COUNTING HYPERGRAPH COLOURINGS IN THE LOCAL LEMMA REGIME

HENG GUO, CHAO LIAO, PINYAN LU, AND CHIHAO ZHANG

Abstract. We give a fully polynomial-time approximation scheme (FPTAS) to count the number of q-colourings for k-uniform hypergraphs with maximum degree Δ if $k \geq 28$ and $q > 357\Delta^{\frac{14}{k-14}}$. We also obtain a polynomial-time almost uniform sampler if $q > 931\Delta^{\frac{16}{k-16/3}}$. These are the first approximate counting and sampling algorithms in the regime $q \ll \Delta$ (for large Δ and k) without any additional assumptions. Our method is based on the recent work of Moitra (STOC, 2017). One important contribution of ours is to remove the dependency of k and Δ in Moitra's approach.

1. Introduction

Hypergraph colouring is a classic and important topic in combinatorics. Its study was initiated by Erdős' seminal result [Erd63], a sufficient upper bound on the number of edges so that a uniform hypergraph is 2-colourable. Many important tools in the probabilistic method have been developed around this subject, such as the Lovász local lemma [EL75], and the Rödl nibble [Röd85].

In this paper, we consider the problem of approximately counting colourings in k-uniform hypergraphs. The most successful approach to approximate counting is Markov chain Monte Carlo (MCMC). See [DFK91, JS93, JSV04] for a few famous examples. Indeed, MCMC has been extensively studied for graph colourings in low-degree graphs. Jerrum [Jer95] showed that the simple and natural Markov chain, Glauber dynamics, mixes rapidly, if $q > 2\Delta$, where q is the number of colours and Δ is the maximum degree of the graph. As a consequence, there is a *fully polynomial-time randomized approximation scheme* (FPRAS) for the number of colourings if $q > 2\Delta$. This result initiated a series of research and the best bound in general requires that $q > (11/6 - \varepsilon)\Delta$ for some small constant $\varepsilon > 0$ [Vig00, CDM+19]. It is conjectured that Glauber dynamics is rapidly mixing if $q > \Delta + 1$, the "freezing" threshold, but current evidences typically require extra conditions in addition to the maximum degree [HV03, DFHV13]. On the flip side, see [GSV15] for some (almost tight) NP-hardness results.

In k-uniform hypergraphs, the Markov chain approach still works, if $q > C\Delta$ for $C = 1$ when $k \geq 4$ and $C = 1.5$ when $k = 3$ [BDK08, BDK06]. However, the local lemma implies that a hypergraph is q-colourable if $q > C\Delta^{1/(k-1)}$ for some constant C. This threshold is much smaller than Δ when Δ is large. Moser and Tardos' algorithmic version of the local lemma [MT10] implies that we can efficiently find a q-colouring under the same condition. Indeed, the study of the algorithmic local lemma has been a highly active area. See [KS11, HSS11, HS13a, HS13b, HV15, AI16, Kol16, CPS17, HLL+17] for various recent development.

In view of the success of algorithmic local lemma, it is natural to wonder, whether we can also randomly generate hypergraph colourings, or equivalently, approximately count their number, beyond the $q \asymp \Delta$ bound and approaching $q \asymp \Delta^{1/(k-1)}$? Unfortunately, designing Markov chains quickly runs into trouble if $q \ll \Delta$. "Freezing" becomes possible in this regime (see [FM11] for examples[1]), and the state space of proper hypergraph colourings may not be connected via changing the colour of a single vertex, the building block move of Glauber dynamics.

A preliminary version of this paper appeared in 50th Annual ACM SIGACT Symposium on the Theory of Computing (STOC), 2018, Los Angeles.

[1]Interestingly, to prove the existence of frozen colourings, we also need to appeal to the local lemma.

The only successful application of MCMC in this regime is due to Frieze et al. [FM11, FA17], which requires that $q > \max\{C_k \log n,\ 500k^3\Delta^{1/(k-1)}\}$ and the hypergraph is simple.[2] Here $q = \Omega(\log n)$ is necessary to guarantee that "frozen" colourings are not prevalent. Furthermore, it is reasonable to believe that simple hypergraphs are much easier algorithmically than general ones, since their chromatic numbers are $O\left(\frac{\Delta}{\log \Delta}\right)^{1/(k-1)}$ [FM13], significantly smaller than the bound implied by the local lemma, and related Glauber dynamics for hypergraph independent sets works significantly better in simple hypergraphs than in general ones [HSZ19].

Our main result is a positive step beyond the freezing barrier in general k-uniform hypergraphs. Our result also answers some open problems raised in [FM11].

Theorem 1. *For integers $\Delta \geq 2$, $k \geq 28$, and $q > 357\Delta^{\frac{14}{k-14}}$, there is an FPTAS for q-colourings in k-uniform hypergraphs with maximum degree Δ.*

When k and Δ are large, our result is better than the Markov chain results [BDK08, BDK06] and gets into the freezing regime. The exponent of our polynomial time bound depends on the constants k and Δ.

Our method is based on an intriguing result shown by Moitra [Moi19] recently, who gave *fully polynomial-time deterministic approximation schemes* (FPTAS) to count satisfying assignments of k-CNF formulas in the local lemma regime. It is not hard to see that Moitra's approach is rather general, and indeed it works for hypergraph colourings if some strong form of the local lemma condition holds, and $k \geq C \log \Delta$ for some constant C, without any requirement on the connectedness of the state space. Unfortunately, the requirement that $k \geq C \log \Delta$ is necessary for a "marking" argument to work in Moitra's approach. This is not an issue for k-CNF formulas, as in that setting the (strong) local lemma condition dictates that $k \geq C \log \Delta$. However, for hypergraph colourings, we generally want k and Δ to be two independent parameters. Marking is no longer possible in our general situation.

We briefly describe Moitra's approach before introducing our modifications. The first observation is that if the maximum degree is much smaller than the local lemma threshold, variables in the target distribution are very close to uniform. As a consequence, if we couple two copies of the Gibbs distribution while giving different colours at a particular vertex, sequentially and in a vertex-wise maximal fashion, the discrepancy in the resulting coupling will be logarithmic with high probability. Then, one can set up a linear program to do binary search for the marginal probability, where the variables to solve mimic the transition probabilities in this coupling. The marking procedure ensures these locally (almost-)uniform properties to hold at any point of the coupling process above, by finding a good set of vertices so that we only couple these vertices and nothing goes awry.

Since marking is no longer possible in our setting, we take an adaptive approach in the coupling procedure to ensure local (almost-)uniform properties, rather than marking what we are going to couple in advance. Although similar in spirit, our proof details are rather different from those by Moitra [Moi19]. Since this coupling (or the analysis thereof) is used repeatedly in the whole algorithm, we have to rework almost all other proofs as well. A crucial technical contribution of ours is to distinguish two kinds of errors that may rise in the linear program.[3] In particular, the coupling process terminating in logarithmic steps with high probability is *not* sufficient to bound the number of certain "bad" partial colourings and a new exponentially small bound is shown (see Lemma 16). Moreover, we also streamline the argument and tighten the bounds at various places. Hopefully these refinement also sheds some light on where the limit of the method is.

The outline above only gives an approximation of the marginal probabilities. Due to the lack of marking, we also need to provide new algorithms for approximate counting and sampling. For approximate counting, we use the local lemma again to find a good ordering of the vertices

[2]A hypergraph is simple if the intersection of any two hyperedges contains at most one vertex.

[3]These two kinds of errors are not to be confused with the type 1 and type 2 errors in [Moi19]. Both types are one kind of error in our analysis.

so that the standard self-reduction goes through. For sampling, we use the marginal algorithm as an oracle, to faithfully simulate the true distribution, in an adaptive fashion similar to the coupling procedure. At the end of this process, not all vertices will be coloured. However we show that with high probability, all remaining connected components have logarithmic sizes and we fill those in by brutal force enumeration. The threshold we obtain for sampling is larger than the one for approximate counting.

Theorem 2. *For integers $\Delta \geq 2$, $k \geq 28$, and $q > 931\Delta^{\frac{16}{k-16/3}}$, there is a sampler whose distribution is ε-close in total variation distance to the uniform distribution on all proper colourings, with running time polynomial in the number of vertices and $1/\varepsilon$.*

The correlation decay approach of approximate counting [Wei06, BG08] have been successfully applied to graph colouring problems [LY13, LYZZ17] or hypergraph problems [BGG+19], but it seems difficult to combine the two in our setting. More recently, there are other progresses with respect to approximate counting in the local lemma regime [HSZ19, GJL19, GJ19]. However, these results do not directly apply to our situation either. Indeed, our result can be seen as one step further to linking the local lemma with approximate counting, as we made Moitra's approach applicable in a more general setting, where the constraint size does not have to be directly related to the probability of bad events or the dependency degree. However, there still seem to be a few difficulties, such as constraints that cannot be satisfied by partial assignments, to go further towards the most general abstract setting of the local lemma, and this is an interesting direction for the future.

The paper is organized as follows. Section 2 introduces basic notions as well as the local lemma, and Section 3 introduces the coupling procedure. We give the algorithm of estimating marginal probabilities in Section 4, and use this algorithm to do counting and sampling in Sections 5 and 6, respectively. To maintain flexibility, in Sections 3, 4, 5, and 6, we keep track of various parameters, and all parameters are optimized in Section 7. We conclude in Section 8 by describing the bottleneck of the current approach, and outlining the difficulties for further generalizations.

2. Preliminary

A hypergraph is a pair $H = (V, \mathcal{E})$ where V is the collection of vertices and $\mathcal{E} \subseteq 2^V$ is the set of hyperedges. We say a hypergraph H is k-uniform if every $e \in \mathcal{E}$ satisfies $|e| = k$. Let $q \in \mathbb{N}$ be the number of available colours. A proper colouring of H is an assignment $\sigma \in [q]^V$ so that every hyperedge in $\mathcal{E}$ is not monochromatic, namely that σ satisfies $|\{\sigma(v) : v \in e\}| > 1$ for every $e \in \mathcal{E}$.

Although our goal is to count colourings in k-uniform hypergraphs, as the algorithm progresses, vertices will be pinned to some fixed value. Therefore we will work with a slightly more general problem, namely hypergraph colouring with pinnings. Formally, an instance of hypergraph colouring *with pinnings* is a pair $(H(V, \mathcal{E}), \mathcal{P})$ where $\mathcal{P} = \{P_e \subseteq [q] : e \in \mathcal{E}\}$ and P_e is the set of colours that are already present (pinned) inside the edge e. In the intermediate steps of our algorithms, $\mathcal{P}$ will be induced by pinning a subset of vertices, but it is more convenient to consider this slightly more general setup. For an instance with pinning, a colouring $\sigma \in [q]^V$ is *proper* if for every $e \in \mathcal{E}$, it holds that $|\{\sigma(v) : v \in e\} \cup P_e| > 1$.

Denote by $\mathcal{C}$ the set of all proper colourings of $(H, \mathcal{P})$. For any $\mathcal{C}' \subseteq \mathcal{C}$, we use $\mu_{\mathcal{C}'}$ to denote the uniform distribution over $\mathcal{C}'$. Since there is no weight involved, $\mu_{\mathcal{C}}$ is our targeting Gibbs distribution.

Let μ be a distribution over colourings $([q] \cup \{-\})^V$, where "$-$" denotes that the vertex is not coloured (yet). We say $\mu(\cdot)$ is *pre-Gibbs* with respect to $\mu_{\mathcal{C}}$ if for every $\sigma \in \mathcal{C}$,

$$\frac{1}{|\mathcal{C}|} = \mu_{\mathcal{C}}(\sigma) = \sum_{\substack{\sigma' \in ([q] \cup \{-\})^V \\ \sigma \models \sigma'}} \mu(\sigma') \cdot \mu_{\mathcal{C}}(\sigma \mid \sigma'),$$

where $\sigma \models \sigma'$ means that the full colouring σ is consistent with the partial one σ'. In other words, if we draw a partial colouring σ' from a pre-Gibbs distribution μ, and then complete σ' uniformly conditioned on coloured vertices (with respect to $\mu_{\mathcal{C}}$), the resulting distribution is exactly $\mu_{\mathcal{C}}$. Note that in our definition we do not require the support of μ to be all partial colourings.

2.1. **Lovász Local Lemma.** Let $(H(V,\mathcal{E}),\mathcal{P})$ be an instance of hypergraph colourings and $q \in \mathbb{N}$ be a non-negative integer. We use Δ to denote the maximum degree of H. Although we consider k-uniform hypergraphs in Theorem 1, in both the sampling and the counting procedure we will pin vertices gradually. Those pinning operations reduce the size of edges, but in our algorithms we make sure that the size of edges will not go down too much. Throughout the section, for every $e \in \mathcal{E}$, we assume $k' \le |e| \le k$. Instances of this kind will emerge in Theorem 21 and Theorem 23.

Let $\mathrm{Lin}(H)$ be the line graph of H, that is, vertices in $\mathrm{Lin}(H)$ are hyperedges in H and two hyperedges are adjacent if they share some vertex in H. The "dependency graph" of our problem is simply the line graph of H. For $e \in \mathcal{E}$, let $\Gamma(e)$ be the neighbourhood of e, namely the set $\{e' \mid e \cap e' \neq \emptyset\}$. It is clear that the maximum degree of $\mathrm{Lin}(H)$ is at most $k(\Delta-1)$. Hence $|\Gamma(e)| \le k(\Delta-1)$ for any $e \in \mathcal{E}$. With a little abuse of notation, for $v \in V$, let $\Gamma(v)$ be the set of edges in $\mathcal{E}$ incident to v, i.e., $\Gamma(v) := \{e \in \mathcal{E} : v \in e\}$. Furthermore, for any event B depending a set of vertices $\mathrm{ver}(B)$, let $\Gamma(B)$ be the set of dependent sets of B, i.e., $\Gamma(B) = \{e \mid e \cap \mathrm{ver}(B) \neq \emptyset\}$.

The (asymmetric) Lovász Local Lemma (proved by Lovász and published by Spencer [Spe77]) states a sufficient condition for the existence of a proper colouring. Note that in the following $\mathbf{Pr}[\cdot]$ refers to the product distribution where every vertex is coloured uniformly and independently.

Theorem 3. *If there exists an assignment $x : \mathcal{E} \to (0,1)$ such that for every $e \in \mathcal{E}$ we have*

$$\mathbf{Pr}[e \text{ is monochromatic}] \le x(e) \prod_{e' \in \Gamma(e)} (1 - x(e')), \tag{1}$$

then a proper colouring exists.

When the condition of Theorem 3 is met, we actually have good control over any event in the uniform distribution $\mu_{\mathcal{C}}$ due to the next theorem, shown in [HSS11].

Theorem 4. *If (1) holds for every $e \in \mathcal{E}$, then for any event B, it holds that*

$$\mu_{\mathcal{C}}(B) \le \mathbf{Pr}[B] \prod_{e \in \Gamma(B)} (1 - x(e))^{-1}.$$

Theorem 4 also allows us to have some quantitative control over the marginal probabilities.

Lemma 5. *If $k' \le |e| \le k$ for any $e \in \mathcal{E}$, $t \ge k$ and $q \ge (et\Delta)^{\frac{1}{k'-1}}$, then for any $v \in V$ and any colour $c \in [q]$,*

$$\mathbf{Pr}_{\sigma \sim \mu_{\mathcal{C}}}[\sigma(v) = c] \le \frac{1}{q}\left(1 + \frac{4}{t}\right).$$

Proof. Let $x(e) = \frac{1}{t\Delta}$ for every $e \in \mathcal{E}$. We first verify that (1) holds. Since $|\Gamma(e)| \le k(\Delta-1)$ and $t \ge k$,

$$x(e) \prod_{e' \in \Gamma(e)} (1 - x(e')) \ge \frac{1}{t\Delta}\left(1 - \frac{1}{t\Delta}\right)^{k(\Delta-1)} \ge \frac{1}{et\Delta} \ge q^{1-k'} \ge \mathbf{Pr}[e \text{ is monochromatic}].$$

Hence, Theorem 4 applies. Then,

$$\mathbf{Pr}_{\sigma \sim \mu_{\mathcal{C}}}[\sigma(v) = c] \le \frac{1}{q}\left(1 - \frac{1}{t\Delta}\right)^{-\Delta} \le \frac{1}{q}\exp\left(\frac{2}{t}\right) \le \frac{1}{q}\left(1 + \frac{4}{t}\right). \qquad \square$$

Unfortunately, Theorem 4 does not give lower bounds directly. We will instead bound the probability of blocking v to have colour c.

Lemma 6. *If $k' \le |e| \le k$ for any $e \in \mathcal{E}$, $t \ge k$, and $q \ge (et\Delta)^{\frac{1}{k'-1}}$, then for any $v \in V$ and any colour $c \in [q]$,*

$$\mathop{\mathbf{Pr}}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v) = c] \ge \frac{1}{q}\left(1 - \frac{1}{t}\right).$$

Proof. Fix v and c. For every $e \in \Gamma(v)$, let Block_e be the event that vertices in e other than v all have the colour c. Clearly, conditioned on none of Block_e occurring, the probability of v coloured c is larger than $1/q$. Hence we have that

$$\mathop{\mathbf{Pr}}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v) = c] \ge \frac{1}{q}\left(1 - \sum_{e\in\Gamma(v)} \mu_{\mathcal{C}}(\mathrm{Block}_e)\right). \tag{2}$$

Clearly $\mathbf{Pr}[\mathrm{Block}_e] = q^{1-|e|} \le q^{1-k'}$. Again let $x(e) = \frac{1}{t\Delta}$ for every $e \in \mathcal{E}$ and (1) holds. Since $|\Gamma(\mathrm{Block}_e)| \le k(\Delta - 1) + 1$ and $t \ge k$, by Theorem 4,

$$\mu_{\mathcal{C}}(\mathrm{Block}_e) \le q^{1-k'}\left(1 - \frac{1}{t\Delta}\right)^{-k(\Delta-1)-1} \le \frac{1}{t\Delta}. \tag{3}$$

Plugging (3) into (2) yields

$$\mathop{\mathbf{Pr}}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v) = c] \ge \frac{1}{q}\left(1 - \frac{1}{t}\right). \qquad \square$$

Combining Lemma 5 and Lemma 6, we obtain the following result.

Lemma 7. *If $k' \le |e| \le k$ for any $e \in \mathcal{E}$, $t \ge k$ and $q \ge (et\Delta)^{\frac{1}{k'-1}}$, then for any $v \in V$ and any colour $c \in [q]$,*

$$\frac{1}{q}\left(1 - \frac{1}{t}\right) \le \mathbf{Pr}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v) = c] \le \frac{1}{q}\left(1 + \frac{4}{t}\right).$$

3. The coupling

Recall that a partial colouring is an assignment $\sigma \in ([q] \cup \{-\})^V$ where "$-$" denotes an unassigned colour. Fix a vertex $v \in V$ and two distinct colours $c_1, c_2 \in [q]$, we define two initial partial colourings X_0 and Y_0 that assign v with colours c_1 and c_2 respectively and let all other vertices be unassigned. We use $\mathcal{C}_1$ and $\mathcal{C}_2$ to denote the set of proper colourings with v fixed to be c_1 and c_2 respectively. For a partial colouring X, we use $\mathcal{C}_X$ to denote the set of proper colourings consistent with X.

Moitra [Moi19] introduced the following intriguing idea (in the setting of CNF) to compute the ratio of marginal probabilities on v. Couple $\mu_{\mathcal{C}_1}$ and $\mu_{\mathcal{C}_2}$ in a sequential way. Start from v, where the colours differ, and proceed in a breadth-first search manner, vertex by vertex. At each vertex we draw a colour from $\mu_{\mathcal{C}_1}$ and $\mu_{\mathcal{C}_2}$, respectively, conditioned on all the existing colours, and couple them maximally. The process ends when the set of vertices coupled successfully form a cut separating v from uncoloured vertices. If every vertex we encounter has its marginal distribution close enough to the uniform distribution, then this coupling process terminates quickly with high probability. These local almost-uniform properties are guaranteed by Lemma 7. Then Moitra sets up a clever linear program (LP), where the variables mimic transition probabilities during the coupling (but in some conditional way), and shows that the LP is sufficient to recover the marginal distribution at v by a binary search.

We apply the same idea here for hypergraph colourings. However, one needs to carefully implement the coupling to guarantee that all marginal distributions encountered are close enough to uniform. Formally, we describe our coupling process in Algorithm 1. The coupling process applies to hypergraphs with edge size between k_1 and k for some parameter $0 < k_1 \le k$. There is another parameter $0 < k_2 < k_1$ and all these parameters will be set in Section 7. The output is a pair of partial colourings (X, Y) extending X_0 and Y_0 respectively. Notice that in order to implement the coupling process, we fix an arbitrary ordering of edges and vertices in advance.

Algorithm 1 The coupling process

Input: *A hypergraph $H(V, \mathcal{E})$ with pinnings $\mathcal{P}$ and $k_1 \le |e| \le k$ for every $e \in \mathcal{E}$, two partial colourings X_0 and Y_0.*
Output: *$V_{\mathrm{col}} \subseteq V$, a partition $V_1 \sqcup V_2 = V$, and two partial colourings X, Y defined on V_{col}.*
$V_1 \leftarrow \{v\}$, $V_2 \leftarrow V \setminus V_1$, $V_{\mathrm{col}} \leftarrow \{v\}$;
$X \leftarrow X_0$, $Y \leftarrow Y_0$;
while $\exists e \in \mathcal{E}$ s.t. $e \cap V_1 \neq \varnothing$ and $e \cap V_2 \neq \varnothing$ **do**
  Let e be the first such hyperedge;
  Let u be the first vertex in $e \cap V_2$;
  Sample a pair of colours (c_x, c_y) according to the maximal coupling of the marginal distribution at u conditioned on X and Y respectively;
  Extend X and Y by colouring u with c_x and c_y, respectively;
  $V_{\mathrm{col}} \leftarrow V_{\mathrm{col}} \cup \{u\}$;
  if $c_x \neq c_y$ **then**
    $V_1 \leftarrow V_1 \cup \{u\}$, $V_2 \leftarrow V_2 \setminus \{u\}$;
  end if
  for $e \in \Gamma(u) \cap \mathcal{E}$ s.t. e is satisfied by both X and Y **do**
    $\mathcal{E} \leftarrow \mathcal{E} \setminus \{e\}$;
  end for
  for $e \in \Gamma(u) \cap \mathcal{E}$ s.t. $e \cap V_1 \neq \varnothing$, $e \cap V_2 \neq \varnothing$, and $|e \cap V_{\mathrm{col}}| = k_2$ **do**
    $V_1 \leftarrow V_1 \cup (e \setminus V_{\mathrm{col}})$, $V_2 \leftarrow V \setminus V_1$;
    $\mathcal{E} \leftarrow \mathcal{E} \setminus \{e\}$;
  end for
end while

The set V_{col} consists of all coloured vertices. Intuitively, the set V_1 contains vertices that have failed the coupling and V_2 is its complement. Once a hyperedge is satisfied by both partial colourings X and Y, it has no effect any more and is thus removed.

The main difference from Moitra's coupling [Moi19] is that we cannot choose what vertices to couple in advance ("marking"). Instead, we take an adaptive approach to ensure that no hyperedge becomes too small. Once k_2 vertices of a hyperedge are coloured, all the rest vertices are considered "failed" in the coupling (namely they are added to V_1). However these failed vertices are left uncoloured.

Algorithm 1 outputs a pair of partial colourings X, Y defined on V_{col} and a partition of vertices $V = V_1 \sqcup V_2$. For any edge e in the original $\mathcal{E}$ such that $e \cap V_1 \neq \varnothing$ and $e \cap V_2 \neq \varnothing$, it is removed because either it is satisfied by both X and Y, or k_2 vertices in e have been coloured. In the latter case, all vertices in e are either coloured or in V_1, namely $e \subset V_1 \cup V_{\mathrm{col}}$. Hence all edges intersecting V_1 and $V_2 \setminus V_{\mathrm{col}}$ are satisfied by both X and Y. This fact will be useful later.

For $u \in V$, let $\Gamma_{\mathrm{ver}}(u)$ denote the neighbouring vertices of u (including u), namely $\Gamma_{\mathrm{ver}}(u) = \{w \mid \exists e \in \mathcal{E}, \{u, w\} \subseteq e\}$, and let $\Gamma_{\mathrm{ver}}(U) = \bigcup_{u \in U} \Gamma_{\mathrm{ver}}(u)$ for a subset $U \subseteq V$. The following lemma summarizes some properties of this random process.

Lemma 8. *The following properties of Algorithm 1 hold:*

(1) *All coloured vertices are either in V_1 or incident to V_1, namely $V_{\mathrm{col}} \subseteq \Gamma_{\mathrm{ver}}(V_1)$;*
(2) *The distributions of X and Y are pre-Gibbs with respect to $\mu_{\mathcal{C}_1}$ and $\mu_{\mathcal{C}_2}$ respectively.*

Proof. For (1), notice that whenever we add a vertex u into V_{col}, it must hold that $u \in e$ for some $e \cap V_1 \neq \varnothing$ at the time. The claim follows from a simple induction.

For (2), we only prove the lemma for X. The proof for Y is similar. The partial colouring X is generated in the following way: at each step either the process ends, or the next uncoloured vertex u is chosen and extend X to u with the correct (conditional) marginal probability and repeat. Our decisions (whether or not to halt, and what is the next u) depend on Y in addition to the partial colouring X so far.

An intermediate state $\mathcal{S}$ of Algorithm 1 consists of partial colourings X, Y, V_{col}, and V_1.[4] Our claim is that, conditioned on any valid $\mathcal{S}$, the distribution of the final output (on the X side) of Algorithm 1 is pre-Gibbs with respect to $\mu_{\mathcal{C}_X}$. The lemma clearly follows from the claim by setting $\mathcal{S}$ to the initial state of Algorithm 1.

We induct on the maximum possible future steps of $\mathcal{S}$. The base case is that $\mathcal{S}$ will halt immediately. Thus the output is simply X and completing it yields the uniform distribution on $\mathcal{C}_X$. That is, the output is pre-Gibbs.

For the induction step, $\mathcal{S}$ will not halt but rather, extend the colourings to some vertex u which is deterministically selected by our algorithm. Let $\tau_{\mathcal{S}}(\cdot)$ denote the measure on colourings obtained by completing the output of Algorithm 1 conditioned on $\mathcal{S}$. Let $X^{u\leftarrow c}$ be a partial colouring defined on $V_{\text{col}} \cup \{u\}$ by extending X to u with colour c, and $\mathcal{S}'$ be an internal state consistent with $X^{u\leftarrow c}$, denoted by $\mathcal{S}' \models X^{u\leftarrow c}$. Moreover, let $q(\mathcal{S}')$ be the probability of transiting from $\mathcal{S}$ to $\mathcal{S}'$. Since the marginal probability at u only depends on the previous partial colourings X', we have that

$$\sum_{\mathcal{S}' \models X^{u\leftarrow c}} q(\mathcal{S}') = \mu_{C_X}(X^{u\leftarrow c}), \tag{4}$$

where $\mu_{C_X}(X^{u\leftarrow c})$ is in fact the marginal probability of the colour c at u conditioned on X. By our induction hypothesis, conditioned on $\mathcal{S}'$, the final output is pre-Gibbs with respect to $\mathcal{C}_{X^{u\leftarrow c}}$. That is,

$$\tau_{\mathcal{S}'}(\cdot) = \mu_{\mathcal{C}_{X^{u\leftarrow c}}}(\cdot). \tag{5}$$

For $\sigma \in \mathcal{C}_X$, suppose $X^{u\leftarrow c}$ is the partial colouring of σ restricted to $V_{\text{col}} \cup \{u\}$. Then we have that

$$\begin{aligned}
\tau_{\mathcal{S}}(\sigma) &= \sum_{\mathcal{S}' \models X^{u\leftarrow c}} q(\mathcal{S}')\tau_{\mathcal{S}'}(\sigma) \\
&= \sum_{\mathcal{S}' \models X^{u\leftarrow c}} q(\mathcal{S}')\mu_{\mathcal{C}_{X^{u\leftarrow c}}}(\sigma) \\
&= \mu_{\mathcal{C}_{X^{u\leftarrow c}}}(\sigma) \sum_{\mathcal{S}' \models X^{u\leftarrow c}} q(\mathcal{S}') \\
&= \mu_{\mathcal{C}_{X^{u\leftarrow c}}}(\sigma)\mu_{C_X}(X^{u\leftarrow c}) \\
&= \mu_{\mathcal{C}_X}(\sigma),
\end{aligned}$$

where in the second line we use (5), and in the fourth line we use (4). The claim follows. $\square$

Therefore, the output of Algorithm 1 is a coupling of two pre-Gibbs measures such that they are defined on the same set of vertices V_{col}. We use $\mu_{\text{cp}}(\cdot,\cdot)$ to denote this joint distribution.

It is possible to show that the final size of $|V_1|$ is $O(\log |V|)$ with high probability. This fact will not be directly used, and is indeed not strong enough for the algorithm and its analysis in the next section. We will omit its proof. What we will show eventually is that, conditioned on a randomly chosen colouring from $\mathcal{C}_1$ or $\mathcal{C}_2$, the probability that the coupling process terminates decays exponentially with the depth. There are two levels of randomness here, and they will be separated, since the linear program later will only be able to certify the second kind randomness.

Later, in Section 6, when we do sampling, we will consider a similar procedure, Algorithm 2, and we will show that the connected components produced by Algorithm 2 are $O(\log |V|)$ with high probability (Lemma 22). This is in the same vein as $|V_1|$ being size $O(\log |V|)$ with high probability in Algorithm 1.

[4]We note that actually V_{col} and V_1 are completely determined by X and Y, but we do not need this fact here. The reason for V_{col} is obvious, and V_1 can be deduced from X, Y by simulating the whole process from start.

4. Computing the marginals

In the previous section, we introduced a random process to generate a joint distribution of partial colourings $\mu_{\mathrm{cp}}(\cdot,\cdot)$, whose marginal distributions are pre-Gibbs. Recall that we fixed $X(v) = c_1$ and $Y(v) = c_2$. Let q_i denote the marginal probability in $\mu_{\mathcal{C}}$ of v being coloured by c_i, for $i = 1, 2$. That is, $\mathrm{q}_i = \frac{|\mathcal{C}_i|}{|\mathcal{C}|}$ for $i = 1, 2$. The coupling naturally induces an (imaginary) sampler to uniformly sample from $\mathcal{C}_1 \cup \mathcal{C}_2$ as follows:

Step 1: Sample $(X, Y) = (x, y)$ using Algorithm 1;
Step 2: Let $v \leftarrow c_1$ with probability $\frac{\mathrm{q}_1}{\mathrm{q}_1+\mathrm{q}_2}$ and $v \leftarrow c_2$ otherwise;
Step 3: If v is coloured by c_1, uniformly output a colouring in $\mathcal{C}_x$, otherwise uniformly output a colouring in $\mathcal{C}_y$.

We denote this sampler by $\mathbb{S}$. The output of $\mathbb{S}$ is uniform over $\mathcal{C}_1 \cup \mathcal{C}_2$ is because by Lemma 8, the output distribution of Algorithm 1, projected to either side, is pre-Gibbs. Then we choose the final colouring proportional to the correct ratio.

One can represent the coupling process (Algorithm 1) as traversing a (deterministic) *coupling tree* $\mathcal{T}$ constructed as follows: each vertex in $\mathcal{T}$ represents a pair of partial colourings (x, y)[5] defined on some V_{col} that have appeared in the coupling. We write $(x, y) \in \mathcal{T}$ if (x, y) is a pair of partial colourings represented by some vertex in $\mathcal{T}$. Although the intermediate state of Algorithm 1 consists of partial colourings x, y together with V_{col} and V_1, we can actually deduce V_{col} from x, y, as well as V_1 by simulating Algorithm 1 from the start given x and y. Thus the pair (x, y) determines either that the coupling should halt or the next vertex u to extend to. In the coupling tree $\mathcal{T}$, (x, y) either is a leaf or has q^2 children, which correspond to the q^2 possible ways to extend (x, y) by colouring u. The root of the tree is the initial pair (x_0, y_0) defined on $\{v\}$.

In the following, we identify a collection of conditional marginal probabilities that keeps the information of the coupling process.

First, consider a pair of partial colourings $(x, y) \in \mathcal{T}$ which is a leaf, and any two proper colourings σ_x, σ_y such that $\sigma_x \models x$ and $\sigma_y \models y$. In the probability space induced by the sampler introduced above, define

$$\mathrm{p}^x_{x,y} := \mathbf{Pr}_{(X,Y)\sim\mu_{\mathrm{cp}}}\left[X = x, Y = y \mid \mathbb{S} \text{ outputs } \sigma_x\right];$$
$$\mathrm{p}^y_{x,y} := \mathbf{Pr}_{(X,Y)\sim\mu_{\mathrm{cp}}}\left[X = x, Y = y \mid \mathbb{S} \text{ outputs } \sigma_y\right].$$

These quantities are well defined and independent of the particular choices of σ_x and σ_y. Essentially we only condition on the random choice at step 2 of $\mathbb{S}$. Once that choice is made, the output is uniform over $\mathcal{C}_x$ or $\mathcal{C}_y$.

Perhaps a clearer way of seeing this independence is to give more explicit expressions to $p^x_{x,y}$ and $p^y_{x,y}$. By Bayes' rule,

$$\begin{aligned}
\mathrm{p}^x_{x,y} &= \frac{\mathbf{Pr}_{(X,Y)\sim\mu_{\mathrm{cp}}}\left[\mathbb{S} \text{ outputs } \sigma_x \mid X = x, Y = y\right]\mu_{\mathrm{cp}}(x, y)}{\mathbf{Pr}\left[\mathbb{S} \text{ outputs } \sigma_x\right]} \\
&= \mathrm{q}_1 \cdot \frac{|\mathcal{C}_1 \cup \mathcal{C}_2|}{|\mathcal{C}_x|} \cdot \mu_{\mathrm{cp}}(x, y); \qquad (6)
\end{aligned}$$

$$\begin{aligned}
\mathrm{p}^y_{x,y} &= \frac{\mathbf{Pr}_{(X,Y)\sim\mu_{\mathrm{cp}}}\left[\mathbb{S} \text{ outputs } \sigma_y \mid X = x, Y = y\right]\mu_{\mathrm{cp}}(x, y)}{\mathbf{Pr}\left[\mathbb{S} \text{ outputs } \sigma_y\right]} \\
&= \mathrm{q}_2 \cdot \frac{|\mathcal{C}_1 \cup \mathcal{C}_2|}{|\mathcal{C}_y|} \cdot \mu_{\mathrm{cp}}(x, y). \qquad (7)
\end{aligned}$$

Combining two identities above we obtain

$$\mathrm{q}_1 \cdot \mathrm{p}^y_{x,y} \cdot |\mathcal{C}_y| = \mathrm{q}_2 \cdot \mathrm{p}^x_{x,y} \cdot |\mathcal{C}_x|\,. \qquad (8)$$

[5] We use small letters x, y to denote particular partial colourings, and reserve capital X, Y to denote random ones.

A crucial observation is that, for every pair of partial colourings (x, y) that is a leaf of $\mathcal{T}$ with corresponding V_{col}, V_1, V_2, the ratio $\frac{|\mathcal{C}_x|}{|\mathcal{C}_y|}$ can be computed in $q^{|V_1 \setminus V_{\text{col}}|}$ time. This is because when Algorithm 1 terminates, all edges intersecting V_1 and $V_2 \setminus V_{\text{col}}$ are satisfied by both x and y. The numbers of ways colouring blank vertices in V_2 cancel out, and we only need to enumerate all colourings for blank vertices inside V_1. Let $r_{x,y} = \frac{|\mathcal{C}_x|}{|\mathcal{C}_y|}$.

Next, consider an internal (x, y) in the coupling tree $\mathcal{T}$. We interpret $\mathrm{p}^x_{x,y}$ and $\mathrm{p}^y_{x,y}$ as the probability that the coupling process has ever arrived at an internal pair of partial colourings (x, y) conditioned on the output of $\mathbb{S}$ being σ_x and σ_y for any σ_x, σ_y such that $\sigma_x \models x$ and $\sigma_y \models y$, respectively. Note that the definition is consistent with our previous definition when (x, y) is a leaf of $\mathcal{T}$. Recall that (x_0, y_0) is the root of $\mathcal{T}$, namely x_0 or y_0 only colours v with c_1 or c_2, respectively. For (x_0, y_0), we have that

$$p^{x_0}_{x_0,y_0} = p^{y_0}_{x_0,y_0} = 1. \tag{9}$$

Moreover, for an internal (x, y) whose children are defined on $V'_{\text{col}} = V_{\text{col}} \cup \{u\}$, it holds that

$$\text{for every } c \in [q],\ \mathrm{p}^x_{x,y} = \sum_{c' \in [q]} \mathrm{p}^{x^{u\leftarrow c}}_{x^{u\leftarrow c},y^{u\leftarrow c'}}; \tag{10}$$

$$\text{for every } c \in [q],\ \mathrm{p}^y_{x,y} = \sum_{c' \in [q]} \mathrm{p}^{y^{u\leftarrow c}}_{x^{u\leftarrow c'},y^{u\leftarrow c}}. \tag{11}$$

where we use $x^{u\leftarrow c}$ to denote the partial colouring that extends x by assigning colour c to the vertex u. To see why (10) holds, we note that conditioned on the event that $\mathbb{S}$ outputs some $\sigma_x \models x$, the colouring on u is $\sigma_x(u)$ and all the randomnesses are from the choice of colours in Y on u. The identity (11) holds for the same reason, by reversing the roles of x and y.

In fact, when the coupling process is at some internal node of the coupling tree, say (x, y), defined on V_{col}, and the next step is to sample the colour on a vertex u, one can recover the distribution of the colour on u in the next step from the values

$$\left\{\mathrm{p}^{x^{u\leftarrow c}}_{x^{u\leftarrow c},y^{u\leftarrow c'}}, \mathrm{p}^{y^{u\leftarrow c}}_{x^{u\leftarrow c'},y^{u\leftarrow c}} \;:\; c, c' \in [q]\right\}$$

by solving linear constraints using Bayes' rule. Therefore, the collection $\left\{\mathrm{p}^x_{x,y}, \mathrm{p}^y_{x,y} \;:\; (x, y) \in \mathcal{T}\right\}$ encodes all information of the coupling process.

4.1. **The linear program.** The values $\mathrm{p}^x_{x,y}$ and $\mathrm{p}^y_{x,y}$ are unknown and we are going to impose a few necessary linear constraints on them. The basic constraints are derived from (8), (9), (10), and (11). To this end, for every node (x, y) in $\mathcal{T}$, we introduce two variables $\underline{p}^x_{x,y}$ and $\underline{p}^y_{x,y}$, aiming to mimic $\mathrm{p}^x_{x,y}$ and $\mathrm{p}^y_{x,y}$.

The full coupling tree $\mathcal{T}$ is too big, and we will truncate it up to some depth $L > 0$. The quantity L will be set later. We will perform a binary search to estimate the ratio $\frac{\mathrm{q}_1}{\mathrm{q}_2}$ using the truncated coupling tree. Thus, we introduce two variables $\overline{r}$ and $\underline{r}$ as our guesses for upper and lower bounds of $\frac{\mathrm{q}_1}{\mathrm{q}_2}$. Let $\mathcal{T}_L$ be the coupling tree truncated at depth L, and denote by $\mathcal{L}(\mathcal{T})$ the leaves of a tree $\mathcal{T}$. Since the coupling procedure colours one vertex at a time, for any node $(x, y) \in \mathcal{T}_L$, we have that $|V_{\text{col}}| \leq L$ where V_{col} is determined by (x, y). Formally, we have three types of constraints.

Constraints 1: For every leaf $(x, y) \in \mathcal{L}(\mathcal{T}_L)$ with corresponding $|V_{\text{col}}| < L$, we have the constraints:

$$\begin{aligned} \underline{r} \cdot \underline{p}^y_{x,y} &\leq \underline{p}^x_{x,y} \cdot r_{x,y}; \\ \underline{p}^x_{x,y} \cdot r_{x,y} &\leq \overline{r} \cdot \underline{p}^y_{x,y}; \\ 0 \leq \underline{p}^x_{x,y}, &\underline{p}^y_{x,y} \leq 1. \end{aligned}$$

Constraints 1 are relaxed versions of identity (8). It will be clear soon that these constraints are the most critical ones, as they guarantee that we can recover the marginal probability on v from these variables. However, in order to compute $r_{x,y}$, one needs $\exp(L)$ amount of time. This forces us to truncate at only logarithmic depth in the coupling tree in order to get a polynomial time algorithm, but we will show later that this is enough.

Constraints 2: For the root $(x_0, y_0) \in \mathcal{T}$, we have

$$\underline{p}^{x_0}_{x_0,y_0} = \underline{p}^{y_0}_{x_0,y_0} = 1.$$

Moreover, for every non-leaf $(x, y) \in \mathcal{T}$ with corresponding $|V_{\text{col}}| < L$, let u be the next vertex to couple. We have the following constraints:

$$\begin{aligned}
\text{for every } c \in [q],\ \underline{p}^{x}_{x,y} &= \sum_{c' \in [q]} \underline{p}^{x^{u\leftarrow c}}_{x^{u\leftarrow c},y^{u\leftarrow c'}};\\
\text{for every } c \in [q],\ \underline{p}^{y}_{x,y} &= \sum_{c' \in [q]} \underline{p}^{y^{u\leftarrow c}}_{x^{u\leftarrow c'},y^{u\leftarrow c}};\\
0 \leq \underline{p}^{x}_{x,y}&, \underline{p}^{y}_{x,y} \leq 1.
\end{aligned}$$

These constraints faithfully realize the properties (9), (10), and (11).

Constraints 3: For every $c, c' \in [q]$ that $c \neq c'$, we add constraints:

$$\begin{aligned}
\underline{p}^{x^{u\leftarrow c}}_{x^{u\leftarrow c},y^{u\leftarrow c'}} &\leq \frac{5}{t^*} \cdot \underline{p}^{x}_{x,y};\\
\underline{p}^{y^{u\leftarrow c'}}_{x^{u\leftarrow c},y^{u\leftarrow c'}} &\leq \frac{5}{t^*} \cdot \underline{p}^{y}_{x,y}.
\end{aligned}$$

We will eventually set $t^* = 5\big(e^2 k^3 \Delta^3\big)^{\frac{1}{1-\beta}}$ in Lemma 18, where the parameter $0 < \beta < 1$ will become clear in Definition 15.

These constraints reflect the fact that the coupling at individual vertices is very likely to succeed, due to Lemma 7. Assume the conditions of Lemma 7 are met with $t = t^*$. We claim the following property of those true values $\{\mathrm{p}^{x}_{x,y}\}$.

Claim 9.

$$\frac{\mathrm{p}^{x^{u\leftarrow c}}_{x^{u\leftarrow c},y^{u\leftarrow c}}}{\mathrm{p}^{x}_{x,y}} \geq 1 - \frac{5}{t^*}.$$

The claim implies that these true values satisfy **Constraints 3** since they also satisfy **Constraints 2**. We use (6) to show the claim. By Lemma 7,

$$\frac{|C_x|}{|C_{x^{u\leftarrow c}}|} = \frac{1}{\mathbf{Pr}_{\sigma\sim\mu_{C_x}}[\sigma(u) = c]} \geq \frac{qt^*}{t^*+4}.$$

Again by Lemma 7, the coupling at u with any colour c succeeds with probability at least $\frac{1}{q}\left(1 - \frac{1}{t^*}\right)$. Thus the ratio $\frac{\mu_{\text{cp}}(x^{u\leftarrow c}, y^{u\leftarrow c})}{\mu_{\text{cp}}(x,y)}$, which can be viewed as the probability of coupling u successfully with colour c conditioned on reaching (x, y), is at least $\frac{1}{q}\left(1 - \frac{1}{t^*}\right)$. Combine these facts with (6),

$$\frac{\mathrm{p}^{x^{u\leftarrow c}}_{x^{u\leftarrow c},y^{u\leftarrow c}}}{\mathrm{p}^{x}_{x,y}} = \frac{|C_x|}{|C_{x^{u\leftarrow c}}|} \cdot \frac{\mu_{\text{cp}}(x^{u\leftarrow c}, y^{u\leftarrow c})}{\mu_{\text{cp}}(x,y)} \geq \frac{qt^*}{t^*+4} \cdot \frac{1}{q}\left(1 - \frac{1}{t^*}\right) = 1 - \frac{5}{t^*+4} \geq 1 - \frac{5}{t^*}.$$

Similar inequalities hold for $\{p^{y}_{x,y}\}$ due to (7).

4.2. **Analysis of the LP.** In this subsection, we show that the LP can be used to obtain an efficient and accurate estimator of marginals.

Theorem 10. *Let $\Delta \geq 2$ and $k > 0$ be two integers. Let $0 < \beta < 1$ be a constant. Let $0 < k_2 < k_1 \leq k$ be integers. Let $H = (V, \mathcal{E})$ be a hypergraph with pinnings $\mathcal{P}$, maximum degree Δ such that $k_1 \leq |e| \leq k$ for every $e \in \mathcal{E}$. If*

$$q > \max\left\{(ek\Delta)^{\frac{1}{k_1-2}}, \beta^{\frac{-1}{k_2-1}}, C\Delta^{\frac{3}{\beta(k_2-1)}}, C\Delta^{\frac{4-\beta}{(1-\beta)(k_1-k_2-1)}}\right\}$$

where

$$C > \max\left\{\left(\frac{e^{\beta+3}k^3}{\beta^\beta}\cdot\binom{k}{k_2}\right)^{\frac{1}{\beta(k_2-1)}}, \left(5e\left(e^2k^3\right)^{\frac{1}{1-\beta}}\right)^{\frac{1}{k_1-k_2-1}}\right\},$$

then there is a deterministic algorithm that, for every $v \in V$, $c \in [q]$ and $\varepsilon > 0$, it computes a number $\widehat{p}$ satisfying

$$e^{-\varepsilon}\cdot\widehat{p} \leq \mathbf{Pr}_{\sigma\sim\mu_C}\left[\sigma(v) = c\right] \leq e^{\varepsilon}\cdot\widehat{p}.$$

in time $\mathrm{poly}(\frac{1}{\varepsilon})$.

Before diving into the proof details, let us first imagine that we set up the LP for the whole coupling tree. To do this would require exponential amount of time, but we show that this indeed can be used to estimate the marginals to arbitrary precision. We use $\left\{\widehat{p}^x_{x,y}, \widehat{p}^y_{x,y}\right\}_{(x,y)\in\mathcal{T}}$ to denote a solution of this LP. Due to **Constraints 2**, a simple induction shows that for every $L \leq |V|$ and $\sigma \in \mathcal{C}_1$,

$$\sum_{(x,y)\in\mathcal{L}(T_L):\ \sigma\models x} \widehat{p}^x_{x,y} = 1.$$

In particular, when $L = |V|$, this means that

$$\sum_{(x,y)\in\mathcal{L}(T):\ \sigma\models x} \widehat{p}^x_{x,y} = 1.$$

Similar equalities hold on the Y side. Using this, we rewrite the ratio $\frac{|\mathcal{C}_1|}{|\mathcal{C}_2|}$ as follows:

$$\begin{aligned}\frac{|\mathcal{C}_1|}{|\mathcal{C}_2|} &= \frac{\sum_{\sigma\in\mathcal{C}_1} 1}{\sum_{\sigma\in\mathcal{C}_2} 1} = \frac{\sum_{\sigma\in\mathcal{C}_1}\sum_{(x,y)\in\mathcal{L}(\mathcal{T}):\sigma\models x}\widehat{p}^x_{x,y}}{\sum_{\sigma\in\mathcal{C}_2}\sum_{(x,y)\in\mathcal{L}(\mathcal{T}):\sigma\models y}\widehat{p}^y_{x,y}} \\ &= \frac{\sum_{(x,y)\in\mathcal{L}(\mathcal{T})}\sum_{\sigma\models x}\widehat{p}^x_{x,y}}{\sum_{(x,y)\in\mathcal{L}(\mathcal{T})}\sum_{\sigma\models y}\widehat{p}^y_{x,y}} \\ &= \frac{\sum_{(x,y)\in\mathcal{L}(\mathcal{T})}\widehat{p}^x_{x,y}\,|C_x|}{\sum_{(x,y)\in\mathcal{L}(\mathcal{T})}\widehat{p}^y_{x,y}\,|C_y|}.\end{aligned}$$

Recall $r_{x,y} = \frac{|C_x|}{|C_y|}$. By **Constraints 1**, we know that for any $(x,y) \in \mathcal{L}(\mathcal{T})$,

$$\underline{r} \leq \frac{\widehat{p}^x_{x,y}\,|C_x|}{\widehat{p}^y_{x,y}\,|C_y|} \leq \overline{r}.$$

It implies that

$$\underline{r} \leq \frac{|\mathcal{C}_1|}{|\mathcal{C}_2|} \leq \overline{r}.$$

Unfortunately, as the size and the computational cost of setting up the LP is exponential in L, we have to truncate the tree at a suitable place. The rest of our task is to show that the error caused by the truncation is small. One may notice that in the analysis above we do not use **Constraints** 3. Indeed, these constraints are used to bound the truncation error.

Intuitively, the truncation error comes from the proper colourings so that the coupling does not halt at depth L (since we cannot impose **Constraints 1** for these nodes). A naive approach would then try to show that conditioned on any proper colouring as the final output, the coupling will terminate quickly. This is unfortunately not true and there exist "bad" colourings so that the coupling does not terminate at level L with high probability. For example, given a predetermined ordering of vertices and edges, a proper colouring $\sigma \in \mathcal{C}_1$ may render all vertices encountered in Algorithm 1 with the same colour. Hence conditioned on this σ on the X side, Algorithm 1 will not stop until all edges are enumerated.

We will show, nonetheless, that the fraction of "bad" colourings is small. Let us formally define bad colourings first. We need to use the notion of $\{2,3\}$-trees. This notion dates back to Alon's parallel local lemma algorithm [Alo91].

Definition 11 ($\{2,3\}$-tree)**.** *Let $G = (V, E)$ be a graph. A set of vertices $T \subseteq V$ is a* $\{2,3\}$-tree *if (1) for any $u, v \in T$, $\mathrm{dist}_G(u, v) \geq 2$; (2) if one adds an edge between every $u, v \in T$ such that $\mathrm{dist}_G(u, v) = 2$ or 3, then T is connected.*

We will need to count the number of $\{2,3\}$-trees later for union bounds. The following lemma, due to Borgs et al. [BCKL13], counts the number of connected induced subgraphs in a graph.

Lemma 12. *Let $G = (V, E)$ be a graph with maximum degree d and $v \in V$ be a vertex. The number of connected induced subgraphs of size ℓ containing v is at most $\frac{(ed)^{\ell-1}}{2}$.*

Corollary 13. *Let $G = (V, E)$ be a graph with maximum degree d and $v \in V$ be a vertex. Then the number of $\{2,3\}$-trees in G of size ℓ containing v is at most $\frac{(ed^3)^{\ell-1}}{2}$.*

Proof. Let $G' = (V, E')$ be the graph with vertex set V and $(u, v) \in E'$ if $\mathrm{dist}_G(u, v) = 2, 3$. The degree of G' is at most d^3 and any $\{2,3\}$-tree in G is a connected set of vertices in G'. Therefore, the number of $\{2,3\}$-trees in G containing v of size ℓ can be bounded by the number of induced subgraphs in G' containing v of size ℓ. Lemma 12 then concludes the proof. $\square$

Recall that $\mathrm{Lin}(H)$ is the line graph of H, that is, vertices in $\mathrm{Lin}(H)$ are hyperedges in H and two hyperedges are adjacent if they share some vertex in H. Let $\mathrm{L}^2(H)$ be a graph whose vertices are hyperedges in H and two hyperedges are adjacent in $\mathrm{L}^2(H)$ if their distance is at most 2 in $\mathrm{Lin}(H)$. Any connected subgraph in $\mathrm{L}^2(H)$ contains a large $\{2,3\}$-tree in $\mathrm{Lin}(H)$.

Lemma 14. *Let B be a set of hyperedges which induces a connected subgraph in $\mathrm{L}^2(H)$, and $e^* \in B$ be an arbitrary hyperedge. There exists a $\{2,3\}$-tree $T \subseteq B$ such that $e^* \in T$ in $\mathrm{Lin}(H)$ and $|T| \geq \frac{|B|}{k\Delta}$.*

Proof. We construct T greedily starting from $T_0 := \{e^*\}$. Given T_i, let $B \leftarrow B \setminus \Gamma(T_i)$, and then let T_{i+1} be T_i plus the first hyperedge in B which has distance ≤ 3 from T_i. If no such hyperedge exists, the process stops.

We claim that when the process stops, all hyperedges in B are removed. If there is a nonempty subset $B' \subset B$ remaining, choose an arbitrary $e \in B'$. Since B is connected in $\mathrm{L}^2(H)$, there is a shortest path $P \subset B$ from e to some $e' \in T$ in $\mathrm{L}^2(H)$. Assume that P is $e \to \cdots \to e_1 \to e_2 \to e'$ (where e_1 is possible to be e). The minimality of $|P|$ implies that $e_1, e_2 \notin T$. If $\mathrm{dist}_{\mathrm{Lin}(H)}(T, e_2) = 1$, then $\mathrm{dist}_{\mathrm{Lin}(H)}(T, e_1) \leq 1 + \mathrm{dist}_{\mathrm{Lin}(H)}(e_1, e_2) \leq 3$ and it contradicts the construction of T as e_1 would be added to T. Otherwise $\mathrm{dist}_{\mathrm{Lin}(H)}(T, e_2) = 2$, and again it contradicts the construction of T as e_2 would be added to T.

For the size of T, notice that in every step of the process, at most $k\Delta$ hyperedges are removed. Hence $|T| \geq \frac{|B|}{k\Delta}$. $\square$

We now define *bad colourings*. Let e_0 be the first edge in $\Gamma(v)$. Recall that in the coupling process we would attempt to colour at most k_2 vertices in an edge, where $0 < k_2 < k_1$. We will have another parameter $0 < \beta < 1$, which denotes the fraction of (partially) monochromatic hyperedges[6] in a bad colouring. All parameters will be set in Section 7.

[6]A hyperedge is (partially) monochromatic if every vertex in the hyperedge is either of the same colour or not coloured.

Definition 15 (bad colourings)**.** *Let $\ell > 0$ be an integer and $\beta > 0$ be a constant. A colouring $\sigma \in \mathcal{C}_1$ is* ℓ-bad *if there exist a $\{2,3\}$-tree T in $\mathrm{Lin}(H)$ and vertices V_{col} such that*

(1) *$|T| = \ell$ and $e_0 \in T$;*
(2) *for every $e \in T$, $|e \cap V_{\mathrm{col}}| = k_2$;*
(3) *the partial colouring of σ restricted to V_{col} makes at least $\beta\ell$ hyperedges in T (partially) monochromatic.*

We say $\sigma \in \mathcal{C}_1$ is ℓ-good *if it is not ℓ-bad.*

Note that since T is a $\{2,3\}$-tree in $\mathrm{Lin}(H)$ in Definition 15, all hyperedges in T are disjoint.

We show that the fraction of bad proper colourings among all proper colourings in $\mathcal{C}_1$ is small. This allows us to throw away bad colourings in the estimates later.

Lemma 16. *Let $\Delta \geq 2$ and $0 < k_2 < k_1 \leq k$ all be integers. Let $0 < \beta < 1$ be a constant. Let $H(V, \mathcal{E})$ be a hypergraph with pinnings $\mathcal{P}$, where the maximum degree is Δ and $k_1 \leq |e| \leq k$ for every $e \in \mathcal{E}$. If $q^{1-k_2} < \beta$, $q > (ek\Delta)^{\frac{1}{k_1-2}}$, and $q > C\Delta^{\frac{3}{\beta(k_2-1)}}$ where $C^{\beta(k_2-1)} \geq \frac{e^{\beta+3}k^3}{\beta^\beta} \cdot \binom{k}{k_2}$, then we have*

$$\frac{|\{\sigma \in \mathcal{C}_1 \,:\, \sigma \text{ is } \ell\text{-bad}\}|}{|\mathcal{C}_1|} \leq e^{-\ell}.$$

Proof. Fix a $\{2,3\}$-tree $T = \{e_1, e_2, \cdots, e_\ell\}$ in $\mathrm{Lin}(H)$ of size ℓ and V_{col} such that for every $e \in T$, $|e \cap V_{\mathrm{col}}| = k_2$. We say σ is ℓ-bad with respect to T and V_{col} if σ, T, and V_{col} satisfy the requirments in Definition 15. Denote by $Z_{V_{\mathrm{col}}}$ or simply Z the number of (partially) monochromatic hyperedges by first drawing from $\mu_{\mathcal{C}_1}$ and then revealing the colours of vertices in V_{col}. We use Theorem 4 to bound the probability that $Z \geq \beta\ell$.

Indeed, $\mu_{\mathcal{C}_1}$ can be viewed as the uniform distribution over proper colourings of an instance where v is pinned to colour c_1. In this instance, we have that $k_1 - 1 \leq |e| \leq k$ for every $e \in \mathcal{E}$. Hence, in the product distribution $\mathbf{Pr}\left[e \text{ is monochromatic}\right] \leq q^{2-k_1} \leq \frac{1}{ek\Delta}$ for every $e \in \mathcal{E}$ by assumption. We set $x(e) = \frac{1}{k\Delta}$ in Theorem 4 and verify (1):

$$x(e) \prod_{e' \in \Gamma(e)} (1 - x(e')) \geq \frac{1}{k\Delta}\left(1 - \frac{1}{k\Delta}\right)^{k\Delta-1} \geq \frac{1}{ek\Delta} \geq \mathbf{Pr}\left[e \text{ is monochromatic}\right].$$

In the product distribution (where all vertices are independent), for $e \in T$, the vertices in $e \cap V_{\mathrm{col}}$ are monochromatic with probability $p^* := q^{1-k_2} < \beta$. Since T is a $\{2,3\}$-tree in $\mathrm{Lin}(H)$, all edges are disjoint and these events are independent in the product distribution. Hence, by a multiplicative Chernoff bound with mean $p^*\ell$ and $\gamma = \frac{\beta}{p^*} - 1 > 0$,

$$\mathbf{Pr}\left[Z \geq \beta\ell\right] = \mathbf{Pr}\left[Z \geq (1+\gamma)p^*\ell\right] \leq \left(\frac{e^\gamma}{(1+\gamma)^{1+\gamma}}\right)^{p^*\ell} \leq \left(\frac{ep^*}{\beta}\right)^{\beta\ell}.$$

For each edge $e \in T$, there are at most $k(\Delta - 1) + 1 \leq k\Delta - 1$ edges that intersect with e (including itself). The random variable Z thus depends on at most $(k\Delta - 1)\ell$ hyperedges in $\mu_{\mathcal{C}_1}$. By Theorem 4 with $x(e) = \frac{1}{k\Delta}$,

$$\begin{aligned}\mu_{\mathcal{C}_1}(Z \geq \beta\ell) &\leq \mathbf{Pr}\left[Z \geq \beta\ell\right] \cdot \left(1 - \frac{1}{k\Delta}\right)^{-(k\Delta-1)\ell} \\ &\leq \left(\frac{ep^*}{\beta}\right)^{\beta\ell} \cdot e^\ell = \left(\frac{e^{1+1/\beta}p^*}{\beta}\right)^{\beta\ell}.\end{aligned}$$

To finish the argument, we still need to account for all $\{2,3\}$-trees and V_{col} by an union bound. Since the maximum degree in $\mathrm{Lin}(H)$ is $k\Delta$, the total number of $\{2,3\}$-trees containing e_0 of size ℓ, by Corollary 13, is at most $\frac{\left(e(k\Delta)^3\right)^\ell}{2}$. For a fixed T, since all edges in T are disjoint, the number of possible V_{col} is at most $\binom{k}{k_2}^\ell$.

Putting everything together, we have that

$$\Pr_{\sigma\sim\mu_{\mathcal{C}_1}}[\sigma \text{ is } \ell\text{-bad}] \le \left(\frac{e^{1+1/\beta}p^*}{\beta}\right)^{\beta\ell} \cdot \frac{(e(k\Delta)^3)^\ell}{2} \cdot \binom{k}{k_2}^\ell$$
$$\le \left(\frac{e^{\beta+1}}{\beta^\beta} \cdot ek^3 \cdot \binom{k}{k_2} \cdot q^{\beta-\beta k_2}\Delta^3\right)^\ell.$$

By assumption,

$$q^{\beta k_2-\beta} \ge C^{\beta(k_2-1)}\Delta^3 \ge \frac{e^{\beta+2}}{\beta^\beta}\cdot ek^3\cdot\binom{k}{k_2}\cdot\Delta^3.$$

Combining these two inequalities finishes the proof. □

Let $(x,y)\in\mathcal{T}$ be a pair of partial colourings defined on V_{col}. We are now going to prove some structural properties of (x,y). Say an edge $e\in\mathcal{E}$ such that $e\cap V_{\text{col}}\neq\varnothing$ is *blocked* by (x,y) if one of the following holds

(1) $x(u)\neq y(u)$ for some $u\in e$.
(2) $|e\cap V_{\text{col}}|=k_2$ and e is not satisfied by both x and y.

These two cases are called type 1 and type 2 errors respectively in [Moi19]. Notice that all edges in $\Gamma(v)$ are always blocked, and in particular, e_0 is always blocked.

Let us denote the set of edges blocked by (x,y) as $\mathcal{B}_{x,y}$. Then $\mathcal{B}_{x,y}$ always contains a large $\{2,3\}$-tree.

Lemma 17. *Let $(x,y)\in\mathcal{T}$ be a pair of partial colourings in the coupling tree defined on V_{col} with corresponding V_1. Assume $|V_{\text{col}}|=L$. There exists a $\{2,3\}$-tree $T\subseteq\mathcal{B}_{x,y}$ in $\text{Lin}(H)$ of size at least $\frac{L}{k^3\Delta^2}$ containing e_0.*

Proof. We first claim that $\mathcal{B}_{x,y}$ is connected in $\mathrm{L}^2(H)$ by inducting on L. Once an edge is blocked during Algorithm 1, it will remain blocked till the end. If u is the next vertex to be coloured in Algorithm 1, then u must be adjacent to some vertex $u'\in V_1$, and u' is in some edge e blocked by the current (x,y). Therefore any newly blocked edge caused by colouring u has distance at most 2 to e.

Since e_0 is always blocked, $e_0\in\mathcal{B}_{x,y}$. By Lemma 14, there exists a $\{2,3\}$-tree $T\subseteq\mathcal{B}_{x,y}$ in $\text{Lin}(H)$ such that $|T|\ge\frac{|\mathcal{B}_{x,y}|}{k\Delta}$. Next we claim that $|\mathcal{B}_{x,y}|\ge\frac{L}{k^2\Delta}$. This is because that every vertex in V_1 belongs to some blocked edge. Hence $|V_1|\le k\,|\mathcal{B}_{x,y}|$. By item (1) of Lemma 8, $V_{\text{col}}\subseteq\Gamma_{\text{ver}}(V_1)$. It implies that $L=|V_{\text{col}}|\le|\Gamma_{\text{ver}}(V_1)|\le k\Delta\,|V_1|$. Combining these facts yields the lemma. □

Recall that $\mathcal{T}_L$ is the tree obtained from $\mathcal{T}$ by truncating at depth L, and $\mathcal{L}(\mathcal{T}_L)$ is its leaves. Because of **Constraints 2**, for every proper colouring $\sigma\in\mathcal{C}_1$, it holds that

$$\sum_{(x,y)\in\mathcal{L}(\mathcal{T}_L):\ \sigma\models x} p^x_{x,y}=1. \tag{12}$$

However, in **Constraints 1**, our linear program only contains constraints for those $\underline{p}^x_{x,y}$ and $\underline{p}^y_{x,y}$ whose V_{col} is of size strictly smaller than L. The next lemma shows that, for a ℓ-good colouring σ, solving $\underline{p}^x_{x,y},\underline{p}^y_{x,y}$ provides a good approximation for the identity (12).

Lemma 18. *Let $0<\beta<1$ be a constant. Let $H=(V,\mathcal{E})$ be a hypergraph with pinnings $\mathcal{P}$ and maximum degree Δ such that $|e|\le k$ for all $e\in\mathcal{E}$. Let $\sigma\in\mathcal{C}_1$ be ℓ-good where ℓ is an integer. If $\{\widehat{p}^x_{x,y}\}$ is a collection of values satisfying all our linear constraints, with $t^*=5\left(e^2k^3\Delta^3\right)^{\frac{1}{1-\beta}}$ in **Constraints 3** up to level $L=k^3\Delta^2\ell$, then it holds that*

$$\sum_{\substack{(x,y)\in\mathcal{L}(\mathcal{T}_L):\ |V_{\text{col}}|<L\\ \text{and } \sigma\models x}} \widehat{p}^x_{x,y}\ge 1-e^{-\ell}. \tag{13}$$

Proof. We construct a new coupling process similar to Algorithm 1, and show the left-hand side of (13) is the probability of an event defined by the new process. We modify $\mathbb{S}$ in the following two ways: (1) condition on the final output being σ; (2) use probabilities induced by $\left\{\widehat{p}^{x}_{x,y}\right\}$ instead of $\left\{\mathrm{p}^{x}_{x,y}\right\}$. To be more specific, consider each step where one needs to extend (x, y) defined on V_{col} to a new vertex u. Call the new colourings (x', y'). Since the output σ is fixed, we simply reveal $x'(u) = \sigma(u)$. In the original $\mathbb{S}$, the colour of $y'(u)$ is drawn according to an optimal coupling of (x', y') on u. Here, we set $y'(u)$ to colour c with probability $\frac{\widehat{p}^{x^{u\leftarrow\sigma(u)}}_{x^{u\leftarrow\sigma(u)},y^{u\leftarrow c}}}{\widehat{p}^{x}_{x,y}}$. This is well-defined since $\left\{\widehat{p}^{x}_{x,y}\right\}$ satisfies **Constraints 2**. If this process reaches depth L, then it stops.

The output of the new coupling defines a distribution over pairs of partial colourings (x, y) such that $\sigma \models x$ and we denote it by $\widehat{\mu}$. We claim that

$$\sum_{\substack{(x,y)\in\mathcal{L}(\mathcal{T}_L):\ |V_{\mathrm{col}}|=L \\ \text{and } \sigma\models x}} \widehat{p}^{x}_{x,y} \le \sum_{\substack{\{2,3\}\text{-tree } T: \\ |T|=\ell,\ e_0\in T}} \Pr_{(X,Y)\sim\widehat{\mu}} \left[T \subseteq \mathcal{B}_{X,Y}\right]. \tag{14}$$

Each summand on the left-hand side of (14) is the probability that our new coupling reaches some (x, y) with $|V_{\mathrm{col}}| = L$. Lemma 17 implies that the set $\mathcal{B}_{x,y}$ of blocked edges contains a $\{2, 3\}$-tree T of size at least $\frac{L}{k^3\Delta^2} = \ell$. Thus the probability of reaching vertices of depth L is upper bounded by the right-hand side of (14).

Fix a $\{2, 3\}$-tree T of size ℓ. Since σ is ℓ-good, whatever the choice of V_{col} is, at least a $(1 - \beta)$ fraction of hyperedges in T must not be monochromatic on the X side. However, if $T \subseteq \mathcal{B}_{X,Y}$, then at least $\lfloor(1 - \beta)\,|T|\rfloor$ hyperedges satisfy (1) $\sigma(v) \neq Y(v)$ for some $v \in e \cap V_{\mathrm{col}}$, or (2) $|e \cap V_{\mathrm{col}}| = k_2$ and $\sigma|_{V_{\mathrm{col}}} = X|_{V_{\mathrm{col}}}$ satisfies e but Y does not satisfy e. It is clear that case (2) implies case (1), since if one partial colouring satisfies e and another one does not, then they must differ at some $v \in e\cap V_{\mathrm{col}}$. We use $T' = \left\{e_1, e_2, \ldots, e_{|T'|}\right\}$ to denote these hyperedges in T. For each hyperedge in T', there must be at least one vertex on which the (modified) coupling fails, which happens with probability at most $5/t^*$ due to **Constraints 3**. Since T is a $\{2, 3\}$-tree in $\mathrm{Lin}(H)$, all of these failed couplings are for distinct vertices and thus happen independently. Hence, in this new coupling, the probability that every edge in T' is blocked due to at least one failed vertex is at most $\left(\frac{5}{t^*}\right)^{|T'|} \le \left(\frac{5}{t^*}\right)^{\lfloor(1-\beta)\ell\rfloor}$.

We still need to apply a union bound. The number of $\{2, 3\}$-trees of size ℓ in $\mathrm{Lin}(H)$ and containing e_0 is, by Corollary 13, at most $\frac{\left(ek^3\Delta^3\right)^\ell}{2}$. Therefore the right-hand side of (14) is at most

$$\sum_{\substack{\{2,3\}\text{-tree } T: \\ |T|=\ell,\ e_0\in T}} \Pr_{(X,Y)\sim\widehat{\mu}} \left[T \subseteq \mathcal{B}_{X,Y}\right] \le \left(\frac{5}{t^*}\right)^{\lfloor(1-\beta)\ell\rfloor} \cdot \frac{\left(ek^3\Delta^3\right)^\ell}{2} \le e^{-\ell}, \tag{15}$$

since we have chosen $t^* = 5\left(e^2k^3\Delta^3\right)^{\frac{1}{1-\beta}}$ in **Constraints 3**. The lemma follows by combining (12), (14), and (15). ☐

Note that in Lemma 18 we do not explicitly require a lower bound of q nor a lower bound on the size of the edges. However, these requirements are implicit since we have set t^* to be large in **Constraints 3**.

Lemma 16 and Lemma 18 also hold for any $\sigma \in \mathcal{C}_2$. Now we can prove that any solution to the LP provides accurate estimates.

Lemma 19. *Assume the settings of Lemma 16 and Lemma 18. If the linear program up to level L has a solution $\left\{\widehat{p}^{x}_{x,y}, \widehat{p}^{y}_{x,y}\right\}$ with guessed bounds $\left\{\underline{\widehat{r}}, \widehat{\overline{r}}\right\}$, then it holds*

$$e^{-\gamma}\underline{\widehat{r}} \le \frac{|\mathcal{C}_1|}{|\mathcal{C}_2|} \le e^{\gamma}\widehat{\overline{r}},$$

where $\gamma = 4e^{-\frac{L}{k^3\Delta^2}}$.

Proof. Let $\ell = \frac{L}{k^3\Delta^2}$. Let

$$Z_1 := \sum_{\sigma\in\mathcal{C}_1} \sum_{\substack{(x,y)\in\mathcal{L}(\mathcal{T}):\, |V_{\mathrm{col}}|<L \\ \text{and } \sigma\models x}} \widehat{p}^{x}_{x,y}.$$

Exchange the order of summation:

$$Z_1 = \sum_{(x,y)\in\mathcal{L}(\mathcal{T}):\, |V_{\mathrm{col}}|<L} \sum_{\sigma\in\mathcal{C}_1:\, \sigma\models x} \widehat{p}^{x}_{x,y} = \sum_{(x,y)\in\mathcal{L}(\mathcal{T}):\, |V_{\mathrm{col}}|<L} \widehat{p}^{x}_{x,y}\cdot|\mathcal{C}_x|\,.$$

A similar quantity Z_2 can be defined and bounded by replacing $\widehat{p}^{x}_{x,y}$ with $\widehat{p}^{y}_{x,y}$. **Constraints 1** impose that for any $(x,y)\in\mathcal{L}(\mathcal{T})$ such that $|V_{\mathrm{col}}|<L$,

$$\underline{\widehat{r}} \le \frac{\widehat{p}^{x}_{x,y}\cdot|\mathcal{C}_x|}{\widehat{p}^{y}_{x,y}\cdot|\mathcal{C}_y|} \le \overline{\widehat{r}}.$$

Hence,

$$\underline{\widehat{r}} \le \frac{Z_1}{Z_2} \le \overline{\widehat{r}}. \tag{16}$$

We will relate $|\mathcal{C}_1|$ with Z_1. It is easy to see, by (12), that

$$|\mathcal{C}_1| = \sum_{\sigma\in\mathcal{C}_1} 1 = \sum_{\sigma\in\mathcal{C}_1}\sum_{(x,y)\in\mathcal{L}(\mathcal{T}_L):\, \sigma\models x} \widehat{p}^{x}_{x,y} \ge Z_1. \tag{17}$$

The lower bound is more complicated:

$$\begin{aligned}
|\mathcal{C}_1| = \sum_{\sigma\in\mathcal{C}_1} 1 &\le \left(1-e^{-\ell}\right)^{-1} \sum_{\substack{\sigma\in\mathcal{C}_1: \\ \sigma \text{ is } \ell\text{-good}}} 1 \\
&\le \left(1-e^{-\ell}\right)^{-1}\left(1-e^{-\ell}\right)^{-1} \sum_{\substack{\sigma\in\mathcal{C}_1: \\ \sigma \text{ is } \ell\text{-good}}} \sum_{\substack{(x,y)\in\mathcal{L}(\mathcal{T}):\, |V_{\mathrm{col}}|<L \\ \text{and } \sigma\models x}} \widehat{p}^{x}_{x,y} \\
&\le e^{\gamma} \sum_{\sigma\in\mathcal{C}_1} \sum_{\substack{(x,y)\in\mathcal{L}(\mathcal{T}):\, |V_{\mathrm{col}}|<L \\ \text{and } \sigma\models x}} \widehat{p}^{x}_{x,y} = e^{\gamma} Z_1, \qquad (18)
\end{aligned}$$

where in the first line we use Lemma 16 and in the second line we use Lemma 18. Similar bounds hold with $|\mathcal{C}_2|$ and Z_2. Combining (16), (17), (18), and their counterparts for $|\mathcal{C}_2|$ and Z_2, we have that

$$e^{-\gamma}\underline{\widehat{r}} \le \frac{|\mathcal{C}_1|}{|\mathcal{C}_2|} \le e^{\gamma}\overline{\widehat{r}}. \qquad \square$$

We then set up a binary search, to find $\underline{r}$ and $\overline{r}$ that are close enough to the true ratio. We are now ready to prove the main theorem of this section.

Proof of Theorem 10. Take $L = k^3\Delta^2\left\lceil\log\left(\frac{4}{\varepsilon}\right)\right\rceil$ so that $\gamma = 4e^{-\frac{L}{k^3\Delta^2}} \le \varepsilon$. We claim the true values of $\left\{p^{x}_{x,y}, p^{y}_{x,y}\right\}$ always satisfy our LP. This is trivial for **Constraints 1** and **2**. For **Constraints 3**, recall that $t^* = 5\left(e^2k^3\Delta^3\right)^{\frac{1}{1-\beta}} > k$ and we only need to verify the conditions of Lemma 7 with $t = t^*$. At any point of Algorithm 1, the size of an edge is at least $k_1 - k_2$. Hence we set $k' = k_1 - k_2$ in Lemma 7. By our assumption,

$$q > C\Delta^{\frac{4-\beta}{(1-\beta)(k_1-k_2-1)}} \ge \left(5e\left(e^2k^3\right)^{\frac{1}{1-\beta}}\right)^{\frac{1}{k'-1}}\cdot\Delta^{\frac{4-\beta}{(1-\beta)(k'-1)}} = \left(et^*\Delta\right)^{\frac{1}{k'-1}}.$$

Fix the colour c. It follows from Lemma 19 that for every $c' \in [q]$, we can apply the binary search algorithm to obtain a value $p_{c'}$, which is an estimate of $\frac{\mathbf{Pr}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v)=c']}{\mathbf{Pr}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v)=c]}$ satisfying

$$e^{-\varepsilon}\cdot p_{c'} \le \frac{\mathbf{Pr}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v)=c']}{\mathbf{Pr}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v)=c]} \le e^{\varepsilon}\cdot p_{c'}.$$

We then use $\widehat{p} := \left(\sum_{c'\in[q]} p_{c'}\right)^{-1}$ to estimate $\mathbf{Pr}_{\sigma\sim\mu_{\mathcal{C}}}[\sigma(v)=c]$.

For the running time, we treat Δ, k, and q as constants. The size of the linear program in the WHILE loop is $\exp(O(L))$. This is because the coupling tree $\mathcal{T}$ is q^2-ary, and therefore it has at most $\exp(O(L))$ vertices up to depth L, and we have a pair of variables $\underline{p}^x_{x,y}$ and $\underline{p}^y_{x,y}$ for each vertex. The number of variables and the number of constraints is at most $\exp(O(L))$. Note that for each set of constraints in **Constraints 1**, we need to enumerate all the possible colourings in V_1 to compute $r_{x,y}$ for every leaf (x,y). This costs at most $\exp(O(L))$ time. Hence it takes $\exp(O(L))$ time to construct an LP of size $\exp(O(L))$, which requires again $\exp(O(L))$ time to solve. Note that with our choice of L, $\exp(O(L)) = \mathrm{poly}\left(\frac{1}{\varepsilon}\right)$. For the WHILE loop, we use binary search to find $\underline{r}$ and $\overline{r}$. Thus the number of loops of the binary search is at most $\log_2 \frac{2}{e^{\varepsilon}} = \mathrm{poly}\left(\frac{1}{\varepsilon}\right)$. Therefore, the total running time of our estimator is $\mathrm{poly}\left(\frac{1}{\varepsilon}\right)$. ☐

5. Approximate counting

Now we give our FPTAS for the number of proper q-colourings of a k-uniform hypergraph H with maximum degree Δ. The next lemma guarantees us a "good" proper colouring σ so that we can use the algorithm in Theorem 10 to compute the marginal probability of σ.

Lemma 20. *Let k_1^C be an integer such that $0 < k_1^C < k-1$. Let $q \ge \left(4(k-k_1^C)\Delta\right)^{\frac{1}{k-k_1^C-1}}$. Let $v_1,\dots,v_n$ be an arbitrary ordering of the vertices of a k-uniform hypergraph $H=(V,\mathcal{E})$. There exists a proper colouring σ such that for every hyperedge $e\in\mathcal{E}$, the partial colouring σ restricted to the first $k-k_1^C$ vertices is not monochromatic. Moreover, σ can be found in deterministic polynomial time.*

Proof. Let $k' = k - k_1^C$. Consider a new hypergraph $H' = (V,\mathcal{E}')$ on the same vertex set V, but for every $e\in\mathcal{E}$, we replace it with its first k' vertices. We set $x(e) = \frac{1}{k'\Delta}$ in Theorem 3 and verify (1) for every $e\in\mathcal{E}'$,

$$x(e)\prod_{e'\in\Gamma(e)}(1-x(e')) \ge \frac{1}{k'\Delta}\left(1-\frac{1}{k'\Delta}\right)^{k'(\Delta-1)} \ge \frac{1}{ek'\Delta} \ge q^{1-k'} \ge \mathbf{Pr}\left[e \text{ is monochromatic}\right].$$

Hence, Theorem 3 implies that there exists a proper colouring σ in H', which satisfies the requirement of the lemma.

In order to find σ, we have left a bit slack in our bound on q. Thus the deterministic algorithm from [MT10] applies. ☐

Theorem 21. *Assume the conditions of Theorem 10 (on q, Δ, k, k_1, k_2, and β) with $k_1 = k_1^C$ hold, together with the conditions of Lemma 20. There is an FPTAS for the number of proper q-colourings of a k-uniform hypergraph $H=(V,\mathcal{E})$ with maximum degree Δ.*

Proof. Let $n = |V|$. Choose an arbitrary ordering of the vertices $v_1,\dots,v_n$ of V. Lemma 20 implies that we can find a proper colouring σ so that any hyperedge is properly coloured by the first $k-k_1^C$ of its vertices. Let $Z = |\mathcal{C}|$ be the number of proper colourings of H. For every $\varepsilon > 0$, we will deterministically compute a number $\widehat{Z}$ in time polynomial in n and $1/\varepsilon$ such that $e^{-\varepsilon}\widehat{Z} \le Z \le e^{\varepsilon}\widehat{Z}$.

As before, let $\mu_{\mathcal{C}}$ be uniform over $\mathcal{C}$, the set of all proper colourings of H. We will actually estimate $\mu_{\mathcal{C}}(\sigma) = \frac{1}{Z}$. To this end, we create a sequence of hypergraphs $\{H_i\}$ with pinnings $\{\mathcal{P}_i\}$ inductively. Let $H_1 = H$ and $\mathcal{P}_1$ be empty. Given $H_i = (V_i,\mathcal{E}_i)$ and $\mathcal{P}_i$, we find the next vertex u_i under the ordering that are contained in at least one hyperedge of H_i. We pin the colour of

u_i to be $\sigma(u_i)$. This induces a pinning $\mathcal{P}_{i+1}$ on all hyperedges in $\mathcal{E}_i$. Then, H_{i+1} is obtained by removing u_i from V_i and removing all hyperedges that are properly coloured under $\mathcal{P}_{i+1}$ from $\mathcal{E}_i$. We also truncate the pinning $\mathcal{P}_{i+1}$ accordingly. If for some $n' \leq n$, $\mathcal{E}_{n'}$ is empty, then this process terminates. Notice that the construction above yields a subset of vertices $u_1, \dots, u_{n'}$ where $n' \leq n$. Their ordering is consistent with the given ordering.

We claim that for any $i \in [n']$, for any $e \in \mathcal{E}_i$, it satisfies that $k_1^C \leq |e| \leq k$. This is because an edge e shrinks in size in the process when vertices are pinned according to σ. However, Lemma 20 guarantees that the edge e will be removed in the process above before $k - k_1^C$ vertices are coloured. Therefore, together with our assumptions, Theorem 10 applies with $k_1 = k_1^C$.

Let p_i be the marginal probability of colour $\sigma(u_i)$ at u_i in H_i with pinning $\mathcal{P}_i$. Let $p_i = \frac{1}{q}$ for all $i \geq n'$. It is easy to see that $Z^{-1} = \mu_{\mathcal{C}}(\sigma) = \prod_{i=1}^{n} p_i$. Thus we can obtain our desired estimate $\widehat{Z}$ by approximating each p_i within $e^{\pm\frac{\varepsilon}{n}}$. To this end, we appeal to Theorem 10 with $\varepsilon' = \frac{\varepsilon}{n}$. □

6. Sampling

Finally we give the algorithm to sample proper colourings almost uniformly. As usual, let $H(V, \mathcal{E})$ be a k-uniform hypergraph with maximum degree Δ, q be the number of colours, and $\mathcal{C}$ be the set of proper colourings. Let $n = |V|$. Algorithm 2 samples a colouring in $\mathcal{C}$ within total variation distance ε from $\mu_{\mathcal{C}}$. Similar to the coupling process in Section 3, we assume that there is an arbitrary fixed ordering of all vertices and hyperedges. There is a parameter $0 < k_1^S < k - 1$ in Algorithm 2, which will be set in Section 7.

Algorithm 2 An almost uniform sampler for proper colourings

1: **Input:** *A k-uniforom hypergraph $H(V, \mathcal{E})$ with maximum degree Δ and* $0 < \varepsilon < 1$
2: **Output:** *A colouring in $\mathcal{C}$*
3: Let X be the partial colouring that $X(v) = -$ for every $v \in V$ initially;
4: **while** $\mathcal{E}$ is nonempty **do**
5: Choose the first uncoloured $v \in V$ such that every $e \in \Gamma(v)$ contains $> k_1^S$ uncoloured vertex;
6: **if** no such vertex v exists **then**
7: **break**
8: **end if**
9: Apply the algorithm in Theorem 10 to compute the marginal distribution on v with precision $\frac{\varepsilon}{2n}$, and extend X with the colour on v according to the distribution;
10: Remove from $\mathcal{E}$ all hyperedges that are now satisfied.
11: **end while**
12: $S \leftarrow$ uncoloured vertices in V;
13: Let $H_S = (S, \mathcal{E}_S)$ where $\mathcal{E}_S := \{e \cap S \,:\, e \in \mathcal{E}\}$;
14: **if** H_S contains a connected component with size at least $k^2\Delta\log\left(\frac{2n\Delta}{\varepsilon}\right)$ **then**
15: **return** an arbitary $x \in \mathcal{C}$
16: **else**
17: **return** a uniformly random proper colouring consistent with X by enumerating all proper colourings of H_S.
18: **end if**

We first assume that at Line 9, the oracle call to Theorem 10 is always within the correct range. This simplification allows us to identify a threshold involving the parameter k_1^S to guarantee small connected components, which will be put together with the conditions of Theorem 10 later.

Lemma 22. *Assume the oracle call to Theorem 10 at Line 9 is within the desired range. If $q > (ek\Delta)^{\frac{1}{k_1^S-1}}$ and $q > C\Delta^{\frac{3}{k-k_1^S-1}}$ where $C^{(k-k_1^S)-1} > e^7k^3$, the condition in line 14 of Algorithm 2 holds with probability at most $\varepsilon/2$.*

Proof. The proof idea is to show the existence of a large components in H_S implies the existence of a large $\{2,3\}$-tree in $\mathrm{Lin}(H)$ whose vertices are edges that are not satisfied but $k-k_1^S$ of their vertices are already coloured. Then we show the probability of the latter event is small.

Now assume that the sampler ends the WHILE loop with a partial colouring X and H_S. We say an edge $e\in\mathcal{E}$ is *bad* if X does not satisfy e and $|e\cap S|=k_1^S$, namely e is partially monochromatic under X but $k-k_1^S$ vertices have been coloured. Also, say a vertex $v\in S$ is *blocked* by an edge $e\in\mathcal{E}$ if $v\in e$ and e is bad.

Fix an arbitrary hyperedge e_0 that is bad, and e_0 is contained in a connected component of size at least L in H_S. We denote the set of vertices of this component by U and its induced hypergraph H_U. It is clear that every vertex in S is blocked by some bad edge. Let $\mathcal{F}$ be the set of all bad edges incident to U. Then $e_0\in\mathcal{F}$. Since every vertex in U is blocked by some edge in $\mathcal{F}$ and every edge in $\mathcal{F}$ contains at most k vertices, $|\mathcal{F}|\geq \frac{L}{k}$.

We claim that $\mathcal{F}$ is connected in $\mathrm{L}^2(H)$. The reason is the following. For any two edges, say $e_1,e_2\in\mathcal{F}$, since H_U is connected, there exists a path in H_U connecting e_1 and e_2. Every vertex along this path must be blocked by some edge in $\mathcal{F}$. Each adjacent pair of vertices along this path corresponds to a pair of edges in $\mathcal{F}$ that have distance at most 2 in $\mathrm{Lin}(H)$.

Lemma 14 implies that $\mathcal{F}$ contains a $\{2,3\}$-tree of size at least $\ell=\frac{L}{k^2\Delta}$ containing e_0. Fix such a $\{2,3\}$-tree $T=\left\{e_1,\dots,e_{|T|}\right\}$. Let $\widehat{\mu}$ be the distribution of our sampler at the end of the WHILE loop. It holds that

$$\Pr_{X\sim\widehat{\mu}}\left[\text{every } e_i\in T \text{ is bad}\right]=\prod_{i=1}^{|T|}\Pr_{X\sim\widehat{\mu}}\left[e_i \text{ is bad}\;\middle|\;\bigwedge_{j<i} e_j \text{ is bad}\right].$$

Since $e_i\cap e_j=\varnothing$ for every $i\neq j$ and Theorem 10 guarantees our estimated marginals are within $e^{\varepsilon/2n}$, for every $1\leq i\leq |T|$, we can apply Lemma 7 with $k'=k_1^S$ and $t=k$,

$$\Pr_{X\sim\widehat{\mu}}\left[e_i \text{ is bad}\;\middle|\;\bigwedge_{j<i} e_j \text{ is bad}\right]\leq q\cdot q^{-(k-k_1^S)}\cdot(1+8/t)^{k/2}\cdot e^{\frac{\varepsilon(k-k_1^S)}{2n}}\leq e^5\cdot q^{1-(k-k_1^S)}.$$

Applying Lemma 7 requires that $q>(ek\Delta)^{\frac{1}{k_1^S-1}}$. By Corollary 13, the number of $\{2,3\}$-trees of size ℓ in $\mathrm{Lin}(H)$ containing e_0 in $\mathcal{F}$ is at most $\frac{\left(ek^3\Delta^3\right)^\ell}{2}$. Then by the union bound, the probability that H_S contains a component with size at least L is at most

$$n\Delta\left(ek^3\Delta^3\right)^\ell\left(e^5\cdot q^{1-(k-k_1^S)}\right)^\ell, \tag{19}$$

where the term $|n\Delta|\geq|\mathcal{E}|$ accounts for the choice of e_0. By assumption,

$$q^{(k-k_1^S)-1}>C^{(k-k_1^S)-1}\Delta^3>e^7k^3\Delta^3.$$

As $L=k^2\Delta\log\left(\frac{2n\Delta}{\varepsilon}\right)$ and $\ell=\frac{L}{k^2\Delta}$, $e^{-\ell}\leq\frac{\varepsilon}{2n\Delta}$. Hence, by (19) the probability in Line 14 is at most

$$n\Delta\left(ek^3\Delta^3\right)^\ell\left(e^5\cdot q^{1-(k-k_1^S)}\right)^\ell\leq n\Delta\cdot e^{-\ell}\leq\frac{\varepsilon}{2}.$$ □

Now we are ready to give the sampling algorithm.

Theorem 23. *Assume the conditions of Theorem 10 (on q, Δ, k, k_1, k_2, and β) with $k_1=k_1^S$ hold, together with the conditions of Lemma 22. For any k-uniform hypergraph $H=(V,\mathcal{E})$ with maximum degree Δ and $\varepsilon>0$, Algorithm 2 outputs a proper colouring whose distribution is within ε total variation distance to the uniform distribution, and the running time is* $\mathrm{poly}(n,\frac{1}{\varepsilon})$ *where $n=|V|$.*

Proof. First we check that the condition of Theorem 10 is met with $k_1=k_1^S$, when it is called in Algorithm 2 at Line 9. This is because whenever we colour a vertex, we make sure that all hyperedges have at least k_1^S uncoloured vertices afterwards. Hence we apply Theorem 10 with the pinnings $\mathcal{P}$ induced by the partial colouring X so far.

We use $\widehat{\mu}(\cdot)$ to denote the distribution of the final output of Algorithm 2. Recall thet $\mu_{\mathcal{C}}$ is the uniform distribution over $\mathcal{C}$. We shall bound the total variation distance $\mathrm{dist}_{TV}(\mu_{\mathcal{C}}, \widehat{\mu})$. To this end, we introduce two intermediate distributions: Let $\mu_1(\cdot)$ be the distribution obtained from the output of Algorithm 2 but ignoring the condition on line 14 in Algorithm 2. Namely, it never checks the size of connected components in H_S and proceed to enumerate all the proper colourings on S in any case. This is unrealistic since doing so would require exponential time. We also define another distribution $\mu_2(\cdot)$, which is the same as $\mu_1(\cdot)$ except at line 9, it uses the true marginal instead of the estimate by calling Theorem 10.

Denote by B the event that the condition on line 14 holds. Let p_{fail} be the probability of event B. By Lemma 22, $p_{\mathrm{fail}} \le \varepsilon/2$.

First note that $\mu_2 = \mu_{\mathcal{C}}$. Consider the distribution of the partial colouring obtained immediately after the WHILE loop, i.e., the partial colouring X. One can apply induction similar to the proof of Lemma 8 to show that it follows a pre-Gibbs distribution. Therefore, conditioned on X, sampling a uniform proper colouring of the remaining vertices results in a uniform proper colouring.

We then bound $\mathrm{dist}_{TV}(\mu_1, \mu_2)$. For a particular partial colouring x, we use E_x to denote the event that the sampler produces x at the end of the WHILE loop, namely $X = x$. It holds that

$$\begin{aligned}\mathrm{dist}_{TV}(\mu_1,\mu_2) &= \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left|\Pr_{Z\sim\mu_1}[Z=\sigma] - \Pr_{Z\sim\mu_2}[Z=\sigma]\right| \\ &= \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left|\sum_{x\colon \sigma\models x}\left(\Pr_{Z\sim\mu_1}[Z=\sigma\mid E_x]\cdot\Pr_{Z\sim\mu_1}[E_x] - \Pr_{Z\sim\mu_2}[Z=\sigma\mid E_x]\cdot\Pr_{Z\sim\mu_2}[E_x]\right)\right|,\end{aligned}$$

where x runs over partial colourings.

The partial colouring x may never appear at the end of the WHILE loop in Algorithm 2. In this case,

$$\mathbf{Pr}_{Z\sim\mu_1}[E_x] = \mathbf{Pr}_{Z\sim\mu_2}[E_x] = 0.$$

Otherwise x can be the partial colouring at the end of the WHILE loop. Since the enumeration steps are identical and correct in both μ_1 and μ_2 conditioned on E_x, we have that

$$\mathbf{Pr}_{Z\sim\mu_1}[Z=\sigma\mid E_x] = \mathbf{Pr}_{Z\sim\mu_2}[Z=\sigma\mid E_x] = \frac{\mathbf{1}_{\sigma\models x}}{|\mathcal{C}_x|},$$

where $\mathcal{C}_x$ is again the set of proper colourings consistent with the partial colouring x.

It implies that

$$\mathrm{dist}_{TV}(\mu_1,\mu_2) = \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left|\sum_{x\colon\sigma\models x}\frac{1}{|\mathcal{C}_x|}\left(\Pr_{Z\sim\mu_1}[E_x] - \Pr_{Z\sim\mu_2}[E_x]\right)\right|. \tag{20}$$

Fix a partial colouring x defined on $V_{\mathrm{col}} \subseteq V$ that is a possible output of the WHILE loop. We note that the order of visiting V_{col} is determined by the random choices of x. Say this order is $v_1, \dots, v_s$. Let

$$p_i := \mathbf{Pr}_{Z\in\mu_C}\left[Z(v_i) = x(v_i)\;\middle|\;\bigwedge_{1\le j<i} Z(v_j) = x(v_j)\right].$$

Hence

$$\Pr_{Z\sim\mu_1}[E_x] - \Pr_{Z\sim\mu_2}[E_x] = \prod_{i=1}^{s}\widehat{p}_i - \prod_{i=1}^{s} p_i,$$

where $\widehat{p}_i$ is our estimate of p_i using Theorem 10 with error $\frac{\varepsilon}{2n}$. Theorem 10 implies that

$$e^{-\frac{\varepsilon}{2n}}\widehat{p}_i \le p_i \le e^{\frac{\varepsilon}{2n}}\widehat{p}_i.$$

Therefore, we have

$$\left| \mathop{\mathbf{Pr}}_{Z\sim\mu_1}[E_x] - \mathop{\mathbf{Pr}}_{Z\sim\mu_2}[E_x] \right| \le \varepsilon \mathop{\mathbf{Pr}}_{Z\sim\mu_2}[E_x]. \tag{21}$$

Plugging (21) into (20), we obtain

$$\mathrm{dist}_{TV}(\mu_1,\mu_2) \le \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left|\sum_{x:\,\sigma\models x}\frac{\varepsilon}{|\mathcal{C}_x|}\mathop{\mathbf{Pr}}_{Z\sim\mu_2}[E_x]\right| = \frac{\varepsilon}{2}\sum_{\sigma\in\mathcal{C}}\mu_2(\sigma) = \frac{\varepsilon}{2}.$$

Finally we bound $\mathrm{dist}_{TV}(\widehat{\mu},\mu_1)$. Since the behaviours of $\widehat{\mu}$ and μ_1 are identical if B does not happen, we have that $\mathbf{Pr}_{Z\sim\widehat{\mu}}\left[Z=\sigma \mid \overline{B}\right] = \mathbf{Pr}_{Z\sim\mu_1}\left[Z=\sigma \mid \overline{B}\right]$. It implies that

$$\begin{aligned}
\mathrm{dist}_{TV}(\widehat{\mu},\mu_1) &= \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left|\mathop{\mathbf{Pr}}_{Z\sim\widehat{\mu}}[Z=\sigma] - \mathop{\mathbf{Pr}}_{Z\sim\mu_1}[Z=\sigma]\right| \\
&= \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left|\mathop{\mathbf{Pr}}_{Z\sim\widehat{\mu}}[Z=\sigma\wedge B] + \mathop{\mathbf{Pr}}_{Z\sim\widehat{\mu}}\left[Z=\sigma \mid \overline{B}\right]\cdot(1-p_{\text{fail}})\right. \\
&\qquad\left. - \mathop{\mathbf{Pr}}_{Z\sim\mu_1}[Z=\sigma\wedge B] - \mathop{\mathbf{Pr}}_{Z\sim\mu_1}\left[Z=\sigma \mid \overline{B}\right]\cdot(1-p_{\text{fail}})\right| \\
&= \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left|\mathop{\mathbf{Pr}}_{Z\sim\widehat{\mu}}[Z=\sigma\wedge B] - \mathop{\mathbf{Pr}}_{Z\sim\mu_1}[Z=\sigma\wedge B]\right| \\
&\le \frac{1}{2}\sum_{\sigma\in\mathcal{C}}\left(\mathop{\mathbf{Pr}}_{Z\sim\widehat{\mu}}[Z=\sigma\wedge B] + \mathop{\mathbf{Pr}}_{Z\sim\mu_1}[Z=\sigma\wedge B]\right) \\
&\le p_{\text{fail}}.
\end{aligned}$$

Combining the above and Lemma 22, we obtain

$$\begin{aligned}
\mathrm{dist}_{TV}(\widehat{\mu},\mu_{\mathcal{C}}) &\le \mathrm{dist}_{TV}(\widehat{\mu},\mu_1) + \mathrm{dist}_{TV}(\mu_1,\mu_2) + \mathrm{dist}_{TV}(\mu_2,\mu_{\mathcal{C}}) \\
&\le p_{\text{fail}} + \frac{\varepsilon}{2} \le \varepsilon.
\end{aligned}$$

It remains to bound the running time of the sampler. The sampler calls subroutines to estimate marginal at most n times and each time the subroutine costs $\mathrm{poly}(n,\frac{1}{\varepsilon})$. Finally, upon the condition on line 14 does not hold, the sampler enumerates proper colourings on connected components of size $O(\log\left(\frac{n}{\varepsilon}\right))$. Therefore, the total running time is $\mathrm{poly}(n,\frac{1}{\varepsilon})$. □

The distribution μ_1 has a small multiplicative error comparing to the uniform distribution $\mu_{\mathcal{C}}$. We remark that there are standard algorithms to turn such a distribution into an exact sampler, dating back to [Bac88, JVV86]. However, since we cannot completely avoid event B, we can only bound the error in the final distribution $\widehat{\mu}$ in terms of total variation distance.

7. Settling all parameters

We have defined the following parameters throughout the paper:

- k_1^C: the number of vertices in a hyperedge that are *not* fixed in approximate counting, Theorem 21;
- k_1^S: the number of vertices in a hyperedge that are *not* fixed in sampling, Theorem 23;
- k_2: the number of vertices in a hyperedge Algorithm 1 would attempt to couple;
- β: the fraction of hyperedges that are monochromatic in Definition 15.

We want our bound for approximate counting to have the form $C\Delta^{\frac{A_1}{k-B_1}}$. By Theorem 21, we want to make sure that, for any $k>0$, subject to $0<k_2<k_1^C<k-1$, and $0<\beta<1$,

$$\frac{A_1}{k-B_1} \geq \frac{3}{\beta(k_2-1)};$$
$$\frac{A_1}{k-B_1} \geq \frac{4-\beta}{(1-\beta)(k_1^C-k_2-1)};$$
$$\frac{A_1}{k-B_1} \geq \frac{1}{k-k_1^C-1}.$$

We assume k_1^C and k_2 are proportional to k. Minimizing A_1 yields the following solutions: $A_1=14, B_1=14, k_1^C=\left\lfloor\frac{13k}{14}\right\rfloor, k_2=\left\lfloor\frac{3k}{7}\right\rfloor, \beta=\frac{1}{2}$. Plugging these values into Theorem 21, we want to satisfy the following constraints:

$$k-k_1^C-2\geq 0, \qquad C\geq\left(5e\left(e^2k^3\right)^{\frac{1}{1-\beta}}\right)^{\frac{1}{k_1^C-k_2-1}},$$
$$q^{k_2-1}>\frac{1}{\beta}, \qquad C\geq\left(\frac{e^{\beta+3}k^3}{\beta^\beta}\cdot\binom{k}{k_2}\right)^{\frac{1}{\beta(k_2-1)}},$$
$$q>(ek\Delta)^{\frac{1}{k_1^C-2}}, \qquad C\geq 4(k-k_1^C)^{\frac{1}{k-k_1^C-1}}.$$

One can verify that $k\geq 28$ and $C\geq 357$ suffice. This yields Theorem 1.

Similarly, we want our bound for sampling to have the form $C\Delta^{\frac{A_2}{k-B_2}}$. By Theorem 23, we want to make sure that, for any $k>0$, subject to $0<k_2<k_1^S<k-1$ and $0<\beta<1$,

$$\frac{A_2}{k-B_2} \geq \frac{3}{\beta(k_2-1)};$$
$$\frac{A_2}{k-B_2} \geq \frac{4-\beta}{(1-\beta)(k_1^S-k_2-1)};$$
$$\frac{A_2}{k-B_2} \geq \frac{3}{k-k_1^S-1}.$$

Similarly to the approximate counting case, minimizing A_2 yields the following solutions: $A_2=16, B_2=\frac{16}{3}, k_1^S=\left\lfloor\frac{13k}{16}\right\rfloor, k_2=\left\lfloor\frac{3k}{8}\right\rfloor, \beta=\frac{1}{2}$. Plugging these values into Theorem 23, we want to satisfy the following constraints:

$$k-k_1^S-2\geq 0, \qquad C\geq\left(5e\left(e^2k^3\right)^{\frac{1}{1-\beta}}\right)^{\frac{1}{k_1^S-k_2-1}},$$
$$q^{k_2-1}>\frac{1}{\beta}, \qquad C\geq\left(\frac{e^{\beta+3}k^3}{\beta^\beta}\cdot\binom{k}{k_2}\right)^{\frac{1}{\beta(k_2-1)}},$$
$$q>(ek\Delta)^{\frac{1}{k_1^S-2}}, \qquad C>\left(e^7k^3\right)^{\frac{1}{(k-k_1^S)-1}}.$$

One can verify that $k\geq 28$ and $C\geq 931$ suffice. This yields Theorem 2. We note that these constraints also hold for $k\geq 6$ and $C\geq 1.2\times 10^{11}$.

8. Concluding remarks

In this paper we give approximate counting and sampling algorithms for hypergraph colourings, when the parameters are in the local lemma regime. One important open question is how to get an optimal constant in the exponent of Δ in Theorem 1 and 2. This constant comes from three places: to bound the number of "bad colourings" (Lemma 16), to bound the error (in the LP) incurred by "good colourings" (Lemma 18), and finally to leave some slack for either counting (Theorem 21) or sampling (Theorem 23). It seems to us that the last slack is difficult to reduce, and a tighter result, if possible, would come from improvements on the first two parts, although our analysis has been pushed to the limit.

Another future direction is to generalize this approach for general constraint satisfaction problems (CSP), or equivalently, the general setup of the (variable version) local lemma. Our analysis relies on some crucial property of hypergraph colourings, that all constraints can be satisfied by partial assignments, ideally with appropriate probabilities. To be more specific, suppose a constraint C contains k variables. We require a property that, when a subset of k' variables are randomly assigned, the probability that C is still not satisfied is roughly $c^{-k'}$ for some constant $c > 1$. This property does not necessarily hold in general, even for symmetric constraints. One such example is when the variables take values from $[q]$, and the constraint is satisfied unless the sum of all its variables is 0 modulo q. We can take q to be large so that the strong local lemma conditions hold, and yet this constraint cannot be satisfied by any subset of variables. In particular, it is problematic to bound our definition of "bad colourings" (Definition 15) when constraints cannot be satisfied by partial assignments. New ideas are required to handle more general settings.

Upon closer look, the success of our approach does not truly rely on that the system is in the local lemma regime. What is essential is that the coupling tree can be truncated at a suitable depth without incurring big error. This turns out to be a special form of the *spatial mixing* property. In the settings of this paper, a strong form of the local lemma condition guarantees that the coupling process succeeds with sufficiently high probability at each step and therefore establishes the desired property. A consequence is that we can use *local* linear constraints to certify the coupling. It remains unclear whether a global correlation decay argument would suffice as well.

Acknowledgements

We thank anonymous referees for many improvements, and in particular, for pointing out the connection to spatial mixing properties.

References

[AI16] Dimitris Achlioptas and Fotis Iliopoulos. Random walks that find perfect objects and the Lovász Local Lemma. *J. ACM*, 63(3):22, 2016. 1

[Alo91] Noga Alon. A parallel algorithmic version of the local lemma. *Random Struct. Algorithms*, 2(4):367–378, 1991. 12

[Bac88] Eric Bach. How to generate factored random numbers. *SIAM J. Comput.*, 17(2):179–193, 1988. 21

[BCKL13] Christian Borgs, Jennifer Chayes, Jeff Kahn, and László Lovász. Left and right convergence of graphs with bounded degree. *Random Struct. Algorithms*, 42(1):1–28, 2013. 12

[BDK06] Magnus Bordewich, Martin E. Dyer, and Marek Karpinski. Stopping times, metrics and approximate counting. In *ICALP*, pages 108–119, 2006. 1, 2

[BDK08] Magnus Bordewich, Martin E. Dyer, and Marek Karpinski. Path coupling using stopping times and counting independent sets and colorings in hypergraphs. *Random Struct. Algorithms*, 32(3):375–399, 2008. 1, 2

[BG08] Antar Bandyopadhyay and David Gamarnik. Counting without sampling: Asymptotics of the log-partition function for certain statistical physics models. *Random Struct. Algorithms*, 33(4):452–479, 2008. 3

[BGG+19] Ivona Bezáková, Andreas Galanis, Leslie Ann Goldberg, Heng Guo, and Daniel Štefankovič. Approximation via correlation decay when strong spatial mixing fails. *SIAM J. Comput.*, 48(2):279–349, 2019. 3

[CDM+19] Sitan Chen, Michelle Delcourt, Ankur Moitra, Guillem Perarnau, and Luke Postle. Improved bounds for randomly sampling colorings via linear programming. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2216–2234. SIAM, 2019. 1

[CPS17] Kai-Min Chung, Seth Pettie, and Hsin-Hao Su. Distributed algorithms for the Lovász local lemma and graph coloring. *Distributed Computing*, 30(4):261–280, 2017. 1

[DFHV13] Martin E. Dyer, Alan M. Frieze, Thomas P. Hayes, and Eric Vigoda. Randomly coloring constant degree graphs. *Random Struct. Algorithms*, 43(2):181–200, 2013. 1

[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

[EL75] Paul Erdős and László Lovász. Problems and results on 3-chromatic hypergraphs and some related questions. *Infinite and finite sets, volume 10 of Colloquia Mathematica Societatis János Bolyai*, pages 609–628, 1975. 1

[Erd63] Paul Erdős. On a combinatorial problem. *Nordisk Mat. Tidskr.*, 11:5–10, 1963. 1

[FA17] Alan M. Frieze and Michael Anastos. Randomly coloring simple hypergraphs with fewer colors. *Inf. Process. Lett.*, 126:39–42, 2017. 2

[FM11] Alan M. Frieze and Páll Melsted. Randomly coloring simple hypergraphs. *Inf. Process. Lett.*, 111(17):848–853, 2011. 1, 2

[FM13] Alan M. Frieze and Dhruv Mubayi. Coloring simple hypergraphs. *J. Comb. Theory, Ser. B*, 103(6):767–794, 2013. 2

[GJ19] Heng Guo and Mark Jerrum. A polynomial-time approximation algorithm for all-terminal network reliability. *SIAM J. Comput.*, 48(3):964–978, 2019. 3

[GJL19] Heng Guo, Mark Jerrum, and Jingcheng Liu. Uniform sampling through the Lovász local lemma. *J. ACM*, 66(3):18:1–18:31, 2019. 3

[GSV15] Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Inapproximability for antiferromagnetic spin systems in the tree nonuniqueness region. *J. ACM*, 62(6):50:1–50:60, 2015. 1

[HLL+17] Kun He, Liang Li, Xingwu Liu, Yuyi Wang, and Mingji Xia. Variable-version Lovász local lemma: Beyond Shearer's bound. In *FOCS*, pages 451–462, 2017. 1

[HS13a] David G. Harris and Aravind Srinivasan. Constraint satisfaction, packet routing, and the Lovász Local Lemma. In *STOC*, pages 685–694, 2013. 1

[HS13b] David G. Harris and Aravind Srinivasan. The Moser-Tardos framework with partial resampling. In *FOCS*, pages 469–478, 2013. 1

[HSS11] Bernhard Haeupler, Barna Saha, and Aravind Srinivasan. New constructive aspects of the Lovász Local Lemma. *J. ACM*, 58(6):28:1–28:28, 2011. 1, 4

[HSZ19] Jonathan Hermon, Allan Sly, and Yumeng Zhang. Rapid mixing of hypergraph independent sets. *Random Struct. Algorithms*, 54(4):730–767, 2019. 2, 3

[HV03] Thomas P. Hayes and Eric Vigoda. A non-Markovian coupling for randomly sampling colorings. In *FOCS*, pages 618–627, 2003. 1

[HV15] Nicholas J. A. Harvey and Jan Vondrák. An algorithmic proof of the Lovász Local Lemma via resampling oracles. In *FOCS*, pages 1327–1346, 2015. 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):157–165, 1995. 1

[JS93] Mark Jerrum and Alistair Sinclair. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.*, 22(5):1087–1116, 1993. 1

[JSV04] Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004. 1

[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. 21

[Kol16] Vladimir Kolmogorov. Commutativity in the algorithmic Lovász local lemma. In *FOCS*, pages 780–787, 2016. 1

[KS11] Kashyap Babu Rao Kolipaka and Mario Szegedy. Moser and Tardos meet Lovász. In *STOC*, pages 235–244, 2011. 1

[LY13] Pinyan Lu and Yitong Yin. Improved FPTAS for multi-spin systems. In *RANDOM*, pages 639–654, 2013. 3

[LYZZ17] Pinyan Lu, Kuan Yang, Chihao Zhang, and Minshen Zhu. An FPTAS for counting proper four-colorings on cubic graphs. In *SODA*, pages 1798–1817, 2017. 3

[Moi19] Ankur Moitra. Approximate counting, the Lovász local lemma, and inference in graphical models. *J. ACM*, 66(2):10:1–10:25, 2019. 2, 5, 6, 14

[MT10] Robin A. Moser and Gábor Tardos. A constructive proof of the general Lovász Local Lemma. *J. ACM*, 57(2), 2010. 1, 17

[Röd85] Vojtěch Rödl. On a packing and covering problem. *Eur. J. Combin.*, 6(1):69 – 78, 1985. 1

[Spe77] Joel Spencer. Asymptotic lower bounds for Ramsey functions. *Discrete Mathematics*, 20:69 – 76, 1977. 4

[Vig00] Eric Vigoda. Improved bounds for sampling colorings. *J. Math. Phys.*, 41(3):1555–1569, 2000. 1

[Wei06] Dror Weitz. Counting independent sets up to the tree threshold. In *STOC*, pages 140–149, 2006. 3

(Heng Guo) School of Informatics, University of Edinburgh, Informatics Forum, Edinburgh, EH8 9AB, United Kingdom.
Email address: hguo@inf.ed.ac.uk

(Chao Liao) Department of Computer Science and Engineering, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai, China.
Email address: chao.liao.95@gmail.com

(Pinyan Lu) ITCS, Shanghai University of Finance and Economics, 100 Wudong Road, Yangpu District, Shanghai, China.
Email address: lu.pinyan@mail.shufe.edu.cn

(Chihao Zhang) John Hopcroft Center for Computer Science, Shanghai Jiao Tong University, 800 Dongchuan Road, Minhang District, Shanghai, China.
Email address: chihao@sjtu.edu.cn