# Approximating the total variation distance between spin systems

Weiming Feng [*]        Hongyang Liu [†]        Minji Yang [*]

## Abstract

 Spin systems form an important class of undirected graphical models. For two Gibbs distributions $\mu$ and $\nu$ induced by two spin systems on the same graph $G = (V, E)$, we study the problem of approximating the total variation distance $d_{\mathrm{TV}}(\mu, \nu)$ with an $\varepsilon$-relative error. We propose a new reduction that connects the problem of approximating the TV-distance to sampling and approximate counting. Our applications include the hardcore model and the antiferromagnetic Ising model in the uniqueness regime, the ferromagnetic Ising model, and the general Ising model satisfying the spectral condition.

 Additionally, we explore the computational complexity of approximating the total variation distance $d_{\mathrm{TV}}(\mu_S, \nu_S)$ between two marginal distributions on an arbitrary subset $S \subseteq V$. We prove that this problem remains hard even when both $\mu$ and $\nu$ admit polynomial-time sampling and approximate counting algorithms.

## 1   Introduction

The total variation distance (TV-distance) is a widely used metric for quantifying the difference between two distributions. For two discrete distributions $\mu$ and $\nu$ defined over the sample space $\Omega$, the TV-distance is given by

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

 Alternatively, the TV-distance can be characterized as the minimum probability of $X \neq Y$, where $X \sim \mu$ and $Y \sim \nu$ form a *coupling* of the two distributions. These different characterizations provide a rich set of tools for analyzing the TV-distance. It is also closely related to other measures of distance between distributions, such as the Wasserstein distance and the KL-divergence [MU17]. Consequently, the TV-distance is a fundamental quantity in many applications, including randomized algorithms, statistical physics, and machine learning.

 The problem of *computing* the TV-distance between two distributions has garnered considerable attention in machine learning and theoretical computer science. A straightforward approach is to compute the TV-distance directly from its definition. This algorithm runs in $O(|\Omega|)$ time, assuming access to the probability mass at every point in $\Omega$. However, the problem becomes particularly interesting when the distributions have succinct representation, as the sample space $\Omega$ can be exponentially large relative to the size of the input. In fact, the problem is intractable for many classes of distributions. For instance, [SV03] showed that when two distributions are specified by circuits that sample from them, deciding whether their TV-distance is small or large is complete for SZK (statistical zero-knowledge). More recently, Bhattacharyya, Gayen,

---

[*]School of Computing and Data Science, The University of Hong Kong, emails: `wfeng@hku.hk` and `ymjessen02@connect.hku.hk`.

[†]State Key Laboratory for Novel Software Technology, New Cornerstone Science Laboratory, Nanjing University, email: `liuhongyang@smail.nju.edu.cn`.

Meel, Myrisiotis, Pavan, and Vinodchandran [BGM+23] proved that even for pairs of product distributions, the exact computation of their TV-distance is **#P**-complete.

On the algorithmic side, the problem of *approximating* the TV-distance between two distributions with $\varepsilon$-relative error was first studied in [BGM+23], where the authors proposed an approximation algorithm for a restricted class of product distributions. Subsequently, Feng, Guo, Jerrum, and Wang [FGJW23] developed a simple FPRAS that works for general product distributions. Later, a deterministic FPTAS was also introduced for the same task [FLL24]. Beyond product distributions, however, the understanding of approximating the TV-distance remains very limited. The only progress was made in [BGM+24], which showed that for two Bayesian networks on DAGs, an FPRAS for the TV-distance exists if *exact* probabilistic inference can be performed in polynomial time. As a consequence, their algorithm can be applied when the underlying DAG has bounded treewidth. However, the requirement of exact inference is a strong assumption. For many natural graphical models without the bounded treewidth property, exact inference is itself a **#P**-complete problem.

In this paper, we further investigate the problem of approximating the TV-distance for spin systems, which are a fundamental class of undirected graphical models. Canonical examples include the hardcore model and the Ising model. Our contributions can be summarized as:

- We give a new algorithm to reduce the TV-distance approximation to sampling and approximate counting. As a result, our algorithm applies to a broad class of hardcore and Ising models, even if the underlying graph has unbounded treewidth.

- We analyze the computational complexity of approximating the TV-distance between two *marginal distributions* of spin systems. We show that this problem is **#P**-hard, even in parameter regimes where both sampling and approximate counting are tractable.

## 1.1 Approximating the TV-distance between two Gibbs distributions

Let $G = (V, E)$ be a graph. A spin system $\mathbb{S}$ (a.k.a. Markov random field) defines a distribution over $\{-1, +1\}^V$ (denoted by $\{\pm\}^V$ in short) in the following way. It defines a weight function $w = w_{\mathbb{S}} := \{\pm\}^V \to \mathbb{R}_{\geq 0}$ that assigns each configuration $\sigma \in \{\pm\}^V$ a weight $w(\sigma)$. The spin system $\mathbb{S}$ induces a Gibbs distribution $\mu = \mu_{\mathbb{S}}$ over $\{\pm\}^V$ such that

$$\forall \sigma \in \{\pm\}^V, \quad \mu(\sigma) \triangleq \frac{w(\sigma)}{Z}, \quad \text{where } Z = Z_{\mathbb{S}} \triangleq \sum_{\tau \in \{\pm\}^V} w(\tau) \text{ is the } \textit{partition function.}$$

The weight function $w$ of a spin system is a product of factors associated with each vertex and edge on graph $G$, which can be computed exactly and efficiently. For a broad class of spin systems, the following sampling and approximate counting oracles with $T_G^{\mathsf{sp}}(\varepsilon), T_G^{\mathsf{ct}}(\varepsilon) = \mathrm{poly}(n/\varepsilon)$ exist, where $n$ denotes the number of vertices in graph $G$.

**Definition 1** (sampling and approximate counting oracles)**.** Let $\mathbb{S}$ be a spin system on graph $G$ with Gibbs distribution $\mu$ and partition function $Z$. Let $T_G^{\mathsf{sp}}, T_G^{\mathsf{ct}} : (0, 1) \to \mathbb{N}$ be two non-increasing cost functions.

- We say $\mathbb{S}$ admits a sampling oracle with cost function $T_G^{\mathsf{sp}}$ if given any $0 < \varepsilon < 1$, it returns a random sample $X \in \{\pm\}^V$ in time $T_G^{\mathsf{sp}}(\varepsilon)$ with $d_{\mathrm{TV}}(X, \mu) \leq \varepsilon$.

- We say $\mathbb{S}$ admits an approx. counting oracle with cost function $T_G^{\mathsf{ct}}$ if given any $0 < \varepsilon < 1$, it returns a random number $\hat{Z}$ in time $T_G^{\mathsf{ct}}(\varepsilon)$ with $\mathbf{Pr}\left[(1 - \varepsilon)Z \leq \hat{Z} \leq (1 + \varepsilon)Z\right] \geq 0.99$.

Given two spin systems $\mathbb{S}^\mu$ and $\mathbb{S}^\nu$ on the same graph $G$ that defines two Gibbs distributions $\mu$ and $\nu$ respectively and an error bound $0 < \varepsilon < 1$, we study the following problem.

*Given the assess to sampling and approximate counting oracles for both $\mathbb{S}^\mu$ and $\mathbb{S}^\nu$, can we efficiently approximate the TV-distance $d_{\mathrm{TV}}(\mu, \nu)$ within relative error $(1 \pm \varepsilon)$?*

In this paper, we focuses on the following two canonical and extensively studied spin systems.

- **Hardcore model**: Let $G = (V, E)$ be a graph. Let $\lambda = (\lambda_v)_{v \in V} \in \mathbb{R}^V_{\geq 0}$ be external fields. A configuration $\sigma \in \{\pm\}^V$ is said to be an independent set if $S_\sigma = \{v \in V \mid \sigma_v = +1\}$ forms an independent set in graph $G$. The hardcore model is specified by the pair $(G, \lambda)$, which defines the weight function $w$ such that

$$\forall \sigma \in \{\pm\}^V, \quad w(\sigma) \triangleq \begin{cases} \prod_{v \in V : \sigma_v = +1} \lambda_v & \text{if } \sigma \text{ is an independent set;} \\ 0 & \text{otherwise.} \end{cases}$$

- **Ising model**: Let $G = (V, E)$ be a graph. Let $J \in \mathbb{R}^{V \times V}$ be a *symmetric interaction matrix* such that $J_{uv} \neq 0$ only if $\{u, v\} \in E$. Let $h \in (\mathbb{R} \cup \{\pm\infty\})^V$ be the *external field*. Define *Hamiltonian* function

$$\forall \sigma \in \{\pm\}, \quad H(\sigma) \triangleq \sum_{\{u,v\} \in E} J_{uv} \sigma_u \sigma_v + \sum_{v \in V} \sigma_v h_v = \frac{1}{2} \langle \sigma, J\sigma \rangle + \langle \sigma, h \rangle.$$

The weight of a configuration $\sigma$ in Ising model is defined by $w(\sigma) \triangleq \exp(H(\sigma))$.

The problem of approximating the TV-distance can be formalized as follows.

**Problem 2** (Hardcore TV-distance approximation)**.** The problem is defined as follows.

- *Input*: two hardcore models $(G, \lambda^\mu)$ and $(G, \lambda^\nu)$ defined on the same graph $G = (V, E)$, which specifies two Gibbs distributions $\mu$ and $\nu$ respectively, and an error bound $0 < \varepsilon < 1$.

- *Output*: a number $\hat{d}$ such that $(1 - \varepsilon) d_{\mathrm{TV}}(\mu, \nu) \leq \hat{d} \leq (1 + \varepsilon) d_{\mathrm{TV}}(\mu, \nu)$.

**Problem 3** (Ising TV-distance approximation)**.** The problem is defined as follows.

- *Input*: two Ising models $(G, J^\mu, h^\mu)$ and $(G, J^\nu, h^\nu)$ defined on the same graph $G = (V, E)$, which specifies two Gibbs distributions $\mu$ and $\nu$ respectively, and an error bound $0 < \varepsilon < 1$.

- *Output*: a number $\hat{d}$ such that $(1 - \varepsilon) d_{\mathrm{TV}}(\mu, \nu) \leq \hat{d} \leq (1 + \varepsilon) d_{\mathrm{TV}}(\mu, \nu)$.

We are interested in the FPRAS (fully polynomial randomized approximation scheme), which solves the above problems with probability at least $2/3$[1] in time $\mathrm{poly}(n/\varepsilon)$.

### 1.1.1 General results for hardcore and Ising models

We need a marginal lower bound condition for general results. For any subset $\Lambda \subseteq V$, any feasible partial configuration $\sigma \in \{\pm\}^\Lambda$, let $\mu^\sigma$ denote the distribution $\mu$ conditional on $\sigma$. For any $v \in V$, let $\mu_v^\sigma$ denote the marginal distribution on $v$ projected from $\mu^\sigma$.

**Condition 4** (marginal lower bound)**.** *Let $0 < b < 1$ be a parameter. We say a distribution $\mu$ over $\{\pm\}^V$ is b-marginally bounded if for any feasible partial configuration $\sigma \in \{\pm\}^\Lambda$ on a subset $\Lambda \subseteq V$, any $v \in V$, any $c \in \{\pm\}$, it holds that $\mu_v^\sigma(c) \geq b$ if $\mu_v^\sigma(c) > 0$.*

---

[1]The success probability can be boosted to $1 - \delta$ by independently running the algorithm for $O(\log \frac{1}{\delta})$ times and taking the median of the output.

The marginal lower bound condition is a natural condition for spin systems. Hardcore and Ising models with constant marginal lower bound were extensively studied in sampling and approximate counting [Jer03, Wei06, Sly10, CLV21].

**Theorem 5** (general result)**.** *Let $0 < b < 1$ be a constant. There exists a randomized algorithm that solves Problem 2 and Problem 3 with probability at least $\frac{2}{3}$ in time*

$$O_b\left(\frac{N^2}{\varepsilon^2}T_G^{sp}\left(\mathrm{poly}(b)\cdot\frac{\varepsilon^2}{N^2}\right) + T_G^{ct}\left(\mathrm{poly}(b)\cdot\frac{\varepsilon}{N}\right)\right),$$

*if two input spin systems are both b-marginally bounded and both admit sampling and approximate counting oracles with cost functions $T_G^{sp}(\cdot), T_G^{ct}(\cdot)$, where $O_b(\cdot)$ hides a factor depending only on b, $N = n = |V|$ for hardcore model, and $N = n + m = |V| + |E|$ for Ising model.*

For the hardcore and Ising models with a constant marginal lower bound, the theorem provides the first polynomial-time reduction from approximating the TV-distance to sampling and approximate counting. Moreover, the above theorem is a simplified version, and our technique yields a stronger result that also applies to some hardcore and Ising models with a smaller marginal lower bound $b = o(1)$. See Remark 38 or a more detailed discussion.

Consider a hardcore model $(G, \lambda)$. Let $\Delta \geq 3$ denote the maximum degree of $G$. The hardcore model is said to satisfy the *uniqueness condition* with a constant gap $0 < \eta < 1$ if

$$\forall v \in V, \quad \lambda_v \leq (1 - \eta)\lambda_c(\Delta), \quad \text{where } \lambda_c(\Delta) \triangleq \frac{(\Delta - 1)^{\Delta - 1}}{(\Delta - 2)^{\Delta}} \approx \frac{e}{\Delta}. \tag{1}$$

Previous works [CFYZ22, CE22, ŠVV09] proved that for hardcore model satisfying the uniqueness condition in (1), it admits $\mathrm{poly}(\frac{n}{\varepsilon})$ time sampling and approximate counting oracles. To obtain a $\mathrm{poly}(\frac{n}{\varepsilon})$ time algorithm for TV-distance estimation, one way is to consider hardcore models in the uniqueness regime with constant marginal lower bound $b = \Omega(1)$, which requires that $\Delta = O(1)$ and for all $v \in V$, either $\Omega(1) \leq \lambda_v \leq (1 - \eta)\lambda_c(\Delta)$ or $\lambda_v = 0$[2].

For Ising models $\mathbb{S} = (G, J, h)$, the following conditions are well-known.

**Condition 6.** *$\mathbb{S}$ satisfies one of the following conditions:*

- *Spectral condition: $\lambda_{\max}(J) - \lambda_{\min}(J) \leq 1 - \eta$ for some constant $\eta > 0$, where $\lambda_{\max}(J)$ and $\lambda_{\min}(J)$ denote the max and min eigenvalues of $J$ respectively.*

- *Ferromagnetic interaction with consistent field condition: $J_{uv} \geq 0$ for all edge $\{u, v\} \in E$ and $h_v \geq 0$ for all $v \in V$.*

- *Anti-ferromagnetic interaction at or within the uniqueness threshold: $J_{uv} = \beta \leq 0$ for all $\{u, v\} \in E$ and $\exp(2\beta) \geq \frac{\Delta - 2}{\Delta}$, where $\Delta$ is the maximum degree of $G$.*

Previous works [AJK$^+$22, FGW23, JS93, CCYZ24, JVV86, ŠVV09] gave $\mathrm{poly}(\frac{n}{\varepsilon})$ time sampling and approximate counting oracles for Ising models satisfying Condition 6.

We have the following corollary for hardcore and Ising models from Theorem 5.

**Corollary 7.** *Let $0 < \eta < 1$ and $\Delta \geq 3$ be two constants. There exists an FPRAS for Problem 2 if two input hardcore models are defined on graph $G$ with maximum degree $\Delta$ and all external fields satisfy that $\Omega(1) \leq \lambda_v^{\pi} \leq (1 - \eta)\lambda_c(\Delta)$ or $\lambda_v^{\pi} = 0$ for all $v \in V$ and $\pi \in \{\mu, \nu\}$.*

*There exists an FPRAS for Problem 3 if two input Ising models both satisfy Condition 6 and the marginal lower bound in Condition 4 with $b = \Omega(1)$.*

---

[2]We allow $\lambda_v = 0$ because Condition 4 only considers the lower bound for spins $c \in \{\pm\}$ with $\mu_v^{\sigma}(c) > 0$.

The condition required by Corollary 7 is arguably significant conditions for TV-distance estimation. Let $\Delta \geq 3$ be a constant. For the hardcore model beyond the uniqueness regime such that $\lambda_v > \lambda_c(\Delta)$ for all $v \in V$, polynomial-time sampling and approximate counting oracles do not exist unless $\mathbf{NP} = \mathbf{RP}$ [Sly10]. Moreover, the technique in [BGM⁺25] can show that unless $\mathbf{NP} = \mathbf{RP}$, there is no FPRAS for Problem 2 if input hardcore models are beyond the uniqueness condition. For the completeness, we give a simplified proof in Appendix A. We can show that this hardness result holds even for *additive-error* approximation.

To see the significance of Condition 6, consider the following family of Ising models. Let $\mathbb{S}^\mu = (G, J, h^\mu)$ and $\mathbb{S}^\nu = (G, J, h^\nu)$ be two Ising models defined on the same graph $G$ and have the same interaction matrix $J$. Assume that $G$ has constant maximum degree $\Delta$; for all edges $\{u, v\}$ in $G$, $J_{uv} = \beta < 0$ is a unified negative constant; and for all vertices $v \in V$, $h_v^\mu$ and $h_v^\nu$ can take values from set $\{\pm\infty, 0\}$. This family of Ising models has a constant marginal lower bound. We have the following two results:

- if $\exp(2\beta) \geq \frac{\Delta-2}{\Delta}$, then FPRAS for TV-distance exists by Corollary 7;

- if $\exp(2\beta) < \frac{\Delta-2}{\Delta}$, there is no FPRAS for TV-distance unless $\mathbf{NP} = \mathbf{RP}$, the hardness result holds even for approximating the TV-distance with *additive* error.

Again, the hardness result can be proved by the technique in [BGM⁺25]. We give a simple proof in Appendix A for the completeness.

### 1.1.2 Improved results for hardcore models in the uniqueness regime

Corollary 7 works for hardcore model in the uniqueness regime. However, it additionally requires a marginal lower bound $b = \Omega(1)$. We give the following improved algorithm that removes the marginal lower bound requirement, thus it works for the *whole* uniqueness regime.

**Theorem 8** (improved algorithm for the whole uniqueness regime)**.** *There exists a randomized algorithm that solves Problem 2 with probability at least $2/3$ in time $\tilde{O}_\eta(\frac{\Delta n^7}{\varepsilon^{5/2}})$ if two input hardcore models both satisfy the uniqueness condition with a constant gap $\eta > 0$, where $n$ is the number of vertices and $\tilde{O}_\eta(\cdot)$ hides a constant factor depending on $\eta$ and a $\mathrm{polylog}(\frac{n}{\varepsilon})$ factor.*

Next, let us further assume that the hardcore model has a *constant* marginal lower bound $b = \Omega(1)$. Theorem 5 and Corollary 7 give an FPRAS in time

$$O_b\left(\frac{N^2}{\varepsilon^2}T_G^{\mathsf{sp}}\left(\mathrm{poly}(b) \cdot \frac{\varepsilon^2}{N^2}\right) + T_G^{\mathsf{ct}}\left(\mathrm{poly}(b) \cdot \frac{\varepsilon}{N}\right)\right) = \tilde{O}\left(\frac{n^4}{\varepsilon^2}\right). \tag{2}$$

The equation holds because $T_G^{\mathsf{sp}}(\delta) = O(\Delta n \log \frac{n}{\delta})$ [CFYZ22, CE22], $T_G^{\mathsf{ct}}(\delta) = \tilde{O}(\frac{\Delta n^2}{\delta^2})$ [ŠVV09], and $\Delta = O(1)$ due to the constant marginal lower bound assumption. For these hardcore models, we can also give a faster algorithm than the general results in Theorem 5. Compared to the running time in (2), the following improved algorithm reduces a factor of $n$ in the running time.

**Theorem 9** (faster algorithm further assuming constant marginal lower bound)**.** *Let $0 < \eta < 1$ and $\Delta \geq 3$ be two constants. There exists an FPRAS in time $\tilde{O}(\frac{n^3}{\varepsilon^2})$ for Problem 2 if two input hardcore models are defined on graph $G$ with maximum degree $\Delta$ and all external fields satisfy that $\Omega(1) \leq \lambda_v^\pi \leq (1 - \eta)\lambda_c(\Delta)$ or $\lambda_v^\pi = 0$ for all $v \in V$ and $\pi \in \{\mu, \nu\}$.*

## 1.2 Approximating the TV-distance between two marginal distributions

One natural extension is to approximate the TV-distance between two marginal distributions on a subset of vertices. We use the hardcore model as an example to state our results on this problem. The same results can be extended to Ising model using a similar reduction.

**Problem 10** (*k*-marginal TV-distance approximation)**.** Let $k : \mathbb{N} \to \mathbb{N}$ be a function.

- *Input*: two hardcore models $(G, \lambda^\mu)$ and $(G, \lambda^\nu)$ defined on the same graph $G = (V, E)$, which specifies two Gibbs distributions $\mu$ and $\nu$ respectively, a subset $S \subseteq V$ such that $|S| = k(n)$, where $n = |V|$, and an error bound $\varepsilon > 0$.

- *Output*: a number $\hat{d}$ such that $\frac{d_{\mathrm{TV}}(\mu_S, \nu_S)}{1+\varepsilon} \leq \hat{d} \leq (1 + \varepsilon) d_{\mathrm{TV}}(\mu_S, \nu_S)$, where $\mu_S$ and $\nu_S$ are marginal distributions on $S$ projected from $\mu$ and $\nu$ respectively.

In particular, if the function $k(n) = n$, then Problem 10 is the same as Problem 2.

We show that the problem is hard when $k(n) = 1$. In this case, the problem is to approximate the TV-distance between two marginal distributions at a single vertex. The hardness result holds even if two input hardcore models are in the uniqueness regime, where both sampling and approximate counting are intractable in polynomial time.

**Theorem 11.** *Let $k(n) = 1$ for all $n \in \mathbb{N}$ be a constant function. The k-marginal TV-distance approximation is **#P**-Hard when two input hardcore models both satisfy the uniqueness condition, the hardness result holds even if $\varepsilon = \mathrm{poly}(n)$, where $n$ is number of vertices in $G$.*

The above theorem is for marginal distributions at one vertex. One can simply lift the result to the marginal distributions on a set of vertices. In particular, we have the following corollary.

**Corollary 12.** *Let $0 < \alpha < 1$ be a constant and $k(n) = n - \lceil n^\alpha \rceil$. The k-marginal TV-distance approximation is **#P**-Hard when two input hardcore models both satisfy the uniqueness condition, the hardness result holds even if $\varepsilon = \mathrm{poly}(n)$, where $n$ is number of vertices in $G$.*

The proofs of the hardness results are given in Section 7. The proof constructs a Turing reduction that exactly counts the number $Z$ of independent sets in graphs with a maximum degree of 3, a problem known to be **#P**-complete [DG00]. Specifically, we show that if one can efficiently solve the problem stated in Theorem 11, then it is possible to efficiently estimate the probability that a vertex $v$ is included in a uniformly random independent set, with an *exponentially* small relative error of $4^{-n}$. This, in turn, solves the *exact* counting problem by using the self-reducibility [JVV86] property of the hardcore model and the fact that $Z \leq 2^n$.

Now, let us compare our hardness results with that in [BGM+25]. The hardness results in [BGM+25] are for approximating the TV-distance between two *entire Gibbs distributions*. In contrast, our hardness results are for approximating the TV-distance between two *marginal distributions*. The hardness results in [BGM+25] are **NP**-hard results. It considered the Gibbs distributions in a parameter regime where sampling and approximate counting are intractable. In contrast, our hardness results are **#P**-hard results. Our hardness results holds even if sampling and approximate counting can both be solved in polynomial time.

Finally, consider approximating the TV-distance between two *marginal distributions* with *additive error* $\varepsilon$. We show that this relaxed problem admits FPRAS if two input hardcore models both satisfy the uniqueness condition.

**Theorem 13.** *There exists a randomized algorithm such that given two hardcore distributions $\mu$ and $\nu$ on the same graph $G = (V, E)$ with $n = |V|$, any subset $S \subseteq V$, and any $0 < \varepsilon < 1$, if $\mu$ and $\nu$ both satisfy (1), it returns a random number $\hat{d}$ in time $\frac{\Delta n^2}{\varepsilon^4} \cdot \mathrm{polylog}(\frac{n}{\varepsilon})$ such that*

$$\mathbf{Pr}\left[\left|\hat{d} - d_{\mathrm{TV}}(\mu_S, \nu_S)\right| \leq \varepsilon\right] \geq \frac{2}{3}.$$

Theorem 13 together with two hardness results give a clear separation between the computational complexity for approximating the TV-distance with relative and additive error.

## 1.3 Related work

There are a series of works on checking whether two given distributions are identical (e.g. [CMR07, DHR08, KMO+11, BGM+25]), which can be viewed as the decision version of computing the TV-distance, i.e., checking whether it is zero.

There are also a long line of works (e.g., see a survey in [Can15]) studying the identical testing problem in access model, where the algorithm can only access the set of random samples from the distributions. This setting is different from our setting, where we assume that all the parameters of the spin systems are given as the input to the algorithm.

A series of works [BGMV20, CK14, Kie18, CR14] studied the algorithm and the hardness of approximating the TV-distance with additive error, which is an easier problem than the relative-error approximation.

[DMR18, AAL23, Kon24, AK24] found *closed-form formulas*, which approximates the TV-distance between two high-dimensional distributions with a *fixed* relative error. We study *algorithms* achieving an *arbitrary* $\varepsilon$-relative error approximation for spin systems.

## 2 Algorithm overview

In this section, we give an overview of our algorithm. Let $G = (V, E)$ be a graph. Let $\mathbb{S}^\mu$ and $\mathbb{S}^\nu$ be two spin systems (either two hardcore models or two Ising models) defined on the same graph $G$. Let $\mu$ and $\nu$ denote Gibbs distributions of $\mathbb{S}^\mu$ and $\mathbb{S}^\nu$ respectively.

We first introduce the distance between parameters of two spin systems. For a vector $a \in \mathbb{R}^V$, denote $\|a\|_\infty = \max_{v \in V} |a_v|$. For a matrix $A \in \mathbb{R}^{V \times V}$, denote $\|A\|_{\max} = \max_{u,v \in V} |A_{uv}|$.

**Definition 14** (parameter distance). The parameter distance $d_{\mathrm{par}}(\mathbb{S}^\mu, \mathbb{S}^\nu)$ between $\mathbb{S}^\mu$ and $\mathbb{S}^\mu$, which is denote by $d_{\mathrm{par}}(\mu, \nu)$ for simplicity, is defined by

- Hardcore model: for two hardcore models $\mathbb{S}^\mu = (G, \lambda^\mu)$ and $\mathbb{S}^\nu = (G, \lambda^\nu)$,

$$d_{\mathrm{par}}(\mu, \nu) \triangleq \|\lambda^\mu - \lambda^\nu\|_\infty.$$

- Soft-Ising model: for two soft-Ising models $\mathbb{S}^\mu = (G, J^\mu, h^\mu)$ and $\mathbb{S}^\nu = (G, J^\nu, h^\nu)$,

$$d_{\mathrm{par}}(\mu, \nu) \triangleq \max\left\{\|J^\mu - J^\nu\|_{\max}, \max_v \frac{|h^\mu(v) - h^\nu(v)|}{\deg_v + 1}\right\},$$

where $\deg_v$ is the degree of $v$ in graph $G$.

The above parameter distance can be computed easily. The following lemma gives the relation between parameter distance $d_{\mathrm{par}}(\mu, \nu)$ and the total variation distance $d_{\mathrm{TV}}(\mu, \nu)$. An Ising model $(G = (V, E), J, h)$ is said to be *soft* if $h \in \mathbb{R}^V$ (instead of $h \in (\mathbb{R} \cup \{\pm\infty\})^V$). We now focus on soft-Ising model in this overview. We will show how to reduce general Ising models to soft models in Section 6.

**Lemma 15** (TV-distance lower bound). *It holds that* $d_{\mathrm{TV}}(\mu, \nu) \geq C_{\mathrm{par}}^{TV} \cdot d_{\mathrm{par}}(\mu, \nu)$ *such that*

- *Hardcore model: if both $\mu$ and $\nu$ satisfy the uniqueness condition in* (1)*,* $C_{\mathrm{par}}^{TV} = \frac{1}{5000}$.

- *Hardcore model: if both $\mu$ and $\nu$ are b-marginally bounded,* $C_{\mathrm{par}}^{TV} = b^3$.

- *Soft-Ising model: if both $\mu$ and $\nu$ are b-marginally bounded,* $C_{\mathrm{par}}^{TV} = \frac{b^2}{2}$.

By Lemma 15, if $d_{\mathrm{par}}(\mu, \nu)$ is large, the TV-distance $d_{\mathrm{TV}}(\mu, \nu)$ is also large. Let $n = |V|$ and $m = |E|$ denote the numbers of vertices and edges in $G$. Define the a threshold $\theta = \frac{c_b}{\mathrm{poly}(n)}$ for the parameter distance for soft-Ising and hardcore models, where $c_b$ is some parameter depending only on the marginal lower bound $b$[3]. The specific value of $\theta$ can be found in Section 6. Our algorithm first compute $d_{\mathrm{par}}(\mu, \nu)$ in time $O(m + n)$, and then compares $d_{\mathrm{par}}(\mu, \nu)$ to the threshold $\theta$. The algorithm considers the following two cases.

- Case $d_{\mathrm{par}}(\mu, \nu) \geq \theta$: By Lemma 15, the TV-distance $d_{\mathrm{TV}}(\mu, \nu) = C_{\mathrm{par}}^{\mathrm{TV}} d_{\mathrm{par}}(\mu, \nu) \geq \Omega_b(\frac{1}{\mathrm{poly}(n)})$ is large. In this case, the task of approximating $d_{\mathrm{TV}}(\mu, \nu)$ with relative error is the same (up to $O_b(\mathrm{poly}(n))$ running time) as the task of approximating $d_{\mathrm{TV}}(\mu, \nu)$ with *additive* error. We give a general FPRAS that achieves the additive-error approximation assuming polynomial-time sampling and approximate counting oracles for both $\mu$ and $\nu$.

- Case $d_{\mathrm{par}}(\mu, \nu) < \theta$: The algorithm for this case is our main technical contribution. Let $w_\mu(\cdot)$ and $w_\nu(\cdot)$ denote the weight functions for spin systems $\mathbb{S}^\mu$ and $\mathbb{S}^\nu$ respectively. By the definition of parameter distance, we know that for any $\sigma \in \{\pm\}^V$, $w_\mu(\sigma) \approx w_\nu(\sigma)$. We will utilize this property to design efficient approximation algorithm for TV-distance.

In Section 2.1 and Section 2.2, we will explain the main ideas of our algorithms for the above two cases. The formal description of the algorithms are given in Sections 4 and 5.

## 2.1 Warm-up: additive-error approximation algorithm

Let us first consider the easy case. This case is solved by improving the algorithm in [BGMV20]. By the definition of total variation distance, we can write

$$d_{\mathrm{TV}}(\mu, \nu) = \sum_{\sigma \in \{\pm\}^V : \mu(\sigma) > \nu(\sigma)} |\mu(\sigma) - \nu(\sigma)| = \sum_{\sigma \in \{\pm\}^V} \mu(\sigma) \max\left(0, 1 - \frac{\nu(\sigma)}{\mu(\sigma)}\right). \qquad (3)$$

Define the random variable $X \triangleq \max\left(0, 1 - \frac{\nu(\sigma)}{\mu(\sigma)}\right)$, where $\sigma \sim \mu$. Note that $0 \leq X \leq 1$. We have $\mathbf{E}[X] = d_{\mathrm{TV}}(\mu, \nu)$ and $\mathbf{Var}[X] \leq 1$. Additive-error approximation can be achieved if random samples of the variable $X$ can be efficiently generated.

However, since we can only access sampling and approximate counting oracles, we cannot compute $\mu(\sigma)$ and $\nu(\sigma)$ exactly. Instead, our algorithm uses an alternative estimator $\hat{X}$ to approximate the random variable $X$. Let $Z_\mu, w_\mu(\cdot)$ and $Z_\nu, w_\nu(\cdot)$ denote the partition functions and weight functions of $\mathbb{S}^\mu$ and $\mathbb{S}^\nu$, respectively. We first call the approximate counting oracle to obtain $\hat{Z}_\mu$ and $\hat{Z}_\nu$, which approximate $Z_\mu$ and $Z_\nu$ with a relative error of $O(\varepsilon)$. The estimator $\hat{X}$ is then defined by the following process:

- Call sampling oracle to generate approximate sample $\sigma$ from $\mu$;

- $\hat{X} = \max\left(0, 1 - \frac{\hat{\nu}(\sigma)}{\hat{\mu}(\sigma)}\right)$, where $\hat{\nu}(\sigma) = w_\nu(\sigma)/\hat{Z}_\nu$ and $\hat{\mu}(\sigma) = w_\mu(\sigma)/\hat{Z}_\mu$.

The error of $\hat{X}$ arises from the errors in the approximate sample $\sigma$ and the probabilities $\hat{\nu}(\cdot)$ and $\hat{\mu}(\cdot)$. To illustrate the main idea of the algorithm, let us ignore the sampling error and assume $\sigma \sim \mu$. Note that the true value $\frac{\nu(\sigma)}{\mu(\sigma)} = \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \cdot \frac{Z_\mu}{Z_\nu}$. By the assumption of approximate counting oracle, $\frac{\hat{\nu}(\sigma)}{\hat{\mu}(\sigma)} \geq (1 - O(\varepsilon))\frac{\nu(\sigma)}{\mu(\sigma)}$, which implies $\mathbf{E}[\hat{X}] \leq \mathbf{E}_{\sigma \sim \mu}[\max(0, 1 - (1 - O(\varepsilon))\frac{\nu(\sigma)}{\mu(\sigma)})]$, which is at most $\mathbf{E}_{\sigma \sim \mu}[\max(0, 1 - \frac{\nu(\sigma)}{\mu(\sigma)})] + O(\varepsilon)\mathbf{E}_{\sigma \sim \mu}\frac{\nu(\sigma)}{\mu(\sigma)} = \mathbf{E}[X] + O(\varepsilon)$. A similar analysis gives the

---

[3]The value of $b$ is not given in the input, but we can compute $b$ efficiently (see Lemma 35).

lower bound $\mathbf{E}[\hat{X}] \geq \mathbf{E}[X] - O(\varepsilon)$. Since $0 \leq \hat{X} \leq 1$ so the variance is at most 1, this achieves the additive error approximation of the TV-distance.

The algorithm in Theorem 13 estimates the TV-distance between two marginal distributions with additive error, which follows a similar approach. The key difference is that it requires approximating the marginal probabilities $\nu_S(\sigma)$ and $\mu_S(\sigma)$ for $\sigma \in \{\pm\}^S$, where $S \subseteq V$. For the hardcore model within the uniqueness regime, we have not only an efficient approximate counting oracle but also efficient oracles for approximating the conditional partition function $Z^\sigma = \sum_{\tau \in \{\pm\}^V : \tau_S = \sigma} w(\tau)$, which give the efficient approximation of $\mu_S(\sigma)$ and $\nu_S(\sigma)$.

Additive-error approximation algorithms were also studied in [BGMV20], where a similar algorithm was proposed. Their approach assumes that the values of $\mu(\sigma)$ and $\nu(\sigma)$ can be computed exactly, allowing the use of the estimator $X$. In contrast, our algorithm demonstrates that approximated values of $\mu(\sigma)$ and $\nu(\sigma)$ suffice. Consequently, our additive-error algorithm (see Theorem 17 for a formal statement) improves upon some results in [BGMV20]. For instance, while [BGMV20] applies to ferromagnetic Ising models on bounded tree-width graphs with consistent external fields, our result removes the bounded tree-width restriction and applies to general graphs (see Corollary 18).

## 2.2 Approximation algorithm for instances with small parameter distance

Consider the case where $d_{\mathrm{par}}(\mu, \nu) < \theta$. The total variation distance $d_{\mathrm{TV}}(\mu, \nu)$ can be very small (e.g., $\exp(-\Omega(n))$). We cannot use additive-error approximation algorithm to efficiently achieve a relative-error approximation. The main challenge lies in the term $1 - \nu(\sigma)/\mu(\sigma)$ in (3). With approximate counting oracles, we can approximate $\nu(\sigma)/\mu(\sigma)$ with relative error, but there is no guarantee of relative accuracy for $1 - \nu(\sigma)/\mu(\sigma)$. Due to this difficulty, previous works mainly studied graphical models on bounded treewidth graphs [BGM+24] and product distributions [FGJW23], where $\nu(\sigma)$ and $\mu(\sigma)$ can be computed *exactly*.

We overcome this challenge by designing an alternative estimator with good concentration property. We use the hardcore model as an example to illustrate the main idea. Assume both $\mu$ and $\nu$ are hardcore models $(G, \lambda^\mu)$ and $(G, \lambda^\nu)$ satisfying uniqueness condition in (1). For the simplicity of the overview, let us further assume that for any for $v \in V$, $0 < \lambda_v^\mu \leq \lambda_v^\nu$. In words, $\nu$ is obtained by increasing some external fields of $\mu$ by a small amount. We will deal with the general case in later technical sections.

**Basic case without small external fields** Let us start with a simple case where $\lambda_v^\mu = \Theta(\frac{1}{\Delta})$ for all $v \in V$. The uniqueness condition guarantees that $\lambda_v^\mu = O(\frac{1}{\Delta})$. The assumption additionally requires that every $\lambda_v^\mu$ cannot be too small. By the definitions of total variation distance and Gibbs distributions, we can compute

$$d_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2} \sum_{\sigma \in \{\pm\}^V} \mu(\sigma) \left| 1 - \frac{\nu(\sigma)}{\mu(\sigma)} \right| = \frac{Z_\mu}{2Z_\nu} \cdot \sum_{\sigma \in \{\pm\}^V} \mu(\sigma) \left| \frac{Z_\nu}{Z_\mu} - \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \right|. \tag{4}$$

The first term $\frac{Z_\mu}{2Z_\nu}$ can be approximated with relative error by approximate counting oracle. For the second term, we consider the following random variable, which also appears in the previous work of approximate counting [ŠVV09],

$$W \triangleq \frac{w_\nu(\sigma)}{w_\mu(\sigma)}, \text{ where } \sigma \sim \mu.$$

Note that $\mathbf{E}[W] = \frac{Z_\nu}{Z_\mu}$. By (4), our task is reduced to estimate the value of $\mathbf{E}[|\mathbf{E}[W] - W|]$. We are in the case that $0 \leq \lambda_v^\nu - \lambda_v^\mu \leq d_{\mathrm{par}}(\mu, \nu) < \theta = \frac{1}{\mathrm{poly}(n)}$, so that for any $\sigma \in \{\pm\}^V$, $\frac{w_\mu(\sigma)}{w_\nu(\sigma)} \approx 1$.

We will utilize this property to show that $W$ has a good concentration property. Formally,

$$1 \le \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \le \prod_{v \in V: \sigma_v = +1} \frac{\lambda_v^\nu}{\lambda_v^\mu} \le \prod_{v \in V: \sigma_v = +1} \left( 1 + \frac{d_{\mathrm{par}}(\mu, \nu)}{\lambda_v^\mu} \right)$$

$$(\star) \quad \le 1 + O\left( \sum_{v \in V: \sigma_v = +1} \frac{d_{\mathrm{par}}(\mu, \nu)}{\lambda_v^\mu} \right) \le 1 + O(n\Delta) \cdot d_{\mathrm{par}}(\mu, \nu). \tag{5}$$

In inequality $(\star)$, we use the fact that $\lambda_v^\mu = \Omega(\frac{1}{\Delta})$ so that $\frac{d_{\mathrm{par}}(\mu, \nu)}{\lambda_v^\mu} = O(\Delta\theta)$, since we choose $\theta = \frac{1}{\mathrm{poly}(n)}$ small enough, then $\frac{d_{\mathrm{par}}(\mu, \nu)}{\lambda_v^\mu} \ll \frac{1}{n}$ and the inequality can be verified as follows

$$\prod_{v \in V: \sigma_v = +1} \left( 1 + \frac{d_{\mathrm{par}}(\mu, \nu)}{\lambda_v^\mu} \right) \le \exp\left( \underbrace{\sum_{v \in V: \sigma_v = +1} \frac{d_{\mathrm{par}}(\mu, \nu)}{\lambda_v^\mu}}_{\ll n \cdot (1/n) = o(1)} \right) = 1 + O\left( \sum_{v \in V: \sigma_v = +1} \frac{d_{\mathrm{par}}(\mu, \nu)}{\lambda_v^\mu} \right).$$

By Lemma 15, $d_{\mathrm{TV}}(\mu, \nu) = \Omega(d_{\mathrm{par}}(\mu, \nu))$. This implies $W$ enjoys a very good concentration property such that

$$1 \le W \le 1 + O(n\Delta) \cdot d_{\mathrm{TV}}(\mu, \nu) \quad \implies \quad \mathbf{Var}\,[W] \le O(n^2\Delta^2) \cdot (d_{\mathrm{TV}}(\mu, \nu))^2. \tag{6}$$

Now, our algorithm can be outline as follows.

- Call approximate counting oracles to obtain $\hat{Z}_\mu$ and $\hat{Z}_\nu$ with relative error $O(\varepsilon)$.

- Draw $T = \mathrm{poly}(n/\varepsilon)$ samples $W_1, \ldots, W_T$ from $W$ independently.

- Compute $\bar{W} = \frac{1}{T} \sum_{i=1}^{T} W_i$.  (approximate $\mathbf{E}\,[W]$)

- Compute $\bar{E} = \frac{1}{T} \sum_{i=1}^{T} |W_i - \bar{W}|$.  (approximate $\mathbf{E}\,[|W - \mathbf{E}\,[W]|]$)

- Return $\hat{d} = \frac{\hat{Z}_\mu}{2\hat{Z}_\nu} \bar{E}$.

Using the variance bound in (6), we can prove that with high probability, $\bar{E}$ approximates $\mathbf{E}\,[|W - \mathbf{E}\,[W]|]$ with an additive error of $O(\varepsilon) \cdot d_{\mathrm{TV}}(\mu, \nu)$. By (5), we can also verify that $\frac{Z_\mu}{Z_\nu} = O(1)$. Hence, we can bound the error as follows

$$\hat{d} \le (1 + O(\varepsilon)) \frac{Z_\nu}{2Z_\mu} \cdot (\mathbf{E}\,[|W - \mathbf{E}\,[W]|] + O(\varepsilon) \cdot d_{\mathrm{TV}}(\mu, \nu)) \le (1 + O(\varepsilon)) d_{\mathrm{TV}}(\mu, \nu).$$

A similar analysis gives the lower bound $\hat{d} \ge (1 - O(\varepsilon)) d_{\mathrm{TV}}(\mu, \nu)$. This achieves the relative-error approximation of the TV-distance. The above technique can be generalized to Ising models and hardcore models with a marginal lower bound.

**General case containing small external fields**  For the general case, there may exist vertex $v \in V$ such that the external field $\lambda_v^\mu \ll d_{\mathrm{par}}(\mu, \nu)$. In this case, the inequality $(\star)$ in (5) may not hold. In fact, it will cause a fundamental problem to the above algorithm. Consider the case that $\lambda_v^\mu = \exp(-n)$ and $\lambda_v^\nu = \lambda_v^\mu + D$ for all $v \in V$, where $D > 0$. If we draw polynomial number of samples $\sigma \sim \mu$, typically, every sample $\sigma$ corresponds to the empty set, i.e., $\sigma_u = -1$

for all $u \in V$. Hence, with high probability, all $W_i$ in the above algorithm are $\frac{w_\nu(\varnothing)}{w_\mu(\varnothing)} = 1$ and the algorithm will return $\hat{d} = 0$. However, the true TV-distance $d_{\mathrm{TV}}(\mu, \nu) > 0$ is positive.

Let us first consider a special case such that for all $v \in V$, $\lambda_v^\mu < \kappa$ and $\lambda_v^\nu < \kappa$, where $\kappa = \mathrm{poly}(\frac{\varepsilon}{n}) \ll \frac{1}{n}$ is very small. The total variation distance is the sum of $\frac{1}{2}|\mu(\sigma) - \nu(\sigma)|$ for all independent sets $\sigma \in \{\pm\}^V$. In this special case, all external fields are tiny, so that large independent sets appear with very low probability. Let $\|\sigma\|_+$ be the number of $+1$ in $\sigma$. We can show that there is a constant $t = O(1)$ depending on $\kappa$ such that

$$\frac{1}{2} \sum_{\sigma \in \{\pm\}^V : \|\sigma\|_+ \geq t+1} |\mu(\sigma) - \nu(\sigma)| \leq O(\varepsilon) \cdot d_{\mathrm{TV}}(\mu, \nu).$$

In words, to approximate the total variation distance, we only need to consider the independent sets with size at most $t$. Now, we define a distribution $\mu'$ as the distribution $\mu$ restricted on the independent sets with size at most $t$. Similarly, we can define $\nu'$ from $\nu$. Since $t$ is a constant, we can enumerate all independent sets with size at most $t$ and compute the total variation distance between $\mu'$ and $\nu'$. We can show that $d_{\mathrm{TV}}(\mu', \nu')$ approximate $d_{\mathrm{TV}}(\mu, \nu)$ with $\varepsilon$ relative error.

For the most general case, we divide the vertices into two groups. The big group $B$ contains all vertices $v \in V$ such that $\min\{\lambda_v^\mu, \lambda_v^\nu\} > \kappa$. The small group $S$ contains all vertices $v \in V$ such that $\min\{\lambda_v^\mu, \lambda_v^\nu\} \leq \kappa$ for a small $\kappa = \mathrm{poly}(\frac{\varepsilon}{n})$. The total variation distance is

$$d_{\mathrm{TV}}(\mu, \nu) = \frac{1}{2} \sum_{\sigma \in \{\pm\}^V} |\mu(\sigma) - \nu(\sigma)| = \frac{1}{2} \sum_{x \in \{\pm\}^B} \sum_{y \in \{\pm\}^S} |\mu_B(x)\mu_S^x(y) - \nu_B(x)\nu_S^x(y)|$$

$$= \sum_{x \in \{\pm\}^B} \mu_B(x) \cdot \underbrace{\frac{1}{2} \sum_{y \in \{\pm\}^S} \left| \frac{\nu_B(x)}{\mu_B(x)} \nu_S^x(y) - \mu_S^x(y) \right|}_{f(x)}.$$

In a high level, our algorithm will draw independent samples $x \sim \mu_B$ from the marginal distribution and approximately compute the value of $f(x)$. The algorithm finally outputs the average value of $f(x)$ over all samples. To make the idea work, we need to deal with the following two main technical challenging.

- We need to bound the variance of $f(x)$ where $x \sim \mu_B$ to show that polynomial number of samples are sufficient for approximation. To achieve a good bound, we use the *Poincaré inequality* for hardcore model in uniqueness regime [CFYZ21] to control the variance.

- Given a sample $x \sim \mu_B$, we also need to approximately compute the value of $f(x)$. We need to (1) approximate two distributions $\mu_S^x$ and $\nu_S^x$ over $\{\pm\}^S$; and (2) approximate the ratio $\frac{\nu_B(x)}{\mu_B(x)}$. For the first task, note that both $\mu_S^x$ and $\nu_S^x$ are hardcore distributions on the induced subgraph $G[S]$ such that for all $v \in S$, the external fields are small. Hence, we can approximate them using distributions over independent sets of size at most $t = O(1)$. A similar idea will also be use for the second task: to estimate marginal probabilities $\mu_B(x)$ and $\nu_B(x)$, we also need to consider the total weight of all independent sets $I$ in $G[S]$ such that $I \cup \{v \in B \mid x_v = +1\}$ forms a independent set in $G$. Again, we show that the approximation algorithm only needs to consider all such $I$ with $|I| \leq t = O(1)$.

All the technical details for general case are given in Section 5.2.

**Organization of the paper** Lemma 15 is proved in Section 3. The additive error approximation algorithm is given in Section 4. The algorithm for instances with small parameter distance

in given in Section 5. In Section 6, we put all the pieces together to prove all algorithmic results. Our hardness results on approximating TV-distance for marginal distributions is proved in Section 7.

## 3 Parameter distance v.s. total variation distance

In this section, we prove Lemma 15. We first prove the result for the hardcore model in Section 3.1, and then prove the result for the soft-Ising model in Section 3.2.

### 3.1 Analysis for the hardcore model

Recall that our problem setting is: Let $G = (V, E)$ be a graph, and $(G, \lambda^\mu)$ and $(G, \lambda^\nu)$ are two hardcore models on the same graph, satisfying the uniqueness condition in (1). Let $\mu$ and $\nu$ denote distributions of $(G, \lambda^\mu)$ and $(G, \lambda^\nu)$ respectively. The parameter distance $d_{\mathrm{par}}(\mu, \nu)$ for hardcore model is defined by $d_{\mathrm{par}}(\mu, \nu) \triangleq \|\lambda^\mu - \lambda^\nu\|_\infty$. Now we prove the first hardcore model part of Lemma 15:

$$d_{\mathrm{TV}}(\mu, \nu) \geq C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu), \quad \text{where } C_{\mathrm{par}}^{\mathrm{TV}} = \frac{1}{5000}.$$

Let $i \in V$ be one vertex such that $|\lambda_i^\mu - \lambda_i^\nu| = d_{\mathrm{par}}(\mu, \nu)$. Without loss of generality, we can assume that $\lambda_i^\mu - \lambda_i^\nu = d_{\mathrm{par}}(\mu, \nu)$. Otherwise, we can flip the roles of $\mu$ and $\nu$ in the following proof. Define two collections of independent sets of $V$:

$$H_1 := \{S \text{ is an independent set of } G \mid i \in S\},$$
$$H_2 := \{S \setminus \{i\} \mid S \in H_1\}.$$

In this proof, we use $\mu(S)$ to denote the probability of $\sigma \in \{\pm\}^V$ in $\mu$ such that $\sigma_i = +1$ if and only if $i \in S$. Define $\mu(H_1) = \sum_{S \in H_1} \mu(S)$ and $\mu(H_2), \nu(H_1), \nu(H_2)$ in a similar way. It is easy to verify that

$$\frac{\mu(H_1)}{\mu(H_2)} = \lambda_i^\mu, \quad \frac{\nu(H_1)}{\nu(H_2)} = \lambda_i^\nu.$$

Then, we consider two cases depending on the value of $|\mu(H_2) - \nu(H_2)|$. The simple case is $|\mu(H_2) - \nu(H_2)| \geq C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu)$, then $d_{\mathrm{TV}}(\mu, \nu) \geq |\mu(H_2) - \nu(H_2)| \geq C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu)$.

The main case is $|\mu(H_2) - \nu(H_2)| < C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu)$. In this case we show that $|\mu(H_1) - \nu(H_1)| \geq C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu)$, which also implies $d_{\mathrm{TV}}(\mu, \nu) \geq |\mu(H_1) - \nu(H_1)| \geq C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu)$. We can lower bound the value of $|\mu(H_1) - \nu(H_1)|$ as follows:

$$\begin{aligned}
\mu(H_1) - \nu(H_1) &= \lambda_i^\mu \cdot \mu(H_2) - \lambda_i^\nu \cdot \nu(H_2) \\
&= (\lambda_i^\nu + d_{\mathrm{par}}(\mu, \nu))\mu(H_2) - \lambda_i^\nu \cdot \nu(H_2) \\
&= \lambda_i^\nu(\mu(H_2) - \nu(H_2)) + d_{\mathrm{par}}(\mu, \nu) \cdot \mu(H_2) \\
&> d_{\mathrm{par}}(\mu, \nu) \cdot \mu(H_2) - \lambda_i^\nu \cdot C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu). \quad (7)
\end{aligned}$$

Because $(G, \lambda^\mu)$ and $(G, \lambda^\nu)$ satisfy the uniqueness condition, we have for all vertex $i \in V$, all $x \in \{\mu, \nu\}$, $\lambda_i^x \leq \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta} \leq 4$ because $\Delta \geq 3$. Let $N(i)$ denotes the set of neighbors of $i$ in graph $G = (V, E)$. By definition of $H_2$, $\mu(H_2)$ is the probability of $j \notin S$ for all $j \in N(i) \cup \{i\}$

and $S \sim \mu$ is a random independent set from the hardcore model $(G, \lambda^\mu)$. Suppose there is a total ordering $<$ among all vertices in $V$. We have

$$\mu(H_2) = \mathbf{Pr}_{S \sim \mu}\left[i \notin S\right] \prod_{j \in N(i)} \mathbf{Pr}_{S \sim \mu}\left[j \notin S \mid (i \notin S) \wedge (\forall k \in N(i) \text{ with } k < j, k \notin S)\right]$$

$$\geq \frac{1}{1 + \lambda_i^\mu} \prod_{j \in N(i)} \left(\frac{1}{1 + \lambda_j^\mu}\right) \geq \left(\frac{1}{1 + 4/(\Delta - 2)}\right)^{\Delta + 1}$$

$$= \left(1 - \frac{4}{\Delta + 2}\right)^{\Delta + 1} > \frac{1}{1000}. \tag{8}$$

Also note that $\lambda_i^\nu \leq 4$ by the uniqueness condition. Recall $C_{\mathrm{par}}^{\mathrm{TV}} = \frac{1}{5000}$. By (7) and (8), we have

$$\mu(H_1) - \nu(H_1) > d_{\mathrm{par}}(\mu, \nu) \cdot \mu(H_2) - \lambda_i^\nu \cdot C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu)$$

$$> d_{\mathrm{par}}(\mu, \nu) \left(\frac{1}{1000} - \frac{4}{5000}\right) = C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu).$$

Next, we prove the second part of the hardcore model with $C_{\mathrm{par}}^{\mathrm{TV}} = b^3$. Now, we do not have the uniqueness condition but we have a marginal lower bound $b$ and the the hardcore model is soft. The proof is similar. We only need to show how to lower bound the value in (7). We first upper bound $\lambda_i^\nu$ in (7). Consider the pinning that all neighbors of $i$ are fixed as $-1$, then the probability of $i$ taking $-1$ is $\frac{1}{1 + \lambda_i^\nu} \geq b$. Thus, $\lambda_i^\nu \leq \frac{1-b}{b}$. Next, we lower bound the value of $\mu(H_2)$. Let $N(i)$ denote all neighbors of $i$. Note that $i$ takes value $+$ only if all neighbors of $i$ take the value $-$. Recall that we assume that $\lambda_i^\mu - \lambda_i^\nu = d_{\mathrm{par}}(\mu, \nu)$ in the beginning of the proof. We can assume that $d_{\mathrm{par}}(\mu, \nu) > 0$, otherwise $d_{\mathrm{TV}}(\mu, \nu) = 0$ and the lemma is trivial. We have $\lambda_i^\mu > 0$ so that $\mu_i(+) > 0$. By the marginal lower bound, $\mu_i(+) \geq b$. The vertex $i$ takes value $+$ only if all neighbors take the value $-$. We have

$$b \leq \mu_i(+) \leq \mathbf{Pr}_{S \sim \mu}\left[\forall j \in N(i), j \notin S\right].$$

On the other hand, the value of $\mu(H_2)$ can be lower bound by

$$\mu(H_2) = \mathbf{Pr}_{S \sim \mu}\left[\forall j \in N(i) \cup \{i\}, j \notin S\right]$$
$$= \mathbf{Pr}_{S \sim \mu}\left[\forall j \in N(i), j \notin S\right] \mathbf{Pr}_{S \sim \mu}\left[i \notin S \mid \forall j \in N(i), j \notin S\right]$$
$$\geq b^2.$$

We can set $C_{\mathrm{par}}^{\mathrm{TV}} = b^3$ so that

$$\mu(H_1) - \nu(H_1) > d_{\mathrm{par}}(\mu, \nu) \cdot \mu(H_2) - \lambda_i^\nu \cdot C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu)$$

$$\geq d_{\mathrm{par}}(\mu, \nu) \left(b^2 - \frac{1-b}{b} \cdot b^3\right) = C_{\mathrm{par}}^{\mathrm{TV}} \cdot d_{\mathrm{par}}(\mu, \nu).$$

## 3.2 Analysis for the soft-Ising model

Recall that our problem setting is: Let $G = (V, E)$ be a graph, and $(G, J^\mu, h^\mu)$ and $(G, J^\nu, h^\nu)$ are two soft-Ising models on the same graph. Let $\mu$ and $\nu$ denote distributions of $(G, J^\mu, h^\mu)$ and $(G, J^\nu, h^\nu)$ respectively. The parameter distance $d_{\mathrm{par}}(\mu, \nu)$ for soft-Ising model is defined by

$$d_{\mathrm{par}}(\mu, \nu) \triangleq \max\left\{\|J^\mu - J^\nu\|_{\max}, \max_v \frac{|h_v^\mu - h_v^\nu|}{\deg_v + 1}\right\}.$$

Now we prove the soft-Ising model part of Lemma 15: if both $\mu$ and $\nu$ are $b$-marginally bounded for $0 < b < 1$, then

$$d_{\mathrm{TV}}(\mu, \nu) \geq f(b) \cdot d_{\mathrm{par}}(\mu, \nu), \quad \text{where } f(b) = \frac{1}{2} b^2.$$

Before proving the above, we first present the following lemma.

**Lemma 16.** *Let $\mu$ and $\nu$ be any two distributions on $\{\pm\}^V$, and $S \subseteq V$ be a subset of vertices. Let $0 < \delta < 1$. If for any $\sigma \in \{\pm\}^S$, $d_{\mathrm{TV}}(\mu^\sigma, \nu^\sigma) \geq \delta$, then $d_{\mathrm{TV}}(\mu, \nu) \geq \delta/2$.*

*Proof.* For any two distributions $p, q$ on space $\Omega$, by the definition of total variation distance,

$$\sum_{x \in \Omega} \min(p(x), q(x)) = 1 - d_{\mathrm{TV}}(p, q),$$

$$\sum_{x \in \Omega} \max(p(x), q(x)) = 1 + d_{\mathrm{TV}}(p, q).$$

Under the assumption of the lemma, we have

$$
\begin{aligned}
1 - d_{\mathrm{TV}}(\mu, \nu) &= \sum_{\sigma \in \{\pm\}^V} \min(\mu(\sigma), \nu(\sigma)) \\
&= \sum_{\sigma \in \{\pm\}^S} \sum_{\tau \in \{\pm\}^{V \setminus S}} \min(\mu(\sigma)\mu(\tau \mid \sigma), \nu(\sigma)\nu(\tau \mid \sigma)) \\
&\leq \sum_{\sigma \in \{\pm\}^S} \max(\mu(\sigma), \nu(\sigma)) \sum_{\tau \in \{\pm\}^{V \setminus S}} \min(\mu(\tau \mid \sigma), \nu(\tau \mid \sigma)) \\
(\text{since } d_{\mathrm{TV}}(\mu^\sigma, \nu^\sigma) \geq \delta) \quad &\leq \sum_{\sigma \in \{\pm\}^S} \max(\mu(\sigma), \nu(\sigma))(1 - \delta) \\
&= (1 + d_{\mathrm{TV}}(\mu_S, \nu_S))(1 - \delta), \\
&\leq 1 - \delta + d_{\mathrm{TV}}(\mu_S, \nu_S)
\end{aligned}
$$

which implies

$$d_{\mathrm{TV}}(\mu, \nu) \geq \delta - d_{\mathrm{TV}}(\mu_S, \nu_S). \tag{9}$$

On the other hand, we have

$$d_{\mathrm{TV}}(\mu, \nu) \geq d_{\mathrm{TV}}(\mu_S, \nu_S). \tag{10}$$

Then Lemma 16 follows by combining (9) and (10). □

According to the definition of the parameter distance, there are 2 cases:

- It exists $\{u, v\} \in E$ such that $|J_{u,v}^\mu - J_{u,v}^\nu| = d_{\mathrm{par}}(\mu, \nu)$.

- It exists $v \in V$ such that $|h_v^\mu - h_v^\nu| = d_{\mathrm{par}}(\mu, \nu) \cdot (\deg_v + 1)$.

For the first case, we will show that

$$\forall \sigma \in \{\pm\}^{V \setminus \{u,v\}}, d_{\mathrm{TV}}(\mu^\sigma, \nu^\sigma) \geq 2f(b) d_{\mathrm{par}}(\mu, \nu), \tag{11}$$

and for the second case, we will show that

$$\forall \sigma \in \{\pm\}^{V \setminus \{v\}}, d_{\mathrm{TV}}(\mu^\sigma, \nu^\sigma) \geq 2f(b) d_{\mathrm{par}}(\mu, \nu). \tag{12}$$

Assuming (11) and (12) hold, using Lemma 16, we can directly derive the result of the total variation distance lower bound for soft-Ising model claimed in Lemma 15.

Next, we will prove (11) and (12), respectively to complete the entire proof.

*Proof.* [Proof of eq.(11)] Recall that $\mu, \nu$ are both $b$-marginal bounded which is defined in Condition 4. Thus for any $\sigma \in \{\pm\}^{V \setminus \{u,v\}}$, for any $\tau \in \{\pm\}^{\{u,v\}}$, we have $\mu^\sigma(u = \tau_u, v = \tau_v) = \mu_u^\sigma(\tau_u) \cdot \mu_v^{\sigma \wedge u \leftarrow \tau_u}(\tau_v) \geq b^2$. The same local lower bound also holds for the distribution $\nu$.

Define $c_u^\mu = h_u^\mu + \sum_{\{u,w\} \in E, w \neq v} J_{u,w}^\mu \sigma_w$ and $c_v^\mu = h_v^\mu + \sum_{\{v,w\} \in E, w \neq u} J_{v,w}^\mu \sigma_w$ be the influence coefficient of the external field $\sigma$ on $u, v$ in the distribution $\mu$. Similarly, let $c_u^\nu$ and $c_v^\nu$ denote the corresponding influence coefficients in the distribution $\nu$.

According to the definition of the Ising model, when the external field $\sigma$ is fixed, the local distribution of $u, v$ in the distribution $\mu$ is follows:

$$\mu^\sigma(u = +, v = +) = \exp(J_{u,v}^\mu + c_u^\mu + c_v^\mu - s^\mu),$$
$$\mu^\sigma(u = +, v = -) = \exp(-J_{u,v}^\mu + c_u^\mu - c_v^\mu - s^\mu),$$
$$\mu^\sigma(u = -, v = +) = \exp(-J_{u,v}^\mu - c_u^\mu + c_v^\mu - s^\mu),$$
$$\mu^\sigma(u = -, v = -) = \exp(J_{u,v}^\mu - c_u^\mu - c_v^\mu - s^\mu),$$

where $s^\mu = \log(\exp(J_{u,v}^\mu + c_u^\mu + c_v^\mu) + \exp(-J_{u,v}^\mu + c_u^\mu - c_v^\mu)) + \exp(-J_{u,v}^\mu - c_u^\mu + c_v^\mu) + \exp(J_{u,v}^\mu - c_u^\mu - c_v^\mu))$.

Using the same method, we define $s^\nu$, and the local distribution of $u, v$ in the distribution $\nu$ can be expressed in terms of $c_u^\nu, c_v^\nu$ and $s^\nu$, as in the above equation.

Let, $p = J_{u,v}^\mu + c_u^\mu + c_v^\mu - s^\mu$ and $q = J_{u,v}^\nu + c_u^\nu + c_v^\nu - s^\nu$, then

$$d_{\text{TV}}(\mu^\sigma, \nu^\sigma) \geq |\mu^\sigma(u = +, v = +) - \nu^\sigma(u = +, v = +)|$$
$$= |\exp(p) - \exp(q)| = \int_{x = \min\{p,q\}}^{\max\{p,q\}} \exp(x) dx$$
$$\geq \int_{x = \min\{p,q\}}^{\max\{p,q\}} \exp(p) dx = \exp(p) \cdot |p - q|$$

(by marginal lower bound) $\quad \geq b^2 \cdot |p - q|. \tag{13}$

Based on this, we can derive the following lower bound on the total variation distance between $\mu^\sigma$ and $\nu^\sigma$:

$$d_{\text{TV}}(\mu^\sigma, \nu^\sigma) \geq b^2 \cdot |(J_{u,v}^\mu + c_u^\mu + c_v^\mu - s^\mu) - (J_{u,v}^\nu + c_u^\nu + c_v^\nu - s^\nu)|. \tag{14}$$

Using the same method, consider the distribution differences in $\mu^\sigma$ and $\nu^\sigma$ for the remaining three cases $(u = +, v = -), (u = -, v = +)$ and $(u = -, v = -)$, we can also obtain the following lower bounds:

$$d_{\text{TV}}(\mu^\sigma, \nu^\sigma) \geq b^2 \cdot |-(-J_{u,v}^\mu + c_u^\mu - c_v^\mu - s^\mu) + (-J_{u,v}^\nu + c_u^\nu - c_v^\nu - s^\nu)|, \tag{15}$$
$$d_{\text{TV}}(\mu^\sigma, \nu^\sigma) \geq b^2 \cdot |-(-J_{u,v}^\mu - c_u^\mu + c_v^\mu - s^\mu) + (-J_{u,v}^\nu - c_u^\nu + c_v^\nu - s^\nu)|, \tag{16}$$
$$d_{\text{TV}}(\mu^\sigma, \nu^\sigma) \geq b^2 \cdot |(J_{u,v}^\mu - c_u^\mu - c_v^\mu - s^\mu) - (J_{u,v}^\nu - c_u^\nu - c_v^\nu - s^\nu)|. \tag{17}$$

Note that the absolute value operation satisfies the triangle inequality, by combining inequalities (14), (15), (16) and (17), we have

$$4 d_{\text{TV}}(\mu^\sigma, \nu^\sigma) \geq b^2 \cdot 4 |J_{u,v}^\mu - J_{u,v}^\nu|.$$

Then for the case that $|J_{u,v}^\mu - J_{u,v}^\nu| = d_{\text{par}}(\mu, \nu)$, the lower bound of local total variation distance in (11) is proven. $\qquad \square$

*Proof.* [Proof of eq.(12)] For the case that $|h_v^\mu - h_v^\nu| = d_{\text{par}}(\mu, \nu) \cdot (\deg_v + 1)$. Define $c^\mu = h_v^\mu + \sum_{\{u,v\} \in E} J_{u,v}^\mu \sigma_v$ and $c^\nu = h_v^\nu + \sum_{\{u,v\} \in E} J_{u,v}^\nu \sigma_v$ be the influence coefficient of the external

field $\sigma$ on $v$ in the distribution $\mu$ and $\nu$ respectively, note that for each $\{u, v\} \in E$, $|J_{u,v}^{\mu} - J_{u,v}^{\nu}|$ is bounded by $d_{\mathrm{par}}(\mu, \nu)$, then

$$
\begin{aligned}
|c^{\mu} - c^{\nu}| &\geq |h_v^{\mu} - h_v^{\nu}| - \sum_{\{u,v\} \in E} |J_{u,v}^{\mu} - J_{u,v}^{\nu}| \\
&\geq (\deg_v + 1) d_{\mathrm{par}}(u, v) - \sum_{\{u,v\} \in E} d_{\mathrm{par}}(\mu, \nu) = d_{\mathrm{par}}(\mu, \nu).
\end{aligned}
$$

According to the definition of the Ising model, when the external field $\sigma$ is fixed, the local distribution on $v$ in the distribution $\mu$ is follows:

$$
\mu^{\sigma}(v = +) = \exp(c^{\mu} - s^{\mu}), \quad \mu^{\sigma}(v = -) = \exp(-c^{\mu} - s^{\mu}),
$$

where $s^{\mu} = \log(\exp(c^{\mu}) + \exp(-c^{\mu}))$.

Using the same method, we define $s^{\nu}$, and the local distribution of $u, v$ in the distribution $\nu$ can be expressed in terms of $c^{\nu}$ and $s^{\nu}$, as in the above equation.

Recall that $\mu, \nu$ are both $b$-marginal bounded, applying the same method in (13), we can derive the following lowerbound on the total variation distance between $\mu^{\sigma}$ and $\nu^{\sigma}$:

$$
d_{\mathrm{TV}}(\mu^{\sigma}, \nu^{\sigma}) \geq b \cdot |(c^{\mu} - s^{\mu}) - (c^{\nu} - s^{\nu})|, \tag{18}
$$
$$
d_{\mathrm{TV}}(\mu^{\sigma}, \nu^{\sigma}) \geq b \cdot |-(-c^{\mu} - s^{\mu}) + (-c^{\nu} - s^{\nu})|. \tag{19}
$$

By combining the above inequalities, we have

$$
2 d_{\mathrm{TV}}(\mu^{\sigma}, \nu^{\sigma}) \geq b \cdot 2 |c^{\mu} - c^{\nu}| \geq 2b \cdot d_{\mathrm{par}}(\mu^{\sigma}, \nu^{\sigma}).
$$

Note $0 < b < 1$, then the lower bound of local total variation distance in (12) is proven. $\qquad \square$

## 4 Additive-error approximation algorithm

### 4.1 TV-distance between two Gibbs distributions

We first present an algorithm can achieves the additive-error approximation to the total variation distance between two *general* Gibbs distributions, which covers Ising and hardcore models as special cases. Let $\mu$ over $\{0, 1\}^V$ be a Gibbs distribution over graph $G = (V, E)$. For any configuration $\sigma \in \{0, 1\}^V$, $\mu(\sigma) = w_{\mu}(\sigma) / Z_{\mu}$, where $w_{\mu}(\cdot)$ is the weight function and $Z_{\mu} = \sum_{\tau \in \{0,1\}^V} w_{\mu}(\sigma)$. Let $T_G^{\mathsf{wt}} \in \mathbb{N}$. We say the Gibbs distribution $\mu$ admits a weight oracle with cost $T_G^{\mathsf{wt}}$ is given any $\sigma \in [q]^V$, it returns the exact weight $w_{\mu}(\sigma)$ in time $T_G^{\mathsf{wt}}$. Note that both Ising and hardcore models, as well as most Gibbs distributions, admit weight oracle with cost $T_G^{\mathsf{wt}} = O(|V| + |E|)$. Recall the sampling and approximate counting oracles are defined in Definition 1.

**Theorem 17.** *There exists an algorithm such that given two general Gibbs distributions $\mu$ and $\nu$ on the same graph $G = (V, E)$ and an error bound $\varepsilon > 0$, if $\mu$ and $\nu$ both admit weight, sampling, and approximate counting oracles with cost $T_G^{\mathsf{wt}}$ and cost functions $T_G^{\mathsf{sp}}(\cdot)$ and $T_G^{\mathsf{ct}}(\cdot)$ respectively, then it returns a random $\hat{d}$ in time $O(T_G^{\mathsf{ct}}(\frac{\varepsilon}{4}) + \frac{1}{\varepsilon^2}(T_G^{\mathsf{wt}} + T_G^{\mathsf{sp}}(\frac{\varepsilon}{4})))$ such that*

$$
\mathbf{Pr}\left[d_{\mathrm{TV}}(\mu, \nu) - \varepsilon \leq \hat{d} \leq d_{\mathrm{TV}}(\mu, \nu) + \varepsilon\right] \geq \frac{2}{3}.
$$

*Proof.* Define a random variable $X \in [0, 1]$ such that $X = \max \left( 0, 1 - \frac{\nu(\sigma)}{\mu(\sigma)} \right)$, where $\sigma \sim \mu$.

$$d_{\mathrm{TV}}(\mu, \nu) = \sum_{\sigma : \mu(\sigma) > \nu(\sigma)} |\mu(\sigma) - \nu(\sigma)| = \sum_{\sigma : \mu(\sigma) > \nu(\sigma)} \mu(\sigma) \left| 1 - \frac{\nu(\sigma)}{\mu(\sigma)} \right|$$

$$= \sum_{\sigma} \mu(\sigma) \max \left( 0, 1 - \frac{\nu(\sigma)}{\mu(\sigma)} \right) = \mathbf{E}\left[ X \right].$$

By the definition of $X$, we have $0 \leq X \leq 1$, so $\mathbf{Var}\left[ X \right] \leq 1$. Ideally, we want to draw independent samples of $X$ and take average to approximate $d_{\mathrm{TV}}(\mu, \nu)$. However, the main issue is that given a $\sigma \sim \mu$, we cannot compute neither $\mu(\sigma)$ nor $\nu(\sigma)$ exactly. Alternatively, we will define another random variable $\hat{X} \in [0, 1]$ that approximate the random variable $X$.

Call the approximate counting oracles of $\mu$ and $\nu$ to obtain $\hat{Z}_\mu$ and $\hat{Z}_\nu$ that approximate partition functions $Z_\mu$ and $Z_\nu$ with relative error bound $\frac{\varepsilon}{4}$. We may assume both counting oracles succeed, which happens with probability at least 0.98. Define the random variable $\hat{X} \in [0, 1]$ by the following process.

1. Call sampling oracle of $\mu$ to obtain one sample $\sigma \in \{\pm\}^V$ such that $d_{\mathrm{TV}}(\sigma, \mu) \leq \frac{\varepsilon}{4}$.

2. Call weight oracles of both $\mu$ and $\nu$ to obtain exact weights $w_\mu(\sigma), w_\nu(\sigma)$. Compute $\hat{\mu}(\sigma) = w_\mu(\sigma)/\hat{Z}_\mu$ and $\hat{\nu}(\sigma) = w_\nu(\sigma)/\hat{Z}_\nu$.

3. Define $\hat{X} = \max(0, 1 - \hat{\nu}(\sigma)/\hat{\mu}(\sigma))$, in particular, $\hat{X} = 0$ if $\hat{\mu}(\sigma) = 0$.

Let $T = \frac{64}{\varepsilon^2}$. Our algorithm is the following simple process:

- Draw $T$ independent samples $\hat{X}_1, \hat{X}_2, \ldots, \hat{X}_T$ of random variable $\hat{X}$.

- Output the average $\hat{d} = \frac{1}{T} \sum_{i=1}^{T} \hat{X}_i$.

It is easy to see the running time of our algorithm is $2T_G^{\mathsf{ct}}(\frac{\varepsilon}{4}) + T \cdot (T_G^{\mathsf{sp}}(\frac{\varepsilon}{4}) + 2T_G^{\mathsf{wt}} + O(1))$.

To prove the correctness of our algorithm, we only need to show that

$$\left| \mathbf{E}\left[ \hat{X} \right] - \mathbf{E}\left[ X \right] \right| = \left| \mathbf{E}\left[ \hat{X} \right] - d_{\mathrm{TV}}(\mu, \nu) \right| \leq \frac{7\varepsilon}{8}. \tag{20}$$

Note that $0 \leq \hat{X} \leq 1$ so that $\mathbf{Var}\left[ \hat{X} \right] \leq 1$. By Hoeffding's inequality, it is easy to show that with probability at least 0.9, $|\hat{d} - \mathbb{E}[\hat{X}]| \leq \varepsilon/8$. Combining with (20) proves the theorem.

Now, we only need to verify (20). We introduce a new random variable $X^*$ in analysis. In the definition of $\hat{X}$, assume we replace the sample $\sigma$ in Item 1 with a perfect sampler of the distribution $\mu$. Let $X^*$ denote the resulting random variable. We first compare $\mathbf{E}\left[ X^* \right]$ with $\mathbf{E}\left[ X \right]$. The difference between $X^*$ and $X$ comes from the error of computing the ratio of $\mu(\sigma)$ and $\nu(\sigma)$. Note that $\frac{\nu(\sigma)}{\mu(\sigma)} = \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \cdot \frac{Z_\mu}{Z_\nu}$ and $\frac{\hat{\nu}(\sigma)}{\hat{\mu}(\sigma)} = \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \cdot \frac{\hat{Z}_\mu}{\hat{Z}_\nu}$. By the definition of approximate counting oracle, for $\sigma$ with $\mu(\sigma) > 0$,

$$\left( 1 - \frac{5\varepsilon}{8} \right) \frac{\nu(\sigma)}{\mu(\sigma)} \leq \frac{\hat{\nu}(\sigma)}{\hat{\mu}(\sigma)} \leq \left( 1 + \frac{5\varepsilon}{8} \right) \frac{\nu(\sigma)}{\mu(\sigma)}.$$

We can compute the expectation as

$$\mathbf{E}\left[X^*\right] = \sum_{\sigma:\mu(\sigma)>0} \mu(\sigma) \max\left(0, 1 - \frac{\hat{\nu}(\sigma)}{\hat{\mu}(\sigma)}\right) \leq \sum_{\sigma:\mu(\sigma)>0} \mu(\sigma) \max\left(0, 1 - \left(1 - \frac{5\varepsilon}{8}\right)\frac{\nu(\sigma)}{\mu(\sigma)}\right)$$

$$\leq \sum_{\sigma:\mu(\sigma)>0} \mu(\sigma)\left(\max\left(0, 1 - \frac{\nu(\sigma)}{\mu(\sigma)}\right) + \frac{5\varepsilon}{8}\frac{\nu(\sigma)}{\mu(\sigma)}\right) \leq d_{\mathrm{TV}}\left(\mu, \nu\right) + \frac{5\varepsilon}{8} \sum_{\sigma:\mu(\sigma)>0} \nu(\sigma)$$

$$\leq d_{\mathrm{TV}}\left(\mu, \nu\right) + \frac{5\varepsilon}{8}.$$

Using the same way, we could verify the other direction. We have

$$\left|\mathbf{E}\left[X^*\right] - \mathbf{E}\left[X\right]\right| = \left|\mathbf{E}[X^*] - d_{\mathrm{TV}}\left(\mu, \nu\right)\right| \leq \frac{5\varepsilon}{8}. \tag{21}$$

Next, we compare $X^*$ to $\hat{X}$. The only difference is that they sample $\sigma$ from different distributions. Let $\mu'$ be the distribution defined by the approximate sampling oracle. Define a function $f$ such that for any $\sigma \in \{\pm\}^V$, $f(\sigma) = \max(0, 1 - \hat{\nu}(\sigma)/\hat{\mu}(\sigma))$, where we set $f(\sigma) = 0$ if $\hat{\mu}(\sigma) = 0$. We have $\left|\mathbf{E}(\hat{X}) - \mathbf{E}(X^*)\right| = \left|\mathbf{E}_{\sigma\sim\mu}[f(\sigma)] - \mathbf{E}_{\sigma\sim\mu'}[f(\sigma)]\right|$. Define $A = \{\sigma \mid \mu(\sigma) > \mu'(\sigma)\}$ and $B = \{\sigma \mid \mu(\sigma) < \mu'(\sigma)\}$. We can write

$$\left|\mathbf{E}(\hat{X}) - \mathbf{E}(X^*)\right| = \left|\sum_{\sigma\in A}(\mu'(\sigma) - \mu(\sigma))f(\sigma) + \sum_{\sigma\in B}(\mu'(\sigma) - \mu(\sigma))f(\sigma)\right|$$

$$(\star) \quad \leq \max\left(\sum_{\sigma\in A}(\mu(\sigma) - \mu'(\sigma))f(\sigma), \sum_{\sigma\in B}(\mu'(\sigma) - \mu(\sigma))f(\sigma)\right)$$

$$(\text{by } f(\sigma) \in [0,1]) \quad \leq d_{\mathrm{TV}}\left(\mu, \mu'\right) \leq \frac{\varepsilon}{4}, \tag{22}$$

where in inequality $(\star)$, we use $\sum_{\sigma\in A}(\mu'(\sigma) - \mu(\sigma))f(\sigma) \leq 0$ and $\sum_{\sigma\in B}(\mu'(\sigma) - \mu(\sigma))f(\sigma) \geq 0$. Finally, (20) holds due to (21) and (22). $\qquad\square$

Theorem 17 implies the following corollary for concrete models.

**Corollary 18.** *There exist an FPRAS for approximating TV-distances with additive error for following models: (1) Hardcore model satisfying uniqueness condition; (2) Ising model with spectral condition; (3) Ferromagnetic interaction with consistent field condition; and (4) Anti-ferromagnetic interaction at or within the uniqueness threshold.*

The definitions of the conditions can be found in (1) and Condition 6. In contrast to the relative-error approximation, the above corollary does not require a marginal lower bound for the Ising model. Furthermore, since the proof of Theorem 17 does not use Lemma 15, Corollary 18 holds for general (not necessarily soft) Ising models.

## 4.2 TV-distance between two marginal distributions

In this subsection, we present an algorithm can achieves the additive-error approximation to the total variance distance between two marginal distributions. Let $\mu$ over $\{\pm\}^V$ be a Gibbs distribution over graph $G = (V, E)$, and $S$ is a subset of $V$. For any configuration $\sigma$, let $\sigma$ be a partial configuration over $\{\pm\}^S$. Recall that $\mu_S(\sigma) \propto Z^\sigma$, where $Z^\sigma$ is the conditional partition function of $\sigma$ defined as

$$Z^\sigma \triangleq \sum_{\tau\in\{\pm\}^V:\tau_S=\sigma} w_\mu(\tau).$$

**Definition 19** (approximate conditional counting oracle). Let $\mathbb{S}$ be a spin system on graph G with Gibbs distribution $\mu$. Let $T_G^{\mathsf{ct}} : (0,1) \to \mathbb{N}$ be a function. We say $\mathbb{S}$ admits a conditional counting oracle with cost function $T_G^{\mathsf{ct}}(\cdot)$ if given any $0 < \varepsilon < 1$, and any partial configuration $\sigma \in \{\pm\}^S$ on a subset $S \subseteq V$, it returns a random number $\hat{Z}_\mu^\sigma$ in time $T_G^{\mathsf{ct}}(\varepsilon)$ such that $Z_\mu^\sigma(1 - \varepsilon) \leq \hat{Z}_\mu^\sigma(\sigma) \leq Z_\mu^\sigma(1 + \varepsilon)$ with probability at least 0.99.

The oracle above is stronger than the approximate counting oracle in Definition 1. The approximate counting oracle only answers the query for $S = \varnothing$, while the conditional counting oracle can answer the query for any subset $S \subseteq V$.

**Theorem 20.** *There exists an algorithm such that given two general Gibbs distributions $\mu$ and $\nu$ on the same graph $G = (V, E)$ with $n = |V|$, any subset $S \subseteq V$, and any $\varepsilon > 0$, if $\mu$ and $\nu$ both admit sampling and conditional counting oracles with cost functions $T_G^{\mathsf{sp}}(\cdot)$ and $T_G^{\mathsf{ct}}(\cdot)$ respectively, then it returns a random number $\hat{d}$ in time $O(\frac{1}{\varepsilon^2} \log \frac{1}{\varepsilon}) \cdot (T_G^{\mathsf{ct}}(\frac{\varepsilon}{8}) + T_G^{\mathsf{sp}}(\frac{\varepsilon}{8}))$ such that*

$$\mathbf{Pr}\left[\left|\hat{d} - d_{\mathrm{TV}}(\mu_S, \nu_S)\right| \leq \varepsilon\right] \geq \frac{2}{3}.$$

*Proof.* Define a random variable $Y \in [0, 1]$ such that $Y = \max\left(0, 1 - \frac{\nu_S(\sigma)}{\mu_S(\sigma)}\right)$, where $\sigma \sim \mu_S$.

$$\begin{aligned}
d_{\mathrm{TV}}(\mu_S, \mu_S) &= \sum_{\sigma \in \{\pm\}^S} \max(0, \mu_S(\sigma) - \nu_S(\sigma)) \\
&= \sum_{\sigma \in \{\pm\}^S} \mu_S(\sigma) \max\left(0, 1 - \frac{\nu_S(\sigma)}{\mu_S(\sigma)}\right) = \mathbf{E}[Y].
\end{aligned}$$

The random variable $Y$ satisfies that $0 \leq 1 \leq Y$ so $\mathbf{Var}[Y] \leq 1$. Similar to the proof of Theorem 17, we want to draw independent sample of $Y$ and take average to approximate $d_{\mathrm{TV}}(\mu_S, \nu_S)$. However, here $\mu_S(\sigma) = Z_\mu^\sigma / Z_\mu$ and $\nu_S(\sigma) = Z_\nu^\sigma / Z_\nu$. An additional problem is that we cannot exactly compute the weight $Z_\mu^\sigma$ and $Z_\nu^\sigma$ for each partial configuration $\sigma \sim \mu_S$.

First note that we can boost the success probability of conditional counting oracle from 0.99 to $1 - \delta$ by calling it independently for $O(\log \frac{1}{\delta})$ times and take the median. Let $\delta = \frac{\varepsilon^2}{320}$. Call conditional counting oracles with $S = V$ to obtain $\hat{Z}_\mu$ and $\hat{Z}_\nu$ that approximate partition functions $Z_\mu$ and $Z_\nu$ with relative error bound $\frac{\varepsilon}{8}$. We may assume both counting oracles succeed, which happens with probability at least $1 - \delta$. Similarly, we first define the random variable $\hat{Y}$ by the following process:

1. Call sampling oracle of $\mu$ to obtain one sample $\sigma \in \{\pm\}^V$ such that $d_{\mathrm{TV}}(\sigma, \mu) \leq \frac{\varepsilon}{8}$, and we use $\sigma_S \in \{\pm\}^S$ as our sample.

2. Call conditional counting oracles to obtain $\hat{Z}_\mu^{\sigma_S}$ and $\hat{Z}_\nu^{\sigma_S}$ with an error bound $\varepsilon/8$ and success probability $1 - \delta$. Compute $\hat{\mu}_S(\sigma_S) = \hat{Z}_\mu^{\sigma_S} / \hat{Z}_\mu$ and $\hat{\nu}_S(\sigma_S) = \hat{Z}_\nu^{\sigma_S} / \hat{Z}_\nu$.

3. Define $\hat{Y} = \max(0, 1 - \hat{\nu}_S(\sigma_S)/\hat{\mu}_S(\sigma_S))$, and in particular, $\hat{Y} = 0$ if $\hat{\mu}_S(\sigma) = 0$.

Let $T = \frac{64}{\varepsilon^2}$, we present our algorithm by following process:

- Draw T samples $\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_n$ of random variable $\hat{Y}$.

- Output the average $\hat{d} = \frac{1}{T} \sum_{i=1}^T \hat{Y}_i$.

The running time of our algorithm is $(T_G^{\mathsf{ct}}(\frac{\varepsilon}{8}) + T_G^{\mathsf{sp}}(\frac{\varepsilon}{8})) \cdot T \cdot O(\log \frac{1}{\delta})$.

We now analyze the approximation error. First note that the probability that all conditional counting oracles success is $(1 - \frac{\varepsilon^2}{320})^{2T+2} > 0.98$ and $d_{\mathrm{TV}}(\sigma, \mu) \leq \frac{\varepsilon}{8}$ implies $d_{\mathrm{TV}}(\sigma_S, \mu_S) \leq \frac{\varepsilon}{8}$. Compared to the proof of Theorem 17, the difference is that we can only compute $\hat{Z}_\mu^{\sigma_S}$ and $\hat{Z}_\nu^{\sigma_S}$ approximately. However, the error can still be bounded. We prove that: for $\sigma_S$ with $\mu_S(\sigma_S) > 0$, $\left(1 - \frac{5\varepsilon}{8}\right) \frac{\nu_S(\sigma_S)}{\mu_S(\sigma_S)} \leq \frac{\hat{\nu}_S(\sigma_S)}{\hat{\mu}_S(\sigma_S)} \leq \left(1 + \frac{5\varepsilon}{8}\right) \frac{\nu_S(\sigma_S)}{\mu_S(\sigma_S)}$. Since $\frac{\hat{\nu}_S(\sigma_S)}{\hat{\mu}_S(\sigma_S)} = \frac{\hat{Z}_\nu^{\sigma_S}}{\hat{Z}_\mu^{\sigma_S}} \frac{\hat{Z}_\mu}{\hat{Z}_\nu}$ and $\frac{\nu_S(\sigma_S)}{\mu_S(\sigma_S)} = \frac{Z_\nu^{\sigma_S}}{Z_\mu^{\sigma_S}} \frac{Z_\mu}{Z_\nu}$,

$$\frac{\hat{\nu}_S(\sigma_S)}{\hat{\mu}_S(\sigma_S)} = \frac{\hat{Z}_\nu^{\sigma_S}}{\hat{Z}_\mu^{\sigma_S}} \frac{\hat{Z}_\mu}{\hat{Z}_\nu} \leq \frac{Z_\nu^{\sigma_S}(1 + \frac{\varepsilon}{8})}{Z_\mu^{\sigma_S}(1 - \frac{\varepsilon}{8})} \frac{Z_\mu(1 + \frac{\varepsilon}{8})}{Z_\nu(1 - \frac{\varepsilon}{8})} < \left(1 + \frac{5\varepsilon}{8}\right) \frac{\nu_S(\sigma_S)}{\mu_S(\sigma_S)}.$$

The other side of the inequality can be proved similarly. The rest of the proof follows from the proof of Theorem 17. $\qquad\square$

Theorem 13 is a simple corollary of Theorem 20, which is proved in section 6.3.

## 5  The algorithm for instances with small parameter distance

Let $\mu$ and $\nu$ be two general Gibbs distributions (including hardcore and Ising models) on the same graph $G = (V, E)$ with $n = |V|$. Let $w_\mu(\sigma)$ and $w_\nu(\sigma)$ be the weights of $\mu$ and $\nu$ on configuration $\sigma \in \{\pm\}^V$. In this section, we focus on the case where the parameter distance $d_{\mathrm{par}}(\mu, \nu)$ is small. We will first give a basic algorithm for instances satisfying Condition 21, and then verify the Condition 21 for Ising models with small parameter distance. We next give a more advanced algorithm for hardcore models with small parameter distance.

### 5.1  Basic algorithm

Define random variable

$$W \triangleq \frac{w_\nu(\sigma)}{w_\mu(\sigma)}, \quad \text{where } \sigma \sim \mu. \tag{23}$$

**Condition 21.** *Let $K, L \geq 1$ be two parameters. Two Gibbs distributions $\mu$ and $\nu$ satisfy that*

- *$\nu$ is absolutely continuous with respect to $\mu$: for all $\sigma \in \{\pm\}^V$, if $\mu(\sigma) = 0$, then $\nu(\sigma) = 0$;*
- *$\sqrt{\mathbf{Var}[W]} \leq K d_{\mathrm{TV}}(\mu, \nu)$;*
- *$\mathbf{E}[W] \geq \frac{1}{L}$.*

**Theorem 22.** *There exists an algorithm such that given two Gibbs distributions $\mu$ and $\nu$ on the same graph $G = (V, E)$, and any $0 < \varepsilon < 1$, if $\mu$ and $\nu$ satisfy Condition 21 with $K$ and $L$, and both admit sampling and approximate counting oracles with cost functions $T_G^{\mathsf{sp}}(\cdot)$ and $T_G^{\mathsf{ct}}(\cdot)$, then it returns a random number $\hat{d}$ in time $O(T_G^{\mathsf{ct}}(\frac{\varepsilon}{4}) + T \cdot T_G^{\mathsf{sp}}(\frac{1}{100T}))$, where $T = O(\frac{L^2 K^2}{\varepsilon^2})$, such that*

$$\mathbf{Pr}\left[(1 - \varepsilon) d_{\mathrm{TV}}(\mu, \nu) \leq \hat{d} \leq (1 + \varepsilon) d_{\mathrm{TV}}(\mu, \nu)\right] \geq \frac{2}{3}.$$

*Proof.* Since $\nu$ is absolutely continuous with respect to $\mu$ ($\nu \ll \mu$), we can compute that:

$$\begin{aligned}
d_{\mathrm{TV}}(\mu, \nu) &= \frac{1}{2} \sum_{\sigma \in \{\pm\}^V} |\mu(\sigma) - \nu(\sigma)| = \frac{1}{2} \sum_{\sigma \in \{\pm\}^V : \mu(\sigma) > 0} \mu(\sigma) \left|1 - \frac{\nu(\sigma)}{\mu(\sigma)}\right| \\
&= \frac{1}{2} \sum_{\sigma \in \{\pm\}^V : \mu(\sigma) > 0} \mu(\sigma) \left|1 - \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \frac{Z_\mu}{Z_\nu}\right| = \frac{Z_\mu}{2 Z_\nu} \sum_{\sigma \in \{\pm\}^V : \mu(\sigma) > 0} \mu(\sigma) \left|\frac{Z_\nu}{Z_\mu} - \frac{w_\nu(\sigma)}{w_\mu(\sigma)}\right| \\
&= \frac{Z_\mu}{2 Z_\nu} \mathbf{E}\left[|\mathbf{E}[W] - W|\right], \tag{24}
\end{aligned}$$

20

where in the last step we use the fact that if $\nu \ll \mu$, then $\mathbf{E}\left[W\right] = \sum_{\sigma \in \{\pm\}^V : \mu(\sigma) > 0} \frac{w_\mu(\sigma)}{Z_\mu} \frac{w_\nu(\sigma)}{w_\mu(\sigma)} = \frac{Z_\nu}{Z_\mu}$. We next use sampling oracles to define a random variable $\hat{W} \in [0, +\infty)$, which serves as an approximation of the random variable $W$ in (23). Define

$$T \triangleq \left\lceil \frac{10^4 L^2 K^2}{\varepsilon^2} \right\rceil.$$

1. Call the sampling oracle of $\mu$ to obtain a random $\sigma \in \{\pm\}^V$ such that $d_{\mathrm{TV}}\left(\sigma, \mu\right) \leq \frac{1}{100T}$.

2. Compute $\hat{W} = \frac{w_\nu(\sigma)}{w_\mu(\sigma)}$. In particular, if $w_\mu(\sigma) = 0$[4], then we set $\hat{W} = 0$.

Given the random variable $\hat{W}$, our algorithm is given by the following processes:

> **Basic algorithm for instances satisfying Condition 21**
>
> - Call approximate counting oracles to obtain $\hat{Z}_\mu$ and $\hat{Z}_\nu$ with error $\frac{\varepsilon}{4}$.
>
> - Draw $T$ samples $\hat{W}_1, \ldots, \hat{W}_T$ from $\hat{W}$ independently.
>
> - Compute $\bar{W} = \frac{1}{T} \sum_{i=1}^T \hat{W}_i$.
>
> - Compute $\bar{E} = \frac{1}{T} \sum_{i=1}^T |\hat{W}_i - \bar{W}|$.
>
> - Return $\hat{d} = \frac{\hat{Z}_\mu}{2\hat{Z}_\nu} \bar{E}$.

The total running time of the above algorithm is

$$2T_G^{\mathsf{ct}}\left(\frac{\varepsilon}{4}\right) + O\left(T \cdot T_G^{\mathsf{sp}}\left(\frac{1}{100T}\right)\right).$$

We remark that if both sampling and approximate counting can be solved in polynomial time, i.e., for any $\delta \in (0, 1)$, $T_G^{\mathsf{ct}}(\delta), T_G^{\mathsf{sp}}(\delta) = \mathrm{poly}(\frac{n}{\delta})$, where $n$ is the number of vertices in $G$, then the above running time is $\mathrm{poly}(\frac{nLK}{\varepsilon})$.

We now prove the correctness of the algorithm. First, due to the definition of approximate counting oracle, with probability at least 0.98, we can bound the error from $\hat{Z}_\mu \hat{Z}_\nu$ as follows:

$$\left(1 - \frac{3\varepsilon}{4}\right) \frac{Z_\mu}{Z_\nu} \leq \frac{(1 - \varepsilon/4)Z_\mu}{(1 + \varepsilon/4)Z_\nu} \leq \frac{\hat{Z}_\mu}{\hat{Z}_\nu} \leq \frac{(1 + \varepsilon/4)Z_\mu}{(1 - \varepsilon/4)Z_\nu} < \left(1 + \frac{3\varepsilon}{4}\right) \frac{Z_\mu}{Z_\nu}. \tag{25}$$

Suppose we can access a perfect sampler of $\mu$ and we draw perfect samples $W_1, \ldots, W_T$ of $W$. For each pair of $W_i$ and $\hat{W}_i$, there exists a coupling of $W_i, \hat{W}_i$ such that $\mathbf{Pr}\left[W_i \neq \hat{W}_i\right] \leq \frac{1}{100T}$. Then with probability at least 0.99, $W_i = \hat{W}_i$ for all $1 \leq i \leq T$. Consider an ideal algorithm that can use the perfect samples $W_1, \ldots, W_T$. Our real algorithm can be coupled successfully with the ideal algorithm with probability at least 0.99. If we can show the ideal algorithm outputs correct result with probability at least 0.96, then our real algorithm outputs correct result with probability at least $0.95 > 2/3$.

Now we assume all $W_i$ are perfect samples of $W$. We compute $\bar{W} = \frac{1}{T} \sum_{i=1}^T W_i$ and similarly $\bar{E}$ and $\hat{d}$. We only need to prove that $(1 - \varepsilon)d \leq \hat{d} \leq (1 + \varepsilon)d$ with probability at least 0.9, where $d = d_{\mathrm{TV}}\left(\mu, \nu\right)$. For every random variable $W_i$, by triangle inequality, we have

$$|\mathbf{E}\left[W\right] - W_i| - |\bar{W} - \mathbf{E}\left[W\right]| \leq |\bar{W} - W_i| \leq |\mathbf{E}\left[W\right] - W_i| + |\bar{W} - \mathbf{E}\left[W\right]|.$$

---

[4]This case can happen because our sampling oracle is approximate.

Note that $\bar{E} = \frac{1}{T}\sum_{i=1}^{T}|\bar{W} - W_i|$. We have

$$\left(\frac{1}{T}\sum_{i=1}^{T}|\mathbf{E}\,[W] - W_i|\right) - |\bar{W} - \mathbf{E}\,[W]\,| \le \bar{E} \le \left(\frac{1}{T}\sum_{i=1}^{T}|\mathbf{E}\,[W] - W_i|\right) + |\bar{W} - \mathbf{E}\,[W]\,|. \quad (26)$$

By definition of $\bar{W}$, we have $\mathbf{E}\,[\bar{W}] = \mathbf{E}\,[W]$ and $\mathbf{Var}\,[\bar{W}] = \frac{\mathbf{Var}[W]}{T}$. By Chebyshev's inequality,

$$\mathbf{Pr}\left[|\bar{W} - \mathbf{E}\,[W]\,| \ge \frac{\varepsilon d}{10L}\right] \le \frac{100L^2 \mathbf{Var}\,[\bar{W}]}{\varepsilon^2 d^2} = \frac{100L^2 \mathbf{Var}\,[W]}{T\varepsilon^2 d^2} \le \frac{100L^2 K^2 d^2}{T\varepsilon^2 d^2} \le 0.01, \quad (27)$$

where the second inequality follows from Condition 21. Next, consider the random variable

$$R \triangleq |\mathbf{E}\,[W] - W|.$$

By the definition of $R$ and the variance bound in Condition 21, we know that

$$\mathbf{Var}\,[R] \le \mathbf{E}\,[R^2] = \mathbf{E}\left[(\mathbf{E}\,[W] - W)^2\right] = \mathbf{Var}\,[W] \le K^2 d^2.$$

Note that $\frac{1}{T}\sum_{i=1}^{T}|\mathbf{E}\,[W] - W_i|$ is the average of $T$ i.i.d. random samples of $R$. Denote $\bar{R} \triangleq \frac{1}{T}\sum_{i=1}^{T}|\mathbf{E}\,[W] - W_i|$. Note that $\mathbf{Var}\,[\bar{R}] = \frac{\mathbf{Var}[R]}{T}$. By Chebyshev's inequality, we have

$$\mathbf{Pr}\left[|\bar{R} - \mathbf{E}\,[R]| \ge \frac{\varepsilon d}{10L}\right] \le \frac{100L^2 \mathbf{Var}\,[R]}{T\varepsilon^2 d^2} \le \frac{100L^2 K^2 d^2}{T\varepsilon^2 d^2} \le 0.01. \quad (28)$$

Combining (26), (27), (28), and a union bound, we have

$$\mathbf{Pr}\left[|\bar{E} - \mathbf{E}\,[R]\,| \le \frac{\varepsilon d}{5L}\right] \ge 0.98. \quad (29)$$

Assume two good events in (29) and (25) both hold, which happens with probability at least 0.96. The final output $\hat{d} = \frac{\hat{Z}_\mu}{2\hat{Z}_\nu}\bar{E}$ satisfies

$$\hat{d} = \frac{\hat{Z}_\mu}{2\hat{Z}_\nu}\bar{E} \le \frac{(1 + 3\varepsilon/4)Z_\mu}{2Z_\nu} \cdot \left(\mathbf{E}\,[R] + \frac{\varepsilon d}{5L}\right)$$
$$\le \left(1 + \frac{3\varepsilon}{4}\right)\frac{Z_\mu}{2Z_\nu}\mathbf{E}\,[R] + \frac{Z_\mu}{Z_\nu L} \cdot \frac{\varepsilon}{10} \cdot \left(1 + \frac{3\varepsilon}{4}\right) \cdot d.$$

By (24), we have $\frac{Z_\mu}{2Z_\nu}\mathbf{E}\,[R] = d$. By Condition 21, we have $\frac{Z_\mu}{Z_\nu L} = \frac{1}{L\mathbf{E}[W]} \le 1$. Therefore,

$$\hat{d} \le \left(1 + \frac{3\varepsilon}{4}\right)d + \frac{\varepsilon}{10} \cdot \left(1 + \frac{3\varepsilon}{4}\right) \cdot d < (1 + \varepsilon)d.$$

A similar argument gives a lower bound $\hat{d} \ge (1 - \varepsilon)d$. $\qquad\square$

## 5.2 Advanced algorithm for hardcore model

We give the following algorithm for hardcore models with small parameter distance.

**Theorem 23.** *Let $\theta = 10^{-10}\frac{\varepsilon^{1/4}}{n^{5/2}}$. There exists an algorithm such that given two hardcore models $\mu$ and $\nu$ on the same graph $G = (V, E)$, and any $0 < \varepsilon < 1$, if $\mu$ and $\nu$ both satisfy uniqueness condition in (1) and $d_{\mathrm{par}}(\mu, \nu) < \theta$, then it returns a random number $\hat{d}$ in time $\tilde{O}\left(\frac{n^7}{\varepsilon^2} + \frac{n^{6.5}}{\varepsilon^{9/4}}\right)$ such that $(1 - \varepsilon)d_{\mathrm{TV}}\,(\mu, \nu) \le \hat{d} \le (1 + \varepsilon)d_{\mathrm{TV}}\,(\mu, \nu)$ with probability at least $2/3$.*

Let $\lambda^\mu$ and $\lambda^\nu$ be external fields of two hardcore models $\mu$ and $\nu$, respectively. For simplicity of the notation, we denote

$$D = d_{\mathrm{par}}(\mu, \nu) = \|\lambda^\mu - \lambda^\nu\|_\infty < \theta = 10^{-10} \frac{\varepsilon^{1/4}}{n^{5/2}},$$

$$d = d_{\mathrm{TV}}(\mu, \nu).$$

As discussed in the algorithm overview, we divide the vertices of $G$ into two parts: the "big" vertices and the "small" vertices. Define the threshold parameter

$$\kappa \triangleq 10^{-9} \frac{\varepsilon^{1/4}}{n^{3/2}} = \Theta\left(\frac{\varepsilon^{1/4}}{n^{3/2}}\right).$$

Define two sets of vertices $B$ and $S$ in graph $G$:

$$B = \{v \in V \mid \min\{\lambda_v^\mu, \lambda_v^\nu\} \geq \kappa\},$$
$$S = V \setminus B = \{v \in V \mid \min\{\lambda_v^\mu, \lambda_v^\nu\} < \kappa\}.$$

Recall $\mu_B$ and $\nu_B$ are the marginal distributions of $\mu$ and $\nu$ on $B$, respectively. Let $\Omega_B \subseteq \{\pm\}^B$ be the support of both $\mu_B$ and $\nu_B$. By the definition of $B$, for any $x \in \Omega_B$, all vertices $v \in B$ with $x_v = +1$ forms an independent set in $G$. For any $x \in \Omega_B$, let $\mu_S^x$ and $\nu_S^x$ be the marginal distributions of $\mu$ and $\nu$ on $S$ conditioned on $x$. The TV-distance between $\mu$ and $\nu$ can be represented by

$$d = \frac{1}{2} \sum_{\sigma \in \{\pm\}^V} |\mu(\sigma) - \nu(\sigma)| = \frac{1}{2} \sum_{x \in \Omega_B} \sum_{y \in \{\pm\}^S} |\mu_B(x)\mu_S^x(y) - \nu_B(x)\nu_S^x(y)|$$
$$= \frac{1}{2} \sum_{x \in \Omega_B} \mu_B(x) \sum_{y \in \{\pm\}^S} \left| \frac{\nu_B(x)}{\mu_B(x)} \nu_S^x(y) - \mu_S^x(y) \right|. \tag{30}$$

Define the function $f : \Omega_B \to \mathbb{R}$ as

$$f(x) \triangleq \frac{1}{2} \sum_{y \in \{\pm\}^S} \left| \frac{\nu_B(x)}{\mu_B(x)} \nu_S^x(y) - \mu_S^x(y) \right|. \tag{31}$$

The calculation shows that $d_{\mathrm{TV}}(\mu, \nu) = \mathbf{E}_{x \sim \mu_B}[f(x)]$. In a high-level view, our algorithm wants to draw i.i.d. samples $x \sim \mu_B$ and compute values $f(x)$ and then output the average value. Formally, we have the following lemmas. Let $n$ denote the number of vertices in $G$.

**Lemma 24.** *The variance* $\mathbf{Var}_{x \sim \mu_B}[f(x)] = O_\eta(d^2) \cdot (n^3 + n/\kappa)$, *where* $O_\eta$ *holds a constant depending only on the gap* $\eta$ *in the uniqueness condition in* (1).

**Lemma 25.** *There exists a randomized data structure satisfies that*

- *the data structure can be constructed in time* $\tilde{O}(\frac{n^7}{\varepsilon^2} + \frac{n^{6.5}}{\varepsilon^{9/4}})$ *and the construction succeeds with probability at least 0.99;*

- *if the data structure is constructed successfully, then given any* $x \in \Omega_B$, *it deterministically answers an* $\hat{f}(x) \geq 0$ *in time* $O(n^4)$ *such that*

$$|\hat{f}(x) - f(x)| \leq \frac{\varepsilon}{50} \cdot d.$$

Lemma 24 can be used to control the variance of $f(x)$. Lemma 25 is the main technical part of our algorithm. We first assume both Lemma 24 and Lemma 25 hold and prove Theorem 23. The proofs of Lemma 24 and Lemma 25 are given in Section 5.2.1 and 5.2.2, respectively.

*Proof of Theorem 23.* Let $T = O(\frac{n^3 + n/\kappa}{\varepsilon^2}) = O(\frac{n^3}{\varepsilon^2} + \frac{n^{5/2}}{\varepsilon^{9/4}})$ be large enough.

---

**The algorithm for hardcore model**

- Construct the data structure in Lemma 25;

- Draw $T$ independent approximate samples $x_1, x_2, \ldots, x_T$ from the marginal distribution $\mu_B$ with $d_{\mathrm{TV}}(\mu_B, x_i) \leq \frac{1}{100T}$.

- Use the data structure to compute $\hat{f}(x_1), \hat{f}(x_2), \ldots, \hat{f}(x_T)$.

- Return $\hat{d} = \frac{1}{T} \sum_{i=1}^{T} \hat{f}(x_i)$.

---

Consider an ideal algorithm $\mathcal{A}^*$ that draw perfect samples $x_1, \ldots, x_T$ and exactly compute the values of $f(x_1), \ldots, f(x_T)$. Let $d^*$ denote the output $\frac{1}{T} \sum_{i=1}^{T} f(x_i)$. We have $\mathbf{E}[d^*] = d$ and the variance of $d^*$ is $\frac{\mathbf{Var}_{\mu_B}[f]}{T}$. By our choice of $T$ and the variance bound in Lemma 24, using Chebyshev's inequality, we have

$$\mathbf{Pr}\left[|d^* - d| \geq \frac{\varepsilon}{10}d\right] \leq 0.01.$$

Consider our real algorithm $\mathcal{A}$. All the approximate samples $x_1, \ldots, x_T$ in $\mathcal{A}$ can be coupled successfully with perfect samples with probability at least $1 - T \cdot \frac{1}{100T} = 0.99$. Also that the data structure in $\mathcal{A}$ is constructed successfully, which happen with probability at least 0.99. By Lemma 25, there exists a coupling of $\mathcal{A}$ and $\mathcal{A}^*$ such that with probability at least 0.98, $|\hat{d} - d^*| \leq \frac{\varepsilon}{50}d$. Hence, using a union bound, we have

$$\mathbf{Pr}\left[(1 - \varepsilon)d \leq \hat{d} \leq (1 + \varepsilon)d\right] \geq 0.97 > \frac{2}{3}.$$

The total running time is bounded by

$$\tilde{O}\left(\frac{n^7}{\varepsilon^2} + \frac{n^{6.5}}{\varepsilon^{9/4}}\right) + O\left(n^4 \cdot T\right) = \tilde{O}\left(\frac{n^7}{\varepsilon^2} + \frac{n^{6.5}}{\varepsilon^{9/4}}\right). \qquad \square$$

### 5.2.1 Analyze the variance of the estimator (Proof of Lemma 24)

Before we prove the lemma, we first remark that one can show for any $x \in \Omega_B$, $|f(x) - 1| \leq O(\frac{n}{\kappa}d)$, which implies the $O(\frac{n^2}{\kappa^2}d^2)$ variance bound. This bound gives a polynomial-time algorithm but the degree of polynomial is higher. If we use this bound, then in the proof of Theorem 23, it requires $O(\frac{n^2}{\kappa^2\varepsilon^2})$ samples $x \sim \mu_B$ and Lemma 25 computes each value of $\hat{f}(x)$ in a super-linear time. Alternatively, we give a more technical analysis to achieve a better dependency on $\kappa$, which gives a better running time of our algorithm.

We bound the variance of $f(x)$ by bounding the second moment of $\mathbf{E}_{\mu_B}[f^2] = \mathbf{E}_{x \sim \mu_B}[f^2(x)]$.

We first need the following upper bound on the value of $f(x)$:

$$f(x) = \frac{1}{2} \sum_{y \in \{\pm\}^S} \left| \frac{\nu_B(x)}{\mu_B(x)} \nu_S^x(y) - \mu_S^x(y) \right|$$

(by triangle inequality) $\quad \leq \frac{1}{2} \left| \frac{\nu_B(x)}{\mu_B(x)} - 1 \right| \sum_{y \in \{\pm\}^S} \nu_S^x(y) + \frac{1}{2} \sum_{y \in \{\pm\}^S} |\nu_S^x(y) - \mu_S^x(y)|$

$$= \frac{1}{2} \left| \frac{\nu_B(x)}{\mu_B(x)} - 1 \right| + d_{\mathrm{TV}}\left( \nu_S^x, \mu_S^x \right).$$

We have the following upper bound on the $d_{\mathrm{TV}}\left( \nu_S^x, \mu_S^x \right)$.

**Lemma 26.** *for any $x \in \Omega_B$, it holds that $d_{\mathrm{TV}}\left( \nu_S^x, \mu_S^x \right) \leq 4nD$.*

The proof of Lemma 26 will be given later. Assume Lemma 26 holds. Since $f(x) \geq 0$, we can upper bound

$$\mathbf{E}_{\mu_B}\left[ f^2 \right] \leq \frac{1}{4} \mathbf{E}_{x \sim \mu_B}\left[ \left( \frac{\nu_B(x)}{\mu_B(x)} - 1 \right)^2 \right] + 4nD \mathbf{E}_{x \sim \mu_B}\left[ \left| \frac{\nu_B(x)}{\mu_B(x)} - 1 \right| \right] + 16n^2 D^2$$

$$= \frac{1}{4} \mathbf{E}_{x \sim \mu_B}\left[ \left( \frac{\nu_B(x)}{\mu_B(x)} - 1 \right)^2 \right] + 8nD \cdot d_{\mathrm{TV}}\left( \nu_B, \mu_B \right) + 16n^2 D^2$$

$$\leq \frac{1}{4} \mathbf{E}_{x \sim \mu_B}\left[ \left( \frac{\nu_B(x)}{\mu_B(x)} - 1 \right)^2 \right] + 4 \cdot 10^8 n^2 d^2, \tag{32}$$

where the last inequality follows from the fact $d = d_{\mathrm{TV}}\left( \mu, \nu \right) > d_{\mathrm{TV}}\left( \mu_B, \nu_B \right)$ and $d \geq \frac{1}{5000} D$.

Define a function $h(x) = \frac{\nu_B(x)}{\mu_B(x)}$. We have $\mathbf{E}_{\mu_B}[h] = 1$. Our task is reduced to bound the variance $\mathbf{Var}_{\mu_B}[h] = \mathbf{Var}_{x \sim \mu_B}[h(x)]$. We will use the following Poincaré inequality (i.e. the spectral gap of the Glauber dynamics) for the marginal distribution $\mu_B$. For any subset $\Lambda \subseteq V$, recall that $\Omega_\Lambda$ denotes the support of $\mu_\Lambda$.

**Lemma 27** (CFYZ21). *Since the hardcore model $(G, \lambda^\mu)$ is in the uniqueness regime in (1) with constant gap $\eta > 0$. For any function $g : \Omega_B \to \mathbb{R}$, it holds that*

$$\mathbf{Var}_{\mu_B}[g] \leq C_\eta \sum_{v \in B} \sum_{\sigma \in \Omega_{B-v}} \mu_{B-v}(\sigma) \mathbf{Var}_{\mu_B^\sigma}[g],$$

*where $B - v$ denote the set $B \setminus \{v\}$ and $C_\eta$ is a constant depending on $\eta$.*

The Poincaré inequality in [CFYZ21] is stated from the entire Gibbs distribution $\mu$. One can lift it to marginal distribution $\mu_B$. For the completeness, we give a proof in Appendix B.

Fix a vertex $v \in B$ and a $\sigma \in \Omega_{B-v}$. Let $\sigma^{v+}$ denote a configuration in $\{\pm\}^B$ obtained by extending $\sigma$ further by setting $v$ to $+1$. Define $\sigma^{v-}$ similarly. By the definition of variance and the definition of the function $h = \frac{\nu_B}{\mu_B}$, we can write

$$\mathbf{Var}_{\mu_B^\sigma}[h] = \mu_v^\sigma(+1) \mu_v^\sigma(-1) (h(\sigma^{v+}) - h(\sigma^{v-}))^2$$

$$= \left( \frac{\nu_{B-v}(\sigma)}{\mu_{B-v}(\sigma)} \right)^2 \mu_v^\sigma(+1) \mu_v^\sigma(-1) \left( \frac{\nu_v^\sigma(+1)}{\mu_v^\sigma(+1)} - \frac{\nu_v^\sigma(-1)}{\mu_v^\sigma(-1)} \right)^2.$$

Note that $\mathbf{Var}_{\mu_B^\sigma}[h] = 0$ if either $\mu_v^\sigma(+1) = 0$. We assume that $\mu_v^\sigma(+1) > 0$, then for every neighbor $u$ of vertex $v$, it must hold that if $u \in B$, then $\sigma_u = -1$. We claim the following bounds.

**Claim 28.** *The following bounds hold*

- $\lambda_v^\mu/10 \leq \mu_v^\sigma(+1) \leq \lambda_v^\mu$;

- $|\mu_v^\sigma(+1) - \nu_v^\sigma(+1)| \leq 80nD\mu_v^\sigma(+1) + 4D$;

- $\frac{\nu_{B-v}(\sigma)}{\mu_{B-v}(\sigma)} \leq 2$.

Assume the above claim holds, which will be proved later. Let $p = \mu_v^\sigma(+1)$. By our choices of parameters, since $D \leq \theta$ is sufficient small, $\mu_v^\sigma(+1) \geq \frac{\kappa}{10} \geq 100nD\mu_v^\sigma(+1) + 4D$. We have

$$\left(\frac{\nu_v^\sigma(+1)}{\mu_v^\sigma(+1)} - \frac{\nu_v^\sigma(-1)}{\mu_v^\sigma(-1)}\right)^2 \leq \left(\frac{4D + 80nD\mu_v^\sigma(+1)}{p(1-p)}\right)^2.$$

Using Claim 28, the variance can be bounded by

$$\mathbf{Var}_{\mu_B^\sigma}[h] \leq 4p(1-p)\frac{(80nDp + 4D)^2}{p^2(1-p)^2} = O(D^2) \cdot \frac{(20np + 1)^2}{p(1-p)} \leq O(D^2) \cdot (n^2 p + n + 1/p)$$

$$\leq O(D^2) \cdot \left(n^2 + \frac{1}{\kappa}\right).$$

Using Lemma 27 and the above upper bound, we have

$$\mathbf{Var}_{\mu_B}[h] \leq C_\eta \sum_{v \in B} \sum_{\sigma \in \Omega_{B-v}} \mu_{B-v}(\sigma)\mathbf{Var}_{\mu_B^\sigma}[h]$$

$$\leq O_\eta(D^2) \sum_{v \in B} \left(n^2 + \frac{1}{\kappa}\right)$$

$$\text{(by Lemma 15 )} \quad = O_\eta(d^2) \cdot \left(n^3 + \frac{n}{\kappa}\right). \tag{33}$$

Hence, using (32), the variance of $f(x)$ is at most

$$\mathbf{Var}_{\mu_B}[f] \leq \mathbf{E}_{\mu_B}[f^2] \leq \mathbf{Var}_{\mu_B}[h] + O_\eta(n^2 d^2) = O_\eta(d^2) \cdot \left(n^3 + \frac{n}{\kappa}\right).$$

**Proofs of technical lemmas and claims** We first give two general lemmas (Lemma 29 and Lemma 30) that will be used in later proofs. We then prove all technical lemmas and claims appeared in the proof of Lemma 24.

Define the conditional partition function. Fix a configuration $x \in \{\pm\}^B$. Define $Z_{S,\mu}^x$ as the conditional partition function defined by

$$Z_{S,\mu}^x \triangleq \sum_{y \in \{\pm\}^S : \mu(x+y) > 0} \prod_{v \in S : y_v = +1} \lambda_v^\mu.$$

Intuitively, $Z_{S,\mu}^x$ is the total weights of $y \in \{\pm\}^S$ such that $x + y$ is a valid configuration (forms an independent set in $G$), where $x + y \in \{\pm\}^V$ is the concatenation of $x$ and $y$. Alternatively, $Z_{S,\mu}^x$ can be interpreted as follows. Let $N_G(v)$ denote the set of neighbors of $v$ in $G$. Given $x$, one can remove all vertices $v \in S$ from $S$ such that there exists $u \in N_G(v) \cap B$ with $x_u = +1$. Let $S^x \subseteq S$ denote the set of remaining vertices:

$$S^x = S \setminus \{v \in S \mid \exists u \in N_G(v) \cap B \text{ s.t. } x_u = +1\}. \tag{34}$$

Then $Z_{S,\mu}^x$ is the partition function for the hardcore model in induced subgraph $G[S^x]$. The following property of the conditional partition function will be used in our proofs.

**Lemma 29.** *Suppose $\kappa + \theta < 1/(10n)$. For any $x \in \{\pm\}^B$, it holds that*

- $1 \le Z^x_{S,\mu}, Z^x_{S,\nu} < 2$;

- $|Z^x_{S,\mu} - Z^x_{S,\nu}| \le 2nD$.

*Proof.* Since the empty set contributes the weight 1 to both $Z^x_{S,\mu}$, so $Z^x_{S,\mu} \ge 1$. For the upper bound, Let $\lambda_{\max} = \max_{v \in S} \max\{\lambda^\mu_v, \lambda^\nu_v\}$. By the definition of $S$, we have $\lambda_{\max} \le \kappa + D < \kappa + \theta < \frac{1}{10n}$. Hence $Z^x_{S,\mu} \le (1 + 1/(10n))^n < 2$. The same bound holds for $Z^x_{S,\nu}$.

For any independent set $I$ in graph $G[S^x]$, the difference of the weights is $\left| \prod_{v \in I} \lambda^\nu_v - \prod_{v \in I} \lambda^\mu_v \right|$. We first show the difference of the weights of $I$ is at most $(\lambda_{\max} + D)^{|I|} - \lambda^{|I|}_{\max}$. Assume $\lambda^\nu_v = \lambda^\mu_v + \delta_v$, where $|\delta_v| \le D$. Then

$$
\begin{aligned}
\left| \prod_{v \in I} \lambda^\nu_v - \prod_{v \in I} \lambda^\mu_v \right| &= \left| \prod_{v \in I} (\lambda^\mu_v + \delta_v) - \prod_{v \in I} \lambda^\mu_v \right| = \left| \sum_{A \subseteq I : A \ne \varnothing} \prod_{v \in A} \delta_v \prod_{u \in I \setminus A} \lambda^\mu_u \right| \\
&\le \sum_{A \subseteq I : A \ne \varnothing} \prod_{v \in A} D \prod_{u \in I \setminus A} \lambda_{\max} \\
&= (\lambda_{\max} + D)^{|I|} - \lambda^{|I|}_{\max}.
\end{aligned}
\tag{35}
$$

The number of independent sets of size $k$ in $G[S^x]$ is at most $\binom{n}{k}$, where $n$ is the number of vertices in $G$. We have

$$
\begin{aligned}
|Z^x_{S,\nu} - Z^x_{S,\mu}| &\le \sum_{k=0}^n \binom{n}{k} (\lambda_{\max} + D)^k - \sum_{k=0}^n \binom{n}{k} \lambda^k_{\max} \\
&= (1 + \lambda_{\max} + D)^n - (1 + \lambda_{\max})^n \\
&= (1 + \lambda_{\max})^n \left( \left( 1 + \frac{D}{1 + \lambda_{\max}} \right)^n - 1 \right) \\
\text{by } (\lambda_{\max}, D < 1/(10n)) \quad &\le 2nD.
\end{aligned}
\tag{36}
$$

This proves the lemma. $\qquad \square$

The second general lemma we will use is the following bound on the marginal ratio.

**Lemma 30.** *Suppose $\kappa + \theta < 1/(10n)$ and $\theta/\kappa < 1/(10n)$. For any $x \in \Omega_B$, it holds that*

$$
\left| \frac{\nu_B(x)}{\mu_B(x)} - 1 \right| \le \frac{10nD}{\kappa}.
$$

*Proof.* Define $g(x) = \frac{\nu_B(x)}{\mu_B(x)}$. For $x \in \Omega_B \subseteq \{\pm\}^B$ and $y \in \{\pm\}^S$, we use $x + y$ to denote a full configuration in $\{\pm\}^V$ obtained by concatenating $x$ and $y$. We have

$$
\mu_B(x) = \sum_{y \in \{\pm\}^S} \mu(x + y) = \frac{\prod_{v \in B : x_v = 1} \lambda^\mu_v \cdot Z^x_{S,\mu}}{Z_\mu}, \text{ and}
$$

$$
\nu_B(x) = \sum_{y \in \{\pm\}^S} \nu(x + y) = \frac{\prod_{v \in B : x_v = 1} \lambda^\nu_v \cdot Z^x_{S,\nu}}{Z_\nu}.
$$

Then we have

$$
g(x) = \frac{\nu_B(x)}{\mu_B(x)} = \left( \prod_{v \in B : x_v = 1} \frac{\lambda^\nu_v}{\lambda^\mu_v} \right) \cdot \frac{Z^x_{S,\nu}}{Z^x_{S,\mu}} \cdot \frac{Z_\mu}{Z_\nu}.
$$

Denote $\alpha = \prod_{v \in B : x_v = 1} \frac{\lambda_v^\nu}{\lambda_v^\mu}$, $\beta = \frac{Z_{S,\nu}^x}{Z_{S,\mu}^x}$ and $\gamma = \frac{Z_\mu}{Z_\nu}$. We analyze each term one by one.

We have $D/\kappa \leq \theta/\kappa \leq 1/(10n)$ and

$$\alpha \leq \prod_{v \in B : x_v = 1} \frac{\lambda_v^\mu + D}{\lambda_v^\mu} \leq \left(1 + \frac{D}{\kappa}\right)^{|B|} \leq \left(1 + \frac{D}{\kappa}\right)^n \leq 1 + \frac{2nD}{\kappa}.$$

A similar argument gives a lower bound $\alpha \geq 1 - \frac{2nD}{\kappa}$.

For the second term $\beta$, using Lemma 29, we have

$$|\beta - 1| = \left| \frac{Z_{S,\nu}^x}{Z_{S,\mu}^x} - 1 \right| = \frac{|Z_{S,\nu}^x - Z_{S,\mu}^x|}{Z_{S,\mu}^x} \leq |Z_{S,\nu}^x - Z_{S,\mu}^x| \leq 2nD.$$

Finally, for $\gamma = \frac{Z_\mu}{Z_\nu}$, we have the following equivalent form

$$\gamma = \frac{Z_\mu}{Z_\nu} = \frac{\sum_{x \in \Omega_B} \left( \prod_{v \in B : x_v = +1} \lambda_v^\mu \cdot Z_{S,\mu}^x \right)}{\sum_{x \in \Omega_B} \left( \prod_{v \in B : x_v = +1} \lambda_v^\nu \cdot Z_{S,\nu}^x \right)}.$$

Using the bound for $\alpha$ and $\beta$, since $\theta/\kappa < 1/(10n)$ and $\theta + \kappa < 1/(10n)$, we have

$$1 - \frac{4nD}{\kappa} < \gamma \leq \left(1 + \frac{2nD}{\kappa}\right)(1 + 2nD) < 1 + \frac{4nD}{\kappa}. \tag{37}$$

Combining the bounds for $\alpha$, $\beta$ and $\gamma$, since $D/\kappa \leq \theta/\kappa < 1/(10n)$, we have

$$1 - \frac{10nD}{\kappa} < g(x) \leq \left(1 + \frac{2nD}{\kappa}\right)(1 + 2nD)\left(1 + \frac{4nD}{\kappa}\right) < 1 + \frac{10nD}{\kappa}. \qquad \square$$

*Proof of Lemma 26.* For any $y \in \{\pm\}^S$ such that $y + x$ forms an independent set, we can bound

$$|\nu_S^x(y) - \mu_S^x(y)| \leq \left| \frac{\prod_{v \in S : y_v = +1} \lambda_v^\nu}{Z_{S,\nu}^x} - \frac{\prod_{v \in S : y_v = +1} \lambda_v^\mu}{Z_{S,\mu}^x} \right|$$

$$(\text{by } Z_{S,\mu}^x, Z_{S,\nu}^x \geq 1) \quad \leq \left| Z_{S,\mu}^x \prod_{v \in S : y_v = +1} \lambda_v^\nu - Z_{S,\nu}^x \prod_{v \in S : y_v = +1} \lambda_v^\mu \right|$$

$$(\text{triangle ineq.}) \quad \leq Z_{S,\mu}^x \left| \prod_{v \in S : y_v = +1} \lambda_v^\nu - \prod_{v \in S : y_v = +1} \lambda_v^\mu \right| + |Z_{S,\mu}^x - Z_{S,\nu}^x| \prod_{v \in S : y_v = +1} \lambda_v^\mu. \tag{38}$$

We bound each term separately. Recall that $\lambda_{\max} = \max_{v \in S} \max\{\lambda_v^\mu, \lambda_v^\nu\} < \kappa + \theta$. By Lemma 29, we have $Z_{S,\mu}^x < 2$. Using (35), we have

$$Z_{S,\mu}^x \left| \prod_{v \in S : y_v = +1} \lambda_v^\nu - \prod_{v \in S : y_v = +1} \lambda_v^\mu \right| \leq 2\left((\lambda_{\max} + D)^{\|y\|_+} - \lambda_{\max}^{\|y\|_+}\right), \tag{39}$$

where $\|y\|_+$ is the number of $+1$s in $y$. For the second term, using Lemma 29, we have

$$\prod_{v \in S : y_v = +1} \lambda_v^\mu \left| Z_{S,\mu}^x - Z_{S,\nu}^x \right| \leq 2nD\lambda_{\max}^{\|y\|_+}. \tag{40}$$

The number of independent sets of size $k$ is at most $\binom{n}{k}$. We have

$$
\begin{aligned}
d_{\mathrm{TV}}\left(\nu_S^x, \mu_S^x\right) &= \frac{1}{2} \sum_{y \in \{\pm\}^S} |\nu_S^x(y) - \mu_S^x(y)| \\
&\leq \sum_{k=0}^n \binom{n}{k}\left((\lambda_{\max} + D)^k - \lambda_{\max}^k\right) + nD \sum_{k=0}^n \binom{n}{k} \lambda_{\max}^k
\end{aligned}
$$

(by the same calculation as (36)) $\quad \leq 2nD + nD(1 + \lambda_{\max})^n$

(by $\lambda_{\max} < 1/(10n)$) $\quad < 4nD.$

This proves the upper bound on TV distance. $\qquad\square$

*Proof of Claim 28.* Recall that $\sigma$ is a configuration in $\Omega_{B-v}$, and $\sigma^{v+}, \sigma^{v-}$ are configurations in $\{\pm\}^B$. For simplicity of the notation, we use $Z_\mu^+$ to denote $Z_{S,\mu}^{\sigma^{v+}}$ and $Z_\mu^-$ to denote $Z_{S,\mu}^{\sigma^{v-}}$. Similarly, we define $Z_\nu^+, Z_\nu^-$. We have

$$
\begin{aligned}
|\mu_v^\sigma(+1) - \nu_v^\sigma(+1)| &= \left| \frac{\lambda_v^\mu Z_\mu^+}{Z_\mu^- + \lambda_v^\mu Z_\mu^+} - \frac{\lambda_v^\nu Z_\nu^+}{Z_\nu^- + \lambda_v^\nu Z_\nu^+} \right| = \left| \frac{\lambda_v^\mu Z_\mu^+ Z_\nu^- - \lambda_v^\nu Z_\nu^+ Z_\mu^-}{(Z_\mu^- + \lambda_v^\mu Z_\mu^+)(Z_\nu^- + \lambda_v^\nu Z_\nu^+)} \right| \\
&\leq |\lambda_v^\mu Z_\mu^+ Z_\nu^- - \lambda_v^\nu Z_\nu^+ Z_\mu^-| \\
&\leq Z_\nu^+ Z_\mu^- |\lambda_v^\mu - \lambda_v^\nu| + \lambda_v^\mu |Z_\mu^+ Z_\nu^- - Z_\nu^+ Z_\mu^-|.
\end{aligned}
$$

Using Lemma 29, we have $Z_\nu^+ Z_\mu^- |\lambda_v^\mu - \lambda_v^\nu| \leq 4D$. We also know that $|Z_\mu^+ - Z_\nu^+| \leq 2nD$ and $|Z_\mu^- - Z_\nu^-| \leq 2nD$. Therefore, by using the triangle inequality, we have

$$
|Z_\mu^+ Z_\nu^- - Z_\nu^+ Z_\mu^-| \leq Z_\mu^+ |Z_\nu^- - Z_\mu^-| + Z_\mu^- |Z_\mu^+ - Z_\nu^+| \leq 8nD.
$$

Therefore, we have

$$
|\mu_v^\sigma(+1) - \nu_v^\sigma(+1)| \leq 4D + 8nD\lambda_v^\mu.
$$

Next, we bound the value of $\mu_v^\sigma(+1)$. To sample from the distribution $\mu_v^\sigma$, one can first sample all the neighbors $u \in N_G(v) \setminus B$ of $v$, then sample the value of $v$ further conditional on the configuration of $N_G(v)$. Suppose with probability $q$, all vertices $u \in N_G(v) \setminus B$ are sampled to be $-1$. Note that $q \geq (\frac{1}{1+\lambda_{\max}})^n \geq (\frac{1}{1+\kappa+D})^n$. Since $\kappa + \theta < 1/(4n)$ and $D < \theta$, we have $q \geq \frac{1}{2}$. Conditional on this event, $v$ takes value $+1$ with probability $\frac{\lambda_v^\mu}{1+\lambda_v^\mu} \geq \frac{\lambda_v^\mu}{5}$, where we use the fact that $\lambda_v^\mu \leq \lambda_c(\Delta) \leq 4$. We have the lower bound $\mu_v^\sigma(+1) \geq \frac{1}{10}\lambda_v^\mu$. On the other hand, we have the upper bound $\mu_v^\sigma(+1) \leq \frac{\lambda_v^\mu}{1+\lambda_v^\mu} \leq \lambda_v^\mu$. Combining together, we have

$$
\frac{\lambda_v^\mu}{10} \leq \mu_v^\sigma(+1) \leq \lambda_v^\mu.
$$

The above bound implies

$$
|\mu_v^\sigma(+1) - \nu_v^\sigma(+1)| \leq 4D + 80nD\mu_v^\sigma(+1).
$$

For the last bound, using Lemma 30, we have

$$
\frac{\nu_{B-v}(\sigma)}{\mu_{B-v}(\sigma)} = \frac{\nu_B(\sigma^{v+}) + \nu_B(\sigma^{v-})}{\mu_B(\sigma^{v+}) + \mu_B(\sigma^{v-})} \leq 1 + \frac{10nD}{\kappa} < 2,
$$

where the last inequality follows from the fact that $D/\kappa < 1/(10n)$. $\qquad\square$

### 5.2.2 Approximate the value of the estimator (Proof of Lemma 25)

Define $\|y\|_+$ be the number of $+1$s in $y$. Let $t \geq 0$ be an integer. The specific choice of $t$ will be fixed later. Define the truncated function $f_t$ of $f$ defined by

$$f_t(x) \triangleq \frac{1}{2} \sum_{y \in \{\pm\}^S : \|y\|_+ \leq t} \left| \frac{\nu_B(x)}{\mu_B(x)} \nu_S^x(y) - \mu_S^x(y) \right|.$$

Compared to $f$ in (31), $f_t$ only includes the size at most $t$ independent sets of the induced subgraph $G[S^x]$, where $S^x$ is defined in (34). We have the following relation between $f_t$ and $f$.

**Lemma 31.** *Suppose $\kappa + \theta < 1/(10n)$ and $\theta/\kappa < 1/(10n)$. For any integer $t \geq 0$, any $x \in \Omega_B$,*

$$0 \leq f(x) - f_t(x) \leq 10^6 \left(1 + \frac{n}{10}\right)^{t+1} \kappa^t n^{t+2} \cdot d \triangleq \eta(\kappa, t) \cdot d. \tag{41}$$

*Proof.* First, by definition, $f(x) - f_t(x) \geq 0$. Recall that $g(x) = \frac{\nu_B(x)}{\mu_B(x)}$. Let $\Omega_S^x \subseteq \{\pm\}^S$ be the set over all $y$ such that $x + y$ forms an independent set in $G$. We have

$$f(x) - f_t(x) = \frac{1}{2} \sum_{y \in \Omega_S^x : \|y\|_+ \geq t+1} |g(x)\nu_S^x(y) - \mu_S^x(y)|$$

$$\text{(by Lemma 30)} \quad \leq \frac{1}{2} \sum_{y \in \Omega_S^x : \|y\|_+ \geq t+1} |\nu_S^x(y) - \mu_S^x(y)| + \frac{5nD}{\kappa} \cdot \sum_{y \in \Omega_S^x : \|y\|_+ \geq t+1} \nu_S^x(y)$$

Recall that $\lambda_{\max} = \max_{v \in S} \max\{\lambda_v^\mu, \lambda_v^\nu\}$. Using the result in (38), (39) and (40), we have

$$\sum_{y \in \Omega_S^x : \|y\|_+ \geq t+1} |\nu_S^x(y) - \mu_S^x(y)| \leq \sum_{k=t+1}^{n} \binom{n}{k} \left((\lambda_{\max} + D)^k - \lambda_{\max}^k\right) + nD \sum_{k=t+1}^{n} \binom{n}{k} \lambda_{\max}^k.$$

Let $\phi = \kappa + D$. Then $\lambda_{\max} \leq \phi$ and $(\lambda_{\max} + D)^k - \lambda_{\max}^k \leq (\phi + D)^k - \phi^k = \phi^k((1 + D/\phi)^k - 1)$. Note that $D/\phi < \theta/\kappa < 1/(4n)$. The last term is at most $\phi^k \cdot (2kD)/\phi \leq \phi^{k-1} \cdot (2nD)$. Then

$$f(x) - f_t(x) \leq \frac{nD}{\kappa} \sum_{k=t+1}^{n} \binom{n}{k} \phi^k + \frac{nD}{2} \sum_{k=t+1}^{n} \binom{n}{k} \phi^k + \frac{5nD}{\kappa} \sum_{k=t+1}^{n} \binom{n}{k} \phi^k \leq \frac{8nD}{\kappa} \sum_{k=t+1}^{n} \binom{n}{k} \phi^k$$

$$\leq \frac{8nD}{\kappa} \sum_{k \geq t+1} (n\phi)^k$$

Note that $\phi < \kappa + \theta < 1/(4n)$ and $\phi n < 1/10$. Also note that $\phi \leq (1 + 1/n)\kappa$. The last term is at most $\frac{10nD}{\kappa} \left(1 + 1/(10n)\right)^{t+1} \kappa^{t+1} n^{t+1}$. The lemma holds by using $D \leq 5000d$. $\square$

Next, we give our algorithm to approximate $f_t(x)$. We can expand $f_t$ as follows

$$f_t(x) \triangleq \frac{1}{2} \sum_{y \in \{\pm\}^S : \|y\|_+ \leq t} \left| \frac{\nu_B(x)}{\mu_B(x)} \nu_S^x(y) - \mu_S^x(y) \right|$$

$$= \frac{1}{2} \sum_{y \in \{\pm\}^S : \|y\|_+ \leq t} \left| \left( \prod_{v \in B : x_v = 1} \frac{\lambda_v^\nu}{\lambda_v^\mu} \right) \cdot \frac{Z_{S,\nu}^x}{Z_{S,\mu}^x} \cdot \frac{Z_\mu}{Z_\nu} \cdot \nu_S^x(y) - \mu_S^x(y) \right|. \tag{42}$$

We now introduce the approximation of $Z_{S,\nu}^x, Z_{S,\mu}^x, \mu_S^x, \nu_S^x$ in the above formula. Let $\Omega^x \subseteq \{\pm\}^S$ be the set of all $y \in \{\pm\}^S$ such that $x + y$ forms an independent set in graph $G$ For

30

any integer $t \geq 0$, let $\Omega_t^x$ be all $y \in \Omega^x$ such that $\|y\|_+ \leq t$. Define $\tilde{\mu}_{S,t}^x$ as the distribution $\mu_S^x$ restricted on $\Omega_t^x$. Formally,

$$\forall y \in \Omega_t^x, \quad \tilde{\mu}_{S,t}^x(y) = \frac{\prod_{v \in S: y_v = +1} \lambda_v^\mu}{\tilde{Z}_{S,\mu}^x(t)} \quad \text{and} \quad \tilde{Z}_{S,\mu}^x(t) = \sum_{\tau \in \Omega_t^x} \prod_{v \in S: \tau_v = +1} \lambda_v^\mu. \quad (43)$$

Similarly, we can define $\tilde{\nu}_{S,t}^x$ and $\tilde{Z}_{S,\nu}^x(t)$ from $\nu_S^x$. The following approximation lemmas hold.

**Lemma 32.** *Suppose $\kappa + \theta < 1/(10n)$ and $\theta/\kappa < 1/(10n)$. For any integer $t \geq 0$ and any $x \in \Omega_B$, it holds that*

$$\left| \frac{Z_{S,\nu}^x}{Z_{S,\mu}^x} - \frac{\tilde{Z}_{S,\nu}^x(t)}{\tilde{Z}_{S,\mu}^x(t)} \right| \leq \eta(\kappa, t) \cdot d \quad \text{and} \quad \left| \frac{Z_{S,\mu}^x}{Z_{S,\nu}^x} - \frac{\tilde{Z}_{S,\mu}^x(t)}{\tilde{Z}_{S,\nu}^x(t)} \right| \leq \eta(\kappa, t) \cdot d,$$

*where $\eta(\kappa, t)$ is defined in* (41).

*Proof.* Due to the symmetry, we only consider the first term in our proof. Define $\Omega_{>t}^x = \Omega^x \setminus \Omega_t^x$. Because $Z_{S,\mu}^x, \tilde{Z}_{S,\mu}^x(t) \geq 1$, similar to the proof of Lemma 31, we have

$$\left| \frac{Z_{S,\nu}^x}{Z_{S,\mu}^x} - \frac{\tilde{Z}_{S,\nu}^x(t)}{\tilde{Z}_{S,\mu}^x(t)} \right| \leq \left| \tilde{Z}_{S,\mu}^x(t) Z_{S,\nu}^x - Z_{S,\mu}^x \tilde{Z}_{S,\nu}^x(t) \right|$$

$$= \left| \sum_{\tau \in \Omega_t^x} \prod_{v \in S: \tau_v = +1} \lambda_v^\mu \sum_{\varphi \in \Omega^x} \prod_{u \in S: \varphi_u = +1} \lambda_u^\nu - \sum_{\tau \in \Omega_t^x} \prod_{v \in S: \tau_v = +1} \lambda_v^\nu \sum_{\varphi \in \Omega^x} \prod_{u \in S: \varphi_u = +1} \lambda_u^\mu \right|$$

$$= \left| \sum_{\tau \in \Omega_t^x} \sum_{\varphi \in \Omega_{>t}^x} \left( \prod_{v \in S: \tau_v = +1} \lambda_v^\mu \prod_{u \in S: \varphi_u = +1} \lambda_u^\nu - \prod_{v \in S: \tau_v = +1} \lambda_v^\nu \prod_{u \in S: \varphi_u = +1} \lambda_u^\mu \right) \right|$$

$$\leq \sum_{i=0}^t \sum_{j=t+1}^n \binom{n}{i} \binom{n}{j} \left( (\lambda_{\max} + D)^{i+j} - \lambda_{\max}^{i+j} \right).$$

In the proof of Lemma 31, we already show that $(\lambda_{\max} + D)^k - \lambda_{\max}^k \leq \phi^{k-1} \cdot (2kD)$, where $\phi = \kappa + \theta$. Note that $\binom{n}{j} \leq n^j/(t+1)!$ and for any $k$, there are at most $(t+1)$ pairs $(i,j)$ s.t. $i + j = k$. We have

$$\left| \frac{Z_{S,\nu}^x}{Z_{S,\mu}^x} - \frac{\tilde{Z}_{S,\nu}^x(t)}{\tilde{Z}_{S,\mu}^x(t)} \right| \leq \frac{2(n+t)D}{\phi t!} \sum_{k=t+1}^{n+t} n^k \cdot \phi^k \leq \frac{4nD}{\kappa} \sum_{k \geq t+1} (n\phi)^t.$$

The last term is at most $\eta(\kappa, t) \cdot d$. $\square$

**Lemma 33.** *Suppose $\kappa + \theta < 1/(10n)$ and $\theta/\kappa < 1/(10n)$. For any integer $t \geq 0$, it holds that*

$$\sum_{y \in \Omega_t^x} |\tilde{\mu}_{S,t}^x(y) - \mu_S^x(y)| \leq \eta(\kappa, t) \cdot d \quad \text{and} \quad \sum_{y \in \Omega_t^x} |\tilde{\nu}_{S,t}^x(y) - \nu_S^x(y)| \leq \eta(\kappa, t) \cdot d.$$

*Proof.* Due to the symmetry of $\mu$ and $\nu$, we only prove the first term here. We know that

31

$$\tilde{Z}^x_{S,\mu}(t) \leq Z^x_{S,\mu},$$

$$
\sum_{y \in \Omega_t^x} |\tilde{\mu}^x_{S,t}(y) - \mu^x_S(y)| = \sum_{y \in \Omega_t^x} \left| \frac{\prod_{v \in S: y_v = 1} \lambda_v^\mu}{\tilde{Z}^x_{S,\mu}(t)} - \frac{\prod_{v \in S: y_v = 1} \lambda_v^\mu}{Z^x_{S,\mu}} \right| = \frac{Z^x_{S,\mu} - \tilde{Z}^x_{S,\mu}(t)}{\tilde{Z}^x_{S,\mu}(t) Z^x_{S,\mu}} \sum_{y \in \Omega_t^x} \prod_{v \in S: y_v = 1} \lambda_v^\mu
$$

$$
= \frac{Z^x_{S,\mu} - \tilde{Z}^x_{S,\mu}(t)}{\tilde{Z}^x_{S,\mu}(t) Z^x_{S,\mu}} \tilde{Z}^x_{S,\mu}(t) = \frac{Z^x_{S,\mu} - \tilde{Z}^x_{S,\mu}(t)}{Z^x_{S,\mu}} \leq Z^x_{S,\mu} - \tilde{Z}^x_{S,\mu}(t)
$$

$$
= \sum_{y \in \Omega_{>t}^x} \prod_{v \in S: y_v = 1} \lambda_v^\mu \leq \sum_{y \in \Omega_{>t}^x} \lambda_{\max}^{\|y\|_+} \leq \sum_{k=t+1}^{n} \binom{n}{k} \phi^k \leq \frac{1}{(t+1)!} \sum_{k \geq t+1} (n\phi)^k.
$$

The last term is at most $\eta(\kappa, t) \cdot d$. $\qquad\qquad\square$

Now, we are ready to give our algorithm. First consider the ratio $R = \frac{Z_\nu}{Z_\mu}$ in (42). We have the following algorithm to approximate $R$.

**Lemma 34.** *If $\eta(\kappa, t) \leq \frac{\varepsilon}{200}$, $\theta/\kappa < 1/(10n)$, and $\theta + \kappa < 1/(10n)$, where $\eta(\kappa, t)$ is defined in (41), then there exists a randomized algorithm that computes a random number $\tilde{R}$ in time $T' \cdot T_G^{\mathsf{sp}}(T'/10^3) + T' \cdot O(n^t)$, where $T' = O(\frac{n^3 + n/\kappa}{\varepsilon^2})$ and $T_G^{\mathsf{sp}}(\cdot)$ is the cost of sample oracle for $\mu$, such that with probability at least $0.99$,*

$$\left| \tilde{R} - \frac{Z_\nu}{Z_\mu} \right| \leq \frac{\varepsilon}{100} d. \tag{44}$$

We first prove Lemma 25 assuming Lemma 34. Lemma 34 will be proved later.

*Proof of Lemma 25.* We fix the parameter $t = 4$. Recall that $\kappa = 10^{-9} \frac{\varepsilon^{1/4}}{n^{3/2}}$ and $\theta = 10^{-10} \frac{\varepsilon^{1/4}}{n^{5/2}}$. We can verify that $\eta(\kappa, t) = 10^6 \left(1 + \frac{n}{10}\right)^{t+1} \kappa^t n^{t+2} \leq \frac{\varepsilon}{200}$, $\theta + \kappa < 1/(10n)$, and $\theta/\kappa < 1/(10n)$.

In the construction step, we use Lemma 34 to compute the random number $\tilde{R}$. We say the construction step succeeds if (44) is satisfied, which happens with probability at least $0.99$.

In the query step, given any $x \in \Omega_B$, our data structure answers the following $\hat{f}(x)$:

$$\hat{f}(x) = \frac{1}{2} \sum_{y \in \Omega_t^x} \left| \left( \prod_{v \in B: x_v = 1} \frac{\lambda_v^\nu}{\lambda_v^\mu} \right) \cdot \frac{\tilde{Z}^x_{S,\nu}(t)}{\tilde{Z}^x_{S,\mu}(t)} \cdot \tilde{R} \cdot \tilde{\nu}^x_{S,t}(y) - \tilde{\mu}^x_{S,t}(y) \right|. \tag{45}$$

We bound the approximation error of $\hat{f}$. Define $A = \left( \prod_{v \in B: x_v = 1} \frac{\lambda_v^\nu}{\lambda_v^\mu} \right) \cdot \frac{\tilde{Z}^x_{S,\nu}(t)}{\tilde{Z}^x_{S,\mu}(t)} \cdot \tilde{R}$ and $B = \left( \prod_{v \in B: x_v = 1} \frac{\lambda_v^\nu}{\lambda_v^\mu} \right) \cdot \frac{Z^x_{S,\nu}}{Z^x_{S,\mu}} \cdot \frac{Z_\nu}{Z_\mu}$. By Lemma 29 and the analysis in the proof of Lemma 30, we have $\left( \prod_{v \in B: x_v = 1} \frac{\lambda_v^\nu}{\lambda_v^\mu} \right) < (1 + \theta/\kappa)^n < 2$, $\frac{Z^x_{S,\nu}}{Z^x_{S,\mu}} \leq 2$, and $\frac{Z_\nu}{Z_\mu} \leq 2$. Also note that $d \leq 1$. Using Lemma 32 and our assumption on $\tilde{R}$, it holds that

$$|A - B| \leq \frac{\varepsilon}{25} d.$$

Using triangle inequality, we can bound

$$|f^t(x) - \hat{f}(x)| \leq B \sum_{y \in \Omega_t^x} |\tilde{\nu}^x_{S,t}(y) - \nu^x_S(y)| + |A - B| \sum_{y \in \Omega_t^x} \tilde{\nu}^x_{S,t}(y) + \sum_{y \in \Omega_t^x} |\tilde{\mu}^x_{S,t}(y) - \mu^x_S(y)|.$$

Using the fact $B \leq 8$ and Lemma 33, we have $|f^t(x) - \hat{f}(x)| \leq \frac{\varepsilon}{10} d$. By Lemma 31,

$$|f(x) - \hat{f}(x)| \leq \frac{\varepsilon}{8} d.$$

The construction step takes time $T' \cdot T_G^{\mathsf{sp}}(T'/10^3) + T' \cdot O(n^4)$, where $T' = O(\frac{n^3}{\varepsilon^2} + \frac{n^{5/2}}{\varepsilon^{9/4}})$. Since the hardcore model satisfies the uniqueness condition in (1), we have $T_G^{\mathsf{sp}}(T'/10^3) = O_\eta(\Delta n \log \frac{n}{\varepsilon})$ [CFYZ22, CE22]. The total construction time is $\tilde{O}(\frac{n^7}{\varepsilon^2} + \frac{n^{6.5}}{\varepsilon^{9/4}})$. For each query, the running time is dominated by computing distributions $\tilde{\mu}_{S,t}^x$ and $\tilde{\nu}_{S,t}^x$, which is $O(n^t) = O(n^4)$. $\quad\square$

Finally, we prove Lemma 34.

*Proof of Lemma 34.* Recall our definition of $\mu_B$ and $\nu_B$ is that for any $x \in \Omega_B$,

$$\mu_B(x) = \frac{\prod_{v \in B: x_v = 1} \lambda_v^\mu Z_{S,\mu}^x}{Z_\mu}, \quad \nu_B(x) = \frac{\prod_{v \in B: x_v = 1} \lambda_v^\nu Z_{S,\nu}^x}{Z_\nu}.$$

We can compute that

$$\frac{Z_\nu}{Z_\mu} = \sum_{x \in \Omega_B} \mu_B(x) \frac{\prod_{v \in B: x_v = 1} \lambda_v^\nu Z_{S,\nu}^x}{\prod_{v \in B: x_v = 1} \lambda_v^\mu Z_{S,\mu}^x} = \mathbf{E}_{x \sim \mu_B} \Big[ \underbrace{\frac{\prod_{v \in B: x_v = 1} \lambda_v^\nu Z_{S,\nu}^x}{\prod_{v \in B: x_v = 1} \lambda_v^\mu Z_{S,\mu}^x}}_{\triangleq Q(x)} \Big].$$

We estimate $Z_\nu/Z_\mu$ by sampling $x$ from $\mu_B$, approximating $Q(x)$ and taking the average. We propose the algorithm as follow. Let $T' = O(\frac{n^3 + n/\kappa}{\varepsilon^2})$ be a sufficiently large integer.

- Draw $T'$ independent approximate samples $x_1, \cdots, x_{T'}$ from $\mu_B$ with $d_{\mathrm{TV}}(\mu_B, x_i) \le \frac{1}{1000T'}$.

- Compute $\tilde{Q}(x) = \frac{\prod_{v \in B: x_v = 1} \lambda_v^\nu \tilde{Z}_{S,\nu}^x(t)}{\prod_{v \in B: x_v = 1} \lambda_v^\mu \tilde{Z}_{S,\mu}^x(t)}$ for $x = x_1, \cdots x_{T'}$, where $\tilde{Z}_{S,\mu}^x(t), \tilde{Z}_{S,\nu}^x(t)$ are defined in (43).

- Compute $\tilde{R} = \frac{1}{T'} \sum_{i=1}^{T'} \tilde{Q}(x_i)$.

First, for any $x \in \Omega_B$, by Lemma 32, since $\eta(\kappa, t) \le \frac{\varepsilon}{200}$ and $D/\kappa \le \theta/\kappa \le 1/(10n)$,

$$\left| Q(x) - \tilde{Q}(x) \right| = \left| \frac{\prod_{v \in B: x_v = 1} \lambda_v^\nu Z_{S,\nu}^x}{\prod_{v \in B: x_v = 1} \lambda_v^\mu Z_{S,\mu}^x} - \frac{\prod_{v \in B: x_v = 1} \lambda_v^\nu \tilde{Z}_{S,\nu}^x(t)}{\prod_{v \in B: x_v = 1} \lambda_v^\mu \tilde{Z}_{S,\mu}^x(t)} \right| = \left| \frac{Z_{S,\nu}^x}{Z_{S,\mu}^x} - \frac{\tilde{Z}_{S,\nu}^x(t)}{\tilde{Z}_{S,\mu}^x(t)} \right| \prod_{v \in B: x_v = 1} \frac{\lambda_v^\nu}{\lambda_v^\mu}$$

$$\le \frac{\varepsilon}{200} d \left( \frac{\kappa + D}{\kappa} \right)^n \le \frac{\varepsilon d}{150}.$$

Recall $h(x) = \nu_B(x)/\mu_B(x)$, the variance of $Q(x)$ is

$$\mathbf{Var}_{\mu_B}[Q] = \mathbf{Var}_{x \sim \mu_B} \left[ \frac{\prod_{v \in B: x_v = 1} \lambda_v^\nu Z_{S,\nu}^x}{\prod_{v \in B: x_v = 1} \lambda_v^\mu Z_{S,\mu}^x} \right] = \mathbf{Var}_{x \sim \mu_B} \left[ \frac{Z_\nu \nu_B(x)}{Z_\mu \mu_B(x)} \right] = \frac{Z_\nu^2}{Z_\mu^2} \mathbf{Var}_{\mu_B}[h].$$

In (37), we showed that $\frac{Z_\mu}{Z_\nu} \le 1 + \frac{4nD}{\kappa}$, which implies $\frac{Z_\mu^2}{Z_\nu^2} \le \left(1 + \frac{4nD}{\kappa}\right)^2 \le e^{4/5}$. In (33), we also proved that $\mathbf{Var}_{\mu_B}[h] \le O_\eta(d^2) \cdot \left(n^3 + \frac{n}{\kappa}\right)$. We can conclude that

$$\mathbf{Var}_{\mu_B}[Q] \le O_\eta(d^2) \cdot \left(n^3 + \frac{n}{\kappa}\right).$$

Assume that we have an ideal algorithm that draw perfect samples $x_1, \ldots, x_{T'}$ and exactly compute $Q(x_1), \cdots, Q(x_{T'})$ and compute $R^* = \frac{1}{T'} \sum_{i=1}^{T'} Q(x_i)$. Note $T' = O(\frac{n^3 + n/\kappa}{\varepsilon^2})$. By Chebyshev's inequality, if $T'$ is sufficiently large, we have

$$\mathbf{Pr}\left[ |R^* - Z_\nu/Z_\mu| \ge \frac{\varepsilon d}{300} \right] \le 0.005.$$

Note that $\left|Q(x) - \tilde{Q}(x)\right| \leq \frac{\varepsilon d}{150}$ for all $x \in \Omega_B$. All the approximate samples $x_1, \cdots x_{T'}$ can be coupled successfully with perfect samples with probability at least $1 - T' \cdot \frac{1}{1000T'} = 0.999$. We can couple our algorithm with ideal algorithm such that $|R^* - \tilde{R}| \leq \frac{\varepsilon d}{150}$ with probability at least $0.99$. By a union bound, with probability at least $0.99$,

$$\left|\tilde{R} - \frac{Z_\nu}{Z_\mu}\right| \leq \frac{\varepsilon d}{150} + \frac{\varepsilon d}{300} = \frac{\varepsilon d}{100}.$$

The running time of our algorithm is $T' \cdot T_G^{\mathsf{sp}}(T'/10^3) + T' \cdot O(n^t)$ because each $\tilde{Q}(x)$ can be computed in time $O(n^t)$. $\qquad\square$

# 6 Proofs of algorithmic results

## 6.1 The general algorithm (Proof of Theorem 5)

**Compute marginal lower bound** In our main theorem, the input instance is promised to be $b$-marginally bounded for some parameter $b$. However, the specific value of $b$ is not given to the algorithm. The following algorithm computes the tight value of marginal lower bound.

**Lemma 35.** *There exists an algorithm such that given any hardcore or Ising model $\mu$ in graph $G = (V, E)$, it returns a value $b$ in time at most $(1/b)^{O(1/b)}n$ for hardcore model and in time at most $O(n + m)$ for Ising model such that*

$$b = \min\{\mu_v^\sigma(c) \mid v \in V, \sigma \text{ is a feasible partial pinning, and } \mu_v^\sigma(c) > 0\}. \tag{46}$$

The proof of Lemma 35 will be given later. With this lemma, we can assume that the value of $b$ is known to the algorithm.

**Pre-processing step** For Ising model, we need the following pre-processing step to reduce the general Ising model to the soft-Ising model. Recall that an Ising model $(G = (V, E), J, h)$ is said to be *soft* if $h \in \mathbb{R}^V$ instead of $h \in (\mathbb{R} \cup \{\pm\infty\})^V$. There are three cases:

1. Case 1: If there exists $v$ such that $(h_v^\mu = +\infty, h_v^\nu = -\infty)$ or $(h_v^\mu = -\infty, h_v^\nu = +\infty)$, we can direct compute that $d_{\mathrm{TV}}(\mu, \nu) = 1$.

2. Case 2: If there exists $v$ such that $(h_v^\mu = \pm\infty, h_v^\nu \neq \pm\infty)$ or $(h_v^\mu \neq \pm\infty, h_v^\nu = \pm\infty)$, without loss of generality, we consider the case $(h_v^\mu = +\infty, h_v^\nu \neq \pm\infty)$. We have

$$d_{\mathrm{TV}}(\mu, \nu) \geq |\mu_v(-) - \nu_v(-)| = |0 - \nu_v(-)| = \nu_v(-) \geq b,$$

where the last inequality due to the marginal lower bound. We use additive-error algorithm in Theorem 17 with additive error $b\varepsilon$ in this case. The running time is

$$O\left(T_G^{\mathsf{ct}}\left(\frac{b\varepsilon}{4}\right) + \frac{1}{b^2\varepsilon^2}\left(T_G^{\mathsf{wt}} + T_G^{\mathsf{sp}}\left(\frac{b\varepsilon}{4}\right)\right)\right) = O\left(T_G^{\mathsf{ct}}\left(\frac{b\varepsilon}{4}\right) + \frac{1}{b^2\varepsilon^2}\left(T_G^{\mathsf{sp}}\left(\frac{b\varepsilon}{4}\right)\right)\right), \tag{47}$$

where the equation holds because $T_G^{\mathsf{wt}} = O(n + m)$ and we can assume $T_G^{\mathsf{sp}}(\cdot), T_G^{\mathsf{ct}}(\cdot)$ is at least $\Omega(n + m)$ since the algorithm needs to read all vertices and all edges.

3. Case 3: For all $v \in V$, if $h_v^\nu = \pm\infty$ or $h_v^\mu = \pm\infty$, then $h_v^\nu = h_v^\mu$. These vertices are fixed to some value with probability 1. By the standard self-reducibility, one can remove all these vertices and change external fields of neighbors to obtain two soft-Ising models. Formally,

one can go through all vertices $v \in V$ whose value is fixed as $c \in \{\pm\}$, for every free neighbor $u$ of $v$, update $h_u^\mu \leftarrow h_u^\mu + J_{uv}^\mu c$ and $h_u^\nu \leftarrow h_u^\nu + J_{uv}^\nu c$.

We remark that (1) the new soft-Ising models also have the same marginal lower bound because we only remove vertices whose marginal lower bound is 1; (2) to sample from the soft-Ising model, one can call sampling oracle on original model and do a projection; to approximately count the partition function, one can use the approximating counting oracle on the original model, because two partition functions differ only by an easy-to-compute factor.

We also do the pre-processing step for hardcore model. A hardcore model $(G, \lambda)$ is said to be soft if $\lambda \in \mathbb{R}_{>0}^V$ instead of $\lambda \in \mathbb{R}_{\geq 0}^V$. There two cases.

1. Case 1: There exists $v$ such that $(\lambda_v^\mu = 0, \lambda_v^\nu > 0)$ or $(\lambda_v^\mu > 0, \lambda_v^\nu = 0)$, then $d_{\mathrm{TV}}(\mu, \nu) \geq b$. We use additive-error algorithm to solve the problem in time (47).

2. Case 2: For all $v$, $\lambda_v^\mu = 0$ if and only if $\lambda_v^\nu = 0$. We can simply remove all such vertices and work on the soft-hardcore model on the remaining graph. Again, the marginal lower bound and sampling/approximate counting oracles also work for new soft-hardcore model.

**The main algorithm** Since we work on soft models, $b \leq \frac{1}{2}$. Define the parameters

$$
C_{\mathrm{par}}^{\mathrm{TV}} = \begin{cases} b^3 & \text{hardcore model,} \\ \frac{b^2}{2} & \text{Ising model.} \end{cases}, \text{ and } \theta = \begin{cases} \frac{b}{2(1-b)n} & \text{hardcore model,} \\ \frac{1}{2(n+3m)} & \text{Ising model.} \end{cases}
$$

The algorithm computes the parameter distance $d_{\mathrm{par}}(\mu, \nu)$ in time $O(n + m)$. If $d_{\mathrm{par}}(\mu, \nu) \geq \theta$, then by Lemma 15, $d_{\mathrm{TV}}(\mu, \nu) \geq \theta C_{\mathrm{par}}^{\mathrm{TV}}$, we use additive-error algorithm in Theorem 17 with additive error $\theta C_{\mathrm{par}}^{\mathrm{TV}} \varepsilon$. Similar to (47), the running time is

$$
O\left( T_G^{\mathsf{ct}}\left( \frac{\theta C_{\mathrm{par}}^{\mathrm{TV}} \varepsilon}{4} \right) + \frac{1}{(\theta C_{\mathrm{par}}^{\mathrm{TV}} \varepsilon)^2} \left( T_G^{\mathsf{sp}}\left( \frac{\theta C_{\mathrm{par}}^{\mathrm{TV}} \varepsilon}{8} \right) \right) \right). \tag{48}
$$

Next, assume that $d_{\mathrm{par}}(\mu, \nu) < \theta$. We use the basic algorithm in Theorem 22. We have the following two lemmas for the soft-hardcore and soft-Ising models.

**Lemma 36.** *Let $\mu$ and $\nu$ be two soft-hardcore models satisfying $b$-marginal lower bound and $d_{\mathrm{par}}(\mu, \nu) \leq \theta$. Then $\mu$ and $\nu$ satisfy Condition 21 with $K = 4n/(bC_{\mathrm{par}}^{TV})$ and $L = 2$.*

**Lemma 37.** *Let $\mu$ and $\nu$ be two soft-Ising models satisfying $d_{\mathrm{par}}(\mu, \nu) \leq \theta$ and $d_{\mathrm{TV}}(\mu, \nu) \geq C_{\mathrm{par}}^{TV} d_{\mathrm{par}}(\mu, \nu)$. Then $\mu$ and $\nu$ satisfy Condition 21 with $K = 4(n + m)/C_{\mathrm{par}}^{TV}$ and $L = 2$.*

Assuming the above two lemmas, we can use Theorem 22 to solve the problem in time

$$
O\left( T_G^{\mathsf{ct}}\left( \frac{\varepsilon}{4} \right) + T \cdot T_G^{\mathsf{sp}}\left( \frac{1}{100T} \right) \right), \quad \text{where } T = O\left( \frac{L^2 K^2}{\varepsilon^2} \right). \tag{49}
$$

The final running time of our algorithm is dominated by the maximum of (47), (48), (49), and the running time in Lemma 35. Since both $T_G^{\mathsf{sp}}, T_G^{\mathsf{ct}}$ are non-increasing functions, the running time of our algorithm is at most

$$
C_b \cdot \frac{N^2}{\varepsilon^2} T_G^{\mathsf{sp}}\left( \frac{\varepsilon^2}{C_b N^2} \right) + T_G^{\mathsf{ct}}\left( \frac{\varepsilon}{C_b N} \right) + C_b' N,
$$

where $C_b, C_b' \geq 1$ are parameters depending only on $b$ and for hardcore model, $N = n$; for Ising model, $N = n + m$. For hardcore model

$$C_b = \text{poly}\left(\frac{1}{b}\right), \quad C_b' = \left(\frac{1}{b}\right)^{O(\frac{1}{b})};$$

and for Ising model,

$$C_b = \text{poly}\left(\frac{1}{b}\right), \quad C_b' = O(1).$$

The parameter $C_b'$ comes from the running time in Lemma 35.

**Remark 38.** Theorem 5 presents a simplified version that assumes the marginal lower bound $b$ to be a constant. However, our algorithm applies to both the Ising and hardcore models with an arbitrary marginal lower bound $b$, where $b$ may depend on the size of the input. The running time of our algorithm is given by

$$\text{poly}\left(\frac{1}{b}\right) \cdot \frac{N^2}{\varepsilon^2} T_G^{\mathsf{sp}}\left(\frac{\text{poly}(b)\varepsilon^2}{N^2}\right) + T_G^{\mathsf{ct}}\left(\frac{\text{poly}(b)\varepsilon}{N}\right) + C_b' N,$$

- **Ising model:** The parameter $C_b' = O(1)$, so a polynomial-time reduction from TV-distance estimation to sampling and approximate counting exists for the Ising model with marginal lower bound $b \geq \frac{1}{\text{poly}(n)}$.

- **Hardcore model:** The parameter $C_b' = (1/b)^{O(1/b)}$, which comes from the running time in Lemma 35. One can improve the last term $C_b' N$ to $\text{poly}(n) \cdot T_G^{\mathsf{ct}}(\frac{1}{10})$ by assuming a slightly stronger approximate counting oracle. In Theorem 5, we only assume approximate counting oracles for $\mu$ and $\nu$. If we further assume that approximate counting oracles work for all conditional distributions induced by $\mu$ and $\nu$, then by going through the proof of Lemma 35, we can obtain an algorithm that computes $b'$ in time $\text{poly}(n) \cdot T_G^{\mathsf{ct}}(\frac{1}{10})$, such that with high probability, the value $b'$ satisfies $\frac{b}{2} \leq b' \leq b$ for $b$ defined in (46). This $b'$ is also a marginal lower bound and provides a constant approximation to the true lower bound. We can then use this $b'$ in the remainder of the algorithm. All the subsequent proofs follow for this $b'$. Hence, given the stronger approximate counting oracle, the polynomial-time reduction exists for the hardcore model with marginal lower bound $b \geq \frac{1}{\text{poly}(n)}$.

Finally, we give the proofs of technical lemmas. We need the following property about the soft hardcore model.

**Lemma 39.** *Let $0 < b < 1$. Suppose a soft-hardcore model $(G, \lambda)$ is $b$-marginally bounded. For any vertex $v \in V$, let $\deg_v^{free}$ denote the number of free neighbors $u$ of $v$ such that $\lambda_u > 0$. For any $v \in V$, it holds that $\deg_v^{free} \leq \frac{\ln b}{\ln(1-b)}$.*

*Proof.* Let $N^2(v)$ denote the set of vertices with distance 2 to vertex $v \in V$ in graph $G$. Let $\sigma$ be a pinning that fixes all vertices in $N^2(v)$ to the value $-1$. Let $N^{\text{free}}(v) = \{v_1, v_2, \ldots, v_\ell\}$ denote the set of free neighbors of $v$, where $\ell = \deg_v^{\text{free}}$. Let $\mu$ denote the Gibbs distribution. Conditional on $\sigma$, $v$ takes value $+$ only if all free neighbors of $v$ take the value $-$. Since the hardcore model is soft, $v$ takes $+$ with a positive probability so that the marginal lower bound appiles. We have

$$b \leq \mu_v^\sigma(+) \leq \prod_{j=1}^\ell \mu_{v_j}(- \mid \sigma \text{ and } (\forall k < j, v_k \leftarrow -)).$$

36

For any free neighbor $v_j$, conditional on $\sigma$ and $v_k \leftarrow -$ for all $k < j$, $v_j$ takes $+$ with a positive probability so that $v_j$ takes $+$ with probability at least $b$. Hence,

$$b \leq (1-b)^{\deg_v^{\text{free}}}.$$

Note that $1 - b < 1$. This proves the upper bound of the free degree. $\qed$

*Proof of Lemma 35.* For hardcore model, if $\lambda_u = 0$ for some $u \in V$, we fix $u = -$ and consider the remained subgraph. If the subgraph has no vertices, we just return $b = 1$.

When $c = -$, for each any partial configuration $\tau$ on $\Lambda \subseteq V$ where $v \notin \Lambda$, consider the marginal probability $\mu_v^\tau(-)$. Then $\mu_v^\tau(-)$ is a convex combination of $\mu_v^\sigma(-)$'s, where $\sigma$ is a partial configuration on $V \setminus \{v\}$. For any $\sigma$, $\mu_v^\sigma(-)$ is positive. Due to the conditional independence, we only need to consider the worst pinning of on $N(v)$, where $N(v)$ is the set of all neighbors of the vertex $v$. It is easy to see $\mu_v^\sigma(-) \geq \frac{1}{1+\lambda_v^\mu}$, and equality is achieved when $\sigma$ fixes the values of all neighbors to be $-$.

Now we consider the case $c = +$ and fix a vertex $v$. Consider a partial configuration $\sigma \in \{\pm\}^\Lambda$ for subset $\Lambda \subseteq V \setminus \{v\}$. If there exists a vertex $u \in N(v)$ such that $u \in \Lambda$, to make $\mu_v^\sigma(+)$ nonzero, $\sigma(u)$ must be $-$. Assume $u \in N(v)$ and $\sigma_v = -$. We consider another subset $\sigma'$ on $\Lambda' = \Lambda \setminus \{u\}$ such that $\sigma'_{\Lambda'} = \sigma_{\Lambda'}$. Define the notation

$$w_\mu(\sigma, v = +) = \sum_{\tau \in \{\pm\}^V : \tau_\Lambda = \sigma \wedge \tau_v = +} w_\mu(\tau).$$

We have $w_\mu(\sigma, v = +) = w_\mu(\sigma', v = +)$, because $\tau_v = +$ forces all vertices in $N(v)$ to take the value $-$. On the other hand, $w_\mu(\sigma, v = -) \leq w_\mu(\sigma', v = -)$, because $u$ in $\sigma_0$ is free and it can either take $-$ or $+$. Our goal is to find a condition such that $v$ takes $+$ with the minimum positive probability. We can assume $N(v) \cap \Lambda = \varnothing$. Again, $\mu_v^\sigma(+)$ is a convex combination over all $\mu_v^\tau(+)$, where the feasible partial configuration $\tau \in \{\pm\}^{V \setminus (N(v) \cup v)}$ that fixes the value of all vertices except $N(v) \cup \{v\}$. We only need to consider the worst case of $\tau$. Note that

$$\frac{w_\mu(\tau, v = +)}{w_\mu(\tau, v = -)} = \frac{\lambda_v^\mu}{\sum_{\rho \in \{\pm\}^{N(v)}, w_\mu(\tau, \rho, -) > 0} \prod_{u \in N(v), \rho_u = +} \lambda_u^\mu}.$$

It is easy to verify when $\tau = \tau_0$ such that for all $u \in V \setminus (N(v) \cup v)$, $\tau_0(u) = -$, the above ratio obtains its minimum, because other $\tau$ may forbid some possible $\rho$ in the summation. Our algorithm for computing the value of $b$ is:

- Compute $b_0 = \frac{1}{1 + \max_{v \in V}(\lambda_v^\mu)}$.

- For each $v \in V$, compute $m_v = \mu_v^{\tau_0}(+)$ by enumerating all independent sets of $N(v)$.

- Output $b = \min\{b_0, \min_{v \in V}\{m_v\}\}$.

By lemma 39, because we already remove all vertices with zero $\lambda_v$, we have $N(v) \leq \frac{\ln(b)}{\ln(1-b)}$. Let $k = \max_{v \in V} |N(v)| = O(\frac{1}{b} \log \frac{1}{b})$. The running time of the algorithm is

$$O(n2^k k^2) = \left(\frac{1}{b}\right)^{O(\frac{1}{b})} n.$$

The running time $O(k2^k)$ is for the exact computation of the $\mu_v^{\tau_0}(+)$. As stated in Remark 38, given the approximate counting oracle for conditional distributions, we can compute an approximate value $p_v$ such that $\frac{1}{2}\mu_v^{\tau_0}(+) \leq p_v \leq \mu_v^{\tau_0}(+)$ with high probability.

For Ising model, similar to the pre-processing step, we can first remove all $v$ with $|h_v| = \infty$ and then change the external fields on all neighbors of $v$. After this step, we only need to consider a soft-Ising model $(G, J, h)$ in the remaining graph $G$.

We also analyze when $\mu_v^\sigma(c)$ obtains the minimum. Since we deal with soft-Ising model, any $\sigma \in \{\pm\}^V$ appears with positive probability. Since $\mu_v^\sigma(c)$ is a convex combination of $\mu_v^\tau(c)$, where $\tau$ is a pinning on $V \setminus \{v\}$. Due to the conditional independence, if all $N(v)$ is fixed, then other vertices do not influence on the marginal distribution at $v$. Hence, to minimize $\mu_v^\sigma(c)$, we only need to consider $\sigma \in \{\pm\}^{N(v)}$. The marginal distribution can be written as

$$\frac{\mu_v^\sigma(c)}{\mu_v^\sigma(-c)} = \frac{\exp(\sum_{u \in N(v)} J_{vu}\sigma_u c + h_v c)}{\exp(-\sum_{u \in N(v)} J_{vu}\sigma_u c - h_v c)} = \exp\left(2 \sum_{u \in N(v)} J_{vu}\sigma_u c + 2h_v c\right).$$

To find a $\sigma$ that minimizes $\sum_{u \in N(v)} J_{vu}^\mu \sigma_u c$, we greedily assign $\sigma_u \in \{\pm\}$ according to the sign of $J_{vu}$. The final result is for any $c \in \{\pm\}$,

$$\min_{\sigma \in \{\pm\}^{N(v)}} \mu_v^\sigma(c) = \frac{g(v,c)}{g(v,c) + 1/g(v,c)}, \text{ where } g(v,c) = \exp\left(h_v c - \sum_{u \in N(v)} |J_{vu}|\right).$$

Our algorithm is:

- For each $v \in V$, compute $g(v) = \min_{c \in \{\pm\}} \frac{g(v,c)}{g(v,c) + 1/g(v,c)}$.

- Output $b = \min_{v \in V}\{g(v)\}$.

The running time is $O(n + m)$. $\qquad\square$

*Proof of Lemma 36.* For all $\sigma \in \{\pm\}^V$, if $\mu(\sigma) > 0$, $\sigma$ corresponds to an independent set of $G$. Because $\nu$ is soft-hardcore, then $\nu(\sigma) > 0$, so $\nu$ is absolutely continuous with respect to $\mu$.

For each $v \in V$, consider $\sigma = (-1)^{V \setminus v}$. Because $\mu$ and $\nu$ are both soft-hardcore models and satisfy the $b$-marginal lower bound, $\mu_v^\sigma(+), \mu_v^\sigma(-), \nu_v^\sigma(+), \nu_v^\sigma(-) \geq b$. We have $\frac{\lambda_u^x}{1+\lambda_u^x}, \frac{1}{\lambda_u^x+1} \geq b$ for all $v \in V$, $x \in \{\mu, \nu\}$, which means

$$\frac{b}{1-b} \leq \lambda_v^x \leq \frac{1-b}{b}.$$

The above inequality means that $b \leq \frac{1}{2}$. We can compute the ratio of the weight

$$\frac{w_\nu(\sigma)}{w_\mu(\sigma)} = \prod_{v:\sigma(v)=+} \frac{\lambda_v^\nu}{\lambda_v^\mu} \leq \prod_{v:\sigma(v)=+} \frac{\lambda_v^\mu + d_{\text{par}}(\mu,\nu)}{\lambda_v^\mu}$$

$$\leq \left(\frac{\lambda_v^\mu + d_{\text{par}}(\mu,\nu)}{\lambda_v^\mu}\right)^n \leq \left(1 + \frac{(1-b)d_{\text{par}}(\mu,\nu)}{b}\right)^n.$$

For hardcore model, $d_{\text{par}}(\mu,\nu) \leq \theta = \frac{b}{2(1-b)n}$, so

$$\frac{w_\nu(\sigma)}{w_\mu(\sigma)} \leq 1 + \frac{3n(1-b)d_{\text{par}}(\mu,\nu)}{b}.$$

Similarly, $\frac{w_\nu(\sigma)}{w_\mu(\sigma)} \geq 1 - \frac{n(1-b)d_{\text{par}}(\mu,\nu)}{b}$. By Lemma 15, $d_{\text{TV}}(\mu,\nu) \geq C_{\text{par}}^{\text{TV}} d_{\text{par}}(\mu,\nu)$, then

$$\sqrt{\mathbf{Var}[W]} \leq \frac{4n(1-b)d_{\text{par}}(\mu,\nu)}{b} \leq \frac{4n(1-b)}{bC_{\text{par}}^{\text{TV}}} d_{\text{TV}}(\mu,\nu) < \frac{4n}{bC_{\text{par}}^{\text{TV}}} d_{\text{TV}}(\mu,\nu), \text{ and}$$

$$\mathbf{E}[W] \geq \min_\sigma \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \geq 1 - \frac{1}{2} = \frac{1}{2}.$$

This verifies Condition 21. $\qquad\square$

*Proof of Lemma 37.* Since both $\mu$ and $\nu$ are soft-Ising models, the absolutely continuous condition $\nu \ll \mu$ holds. Let $J^\mu, h^\mu$ and $J^\nu, h^\nu$ be the interaction matrices and external field vectors of $\mu$ and $\nu$, respectively. Denote $D = d_{\mathrm{par}}(\mu, \nu)$. For any $\sigma \in \{\pm\}^V$, we have

$$\frac{w_\nu(\sigma)}{w_\mu(\sigma)} = \exp\left(\sum_{u \in V}(h_u^\nu - h_u^\mu)\sigma_u + \sum_{\{u,v\} \in E}(J_{uv}^\nu - J_{uv}^\mu)\sigma_u\sigma_v\right).$$

By the definition of parameter distance in Definition 14, we have

$$\left|\sum_{u \in V}(h_u^\nu - h_u^\mu)\sigma_u + \sum_{\{u,v\} \in E}(J_{uv}^\nu - J_{uv}^\mu)\sigma_u\sigma_v\right| \le \sum_{u \in V}|h_u^\nu - h_u^\mu| + \sum_{\{u,v\} \in E}|J_{uv}^\nu - J_{uv}^\mu|$$

$$\le \left(\sum_{u \in V}(\deg_u + 1) + \sum_{\{u,v\} \in E}1\right)D$$

$$= (n + 3m)D,$$

it implies

$$\exp(-(n + 3m)D) \le \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \le \exp((n + 3m)D).$$

For soft-Ising model, $\theta = \frac{1}{2(n+3m)}$ and $D < \theta$, so that

$$1 - (n + 3m)D \le \frac{w_\nu(\sigma)}{w_\mu(\sigma)} \le 1 + 3(n + 3m)D.$$

Hence, $\sqrt{\mathbf{Var}[W]} \le 4(n + 3m)D \le \frac{4(n+3m)}{C_{\mathrm{par}}^{\mathrm{TV}}}d_{\mathrm{TV}}(\mu, \nu)$, and $\mathbf{E}[W] \ge 1 - \frac{1}{2} = \frac{1}{2}$. $\square$

## 6.2 The improved algorithms for hardcore model in the uniqueness regime

*Proof of Theorem 8.* Consider a hardcore model $(G, \lambda)$, where $\lambda \in \mathbb{R}_{\ge 0}^V$, that satisfies the uniqueness condition. Define threshold $\theta$ for hardcore model as

$$\theta = 10^{-10}\frac{\varepsilon^{1/4}}{n^{5/2}} = \Theta\left(\frac{\varepsilon^{1/4}}{n^{5/2}}\right).$$

If $d_{\mathrm{par}}(\mu, \nu) > \theta$, then by Lemma 15, we know that $d_{\mathrm{TV}}(\mu, \nu) \ge \frac{\theta}{5000}$. we use the algorithm Theorem 17 to achieve the additive error $\frac{\varepsilon D}{5000}$. For the hardcore model in the uniqueness regime, we have $T_G^{\mathsf{sp}}(\delta) = O_\eta(\Delta n \log\frac{n}{\delta})$ and $T_G^{\mathsf{ct}}(\delta) = \tilde{O}_\eta(\frac{\Delta n^2}{\delta^2})$. Note that $T_G^{\mathsf{wt}} = O(n)$. The running time for this case is at most

$$O\left(T_G^{\mathsf{ct}}\left(\frac{\varepsilon D}{20000}\right) + \frac{1}{\varepsilon^2 D^2}\left(n + T_G^{\mathsf{sp}}\left(\frac{\varepsilon D}{20000}\right)\right)\right) = \tilde{O}_\eta\left(\frac{\Delta n^2}{\varepsilon^2 \theta^2}\right) = \tilde{O}_\eta\left(\frac{\Delta n^7}{\varepsilon^{5/2}}\right).$$

If $d_{\mathrm{par}}(\mu, \nu) \le \theta$, we use Theorem 23 with running time $\tilde{O}_\eta\left(\frac{n^7}{\varepsilon^2} + \frac{n^{6.5}}{\varepsilon^{9/4}}\right)$. The over all running time is the maximum of two cases, which is $\tilde{O}_\eta\left(\frac{\Delta n^7}{\varepsilon^{5/2}}\right)$. $\square$

The choice of $\theta$ is closely related to the choices of $t$ and $\kappa$ in the proof of Lemma 25. We choose the parameters to minimize the exponent on $n$ in the running time of Theorem 8.

*Proof of Theorem 9.* Now, we further assume that $\Delta = O(1)$ and $\lambda_v^\pi = \Omega(1)$ or $0$ for all $v \in V$ and $\pi \in \{\mu, \nu\}$. We do a similar pre-processing step as that in Section 6.1. Suppose there exists $v$ such that $(\lambda_v^\mu = \Omega(1), \lambda_v^\nu = 0)$ or $(\lambda_v^\mu = \Omega(1), \lambda_v^\nu = 0)$. Say we are in the first case. Then $\nu_v(+) = 0$ and $\mu_v(-) \geq \frac{\lambda_v^\mu}{1+\lambda_v^\mu}(\frac{1}{1+\lambda_c(\Delta)})^\Delta = \Omega(1)$. The total variation $d_{\mathrm{TV}}(\mu, \nu) = \Omega(1)$. We can use Theorem 17 to solve the problem in time $\tilde{O}(\frac{n^2}{\varepsilon^2})$. For all $v \in V$ with $\lambda_v^\mu = \lambda_v^\nu = 0$, we can remove $v$. Hence, we can assume $\Omega(1) = \lambda_v^\pi \leq (1-\eta)\lambda_c(\Delta)$ for all $v \in V$ and $\pi \in \{\mu, \nu\}$.

In this case, we use a different threshold $\theta_0 = \Theta(\frac{1}{\Delta n}) = \Theta(\frac{1}{n})$ because $\Delta = O(1)$. Suppose $d_{\mathrm{par}}(\mu, \nu) \leq \theta_0$, then by (6), we have $\mathbf{Var}[W] = O(n^2)$. It is easy to verify that $\frac{Z_\nu}{Z_\mu} = \Theta(1)$. Theorem 22 gives an algorithm in time $\tilde{O}_\eta(n^3/\varepsilon^2)$. Let us assume $D = d_{\mathrm{par}}(\mu, \nu) > \theta_0$. If we directly apply Theorem 17 to achieve the additive error $\frac{\varepsilon D}{5000}$, then the running time would be

$$O\left(T_G^{\mathsf{ct}}\left(\frac{\varepsilon D}{20000}\right) + \frac{1}{\varepsilon^2 D^2}T_G^{\mathsf{sp}}\left(\frac{\varepsilon D}{20000}\right)\right).$$

The second term $\frac{1}{\varepsilon^2 D^2}T_G^{\mathsf{sp}}\left(\frac{\varepsilon D}{20000}\right) = \tilde{O}_\eta(n^3/\varepsilon^2)$. But the bottleneck is the first term $T_G^{\mathsf{ct}}\left(\frac{\varepsilon D}{20000}\right) = \tilde{O}_\eta(\frac{n^2}{\varepsilon^2 D^2}) = \tilde{O}_\eta(\frac{n^2}{\varepsilon^2 \theta_0^2}) = \tilde{O}_\eta(n^4/\varepsilon^2)$. However, we can improve the first term by noting that the algorithm in Theorem 17 only needs to approximate the ratio $\frac{Z_\nu}{Z_\mu}$ with relative-error $O(\varepsilon D)$ and we show such ratio can be approximated in time $\tilde{O}_\eta(n^3/\varepsilon^2)$. We can construct a sequence of $\lambda^{(0)}, \lambda^{(1)}, \cdots, \lambda^{(\ell)}$ such that $\lambda^{(0)} = \lambda^\mu$, $\lambda^{(\ell)} = \lambda^\nu$ and other $\lambda^{(i)}$ are defined as follows. For any $v \in V$, let $\delta_v = \frac{\lambda_v^\nu}{\lambda_v^\mu}$. For any $1 \leq i \leq \ell$, $\lambda_v^{(i)}$ is defined by $\lambda_v^{(i)} = \lambda_v^\mu \delta_v^{i/\ell}$. We choose $\ell$ such that $\ell = \Theta(1 + nD)$. Note that $\delta_v = 1 \pm O(D)$. We have $\delta_v^{1/\ell} = 1 \pm O(\frac{1}{n})$. Let $w_i$ be the weight function induced by $\lambda^{(i)}$. Let $Z_i$ be the partition function induced by $w_i$. Let $\mu_i$ be the Gibbs distribution induced by $w_i$. We further define $Z_{\ell+1}$ by setting $\lambda_v^{(\ell+1)} = \lambda_v^\mu \delta_v^{(\ell+1)/\ell}$. Define random variable $W_i$ as

$$W_i = \frac{w_i(X)}{w_{i-1}(X)}, \quad \text{where } X \sim \mu_{i-1}.$$

Define $W \triangleq \prod_{i=1}^\ell W_i$, where $W_i$'s are mutually independent. It is easy to verify that

$$\mathbf{E}[W] = \frac{Z_\nu}{Z_\mu} = \frac{Z_\ell}{Z_0}, \quad \text{and } \mathbf{Var}[W] \leq \mathbf{E}[W^2] = \prod_{i=1}^\ell \mathbf{E}[W_i^2] = \frac{Z_{\ell+1}Z_\ell}{Z_0 Z_1}.$$

We have the following bound

$$\frac{\mathbf{Var}[W]}{(\mathbf{E}[W])^2} \leq \frac{\mathbf{E}[W^2]}{(\mathbf{E}[W])^2} \leq \frac{Z_{\ell+1}}{Z_\ell} \cdot \frac{Z_0}{Z_1} = O(1).$$

The last equality follows from the fact that $\delta_v^{1/\ell} = 1 \pm O(\frac{1}{n})$. Hence, to achieve $O(\varepsilon D)$ relative-error, we can draw $O(\frac{1}{\varepsilon^2 D^2})$ samples of $W$, each sample costs $\tilde{O}_\eta(n\ell) = \tilde{O}_\eta(n + n^2 D)$ time. The total running time is

$$\tilde{O}_\eta\left(\frac{n + n^2 D}{\varepsilon^2 D^2}\right) = \tilde{O}_\eta\left(\frac{n}{\varepsilon^2 \theta_0^2} + \frac{n^2}{\varepsilon^2 \theta_0}\right) = O_\eta\left(\frac{n^3}{\varepsilon^2}\right).$$

## 6.3 The algorithm for marginal distributions (Proof of Theorem 13)

Theorem 13 is a simple corollary of Theorem 20.

*Proof of Theorem 13.* For hardcore model $(G, \lambda^\mu)$, $Z_\mu^\sigma$ where $\sigma \in \{\pm\}^S$ is the partition function of $(G[\Lambda], \lambda_\Lambda^\mu)$. The set $\Lambda$ is obtained from $V$ by removing all vertices in $S$ together with all neighbors $u$ of vertices $v \in S$ such that $\sigma_v = +1$. If $(G, \lambda^\mu)$ satisfies the uniqueness condition, then $(G[\Lambda], \lambda_\Lambda^\mu)$ also satisfies the uniqueness condition. Hence, by the previous results in [CFYZ22, CE22] and [ŠVV09], for both $\mu$ and $\nu$, the approximate conditional counting oracle with $T_G^{\mathsf{ct}}(\varepsilon) = O(\frac{\Delta n^2}{\varepsilon^2} \mathrm{polylog} \frac{n}{\varepsilon})$ exists and the sampling oracle with $T_G^{\mathsf{sp}}(\varepsilon) = O(\Delta n \mathrm{polylog} \frac{n}{\varepsilon})$ exists. The theorem follows from Theorem 20. $\qquad\square$

# 7 Proofs of #P-hardness results

In this section, we prove Theorem 11 and Corollary 12. Our starting point is the #P-hardness for exactly counting the number of independent sets in a graph.

**Proposition 40** ([DG00, Theorem 4.2])**.** *The following problem #Ind(3) is #P-complete.*

- *Input: a graph $G = (V, E)$ with maximum degree $\Delta = 3$;*

- *Output: the exact number of independent sets in $G$.*

The above problem is exactly computing the partition function of $(G, \lambda)$ with $\lambda_v = 1$ for all $v \in V$. Let $n = |V|$ and $V = \{1, 2, \ldots, n\}$. Let $\mu_{G, \mathbf{1}}$ denote the uniform distribution over all independent sets in $G$, which is the hardcore distribution in $G$ when $\lambda_v = 1$ for all $v \in V$. Define

$$p_i = \mathbf{Pr}_{X \sim \mu_{G, \mathbf{1}}} [X_i = 0 \mid \forall 1 \le j \le i - 1, X_j = 0], \tag{50}$$

which is the probability that the vertex $i$ is not in a random independent set $X$ conditional on all $j < i$ not being in $X$. By definition, the total number of independent set is

$$Z = \frac{1}{\mu_{G, \mathbf{1}}(\mathbf{0})} = \prod_{i=1}^{n} \frac{1}{p_i}.$$

Suppose for any $i \in [n]$, we can compute $\hat{p}_i$ such that

$$(1 - 4^{-n}) p_i \le \hat{p}_i \le (1 + 4^{-n}) p_i. \tag{51}$$

Let $\hat{Z} = \prod_{i=1}^{n} \frac{1}{\hat{p}_i}$ and it holds that $(1 - 3^{-n}) Z \le \hat{Z} \le (1 + 3^{-n}) Z$. Note that $Z \le 2^n$. We have

$$|\hat{Z} - Z| \le 3^{-n} Z \le 1.5^{-n} < 0.01.$$

We can round $\hat{Z}$ to the nearest integer to recover $Z$. Hence, #Ind(3) can be reduced to the following high-accuracy marginal estimation problem.

**Problem 41.** The high-accuracy marginal estimation problem is defined by

- Input: a graph $G = (V, E)$ with $n$ vertices and maximum degree $\Delta = 3$;

- Output: $n$ numbers $(\hat{p}_i)_{i \in [n]}$ such that for all $i \in [n]$, $(1 - 4^{-n}) p_i \le \hat{p}_i \le (1 + 4^{-n}) p_i$.

## 7.1 Hardness of approximating the TV-distance on a single vertex

We first prove Theorem 11. Let $k(n) = 1$ for all $n \in \mathbb{N}$ be a constant function. We show that if there is a $\text{poly}(n)$ time algorithm for Problem 10 if the input error bound $\varepsilon = \text{poly}(n)$ and both two input hardcore models satisfy the uniqueness condition, then Problem 41 can also be solved in $\text{poly}(n)$ time. Theorem 11 follows from Proposition 40.

Fix an integer $i \in [n]$. Let $G_i$ denote the induced graph $G[S_i]$, where $S_i = \{j \in [n] \mid j \geq i\}$ is the set of vertices with label at least $i$. Let $\mu^{(i)}$ denote the uniform distribution over all independent set in graph $G_i$. In other words, $\mu^{(i)}$ is the Gibbs distribution of hardcore model $(G_i, \mathbf{1})$. Then $p_i$ in (50) is the marginal distribution on vertex $i$ projected from $\mu^{(i)}$. If the maximum degree of $G_i$ is at most 2, then $G_i$ is a set of disconnected lines or circles and $p_i$ can be computed exactly in polynomial time. We can assume the maximum degree of $G_i$ is 3.

Let $\alpha \geq 0$. Define vector $\lambda^\alpha \in \mathbb{R}^{S_i}$ by

$$\lambda_j^\alpha = \begin{cases} \frac{\alpha}{1-\alpha} & \text{if } j = i; \\ 0 & \text{if } j \neq i. \end{cases} \tag{52}$$

Let $\nu^\alpha$ denote the Gibbs distribution of $(G_i, \lambda^\alpha)$. Note that $\lambda_c(3) = 4 > 1$. The following observation is easy to verify.

**Observation 42.** *Both $\mu^{(i)}$ and $\nu^\alpha$ satisfies the uniqueness condition in (1) if $\alpha \leq \frac{1}{2}$.*

By the definition of $\nu^\alpha$, it is easy to see $\nu_i^\alpha(+1) = \alpha$ and

$$d_{\text{TV}}\left(\mu_i^{(i)}, \nu_i^\alpha\right) = |\mu_i^{(i)}(+1) - \alpha| = |p_i - \alpha|.$$

Let $\mathcal{A}(\alpha)$ be the algorithm such that given $\alpha \in [0, \frac{1}{2}]$, it returns a number $\hat{d}$ such that $\frac{d_{\text{TV}}(\mu_i^{(i)}, \nu_i^\alpha)}{1+\varepsilon} \leq \hat{d} \leq (1+\varepsilon)d_{\text{TV}}(\mu_i^{(i)}, \nu_i^\alpha)$, where $\varepsilon = \text{poly}(n)$. By Observation 42, if the polynomial-time algorithm for the problem in Theorem 11 exists, then $\mathcal{A}(\alpha)$ runs in $\text{poly}(n)$ time. We then can use the following algorithm to solve Problem 10 for $p_i$ in $\text{poly}(n)$ time. Thus, the hardness result in Theorem 11 follows from Proposition 40.

---
**Algorithm 1:** Algorithm for high-accuracy marginal estimation

**1** Let $\alpha \leftarrow \frac{1}{2}$ and $\varepsilon = \text{poly}(n)$ be the parameter assumed by algorithm $\mathcal{A}$;
**2 for** *t from 1 to* $50n(1+\varepsilon)^2$ **do**
**3** $\quad \hat{d} \leftarrow \mathcal{A}(\alpha)$;
**4** $\quad \alpha \leftarrow \alpha - \hat{d}/(1+\varepsilon)$;
**5** $\quad$ if the bit length of $\alpha$ is more than $100n$, then round $\alpha$ up to the nearest number that has bit length at most $100n$;
**6 return** $\hat{p}_i = \alpha$.

---

The above algorithm runs in $\text{poly}(n)$ time. We show that the output $\hat{p}$ satisfies (51).

Let $\alpha_t$ be the value of $\alpha$ after the $t$-th iteration. We first show that $p_i \leq \alpha_t$ for all $t$. At the beginning, $\alpha_0 = 1/2$. Since $\mu^{(i)}$ is a uniform distribution over all independent sets, we have $\mu_i^{(i)}(+1) \leq \frac{1}{2}$. By the assumption of algorithm $\mathcal{A}$, $\mathcal{A}(\alpha_t)/(1+\varepsilon) \leq d_{\text{TV}}(\mu_i^{(i)}, \nu_i^{\alpha_t}) = \alpha_t - p_i$. Hence, $\alpha_{t+1} \geq \alpha_t - \mathcal{A}(\alpha_t)/(1+\varepsilon) \geq p_i$.

We next bound the value of $\alpha_t - p_i$. At the beginning, $\alpha_0 = \frac{1}{2}$ so that $\alpha_0 - p_i \leq \frac{1}{2}$. Note that $\alpha_{t+1} < \alpha_t - \frac{\mathcal{A}(\alpha_t)}{1+\varepsilon} + 2^{-90n} \leq \alpha_t - \frac{\alpha_t - p_i}{(1+\varepsilon)^2} + 2^{-90n}$, where $2^{-90n}$ is an upper bound of rounding

42

error. The inequality implies that

$$\alpha_{t+1} - p_i \leq \left(1 - \frac{1}{(1+\varepsilon)^2}\right)(\alpha_t - p_i) + 2^{-90n}.$$

Note that $\hat{p}_i = \alpha_{50n(1+\varepsilon)^2}$. Solving the recurrence implies that

$$0 \leq \hat{p}_i - p_i \leq \exp\left(-\frac{50n(1+\varepsilon)^2}{(1+\varepsilon)^2}\right) \cdot \frac{1}{2} + (1+\varepsilon)^2 \cdot 2^{-90n} < 2^{-40n}.$$

Note that $p_i$ is at least $1/2^n$. Hence, the output $\hat{p}$ satisfies (51).

**Remark.** In the above proof, while the TV-distance between the two Gibbs distributions $\mu^{(i)}$ and $\nu^\alpha$ is large (because their parameter distance is 1), the TV distance between their marginal distributions at vertex $i$ can be arbitrarily small. This highlights the distinction between the TV-distance of marginal distributions and the TV-distance of the entire distribution.

## 7.2 Hardness of approximating the TV-distance on a subset of vertices

We now prove Corollary 12. Let $G_i$ be the graph defined as above. Let $n_i = n - i + 1$ denote the number of vertices in $G_i$. One can construct a graph $G_i'$ with by adding a set $\Lambda$ of $\ell$ isolated vertices to $G_i$. Let $N = n_i + \ell$. Let $k(\cdot)$ be the function in Corollary 12. Note that $k(N) = N - \lceil N^\alpha \rceil$. Since $\alpha$ is a constant, one can set $\ell = n^{\Omega(1/\alpha)} = \text{poly}(n)$ so that $k(N) \leq \ell$. Hence, the size of $G_i'$ is a polynomial in $n$.

Let $\mu_{\text{new}}^{(i)}$ be the hardcore model on $G_i'$ such that the external fields on vertices in $G_i$ are 1 and the external fields on $\lambda$ are 0. Let $\nu_{\text{new}}^\alpha$ be the hardcore model on $G_i'$ such that the external fields on vertices in $G_i$ are $\lambda^\alpha$ and the external fields on $\Lambda$ are 0. Let $S$ be a subset of vertices containing vertex $i$ and $k(N) - 1$ vertices in $\Lambda$. It holds that

$$d_{\text{TV}}\left(\mu_{\text{new},S}^{(i)}, \nu_{\text{new},S}^\alpha\right) = d_{\text{TV}}\left(\mu_i^{(i)}, \nu_i^\alpha\right).$$

In words, the total variation distance between marginal distributions on $S$ projected from $\mu_{\text{new}}^{(i)}$ and $\nu_{\text{new}}^\alpha$ is the same as the total variation distance between marginal distributions on vertex $i$ projected from $\mu^{(i)}$ and $\nu^\alpha$. Note that both $\mu_{\text{new}}^{(i)}$ and $\nu_{\text{new}}^\alpha$ satisfy the uniqueness condition. Corollary 12 can be verified by going through the same reduction for Theorem 11.

# Acknowledgment

# References

[AAL23] Jamil Arbas, Hassan Ashtiani, and Christopher Liaw. Polynomial time and private learning of unbounded gaussian mixture models. In *ICML*, volume 202 of *Proceedings of Machine Learning Research*, pages 1018–1040. PMLR, 2023.

[AJK+22] Nima Anari, Vishesh Jain, Frederic Koehler, Huy Tuan Pham, and Thuy-Duong Vuong. Entropic independence: optimal mixing of down-up random walks. In *STOC*, pages 1418–1430. ACM, 2022.

[AK24] Ariel Avital and Aryeh Kontorovich. Sharp bounds on aggregate expert error. *arXiv preprint arXiv:2407.16642*, 2024.

[BGM$^+$23] Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, Dimitrios Myrisiotis, A. Pavan, and N. V. Vinodchandran. On approximating total variation distance. In *IJCAI*, pages 3479–3487. ijcai.org, 2023.

[BGM$^+$24] Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, Dimitrios Myrisiotis, A. Pavan, and N. V. Vinodchandran. Total variation distance meets probabilistic inference. In *ICML*. OpenReview.net, 2024.

[BGM$^+$25] Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, Dimitrios Myrisiotis, A. Pavan, and N. V. Vinodchandran. Computational explorations of total variation distance. In *ICLR*. OpenReview.net, 2025.

[BGMV20] Arnab Bhattacharyya, Sutanu Gayen, Kuldeep S. Meel, and N. V. Vinodchandran. Efficient distance approximation for structured high-dimensional distributions via learning. In *NeurIPS*, 2020.

[Can15] Clément L. Canonne. A survey on distribution testing: Your data is big. but is it blue? *Electron. Colloquium Comput. Complex.*, TR15-063, 2015.

[CCYZ24] Xiaoyu Chen, Zongchen Chen, Yitong Yin, and Xinyuan Zhang. Rapid mixing at the uniqueness threshold. *CoRR*, abs/2411.03413, 2024.

[CE22] Yuansi Chen and Ronen Eldan. Localization schemes: A framework for proving mixing bounds for markov chains (extended abstract). In *FOCS*, pages 110–122. IEEE, 2022.

[CFYZ21] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Rapid mixing of glauber dynamics via spectral independence for all degrees. In *FOCS*, pages 137–148. IEEE, 2021.

[CFYZ22] Xiaoyu Chen, Weiming Feng, Yitong Yin, and Xinyuan Zhang. Optimal mixing for two-state anti-ferromagnetic spin systems. In *FOCS*, pages 588–599. IEEE, 2022.

[CK14] Taolue Chen and Stefan Kiefer. On the total variation distance of labelled markov chains. In *LICS*, pages 33:1–33:10. ACM, 2014.

[CLV21] Zongchen Chen, Kuikui Liu, and Eric Vigoda. Optimal mixing of Glauber dynamics: Entropy factorization via high-dimensional expansion. In *STOC*, pages 1537–1550. ACM, 2021.

[CMR07] Corinna Cortes, Mehryar Mohri, and Ashish Rastogi. $L_p$ distance and equivalence of probabilistic automata. *Int. J. Found. Comput. Sci.*, 18(4):761–779, 2007.

[CR14] Clément L. Canonne and Ronitt Rubinfeld. Testing probability distributions underlying aggregated data. In *ICALP*, volume 8572 of *Lecture Notes in Computer Science*, pages 283–295. Springer, 2014.

[DG00] Martin E. Dyer and Catherine S. Greenhill. On markov chains for independent sets. *J. Algorithms*, 35(1):17–49, 2000.

[DHR08] Laurent Doyen, Thomas A. Henzinger, and Jean-François Raskin. Equivalence of labeled markov chains. *Int. J. Found. Comput. Sci.*, 19(3):549–563, 2008.

[DMR18]   Luc Devroye, Abbas Mehrabian, and Tommy Reddad. The total variation distance between high-dimensional gaussians with the same mean. *arXiv preprint arXiv:1810.08693*, 2018.

[FGJW23]  Weiming Feng, Heng Guo, Mark Jerrum, and Jiaheng Wang. A simple polynomial-time approximation algorithm for the total variation distance between two product distributions. *TheoretiCS*, 2, 2023.

[FGW23]   Weiming Feng, Heng Guo, and Jiaheng Wang. Swendsen-Wang dynamics for the ferromagnetic Ising model with external fields. *Inf. Comput.*, 294:105066, 2023.

[FLL24]   Weiming Feng, Liqiang Liu, and Tianren Liu. On deterministically approximating total variation distance. In *SODA*, pages 1766–1791. SIAM, 2024.

[GSV16]   Andreas Galanis, Daniel Stefankovic, and Eric Vigoda. Inapproximability of the partition function for the antiferromagnetic ising and hard-core models. *Comb. Probab. Comput.*, 25(4):500–559, 2016.

[Jer03]   Mark Jerrum. *Counting, sampling and integrating: algorithms and complexity.* Springer Science & Business Media, 2003.

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

[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.

[Kie18]   Stefan Kiefer. On computing the total variation distance of hidden markov models. In *ICALP*, volume 107 of *LIPIcs*, pages 130:1–130:13. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018.

[KMO$^+$11]  Stefan Kiefer, Andrzej S. Murawski, Joël Ouaknine, Björn Wachter, and James Worrell. Language equivalence for probabilistic automata. In *CAV*, volume 6806, pages 526–540. Springer, 2011.

[Kon24]   Aryeh Kontorovich. On the tensorization of the variational distance. *arXiv preprint arXiv:2409.10368*, 2024.

[MU17]    Michael Mitzenmacher and Eli Upfal. *Probability and computing: randomization and probabilistic techniques in algorithms and data analysis.* Cambridge university press, 2017.

[Sly10]   Allan Sly. Computational transition at the uniqueness threshold. In *FOCS*, pages 287–296. IEEE Computer Society, 2010.

[SS12]    Allan Sly and Nike Sun. The computational hardness of counting in two-spin models on d-regular graphs. In *FOCS*, pages 361–369. IEEE Computer Society, 2012.

[SV03]    Amit Sahai and Salil Vadhan. A complete problem for statistical zero knowledge. *J. ACM*, 50(2):196–249, 2003.

[ŠVV09]   Daniel Štefankovič, Santosh S. Vempala, and Eric Vigoda. Adaptive simulated annealing: A near-optimal connection between sampling and counting. *J. ACM*, 56(3):18:1–18:36, 2009.

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

# A  NP-hardness of approximating the TV-distance

In this section, we prove the hardness results of approximating the TV-distance between two Gibbs distributions beyond the uniqueness threshold. The proof is based on the technique developed in [BGM$^+$25]. We define the following instance family for TV-distance approximation.

**Problem 43.** Let $\Delta \geq 3$ and $\lambda > \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ be two constants.

- *Input*: two hardcore models $(G, \lambda^\mu)$ and $(G, \lambda^\nu)$ defined on the same graph $G = (V, E)$ with maximum degree at most $\Delta$, which specifies two Gibbs distributions $\mu$ and $\nu$ respectively, and an error bound $0 < \varepsilon < 1$. There exists a vertex $v^* \in V$ such that the external fields $\lambda_v^\mu = \lambda_v^\nu = \lambda$ for all $v \neq v^*$, $\lambda_{v^*}^\nu = \infty$, and $\lambda_{v^*}^\mu = \lambda$.

- *Output*: a number $\hat{d}$ such that $|d_{\mathrm{TV}}(\mu, \nu) - \hat{d}| \leq \varepsilon$.

The two input hardcore models are not in the uniqueness regime because $\lambda > \lambda_c(\Delta)$. The output only requires to approximate the TV-distance up to an additive error of $\varepsilon$, which is weaker requirement than the relative-error approximation. The hardness result for this problem implies the hardness of relative-error approximation.

**Theorem 44.** *There is no FPRAS for Problem 43 unless $\mathbf{NP} = \mathbf{RP}$.*

To prove Theorem 44, we need the following lemma, which can be abstracted from the proofs in [BGM$^+$25].

**Lemma 45** ([BGM$^+$25])**.** *Let $\mu$ be a distribution over $\{\pm\}^V$. Let $v \in V$ and $c \in \{\pm\}$ with $\mu_v(c) > 0$. Let $\mu^{vc}$ be the distribution over $\{\pm\}^V$ obtained from $\mu$ by conditioning on $v$ taking value $c$. Then, $d_{\mathrm{TV}}(\mu, \mu^{vc}) = \mu_v(-c)$.*

*Proof.* For any $\sigma \in \{\pm\}^V$ with $\sigma_v = c$, we have $\mu^{vc}(\sigma) \geq \mu(\sigma)$. For any $\tau \in \{\pm\}^V$ with $\tau_v = -c$, we have $\mu^{vc}(\tau) = 0 \leq \mu(\tau)$. Therefore, $d_{\mathrm{TV}}(\mu, \mu^{vc}) = \sum_{\tau \in \{\pm\}^V : \tau_v = -c} \mu(\tau) = \mu_v(-c)$. □

*Proof.* [Proof of Theorem 44] Let $\Delta \geq 3$ and $\lambda > \lambda_c(\Delta) = \frac{(\Delta-1)^{\Delta-1}}{(\Delta-2)^\Delta}$ be two constants. By the hardness results in [SS12, GSV16], unless $\mathbf{NP} = \mathbf{RP}$, there is no FPRAS for approximating the partition function of the hardcode model $\mathbb{S} = (G, (\lambda_v)_{v \in V})$ with $\varepsilon$-relative error, where $\lambda_v = \lambda$ for all $v \in V$ and $G$ has the maximum degree $\Delta$. By the standard counting-to-sampling reduction [JVV86], approximating the partition function $Z_\mathbb{S}$ is equivalent to approximating the probability of $\pi(\sigma^\varnothing)$, where $\pi$ is the Gibbs distribution of $\mathbb{S}$ and $\sigma_v^\varnothing = -1$ for all $v \in V$. In other words, $\sigma^\varnothing$ corresponds to the empty set. Let us number all the vertices in $V$ as $\{1, 2, \cdots, n\}$. To approximate $\pi(\sigma^\varnothing)$, we need to approximate the probability $p_i \triangleq \pi_i(-1 \mid \forall j < i, j$ takes value $-1)$ with relative error $O(\frac{\varepsilon}{n})$. This probability is the same as the marginal probability of $i$ in the induced subgraph $G[V \setminus \{1, 2, \cdots, i-1\}]$. We show how to approximate $p_1 = \pi_1(-1)$. We let $\mathbb{S}^\mu = \mathbb{S}$ and define $\mathbb{S}^\nu$ for $\mathbb{S}$ by changing $\lambda_1^\nu$ to $\infty$. By Lemma 45, we have $d_{\mathrm{TV}}(\mu, \nu) = \mu_1(-1) = \pi_1(-1)$. It is easy to verify that $\pi_1(-1) \geq \frac{1}{1+\lambda} = \Omega(1)$ has a constant lower bound. Hence, if we can solve Problem 43 with $O(\varepsilon/n)$-additive error, we can approximate $\pi_1(-1)$ with $O(\varepsilon/n)$-relative error. The same argument can be applied to other probabilities by considering the instances in induced subgraphs. Hence, if Problem 43 admits an FPRAS, then there is an FPRAS for approximating the partition function of $\mathbb{S}$. □

For the Ising model, one can also verify the instance family stated after Corollary 7 is a hard instance family for approximating the TV-distance with additive error. The proof is similar to the one for the hardcore model. Let $\Delta \geq 3$ be a constant and $\beta < 0$ be a constant with $\exp(2\beta) < \frac{\Delta-2}{\Delta}$. Let $\mathbb{S} = (G, J, h)$ such that $J_{uv} = \beta$ for all $\{u, v\} \in E$ and $h_v = 0$ for all $v \in V$. The starting point is the NP-hardness of approximating the partition function of $\mathbb{S}$ [SS12, GSV16]. Then one can go through the same reduction to estimate $p_i \triangleq \pi_i(-1 \mid \forall j < i, j \text{ takes value } -1)$. We can construct $\mathbb{S}^\mu = (G, J, h^\mu)$ and $\mathbb{S}^\nu = (G, J, h^\nu)$ such that for all $j < i$, $h_j^\mu = h_j^\nu = -\infty$, for all $k > i$, $h_k^\mu = h_k^\nu = 0$, and $h_i^\mu = 0, h_i^\nu = \infty$. By Lemma 45, we have $d_{\mathrm{TV}}(\mu, \nu) = p_i$. One can verify $p_i = \Omega(1)$ so that $O(\varepsilon/n)$-additive error approximation of $d_{\mathrm{TV}}(\mu, \nu)$ is equivalent to $O(\varepsilon/n)$-relative error approximation of $p_i$. The hardness result follows from the same argument as in the proof of Theorem 44.

# B  Poincaré inequality for marginal distribution

The Poincaré inequality for hardcore model in the uniqueness regime is established in [CFYZ21]. The paper considers the hardcore model $(G, \lambda)$ such that all $\lambda_v$ are the same. By verifying the technical condition in [CFYZ21, Theorem 1.9], the following Poincaré inequality also holds for hardcore model with different $\lambda_v$ satisfying (1). For any function $g : \Omega \to \mathbb{R}$, where $\Omega$ is the support of the distribution, we have

$$\mathbf{Var}_\mu[g] \leq C_\eta \sum_{v \in V} \sum_{\sigma \in \Omega_{V-v}} \mu_{V-v}(\sigma) \mathbf{Var}_{\mu^\sigma}[g].$$

Let $g' : \Omega_B \to \mathbb{R}$ be an arbitrary function. Define $g(x) = g'(x_B)$ for all $x \in \Omega$. It holds that $\mathbf{E}_\mu[g] = \mathbf{E}_{\mu_B}[g']$ and $\mathbf{E}_\mu[g^2] = \mathbf{E}_{\mu_B}[(g')^2]$. We have

$$\mathbf{Var}_{\mu_B}[g'] = \mathbf{Var}_\mu[g] \leq C_\eta \sum_{v \in V} \sum_{\sigma \in \Omega_{V-v}} \mu_{V-v}(\sigma) \mathbf{Var}_{\mu^\sigma}[g]$$

$$(g(\sigma) \text{ depends on } \sigma_B) \quad = C_\eta \sum_{v \in B} \sum_{\sigma \in \Omega_{V-v}} \mu_{V-v}(\sigma) \mathbf{Var}_{\mu^\sigma}[g].$$

Fix a partial assignment $\tau \in \Omega_{B-v}$. Note that $\mathbf{E}_{\mu^\tau}[g] = \mathbf{E}_{\mu_B^\tau}[g']$ and $\mathbf{E}_{\mu^\tau}[g^2] = \mathbf{E}_{\mu_B^\tau}[(g')^2]$. We have $\mathbf{Var}_{\mu^\tau}[g] = \mathbf{Var}_{\mu_B^\tau}[g']$. Let $X \sim \mu^\tau$ and $Y = g(X)$. By the law of total variance,

$$\begin{aligned}
\mathbf{Var}_{\mu_B^\tau}[g'] = \mathbf{Var}_{\mu^\tau}[g] = \mathbf{Var}[Y] &= \mathbf{E}[\mathbf{Var}[Y \mid X_{V-v}]] + \mathbf{Var}[\mathbf{E}[Y \mid X_{V-v}]] \\
&\geq \mathbf{E}[\mathbf{Var}[Y \mid X_{V-v}]] \\
&= \sum_{\sigma \in \Omega_{V-v}^\tau} \mu_{V-v}^\tau(\sigma) \mathbf{Var}_{\mu^\sigma}[g],
\end{aligned}$$

where $\Omega_{V-v}^\tau \subseteq \{\pm\}^{V-v}$ is the support of $\mu_{V-v}^\tau$. Combining with the above inequality, we have

$$\begin{aligned}
\mathbf{Var}_{\mu_B}[g'] &\leq C_\eta \sum_{v \in B} \sum_{\tau \in \Omega_{B-v}} \mu_{B-v}(\tau) \sum_{\sigma \in \Omega_{V-v}^\tau} \mu_{V-v}^\tau(\sigma) \mathbf{Var}_{\mu^\sigma}[g] \\
&\leq C_\eta \sum_{v \in B} \sum_{\tau \in \Omega_{B-v}} \mu_{B-v}(\tau) \mathbf{Var}_{\mu_B^\tau}[g'].
\end{aligned}$$

This proves the Poincaré inequality for the marginal distribution.

Alternatively, one can also use the fact the the Poincaré inequality is equivalent to the decay of $\chi^2$-divergence in the down walk of Glauber dynamics. The results follows from the data processing inequality. See [FGJW23, Section 6.1] for more details.