A Sampling Lovász Local Lemma
for Large Domain Sizes

Chunyang Wang
State Key Laboratory for Novel Software Technology
New Cornerstone Science Laboratory
Nanjing University
Nanjing, China
wcysai@smail.nju.edu.cn

Yitong Yin
State Key Laboratory for Novel Software Technology
New Cornerstone Science Laboratory
Nanjing University
Nanjing, China
yinyt@nju.edu.cn

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 [1], [2], thus establishing an almost tight sampling Lovász local lemma for large domain sizes.

Index Terms—sampling, constraint satisfaction problem, Lovász Local Lemma

I. 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* [3], 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:

$$ep(D+1) \le 1, \tag{1}$$

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 [4].

A full version of the paper is available at https://arxiv.org/abs/2307.14872v2.

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 [5]–[11], 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* [1], [2], [12]–[24], 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 [1], [2] 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 [23] 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 [5] 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 $pD^4 \lesssim 1$ [6], [7], [9], 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 [13] 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 [14], [18], [22], [23], [25], and also enabled the development of several Markov chain Monte Carlo approaches [15]–[17], [19], [26], [27] 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 [22], [23], which is similar to the source of slackness in [9].

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.

A. 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 $\Phi = (V, Q, C)$. Here, V is a set of $n = |V|$ variables; each variable $v \in V$ is endowed with a finite domain Q_v, altogether $Q \triangleq \bigotimes_{v \in V} Q_v$; and C is a set of constraints, where each constraint $c \in C$ is a function $c : \bigotimes_{v \in \mathsf{vbl}(c)} Q_v \to \{\mathtt{True}, \mathtt{False}\}$ defined on a subset of variables, denoted by $\mathsf{vbl}(c) \subseteq V$. A CSP formula $\Phi = (V, Q, C)$ is called *atomic* if each constraint $c \in C$ is violated by exactly one assignment in the domain of c, i.e., $|c^{-1}(\mathtt{False})| = 1$. The atomicity of constraints is typical for many classical constraint satisfaction prob-

lems, including k-CNF and hypergraph colorings[1], and is a natural assumption in previous studies of LLL [16], [17], [19], [28]–[35].

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

- *width* $k \triangleq \max_{c \in C} |\mathsf{vbl}(c)|$;
- *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*
  $D \triangleq \max_{c \in C} |\{c' \in C \setminus \{c\} \mid \mathsf{vbl}(c) \cap \mathsf{vbl}(c') \neq \emptyset\}|$;
- *maximum violation probability* $p \triangleq \max_{c \in C} \mathcal{P}[\neg c]$.

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

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

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

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 I.1 (counting LLL). *Assume Condition 1 for* $\Phi = (V, Q, C)$. *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 $O\left(\left(\frac{n}{\varepsilon}\right)^{\mathrm{poly}(k, D, \log q_{\max})}\right)$, *where* Z_Φ *represents the number of solutions to* Φ.

Theorem I.2 (sampling LLL). *Assume Condition 1 for* $\Phi = (V, Q, C)$. *There exists an algorithm that given any* Φ *and any* $\varepsilon \in (0, 1)$, *outputs a random assignment* $X \in Q$ *distributed as* $\hat{\mu}$, *such that*

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

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

Remark I.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

---

[1]For 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.

$pD^2 \gtrsim 1$ lower bound [1], [2]. Conversely, with $q = 2$ (the Boolean domain), ζ achieves its maximum value of $\log_{4/3}(9/4) \approx 2.82$. In this scenario, Condition 1 becomes $pD^{4.82} \lesssim 1$, improving upon the previous best upper bound of $pD^5 \lesssim 1$ for counting/sampling LLL [23].

Remark I.4 (time complexity). The $O\left(\left(\frac{n}{\varepsilon}\right)^{\text{poly}(k,D,\log q)}\right)$ time complexities of Theorems I.1 and I.2 align with previous results for deterministic counting LLL [13], [14], [18], [23], [36]. 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.

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 edges. Hypergraph colorings are foundational combinatorial objects that have been a key to the discovery of the Lovász Local lemma [3]. 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 [37], [38], which is also necessary for the irreducibility of local Markov chains [39]. For *simple* hypergraphs, where any two distinct hyperedges share at most one vertex, it has been shown that local Markov chains rapidly mix when $q \geq \max\{C_k \log n, 500k^3\Delta^{\frac{1}{k-3}}\}$ [39], [40], for hypergraphs with n vertices. The first result for efficiently approximate counting/sampling hypergraph colorings under LLL condition is [14]. 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 [16]–[19], resulting in the current best upper bound $q \geq 310\Delta^{\frac{3}{k-3}}$ for $k \geq 24$ in [17], [19]. On the lower bound side, it was proved in [2] that no efficient algorithm exists for the problem when $q \gtrsim \Delta^{\frac{2}{k}}$ unless **NP=RP**.

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

Corollary I.5 (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 I.5 grows like $q_0 = \exp(O(1/\zeta))$ in ζ.

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 [13] 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 [15], [17]–[19], [23], with the current best bound given in [23] as $k \geq 5 \log d + 5 \log k + O(1)$. The lower bound proved in [1] says that unless **NP=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 I.6 (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 $X \in \{\texttt{True}, \texttt{False}\}^V$ 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.

B. Technique overview

We then outline our general approach for the main results (Theorems I.1 and I.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 [13] and was subsequently generalized and refined in [14], [18], [41].

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.

[2]See the last paragraph of Section 1.2 in [18] for a discussion on the specific barriers of this paradigm.

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 [42]–[56]. For sampling CSP solutions, the correlation decay properties have also been used explicitly or implicitly in various previous algorithmic results [1], [12]–[23], [25]–[27], [36], [41].

Fix the variable set V with domain Q, and for any set of constraints C use μ_C to denote the uniform distribution over solutions to $\Phi = (V, Q, C)$. 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 III.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 $\mathsf{vbl}(c_0)$ to be the unique violating assignment $\pi = \mathtt{False}(c_0)$.

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

$$\mu_C = \sum_{\rho \in Q_{\mathsf{vbl}(c_0)}} \mu_C(\rho) \cdot \mu_C(\cdot \mid \rho),$$

which inspire us to achieve the coupling in this case as: first, sample $\rho \sim \mu_{C, \mathsf{vbl}(c_0)}$ according to the marginal distribution $\mu_{C, \mathsf{vbl}(c_0)}$ induced by μ_C on $\mathsf{vbl}(c_0)$, 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 $\neg c_0$ from the distribution, albeit potentially introducing a discrepancy in the assignment over $\mathsf{vbl}(c_0)$. 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 [43]. 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 [43], 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 [57]. 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.

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 [13], [14], [18], [41], 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 [1], [2], 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 as-

signments. 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 $Y \sim \mu_C$.

II. PRELIMINARIES AND NOTATIONS

A. CSP formulas defined by atomic bad events

1) Basic definitions of 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 $c : \bigotimes_{v \in \mathsf{vbl}(c)} Q_v \to \{\mathtt{True}, \mathtt{False}\}$ defined on a subset of variables, denoted by $\mathsf{vbl}(c) \subseteq V$. An assignment $\sigma \in Q$ is called *satisfying* for Φ if

$$\Phi(\sigma) \triangleq \bigwedge_{c \in C} c\left(\sigma_{\mathsf{vbl}(c)}\right) = \mathtt{True}.$$

In the context of LLL, each constraint c can be interpreted as a bad event A_c, which happens when the assignment on $\mathsf{vbl}(c)$ violates c. For any subset of of constraints $\mathcal{E} \subseteq C$, denote $\mathsf{vbl}(\mathcal{E}) \triangleq \bigcup_{c \in \mathcal{E}} \mathsf{vbl}(c)$. For any subset of variables $\Lambda \subseteq V$, denote $Q_\Lambda \triangleq \bigotimes_{v \in \Lambda} Q_v$. Also, we say Φ is *satisfiable* is 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 $Q_{\mathsf{vbl}(c)}$. For atomic constraints c, we use $\mathtt{False}(c)$ to denote the only violating configuration of c in $Q_{\mathsf{vbl}(c)}$. Moreover, when all constraints in C are atomic, we say Φ is atomic.

2) Notations for (partial) assignments: For a (partial) assignment $\sigma \in Q_\Lambda$ specified over some subset of variables $\Lambda \subseteq V$, we use $\Lambda(\sigma) = \Lambda$ to denote the set of variables σ is assigned on. 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 some $v \in V$.

133

For any two assignments σ, τ such that $\Lambda(\sigma) \cap \Lambda(\tau) = \emptyset$, we define $\sigma \wedge \tau \in Q_{\Lambda(\sigma) \cup \Lambda(\tau)}$ 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 the symbol $\varnothing$ to specifically denote an empty assignment, with the intention to distinguish from $\emptyset$, which is the symbol for an empty set.

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

Definition II.1 (simple notations for events)**.** For the simplicity of notations, we will use:

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

Note that under this definition, the notation $\sigma \wedge \tau$ as a concatention 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 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.

$$\mu_\Phi \triangleq \mathcal{P}\left(\cdot \mid C \right).$$

μ_Φ is well-defined only when Φ is satisfiable.

When the variable set V and the domain 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:

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

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.

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

1) removing all the constraints that have already been satisfied by σ;
2) for the remaining constraints, replacing the variables $v \in \Lambda(\sigma)$ with their values $\sigma(v)$.

Note that Φ^σ is well-defined only when σ does not violate any constraint in C. It is easy to see that when Φ^σ is well-defined, it holds that $\mu_{\Phi^\sigma} = \mu_{V \setminus \Lambda(\sigma)}^\sigma$. Moreover, if Φ is atomic, then Φ^σ is atomic, and each of the remaining constraints $c \in C^\sigma$ pinned from some original constraint $c \in C$ can be uniquely determined by identifying the unassigned subset of variables $\mathsf{vbl}(c') = (V \setminus \Lambda(\sigma)) \cap \mathsf{vbl}(c) \subseteq \mathsf{vbl}(c)$. When Φ^σ is well-defined, we directly use C^σ to denote the set of pinned constraints of C under σ. Also, when σ is clear from the context, for each pinned constraint $c \in C^\sigma$, we use c^O to denote its original constraint in C. Also, we use C^* to denote the set of all possible constraints pinned from some constraint in C.

B. Lovász Local lemma

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

Theorem II.2 ([3])**.** *Given a CSP formula $\Phi = (V, Q, C)$, if there is a function $x : C \to (0, 1)$ such that:*

$$\forall c \in C : \quad \mathcal{P}[\neg c] \leq x(c) \prod_{\substack{c' \in C \setminus \{c\} \\ \mathsf{vbl}(c) \cap \mathsf{vbl}(c') \neq \emptyset}} (1 - x(c')), \quad (3)$$

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 II.3 ([58, Theorem 2.1])**.** *Given a CSP formula $\Phi = (V, Q, C)$, if (3) holds, then for any event $\mathcal{A}$ that is determined by the assignment on a subset of variables $\mathsf{vbl}(\mathcal{A}) \subseteq V$,*

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

C. Dependency graph and 2-trees

The dependency graph is a key notion for the Lovász local lemma.

Definition II.4 (dependency graph)**.** Let $\Phi = (V, Q, C)$ 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 $\mathsf{vbl}(c) \cap \mathsf{vbl}(c') \neq \emptyset$.
- $G_\Phi^2(\mathcal{E})$ 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 $\mathrm{dist}_{G_\Phi}(c, c') \leq 2$.

The notion of 2-tree is an important combinatorial structure introduced by Alon [6], which has played key roles in algorithmic and sampling LLL.

Definition II.5 (2-tree). Let $G = (V, E)$ be a graph and $\mathrm{dist}_G(\cdot, \cdot)$ denote the shortest path distance in G. A 2-*tree* in G is a subset of vertices $T \subseteq V$ such that:
- for any $u, v \in T$, $\mathrm{dist}_G(u, v) \geq 2$;
- T is connected if an edge is added between each $u, v \in T$ such that $\mathrm{dist}_G(u, v) = 2$.

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

Lemma II.6 ([15, 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 $\frac{(\mathrm{e}d^2)^{\ell-1}}{2}$.*

D. 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

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

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 II.7 ([59, Proposition 4.7]). *Let $\mathcal{D} : (X, Y)$ be any coupling of μ and ν, then*

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

III. 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 III.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 $K \geq 1$,*

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

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

Theorem III.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 III.1.

A. A constraint-wise coupling

First, we present the construction of the coupling in Theorem III.1. Fix an instance of atomic CSP formula $\Phi = (V, Q, C)$. The coupling is defined by a recursive procedure $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$, 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 $\mathcal{E}^\sigma$ and $\mathcal{F}^\tau$ 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 $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ 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.

Remark III.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 I-B. The coupling procedure maintains a pair of "sub-instances with boundaries" $(\mathcal{E}, \sigma)$ and $(\mathcal{F}, \tau)$, which initially are $(C \setminus \{c_0\}, \emptyset)$ and $(C, \emptyset)$.

Algorithm 1: 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** $\mathcal{F}^\tau \nsubseteq \mathcal{E}^\sigma$ **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** Couple $(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$;

8   **else**

9   let $\pi = \texttt{False}(c)$ and draw a random $\rho \sim \mu_{\mathcal{F}, \mathsf{vbl}(c)}^\tau$;

10   **return** 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** Couple $(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau)$;

15   **else**

16   draw a random $\pi \sim \mu_{\mathcal{E}, \mathsf{vbl}(c)}^\sigma$ and let $\rho = \texttt{False}(c)$;

17   **return** Couple $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$;

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 $c \in \mathcal{E}^\sigma \triangle \mathcal{F}^\tau$, 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 $(\mathcal{F}, \tau)$.

- 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 $\mathsf{vbl}(c)$ to be $\pi = \texttt{False}(c)$. On the other hand, the joint distribution $\mu_{\mathcal{F}}^\tau$ can be decomposes based on the assignment on $\mathsf{vbl}(c)$ as:

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

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 $\pi = \texttt{False}(c)$ and $\rho \sim \mu_{\mathcal{F}, \mathsf{vbl}(c)}^\tau$.

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 III.3 (soundness of coupling). *Assume Condition 1. For any constraint $c_0 \in C$, the procedure* Couple$(C \setminus \{c_0\}, C, \varnothing, \varnothing)$ *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 Couple$(\mathcal{E}, \mathcal{F}, \sigma, \tau)$:

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

The base case is $(\mathcal{E}, \mathcal{F}, \sigma, \tau) = (C \setminus \{c_0\}, C, \varnothing, \varnothing)$. Then (4) holds by Condition 1 and Theorem II.2.

For the induction step, assume the current call is Couple$(\mathcal{E}, \mathcal{F}, \sigma, \tau)$, then by the induction hypothesis, both $\mathcal{E}^\sigma$ and $\mathcal{F}^\tau$ are well-defined and $\Lambda(\sigma) = \Lambda(\tau)$. We then only prove the case when $\mathcal{F}^\tau \nsubseteq \mathcal{E}^\sigma$. 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 Couple$(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$ on Line 7.

For any direct recursive call Couple$(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ such that $\pi = \texttt{False}(c)$ and $\rho \in Q_{\mathsf{vbl}(c)}$ on Line 10,

136

$\Lambda(\sigma \wedge \pi) = \Lambda(\tau \wedge \rho)$ trivially holds. Also, it must hold that $\mu_{\mathcal{E}}^{\sigma}(\neg c) > 0, \mu_{\mathcal{F}}^{\tau}(\rho) > 0$ by Lines 8-9 of Algorithm 1, meaning π implies $\neg c$. Therefore, we have

$$\mathcal{P}\left[\mathcal{E} \wedge \sigma \wedge \pi\right] = \mathcal{P}\left[\mathcal{E} \wedge \sigma \wedge \neg c\right]$$
$$= \mu_{\mathcal{E}}^{\sigma}(\neg c) \cdot \mathcal{P}\left[\mathcal{E} \wedge \sigma\right] > 0;$$
$$\mathcal{P}\left[\mathcal{F} \wedge \tau \wedge \rho\right] = \mu_{\mathcal{F}}^{\tau}(\rho) \cdot \mathcal{P}\left[\mathcal{F} \wedge \tau\right] > 0,$$

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 $\mathcal{E}^{\sigma} \triangle \mathcal{F}^{\tau}$ 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 $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$ terminates eventually.

At last, we apply a structural induction in the bottom-up order of recursion, to prove that each possible recursive call of $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ 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 $\mu_{\mathcal{E}}^{\sigma}(c)$, $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ returns the output of $\mathsf{Couple}(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$. By the induction hypothesis, the random pair of assignments (X_1, Y_1) returned by $\mathsf{Couple}(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$ 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 $\mu_{\mathcal{E}}^{\sigma}(\neg c)$, $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ returns the output of $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ for $\pi = \mathtt{False}(c)$ and $\rho \sim \mu_{\mathcal{F}, \mathsf{vbl}(c)}^{\tau}$. By the induction hypothesis, the random pair of assignments (X_2, Y_2) returned by $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ at Line 10 follow the marginal distributions:

$$X_2 \sim \mu_{\mathcal{E}}^{\sigma}(\cdot \mid \pi) = \mu_{\mathcal{E}}^{\sigma}(\cdot \mid \neg c),$$
$$Y_2 \sim \sum_{\rho \in Q_{\mathsf{vbl}(c)}} \mu_{\mathcal{F}}^{\tau}(\rho) \cdot \mu_{\mathcal{F}}^{\tau}(\cdot \mid \rho) = \mu_{\mathcal{F}}^{\tau}.$$

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

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

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

B. 2-tree witness for discrepancy

We need to bound the decay of correlation stated in Theorem III.1. Throughout, we fix the input atomic CSP formula $\Phi = (V, Q, C)$ and an arbitrary constraint $c_0 \subseteq C$ assumed in Theorem III.1. Our goal is to bound the discrepancy of the coupling measured in Hamming distance between the random pair (X, Y) produced by $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$. To achieve this goal, we introduce a notion of witness for the discrepancy. Inspired by the 2-tree structure [6], 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 III.4 (witness for discrepancy). Given a run of Algorithm 1 from $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$, the *witness set* $B = B_{c_0} \subseteq C$ is a set of (unpinned) constraints, constructed as follows:

- Initially, $B = \emptyset$.
- Whenever $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ 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:

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

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 $d_{\mathrm{Ham}}(X, Y)$ is always upper bounded by $k \cdot (D + 1) \cdot |B|$ for the output (X, Y) of $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$.

Recall the definitions of dependency graph (Definition II.4) and 2-tree (Definition II.5), we have the following observation.

Lemma III.5. *During the process in Definition III.4, the set B is always a 2-tree in the dependency graph G_Φ.*

Proof. Note that by Definition III.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 $(\mathcal{E}, \mathcal{F}, \sigma, \tau) = (C \setminus \{c_0\}, C, \varnothing, \varnothing)$, 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 $\mathsf{Couple}(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau)$ with B remains unchanged, then the claim trivially holds.
- We recurse to $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ for $\pi = \mathtt{False}(c)$ and some $\rho \in Q_{\mathsf{vbl}(c)}$ 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 $(\mathcal{E}, \mathcal{F}, \sigma, \tau) = (C \setminus \{c_0\}, C, \varnothing, \varnothing)$, since otherwise after the first step c_0 is satisfied in both $\mathcal{E}^\sigma$ and $\mathcal{F}^\tau$, 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 $\mathsf{vbl}(c) \neq \mathsf{vbl}(c^O)$, since this is the only way that c can be added into $\mathcal{E}^\sigma \triangle \mathcal{F}^\tau$. Take any variable $v \in \mathsf{vbl}(c^O) \setminus \mathsf{vbl}(c)$, 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 $\mathsf{vbl}(c') \cap \mathsf{vbl}(c^*) \neq \emptyset$ 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. $\square$

To establish Theorem III.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 III.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 III.6 (random process simulating K-truncated Algorithm 1)**.** Let $X \sim \mu_{C \setminus \{c_0\}}$ and $Y \sim \mu_C$ be drawn independently beforehand. Define the random process $P^{\mathrm{cp}} = P_K^{\mathrm{cp}} = \{(\mathcal{E}_t, \mathcal{F}_t, \sigma_t, \tau_t, B_t)\}_{t \geq 0}$ starting from the initial state $(\mathcal{E}_0, \mathcal{F}_0, \sigma_0, \tau_0, B_0) = (C \setminus \{c_0\}, C, \varnothing, \varnothing, \emptyset)$ 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 $\mathcal{F}_t^{\tau_t} \not\subseteq \mathcal{E}_t^{\sigma_t}$. Let c be the smallest pinned constraint in $\mathcal{F}_t^{\tau_t} \setminus \mathcal{E}_t^{\sigma_t}$. Set $(\mathcal{E}_{t+1}, \mathcal{F}_{t+1}, \sigma_{t+1}, \tau_{t+1}, B_{t+1})$ as $(\mathcal{E}_t \cup \{c\}, \mathcal{F}_t, \sigma_t, \tau_t, B_t)$ if c is satisfied by $X_{\mathsf{vbl}(c)}$ and as $(\mathcal{E}_t, \mathcal{F}_t, \sigma_t \wedge X_{\mathsf{vbl}(c)}, \tau_t \wedge Y_{\mathsf{vbl}(c)}, B_t \bowtie c)$ otherwise.
3) Otherwise $\mathcal{F}_t^\tau \subseteq \mathcal{E}_t^\sigma$. Let c be the smallest pinned constraint in $\mathcal{E}_t^\sigma \setminus \mathcal{F}_t^\tau$. Set $(\mathcal{E}_{t+1}, \mathcal{F}_{t+1}, \sigma_{t+1}, \tau_{t+1}, B_{t+1})$ as $(\mathcal{E}_t, \mathcal{F}_t \cup \{c\}, \sigma_t, \tau_t, B_t)$ if c is satisfied by $Y_{\mathsf{vbl}(c)}$ and as $(\mathcal{E}_t, \mathcal{F}_t, \sigma_t \wedge X_{\mathsf{vbl}(c)}, \tau_t \wedge Y_{\mathsf{vbl}(c)}, B_t \bowtie c)$ otherwise.

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

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

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

Remark III.7. Note that for atomic CSP Φ satisfying Condition 1, due to Theorem II.2 and total probability, the random process P^{cp} and the distribution μ^{cp} in Definition III.6 are well-defined.

Remark III.8 (explicitly identified randomness)**.** In Definition III.6, all randomness used by the process P^{cp} is identified with the two pre-generated random assignments $X \sim \mu_{C \setminus c_0}$ and $Y \sim \mu_C$. 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 III.6.

Lemma III.9. *Assume Condition 1. For the output (X, Y) of* $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$,

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

Proof. Note that by Definition III.4, each time we recursively call $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho)$ 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 $d_{\mathrm{Ham}}(X, Y)$ is upper bounded by the total number of variables assigned in σ/τ during $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$. Therefore, we have

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

We claim that the process in Definition III.6 can be coupled with Algorithm 1 initialized by $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$ so that each time we recursively call $\mathsf{Couple}(\mathcal{E}, \mathcal{F}, \sigma, \tau)$ with witness set B, we will also move to the same tuple $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in Definition III.6. Under the coupling assumed by the claim, it follows that

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

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 III.6, it follows that

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

and prove the strengthened claim using a structural induction in the top-down order. The base case is for the initial call $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$ with witness set $B = \varnothing$. Note that in Definition III.6 we also initially have $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) = (C \setminus \{c_0\}, C, \varnothing, \varnothing, \varnothing)$, and

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

For the induction step, assume we are currently executing $\mathsf{Couple}(C \setminus \{c_0\}, C, \varnothing, \varnothing)$ with witness set B. By the induction hypothesis, we are also at the same tuple $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ in Definition III.6, so that (6) holds. Now it remains to compare Algorithm 1 with Definition III.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 III.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 III.1.

Lemma III.10. *Assume Condition 1. For any integer $K \geq 1$,*

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

Theorem III.1 then follows directly from combining Lemmas III.3, III.9 and III.10. All we need to do now is prove Lemma III.10.

C. Refutation of large witnesses

By the above arguments, it suffices to bound the probability $\mu^{\mathrm{cp}}\left[\mathcal{L}_K^{\mathrm{cp}}\right]$. Consider the random outcome generated by the process in Definition III.6:

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

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 III.11. *Assume Condition 1. For any subset of disjoint constraints $T \subseteq C$,*

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

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

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

To prove the claim, by contradiction we suppose that there exists some $v \in \mathsf{vbl}(c)$ such that both $X(v) \neq \mathtt{False}(c)(v)$ and $Y(v) \neq \mathtt{False}(c)(v)$. Let c' be the pinned constraint chosen by the procedure when c is added into B^{cp} at some step in Definition III.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 X and Y on $\mathsf{vbl}(c) \setminus \Lambda(c')$, 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 III.6, c' will be satisfied by both $X_{\mathsf{vbl}(c')}$ and $Y_{\mathsf{vbl}(c')}$ and will not be added into B^{cp}, a contradiction. The above claim is proved.

Thus, we have

$$\Pr\left[T \subseteq B^{\mathrm{cp}}\right] \leq \Pr\left[\bigwedge_{c \in T} \mathcal{A}_c\right]$$
$$\leq \prod_{c \in T}\left(\prod_{v \in \mathsf{vbl}(c)} \left(\frac{2}{|Q_v|} - \frac{1}{|Q_v|^2}\right) \cdot (1 - ep)^{-2(D+1)}\right). \tag{8}$$

139

Here, the last inequality is by interpreting the probability space for generating $X \sim \mu_C$ and $Y \sim \mu_{C \setminus \{c_0\}}$ 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 II.3 gives to the above last inequality.

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

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

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

$$\frac{\ln \left(2/|Q_v| - 1/|Q_v|^2 \right)}{\ln \left(1/|Q_v| \right)} = 2 - \frac{\ln(2/|Q_v| - 1)}{\ln(1/|Q_v|)}$$

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

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

$$\Pr\left[T \subseteq B^{\mathrm{cp}} \right] \le \prod_{c \in T} \left(p^{\frac{2}{2+\zeta}} \cdot (1 - ep)^{-2(D+1)} \right)$$

$$= p^{\frac{2|T|}{2+\zeta}} \cdot (1 - ep)^{-2(D+1)|T|}. \qquad \square$$

With the help of Lemma III.11, we are ready to prove Lemma III.10, completing the proof of Theorem III.1.

Proof of Lemma III.10. Let $\mathbb{T}_K^{c_0}$ be the set of 2-trees in G_Φ of size K containing c_0. Since $K \ge 1$ we have all $(\mathcal{E}^{\mathrm{cp}}, \mathcal{F}^{\mathrm{cp}}, \sigma^{\mathrm{cp}}, \tau^{\mathrm{cp}}, B^{\mathrm{cp}}) \in \mathcal{V}_K^{\mathrm{cp}}$ satisfy that $c_0 \in B^{\mathrm{cp}}$ by Condition 1 and Definition III.6. Then we have

$$\mu^{\mathrm{cp}}[\mathcal{L}_K^{\mathrm{cp}}] \le \Pr\left[|B^{\mathrm{cp}}| \ge K \right]$$

$$\text{(by Lemma III.5)} \le \sum_{T \in \mathbb{T}_K^{c_0}} \Pr\left[T \subseteq B^{\mathrm{cp}} \right]$$

$$\text{(by Lemma III.11)} \le \sum_{T \in \mathbb{T}_K^{c_0}} p^{\frac{2K}{2+\zeta}} \cdot (1 - ep)^{-2(D+1)K}$$

$$\text{(by Lemma II.6)} \le \frac{(eD^2)^{K-1} p^{\frac{2L}{2+\zeta}}}{2} \cdot (1 - ep)^{-2(D+1)K}$$

$$\le \left(eD^2 p^{\frac{2}{2+\zeta}} \cdot (1 - ep)^{-2(D+1)} \right)^K$$

$$\text{(by Condition 1)} \le 2^{-K}. \qquad \square$$

IV. 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 I.1 and Theorem I.2.

As in previous works for LP-based counting LLL [13], [14], [18], we will set up a linear program to mimic the transition probabilities in the coupling procedure (Definition III.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 III that are sufficient for constructing the LP.

A. Marginal probabilities from coupling procedure

Fix an atomic CSP formula $\Phi = (V, Q, C)$, an arbitrary constraint $c_0 \in C$, and an integer $K \ge 1$. Recall the random process $P^{\mathrm{cp}} = P_K^{\mathrm{cp}} = \{(\mathcal{E}_t, \mathcal{F}_t, \sigma_t, \tau_t, B_t)\}_{t \ge 0}$ and the distribution $\mu^{\mathrm{cp}} = \mu_K^{\mathrm{cp}}$ of its final outcome, both introduced in Definition III.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 IV.1 (one-sided sampler and marginal probabilities). Let X (resp. Y) be drawn as:

- draw $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \sim \mu^{\mathrm{cp}}$;
- 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 $x \in Q^{\mathcal{E} \wedge \sigma} \triangleq \{\pi \in Q \mid \pi \text{ satisfies } \mathcal{E} \wedge \sigma\}$ and $y \in Q^{\mathcal{F} \wedge \tau}$, define:

$$p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X \triangleq \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid X = x \right],$$

$$p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y \triangleq \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid Y = y \right].$$

Note that in above definition, $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^X = \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid X = x \right]$ for *any* $x \in Q^{\mathcal{E} \wedge \sigma}$ (and similarly $p_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}^Y = \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid Y = y \right]$ for any $y \in Q^{\mathcal{F} \wedge \tau}$). 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 IV.2. *Assume Condition 1. Fix any evaluation of $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$. It holds that for all $\boldsymbol{x}, \boldsymbol{x}' \in Q^{\mathcal{E} \wedge \sigma}$:*

$$\Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid X = \boldsymbol{x}\right]$$
$$= \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid X = \boldsymbol{x}'\right],$$

and that for all $\boldsymbol{y}, \boldsymbol{y}' \in Q^{\mathcal{F} \wedge \tau}$:

$$\Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid Y = \boldsymbol{y}\right]$$
$$= \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid Y = \boldsymbol{y}'\right].$$

Moreover, it holds that

$$p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = \mu_C(\mathcal{F} \wedge \tau), \quad p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma). \tag{9}$$

Proof. Following the same routine as the proof of Lemma III.9 (more precisely, following (6)), we can couple the one-sided sampler in Definition IV.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 $\mathsf{X} \sim \mu_{C \setminus \{c_0\}}$ (resp. $\mathsf{Y} \sim \mu_C$) used in the construction of P^{cp} in Definition III.6.

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

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

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

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

and the lemma follows from the law of conditional probability.

We then prove (10). Note that the first case is immediate by Definition IV.1.

Following the proof of Lemma III.9, we have that for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\mathrm{cp}}$,

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

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

$$\Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \wedge X = \boldsymbol{x}\right]$$
$$(\star) \quad = \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}}\right] \cdot \mu_{\mathcal{E}}^{\sigma}(\boldsymbol{x})$$
$$(\text{by } (11)) \quad = \mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma) \cdot \mu_C(\mathcal{F} \wedge \tau) \cdot \mu_{\mathcal{E}}^{\sigma}(\boldsymbol{x})$$
$$(\blacktriangle) \quad = \mu_{C \setminus \{c_0\}}(\boldsymbol{x}) \cdot \mu_C(\mathcal{F} \wedge \tau),$$

finishing the proof of (10). Here, the $\star$ equality is by Definition IV.1, the $\blacktriangle$ equality is by the chain rule and

additionally by that $\mathcal{E} \wedge \sigma$ implies $C \setminus \{c_0\}$ for each $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\mathrm{cp}}$, an argument easily proven through an induction on Definition III.6. $\qquad \square$

The following lists some basic properties of $p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ and $p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$.

Proposition IV.3. *Assume Condition 1. The followings hold for the $p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ and $p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$:*

1) $p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}, p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} \in [0, 1]$. In particular, $p^X_{(C \setminus \{c_0\}, C, \varnothing, \varnothing, \emptyset)} = p^Y_{(C \setminus \{c_0\}, C, \varnothing, \varnothing, \emptyset)} = 1$.

2) For any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{V}^{\mathrm{cp}} \setminus \mathcal{L}^{\mathrm{cp}}$, where $\mathcal{V}^{\mathrm{cp}}$ and $\mathcal{L}^{\mathrm{cp}}$ are defined in Definition III.6,

a) if $\mathcal{F}^{\tau} \nsubseteq \mathcal{E}^{\sigma}$, letting $c \in \mathcal{F}^{\tau} \setminus \mathcal{E}^{\sigma}$ be the smallest and $\pi = \texttt{False}(c)$,

$$p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = p^X_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}$$
$$= \sum_{\substack{\rho \in Q_{\mathsf{vbl}(c)} \\ (\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\mathrm{cp}}}} p^X_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)};$$

$$p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = p^Y_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)} + p^Y_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)},$$
for all $\rho \in Q_{\mathsf{vbl}(c)}$ and
all $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\mathrm{cp}}$.

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

$$p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = p^X_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)} + p^X_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)},$$
for all $\pi \in Q_{\mathsf{vbl}(c)}$ and
all $(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\mathrm{cp}}$;

$$p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = p^Y_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}$$
$$= \sum_{\substack{\pi \in Q_{\mathsf{vbl}(c)} \\ (\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c) \in \mathcal{V}^{\mathrm{cp}}}} p^Y_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}.$$

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

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

These properties follow directly from Definition IV.1 or from Proposition IV.2, and are easy to verify.

The following bounds for $p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ and $p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ with 2-trees follow from a similar argument as in our analysis of the correlation decay property (e.g. the proof of Lemma III.11).

Proposition IV.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*

141

$\mathcal{P}$ is the product distribution. For each $c \in C$, let $\mathcal{A}_c$ be the event defined as (7), that is,

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

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

$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\mathrm{cp}}: \\ B = T \wedge \boldsymbol{x} \in Q^{\mathcal{E} \wedge \sigma}}} p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$$

$$\leq (1 - ep)^{(D+1)K} \Pr_{Y \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid X = \boldsymbol{x} \right],$$

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

$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\mathrm{cp}}: \\ B = T \wedge \boldsymbol{y} \in Q^{\mathcal{F} \wedge \tau}}} p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$$

$$\leq (1 - ep)^{(D+1)K} \Pr_{X \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid Y = \boldsymbol{y} \right].$$

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

Let $(\mathcal{E}^{\mathrm{cp}}, \mathcal{F}^{\mathrm{cp}}, \sigma^{\mathrm{cp}}, \tau^{\mathrm{cp}}, B^{\mathrm{cp}}) \sim \mu^{\mathrm{cp}}$ denote the random outcome of the process P^{cp} constructed in Definition III.6. Following the same routine as the proof of Lemma III.9 (more precisely, following (6)), we can couple the one-sided sampler in Definition IV.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 $X \sim \mu_{C \setminus \{c_0\}}$ (resp. $Y \sim \mu_C$) used in the construction of P^{cp} in Definition III.6. Fix any $\boldsymbol{x} \in Q^{C \setminus \{c_0\}}$. Then we have

$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\mathrm{cp}}: \\ B = T \wedge \boldsymbol{x} \in Q^{\mathcal{E} \wedge \sigma}}} p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$$

$$= \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}^{\mathrm{cp}}: \\ B = T \wedge \boldsymbol{x} \in Q^{\mathcal{E} \wedge \sigma}}} \Pr\left[(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in P^{\mathrm{cp}} \mid X = \boldsymbol{x} \right]$$

$$= \Pr\left[B^{\mathrm{cp}} = T \mid X = \boldsymbol{x} \right]$$

$$(\star) \quad = \Pr_{\substack{X \sim \mu_{C \setminus \{c_0\}} \\ Y \sim \mu_C}} \left[B^{\mathrm{cp}} = T \mid X = \boldsymbol{x} \right]$$

$$(\blacktriangle) \quad \leq \Pr_{\substack{X \sim \mu_{C \setminus \{c_0\}} \\ Y \sim \mu_C}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid X = \boldsymbol{x} \right]$$

$$(\blacksquare) \quad \leq (1 - ep)^{(D+1)K} \Pr_{Y \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid X = \boldsymbol{x} \right].$$

Here, the $\star$ equality is by the coupling between X and X, the $\blacktriangle$ inequality follows the same argument in the proof of Lemma III.11 that $B^{\mathrm{cp}} = T$ implies the event $\bigwedge_{c \in T} \mathcal{A}_c$, and the $\blacksquare$ inequality is by Theorem II.3. $\square$

Proposition IV.4 utilizes the independence between the random assignments X, Y, which are generated as the random source used by the process P^{cp} constructed in Definition III.6 for simulating Algorithm 1. Such explicit identification of randomness helped in proving Lemma III.11 and also will play a key role in constructing the linear program that will be introduced later.

B. Setting up the linear program

Next, we set up the linear program that mimicries the marginal probabilities $p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ and $p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ introduced in Definition IV.1, in order to bootstrap the ratio between the numbers of assignments satisfying C and those satisfying $C \setminus \{c_0\}$.

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

Definition IV.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 III.6. The tree $\mathcal{T}$ is inductively constructed as follows:

1) The root of $\mathcal{T}$ corresponds to $(C \setminus \{c_0\}, C, \varnothing, \varnothing, \emptyset)$, and has depth 0.
2) For $i = 0, 1, \ldots$: 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 $\mathcal{F}^\tau \nsubseteq \mathcal{E}^\sigma$, 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 $\pi = \mathtt{False}(c)$ and each $\rho \in Q_{\mathsf{vbl}(c)}$, 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 $\pi \in Q_{\mathsf{vbl}(c)}$ and $\rho = \mathtt{False}(c)$, 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:
- $\mathcal{L}_{\mathrm{coup}} \triangleq \{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L} \mid \mathcal{E}^\sigma = \mathcal{F}^\tau\}$ as the set of "coupled" leaf nodes in $\mathcal{T}$;
- $\mathcal{L}_{\mathrm{trun}} \triangleq \{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L} \mid |B| = K\}$ 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 IV.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.:

$$(\mathcal{E} \cup \{c\}) \wedge \sigma \wedge \mathcal{F} \wedge \tau,$$

and

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

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 IV.5 is also well defined.

Remark IV.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 III.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 III.5, while others need to be reproved in this stronger setting.

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 IV.5, where each tree node $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)$ is associated with a pair of variables, corresponding to $p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ and $p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$. The constraints of the LP are derived from the properties listed in Propositions IV.3 and IV.4.

Definition IV.8 (linear program induced by the coupling). Let $\mathcal{T} = \mathcal{T}_K(\Phi, c_0)$ be the K-truncated coupling tree. Let $0 \le r_- \le r_+$ be two parameters. The following linear program (for checking feasibility) is defined on the variables $\hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ and $\hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ for all $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T})$:

I. *Range constraints*:
$$\hat{p}^X_{(C \setminus \{c_0\}, C, \varnothing, \varnothing, \emptyset)} = \hat{p}^Y_{(C \setminus \{c_0\}, C, \varnothing, \varnothing, \emptyset)} = 1;$$
$$\hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}, \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} \in [0, 1],$$
$$\forall (\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in V(\mathcal{T}).$$

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 $\mathcal{F}^{\tau} \nsubseteq \mathcal{E}^{\sigma}$, letting c be the smallest constraint in $\mathcal{F}^{\tau} \setminus \mathcal{E}^{\sigma}$ and $\pi = \texttt{False}(c)$,
$$\hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = \hat{p}^X_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)}$$
$$= \sum_{\rho \in Q_{\text{vbl}(c)}} \hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)};$$
$$\hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = \hat{p}^Y_{(\mathcal{E} \cup \{c\}, \mathcal{F}, \sigma, \tau, B)} + \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)},$$
$$\forall \rho \in Q_{\text{vbl}(c)}.$$

b) otherwise $\mathcal{F}^{\tau} \subseteq \mathcal{E}^{\sigma}$, letting c be the smallest constraint in $\mathcal{E}^{\sigma} \setminus \mathcal{F}^{\tau}$ and $\rho = \texttt{False}(c)$,
$$\hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = \hat{p}^X_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)} + \hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)},$$
$$\forall \pi \in Q_{\text{vbl}(c)};$$
$$\hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = \hat{p}^Y_{(\mathcal{E}, \mathcal{F} \cup \{c\}, \sigma, \tau, B)}$$
$$= \sum_{\pi \in Q_{\text{vbl}(c)}} \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma \wedge \pi, \tau \wedge \rho, B \bowtie c)}.$$

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

a) if it is a coupled leaf $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}$,
$$r_- \cdot \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} \le \hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} \le r_+ \cdot \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}.$$

b) if it is an invalid leaf $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{invld}}$,
$$\hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = 0, \text{ if } \sigma \text{ violates } \mathcal{E};$$
$$\hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = 0, \text{ if } \tau \text{ violates } \mathcal{F}.$$

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.
$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{trun}}: \\ B = T \wedge x \in Q^{\mathcal{E} \setminus (C \setminus \{c_0\}) \wedge \sigma}}} \hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$$
$$\le (1 - ep)^{(D+1)K} \Pr_{Y \sim \mathcal{P}} \left[\bigwedge_{c \in T} \mathcal{A}_c \mid X = x \right],$$
$$\forall T \in \mathbb{T}_K^{c_0}, x \in Q;$$

$$\sum_{\substack{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{trun}}:\\ B=T\wedge\boldsymbol{y}\in Q^{\mathcal{F}\setminus C\wedge\tau}}} \hat{p}^Y_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)}$$

$$\leq (1-\mathrm{e}p)^{(D+1)K}\Pr_{X\sim\mathcal{P}}\left[\bigwedge_{c\in T}\mathcal{A}_c\mid Y=\boldsymbol{y}\right],$$

$$\forall T\in\mathbb{T}^{c_0}_K, \boldsymbol{y}\in\mathcal{Q},$$

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

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

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 IV.8, can be calculated explicitly as:

$$\Pr_{Y\sim\mathcal{P}}\left[\bigwedge_{c\in T}\mathcal{A}_c\mid X=\boldsymbol{x}\right]=\prod_{c\in T}\prod_{v\in\mathsf{vbl}(c)}|Q_v|^{-I[\boldsymbol{x}_v\neq\mathtt{False}(c)_v]},$$

$$\Pr_{X\sim\mathcal{P}}\left[\bigwedge_{c\in T}\mathcal{A}_c\mid Y=\boldsymbol{y}\right]=\prod_{c\in T}\prod_{v\in\mathsf{vbl}(c)}|Q_v|^{-I[\boldsymbol{y}_v\neq\mathtt{False}(c)_v]}.$$

The overflow constraints (Item IV.) of Definition IV.8 are derived from Proposition IV.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 [13], [18], [41]. Note, however, that the overflow constraints slightly differ from Proposition IV.4, as the domain of $\boldsymbol{x}/\boldsymbol{y}$ 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 IV.8, we will address this concern later (in Proposition IV.9). Specifically, we will show that there are indeed $\exp(K\cdot\mathrm{poly}(k,\log q_{\max},D))$ distinct overflow constraints, and moreover, it takes approximately that amount of time to both write down the linear program and verify its feasibility.

C. 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 IV.9. *For any $0\leq r_-\leq r_+$, the feasibility of the linear program in Definition IV.8 can be checked within $\exp(K\cdot\mathrm{poly}(k,\log q_{\max},D))$ time.*

Proof of Proposition IV.9. To prove Proposition IV.9, we need to first prove the following lemma, which bounds the depth of the coupling tree $\mathcal{T}$.

Lemma IV.10. $\mathcal{T}$ *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 $\mathcal{E}^\sigma\triangle\mathcal{F}^\tau$ 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 decreases by one; or we assign some values to σ and τ on $\mathsf{vbl}(c)$, and let t increases by at most $D-1$ (note that at most D new elements can be added into $\mathcal{E}^\sigma\triangle\mathcal{F}^\tau$ and that c is removed from $\mathcal{E}^\sigma\triangle\mathcal{F}^\tau$). 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 IV.5 and (5), we have $|B'|>K$, contradicting the truncation condition at Item 2a of Definition IV.5. Therefore, the lemma is proved. $\square$

Note that the branching number of the coupling tree is at most $(q_{\max})^{2k}$. Hence by Lemma IV.10, the size of the coupling tree is bounded by $\exp(K\cdot\mathrm{poly}(k,\log q_{\max},D))$.

Therefore, the total number of constraints of Items I. through III. in Definition IV.8 is also bounded by $\exp(K\cdot\mathrm{poly}(k,\log q,D))$. For Item IV. of Definition IV.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 $T\in\mathbb{T}^{c_0}_K$, for all leaf nodes $(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{trun}}$ satisfying $B=T$, the following properties hold:

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

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

144

assumption, and thus $(c^*)^O$ must be added into B by Definition IV.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 IV.8 lies in $\mathcal{E}^\sigma \triangle \mathcal{F}^\tau$ at the time, and c can added into $\mathcal{E}^\sigma \triangle \mathcal{F}^\tau$ only by either

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

By the above properties, the feasibility of the first equation can be directly verified by enumerating the assignment of $\boldsymbol{x}$ 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 $\boldsymbol{x}$ to verify the first inequality in Item IV. of Definition IV.8. Combining with the bound on the number of 2-trees in Lemma II.6, we have the number of constraints in Item IV. of Definition IV.8 can be bounded by $\exp(K \cdot \mathrm{poly}(k, \log q_{\max}, D))$.

Hence, we have both the number of variables and constraints of the linear program defined in Definition IV.8 is at most $\exp(K \cdot \mathrm{poly}(k, \log q_{\max}, D))$. Then the proposition follows from the standard running time guarantees for linear programming (e.g., [60]).　□

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

Lemma IV.11. *Assume Condition 1. The LP in Definition IV.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

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

and show that this is a feasible solution for $r_- = r_+ = \frac{|Q^{C \setminus \{c_0\}}|}{|Q^C|}$. We remark that following Lemma III.3, this choice of the solution in (12) coincides with the real values of $p^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}, p^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$ defined in Definition IV.1, but extends to all nodes in $V(\mathcal{T})$ instead of just $\mathcal{V}^{\mathrm{cp}}$.

Item I. directly follows from Item 1 of Proposition IV.3. Also, Item II. directly follows from Item 2 of Proposition IV.3.

Item III.b directly follows from (12). To prove Item III.a , note that for $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\mathrm{coup}}$, $\mathcal{E}^\sigma = \mathcal{F}^\tau$ 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 IV.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 IV.4. However, since now the domain for $\boldsymbol{x}$ and $\boldsymbol{y}$ has changed, we need to define

another procedure similar to Definition III.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 the full version of this paper [61, Section B.1].　□

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 IV.12. *Assume Condition 1. If the LP in Definition IV.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 IV.13.

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

$$\sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\mathrm{valid}}: \\ \boldsymbol{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} = 1, \quad \text{for all } \boldsymbol{y} \in Q^C.$$

We only prove the first equality of Claim IV.13, and then the second equality follows analogously. By Item II. of Definition IV.8 and an induction on the increasing depth of $\mathcal{T}$, we have

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

Note that for any $\boldsymbol{x} \in Q^{C \setminus \{c_0\}}$, any $(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\mathrm{invld}}$ such that $\boldsymbol{x} \in Q^{\mathcal{E} \wedge \sigma}$, it must hold that τ violates $\mathcal{F}$. Then, Claim IV.13 holds immediately by applying Item III.b of Definition IV.8.

By summing the equations in Claim IV.13 over all $c \in Q^{C \setminus \{c_0\}}$ and respectively all $c \in Q^C$, we obtain:

$$|Q^{C \setminus \{c_0\}}| = \sum_{\boldsymbol{x} \in Q^{C \setminus \{c_0\}}} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\mathrm{valid}}: \\ \boldsymbol{x} \in Q^{\mathcal{E} \wedge \sigma}}} \hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)},$$

$$|Q^C| = \sum_{\boldsymbol{y} \in Q^C} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\mathrm{valid}}: \\ \boldsymbol{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}.$$
(13)

Claim IV.14. *Assume Condition 1. The followings hold.*

$$\frac{1}{|Q^{C\setminus\{c_0\}}|} \sum_{\boldsymbol{x}\in Q^{C\setminus\{c_0\}}} \sum_{\substack{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{trun}}:\\ \boldsymbol{x}\in Q^{\mathcal{E}\wedge\sigma}}} \hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)} \le 2^{-K},$$

$$\frac{1}{|Q^C|} \sum_{\boldsymbol{y}\in Q^C} \sum_{\substack{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{trun}}:\\ \boldsymbol{y}\in Q^{\mathcal{F}\wedge\tau}}} \hat{p}^Y_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)} \le 2^{-K}.$$

Claim IV.14 can be proved by a similar argument to the analysis of correlation decay in Section III. The proof deferred to the full version of this paper [61, Section B.2].

We then have

$$
\begin{aligned}
& \quad |Q^{C\setminus\{c_0\}}| \\
\text{(by (13))} \quad &= \sum_{\boldsymbol{x}\in Q^{C\setminus\{c_0\}}} \sum_{\substack{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{valid}}:\\ \boldsymbol{x}\in Q^{\mathcal{E}\wedge\sigma}}} \hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)} \\
&= \sum_{\boldsymbol{x}\in Q^{C\setminus\{c_0\}}} \sum_{\substack{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{coup}}:\\ \boldsymbol{x}\in Q^{\mathcal{E}\wedge\sigma}}} \hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)} \\
& \quad + \sum_{\boldsymbol{x}\in Q^{C\setminus\{c_0\}}} \sum_{\substack{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{trun}}:\\ \boldsymbol{x}\in Q^{\mathcal{E}\wedge\sigma}}} \hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)} \\
(\star) \quad &\le \sum_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{coup}}} |Q^{\mathcal{E}\wedge\sigma}| \cdot \hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)} \\
& \quad + 2^{-K} \cdot |Q^{C\setminus\{c_0\}}|,
\end{aligned}
$$

where the $\star$ inequality is by Claim IV.14. Therefore, we have

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

Therefore,

$$
\begin{aligned}
& \frac{|Q^{C\setminus\{c_0\}}|}{|Q^C|} \\
&\le \left(1+2\cdot 2^{-K}\right) \frac{\sum\limits_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{coup}}} |Q^{\mathcal{E}\wedge\sigma}| \cdot \hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)}}{\sum\limits_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{coup}}} |Q^{\mathcal{F}\wedge\tau}| \cdot \hat{p}^Y_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)}} \\
&\le \left(1+2\cdot 2^{-K}\right) r_+,
\end{aligned}
$$

where the last inequality is due to the leaf constraints of the LP (Item III. of Definition IV.8). Symmetrically, we have

$$
\begin{aligned}
& \frac{|Q^{C\setminus\{c_0\}}|}{|Q^C|} \\
&\ge \left(1-2\cdot 2^{-K}\right) \frac{\sum\limits_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{coup}}} |Q^{\mathcal{E}\wedge\sigma}| \cdot \hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)}}{\sum\limits_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)\in\mathcal{L}_{\mathrm{coup}}} |Q^{\mathcal{F}\wedge\tau}| \cdot \hat{p}^Y_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)}} \\
&\ge \left(1-2\cdot 2^{-K}\right) r_-.
\end{aligned}
$$

Together, we have

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

V. Counting and sampling via linear programming

In this section, we utilize the linear program introduced in Section IV to prove Theorems I.1 and I.2.

A. Constraint-wise self-reducibility in the local lemma regime

To obtain the algorithms for approximate counting and sampling required in Theorems I.1 and I.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 V.1 (marginal estimator). *There exists a deterministic algorithm such that for any $\Phi = (V, Q, C)$ satisfying Condition 1, any $c_0 \in C$ and $\varepsilon \in (0,1)$, within time $O\left(\left(\frac{1}{\varepsilon}\right)^{\mathrm{poly}(k,D,\log q_{\max})}\right)$ the algorithm outputs an estimate $\hat{r}$ satisfying*

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

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

Theorem V.2 (dynamic sampler). *There exists a randomized algorithm such that for any $\Phi = (V, Q, C)$ 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 $\Phi^{c_0} = (V, Q, C \setminus \{c_0\})$, 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 $\tau \in Q$ satisfying*

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

Through a constraint-wise self-reduction that maintains the local lemma regime, Theorem I.1 is implied by Theorem V.1 and Theorem I.2 is implied by Theorem V.2.

Proof of Theorem I.1. Let $C = \{c_1, c_2, \ldots, c_m\}$. Note that we have $m \leq n(D+1)$. For each $0 \leq i \leq m$, we let $C_i = \{c_1, c_2, \ldots, c_i\}$, $\Phi_i = (V, Q, C_i)$ 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:

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

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 V.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 I.2. Let $C = \{c_1, c_2, \ldots, c_m\}$. Note that we have $m \leq n(D+1)$. For each $0 \leq i \leq m$, we let $C_i = \{c_1, c_2, \ldots, c_i\}$ and $\Phi_i = (V, Q, C_i)$. We build a sequence of assignments $\sigma_0, \sigma_1, \ldots, \sigma_m \in Q$ as follows:

- Sample $\sigma_0 \sim \mathcal{P}$.
- For each $1 \leq i \leq m$, use the dynamic sampler in Theorem V.2 to obtain σ_i from σ_{i-1} with input formula (V, 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 V.2 and induction, it easily follows from maximally coupling σ_i with a uniform solution of Φ_i and applying the coupling lemma (Lemma II.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. $\square$

B. Construction of the marginal estimator/sampler

It remains to prove Theorems V.1 and V.2. Theorem V.1 is proved by directly applying the linear program constructed and analyzed in Section IV.

Proof of Theorem V.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 IV.12. The running time bound follows from Proposition IV.9. $\square$

We then prove Theorem V.2. The dynamic sampler claimed in Theorem V.2 is constructed as follows.

Definition V.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 IV.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 IV.12, 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 $\left\{\hat{p}^X_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)}, \hat{p}^Y_{(\mathcal{E},\mathcal{F},\sigma,\tau,B)}\right\}_{(\mathcal{E},\mathcal{F},\sigma,\tau,B) \in V(\mathcal{T})}$ 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 $(C \setminus \{c_0\}, C, \varnothing, \varnothing, \emptyset)$. 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 $\sigma^{\text{in}}_{\text{vbl}(c)}$;
   - otherwise, move to $(\mathcal{E}, \mathcal{F}, \sigma \wedge \sigma^{\text{in}}_{\text{vbl}(c)}, \tau \wedge \rho, B \bowtie c)$ for each $\rho \in Q_{\text{vbl}(c)}$ w.p.

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

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

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

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

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

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 $\sigma^{\text{out}} \leftarrow \tau \wedge \sigma^{\text{in}}_{V \setminus \Lambda(\tau)}$.

We then finish the section by proving the following lemma, which directly implies Theorem V.2.

Lemma V.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 V.3 satisfies*

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

Proof. We need the following lemma, whose proof is by a similar argument to the analysis of correlation decay in Section III and deferred to the full version of this paper [61, Section B.2].

Lemma V.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*

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

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

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

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

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

$$\Pr\left[\sigma^{\text{out}} = \boldsymbol{y}\right]$$

$$(\star) \quad = \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \boldsymbol{y} \in Q^{\mathcal{F} \wedge \tau}}} \frac{\mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma)}{|Q^{\mathcal{E} \wedge \sigma}|} \cdot \hat{p}^X_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$$

$$(\blacktriangle) \quad \ge r_- \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \boldsymbol{y} \in Q^{\mathcal{F} \wedge \tau}}} \frac{\mu_{C \setminus \{c_0\}}(\mathcal{E} \wedge \sigma)}{|Q^{\mathcal{E} \wedge \sigma}|} \cdot \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$$

$$(\blacksquare) \quad = r_- \cdot \frac{1}{|Q^{C \setminus \{c_0\}}|} \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\text{coup}}: \\ \boldsymbol{y} \in Q^{\mathcal{F} \wedge \tau}}} \hat{p}^Y_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}.$$

Here, the $\star$ 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 V.5. The $\blacktriangle$ inequality is by Item III.a of Definition IV.8. The $\blacksquare$ 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 IV.6.

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

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

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

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

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

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

Combining (16) and (17), we obtain

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

$$\le \frac{\varepsilon}{2}. \quad (18)$$

148

Now that by (15) and Claim IV.13, for each $\boldsymbol{y} \in \boldsymbol{Q}^C$ we have $v[\boldsymbol{y}] \le \frac{1}{|\boldsymbol{Q}^C|} = \mu_C(\boldsymbol{y})$. Hence,

$$d_{\mathrm{TV}}(\mu_C, v)$$

$$= \sum_{\boldsymbol{y} \in \boldsymbol{Q}^C} \left(\frac{1}{|\boldsymbol{Q}^C|} - \frac{1}{|\boldsymbol{Q}^C|} \cdot \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\mathrm{coup}}: \\ \boldsymbol{y} \in \boldsymbol{Q}^{\mathcal{F} \wedge \tau}}} \hat{p}^{Y}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)} \right)$$

$$(\star) = \frac{1}{|\boldsymbol{Q}^C|} \sum_{\boldsymbol{y} \in \boldsymbol{Q}^C} \sum_{\substack{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B) \in \mathcal{L}_{\mathrm{trun}}: \\ \boldsymbol{y} \in \boldsymbol{Q}^{\mathcal{F} \wedge \tau}}} \hat{p}^{Y}_{(\mathcal{E}, \mathcal{F}, \sigma, \tau, B)}$$

$$(\blacktriangle) \le \frac{\varepsilon}{2}.$$

Here, the $\star$ equality is by Claim IV.13 and the $\blacktriangle$ inequality is by Claim IV.14. Combining with (18) and using the triangle inequality for the total variation distance, we obtain

$$d_{\mathrm{TV}}(\mu_C, \sigma^{\mathrm{out}}) \le d_{\mathrm{TV}}(\mu_C, v) + d_{\mathrm{TV}}(v, \sigma^{\mathrm{out}}) = \varepsilon,$$

finishing the proof of the lemma. □

VI. 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 [1], [2]. Even for the worst-case of Boolean domains, our regime $pD^{4.82} \lesssim 1$ still improves over the previous best regime of $pD^5 \lesssim 1$ [23].

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

[1] I. Bezáková, A. Galanis, L. A. Goldberg, H. Guo, and D. Štefankovič, "Approximation via correlation decay when strong spatial mixing fails," *SIAM J. Comput.*, vol. 48, no. 2, pp. 279–349, 2019.

[2] A. Galanis, H. Guo, and J. Wang, "Inapproximability of counting hypergraph colourings," *ACM Trans. Comput. Theory*, 2022.

[3] P. Erdős and L. Lovász, "Problems and results on 3-chromatic hypergraphs and some related questions," *Infinite and finite sets, volume 10 of Colloquia Mathematica Societatis János Bolyai*, pp. 609–628, 1975.

[4] J. B. Shearer, "On a problem of Spencer," *Combinatorica*, vol. 5, no. 3, pp. 241–245, 1985.

[5] J. Beck, "An algorithmic approach to the Lovász local lemma," *Random Struct. Algorithms*, vol. 2, no. 4, pp. 343–365, 1991.

[6] N. Alon, "A parallel algorithmic version of the local lemma," in *FOCS*, pp. 586–593, IEEE, 1991.

[7] M. Molloy and B. Reed, "Further algorithmic aspects of the local lemma," in *STOC '98 (Dallas, TX)*, pp. 524–529, ACM, 1999.

[8] A. Czumaj and C. Scheideler, "Coloring nonuniform hypergraphs: a new algorithmic approach to the general Lovász local lemma," *Random Struct. Algorithms*, vol. 17, no. 3-4, pp. 213–237, 2000.

[9] A. Srinivasan, "Improved algorithmic versions of the Lovász local lemma," in *SODA*, pp. 611–620, SIAM, 2008.

[10] R. A. Moser, "A constructive proof of the Lovász local lemma," in *STOC*, pp. 343–350, ACM, 2009.

[11] R. A. Moser and G. Tardos, "A constructive proof of the general Lovász local lemma," *J. ACM*, vol. 57, no. 2, p. 11, 2010.

[12] J. Hermon, A. Sly, and Y. Zhang, "Rapid mixing of hypergraph independent sets," *Random Struct. Algorithms*, vol. 54, no. 4, pp. 730–767, 2019.

[13] A. Moitra, "Approximate counting, the Lovász local lemma, and inference in graphical models," *J. ACM*, vol. 66, no. 2, pp. 10:1–10:25, 2019. (Conference version in *STOC'17*).

[14] H. Guo, C. Liao, P. Lu, and C. Zhang, "Counting hypergraph colorings in the local lemma regime," *SIAM J. Comput.*, vol. 48, no. 4, pp. 1397–1424, 2019.

[15] W. Feng, H. Guo, Y. Yin, and C. Zhang, "Fast sampling and counting k-SAT solutions in the local lemma regime," *J. ACM*, vol. 68, no. 6, pp. Art. 40, 42, 2021.

[16] W. Feng, K. He, and Y. Yin, "Sampling constraint satisfaction solutions in the local lemma regime," in *STOC*, pp. 1565–1578, ACM, 2021.

[17] V. Jain, H. T. Pham, and T. D. Vuong, "On the sampling Lovász local lemma for atomic constraint satisfaction problems," *arXiv*, vol. abs/2102.08342, 2021.

[18] V. Jain, H. T. Pham, and T. D. Vuong, "Towards the sampling lovász local lemma," in *FOCS*, pp. 173–183, IEEE, 2021.

[19] K. He, X. Sun, and K. Wu, "Perfect sampling for (atomic) Lovász local lemma," *arXiv*, vol. abs/2107.03932, 2021.

[20] G. Qiu, Y. Wang, and C. Zhang, "A perfect sampler for hypergraph independent sets," in *ICALP*, pp. 103:1–103:16, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[21] W. Feng, H. Guo, and J. Wang, "Improved bounds for randomly colouring simple hypergraphs," in *RANDOM*, pp. 25:1–25:17, Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022.

[22] K. He, C. Wang, and Y. Yin, "Sampling lovász local lemma for general constraint satisfaction solutions in near-linear time," in *FOCS*, pp. 147–158, IEEE, 2022.

[23] K. He, C. Wang, and Y. Yin, "Deterministic counting lovász local lemma beyond linear programming," in *SODA*, pp. 3388–3425, SIAM, 2023.

[24] G. Qiu and J. Wang, "Inapproximability of counting independent sets in linear hypergraphs," *Information Processing Letters*, vol. 184, p. 106448, 2024.

[25] K. He, K. Wu, and K. Yang, "Improved bounds for sampling solutions of random CNF formulas," in *SODA*, pp. 3330–3361, SIAM, 2023.

[26] A. Galanis, L. A. Goldberg, H. Guo, and A. Herrera-Poyatos, "Fast sampling of satisfying assignments from random k-sat," *arXiv*, vol. abs/2206.15308, 2022.

[27] Z. Chen, N. Mani, and A. Moitra, "From algorithms to connectivity and back: Finding a giant component in random k-SAT," in *SODA*, pp. 3437–3470, SIAM, 2023.

[28] D. Achlioptas and F. Iliopoulos, "Random walks that find perfect objects and the Lovász local lemma," in *FOCS*, pp. 494–503, IEEE, 2014.

[29] N. J. A. Harvey and J. Vondrák, "An algorithmic proof of the Lovász local lemma via resampling oracles," in *FOCS*, pp. 1327–1345, IEEE, 2015.

[30] V. Kolmogorov, "Commutativity in the algorithmic Lovász local lemma," in *FOCS*, pp. 780–787, IEEE, 2016.

[31] D. G. Harris and A. Srinivasan, "Algorithmic and enumerative aspects of the moser-tardos distribution," *ACM Trans. Algorithms*, vol. 13, no. 3, 2017.

[32] D. G. Harris and A. Srinivasan, "A constructive Lovász local lemma for permutations," *Theory Comput.*, vol. 13, pp. Paper No. 17, 41, 2017.

[33] D. G. Harris and A. Srinivasan, "The Moser-Tardos framework with partial resampling," *J. ACM*, vol. 66, no. 5, pp. Art. 36, 45, 2019.

[34] D. Achlioptas, F. Iliopoulos, and A. Sinclair, "Beyond the Lovász local lemma: point to set correlations and their algorithmic applications," in *FOCS*, pp. 725–744, IEEE, 2019.

[35] D. G. Harris, "Oblivious resampling oracles and parallel algorithms for the lopsided lovász local lemma," *ACM Trans. Algorithms*, vol. 17, no. 1, 2021.

[36] W. Feng, H. Guo, C. Wang, J. Wang, and Y. Yin, "Towards derandomising markov chain monte carlo," in *FOCS*, pp. 1963–1990, IEEE, 2023.

[37] M. Bordewich, M. Dyer, and M. Karpinski, "Stopping times, metrics and approximate counting," 2006.

[38] M. Bordewich, M. Dyer, and M. Karpinski, "Path coupling using stopping times and counting independent sets and colorings in hypergraphs," *ALGORITHMS*, p. 2008, 2008.

[39] A. Frieze and P. Melsted, "Randomly coloring simple hypergraphs," *Information Processing Letters*, vol. 111, no. 17, pp. 848–853, 2011.

[40] A. Frieze and M. Anastos, "Randomly coloring simple hypergraphs with fewer colors," *Information Processing Letters*, vol. 126, pp. 39–42, 2017.

[41] A. Galanis, L. A. Goldberg, H. Guo, and K. Yang, "Counting solutions to random CNF formulas," in *ICALP*, vol. 168 of *LIPIcs*, pp. 53:1–53:14, 2020.

[42] R. L. Dobrushin, "Prescribing a system of random variables by conditional distributions," *Theory of Probability & Its Applications*, vol. 15, no. 3, pp. 458–486, 1970.

[43] L. A. Goldberg, R. Martin, and M. Paterson, "Strong spatial mixing with fewer colors for lattice graphs," *SIAM J. Comput.*, vol. 35, no. 2, pp. 486–517, 2005.

[44] D. Weitz, "Counting independent sets up to the tree threshold," in *STOC*, pp. 140–149, ACM, 2006.

[45] M. Bayati, D. Gamarnik, D. Katz, C. Nair, and P. Tetali, "Simple deterministic approximation algorithms for counting matchings," in *STOC*, pp. 122–127, 2007.

[46] A. Sinclair, P. Srivastava, and M. Thurley, "Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs," in *SODA*, pp. 941–953, SIAM, 2012.

[47] L. Li, P. Lu, and Y. Yin, "Correlation decay up to uniqueness in spin systems," in *SODA*, p. 67–84, SIAM, 2013.

[48] P. Lu and Y. Yin, "Improved FPTAS for multi-spin systems," in *RANDOM*, pp. 639–654, Springer, 2013.

[49] Y. Yin and C. Zhang, "Approximate counting via correlation decay on planar graphs," in *SODA*, pp. 47–66, SIAM, 2013.

[50] D. Gamarnik, D. Katz, and S. Misra, "Strong spatial mixing of list coloring of graphs," *Random Structures & Algorithms*, vol. 46, no. 4, pp. 599–613, 2015.

[51] V. Patel and G. Regts, "Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials," *SIAM J. Comput.*, vol. 46, no. 6, pp. 1893–1919, 2017.

[52] J. Liu, A. Sinclair, and P. Srivastava, "Correlation decay and partition function zeros: Algorithms and phase transitions," *SIAM J. Comput.*, pp. 200–252, 2019.

[53] N. Anari, K. Liu, and S. Oveis Gharan, "Spectral independence in high-dimensional expanders and applications to the hardcore model," in *FOCS*, pp. 1319–1330, IEEE, 2020.

[54] Z. Chen, K. Liu, and E. Vigoda, "Rapid mixing of glauber dynamics up to uniqueness via contraction," in *FOCS*, pp. 1307–1318, IEEE, 2020.

[55] Z. Chen, A. Galanis, D. Štefankovič, and E. Vigoda, "Rapid mixing for colorings via spectral independence," in *SODA*, p. 1548–1557, SIAM, 2021.

[56] W. Feng, H. Guo, Y. Yin, and C. Zhang, "Rapid mixing from spectral independence beyond the Boolean Domain," in *SODA*, pp. 1558–1577, SIAM, 2021.

[57] R. Bubley and M. Dyer, "Path coupling: A technique for proving rapid mixing in markov chains," in *FOCS*, p. 223, IEEE, 1997.

[58] B. Haeupler, B. Saha, and A. Srinivasan, "New constructive aspects of the lovász local lemma," *J. ACM*, vol. 58, no. 6, pp. 28:1–28:28, 2011. (Conference version in *FOCS*'10).

[59] D. A. Levin and Y. Peres, *Markov chains and mixing times*. American Mathematical Soc., 2017.

[60] L. Khachiyan, "Polynomial algorithms in linear programming," *USSR Computational Mathematics and Mathematical Physics*, vol. 20, no. 1, pp. 53–72, 1980.

[61] C. Wang and Y. Yin, "A sampling Lovász local lemma for large domain sizes," *arXiv*, vol. abs/2307.14872v2, 2024.