathfrak{X}^{\text{lp}2} \sim \mu_{\mathcal{E}}^{\sigma \wedge \pi},$$

finishing the proof of this case and the lemma. \square