Approximate counting of vertices of 0/1 polytopes: a stronger hardness result

Heng Guo
School of Informatics
University of Edinburgh
hguo@inf.ed.ac.uk

Mark Jerrum*
School of Mathematical Sciences
Queen Mary, University of London
m.jerrum@qmul.ac.uk

Abstract

We show that approximately counting the vertices of a bounded 0/1 polytope, presented as a system of rational linear inequalities, is, informally speaking, **NP**-hard. In particular, there is no FPRAS for this problem unless **RP** = **NP**. The proof is by a reduction from approximately counting homomorphisms from a given graph to a particular four-vertex graph. The main proof ideas were found using GPT-5.6 Sol Ultra.

1 Introduction

We are interested in counting vertices of a polytope defined by linear inequalities $Ax \leq b$. We concentrate on 0/1 polytopes, namely polytopes all of whose vertices belong to $\{0,1\}^n$. The polytope is presented by its defining inequalities, rather than by a list of its vertices.

Problem: #Vertices of a 0/1 Polytope.

Instance: An $m \times n$ rational matrix $A \in \mathbb{Q}^{m\times n}$ and a vector $b \in \mathbb{Q}^m$.

Promise: The inequalities $Ax \leq b$ define a bounded 0/1 polytope P.

Output: The number $v(P)$ of vertices of P.

It is not known at the time of writing whether there is a polynomial-time algorithm to test whether a given pair (A, b) defines a 0/1-polytope, so we include a promise to this effect.

We allow the polytope to be empty, in which case $v(P) = 0$. An FPRAS for #Vertices of a 0/1 Polytope is a randomised algorithm that, given an instance and a parameter $0 < \varepsilon \leq 1$, outputs a non-negative random value Z such that

$$\Pr\left[(1-\varepsilon)v(P) \leq Z \leq (1+\varepsilon)v(P)\right] \geq \tfrac{3}{4},$$

and runs in time polynomial in the encoding length of (A, b) and in ε^{-1}.

As with the theory of NP-completeness, we may gain evidence for the non-existence of an FPRAS for a given counting problem using a suitable notion of reduction. In this application, the

*'This work was supported by the UK Engineering and Physical Sciences Research Council grant number UKRI 2771.

appropriate notion is *AP-reduction*, short for (polynomial-time) Approximation-Preserving reduction. In this paper, our reductions are actually *parsimonious*, that is to say, they preserve the number of solutions exactly. As an efficient parsimonious reduction would meet any reasonable definition of approximation-preserving reduction, we do not provide an exact definition of AP-reduction here. For a precise definition of concepts used in this paper and a general introduction to the complexity of approximate counting, refer to Dyer, Goldberg, Greenhill and Jerrum [1].

Our main result is the following.

Theorem 1. *#Vertices of a 0/1 Polytope is hard for #**P** under AP-reductions. In particular, there is no FPRAS for #Vertices of a 0/1 Polytope unless* **RP** = **NP**.

A counting problem Π is said to "hard for #**P** under AP-reductions" if, for every counting problem $\Pi' \in \#\mathbf{P}$, there is an AP-reduction from Π' to Π. Some explanatory comments on the statement of this theorem follow later in the section. For the time being, we merely highlight the conclusion that, under a standard complexity-theoretic assumption, there is no polynomial-time algorithm (even a randomised one) that approximates the number of vertices of a 0/1-polytope.

The proof of Theorem 1 is via an AP-reduction from a hard graph homomorphism counting problem. A homomorphism from a graph G to a graph H is a function $\sigma : V(G) \to V(H)$ such that, for all pairs of vertices $u, v \in V(G)$, it is the case that $\{u, v\} \in E(G)$ implies $\{\sigma(u), \sigma(v)\} \in E(H)$. We denote the set of all homomorphisms from G to H by $\mathrm{Hom}(G, H)$. For any fixed undirected graph H, the homomorphisms-to-H counting problem is defined as follows.

Problem: #Hom(H).

Instance: An undirected graph G.

Output: $|\mathrm{Hom}(G, H)|$, the number of homomorphisms from G to H.

In §2 we present a (deterministic) polynomial-time parsimonious reduction from #Hom(H) to #Vertices of a 0/1 Polytope, for the specific graph H_4 depicted in Figure 1. As #Hom(H_4) is known to be hard for #**P** under AP-reductions [4], Theorem 1 follows immediately, by transitivity of AP-reducibility.

As promised, here are some comments on the terminology used in, and the interpretation of, Theorem 1. In the literature on approximate counting, terminology such as "Π is #SAT-hard [under AP-reducibility]" appears. Since every problem in #**P** is parsimoniously reducible to #SAT (the problem of counting satisfying assignments of a CNF Boolean formula) these formulations are equivalent to the one used in Theorem 1.

At first sight, the conclusion in Theorem 1 looks rather weak. One might imagine that the existence of an FPRAS for #Vertices of a 0/1 Polytope should lead to something more dramatic than **RP** = **NP**, for example to **RP** = #**P** or the collapse of the polynomial hierarchy. However, Valiant and Vazirani [7] showed that every problem in #**P** can be approximated by a polynomial-time randomised Turing machine equipped with an **NP**-oracle. So whereas the complexity gap between existence and exact counting is expected to be huge, the gap is almost non-existent for approximate counting. For the same reason, the phrase "hard for **NP** under AP-reductions" used in [3] is equivalent to the formulation used Theorem 1, since (with a slight abuse of terminology) every problem in #**P** is AP-reducible to any **NP**-complete problem.

The standard definition of AP-reduction allows the algorithm performing the reduction to be randomised. However, the reductions used here are deterministic, so we can equally deduce that

there is no polynomial-time deterministic approximation algorithm (technically FPTAS) for #Vertices of a 0/1 Polytope unless $\mathbf{P} = \mathbf{NP}$. (The definition for FPTAS matches the one for FPRAS above, except that the word "randomised" is deleted, and the random variable Z becomes a deterministic value.) Of course, one needs to check that the reductions used to establish hardness results quoted in this paper are deterministic, but that is indeed the case.

A matrix A is *totally unimodular* (TU) if every square submatrix of A has determinant -1, 0 or 1. It is not difficult to verify that the vertices of the polytope defined by $Ax \leq b$ and $\mathbf{0} \leq x \leq \mathbf{1}$, where A is TU, lie in $\{0,1\}^n$. Many 0/1-polytopes appearing in combinatorial optimisation arise in this way. One such is the bipartite independent set polytope. It follows that, restricted to TU instances, approximating #Vertices of a 0/1 Polytope is as hard as approximating #BIS, the problem of counting independent sets in a bipartite graph. No FPRAS is known for #BIS and the same holds for the wide range of counting problems interreducible with #BIS via AP-reductions. At the same time, there is currently no evidence that #BIS is hard for #**P** under AP-reductions.

As well as observing that approximating #Vertices of a 0/1 Polytope can be as hard as #BIS, Guo and Jerrum [3] showed that the generalisation to polytopes with vertices in $\{0, \frac{1}{2}, 1\}^n$ is hard for #**P** under AP-reductions. Theorem 1 is a common strengthening of these two results.

Two prominent classes of TU matrices are *network matrices* and their transposes. As far as counting problems arising from transposes of network matrices are concerned we have a clear understanding: #BIS is an example in this class, and it is a hardest such under AP-reducibility. For network matrices our understanding is incomplete, and it is possible that all instances arising from these have an FPRAS [3]. We observe in the final section of this paper that the polytope $Q(G)$ used in the main reduction is not TU. Thus the approximation complexity of #Vertices of a 0/1 Polytope when restricted to TU instances remains open. It could be #BIS-equivalent or maximally hard (or somewhere between these).

A related conjecture by Mihail and Vazirani [5] is that the edge expansion of any 0/1-polytope is at least 1, which implies rapid mixing of the simple random walks on the graph of the polytope. Our hardness result does not directly invalidate this conjecture, since the degrees of vertices can be exponentially large in the dimension of the ambient space. Nonetheless, this conjecture is recently found false by Yang [8], who constructed 0/1-polytopes with exponentially small edge expansion.

1.1 Statement of AI use

The main proof ideas were found using GPT-5.6 Sol Ultra. In particular, it generated a hardness proof from #Ret(H_4), which is shown hard in [2]. We will first present a reduction from #Hom(H_4), which is simpler and the hardness of #Hom(H_4) is shown by Kelk [4] in his PhD thesis. The proofs are written or rewritten by the authors, who take responsibility of any error.

2 The reduction

Our reduction rests on the graph H_4 with vertex set $V(H_4) = \{00, 10, 01, 11\}$ depicted in Figure 1. Note that we identify every vertex of H_4 with a vector in $\{0,1\}^2$. For $c, d \in V(H_4)$, including the case $c = d$, we put $\{c, d\}$ in the edge set $E(H_4)$ if and only if

$$|c|_1 + |d|_1 \geq 2, \tag{1}$$

where $|c|_1$ is the Hamming weight of c. Thus 10, 01 and 11 form a reflexive triangle (i.e., a triangle with loops on all vertices), while 00 is an unlooped leaf adjacent only to 11.

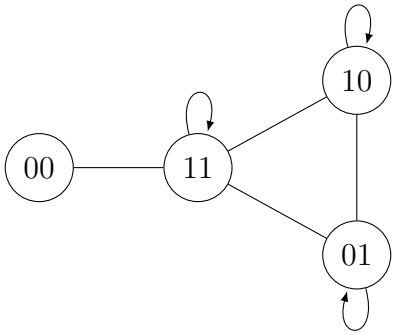

Figure 1: The graph H_4.

Lemma 2. *#Hom(H_4) is AP-reducible to #Vertices of a 0/1 Polytope.*

Proof. Let G be an instance of #Hom(H_4). We construct a polytope $Q(G)$ in a $2|V(G)|$-dimensional ambient space. Let $y \in \mathbb{R}^{V(G)\times\{1,2\}}$ be a vector, and for every $v \in V(G)$ write

$$s_v = y_{v,1} + y_{v,2}.$$

Define the polytope $Q(G) \subseteq \mathbb{R}^{2|V(G)|}$ as the set of vectors y satisfying

$$0 \le y_{v,i} \le 1, \quad \text{for all } v \in V(G) \text{ and } i \in \{1,2\}; \tag{2}$$
$$s_u + s_v \ge 2, \quad \text{for all } \{u,v\} \in E(G). \tag{3}$$

Note that (s_v) are auxiliary variables introduced for convenience in the proof, so (3) should be interpreted as a shorthand for

$$y_{u,1} + y_{u,2} + y_{v,1} + y_{v,2} \ge 2.$$

We aim to show that the vertices of $Q(G)$ are in bijection with homomorphisms from G to H_4. This gives the required AP-reduction from #Hom(H_4) to #Vertices of a 0/1 Polytope. The key observation is that every vertex of $Q(G)$ belongs to $\{0,1\}^{2|V(G)|}$.

Suppose, for a contradiction, that y is a vertex of $Q(G)$ with a fractional coordinate. We call $(y_{v,1}, y_{v,2})$ the *block* corresponding to v. We first claim that every fractional block has exactly one fractional coordinate. Indeed, suppose both coordinates in the block corresponding to v are fractional. For all sufficiently small $\delta > 0$, we can add δ to one coordinate and subtract δ from the other. This preserves s_v, and hence preserves all the inequalities in (3). Both directions of the perturbation satisfy the box constraints in (2). This contradicts the assumption that y is a vertex. The claim follows. In particular, s_v is non-integral whenever the block corresponding to v is fractional.

Let $F \subseteq V(G)$ be the set of vertices whose blocks are fractional. Define a graph T on vertex set F by putting $\{u,v\} \in E(T)$ if $\{u,v\} \in E(G)$ and the corresponding inequality is tight, namely

$$s_u + s_v = 2. \tag{4}$$

Note that a tight inequality cannot have exactly one endpoint in F, since the sum of an integer and a non-integer cannot be 2.

We claim that T is bipartite. Otherwise, let $v_0v_1\cdots v_{2k}v_0$ be an odd cycle in T. The equations in (4) imply

$$s_{v_0} = s_{v_2} = \cdots = s_{v_{2k}} \qquad \text{and} \qquad s_{v_{2k}} = 2 - s_{v_0}.$$

The final edge $v_{2k}v_0$ gives $2s_{v_0} = 2$. Hence $s_{v_0} = 1$, contradicting the fact that s_{v_0} is non-integral.

Now let C be a connected component of T, with bipartition $C = C^+ \cup C^-$. For every $v \in C$, perturb the unique fractional coordinate in its block by $+\delta$ if $v \in C^+$, and by $-\delta$ if $v \in C^-$. The opposite perturbation is obtained by reversing the signs. Every tight edge inequality incident with C has both endpoints in C and is preserved. Every other affected edge inequality has positive slack. Also, all the coordinates being changed lie strictly between 0 and 1. Thus both perturbations are feasible for all sufficiently small $\delta > 0$, again contradicting the assumption that y is a vertex. It follows that F is empty, and hence every vertex of $Q(G)$ is binary.

It remains to show that the reduction mapping G to $Q(G)$ is parsimonious. Observe that there is a natural bijection $\varphi : \{0,1\}^{V(G)\times\{1,2\}} \to V(H_4)^{V(G)}$ that interprets each binary vector $y \in \{0,1\}^{V(G)\times\{1,2\}}$ as a function $\sigma_y : V(G) \to V(H_4)$. Explicitly, for each vertex $v \in V(G)$ we define $\sigma_y(v) = w$, where w is the vertex of H_4 encoded by the pair $(y_{v,1}, y_{v,2})$.

In one direction, suppose that y is a vertex of $Q(G)$. We saw that y is binary vector, so $\varphi(y)$ is a function from $V(G)$ to $V(H_4)$. It is easy to see that the construction of $Q(G)$ — specifically the set of inequalities (3) — forces $\varphi(y)$ to be a homomorphism. This gives an injective map from vertices of $Q(G)$ to $\mathrm{Hom}(G, H_4)$. In the other direction, any homomorphism $\sigma \in \mathrm{Hom}(G, H_4)$ corresponds to a binary vector $y = \varphi^{-1}(\sigma)$ and the construction of $Q(G)$ ensures that $y \in Q(G)$. Any binary vector lying in a 0/1-polytope is a vertex of that polytope. Thus the function φ^{-1} is an injective map from $\mathrm{Hom}(G, H_4)$ to vertices of $Q(G)$. □

Proof of Theorem 1. H_4 is *Graph 39* in Kelk's PhD thesis [4]. On p.71 he shows that #SAT is AP-reducible to #Hom(H_4) and hence #Hom(H_4) is hard for #**P** under AP-reductions. This fact, combined with Lemma 2 and transitivity of AP-reductions, proves the main claim in Theorem 1.

For the subsidiary claim, take any NP-complete problem, say SAT, and consider its counting analogue #SAT. Suppose, for a contradiction, that #Vertices of a 0/1 Polytope has an FPRAS. Let Ψ (a CNF formula) be an instance of SAT and hence also of #SAT. By the first claim, there is an AP reduction from #SAT to #Vertices of a 0/1 Polytope. This, combined with the FPRAS for #Vertices of a 0/1 Polytope, yields an FPRAS for #SAT. Setting $\varepsilon = \frac{1}{2}$ in the definition of FPRAS, we have a randomised algorithm that, with error probability $\frac{1}{4}$, can decide whether Ψ has no satisfying assignments or a positive number. Thus we have a polynomial-time randomised algorithm for SAT with two-sided error probability. A standard technique (see e.g., [6, Ex 1.15]) strengthens this to a randomised algorithm with one-sided errors. So SAT $\in$ **RP** and **RP** = **NP**. □

A possible objection to the above line of argument is that it is based on a result from a non-peer-reviewed source. We can avoid this objection at the expense of a little extra work.

The retraction counting problem is the following.

Problem: #Ret(H).

Instance: An undirected graph G without loops, and a collection of sets $\mathbf{S} = \{S_v \subseteq V(H) : v \in V(G)\}$ such that, for all $v \in V(G)$, we have $|S_v| \in \{1, |V(H)|\}$.

Output: The number of homomorphisms σ from G to H such that, for all $v \in V(G)$, we have $\sigma(v) \in S_v$.

If we think of the vertices of H as being colours, and a function $\sigma : V(G) \to V(H)$ as being a colouring of the vertices of G, then $\#\textsc{Ret}(H)$ counts "H-colourings" of G in which certain vertices of G are forced to take on a specified colour.[1] For the particular H_4 in Figure 1, we have the following analogue of Lemma 2

Lemma 3. *$\#\textsc{Ret}(H_4)$ is AP-reducible to $\#\textsc{Vertices of a 0/1 Polytope}$.*

This is a slightly stronger version of the previous lemma. The advantage of this version is that the hardness proof for $\#\textsc{Ret}(H_4)$ appears in archival form [2]. The proof of Lemma 3 is a mild complication of the proof of Lemma 2.

Proof of Lemma 3. Given an instance $(G, \mathbf{S})$ of $\#\textsc{Ret}(H_4)$ we code retractions as vectors $y \in \mathbb{Q}^{V(G)\times\{1,2\}}$ as before. The polytope $Q(G, \mathbf{S})$ is defined by the same linear inequalities as in the earlier proof, together with some additional equalities coding the sets $S_v \in \mathbf{S}$. For example, the set $S_v = \{01\}$ is coded by the equalities $y_{v,1} = 0$ and $y_{v,2} = 1$. No additional equalities are added when $S_v = V(H_4)$.

Consider a vertex y of $Q(G, \mathbf{S})$. For the same reason as in the earlier proof, it cannot be the case that $y_{v,1}$ and $y_{v,2}$ are both fractional. The blocks and tight edge constraints induce a bipartite subgraph as before. Noting that none of these blocks correspond to vertices v with $|S_v| = 1$, the rest of the argument goes through as before. We deduce that the vertices of $Q(G, \mathbf{S})$ are in bijection with retractions to H_4. □

To complete the alternative proof of Theorem 1 we just need to compare H_4 against the complexity classification of Focke, Goldberg and Živný [2, Theorem 2]. The underlying simple graph of H_4 is a triangle with a pendant vertex. In particular, H_4 is connected and square-free. It is enough to check that H_4 is not in any of the easy or #BIS-equivalent cases in that theorem.

The graph H_4 contains both looped and unlooped vertices. Thus it is neither a reflexive clique nor an irreflexive complete bipartite graph. Also, since H_4 has loops, it is not an irreflexive caterpillar.

It remains to rule out the class $\mathcal{H}_{\text{BIS}}$. Suppose, for a contradiction, that $H_4 \in \mathcal{H}_{\text{BIS}}$. In the notation of [2, Definition 10], there is a positive integer m, reflexive cliques $K_0, \ldots, K_m$, and distinct looped vertices $p_0, \ldots, p_{m+1}$ such that $p_i, p_{i+1} \in K_i$ and

$$K_{i-1} \cap K_i = \{p_i\}.$$

Since H_4 has exactly three looped vertices, we must have $m = 1$. The two cliques K_0 and K_1 intersect only at p_1, so p_0p_2 is not an edge of H_4. This is a contradiction, since the three looped vertices of H_4 form a triangle.

The third case of the trichotomy applies, and $\#\textsc{Ret}(H_4)$ is #SAT-hard (and hence $\#\mathbf{P}$-hard) under approximation-preserving reductions.

[1] This not a "retraction" as usually defined, but equivalent in this context, as explained by Focke et al. [2].

3 Total unimodularity

Recall that a matrix is totally unimodular (TU) if every square submatrix has determinant -1, 0 or 1. It is not to difficult to see that the polytopes $Q(G)$ constructed in the main reduction do not necessarily correspond to TU constraint matrices. In particular, consider the linear inequalities defining the polytope $Q(C_3)$, where C_3 is the cycle on vertex set $\{u, v, w\}$. They include the inequalities

$$\begin{array}{rrrl} y_{u,1} + y_{u,2} & + y_{v,1} + y_{v,2} & & \geq 2 \\ y_{u,1} + y_{u,2} & & + y_{w,1} + y_{w,2} & \geq 2 \\ & y_{v,1} + y_{v,2} & + y_{w,1} + y_{w,2} & \geq 2 \end{array}$$

together with the box constraints. So, written in the form $Ay \leq b$, the matrix A contains the 3×6 submatrix

$$-\begin{pmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

which in turn contains the 3×3 submatrix

$$-\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

As this matrix has determinant 2, the matrix A is not totally unimodular.

It is possible to speculate that escaping from total unimodularity is critical to the proof of Theorem 1.

References

[1] Martin Dyer, Leslie Ann Goldberg, Catherine Greenhill, and Mark Jerrum. The relative complexity of approximate counting problems. *Algorithmica*, 38(3):471–500, 2004. 2

[2] Jacob Focke, Leslie Ann Goldberg, and Stanislav Živný. The complexity of approximately counting retractions to square-free graphs. *ACM Trans. Algorithms*, 17(3):22:1–22:51, 2021. 3, 6

[3] Heng Guo and Mark Jerrum. Counting vertices of integral polytopes defined by facets. *Discrete Comput. Geom.*, 70(3):975–990, 2023. 2, 3

[4] Steven Kelk. *On the relative complexity of approximately counting H-colourings.* PhD thesis, Warwick University, 2003. URL `https://skelk.sdf-eu.org/kelkthesis.pdf`. 2, 3, 5

[5] Milena Mihail and Umesh Vazirani. On the magnification of 0-1 polytopes. Technical Report TR-03-89, Harvard university, 1989. 3

[6] Rajeev Motwani and Prabhakar Raghavan. *Randomized Algorithms.* Cambridge University Press, 1995. 5

[7] L. G. Valiant and V. V. Vazirani. NP is as easy as detecting unique solutions. *Theoret. Comput. Sci.*, 47(1):85–93, 1986. 2

[8] Xiongxin Yang. 0/1-polytopes with exponentially small edge expansion. *arXiv*, abs/2608.01870, 2026. 3