

# DETERMINISTIC COUNTING FROM COUPLING INDEPENDENCE

XIAOYU CHEN, WEIMING FENG, HENG GUO, XINYUAN ZHANG, ZONGRUI ZOU

ABSTRACT. We show that spin systems with bounded degrees and coupling independence admit fully polynomial time approximation schemes (FPTAS). We design a new recursive deterministic counting algorithm to achieve this. As applications, we give the first FPTASes for  $q$ -colourings on graphs of bounded maximum degree  $\Delta \geq 3$ , when  $q \geq (11/6 - \varepsilon_0)\Delta$  for some small  $\varepsilon_0 \approx 10^{-5}$ , or when  $\Delta \geq 125$  and  $q \geq 1.809\Delta$ , and on graphs with sufficiently large (but constant) girth, when  $q \geq \Delta + 3$ . These bounds match the current best randomised approximate counting algorithms by Chen, Delcourt, Moitra, Perarnau, and Postle (2019), Carlson and Vigoda (2024), and Chen, Liu, Mani, and Moitra (2023), respectively.

## 1. INTRODUCTION

The power of randomness is a classical topic in the theory of computing. Randomised algorithms have found many early successes in the field of approximate counting. A striking example is the polynomial-time volume estimation algorithm for convex bodies by Dyer, Frieze, and Kannan [DFK91], whereas deterministic approximation algorithms requires at least exponential membership queries [Ele86, BF87]. However, this lower bound is valid only for membership query models, and does not rule out efficient deterministic approximation algorithms in general. While volume estimation remains difficult for efficient deterministic approximation, deterministic approximate counting algorithms have been quickly catching up with their randomised counterparts for many other problems, since the introduction of the correlation decay technique [Wei06, BG06].

By now, a number of different deterministic approximate counting techniques have been developed. In addition to the correlation decay method, one may utilise zero-freeness of polynomials [Bar16, PR17], linear programming based methods [Moi19, GLLZ19, JPV21], statistical physics related techniques [HPR20, JPP23, JPSS22], or even direct derandomisation of Markov chains [FGW<sup>+</sup>23]. In many occasions, these methods have achieved optimal results, or at least match or even outperform the best randomised algorithms, such as for the hardcore gas model [Wei06, PR17], for Holant problems with log-concave signatures [HLQZ24], or in the local lemma settings [WY23].

Despite all these successes, there is one problem where deterministic algorithms are still lagging behind, namely counting the number of proper colourings. The study of this problem was initiated by Jerrum [Jer95], who showed a rapid mixing bound for Glauber dynamics for  $q$ -colourings on graphs of maximum degree  $\Delta$ , when  $q > 2\Delta$ . This was subsequently improved by Vigoda [Vig00] to  $q > 11/6\Delta$  by considering the flip dynamics. Via more careful analysis, the constant was then improved to  $(11/6 - \varepsilon_0)$  for some  $\varepsilon_0 \approx 10^{-5}$  by Chen, Delcourt, Moitra, Perarnau, and

---

(Xiaoyu Chen, Xinyuan Zhang, Zongrui Zou) STATE KEY LABORATORY FOR NOVEL SOFTWARE TECHNOLOGY, NEW CORNERSTONE SCIENCE LABORATORY, NANJING UNIVERSITY, 163 XIANLIN AVENUE, NANJING, JIANGSU PROVINCE, CHINA.

(Weiming Feng) INSTITUTE FOR THEORETICAL STUDIES, ETH ZÜRICH, CLAUSIUSSTRASSE 47, 8092 ZÜRICH, SWITZERLAND.

(Heng Guo) SCHOOL OF INFORMATICS, UNIVERSITY OF EDINBURGH, INFORMATICS FORUM, EDINBURGH, EH8 9AB, UNITED KINGDOM.

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 947778). Weiming Feng acknowledges the support from Dr. Max Rössler, the Walter Haefner Foundation and the ETH Zürich Foundation.

Postle [CDM<sup>+</sup>19], and to 1.809 by Carlson and Vigoda [CV24] for  $\Delta \geq 125$ . In contrast, on the deterministic side, the first efficient algorithm by Gamarnik and Katz [GK07] requires  $q > 2.844\Delta$  via the correlation decay method. This was later improved by Lu and Yin [LY13] to  $q > 2.581\Delta$ . With the more recent technique using zero-freeness of polynomials, Liu, Srivasatva, and Sinclair [LSS19a] gave a fully polynomial time approximation schemes (FPTAS) when  $q \geq 2\Delta$ , and this bound was very recently improved to  $q \geq (2 - \varepsilon_1)\Delta$  for some  $\varepsilon_1 \approx 0.002$  by Bencs, Berrekkal, and Regts [BBR24].

In this paper, we close this gap between deterministic and randomised algorithms for approximate counting colourings. We introduce a new algorithm that takes inspiration from both the linear programming method and the correlation decay method. We show that this algorithm is efficient as long as coupling independence holds and there is a marginal lower bound. Here coupling independence is a method to establish the so-called spectral independence [ALO20, AL20], a relatively new tool to analyse mixing times of Markov chains. The notion of coupling independence is formally introduced in [CZ23], although it has been implicitly established before that, such as in [FGYZ22, BCC<sup>+</sup>22, Liu21]. Previously, coupling independence is mainly used to analyse Markov chains, and here we show that it also implies deterministic approximate counting algorithms. We also show that contractive coupling for Markov chains can be used to establish coupling independence. Thus, with our technique, the contractive couplings from [CDM<sup>+</sup>19, CV24] imply FPTASes with matching bounds. We describe our main results in more detail in Section 1.1, and give a high-level technical overview in Section 1.2.

**1.1. Main results.** We state our main results in the general context of spin systems. A spin system is specified by the tuple  $\mathcal{S} = (G, q, A_E, A_V)$ . Given a graph  $G = (V, E)$  and an integer  $q > 0$ , a state of the system is a configuration  $\sigma : V \rightarrow [q]$ . Namely, the state space is  $[q]^V$ . The weight of a configuration are characterised by the matrix  $A_E \in \mathbb{R}_{\geq 0}^{q \times q}$  and vector  $A_V \in \mathbb{R}_{\geq 0}^q$ . The Gibbs distribution  $\mu$  of  $\mathcal{S}$  is defined by

$$\mu(\sigma) \propto w(\sigma) := \prod_{\{u,v\} \in E} A_E(\sigma(u), \sigma(v)) \prod_{v \in V} A_V(\sigma(v)).$$

The normalising factor of  $\mu$ , namely the so-called partition function of  $\mathcal{S}$ , is defined by

$$Z := \sum_{\sigma \in [q]^V} w(\sigma).$$

When  $A_E = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix}$  and  $A_v = \begin{pmatrix} 1 \\ \lambda \end{pmatrix}$ , this encodes the hardcore gas model. When  $A_E = J - I$ , where  $J$  is the all-1 matrix and  $I$  is the identity matrix, and  $A_v = \mathbf{1}$  is the all-1 vector,  $\mu$  is uniform over proper  $q$ -colourings, and  $Z$  is the number of them.

To introduce coupling independence, we need to define Hamming and Wasserstein distances. For two configurations  $\sigma$  and  $\tau$ , let their Hamming distance be

$$\text{dist}\{\sigma, \tau\} := |\{v \mid v \in V, \sigma(v) \neq \tau(v)\}|.$$

The Wasserstein distance is defined next.

**Definition 1.** Let  $(\Omega, d)$  be a finite metric space. For any two distributions  $\mu$  and  $\nu$  on  $\Omega$ , the 1-Wasserstein distance (*W1-distance*) with respect to the metric  $d$  between  $\mu$  and  $\nu$  is defined as

$$(1) \quad \mathcal{W}_d(\mu, \nu) := \inf_{\mathcal{C}} \mathbb{E}_{(X,Y) \sim \mathcal{C}}[d(X, Y)],$$

where the infimum is taken over all the possible couplings  $\mathcal{C}$  between  $\mu$  and  $\nu$ .

For two  $\Omega$ -valued random variables  $X, Y$  with distribution  $\mu, \nu$ , we may also use  $\mathcal{W}_d(X, Y)$  to denote  $\mathcal{W}_d(\mu, \nu)$ . When the distance  $d$  is the Hamming distance, we also omit the subscript and write  $\mathcal{W}(\mu, \nu)$ .

**Definition 2** (Coupling independence). We say a Gibbs distribution  $\mu$  satisfies *C-coupling independence* if, for any two partial configurations  $\sigma$  and  $\tau$  on  $\Lambda \subseteq V$  such that  $\text{dist}\{\sigma, \tau\} = 1$ ,

$$\mathcal{W}(\mu^\sigma, \mu^\tau) \leq C,$$

where  $\mu^\sigma$  and  $\mu^\tau$  denote the Gibbs distribution conditional on  $\sigma$  and  $\tau$ , respectively.

In addition to coupling independence (CI), our main theorem also requires marginal lower bound. Let  $\mu_v^\sigma$  be the marginal distribution at  $v$  conditional on  $\sigma$ .

**Definition 3** (Marginal lower bound). We say a Gibbs distribution  $\mu$  over  $[q]^V$  is *b-marginally bounded* if for any partial configuration  $\sigma \in [q]^\Lambda$  on  $\Lambda \subseteq V$ , any vertex  $v \notin \Lambda$ , any spin  $c \in [q]$  with  $\mu_v^\sigma(c) > 0$ ,

$$(2) \quad \mu_v^\sigma(c) \geq b.$$

Now we are ready to state our main theorem.

**Theorem 4.** *Let  $q \geq 2, b > 0, C > 0, \Delta \geq 3$  be constants. There exists a deterministic algorithm such that given a permissive spin system  $\mathcal{S} = (G, q, A_E, A_V)$  and error bound  $0 < \varepsilon < 1$ , if the Gibbs distribution of  $\mathcal{S}$  is *b-marginally bounded* and satisfies *C-coupling independence*, and the maximum degree of  $G$  is at most  $\Delta$ , then it returns  $\hat{Z}$  satisfying  $(1 - \varepsilon)Z \leq \hat{Z} \leq (1 + \varepsilon)Z$  in time  $(\frac{n}{\varepsilon})^{f(q, b, C, \Delta)}$ , where  $f(q, b, C, \Delta) = \Delta^{O(C(\log b^{-1} + \log C + \log \log \Delta))} \log q$  is a constant.*

Being permissive is a mild technical condition we need (see Definition 9). It roughly requires that all the conditional distributions are well-defined. All applications considered in this paper satisfy it. The marginal lower bound is also a mild requirement, since if we treat  $q, \Delta, A_E$ , and  $A_V$  all as constants, then there is a constant  $b$  such that any permissive system is *b-marginally bounded* (see Observation 10). However, for concrete systems, there are usually better lower bound than the generic one in Observation 10. Thus, we choose to make it explicit in the statement of Theorem 4. We also note that the exponent  $f(q, b, C, \Delta)$  is not optimised – our goal is to present the new algorithm as simple and as clearly as possible.

Theorem 4 together with coupling independence from [CLMM23] and the marginal lower bound [GKM15, Lemma 3]<sup>1</sup> directly implies the following result.

**Corollary 5** (Colouring: high-girth graphs). *Let  $q$  and  $\Delta$  be two integers satisfying  $\Delta \geq 3$  and  $q \geq \Delta + 3$ . There exists a constant  $g_0 > 0$  depending only on  $\Delta$  such that the following holds. There exists an FPTAS for the number of proper  $q$ -colourings on graphs  $G$  of maximum degree  $\Delta$  and girth at least  $g_0$ .*

The bound in Corollary 5 matches the rapid mixing result by Chen, Liu, Mani and Moitra [CLMM23]. Prior to our work, no FPTAS is known in this setting.

To establish coupling independence for colourings in general bounded degree graphs, we make use of contractive coupling, a tool typically used to bound the mixing time of Markov chains. Given two copies of a Markov chain, a contractive coupling ensures that after a step, the expected distance decreases multiplicatively. See (17) and (18) for some examples. Previously, contractive couplings have been used to establish spectral independence [BCC<sup>+</sup>22, Liu21]. In fact, their proof implicitly established the stronger result of coupling independence. In Section 5, we give a simpler and more direct argument on how to establish coupling independence from contractive coupling. Using the

<sup>1</sup>The bound yields  $b \geq q^{-1} (\frac{2}{3})^\Delta$  in the setting of Corollary 5.

contractive couplings from [CV24] and [CDM<sup>+</sup>19], as well as the marginal lower bound [LY13, Lemma 3],<sup>2</sup> we have the following result.

**Corollary 6** (Colouring: general graphs). *Let  $q$  and  $\Delta$  be two integers satisfying either*

- $\Delta \geq 125, q \geq 1.809\Delta$ ;
- *or*  $\Delta \geq 3, q \geq (11/6 - \varepsilon_0)\Delta$  for some fixed parameter  $\varepsilon_0 \approx 10^{-5}$ .

*There exists an FPTAS for the number of proper  $q$ -colourings on graphs of maximum degree  $\Delta$ .*

The bounds in Corollary 6 are the same ones as the rapid mixing of Markov chains results by Carlson and Vigoda [CV24] or by Chen, Delcourt, Moitra, Perarnau and Postle [CDM<sup>+</sup>19]. As mentioned before, prior to our work, the best FPTAS [BBR24] requires  $q \geq (2 - \varepsilon_1)\Delta$  for some  $\varepsilon_1 \approx 0.002$ .

Theorem 4 also implies FPTASes for spin systems satisfying the *Dobrushin-Shlosman* condition.

**Definition 7** (Dobrushin-Shlosman condition). Let  $\mu$  be a Gibbs distribution on  $[q]^V$ . The *Dobrushin influence matrix*  $\rho \in \mathbb{R}_{\geq 0}^{V \times V}$  is defined by

$$(3) \quad \forall u, v \in V, \rho(u, v) := \max_{\substack{\sigma, \tau \in [q]^{V-v} \\ \sigma \oplus \tau \subseteq \{u\}}} d_{\text{TV}}(\mu_v^\sigma, \mu_v^\tau),$$

where we use  $d_{\text{TV}}(\cdot, \cdot)$  to denote the TV-distance. The Gibbs distribution  $\mu$  is said to satisfy the *Dobrushin-Shlosman condition* with gap  $\delta \in (0, 1)$  if

$$\|\rho\|_1 = \max_{u \in V} \sum_{v \in V} \rho(u, v) \leq 1 - \delta.$$

**Corollary 8.** *Let  $q \geq 2, \Delta \geq 3, A_E \in \mathbb{R}_{\geq 0}^{q \times q}, A_V \in \mathbb{R}_{\geq 0}^q$ , and  $\delta \in (0, 1)$  be constant parameters. There exists an FPTAS for the partition function of permissive spin systems  $\mathcal{S} = (G, q, A_E, A_V)$  if  $\mathcal{S}$  satisfies the Dobrushin-Shlosman condition with gap  $\delta$  and the maximum degree of  $G$  is at most  $\Delta$ .*

The Dobrushin-Shlosman condition [Dob70, DS85] is a sufficient condition for the uniqueness of the Gibbs measure in infinite graphs. It is well-known that the Dobrushin-Shlosman condition implies rapid mixing of Glauber dynamics [BD97, Hay06]. However, before our result, its implication on deterministic counting algorithms is not well-understood, especially for general multi-spin ( $q > 2$ ) systems.

In addition, Theorem 4 provides a unified framework to derive FPTASes for a few problems for which FPTASes are known before via different methods. Similar to our method, all these FPTASes require a bounded maximum degree  $\Delta$  for the input graph. Examples in this category include  $q$ -colourings for triangle-free graphs if  $q > 1.764\Delta + C$  for some constant  $C > 0$  [LSS19a] (where CI is established in [FGYZ22, CGSV21, CF24]), Holant problems with log-concave signatures [HLQZ24] (where CI is established in [CG24]), antiferromagnetic two-state spin systems in the uniqueness regime [SST14] (where CI is established in [CF24, CLV23]), and ferromagnetic Ising models with non-zero external fields [LSS19b] (where CI is established in [CZ23]).

**1.2. Our technique.** We first show that coupling independence implies decay of total influences. Here, the influence from  $u$  to  $v$  is defined as  $d_{\text{TV}}(\mu_v^\sigma, \mu_v^\tau)$  such that  $\sigma$  and  $\tau$  are *partial configurations* that differ only at  $u$ . Note that this is different from Dobrushin's influence defined in (3). See Definition 11 for the formal definition of influences. Suppose  $C$ -coupling independence holds. By a simple averaging argument, there is  $1 \leq \ell \leq 2C$  such that the total influence of a vertex  $v$  to vertices of distance  $\ell$  from  $v$  is at most  $1/2$ . We repeatedly use this fact to show that, roughly, for

<sup>2</sup>The bound yields  $b \geq q^{-1} e^{-\frac{1}{\alpha-1}}$  in the setting of Corollary 6, where  $\alpha$  is the constant in the assumption  $q \geq \alpha\Delta$ .

every  $2C$  distance, the total influence decays multiplicatively by a factor  $1/2$ , which leads to an exponential decay in the distance from  $v$ . This is in Section 3. Given the decay of total influence, we may choose a sufficiently large but constant  $R$  such that the total influence at distance  $R$  is less than a quantity that is linear in  $R$ . (To be precise, the requirement is given in (10).) This choice is a key guarantee in our error analysis.

Our main algorithm estimates marginal probabilities of a vertex  $v$  under arbitrary conditioning. The basic building block of our algorithm does this task but requires a constant number of other conditional marginal probabilities. These marginal probabilities are either of the same vertex  $v$  but with different boundary conditions at distance  $R$ , or of some other vertices at distance  $R$ . We use the aforementioned bounds for  $R$  to show that the output relative error is roughly half of the relative error of the input marginals. Thus, we achieved a recursive step with constant error decay. To achieve an  $\varepsilon$  relative error, we just need to run this up to  $\log \varepsilon^{-1}$  depth, resulting in polynomial total running time.

On a high level, the structure of our algorithm is very similar to the correlation decay algorithms such as [Wei06, BG06]. Namely, in each step, we use the marginal probabilities in smaller instances to compute the desired one, and the computation is truncated at logarithmic depth. However, a key difference is that in previous algorithms, the recursive step is exact and often via a closed form formula, whereas ours is algorithmic and approximate. This is achieved via a linear programming (LP) algorithm inspired by [Moi19]. We use the LP to simulate a coupling similar to the one in [CLMM23], which is to couple vertices randomly chosen at distance  $R$  one at a time. However, the coupling in [CLMM23] is recursive, and it appears difficult for the LP to handle recursion. We instead only use the LP to simulate a partial coupling up to the recursive point. Because this is a partial coupling, certain quantities in the LP cannot be computed efficiently. We rewrite these quantities in terms of marginal probabilities of smaller instances, and use recursion to solve this issue. To control the evolution of relative errors, we need to find some new linear constraints that can be computed efficiently and characterise the failure probability of the coupling process. These constraints, together with our choice of the radius  $R$ , ensure that the relative error decay by a constant factor each time. The description and analysis of our algorithm are given in Section 4.

Comparing with all previous instantiation of the LP-based approximate counting algorithm [Moi19, GLLZ19, JPV21, WY23, HLQZ24], we do not write a polynomial-sized LP to solve. Instead, our LP is only of constant size, but we recursively construct polynomially many of them. Technically, because each of our sub-instances has only constant size, we are not obliged to use LP. (For example, we could write a quadratic system to solve.) We choose LP just for technical convenience. Moreover, because the (partial) coupling we use the LP to simulate has not been considered in this context before, we need to write a new set of constraints to certify this coupling.

The coupling in [CLMM23] establishes coupling independence for high-girth graphs when  $q \geq \Delta + 3$ . Thus our algorithm works in this setting, resulting in Corollary 5. To apply our algorithm on general bounded degree graphs, we still need to establish coupling independence. To this end, we show that contractive coupling for Markov chains (which is the main technique behind the rapid mixing results of [CDM<sup>+</sup>19] and [CV24]) can be used to establish coupling independence. This argument is given in Section 5.

## 2. PRELIMINARIES

Let  $\mu$  over  $[q]^V$  be a Gibbs distribution on  $G = (V, E)$  with matrix  $A_E$  and vector  $A_V$ . Recall that  $\mu$  is defined by for any configuration  $\sigma \in [q]^V$ ,

$$\mu(\sigma) = \frac{w(\sigma)}{Z}, \text{ where } w(\sigma) = \prod_{\{u,v\} \in E} A_E(\sigma(u), \sigma(v)) \prod_{v \in V} A_V(\sigma(v)) \text{ and } Z = \sum_{\tau \in [q]^V} w(\tau).$$

We call  $w(\sigma)$  the weight of  $\sigma$  and  $Z$  the partition function.

For any subset  $S \subseteq V$ , we use subscript  $S$  to denote the marginal distribution, namely, for any  $\tau \in [q]^S$ ,  $\mu_S(\tau) := \sum_{\tau' \in [q]^{V \setminus S}} \mu(\tau \cup \tau')$ . When  $S = \{v\}$  is a singleton set, we may also write  $\mu_v$  instead of  $\mu_{\{v\}}$ .

Let  $\sigma \in [q]^{V \setminus \Lambda}$  be a partial configuration on  $V \setminus \Lambda$ , where  $\sigma$  is allowed to be infeasible, namely we allow  $\mu_{V \setminus \Lambda}(\sigma) = 0$ . For any  $\tau \in [q]^V$ , define the conditional weight

$$w^\sigma(\tau) = \mathbf{1}_{\tau(V \setminus \Lambda) = \sigma} \cdot \prod_{v \in \Lambda} A_V(\tau(v)) \prod_{\{u,v\} \in E: u \in \Lambda \wedge v \notin \Lambda} A_E(\tau(u), \sigma(v)) \prod_{\{u,v\} \in E: u \in \Lambda \wedge v \in \Lambda} A_E(\tau(u), \tau(v)).$$

Define the conditional distribution  $\mu^\sigma$  over  $[q]^V$  by for any  $\tau \in [q]^V$ ,

$$(4) \quad \mu^\sigma(\tau) = \frac{w^\sigma(\tau)}{Z^\sigma}, \text{ where } Z^\sigma = \sum_{\tau \in [q]^V} w^\sigma(\tau).$$

The above definition works for all partial configurations  $\sigma$ . In particular, if  $\sigma$  is a feasible partial configuration, then  $\mu^\sigma$  is the distribution  $\mu$  conditional on  $\sigma$ . Let  $\mu_S^\sigma$  denote the marginal distribution on  $S$  conditional on the partial configuration  $\sigma$ .

As mentioned before, when  $A_E = J - I$ , where  $J$  is the all-1 matrix and  $I$  is the identity matrix, and  $A_v = \mathbf{1}$  is the all-1 vector,  $\mu$  is uniform over proper  $q$ -colourings, and  $Z$  is the number of them. For technical purposes, we also need to consider list colourings later, where each  $v$  may have a list  $L_v$  of available colours, and no edge can be monochromatic. This can also be modelled as a spin system, with  $q$  being the total number of possible colours of all vertices, and  $A_v$  encoding what colours are available for  $v$ . Let  $\mu$  be the uniform distribution over all proper colourings. Our list colouring instances in fact come from  $\mu^\sigma$  for some partial configuration  $\sigma$ . This is because fixing a colour  $c$  at a vertex  $v$  is equivalent to removing the vertex  $v$  and removing  $c$  from the lists of all neighbours of  $v$ . Note that if  $q - \Delta \geq k$  for some  $k$ , then for any list colouring instances obtained this way, for any  $v$ ,  $|L_v| - \deg_G(v) \geq k$  as well.

More generally, in this paper, we consider the permissive spin systems so that all  $\mu^\sigma$  are well-defined.

**Definition 9.** A spin system  $\mu$  is *permissive* if for any  $\Lambda \subseteq V$ , any  $\sigma \in [q]^\Lambda$ ,  $Z^\sigma > 0$ .

*Remark.* All spin systems with soft constraints ( $A_E(i, j) > 0$  and  $A_V(i) > 0$  for all  $i, j \in [q]$ ) are permissive. Many natural spin systems with hard constraints are also permissive. For example, the hardcore model and any list colouring instance  $(G, L)$  such that  $|L_v| \geq \deg_G(v) + 1$  for any  $v \in V$  are permissive.

Note that in Definition 9,  $\sigma$  is allowed to be infeasible, as  $w^\sigma$  does not consider the weight contributed from inside  $\Lambda$ . This allows  $\mu^\sigma$  to be well-defined, even for infeasible  $\sigma$ . Our definition of coupling independence, Definition 2, indeed allows infeasible partial configurations. Essentially, for any (feasible or infeasible)  $\sigma$  on  $\Lambda \subset V$ , only the values on the boundary of  $\Lambda$  matters for  $\mu^\sigma$ , and the values inside  $\Lambda$  do not.

On the other hand, for any partial configuration  $\sigma$  on  $\Lambda$ , if  $\sigma$  is locally feasible, then  $\sigma$  is also globally feasible. Formally,

$$(5) \quad \prod_{v \in \Lambda} A_V(\sigma(v)) \prod_{\{u,v\} \in E: u \in \Lambda \wedge v \in \Lambda} A_E(\sigma(u), \sigma(v)) > 0 \implies \mu_\Lambda(\sigma) > 0.$$

We use  $\text{supp}(\mu_S^\sigma)$  to denote the support of  $\mu_S^\sigma$ . Formally,

$$\text{supp}(\mu_S^\sigma) := \{\tau \in [q]^S \mid \mu_S^\sigma(\tau) > 0\}.$$

Note that by (5), the set  $\text{supp}(\mu_S^\sigma)$  is easy to compute as we only need to consider local assignments.

Let us also observe that permissive systems always have a marginal lower bound.

**Observation 10.** Let  $q, \Delta, A_E,$  and  $A_V$  be constants. Then there is a constant  $b = b_{q,\Delta,A_E,A_V}$  such that for any  $G$  with maximum degree  $\Delta$ , any permissive spin system  $(G, q, A_E, A_V)$  is  $b$ -marginally bounded.

*Proof.* Given a partial configuration  $\sigma$  on  $\Lambda$  and a vertex  $v \in V \setminus \Lambda$ , let  $S$  be the set of 2-hop neighbours of  $v$  that is not in  $\Lambda$ . For any  $c \in \text{supp}(\mu_v^\sigma)$ , we have that

$$\mu_v^\sigma(c) = \sum_{\tau \in \text{supp}(\mu_S^\sigma)} \mu_S^\sigma(\tau) \mu_v^{\sigma \cup \tau}(c).$$

For any  $c \in \text{supp}(\mu_v^\sigma)$  and  $\tau \in \text{supp}(\mu_S^\sigma)$ , as  $S$  is not adjacent to  $v$  and the system is permissive,  $\mu_v^{\sigma \cup \tau}(c) > 0$ . As both  $\text{supp}(\mu_S^\sigma)$  and  $\text{supp}(\mu_v^\sigma)$  are finite, there is a minimum of  $\mu_v^{\sigma \cup \tau}(c)$ , say  $b$ , over all choices of  $\tau \in \text{supp}(\mu_S^\sigma)$  and  $c \in \text{supp}(\mu_v^\sigma)$ . Note that this  $b$  may depend on  $q, \Delta, A_E,$  and  $A_V$ . It implies that

$$\mu_v^\sigma(c) \geq b \sum_{\tau \in \text{supp}(\mu_S^\sigma)} \mu_S^\sigma(\tau) = b. \quad \square$$

### 3. TOTAL INFLUENCE DECAY

Given a graph  $G = (V, E)$ , let  $B_\ell(v)$  denote the ball of radius  $\ell$  centered at  $v$ . Namely,  $B_\ell(v) := \{u \mid u \in V, \text{dist}_G(u, v) \leq \ell\}$ , where the distance is graph distance in  $G$ . Let  $S_\ell(v)$  denote the sphere of radius  $\ell$  centered at  $v$ , namely,  $S_\ell(v) := \{u \mid u \in V, \text{dist}_G(u, v) = \ell\}$ . In other words,  $S_{\ell+1}(v) = \partial B_\ell(v)$ , where  $\partial B_\ell(v)$  denote the out-boundary of  $B_\ell(v)$ .

**Definition 11** (Total influence decay). Let  $\delta : \mathbb{N} \rightarrow \mathbb{R}$  be a non-increasing function. We say a Gibbs distribution  $\mu$  satisfies *total influence decay* with rate  $\delta$  if for any two partial configurations  $\sigma$  and  $\tau$  on  $\Lambda \subseteq V$  such that they disagree only on some  $v \in \Lambda$ , for any integer  $\ell > 0$ ,

$$(6) \quad \sum_{u \in S_\ell(v)} d_{\text{TV}}(\mu_u^\sigma, \mu_u^\tau) \leq \delta(\ell).$$

The influence defined in (6) was often used in recent works for spectral independence [ALO20, FGYZ22, CGSV21]. The definition in (6) should be distinguished from the definition of Dobrushin's influence in (3). The following theorem shows that the coupling independence implies total influence decay.

**Theorem 12.** Suppose  $\mu$  satisfies  $C$ -coupling independence. Let  $R > 0$  be an integer. For any two partial configurations  $\sigma$  and  $\tau$  on  $\Lambda \subseteq V$  such that they disagree only on some  $v \in \Lambda$ , there is a coupling  $\mathcal{C}$  between  $\mu^\sigma$  and  $\mu^\tau$  such that

$$\mathbb{E}_{(\sigma', \tau') \sim \mathcal{C}} \left[ \sum_{u \in S_R(v)} \mathbf{1}_{\sigma'(u) \neq \tau'(u)} \right] \leq 2C \left( \frac{1}{2} \right)^{\lceil \frac{R}{2C} \rceil}.$$

As a consequence,  $\mu$  satisfies total influence decay with rate  $\delta(x) = 2C \cdot 2^{-\lceil \frac{x}{2C} \rceil}$ .

*Proof.* Let  $T := \lceil \frac{R}{2C} \rceil$ . We construct  $\mathcal{C}$  via a  $T$ -step coupling procedure as follows. The whole procedure will generate a random pair  $X, Y \in [q]^V$  such that  $X \sim \mu^\sigma$  and  $Y \sim \mu^\tau$ .

**Step 1.** Since  $\mu$  satisfies  $C$ -coupling independence, there exists a coupling  $\mathcal{C}_1$  of  $\mu^\sigma$  and  $\mu^\tau$  such that  $\sum_{w \in V} \Pr_{(\sigma', \tau') \sim \mathcal{C}_1}[\sigma'(w) \neq \tau'(w)] \leq C$ . By an averaging argument, there exists an integer  $1 \leq \ell(1) \leq 2C$  such that the expected disagreement on the sphere  $S_{\ell(1)}(v)$  is at most  $\frac{C}{2C} = \frac{1}{2}$ , i.e.,

$$\sum_{w \in S_{\ell(1)}(v)} \Pr_{(\sigma', \tau') \sim \mathcal{C}_1}[\sigma'(w) \neq \tau'(w)] \leq \frac{1}{2}.$$

If there are multiple choices of such  $\ell(1)$ , pick the smallest one. Denote  $B_1 = B_{\ell(1)}(v)$  and  $S_1 = S_{\ell(1)}(v)$ . We use the coupling  $\mathcal{C}_1$  to draw a pair of random configurations and project them to the set  $D = B_1 \setminus \Lambda$ . We then extend  $\sigma, \tau$  on  $\Lambda$  to a pair of random configurations  $X, Y$  on  $\Lambda \cup B_1$ . Formally,  $X(\Lambda) = \sigma$ ,  $Y(\Lambda) = \tau$ , and

$$(7) \quad X(D) \sim \mu_D^\sigma, Y(D) \sim \mu_D^\tau, \text{ and } \mathbb{E}[\text{dist}\{X(S_1), Y(S_1)\}] \leq \frac{1}{2}.$$

**Step  $k$  for  $1 < k < T$ .** Suppose we have obtained a pair of partial configurations  $X$  and  $Y$  on  $B_{k-1} \cup \Lambda$ , where  $B_{k-1} = B_{\ell(k-1)}(v)$  and  $\ell(k-1) \geq 1$  is an integer. Next we extend the two configurations  $X, Y$  onto a larger set  $B_k \cup \Lambda$ , where  $B_k = B_{\ell(k)}(v)$  for some  $\ell(k) > \ell(k-1)$ . Let  $H = V \setminus B_{k-1}$  be the set of vertices outside the ball  $B_{k-1}$ . Let  $\Lambda_H = \Lambda \cap H$ . By conditional independence,  $\mu_H^X = \mu_H^{X(\Lambda_H \cup S_{k-1})}$  and  $\mu_H^Y = \mu_H^{Y(\Lambda_H \cup S_{k-1})}$ . The two pinnings  $X(\Lambda_H \cup S_{k-1})$  and  $Y(\Lambda_H \cup S_{k-1})$  can disagree only at the set  $S_{k-1} = S_{\ell(k-1)}(v)$ , the sphere of radius  $\ell(k-1)$  centered at  $v$ . Suppose the disagreements of  $X(S_{k-1})$  and  $Y(S_{k-1})$  can be listed as  $v_1, v_2, \dots, v_m \in S_{k-1}$ , where  $m = \text{dist}\{X(S_{k-1}), Y(S_{k-1})\}$ . We can define a sequence of pinnings  $\sigma_0, \sigma_1, \dots, \sigma_m$  on  $S_{k-1} \cup \Lambda_H$  such that  $\sigma_0 = X(\Lambda_H \cup S_{k-1})$ ,  $\sigma_i$  is obtained from  $\sigma_{i-1}$  by changing the value at  $v_i$  from  $X(v_i)$  to  $Y(v_i)$ , and so  $\sigma_m = Y(\Lambda_H \cup S_{k-1})$ . By the coupling independence, for any  $i \in [m]$ , one can couple  $\mu^{\sigma_i}$  and  $\mu^{\sigma_{i-1}}$  such that the expected hamming distance is at most  $C$ .<sup>3</sup> By the triangle inequality of the Wasserstein distance, we can couple  $\mu^{\sigma_0}$  and  $\mu^{\sigma_m}$  so that the expected Hamming distance is at most  $mC$ . By projecting this coupling into the subset  $H$ , we obtain a coupling  $\mathcal{C}_k$  of  $\mu_H^X = \mu_H^{\sigma_0}$  and  $\mu_H^Y = \mu_H^{\sigma_m}$  such that  $\mathbb{E}_{(\sigma_k, \tau_k) \sim \mathcal{C}_k}[\text{dist}\{\sigma_k, \tau_k\}] \leq mC$ . Again, by an averaging argument, there exists  $\ell(k)$  such that  $\ell(k-1) + 1 \leq \ell(k) \leq \ell(k-1) + 2C$  and

$$\sum_{w \in S_{\ell(k)}(v)} \Pr_{(\sigma_k, \tau_k) \sim \mathcal{C}_k}[\sigma_k(w) \neq \tau_k(w)] \leq \frac{mC}{2C} = \frac{m}{2} = \frac{\text{dist}\{X(S_{k-1}), Y(S_{k-1})\}}{2}.$$

If there are multiple choices of such  $\ell(k)$ , pick the smallest one. Let  $B_k = B_{\ell(k)}(v)$  and  $S_k = S_{\ell(k)}(v)$ . Similar to Step 1, we use the coupling  $\mathcal{C}_k$  to draw a pair of configurations  $(\sigma_k, \tau_k)$ . Let  $D_k = B_k \setminus (B_{k-1} \cup \Lambda)$ . We further extend  $X, Y$  to the set  $D_k$  by setting  $X(D_k) = \sigma_k(D_k)$  and  $Y(D_k) = \tau_k(D_k)$ . We have

$$X(D_k) \sim \mu_{D_k}^{X(B_{k-1} \cup \Lambda)}, Y(D_k) \sim \mu_{D_k}^{Y(B_{k-1} \cup \Lambda)}.$$

In other words,  $X$  and  $Y$  are now partial configurations on set  $B_{k-1} \cup \Lambda \cup D_k = B_k \cup \Lambda$ . It holds that

$$(8) \quad \mathbb{E}[\text{dist}\{X(S_k), Y(S_k)\} \mid X(B_{k-1} \cup \Lambda), Y(B_{k-1} \cup \Lambda), \ell(k-1)] \leq \frac{\text{dist}\{X(S_{k-1}), Y(S_{k-1})\}}{2}.$$

**Step  $T$ .** The last step is similar to the general step  $k$ . Assume  $X, Y$  are two fixed configurations on  $B_{T-1} \cup \Lambda$ , where  $B_{T-1} = B_{\ell(T-1)}(v)$  and  $\ell(T-1)$  is a fixed integer. Let  $H = V \setminus B_{T-1}$ . Let  $\mathcal{C}_T$  be the coupling obtained in the same way as the general step from the triangle inequality. We use the coupling  $\mathcal{C}_T$  to sample a pair of configurations  $(\sigma_T, \tau_T)$ . We then further extend  $X, Y$  to the set  $D_T = V \setminus (B_{T-1} \cup \Lambda)$  by setting  $X(D_T) = \sigma_T(D_T)$  and  $Y(D_T) = \tau_T(D_T)$ . We have

$$X(D_T) \sim \mu_{D_T}^{X(B_{T-1} \cup \Lambda)}, Y(D_T) \sim \mu_{D_T}^{Y(B_{T-1} \cup \Lambda)}.$$

Note that  $\ell(T-1) \leq 2C(T-1) = 2C(\lceil \frac{R}{2C} \rceil - 1) < R$ , so that  $S_R(v) \subseteq H$ . It holds that

$$(9) \quad \begin{aligned} & \mathbb{E}[\text{dist}\{X(S_R(v)), Y(S_R(v))\} \mid X(B_{T-1} \cup \Lambda), Y(B_{T-1} \cup \Lambda), \ell(T-1)] \\ & \leq C \cdot \text{dist}\{X(S_{T-1}), Y(S_{T-1})\}. \end{aligned}$$

<sup>3</sup>Two pinnings  $\sigma_i, \sigma_{i-1}$  can be improper partial configurations but conditional distributions  $\mu^{\sigma_i}$  and  $\mu^{\sigma_{i-1}}$  are defined in (4). The coupling independence condition holds for all possible pinnings including infeasible ones.



The above inequality holds because the expected disagreement on  $S_R(v)$  is at most the total expected disagreement produced by  $\mathcal{C}_T$ , which is at most  $C \cdot \text{dist}\{X(S_{T-1}), Y(S_{T-1})\}$ .

We now show that the above procedure is a valid coupling between  $\mu^\sigma$  and  $\mu^\tau$ . The procedure can be viewed as follows. Initially,  $X$  and  $Y$  are partial configurations on  $D_0 = \Lambda$  with  $X = \sigma$  and  $Y = \tau$ . In the  $k$ -th step, we extend  $X, Y$  to a new set  $D_k$ . Hence  $X, Y$  are partial configurations on  $\cup_{i \leq k} D_i$  after the  $k$ -th step. The tricky part here is that all  $D_1, D_2, \dots, D_T$  are random variables. The following property is the key to proving the correctness of the coupling: for every step  $1 \leq k \leq T$ , the coupling satisfies

- given  $(D_i)_{i < k}$ ,  $X(D_{0:k-1})$ , and  $Y(D_{0:k-1})$ , the set  $D_k$  is fixed, where  $D_{0:k-1} = \cup_{0 \leq i \leq k-1} D_i$ ;
- $X(D_k) \sim \mu_{D_k}^{X(D_{0:k-1})}$  and  $Y(D_k) \sim \mu_{D_k}^{Y(D_{0:k-1})}$ .

This property is easy to verify from the construction of  $X$  and  $Y$ .

Next we show  $X \sim \mu^\sigma$ . A similar proof shows that  $Y \sim \mu^\tau$ . In every step, we sample  $X(D_k)$  and  $Y(D_k)$  jointly from a coupling. This can be viewed as a two-step process. We first sample  $X(D_k)$ , and given the value of  $X(D_k)$ , the coupling specifies a conditional distribution of  $Y(D_k)$ . Then we sample  $Y(D_k)$  from this conditional distribution using independent randomness  $\mathcal{R}_k$  (e.g.,  $\mathcal{R}_k$  can be a uniform real number in  $(0, 1)$ ). We fix the randomness  $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_T$ . For any configuration  $\rho \in [q]^V$ , we compute the probability that  $X = \rho$ . Initially,  $D_0 = \Lambda$ ,  $X(D_0) = \sigma$  and  $Y(D_0) = \tau$ . Hence,  $X = \rho$  only if  $\rho(\Lambda) = \sigma$ . In the first step, by the property above,  $D_1$  is fixed, and  $X = \rho$  implies that  $X(D_1) = \rho(D_1)$ , which happens with probability  $\mu_{D_1}^{\rho(D_0)}(\rho(D_1))$ . Also,  $Y(D_1)$  is fixed because  $X(D_1) = \rho(D_1)$  and  $\mathcal{R}_1$  is fixed. By induction, in the  $k$ -th step,  $D_k$  is fixed,  $X(D_k) = \rho(D_k)$  with probability  $\mu_{D_k}^{\rho(D_{0:k-1})}(\rho(D_k))$ , and  $Y(D_k)$  is fixed by  $X(D_k)$  and  $\mathcal{R}_k$ . By the chain rule,

$$\Pr[X = \rho \mid \mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_T] = \mathbf{1}_{\rho(\Lambda) = \sigma} \cdot \prod_{k=1}^T \mu_{D_k}^{\rho(D_{0:k-1})}(\rho(D_k)) = \mu^\sigma(\rho).$$

Taking expectation over  $\mathcal{R}_1, \mathcal{R}_2, \dots, \mathcal{R}_T$  in both sides shows that  $X \sim \mu^\sigma$ .

Finally, using (7), (8), and (9), the expected disagreement on  $S_R(v)$  can be bounded as

$$\begin{aligned} \mathbb{E}[\text{dist}\{X(S_R(v)), Y(S_R(v))\}] &\leq C \mathbb{E}[\text{dist}\{X(S_{T-1}), Y(S_{T-1})\}] \\ &\leq \frac{C}{2^{T-2}} \mathbb{E}[\text{dist}\{X(S_1), Y(S_1)\}] \\ &\leq \frac{C}{2^{T-1}}. \end{aligned}$$

The total influence decay consequence follows from the coupling inequality.  $\square$

#### 4. A RECURSIVE MARGINAL ESTIMATOR

We show Theorem 4 in this section. As in the condition of Theorem 4, throughout the section, we assume that the underlying graphs  $G$  for the Gibbs distributions have constant maximum degree  $\Delta$ . Moreover, we assume a marginal lower bound  $0 < b < 1$  as in Definition 3, and exponential total influence decay, namely the bound in (6) with  $\delta(\ell) = \exp(-\Omega(\ell))$ .

Our algorithm requires a parameter  $R$ , the radius at which we couple vertices. For an integer  $k \geq 0$ , let  $H(k)$  be the harmonic sum defined by  $H(k) := \sum_{i=1}^k \frac{1}{i}$  with the convention that  $H(0) = 0$ . We choose a sufficiently large integer  $R$  such that

$$(10) \quad 30\delta(R)H(\Delta^R) < b^4.$$

By Theorem 12, we can take  $\delta(R) = 2C2^{-\lceil \frac{R}{2C} \rceil}$ , where  $C$  is the coupling independence constant. Since  $H(\Delta^R) = O(R \log \Delta)$ , some  $R = O(C(\log b^{-1} + \log C + \log \log \Delta))$  suffices.

In Section 4.1, we introduce an LP based algorithm that takes as inputs estimated ratios of marginal probabilities of some partial configurations, and outputs an estimation of the marginal ratio of a particular vertex, with a better approximation guarantee. The error analysis of this algorithm is given in Section 4.2. These input ratios can be written as the product of two ratios, each of which is regarding a single vertex. Thus, we have a recursive algorithm to estimate the marginal probability, described in Section 4.3.

**4.1. Marginal estimation via linear programming.** Suppose we want to estimate the marginal probability of some vertex  $u$ . It is equivalent to approximate the ratio between marginals where any two different values are assigned to  $u$ . The basic building block of our algorithm is an estimator which takes estimations of marginal ratios of certain partial configurations, and outputs an estimation of the marginal ratio for  $u$  with a better approximation guarantee. We construct a linear program, similar to the one used by Moitra [Moi19], to certify a coupling between assigning  $u$  to two different values. However, the “coupling” we choose is different, and is inspired by the one used by Chen, Liu, Mani, and Moitra [CLMM23]. Given two partial configurations  $\sigma$  and  $\tau$  on  $\Lambda \subseteq V$  which differ at only one vertex  $u \in \Lambda$ , we want to couple  $\mu^\sigma$  and  $\mu^\tau$  by coupling vertices in  $S_R(u) = \partial B_R(u)$  for some radius  $R > 0$ . We choose a vertex  $v$  from  $S_R(u)$  uniformly at random and couple it optimally (in the sense of TV distance) between its two marginal distributions. If the coupling failed, we immediately stop the whole process, and otherwise we continue to couple the next randomly chosen vertex. The process also ends when all of  $S_R(u)$  has been considered. One may have noticed that this is not a complete coupling, but rather a partial one. Formally it is described in Algorithm 1. For  $\sigma \in [q]^\Lambda$  and  $v \notin \Lambda$ , the notation  $\sigma_{v \leftarrow c}$  means a partial configuration which agrees with  $\sigma$  on  $\Lambda$  and assigns  $c$  to  $v$ .

---

**Algorithm 1:** The partial coupling

---

```

1 Partial-Coupling  $(\sigma, \tau)$ ;
   Input : Partial configurations  $\sigma, \tau \in [q]^\Lambda$  that only differ at a vertex  $u \in \Lambda$ 
   Output: A pair of partial configurations  $\sigma'$  and  $\tau'$  over some  $\Lambda' \supseteq \Lambda$ 
2  $\sigma' \leftarrow \sigma, \tau' \leftarrow \tau$ ;
3 while  $S_R(u) \setminus \Lambda \neq \emptyset$  do
4   Choose  $v$  from  $S_R(u) \setminus \Lambda$  uniformly at random;
5   Draw  $(c_1, c_2)$  from the optimal coupling between  $\mu_v^{\sigma'}$  and  $\mu_v^{\tau'}$ ;
6    $\sigma' \leftarrow \sigma'_{v \leftarrow c_1}, \tau' \leftarrow \tau'_{v \leftarrow c_2}$ ;
7    $\Lambda \leftarrow \Lambda \cup \{v\}$ ;
8   if  $c_1 \neq c_2$  then
9     return  $(\sigma', \tau')$ ;
10 return  $(\sigma', \tau')$ ;

```

---

Note that our algorithm does not really need to construct the partial coupling described in Algorithm 1. Instead, we construct linear programs that mimic the coupling process. We start by defining the coupling tree  $\mathcal{T}$  with root  $rt$ . This tree essentially enumerates all possible intermediate states of Algorithm 1. Formally the coupling tree is constructed by Algorithm 2.

Denote by  $\mathcal{V}(\mathcal{T})$  the set of nodes of  $\mathcal{T}$  and  $\mathcal{L}(\mathcal{T})$  the set of leaves. Each node  $w$  of the tree represents an intermediate state of Algorithm 1. Its label  $(\sigma, \tau, \Lambda, D)$  represents the two partial configurations  $\sigma$  and  $\tau$ , the pinned vertex set  $\Lambda$ , and the set  $D$  of differing vertices other than the initial disagreement  $u$ . In fact,  $D$  can only be either some vertex  $v$  or the empty set  $\emptyset$ , because Algorithm 1 stops whenever the first disagreement other than  $u$  is introduced. Thus, if  $D \neq \emptyset$ , then the node must be a leaf. We call a leaf node *good* if its  $T = \emptyset$ . Namely, denote by  $\mathcal{GL}(\mathcal{T}) = \{w \in$

---

**Algorithm 2:** Construction of the coupling tree
 

---

```

1 Coupling-Tree  $(\sigma, \tau)$ ;
   Input      : Partial configurations  $\sigma, \tau \in [q]^\Lambda$  that only differ at a vertex  $u \in \Lambda$ 
   Output     : A coupling tree  $\mathcal{T}$  with its root  $rt$ 
   Parameter : A positive integer  $R$ 
2 Construct a tree  $\mathcal{T}$  containing a single root node  $rt$  with  $\text{label}(rt) \leftarrow (\sigma, \tau, \Lambda, \emptyset)$ ;
3 for  $v \in S_R(u) \setminus \Lambda$  do
4   for  $(c_1, c_2) \in \text{supp}(\mu_v^\sigma) \times \text{supp}(\mu_v^\tau)$  do
5     if  $c_1 = c_2 = c$  then
6        $(\mathcal{T}_{v,c}, rt_{v,c}) \leftarrow \text{Coupling-Tree}(\sigma_{v \leftarrow c}, \tau_{v \leftarrow c})$ ;
7       Append  $\mathcal{T}_{v,c}$  to  $rt$  by connecting  $rt$  and  $rt_{v,c}$ ;
8     else
9       Introduce a leaf node  $w$  with  $\text{label}(w) \leftarrow (\sigma_{v \leftarrow c_1}, \tau_{v \leftarrow c_2}, \Lambda \cup \{v\}, v)$ ;
10      Connect  $rt$  and  $w$ ;
11 return  $(\mathcal{T}, rt)$ ;

```

---

$\mathcal{L}(\mathcal{T}) \mid \text{label}(w) = (*, *, *, \emptyset)$  the set of good leaves, where  $*$  denotes any possible value at that position. The rest of the leaves  $\mathcal{BL}(\mathcal{T}) = \mathcal{L}(\mathcal{T}) \setminus \mathcal{GL}(\mathcal{T})$  are *bad*. Moreover, denote by  $\mathcal{C}(w)$  the set of children of a node  $w$ .

The linear program is introduced in Algorithm 3. For each  $w \in \mathcal{V}(\mathcal{T})$  whose label is  $(\sigma^w, \tau^w, \Lambda^w, *)$ , we call  $r_w := \frac{\mu_{\Lambda^w}(\sigma^w)}{\mu_{\Lambda^w}(\tau^w)}$  its marginal ratio. The goal of the LP is to estimate the marginal ratio  $r = r_{rt} = \frac{\mu_\Lambda(\sigma)}{\mu_\Lambda(\tau)}$  for the root. The LP contains two variables  $x_w$  and  $y_w$  for each node  $w \in \mathcal{V}(\mathcal{T})$  of the coupling tree. Intuitively,  $x_w$  represents  $\frac{z_w \mu_\Lambda(\sigma)}{\mu_{\Lambda^w}(\sigma^w)}$ ,  $y_w$  represents  $\frac{z_w \mu_\Lambda(\tau)}{\mu_{\Lambda^w}(\tau^w)}$ , and  $z_w$  is the probability that the coupling reaches the node  $w$ . See the proof of Lemma 15 where we verify that these values do satisfy the constraints. The choice of the variables are standard to apply Moitra's method [Moi19]. The main change from a typical application of [Moi19] is

- (1) the leaf constraints now allow some errors, and the marginal ratios of the leaves are considered as (recursively given) inputs;
- (2) the overflow constraints are a new kind of constraints to control the error.

To construct the LP, we need the coupling tree  $\mathcal{T}$ , as well as the marginal ratios for all the leaf nodes of  $\mathcal{T}$ . These ratios will be recursively fed and are denoted  $\mathbf{R}$ . We will combine the LP with a binary search to find an estimate to  $r$ , and the other two inputs  $r_-$  and  $r_+$  are our current guesses of the upper and lower bounds of  $r$ . Moreover,  $\ell(w)$  is defined to be  $|S_R(u) \setminus \Lambda^w|$  for a node  $w \in \mathcal{V}(\mathcal{T})$  with label  $(*, *, \Lambda^w, *)$ . The parameter  $\eta$  will be eventually set to

$$(11) \quad \eta := b^{-1} \delta(R).$$

With the coupling tree  $\mathcal{T}$  for  $(\sigma, \tau)$  (in Algorithm 2) and the linear program (in Algorithm 3) in hand, we then estimate the marginal ratio  $r = \frac{\mu_\Lambda(\sigma)}{\mu_\Lambda(\tau)}$  by the binary search mentioned above. We formally state this binary search in Algorithm 4.

The following bound is the main guarantee of the vector  $\mathbf{R}$ .

**Condition 13** ( $\varepsilon$ -error bound). Let  $\varepsilon > 0$  be a parameter. Let  $\mu$  be a Gibbs distribution on  $[q]^V$ . Let  $\mathcal{T}$  be the coupling tree of  $(\sigma, \tau)$ , where  $\sigma, \tau \in [q]^\Lambda$  are two partial configurations on  $\Lambda \subseteq V$  that only differ at some vertex  $u \in \Lambda$ . Then, let  $\mathbf{R} \in \mathbb{R}_{\geq 0}^{\mathcal{L}(\mathcal{T})}$  be a vector defined on the leaves of  $\mathcal{T}$ . For any leaf node  $w \in \mathcal{L}(\mathcal{T})$  with  $\text{label}(w) = (\sigma^w, \tau^w, \Lambda^w, *)$ ,

---

**Algorithm 3:** The linear program
 

---

1 **LP**  $(r_-, r_+, \mathcal{T}, \mathbf{R}, \varepsilon)$ ;

**Input** : Positive real values  $r_- \leq r_+$ , a coupling tree  $\mathcal{T}$ , marginal ratio estimates  $\mathbf{R} \in \mathbb{R}_{>0}^{\mathcal{L}(\mathcal{T})}$ , and their error margin  $\varepsilon$

**Output** : A Boolean value, indicating whether the LP has a feasible solution

**Parameter** :  $\eta > 0$

2 **return** true if and only if the following LP has a feasible solution

(1) Validity constraints:

$$\forall w \in \mathcal{V}(\mathcal{T}), \quad x_w, y_w \geq 0$$

$$x_{rt} = y_{rt} = 1$$

(2) Recursive constraints:

For any non-leaf node  $w$  with  $\text{label}(w) = (\sigma^w, \tau^w, \Lambda^w, \emptyset)$ , and  $v \in S_R(u) \setminus \Lambda^w$ ,

$$\forall c \in \text{supp}(\mu_v^{\sigma^w}), \quad \sum_{\substack{w' \in \mathcal{C}(w): \\ \text{label}(w') = (\sigma_{v \leftarrow c}^w, *, *, *)}} x_{w'} = \frac{x_w}{\ell(w)}$$

$$\forall c \in \text{supp}(\mu_v^{\tau^w}), \quad \sum_{\substack{w' \in \mathcal{C}(w): \\ \text{label}(w') = (*, \tau_{v \leftarrow c}^w, *, *)}} y_{w'} = \frac{y_w}{\ell(w)}$$

(3) Leaf constraints:

$$\forall w \in \mathcal{GL}(\mathcal{T}), \quad r_- R_w^{-1} y_w \leq x_w \leq r_+ R_w^{-1} y_w$$

$$\forall w \in \mathcal{BL}(\mathcal{T}), \quad r_- (1 + \varepsilon)^{-1} R_w^{-1} y_w \leq x_w \leq r_+ (1 + \varepsilon) R_w^{-1} y_w$$

(4) Overflow constraints:

$$\forall w \in \mathcal{V}(\mathcal{T}) \setminus \mathcal{L}(\mathcal{T}), \quad \sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} x_{w'} \leq \frac{\eta}{\ell(w)} x_w$$

$$\forall w \in \mathcal{V}(\mathcal{T}) \setminus \mathcal{L}(\mathcal{T}), \quad \sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} y_{w'} \leq \frac{\eta}{\ell(w)} y_w$$


---

- if  $w \in \mathcal{GL}(\mathcal{T})$ ,  $R_w = \frac{\mu_{\Lambda^w}(\sigma^w)}{\mu_{\Lambda^w}(\tau^w)}$ ;
- if  $w \in \mathcal{BL}(\mathcal{T})$ , it holds that

$$(1 + \varepsilon)^{-1} \leq R_w \cdot \frac{\mu_{\Lambda^w}(\tau^w)}{\mu_{\Lambda^w}(\sigma^w)} \leq (1 + \varepsilon).$$

When Condition 13 holds, we say  $\mathbf{R}$  satisfies the  $\varepsilon$ -error bound.

Our marginal estimator (Algorithm 4) takes  $(\mathcal{T}, \mathbf{R}, \varepsilon)$  as input and outputs an estimate  $\hat{r}$  of  $r$ . The key property of this estimator is that if  $\mathbf{R}$  satisfies the  $\varepsilon$ -error bound, then the error of  $\hat{r}$  shrinks by a constant factor.

**Lemma 14.** *Let  $(\mathcal{T}, \mathbf{R}, \varepsilon)$  be the input of Algorithm 4 such that  $\mathbf{R}$  satisfies the  $\varepsilon$ -error bound (namely Condition 13 holds) for some  $\varepsilon \leq 3b^{-2}$ . Then the output  $\hat{r}$  of Algorithm 4 satisfies*

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

where  $r = \frac{\mu_{\Lambda}(\sigma)}{\mu_{\Lambda}(\tau)}$  and  $\hat{\varepsilon} := 5b^{-2}\eta H(\Delta^R) \cdot \varepsilon$ .

---

**Algorithm 4:** Marginal estimation based on LP
 

---

1 Marginal-estimator  $(\mathcal{T}, \mathbf{R}, \varepsilon)$

**Input** : A coupling tree  $\mathcal{T}$  with root  $rt$ , marginal ratio estimates  $\mathbf{R} \in \mathbb{R}_{>0}^{\mathcal{L}(\mathcal{T})}$  and their error margin  $\varepsilon$

**Output** : An estimate  $\hat{r}$

**Parameters:** An integer  $R > 0$  and a real number  $\eta > 0$

2  $r_{low} \leftarrow b$  and  $r_{upp} \leftarrow b^{-1}$ ;

3  $\hat{\varepsilon} \leftarrow \eta H(\Delta^R) \cdot \varepsilon$ ;

4 **while**  $r_{upp} > (1 + \hat{\varepsilon})^2 r_{low}$  **do**

5     Let  $m \leftarrow (r_{upp} + r_{low})/2$ ;

6     **if** both  $LP(r_{low}, m, \mathcal{T}, \mathbf{R}, \varepsilon)$  and  $LP(m, r_{upp}, \mathcal{T}, \mathbf{R}, \varepsilon)$  are true **then**

7         **return**  $m$ ;

   // Assertion: otherwise either  $LP(r_{low}, m, \mathcal{T}, \mathbf{R}, \varepsilon)$  or  $LP(m, r_{upp}, \mathcal{T}, \mathbf{R}, \varepsilon)$  is true

8     **if**  $LP(r_{low}, m, \mathcal{T}, \mathbf{R}, \varepsilon)$  is true **then**

9          $r_{upp} \leftarrow m$ ;

10    **else**

11          $r_{low} \leftarrow m$ ;

12 **return**  $\hat{r} = \sqrt{r_{low} r_{upp}}$ ;

---

Note that our choices of  $R$  in (10) and  $\eta$  in (11) are stronger than requiring  $5b^{-2}\eta H(\Delta^R) < 1$ . This is because in the full algorithm, we need to rewrite each  $R_w$  into a product of two marginal ratios so that the recursion can continue. Thus, at each recursion step, the error first increases because of this product, and then shrinks by Lemma 14.

Lemma 14 is a direct consequence of the following two lemmas.

**Lemma 15.** *Suppose Condition 13 holds. If the input  $r_-$  and  $r_+$  to Algorithm 3 satisfies  $r_- \leq r \leq r_+$ , then the LP is feasible.*

**Lemma 16.** *Suppose Condition 13 holds with  $\varepsilon \leq 3b^{-2}$ . If Algorithm 3 returns true for parameters  $r_-$  and  $r_+$ , then*

$$(1 + \hat{\varepsilon})^{-1} r_- \leq r \leq (1 + \hat{\varepsilon}) r_+.$$

The proofs of Lemma 15 and Lemma 16 are deferred to the next section, Section 4.2.

*Proof of Lemma 14.* First suppose the binary search terminates early in Line 6 of Algorithm 4. In this case, both  $LP(r_{low}, m, \mathcal{T}, \mathbf{R}, \varepsilon)$  and  $LP(m, r_{upp}, \mathcal{T}, \mathbf{R}, \varepsilon)$  return true. Lemma 16 implies that  $r \leq (1 + \hat{\varepsilon})m$  and  $r \geq (1 + \hat{\varepsilon})^{-1}m$ . Thus  $m$  satisfies the desired approximation bound.

Otherwise, only one of the two LPs is feasible until the while loop ends. By induction  $r \in [r_{low}, r_{upp}]$  throughout the while loop due to Lemma 15. The binary search exits the while loop when  $r_{upp} \leq (1 + \hat{\varepsilon})^2 r_{low}$ . Thus the output satisfies the desired approximation bound.  $\square$

**4.2. Error analysis of the LP.** In this section, we prove Lemma 15 and Lemma 16.

*Proof of Lemma 15.* For each node  $w \in \mathcal{V}(\mathcal{T})$  with  $\text{label}(w) = (\sigma^w, \tau^w, \Lambda^w, *)$ , let  $z_w$  be the probability that Algorithm 1 reaches the node  $w$  in the coupling tree. Furthermore, let  $x_w = \frac{z_w \mu_\Lambda(\sigma)}{\mu_{\Lambda^w}(\sigma^w)}$  and  $y_w = \frac{z_w \mu_\Lambda(\tau)}{\mu_{\Lambda^w}(\tau^w)}$ . It suffices to verify that  $(\mathbf{x}, \mathbf{y})$  is a feasible solution to the LP. We verify the four sets of constraints one by one.

- Validity constraints: it is obvious that  $x_w, y_w \geq 0$  and  $x_{rt} = y_{rt} = z_{rt} = 1$ .
- Recursive constraints: By symmetry, we only verify the first set. For each non-leaf node  $w \in \mathcal{V}(\mathcal{T})$  with  $\text{label}(w) = (\sigma^w, \tau^w, \Lambda^w, \emptyset)$ , it holds that

$$\forall v \in S_R(u) \setminus \Lambda^w, c \in \text{supp}(\mu_v^{\sigma^w}), \quad \sum_{\substack{w' \in \mathcal{C}(w): \\ \text{label}(w') = (\sigma_{v \leftarrow c}^w, *, *, *)}} z_{w'} = z_w \cdot \frac{1}{\ell(w)} \cdot \frac{\mu_{\Lambda^w \cup \{v\}}(\sigma_{v \leftarrow c}^w)}{\mu_{\Lambda^w}(\sigma^w)}.$$

The left hand side is the total probability of reaching  $w' \in \mathcal{C}(w)$  whose first label is  $\sigma_{v \leftarrow c}^w$ . This can only happen by first reaching  $w$ , with probability  $z_w$ , and randomly chosen  $v$ , with probability  $\frac{1}{\ell(w)}$ . Then, as Line 5 of Algorithm 1 is a valid coupling, the probability of getting  $\sigma_{v \leftarrow c}^w$  is  $\frac{\mu_{\Lambda^w \cup \{v\}}(\sigma_{v \leftarrow c}^w)}{\mu_{\Lambda^w}(\sigma^w)}$ . This is exactly the right hand side. The recursive constraints then hold for  $x_w = \frac{z_w \mu_{\Lambda}(\sigma)}{\mu_{\Lambda^w}(\sigma^w)}$ .

- Leaf constraints: For each leaf node  $w \in \mathcal{L}(\mathcal{T})$  with  $\text{label}(w) = (\sigma^w, \tau^w, \Lambda^w, *)$ , we have

$$\frac{x_w}{y_w} = r \cdot \frac{\mu_{\Lambda^w}(\tau^w)}{\mu_{\Lambda^w}(\sigma^w)}.$$

By Condition 13, the term  $\frac{\mu_{\Lambda^w}(\tau^w)}{\mu_{\Lambda^w}(\sigma^w)}$  is either exactly  $R_w^{-1}$  (when  $w$  is a good leaf) or approximated by  $R_w^{-1}$  up to  $1 + \varepsilon$  relative error (when  $w$  is bad). As  $r_- \leq r \leq r_+$ , the leaf constraints hold.

- Overflow constraints: By symmetry, we only verify the first set. For each non-leaf node  $w \in \mathcal{V}(\mathcal{T})$ ,

$$(12) \quad \sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} z_{w'} \leq \frac{b\eta}{\ell(w)} z_w.$$

This is because by Definition 11, the probability that Algorithm 1 reveals a disagreement is bounded by  $\frac{\delta(R)}{\ell(w)} = \frac{b\eta}{\ell(w)}$ . Note that for any  $w' \in \mathcal{C}(w)$ ,  $\Lambda^{w'}$  takes the form  $\Lambda^w \cup \{v\}$  for some  $v \in S_R(u) \setminus \Lambda^w$ , and  $\sigma^{w'}$  takes the form  $\sigma_{v \leftarrow c}^w$  for some  $c \in \text{supp}(\sigma_v^w)$ . By Definition 3, the definition of the marginal lower bound  $b$ , we have

$$\frac{\mu_{\Lambda^{w'}}(\sigma^{w'})}{\mu_{\Lambda^w}(\sigma^w)} = \frac{\mu_{\Lambda^w \cup \{v\}}(\sigma_{v \leftarrow c}^w)}{\mu_{\Lambda^w}(\sigma^w)} \geq b.$$

Therefore, by (12),

$$\begin{aligned} \sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} x_{w'} &= \sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} \frac{z_{w'} \mu_{\Lambda}(\sigma)}{\mu_{\Lambda^{w'}}(\sigma^{w'})} \leq b^{-1} \sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} \frac{z_{w'} \mu_{\Lambda}(\sigma)}{\mu_{\Lambda^w}(\sigma^w)} \\ &\leq \frac{\eta}{\ell(w)} \cdot \frac{z_w \mu_{\Lambda}(\sigma)}{\mu_{\Lambda^w}(\sigma^w)} = \frac{\eta}{\ell(w)} x_w. \quad \square \end{aligned}$$

From the proof above, one can observe that the overflow constraints are weaker than what the values we plug in for  $x_w$  and  $y_w$  really satisfy. However, to write the exact constraints these values satisfy, some marginal probabilities are required. There appears to be no efficient way to compute these marginal probabilities. Thus, we choose to use the marginal lower bound to enforce a weaker set of constraints.

Next we show Lemma 16. Suppose there exists a solution, say,  $(\mathbf{x}, \mathbf{y})$ , to the linear program. For each  $1 \leq i \leq \ell$  where  $\ell = |S_R(u) \setminus \Lambda|$ , define

$$\Gamma_{x,i} := \sum_{w \in \mathcal{BL}(\mathcal{T}): \ell(w)=i-1} \mu_{\Lambda^w}(\sigma^w) x_w,$$

where we use  $\Lambda^w$  and  $\sigma^w$  to denote the corresponding labels for  $w$ . Intuitively,  $\Gamma_{x,i}$  is the sum of  $\mu_{\Lambda^w}(\sigma^w)x_w$  over all bad leaves of a certain depth. Furthermore, let

$$\Gamma_{x,0} := \sum_{w \in \mathcal{GL}(\mathcal{T})} \mu_{\Lambda^w}(\sigma^w)x_w.$$

Similarly, we define  $\Gamma_{y,i}$  for  $0 \leq i \leq \ell$  by replacing  $x_w$  with  $y_w$ . Before proving Lemma 16, we derive some basic properties of  $\Gamma_{\cdot,\cdot}$ .

**Lemma 17.** *Assuming Condition 13, the following holds:*

- (1)  $\sum_{i=0}^{\ell} \Gamma_{x,i} = \mu_{\Lambda}(\sigma)$  and  $\sum_{i=0}^{\ell} \Gamma_{y,i} = \mu_{\Lambda}(\tau)$ ;
- (2) For all  $1 \leq i \leq \ell$ ,  $\Gamma_{x,i} \leq \frac{\eta}{i} \sum_{j=0}^i \Gamma_{x,j}$  and  $\Gamma_{y,i} \leq \frac{\eta}{i} \sum_{j=0}^i \Gamma_{y,j}$ ;
- (3)  $r_- \Gamma_{y,0} \leq \Gamma_{x,0} \leq r_+ \Gamma_{y,0}$ ;
- (4) For any  $1 \leq i \leq \ell$ ,  $(1 + \varepsilon)^{-2} r_- \Gamma_{y,i} \leq \Gamma_{x,i} \leq (1 + \varepsilon)^2 r_+ \Gamma_{y,i}$ .

*Proof.* Note that for any non-leaf node  $w \in \mathcal{V}(\mathcal{T})$ ,

$$\begin{aligned} \sum_{w' \in \mathcal{C}(w)} \mu_{\Lambda^{w'}}(\sigma^{w'})x_{w'} &= \sum_{v \in S_R(u) \setminus \Lambda^w} \sum_{c \in \text{supp}(\mu_v^{\sigma^w})} \sum_{\substack{w' \in \mathcal{C}(w): \\ \text{label}(w') = (\sigma_{v \leftarrow c}^w, *, *, *)}} \mu_{\Lambda^{w \cup \{v\}}(\sigma_{v \leftarrow c}^w)} x_{w'} \\ &= \sum_{v \in S_R(u) \setminus \Lambda^w} \sum_{c \in \text{supp}(\mu_v^{\sigma^w})} \frac{\mu_{\Lambda^{w \cup \{v\}}(\sigma_{v \leftarrow c}^w)} x_w}{\ell(w)} \\ (13) \quad &= \sum_{v \in S_R(u) \setminus \Lambda^w} \frac{\mu_{\Lambda^w}(\sigma^w)x_w}{\ell(w)} = \mu_{\Lambda^w}(\sigma^w)x_w, \end{aligned}$$

where the second equality follows from recursive constraints of  $\mathbf{x}$ , and the fourth equality follows from the definition of  $\ell(w)$ . Recursively applying (13) gives us that

$$\sum_{i=0}^{\ell} \Gamma_{x,i} = \mu_{\Lambda}(\sigma)x_{rt} = \mu_{\Lambda}(\sigma),$$

as  $x_{rt}$  is set to 1 by the LP. A similar proof works for  $\Gamma_{y,\cdot}$ . The first item holds.

For any non-leaf node  $w \in \mathcal{V}(\mathcal{T})$  and  $w' \in \mathcal{C}(w)$ ,  $\sigma^{w'}$  must have the form  $\sigma_{v_{w'} \leftarrow c_{w'}}^{w'}$  for some  $v_{w'} \in S_R(u) \setminus \Lambda^w$  and  $c_{w'} \in \text{supp}(\mu_{v_{w'}}^{\sigma^w})$ , and  $\Lambda^{w'}$  must be  $\Lambda^w \cup \{v_{w'}\}$ . Thus,

$$\sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} \mu_{\Lambda^{w \cup \{v_{w'}\}}(\sigma_{v_{w'} \leftarrow c_{w'}}^{w'})} x_{w'} \leq \sum_{w' \in \mathcal{C}(w) \cap \mathcal{BL}(\mathcal{T})} \mu_{\Lambda^w}(\sigma^w)x_{w'} \leq \frac{\eta}{\ell(w)} \mu_{\Lambda^w}(\sigma^w)x_w,$$

where the first inequality follows from the fact that  $\mu_{\Lambda^{w \cup \{v\}}(\sigma_{v \leftarrow c}^w)} \leq \mu_{\Lambda^w}(\sigma^w)$  for any  $v \in S_R(u) \setminus \Lambda^w$  and  $c \in \text{supp}(\mu_v^{\sigma^w})$ , and the second inequality follows from the overflow constraints in Algorithm 3. Also notice that  $\ell(w) = \ell(w') + 1$ . It implies

$$\Gamma_{x,i} \leq \frac{\eta}{i} \sum_{w \in \mathcal{V}(\mathcal{T}) \setminus \mathcal{L}(\mathcal{T}): \ell(w)=i} \mu_{\Lambda^w}(\sigma^w)x_w.$$

On the other hand, recursively applying (13), we have

$$\sum_{j=0}^i \Gamma_{x,j} = \sum_{w \in \mathcal{V}(\mathcal{T}) \setminus \mathcal{L}(\mathcal{T}): \ell(w)=i} \mu_{\Lambda^w}(\sigma^w)x_w.$$

A similar proof works for  $\Gamma_{y,\cdot}$ . The second item follows.

For the last two items, by the leaf constraints and Condition 13,

$$\begin{aligned}\Gamma_{x,0} &= \sum_{w \in \mathcal{GL}(\mathcal{T})} \mu_{\Lambda^w}(\sigma^w) x_w \leq \sum_{w \in \mathcal{GL}(\mathcal{T})} \mu_{\Lambda^w}(\sigma^w) r_+ \cdot \frac{\mu_{\Lambda^w}(\tau^w)}{\mu_{\Lambda^w}(\sigma^w)} \cdot y_w = r_+ \Gamma_{y,0}; \\ \Gamma_{x,i} &= \sum_{w \in \mathcal{BL}(\mathcal{T}): \ell(w)=i-1} \mu_{\Lambda^w}(\sigma^w) x_w \leq (1+\varepsilon)^2 \sum_{w \in \mathcal{BL}(\mathcal{T}): \ell(w)=i-1} \mu_{\Lambda^w}(\sigma^w) r_+ \cdot \frac{\mu_{\Lambda^w}(\tau^w)}{\mu_{S^w}(\sigma^w)} \cdot y_w \\ &= (1+\varepsilon)^2 r_+ \Gamma_{y,i}.\end{aligned}$$

Similarly,  $\Gamma_{x,0} \geq r_- \Gamma_{y,0}$  and  $\Gamma_{x,i} \geq (1+\varepsilon)^{-2} r_- \Gamma_{y,i}$ .  $\square$

With the help of Lemma 17, we are now ready to give the proof of Lemma 16.

*Proof of Lemma 16.* Let  $\varepsilon > 0$  be input to Algorithm 3. For  $0 \leq \ell \leq \Delta^R$ , define

$$(14) \quad \varepsilon_\ell := 5b^{-2} \eta H(\ell) \cdot \varepsilon,$$

where the function  $H(\ell) = \sum_{i=1}^{\ell} \frac{1}{i}$  is the harmonic sum. Clearly  $\varepsilon_\ell$  is increasing in  $\ell$ .

To prove Lemma 16, by Item (1) of Lemma 17, we only need to show that

$$(15) \quad (1+\varepsilon_\ell)^{-1} r_- \leq \frac{\sum_{i=0}^{\ell} \Gamma_{x,i}}{\sum_{i=0}^{\ell} \Gamma_{y,i}} \leq (1+\varepsilon_\ell) r_+.$$

We do an induction on  $\ell$ . By Item (3) in Lemma 17, (15) holds for  $\ell = 0$ . Now assume (15) holds for  $\ell - 1$ . We prove the upper bound first. By induction hypothesis,

$$\sum_{i=0}^{\ell-1} \Gamma_{x,i} \leq (1+\varepsilon_{\ell-1}) r_+ \sum_{i=0}^{\ell-1} \Gamma_{y,i}$$

Furthermore, by Item (4) of Lemma 17,  $\Gamma_{x,\ell} \leq (1+\varepsilon)^2 r_+ \Gamma_{y,\ell}$ . Therefore,

$$\sum_{i=0}^{\ell} \Gamma_{x,i} \leq r_+ \left( (1+\varepsilon_{\ell-1}) \sum_{i=0}^{\ell-1} \Gamma_{y,i} + (1+\varepsilon)^2 \Gamma_{y,\ell} \right)$$

We claim that

$$(1+\varepsilon_{\ell-1}) \sum_{i=0}^{\ell-1} \Gamma_{y,i} + (1+\varepsilon)^2 \Gamma_{y,\ell} \leq (1+\varepsilon_\ell) \sum_{i=0}^{\ell} \Gamma_{y,i},$$

which would finish the proof. Our choice of parameters satisfy (10) and (11), which implies that

$$\varepsilon_{\ell-1} < \varepsilon_\ell \leq \eta H(\Delta^R) \varepsilon < \varepsilon.$$

Thus, the claim is equivalent to

$$\Gamma_{y,\ell} \leq \frac{\varepsilon_\ell - \varepsilon_{\ell-1}}{2\varepsilon + \varepsilon^2 - \varepsilon_\ell} \sum_{i=0}^{\ell} \Gamma_{y,i}.$$

As  $\varepsilon \leq 3b^{-2}$  and  $0 < b < 1$ ,  $2\varepsilon + \varepsilon^2 - \varepsilon_\ell \leq 5b^{-2}\varepsilon$ , we just need to show

$$\Gamma_{y,\ell} \leq \frac{\varepsilon_\ell - \varepsilon_{\ell-1}}{5b^{-2}\varepsilon} \sum_{i=0}^{\ell} \Gamma_{y,i} = \frac{\eta}{\ell} \sum_{i=0}^{\ell} \Gamma_{y,i}.$$

This is just Item (2) of Lemma 17.

The lower bound in (15) holds by a similar argument. Specifically, we can flip the role of  $x$  and  $y$  and replace  $r_+$  with  $\frac{1}{r_-}$  in the proof above to show the lower bound.  $\square$



4.3. **The full algorithm.** With the Lemma 14 in hand, we are now able to construct the full algorithm for Theorem 4, which is a recursive marginal estimator. It uses Algorithm 2 and Algorithm 4 recursively and do a truncation at some depth  $k$ . Recall that we have set the parameter  $R$  to satisfy (10) and  $\eta = b^{-1}\delta(R)$  in (11). Moreover, set the error parameter  $\varepsilon(k)$  of depth  $k$  to be

$$\varepsilon(k) := \begin{cases} b^{-1}, & \text{if } k = 0; \\ 2^{-k+1}, & \text{if } k \geq 1. \end{cases}$$

The full algorithm is given in Algorithm 5.

---

**Algorithm 5:** Recursive marginal estimator

---

```

1 Recursive-estimator ( $\sigma, \tau, \Lambda, k, u$ );
   Input      : Partial configurations  $\sigma, \tau \in [q]^\Lambda$  that only differ at a vertex  $u \in \Lambda$ , and
                  $k \in \mathbb{N}_{\geq 0}$ 
   Output    : An estimate  $\tilde{r}$ 
   Parameters: A positive integer  $R$ , a positive real  $\eta$ , and a function  $\varepsilon(\cdot)$ 
2  $\mathcal{T} \leftarrow \text{Coupling-Tree}(\sigma, \tau)$ ;
3 if  $k = 0$  then
4   return 1;
5  $\ell \leftarrow |S_R(u) \setminus \Lambda|$ ;
6 if  $\ell = 0$  then
7   return  $r = \frac{\mu_\Lambda(\sigma)}{\mu_\Lambda(\tau)}$ , computed by brute force;
   // Calculate the marginal ratio of the leaves in  $\mathcal{L}(\mathcal{T})$  via recursion
8 for  $w \in \mathcal{L}(\mathcal{T})$  do
9    $(\sigma^w, \tau^w, \Lambda^w, D^w) \leftarrow \text{label}(w)$ ;
10  if  $w \in \mathcal{GL}(\mathcal{T})$  then
11     $R_w \leftarrow$  the marginal ratio of  $r = \frac{\mu_{\Lambda^w}(\sigma^w)}{\mu_{\Lambda^w}(\tau^w)}$  by brute force;
12  else
13     $v \leftarrow D^w$ ;
14     $\gamma^w \leftarrow \sigma_{u \leftarrow \tau^w}^w$ ;
15     $X \leftarrow \text{Recursive-estimator}(\sigma^w, \gamma^w, \Lambda^w, k - 1, u)$ ;
16     $Y \leftarrow \text{Recursive-estimator}(\gamma^w, \tau^w, \Lambda^w, k - 1, v)$ ;
17     $R_w \leftarrow X \cdot Y$ ;
18 return Marginal-estimator ( $\mathcal{T}, \mathbf{R}, 2\varepsilon(k - 1) + \varepsilon(k - 1)^2$ );

```

---

Next we analyse the accuracy and the running time of Algorithm 5.

**Lemma 18** (accuracy). *Let  $\tilde{r}$  be the estimation given by Algorithm 5, it holds that for  $k \geq 0$ ,*

$$(1 + \varepsilon(k))^{-1} \leq \tilde{r}/r \leq 1 + \varepsilon(k).$$

*Proof.* We prove by induction on  $k$ . When  $k = 0$ , Algorithm 5 simply returns 1. Let  $\gamma = \sigma_{\Lambda \setminus \{u\}}$  be the common configuration between  $\sigma$  and  $\tau$ . Then

$$\frac{\mu_\Lambda(\sigma)}{\mu_\Lambda(\tau)} = \frac{\mu_\Lambda(\sigma)}{\mu_{\Lambda \setminus \{u\}}(\gamma)} \cdot \frac{\mu_{\Lambda \setminus \{u\}}(\gamma)}{\mu_\Lambda(\tau)} \leq b^{-1}.$$

Similarly,  $\frac{\mu_\Lambda(\sigma)}{\mu_\Lambda(\tau)} \geq b$ . Since  $\varepsilon(0) = b^{-1}$ , the base case holds.

When  $k \geq 1$ , by the induction hypothesis, we know that  $X$  and  $Y$  obtained at Line 15 and Line 16 in Algorithm 5 are estimates to  $\frac{\mu_{\Lambda^w}(\sigma^w)}{\mu_{\Lambda^w}(\gamma^w)}$  and  $\frac{\mu_{\Lambda^w}(\gamma^w)}{\mu_{\Lambda^w}(\tau^w)}$ , respectively, both with  $(1 + \varepsilon(k - 1))$  relative error. Note that here  $\mu_{\Lambda^w}(\gamma^w) > 0$ . This is because  $\gamma^w$  is locally feasible by its definition, and as the spin system is permissive (recall Definition 9), by (5),  $\gamma^w$  is globally feasible. Since  $R_w = X \cdot Y$ , this implies that the relative error of  $R_w$  to  $\frac{\mu_{\Lambda^w}(\sigma^w)}{\mu_{\Lambda^w}(\tau^w)}$  is

$$(1 + \varepsilon(k - 1))^2 = 1 + 2\varepsilon(k - 1) + \varepsilon(k - 1)^2,$$

which means that  $\mathbf{R}$  satisfies a  $(2\varepsilon(k - 1) + \varepsilon(k - 1)^2)$ -error bound (recall Condition 13).

For  $k = 1$ ,  $(2\varepsilon(0) + \varepsilon(0)^2) \leq 3b^{-2}$  and thus Lemma 14 applies. The lemma holds because in this case, by our choice of parameters in (10) and (11),

$$5b^{-2}\eta H(\Delta^R)(2\varepsilon(0) + \varepsilon(0)^2) \leq 15b^{-4}\eta H(\Delta^R) < 1 = \varepsilon(1).$$

For  $k \geq 2$ , by (10) and (11), again,

$$5b^{-2}\eta H(\Delta^R)(2\varepsilon(k - 1) + \varepsilon(k - 1)^2) \leq 15b^{-2}\eta H(\Delta^R)\varepsilon(k - 1) < \varepsilon(k - 1)/2 = \varepsilon(k).$$

Thus the lemma follows from Lemma 14 as well.  $\square$

Now, we can finish the proof of Theorem 4.

*Proof of Theorem 4.* First note that the assumptions of the marginal lower bound and coupling independence both hold with respect to an arbitrary conditioning. Thus, by standard self-reduction [JVV86], approximating  $Z$  up to relative error  $1 + \varepsilon$  can be reduced to approximate  $n$  marginal probabilities up to relative error  $1 + \frac{\varepsilon}{n}$ . The latter task then can be reduced to approximate  $q - 1$  marginal ratios up to the same relative error  $1 + \frac{\varepsilon}{n}$ , for which we use Algorithm 5. Suppose the running time of Algorithm 5 with relative error  $1 + \frac{\varepsilon}{n}$  is  $T_{\varepsilon/n}$ , then the overall running time is  $O(qnT_{\varepsilon/n})$ .

By Lemma 18, to achieve relative error  $1 + \frac{\varepsilon}{n}$ , we only need to pick  $k = \lceil \log_2(n/\varepsilon) \rceil + 1$ . Denote

- by  $T$  an upper bound of the cost of Algorithm 2 and Algorithm 4 (i.e., the max cost of a single recursive step);
- and by  $N$  maximum number of the nodes of the coupling tree.

Clearly the total cost of Algorithm 5 is bounded by  $T \cdot \sum_{i=0}^k (2N)^i \leq 2T(2N)^k$ . We claim that

$$T = O(k + \log b^{-1}) \cdot \text{poly}(N) \quad \text{and} \quad N \leq 2(q^2 \Delta^R)^{\Delta^R},$$

which implies

$$T_{\varepsilon/n} \leq 2T \cdot (2N)^k = \log b^{-1}(n/\varepsilon) O(\Delta^R \log(q \Delta^R)).$$

By Theorem 12, we can take  $\delta(R) = 2C2^{-\lceil R/2C \rceil}$ , where  $C$  is the coupling independence parameter. To satisfy (10),  $R = O(C(\log b^{-1} + \log C + \log \log \Delta))$  suffices. Thus the bound on the running time in terms of  $b, C, q, \Delta$  is

$$T_{\varepsilon/n} = \left(\frac{n}{\varepsilon}\right)^{\Delta^{O(C(\log b^{-1} + \log C + \log \log \Delta))} \log q}.$$

To show the claim, for  $T$ , it is direct to see that it would take  $O(N)$  time to construct  $\mathcal{T}$  in Algorithm 2. Then, for Algorithm 4, the binary search step requires repeated calls to Algorithm 3 to estimate the marginal ratio. The number of calls is at most  $O(\log \frac{n}{b\varepsilon}) = O(k + \log b^{-1})$ . For each call to Algorithm 3, it costs  $\text{poly}(N)$  time to construct and solve the LP.

For  $N$ , each step of the coupling tree has at most  $\Delta^R$  choices for the vertex  $v$ , and at most  $q^2$  choices for the colours on  $v$ . This continues for at most  $\Delta^R$  times, giving the bound on  $N$ .  $\square$

For Corollary 5, the colouring instance in Corollary 5 has  $O_{\Delta,q}(1)$ -CI, proved in [CLMM23, Theorem 5.10 and Lemma 5.13].<sup>4</sup> Thus Corollary 5 follows from Theorem 4 and Observation 10. The generic marginal lower bound from Observation 10 can be improved using [GKM15, Lemma 3], which implies  $b \geq q^{-1} \left(\frac{2}{3}\right)^\Delta$  in this setting. Moreover, [CLMM23, Theorem 5.9 and Lemma 5.14] directly establish influence decay for colourings in high-girth graphs. One can directly use them to set the parameter  $R$  in (10) to obtain an algorithm with a better running time.

## 5. LOW DISAGREEMENT COUPLING FROM MARKOV CHAIN COUPLING

In this section, we establish coupling independence from contractive coupling for Markov chains. We note that previous results like [BCC<sup>+</sup>22, Liu21] have investigated how to establish spectral independence by using contractive couplings of Markov chains. As a byproduct, their results already implicitly suggested how to establish coupling independence via contractive coupling. However, the proofs of previous works [BCC<sup>+</sup>22, Liu21] rely heavily on Stein’s method (see [BN19, RR19]), which is analytical and arguably unintuitive. In this section, we first give a direct and simple proof that contractive coupling can be used to establish coupling independence.

Given a Markov chain  $P$  and the current state  $X$ , we use  $P(X)$  to denote the next (random) state of  $P$ . Also, recall the Wasserstein distance from Definition 1.

**Lemma 19.** *Let  $(\Omega, d)$  be a finite metric space. Let  $C > 0$  and  $\delta \in (0, 1)$  be two parameters. Let  $P, Q$  be Markov chains on  $\Omega$  with stationary distribution  $\mu$  and  $\nu$ . Suppose the following conditions are satisfied:*

- (*C-disagreement*) for  $X \in \text{supp}(\mu)$ ,  $\mathcal{W}_d(P(X), Q(X)) \leq C$ ;
- (*Q has  $\delta$ -contraction*) for  $X \in \text{supp}(\mu), Y \in \text{supp}(\nu)$ ,  $\mathcal{W}_d(Q(X), Q(Y)) \leq (1 - \delta) d(X, Y)$ .

*Then, it holds that the W1-distance between  $\mu, \nu$  can be bounded by  $\mathcal{W}_d(\mu, \nu) \leq C/\delta$ .*

*Proof.* Let  $X, Y$  be sampled from  $\mu$  and  $\nu$ . Then we have  $P(X) \sim \mu$  and  $Q(Y) \sim \nu$ , which implies

$$\begin{aligned} \mathcal{W}_d(X, Y) &= \mathcal{W}_d(P(X), Q(Y)) \leq \mathcal{W}_d(P(X), Q(X)) + \mathcal{W}_d(Q(X), Q(Y)) \\ &\leq C + (1 - \delta)\mathcal{W}_d(X, Y). \end{aligned}$$

We finish the proof by rearranging terms. □

*Remark.* Note that  $P$  and  $Q$  in Lemma 19 are Markov chains on  $\Omega$ . This does not require the support of  $\mu$  and  $\nu$  to be exactly  $\Omega$ . In fact, it is possible that  $\text{supp}(\mu) \subset \Omega$  or  $\text{supp}(\nu) \subset \Omega$ , and states in  $\Omega \setminus \text{supp}(\mu)$  or  $\Omega \setminus \text{supp}(\nu)$  are transient.

To establish coupling independence, one could apply Lemma 19 with  $\mu^\sigma$  and  $\mu^\tau$  where  $\sigma$  and  $\tau$  are two partial configurations such that  $\text{dist}(\mu, \nu) = 1$ , as in the definition of Definition 2. The first condition typically holds for any Markov chain with local moves. This is because all new disagreement in one step must be near where  $\sigma$  and  $\tau$  disagrees. The second condition is implied by the existence of contractive couplings, and there are plenty available in the literature. We will see Lemma 19 in action in the applications next. In fact, we will design an intermediate distribution and use triangle inequality in the next section.

**5.1. Application: coupling independence for proper colourings.** In this section we show coupling independence for colourings using Lemma 19.

**Theorem 20.** *Let  $G = (V, E)$  be a graph with maximum degree  $\Delta$ . Let  $\mu$  be the uniform distribution over  $q$ -colouring of  $G$ . If either of the following holds:*

- $\Delta \geq 125, q \geq 1.809\Delta$ ; or

---

<sup>4</sup>Theorem 5.10 in [CLMM23] is stated for spectral independence. However, [CLMM23, Lemma 5.13] proved the stronger coupling independence result.

- $\Delta \geq 3, q \geq (11/6 - \epsilon_0)\Delta$  for some fixed parameter  $\epsilon_0 \approx 10^{-5}$ ,

then  $\mu$  satisfies  $O_{\Delta,q}(1)$ -coupling independence. Moreover, the CI constant is  $\text{poly}(q\Delta)$ .

To apply Lemma 19, we consider flip dynamics, for which contractive couplings are known under the conditions of Theorem 20 [CDM<sup>+</sup>19, CV24]. In fact, we need to consider the conditional distribution  $\mu^\sigma$  for some (not necessarily proper) partial colouring  $\sigma$  on  $\Lambda \subset V$ . As mentioned in Section 2, the distribution  $\mu^\sigma$  is the same as the uniform distribution over a list colouring instance, obtained by removing  $\Lambda$  and letting  $L_v = [q] \setminus \{\sigma(u) \mid u \in \Lambda \cap N_G(v)\}$ . Note that if  $q \geq (1 + \alpha)\Delta$  for some  $\alpha > 0$ , then for any list colouring instance obtained this way,  $|L_v| - \deg_G(v) \geq \alpha\Delta$ . Also, as long as  $q \geq \Delta + 1$ , even if  $\sigma$  is not proper, the induced list colouring instance still has solutions. Namely the system is permissive.

We define the flip dynamics for list colouring instances obtained this way. Let the remaining graph be  $G = (V, E)$  and  $\{L_v\}_{v \in V}$  be the lists. Notice  $L_v \subseteq [q]$  for all  $v \in V$ . Let  $X$  be a (not necessarily proper) colouring. We say a path  $w_1 - w_2 - \dots - w_\ell$  is an  $(X(w_1), c)$ -alternating path from  $w_1$  to  $w_\ell$  if for all  $i$ , we have  $X(w_i) \subseteq \{X(w_1), c\}$  and  $X(w_i) \neq X(w_{i+1})$ . The *Kempe component* (or Kempe chain) for  $X$ , a vertex  $v \in V$ , and a colour  $c \in [q]$  is defined by:

$$S_X(v, c) := \{u \in V \mid \text{there is a } (X(v), c)\text{-alternating path from } v \text{ to } u\}.$$

Moreover, let  $S_X(v, X(v)) := \emptyset$ . Let  $\{p_\ell\}_{\ell \geq 1}$  be some parameters. Given a colouring  $X$ , the flip dynamics updates  $X$  to  $X'$  as follows:

- (1) pick a vertex  $v \in V$  and a colour  $c \in [q]$  uniformly at random;
- (2) let  $S = S_X(v, c)$  and  $\ell = |S|$ ;
- (3) if the colouring on  $S$  can be flipped, do so with probability  $p_\ell/\ell$  to obtain  $X'$ ;
- (4) otherwise, let  $X' = X$ .

By flipping, we mean changing the colours of all  $c$ -coloured vertices in  $S$  to  $X(v)$  and change the colours of all  $X(v)$ -coloured vertices in  $S$  to  $c$ . Note that the new colour may not be available in the lists of corresponding vertices. In that case, we simply let  $X' = X$ . Note that a Kempe component of size  $\ell$  is flipped with probability at most  $\frac{p_\ell}{qn}$ . Given a list colouring instance  $(G, L)$ , the above transition rule is defined for all  $X \in [q]^V$  even if  $X$  is improper for the list colouring instance  $(G, L)$ .

Flip dynamics was first considered by Vigoda [Vig00], who designed a set of parameters  $\{p_\ell\}_{\ell \geq 1}$  and a contractive coupling with respect to the Hamming distance when  $q > (11/6)\Delta$ . As a consequence, the flip dynamics is rapidly mixing. Later, Chen, Delcourt, Moitra, Perarnau and Postle [CDM<sup>+</sup>19] improved the rapid mixing regime to  $q \geq (11/6 - \epsilon_0)\Delta$  for some  $\epsilon_0 \approx 10^{-5}$ . They showed that a contractive coupling exists for the flip dynamics with some different  $\{p_\ell\}_{\ell \geq 1}$  by considering either variable length coupling with Hamming distance or using an alternative metric. Very recently, Carlson and Vigoda [CV24] showed that if  $\Delta \geq 125$  and  $q \geq 1.809\Delta$  (note that  $1.809 < 11/6 - \epsilon_0$ ), there is another set of  $\{p_\ell\}_{\ell \geq 1}$  ensuring contractive coupling, by considering a more refined metric.

These alternative metrics can be related to the Hamming distance. Given the metric space  $(\Omega, d)$ , we say that the metric  $d$  is  $\alpha$ -equivalent (to the Hamming distance) for some  $\alpha \geq 1$ , if for all  $x, y \in \Omega$ , it holds that

$$(16) \quad \alpha^{-1} \text{dist}\{x, y\} \leq d\{x, y\} \leq \alpha \text{dist}\{x, y\}.$$

All metrics mentioned above are in fact 2-equivalent.

Moreover, what we really consider is list colouring instances. One can verify (see [CV24, Theorem 4.12 and Section 4.6], [CDM<sup>+</sup>19, Section 6 and Appendix E], and [BCC<sup>+</sup>22, Lemma 4.12]) that the aforementioned contractive couplings extend to list colourings as well, summarised as follows.

**Proposition 21** ([CV24]). *Let  $\Delta \geq 125$  and  $q$  be two integers. There exists a sequence of parameters  $\{p_\ell\}_{\ell \geq 0}$  satisfying  $p_i = 0$  for all  $i > 6$  such the the following holds. Given a list colouring*

instance  $(G, L)$  such that for any  $v \in V$ ,  $L_v \subseteq [q]$  and  $|L_v| - \deg_G(v) \in [0.809\Delta, 5/6\Delta]$ , let  $\Omega \subseteq [q]^V$  denote the set of proper list colourings for  $(G, L)$ , there is a 2-equivalent metric  $d$  on  $\Omega$  such that the flip dynamics  $P$  with  $\{p_\ell\}_{\ell \geq 1}$  satisfies the following: for any two proper list colourings  $X, Y \in \Omega$ ,

$$(17) \quad \mathcal{W}_d(P(X), P(Y)) \leq \left(1 - \frac{10^{-5}}{nq}\right) d(X, Y),$$

where  $n$  is the number of vertices in  $G$ .

**Proposition 22** ([CDM<sup>+</sup>19]). Let  $\Delta \geq 3$  and  $q$  be two integers, and  $\epsilon_0 \approx 10^{-5} > 0$  be a constant. There exists a sequence of parameters  $\{p_\ell\}_{\ell \geq 0}$  satisfying  $p_i = 0$  for all  $i > 6$  such that the following holds. Given a list colouring instance  $(G, L)$  such that for any  $v \in V$ ,  $L_v \subseteq [q]$  and  $|L_v| - \deg_G(v) \geq (5/6 - \epsilon_0)\Delta$ , let  $\Omega \subseteq [q]^V$  denote the set of proper list colourings for  $(G, L)$ , there is a 2-equivalent metric  $d$  on  $\Omega$  such that the flip dynamics  $P$  with  $\{p_\ell\}_{\ell \geq 1}$  satisfies the following: for any two proper colourings  $X, Y \in \Omega$ ,

$$(18) \quad \mathcal{W}_d(P(X), P(Y)) \leq \left(1 - \frac{10^{-9}}{nq}\right) d(X, Y),$$

where  $n$  is the number of vertices in  $G$ .

*Remark* (Domains of colourings). Given a list colouring instance  $(G, L)$ , let  $\Omega = \text{supp}(\mu) = \{\sigma \mid \sigma \in [q]^V \text{ and } \sigma \text{ is a proper list colouring for } (G, L)\}$ , where  $\mu$  denotes the uniform distribution over all proper list colourings. In Proposition 21 and Proposition 22, the metric  $d$  is defined over  $\Omega$  and the contraction results in (17) and (18) hold for all  $X, Y \in \Omega$ , which is sufficient for our proof. We remark that [CDM<sup>+</sup>19, CV24] actually proved stronger results, their metric is defined over a superset of  $\Omega$  and the contraction results holds even for  $X, Y$  beyond  $\Omega$ , because they both used path coupling [BD97] which requires them to prove contraction even for improper list colourings.

Note that in Proposition 21 and Proposition 22, we state the results in terms of W1-distance. This is implied by the contractive couplings from [CV24] and [CDM<sup>+</sup>19]. More precisely, they show that for any  $X, Y \in \Omega$ , there exists a coupling of  $P(X), P(Y)$  such that  $\mathbb{E}[d(P(X), P(Y))] \leq (1 - \frac{C}{nq})d(X, Y)$ , where  $C = 10^{-5}$  or  $10^{-9}$  is the constant in (17) and (18).

Now we are ready to prove Theorem 20.

*Proof of Theorem 20.* Let  $\sigma$  and  $\tau$  be two (not necessarily proper) partial colourings on  $\Lambda \subseteq V$  such that they differ at only  $v$ . Let  $\mu$  denote the uniform distribution over all proper  $q$ -colourings in  $G = (V, E)$ . We goal is to bound  $\mathcal{W}(\mu^\sigma, \mu^\tau)$  to establish the CI for  $\mu$ .

Let  $(G^\sigma, L^\sigma)$  and  $(G^\tau, L^\tau)$  be the two list colouring instances induced by  $\sigma$  and  $\tau$ . Note that  $G^\sigma = G^\tau$  and  $L^\sigma(u) \neq L^\tau(u)$  only if  $u \in N(v)$ , where  $N(v)$  is the set of neighbours of  $v$  in  $G$ . We introduce a middle list colouring instance  $(G', L')$  such that  $G' = G^\sigma = G^\tau = (V', E')$ , where  $V' = V \setminus \Lambda$ , and for all  $u \in V' \setminus N(v)$ ,  $L'(u) = L^\sigma(u) = L^\tau(u)$  and for all  $u \in V' \cap N(v)$ ,  $L'(u) = [q]$ . We use the following bound

$$(19) \quad \mathcal{W}(\mu^\sigma, \mu^\tau) = \mathcal{W}(\mu_{V'}^\sigma, \mu_{V'}^\tau) \leq \mathcal{W}(\mu_{V'}^\sigma, \mu') + \mathcal{W}(\mu', \mu_{V'}^\tau),$$

where  $\mu'$  is the uniform distribution for list colouring  $(G', L')$ .

We show how to bound  $\mathcal{W}(\mu_{V'}^\sigma, \mu')$ . The bound for  $\mathcal{W}(\mu', \mu_{V'}^\tau)$  can be obtained from the same proof. Let the parameters  $\{p_\alpha\}_{\alpha \geq 1}$  and the metric  $d$  be the same as in either Proposition 21 or Proposition 22 for the list colouring  $(G', L')$ . Specifically, if we assume the first condition in Theorem 20, we use Proposition 21, otherwise, we use Proposition 22. The space  $\Omega$  in both propositions is the same one,  $\Omega = \text{supp}(\mu') \subseteq [q]^{V'}$ , which is the set of all proper list colourings for  $(G', L')$ . Let  $P$  and  $Q$  be flip dynamics for  $(G^\sigma, L^\sigma)$  and  $(G', L')$ , respectively, using the same  $\{p_\alpha\}_{\alpha \geq 1}$ . Let  $\mu_P = \mu_{V'}^\sigma$  and  $\mu_Q = \mu'$  denote the uniform distributions of list colourings  $(G^\sigma, L^\sigma)$  and  $(G', L')$  respectively, which are the stationary distributions of  $P$  and  $Q$  respectively. By the

definition of  $(G', L')$ ,  $\text{supp}(\mu_P) \subseteq \text{supp}(\mu_Q)$ . Both  $P$  and  $Q$  can be viewed as Markov chains over the space  $\text{supp}(\mu_Q)$  because  $P(X), Q(X) \in \text{supp}(\mu_Q)$  if  $X \in \text{supp}(\mu_Q)$ . Let  $(\text{supp}(\mu_Q), d)$  be the metric assumed in the two propositions. The contraction results in (17) and (18) hold for  $Q$ .

Next, we use Lemma 19 with  $P$  and  $Q$  to bound  $\mathcal{W}(\mu_P, \mu_Q)$ . For any colouring  $X \in \text{supp}(\mu_P)$ , notice that  $p_\alpha > 0$  only for  $\ell \leq 6$ , and the transitions of flip dynamics  $P$  and  $Q$  can be different only if a neighbour of  $v$  is in the Kempe component. Let  $N(v)$  denotes the set of neighbours of  $v$  in  $G$ . This allows us to couple  $P(X)$  and  $Q(X)$  as follows:

- (1) in both copies, choose the same vertex  $u$  and the same colour  $c$  and let  $S = S_X(u, c)$ ;
- (2) if  $|S| \geq 7$ , let  $P(X) = Q(X) = X$ ;
- (3) if  $v$  is not adjacent to any vertex in  $S$  in graph  $G$ , couple  $P(X)$  and  $Q(X)$  perfectly;
- (4) otherwise, couple  $P(X)$  and  $Q(X)$  independently.

If the coupling goes to the last step, the vertex  $u$  must be within distance 6 from  $v$  in the graph  $G$ . There are at most  $\Delta^7$  such choices. When that happens, there are at most 12 new disagreement. Thus,  $\mathcal{W}(P(X), Q(X)) \leq \frac{12\Delta^7}{|V'|}$ . As  $d$  is 2-equivalent, it implies

$$(20) \quad \mathcal{W}_d(P(X), Q(X)) \leq \frac{24\Delta^7}{|V'|}.$$

This verifies the first condition of Lemma 19.

For the second condition of Lemma 19, we need to bound the distance  $\mathcal{W}_d(Q(X), Q(Y))$  for  $X \in \text{supp}(\mu_P)$  and  $Y \in \text{supp}(\mu_Q)$ . By the definition of the instance  $(G', L')$ , it holds that  $\text{supp}(\mu_P) \subseteq \text{supp}(\mu_Q)$ , and thus  $X \in \text{supp}(\mu_Q)$ . Moreover,  $(G', L')$  has larger colour lists than  $(G^\sigma, L^\sigma)$  and  $(G^\tau, L^\tau)$ . Thus  $(G', L')$  satisfies the condition of either Proposition 21 or Proposition 22, which implies that  $Q$  has  $C/(q|V'|)$ -contraction for some constant  $C > 0$  with respect to  $d$ . Thus, together with (20), we apply Lemma 19 with  $\mu = \mu_P, \nu = \mu_Q$  and  $\Omega = \text{supp}(\mu_Q)$  to derive

$$\mathcal{W}_d(\mu_{V'}^\sigma, \mu') \leq \frac{24\Delta^7/|V'|}{C/(q|V'|)} = \text{poly}(q\Delta).$$

As  $d$  is 2-equivalent to the Hamming distance, we have  $\mathcal{W}(\mu_{V'}^\sigma, \mu') = \text{poly}(q\Delta)$ . Since the same proof works for  $\mathcal{W}(\mu_{V'}^\tau, \mu')$ , the lemma follows from (19).  $\square$

In the proof above, all analysis except the first equation in (19) considers list colouring instances on the subgraph  $G' = G[V']$ . The intermediate instance  $(G', L')$  is introduced because we need to ensure that  $X \in \text{supp}(\mu_Q)$  when applying Proposition 21 or Proposition 22. When verifying the first condition in Lemma 19, we only use the fact that the flip chain is local and the metric  $d$  is 2-equivalent. When verifying the second condition, we use the contraction results in Proposition 21 or Proposition 22 for the flip chain  $Q$  over list colourings of  $(G', L')$ .

Corollary 6 is then a direct consequence of Theorem 4, Theorem 20, and the marginal lower bound [LY13, Lemma 3].

**5.2. Application: coupling independence via the Dobrushin-Shlosman condition.** In this section we show Corollary 8. Recall the Dobrushin-Shlosman condition from Definition 7.

**Theorem 23.** *Let  $\delta \in (0, 1)$  be a parameter. Suppose  $\mu$  is the Gibbs distribution of a permissive spin system  $(G, q, A_E, A_V)$  satisfying the Dobrushin-Shlosman condition with gap  $\delta$ . Then,  $\mu$  satisfies  $\frac{\Delta}{\delta}$ -coupling independence.*

The proof of Theorem 23 follows from Lemma 19 and a well-known contractive coupling for the Glauber dynamics when the 1-norm of the Dobrushin influence matrix  $\rho$  is bounded (see [BD97, Hay06] and references therein). Here Glauber dynamics is a well-studied Markov chain where in each step, we uniformly at random select a variable, and update it conditioning on the rest of the configuration.

**Proposition 24.** Let  $\delta \in (0, 1)$  be a parameter. Suppose  $\mu$  is a distribution over  $[q]^V$  satisfying the Dobrushin-Shlosman condition with gap  $\delta$ . Let  $P$  be the Glauber dynamics for  $\mu$ . Then for every  $X, Y \in \text{supp}(\mu)$ ,

$$\mathcal{W}_{\text{dist}}(P(X), P(Y)) \leq \left(1 - \frac{\delta}{n}\right) \text{dist}(X, Y).$$

*Proof of Theorem 23.* Let  $\sigma$  and  $\tau$  be two (not necessarily feasible) partial configuration such that they differ at only  $v \in V$ . As the system is permissive,  $\mu^\sigma$  and  $\mu^\tau$  are well-defined. To apply Lemma 19, let  $P$  and  $Q$  be Glauber dynamics for  $\mu^\sigma$  and  $\mu^\tau$ , respectively. Consider the coupling where we always choose the same vertex  $u$  to update, and optimally couple the updates at  $u$ . Clearly, only neighbours of  $v$  could be the new disagreement. Thus,

$$\mathcal{W}(P(X), Q(X)) \leq \frac{\Delta}{n}.$$

This verifies the first condition of Lemma 19.

For the second condition of Lemma 19, notice that the Dobrushin influence matrix  $\rho^\tau$  for  $\mu^\tau$  is dominated by the corresponding matrix  $\rho$  for  $\mu$ . Thus,  $\|\rho^\tau\|_1 \leq \|\rho\|_1 \leq 1 - \delta$ . By Proposition 24,  $Q$  has  $\delta/n$ -contraction. Thus, we can apply Lemma 19 to derive

$$\mathcal{W}(\mu^\sigma, \mu^\tau) \leq \frac{\Delta}{\delta}. \quad \square$$

Corollary 8 directly follows from Theorem 4, Observation 10, and Theorem 23.

#### ACKNOWLEDGEMENT

We thank Konrad Anand and Graham Freifeld for some helpful discussions at an early stage of this project. We thank Charlie Carlson, Eric Vigoda, and Hongyang Liu for clarifying some questions regarding the flip dynamics and useful feedback.

#### REFERENCES

- [AL20] Vedat Levi Alev and Lap Chi Lau. Improved analysis of higher order random walks and applications. In *STOC*, pages 1198–1211. ACM, 2020. [2](#)
- [ALO20] Nima Anari, Kuikui Liu, and Shayan Oveis Gharan. Spectral independence in high-dimensional expanders and applications to the hardcore model. In *FOCS*, pages 1319–1330. IEEE, 2020. [2, 7](#)
- [Bar16] Alexander I. Barvinok. *Combinatorics and Complexity of Partition Functions*, volume 30 of *Algorithms and combinatorics*. Springer, 2016. [1](#)
- [BBR24] Ferenc Bencs, Khallil Berrekkal, and Guus Regts. Deterministic approximate counting of colorings with fewer than  $2\delta$  colors via absence of zeros. *arXiv*, abs/2408.04727, 2024. [2, 4](#)
- [BCC<sup>+</sup>22] Antonio Blanca, Pietro Caputo, Zongchen Chen, Daniel Parisi, Daniel Štefankovič, and Eric Vigoda. On mixing of Markov chains: coupling, spectral independence, and entropy factorization. *Electron. J. Probab.*, 27:Paper No. 142, 42, 2022. [2, 3, 19, 20](#)
- [BD97] Russ Bubley and Martin E. Dyer. Path coupling: A technique for proving rapid mixing in markov chains. In *FOCS*, pages 223–231. IEEE Computer Society, 1997. [4, 21, 22](#)
- [BF87] Imre Bárány and Zoltán Füredi. Computing the volume is difficult. *Discret. Comput. Geom.*, 2(4):319–326, 1987. [1](#)
- [BG06] Antar Bandyopadhyay and David Gamarnik. Counting without sampling: new algorithms for enumeration problems using statistical physics. In *SODA*, pages 890–899. ACM, 2006. [1, 5](#)
- [BN19] Guy Bresler and Dheeraj Nagaraj. Stein’s method for stationary distributions of Markov chains and application to Ising models. *Ann. Appl. Probab.*, 29(5):3230–3265, 2019. [19](#)
- [CDM<sup>+</sup>19] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. Improved bounds for randomly sampling colorings via linear programming. In *SODA*, pages 2216–2234. SIAM, 2019. [2, 4, 5, 20, 21](#)
- [CF24] Xiaoyu Chen and Weiming Feng. Rapid mixing via coupling independence for spin systems with unbounded degree. *arXiv*, abs/2407.04672, 2024. [4](#)

- [CG24] Zongchen Chen and Yuzhou Gu. Fast sampling of  $b$ -matchings and  $b$ -edge covers. In *SODA*, pages 4972–4987. SIAM, 2024. [4](#)
- [CGSV21] Zongchen Chen, Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Rapid mixing for colorings via spectral independence. In *SODA*, pages 1548–1557. SIAM, 2021. [4](#), [7](#)
- [CLMM23] Zongchen Chen, Kuikui Liu, Nitya Mani, and Ankur Moitra. Strong spatial mixing for colorings on trees and its algorithmic applications. In *FOCS*, pages 810–845. IEEE, 2023. [3](#), [5](#), [10](#), [19](#)
- [CLV23] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Rapid mixing of Glauber dynamics up to uniqueness via contraction. *SIAM J. Comput.*, 52(1):196–237, 2023. [4](#)
- [CV24] Charlie Carlson and Eric Vigoda. Flip dynamics for sampling colorings: Improving  $(11/6 - \epsilon)$  using a simple metric. *arXiv*, abs/2407.04870, 2024. [2](#), [4](#), [5](#), [20](#), [21](#)
- [CZ23] Xiaoyu Chen and Xinyuan Zhang. A near-linear time sampler for the Ising model with external field. In *SODA*, pages 4478–4503. SIAM, 2023. [2](#), [4](#)
- [DFK91] Martin E. Dyer, Alan M. Frieze, and Ravi Kannan. A random polynomial time algorithm for approximating the volume of convex bodies. *J. ACM*, 38(1):1–17, 1991. [1](#)
- [Dob70] Roland L. Dobrushin. Prescribing a system of random variables by conditional distributions. *Theor. Probability Appl.*, 15(3):458–486, 1970. [4](#)
- [DS85] Roland L. Dobrushin and Senya B. Shlosman. Constructive criterion for the uniqueness of Gibbs field. In *Statistical physics and dynamical systems (Köszeg, 1984)*, volume 10 of *Progr. Phys.*, pages 347–370. Birkhäuser Boston, Boston, MA, 1985. [4](#)
- [Ele86] György Elek. A geometric inequality and the complexity of computing volume. *Discret. Comput. Geom.*, 1(4):289–292, 1986. [1](#)
- [FGW<sup>+</sup>23] Weiming Feng, Heng Guo, Chunyang Wang, Jiaheng Wang, and Yitong Yin. Towards derandomising Markov chain Monte Carlo. In *FOCS*, pages 1963–1990. IEEE, 2023. [1](#)
- [FGYZ22] Weiming Feng, Heng Guo, Yitong Yin, and Chihao Zhang. Rapid mixing from spectral independence beyond the Boolean domain. *ACM Trans. Algorithms*, 18(3):28:1–28:32, 2022. [2](#), [4](#), [7](#)
- [GK07] David Gamarnik and Dmitriy Katz. Correlation decay and deterministic FPTAS for counting list-colorings of a graph. In *SODA*, pages 1245–1254. SIAM, 2007. [2](#)
- [GKM15] David Gamarnik, Dmitriy Katz, and Sidhant Misra. Strong spatial mixing of list coloring of graphs. *Random Struct. Algorithms*, 46(4):599–613, 2015. [3](#), [19](#)
- [GLLZ19] Heng Guo, Chao Liao, Pinyan Lu, and Chihao Zhang. Counting hypergraph colorings in the local lemma regime. *SIAM J. Comput.*, 48(4):1397–1424, 2019. [1](#), [5](#)
- [Hay06] Thomas P. Hayes. A simple condition implying rapid mixing of single-site dynamics on spin systems. In *FOCS*, pages 39–46. IEEE Computer Society, 2006. [4](#), [22](#)
- [HLQZ24] Kun He, Zhidan Li, Guoliang Qiu, and Chihao Zhang. FPTAS for Holant problems with log-concave signatures. *CoRR*, abs/2407.04989, 2024. [1](#), [4](#), [5](#)
- [HPR20] Tyler Helmuth, Will Perkins, and Guus Regts. Algorithmic Pirogov-Sinai theory. *Probab. Theory Related Fields*, 176(3-4):851–895, 2020. [1](#)
- [Jer95] Mark Jerrum. A very simple algorithm for estimating the number of  $k$ -colorings of a low-degree graph. *Random Struct. Algorithms*, 7(2):157–166, 1995. [1](#)
- [JPP23] Matthew Jenssen, Will Perkins, and Aditya Potukuchi. Approximately counting independent sets in bipartite graphs via graph containers. *Random Struct. Algorithms*, 63(1):215–241, 2023. [1](#)
- [JPSS22] Vishesh Jain, Will Perkins, Ashwin Sah, and Mehtaab Sawhney. Approximate counting and sampling via local central limit theorems. In *STOC*, pages 1473–1486. ACM, 2022. [1](#)
- [JPV21] Vishesh Jain, Huy Tuan Pham, and Thuy-Duong Vuong. Towards the sampling Lovász local lemma. In *FOCS*, pages 173–183. IEEE, 2021. [1](#), [5](#)
- [JVV86] Mark Jerrum, Leslie G. Valiant, and Vijay V. Vazirani. Random generation of combinatorial structures from a uniform distribution. *Theor. Comput. Sci.*, 43:169–188, 1986. [18](#)
- [Liu21] Kuikui Liu. From coupling to spectral independence and blackbox comparison with the down-up walk. In *APPROX-RANDOM*, volume 207 of *LIPICs*, pages 32:1–32:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. [2](#), [3](#), [19](#)
- [LSS19a] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. A deterministic algorithm for counting colorings with 2-delta colors. In *FOCS*, pages 1380–1404. IEEE Computer Society, 2019. [2](#), [4](#)
- [LSS19b] Jingcheng Liu, Alistair Sinclair, and Piyush Srivastava. The Ising partition function: zeros and deterministic approximation. *J. Stat. Phys.*, 174(2):287–315, 2019. [4](#)
- [LY13] Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *APPROX-RANDOM*, volume 8096 of *Lecture Notes in Computer Science*, pages 639–654. Springer, 2013. [2](#), [4](#), [22](#)
- [Moi19] Ankur Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. *J. ACM*, 66(2):10:1–10:25, 2019. [1](#), [5](#), [10](#), [11](#)



- [PR17] Viresh Patel and Guus Regts. Deterministic polynomial-time approximation algorithms for partition functions and graph polynomials. *SIAM J. Comput.*, 46(6):1893–1919, 2017. [1](#)
- [RR19] Gesine Reinert and Nathan Ross. Approximating stationary distributions of fast mixing Glauber dynamics, with applications to exponential random graphs. *Ann. Appl. Probab.*, 29(5):3201–3229, 2019. [19](#)
- [SST14] Alistair Sinclair, Piyush Srivastava, and Marc Thurley. Approximation algorithms for two-state anti-ferromagnetic spin systems on bounded degree graphs. *J. Stat. Phys.*, 155(4):666–686, 2014. [4](#)
- [Vig00] Eric Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41(3):1555–1569, 2000. [1](#), [20](#)
- [Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149. ACM, 2006. [1](#), [5](#)
- [WY23] Chunyang Wang and Yitong Yin. A sampling Lovász local lemma for large domain sizes. *arXiv*, abs/2307.14872, 2023. [1](#), [5](#)