Hardness of Learning Boolean Functions from Label Proportions

Venkatesan Guruswami
Department of EECS and Simons Institute for the Theory of Computing,
University of California, Berkeley, CA, USA

Rishi Saket
Google Research India, Banglaore, India

Abstract

In recent years the framework of learning from label proportions (LLP) has been gaining importance in machine learning. In this setting, the training examples are aggregated into subsets or *bags* and only the average label per bag is available for learning an example-level predictor. This generalizes traditional PAC learning which is the special case of unit-sized bags. The computational learning aspects of LLP were studied in recent works [21, 22] which showed algorithms and hardness for learning halfspaces in the LLP setting. In this work we focus on the intractability of LLP learning Boolean functions. Our first result shows that given a collection of bags of size at most 2 which are consistent with an OR function, it is NP-hard to find a CNF of constantly many clauses which *satisfies* any constant-fraction of the bags. This is in contrast with the work of [21] which gave a (2/5)-approximation for learning ORs using a halfspace. Thus, our result provides a separation between constant clause CNFs and halfspaces as hypotheses for LLP learning ORs.

Next, we prove the hardness of satisfying more than $1/2 + o(1)$ fraction of such bags using a t-DNF (i.e. DNF where each term has $\leq t$ literals) for any constant t. In usual PAC learning such a hardness was known [15] only for learning noisy ORs. We also study the learnability of parities and show that it is NP-hard to satisfy more than $(q/2^{q-1} + o(1))$-fraction of q-sized bags which are consistent with a parity using a parity, while a random parity based algorithm achieves a $(1/2^{q-2})$-approximation.

2012 ACM Subject Classification Theory of computation → Problems, reductions and completeness

Keywords and phrases Learning from label proportions, Computational learning, Hardness, Boolean functions

Digital Object Identifier 10.4230/LIPIcs.FSTTCS.2023.37

Funding *Venkatesan Guruswami*: Research supported in part by a Simons Investigator award and NSF grants CCF-2228287 and CCF-2211972.

1 Introduction

In common machine learning applications, one is required to train a classifier using some training set of (vectors, label)-pairs to predict the label of vectors sampled from the same (or a similar) distribution as the training set. A typical approach is to optimize the classifier to predict correctly on the training set to ensure that the classifier has good predictive performance over the target distribution. This optimization view is captured by the *probably approximately correct* (PAC) learning framework [23].

In setting of *learning from label proportions* (LLP), the training set consists of subsets or *bags* of vectors along with the sum or average of the labels of vectors in each bag. The goal is to train a model to predict the labels for vectors. As before, one would want the model to firstly predict as correctly as possible on the training bags. One measure of such performance is the fraction of *satisfied* bags i.e., those on which the predicted average label matches the given average label i.e., the label proportion. Note that traditional PAC learning is the special case of LLP with only unit-sized bags.

LLP is motivated by applications in which only the aggregated labels for bags of vectors are available. This may be to preserve the privacy [20, 24, 16] of labels, due to lack of instrumentation to obtain labels [9] or high labeling costs [8]. Other examples of LLP applications have been in medical image classification [14, 5, 17] where small bag sizes – in the range of 10 to 50 – are typically more relevant (see Sec 1.2 of [4]).

The work of [21] studied LLP from the computational learning perspective on bags of size ≤ 2. The LLP learning goal is the following: given a collection of bags consistent with some function from a target concept class, compute a hypothesis satisfying the most number of bags. With this objective, [21] showed a $(1/2 + o(1))$-factor hardness for LLP learning a halfspace using any function of constantly many halfspaces on bags of size at most 2. From the algorithmic side on such bags [21] gave a $(2/5)$-factor approximation for LLP learning a halfspace using halfspace, based on rounding a semi-definite programming (SDP) relaxation. Subsequently, [22] proved a strengthened $(4/9 + o(1))$-factor hardness for LLP learning a halfspace using any function of constantly many halfspaces on bags of size at most 2, a corresponding $(1/q + o(1))$-factor hardness for bags of size at most any constant $q \in \mathbb{Z}^+$, and extended the algorithmic result of [21] showing a $(1/12)$-approximation on bags of size at most 3.

Since halfspaces capture OR formulas, the algorithmic results of [21, 22] apply to learning OR formulas using halfspaces. Moreover, the $(1/2 + o(1))$-factor hardness of [21] on bags of size ≤ 2 also holds for LLP learning an OR using any function of constantly many halfspaces. Typically however, one would like to learn an OR using an OR or similar Boolean functions such as ℓ-clause CNF formulas (OR is 1-clause CNF), rather than halfspaces. This raises the following question

Can we achieve constant-factor algorithmic approximations for LLP learning OR *using* OR *or constant-clause* CNF*?*

In our first result, we answer the above question in the negative.

▶ **Theorem 1.** *For any constants $\delta > 0, \ell \in \mathbb{Z}^+$, given a collection of bags which are of size at most* 2*, and whose label proportions are consistent with some* OR*, it is NP-hard to compute an ℓ-clause* CNF *that satisfies δ-fraction of the bags.*

The above theorem is proved in Sec. 3. We find the result interesting since it (along with the algorithmic results of [21, 22]) proves a separation between constant clause CNFs – in particular ORs – and halfspaces as hypotheses for learning ORs.

We also study the LLP learnability of OR using as hypothesis ℓ-DNF formulas i.e., DNF where each term is a conjunction of at most ℓ literals. While OR is 1-DNF, for $\ell \geq 2$, ℓ-DNFs are not contained in halfspaces, therefore have the possibility of yielding better approximations. However, our second result below (proved in Sec. 4) essentially rules out this possibility.

▶ **Theorem 2.** *For any constants $\delta > 0, \ell \in \mathbb{Z}^+$, given a collection of bags which are of size at most* 2*, and whose label proportions are consistent with some* OR*, it is NP-hard to compute an ℓ-*DNF *that satisfies $(1/2 + \delta)$-fraction of the bags.*

Note that while the hardness factor achieved above is weaker than that of Theorem 1, no inapproximability is known for the real analogue of Theorem 2 i.e., LLP learning halfspaces using *polynomial thresholds*.

While the works of [21, 22] studied the LLP learnability of halfspaces, the corresponding problem over finite fields has not been studied. In particular, the $\mathbb{F}_2$-version of this problem is equivalent to the LLP learnability of parities using parities over the Boolean domain.

Parities are a fundamental class of Boolean functions which makes this problem of significant interest as well. Our next result however, shows that this is hard to approximate, with the inapproximability growing exponentially as the bag size increases.

▶ **Theorem 3.** *For any constants $\delta > 0, q \in \mathbb{Z}^+$ $(q \geq 2)$, given a collection of bags which are of size at most q, and whose label proportions are consistent with some parity, it is NP-hard to compute a parity that satisfies $(q/2^{q-1} + \delta)$-fraction of the bags.*

The above is proved in Section 5. The hardness factor is asymptotically close (for large q) to the $(1/2^{q-2})$-approximation for this problem using random parities described in Sec. 6.

1.1 Previous Related Work

The formalization of the LLP framework was first done in the work of [25] who proved generalization error bounds for classifiers for any distribution over (bag, label-proportion)-pairs, though their bag-level objective was a relaxed notion – useful for studying LLP with large bag sizes – of the strict bag satisfaction used in [21, 22] and our work. Related recent works [6, 7] have shown bag-to-instance classification generalization error bounds. The study of LLP learnability of specific function classes has nevertheless been fairly sparse, apart from the works of [21, 22] whose contributions have been described earlier in this section.

The learnability of small Boolean formulas has been extensively studied in traditional PAC learning. It is well known that an OR can be efficiently learnt by an OR up to arbitrary accuracy. On the other hand, [15] proved a $(1/2+o(1))$-factor hardness for learning a 2-clause CNF using constant clause CNF, and the same hardness factor for learning a noisy OR using ℓ-DNF for any constant ℓ. The work of [10] proved the same hardness for learning noisy OR with a halfspace as hypothesis. These results were further generalized by [11] who proved the same hardness factors for learning 2-clause CNF and noisy OR using any function of constantly many halfspaces as hypothesis. Similar to OR, parities can also be efficiently learnt by parities using Gaussian elimination over $\mathbb{F}_2$. On the other hand, the $(1/2 + o(1))$-factor hardness for noisy Max-3-Lin by [13] implies the same hardness factor for learning a noisy parity using a parity. Note that all the $(1/2+o(1))$-factor hardness results are tight since one of the constant 0 or 1 functions trivially obtain $(1/2)$-approximation for learning Boolean valued functions. However, this trivial threshold does not hold in the LLP setting since the constant functions are not guaranteed to satisfy even one bag.

The above hardness results carry over to the LLP setting for the special case when all bags are unit-sized. However, we prove hardness of approximating the problems of LLP learning OR and parity *without* any noise, which are tractable in the usual PAC case, thereby showing a qualitative difference between the LLP and PAC settings.

1.2 Overview of Our Techniques

Proof of Theorem 1. Our reduction is from bipartite Label-Cover [1] with N and M as the sizes of the smaller and larger label sets respectively, and is similar to that of [15] for the hardness of learning noisy OR. The high-level approach is to have one coordinate for each vertex-label pair on the larger (right) side of the Label-Cover instance, i.e. the variables are $x_{v,i}$ for $v \in V$ and $i \in [M]$. Fix a random sample of $2t$ vertices $\{\hat{v}_1, \ldots, \hat{v}_t, \tilde{v}_1, \ldots, \tilde{v}_t\}$ from a neighborhood of a left vertex u. For simplicity assume that the projection constraints between u and each of the $2t$ vertices are the same i.e, for each label $j \in [N]$ for u there is a subset $S_j \subseteq [M]$ such that assigning any of the $2t$ vertices with a label from S_j satisfies that edge.

A 2-sized bag with label proportion 1/2 is sampled by letting $J \subseteq [N]$ be a random subset, and for the first point $\mathbf{x}$ setting only the coordinates $\{x_{\hat{v}_r,i} \mid \pi_{\hat{v}_r u}(i) \in J, r \in [t]\}$ to be 1, and for the second point $\mathbf{z}$ only the coordinates $\{z_{\tilde{v}_r,i} \mid \pi_{\tilde{v}_r u}(i) \in \overline{J}, r \in [t]\}$ to be 1. It is easy to see in the YES case that an OR of exactly the coordinates $(v, \rho(v))$ – where ρ is the satisfying labeling – for each right vertex v, satisfies all such bags. For the NO case, we illustrate the analysis of an OR formula $\mathcal{C}$ which satisfies some constant fraction of the bags. From the $o(1)$-Hamming weight of the points, one can assume that $\mathcal{C}$ has no negated literals. If $\mathcal{C}$ has no coordinates of the $2t$ vertices (empty case) then the bag is anyway not satisfied as $\mathcal{C}$ evaluates to 0 on both points.

On the other hand, if a sufficiently large number of these vertices have a corresponding variable in $\mathcal{C}$ (dense case), then elementary probabilistic arguments yield at least two vertices among the $2t$ which have their pre-decided distinguished variables in $\mathcal{C}$ with the same projection, leading to a good randomized labeling to the Label-Cover. A key idea for ensuring that this analysis goes through is to sample t u.a.r. from $\{1, \ldots, 2^T\}$ for some large T so that for each u and most values of t, nearly all samples of the $2t$ vertices are either the empty case or the dense case.

Proof of Theorem 2. The hardness reduction is similar to the above, except that we need to ensure that a significant fraction of vertices have at least one term in which all the positive literals correspond to its coordinates, while none of the negated literals do. Thereafter a similar analysis as the previous case goes through. However, for ensuring this property we introduce an additional distribution over bags of size 1 and label proportion 1, essentially saying that for each vertex the point with all its coordinates to 1 and the rest to 0 should be 1-labeled. Due to this we obtain a $(1/2 + o(1))$-factor hardness in this case.

Proof of Theorem 3. While the reductions above create points in a bag whose active coordinates span the edges of the label-cover, in the parity case we can add homogeneous $\mathbb{F}_2$-linear *folding* constraints which ensure consistency of labels across edges via a reduction from the non-bipartite *Smooth* Label-Cover [12]. It is sufficient to then describe a *dictatorship test* (see Chap. 7 of [18]) on the M coordinates of a single vertex. Our dictatorship test is a distribution over q-sized bags, i.e., q points in the hypercube $\mathbb{F}_2^M$, sampled as follows: independently for each $i \in M$, set exactly one of the q points to 1 in the i'th coordinate and the rest to 0. All these bags have target label proportion $1/q$ which is satisfied by any dictator, i.e., parity given by a single coordinate. On the other hand, any parity over a much larger number K of coordinates will induce a near-uniform distribution over the q-sized vector of labeling to the points of a random bag. In fact, this is close to the uniform distribution over the points of $\mathbb{F}_2^q$ with even or odd (depending only on q and K) number of non-zero coordinates. It is then easy to see that such a distribution will satisfy the $1/q$ label proportion of the bags with probability $\approx q/2^{q-1}$.

The algorithm for this problem first does Gaussian elimination for all the linear constraints given by bags of label proportion $\{0, 1\}$ and obeying the bag-level parity constraints for the remaining bags. It then chooses a random parity from the remaining coordinates. We show that this yields an $1/2^{q-2}$ approximation.

2 Preliminaries

2.1 Problem Definitions

Consider the space $\{0,1\}^d$ for some $d \in \mathbb{Z}^+$ and some function $f : \{0,1\}^d \to \{0,1\}$. For $B \subseteq \{0,1\}^d$, define $\sigma(B, f) := |\{\mathbf{x} \in B \mid f(\mathbf{x}) = 1\}| / |B|$ to be the corresponding label proportion.

An instance $\mathcal{I}$ of LLP-OR$[q]$ is given by a collection $\mathcal{B} := \{(B_j, \sigma_j)\}_{j=1}^m$ of bags and their label proportions where each bag is of size at most q. The goal is to find an OR function $h(\mathbf{x})$ which satisfies the most bags of $\mathcal{B}$ i.e., maximize $|\{j \in [m] \mid \sigma_j = \sigma(B_j, h)\}|$.

An instance $\mathcal{I}$ of LLP-PARITY$[q]$ is similar to the above except that the goal accordingly is to compute a parity maximizing the number of satisfied bags. Since the XOR is simply an addition over $\mathbb{F}_2$ we shall think of the Boolean values as elements of $\mathbb{F}_2$ in this case.

2.2 Label Cover

▶ **Definition 4.** *An instance of $\mathcal{L}$ of* Label-Cover *is given by $(G(U, V, E \subseteq V \times U), M, N, \{\pi_{vu} : [M] \to [N]\}_{e=(v,u)\in E})$ where $G(U, V, E)$ is a bi-regular bipartite graph. A labeling ρ assigning labels from $[N]$ to U and $[M]$ to V satisfies an edge (u, v) iff $\pi_{vu}(\rho(v)) = \rho(u)$. The goal is to find a labeling satisfying the most number of edges.*

The following well-known inapproximability of Label-Cover follows from the the PCP Theorem [3, 2] along with the Parallel Repetition Theorem [19].

▶ **Theorem 5.** *For any constant $\xi > 0$ there exist M and N such that it is NP-hard, given an* Label-Cover *instance $\mathcal{L}(U, V, E, M, N, \{\pi_{vu}\}_{e=(v,u)\in E})$ to distinguish between: (i)* YES case: *there is labeling that satisfies all edges in E, or (ii)* NO case: *Any labeling satisfies at most ξ fraction of the edges of E.*

2.3 Smooth Label Cover

Unlike the standard bipartite version in Definition 4 we also use a non-bipartite version with useful structural properties defined below.

▶ **Definition 6.** *An instance of* Smooth-Label-Cover *$\mathcal{L}(G(V, E), N, M, \{\pi_{ev} \mid e \in E, v \in e\})$ consists of a regular connected (undirected) graph $G(V, E)$ with vertex set V and edge set E. Every edge $e = (v_1, v_2)$ is associated with projection functions $\{\pi_{ev_i}\}_{i=1}^2$ where $\pi_{ev_i} : [M] \to [N]$. A vertex labeling is a mapping defined on $\rho : V \to [M]$. A labeling ρ satisfies edge $e = (v_1, v_2)$ if $\pi_{ev_1}(\rho(v_1)) = \pi_{ev_2}(\rho(v_2))$. The goal is to find a labeling which satisfies the maximum number of edges.*

The following theorem is proved in Appendix A of [12].

▶ **Theorem 7.** *There exists a constant $c_0 > 0$ such that for any constant integer parameters $Q, R \geq 1$, it is* NP*-hard to distinguish between the following two cases for a* Smooth Label Cover *instance $\mathcal{L}(G(V, E), N, M, \{\pi_{ev} \mid e \in E, v \in e\})$ with $M = 7^{(Q+1)R}$ and $N = 2^R 7^{QR}$:*

- (YES Case) *There is a labeling that satisfies every edge.*
- (NO Case) *Every labeling satisfies less than a fraction $2^{-c_0 R}$ of the edges.*

In addition, the instance $\mathcal{L}$ satisfies the following properties:

- (Smoothness) *For any vertex $w \in V$, $\forall i, j \in [M]$, $i \neq j$, $\Pr_{e\sim w}[\pi_{ew}(i) = \pi_{ew}(j)] \leq 1/Q$, where the probability is over a randomly chosen edge incident on w.*
- (Weak Expansion) *For any $\delta > 0$, let $V' \subseteq V$ and $|V'| = \delta \cdot |V|$, then the number of edges among the vertices in $|V'|$ is at least $\delta^2|E|$.*

3 Hardness of LLP Learning OR using ℓ-clause CNF

We prove the following hardness reduction which along with Theorem 5 implies Theorem 1.

▶ **Theorem 8.** *For any constants $\delta > 0$ and $\ell \in \mathbb{Z}^+$, there is a polynomial time reduction from an instance $\mathcal{L}$ of* Label-Cover *to an* LLP-OR[2] *instance $\mathcal{B}$ s.t.*
YES Case: *If $\mathcal{L}$ is YES instance then there is an* OR *consistent with all the bags of $\mathcal{B}$.*
NO Case: *If $\mathcal{L}$ is a NO instance then there is no ℓ-clause* CNF *formula satisfying at least δ-fraction of the bags.*

We begin with the following useful technique. Let $T \geq 10$ be a large integer and consider the set $\mathcal{T} = \{2, 4, \ldots, 2^T\}$. For any $s \in \mathbb{R}^+$ define the subsets $L(s), R(s) \subseteq \mathcal{T}$ as

$$L(s) := \{t \in \mathcal{T} \mid t \leq s/\sqrt{T}\} \quad \text{and} \quad R(s) := \{t \in \mathcal{T} \mid t \geq s \cdot \sqrt{T}\}. \tag{1}$$

We have the following simple lemma:

▶ **Lemma 9.** *For any $s \in \mathbb{R}^+$, $L(s) \cap R(s) = \emptyset$ and $|L(s)| + |R(s)| \geq |\mathcal{T}| - 2\log T$.*

Proof. Since $T > 1$, $L(s) \cap R(s) = \emptyset$ by definition. Let $t' \in \mathcal{T}$ be the smallest element which is larger than $s/\sqrt{T}$ and $t'' \in \mathcal{T}$ be the largest element smaller than $s\sqrt{T}$. By definition, we have that $\mathcal{T} \setminus (L(s) \cup R(s)) = [t', t''] \cap \mathcal{T}$. Note also that $t''/t' \leq T$, and thus $|[t', t''] \cap \mathcal{T}| \leq \log T + 1 \leq 2\log T$ since $T \geq 10$, which completes the proof. ◀

3.1 Hardness Reduction

The hardness reduction is from a Label-Cover instance $\mathcal{L}(U, V, E \subseteq V \times U, M, N, \{\pi_{vu} : [M] \to [N]\}_{e=(v,u)\in E})$. We shall use $\mathcal{T}$ as defined above for some large enough choice of T depending on δ and ℓ. Note that T is a constant compared to $|V|$ which is an increasing value. The underlying space of the vectors is $\mathcal{X} = \{0,1\}^{V\times[M]}$ i.e., a vector $\mathbf{x} \in \mathcal{X}$ is given by $\mathbf{x} = (x_{v,i})_{v\in V, i\in[M]}$. The reduction yields a distribution $\mathcal{D}_{\mathcal{B}}$ over 2-sized bags and all bags B in its support have the label proportion $\sigma = 1/2$. A random bag from $\mathcal{D}_{\mathcal{B}}$ is given by the following steps:

1. Sample t uniformly at random from $\mathcal{T}$.
2. U.a.r. sample a vertex $u \in U$.
3. Independently and u.a.r. sample vertices $V_u^x = \{\hat{v}_1, \ldots, \hat{v}_t\}$ and $V_u^z = \{\tilde{v}_1, \ldots, \tilde{v}_t\}$ from the neighborhood $N(u)$ of u in V.
4. Randomly sample $J \subseteq [N]$, and let $\overline{J} = [N] \setminus J$.
5. Define a point $\mathbf{x} \in \mathcal{X}$ as follows. For each $i \in [M]$ and $v \in V$ set:

$$x_{v,i} = \begin{cases} 1 & \text{if } v \in V_u^x \text{ and } \pi_{vu}(i) \in J \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

6. Define another point $\mathbf{z} \in \mathcal{X}$ as follows. For each $i \in [M]$ and $v \in V$ set:

$$z_{v,i} = \begin{cases} 1 & \text{if } v \in V_u^z \text{ and } \pi_{vu}(i) \in \overline{J} \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

7. Output $(B = \{\mathbf{x}, \mathbf{z}\}, \sigma_B = 1/2)$.

In particular, observe that the points $\mathbf{x}$ and $\mathbf{z}$ are zero outside of the coordinates corresponding to the vertices in $V_u^x \cup V_u^z$.

YES Case. Consider a labeling ρ to the vertices of $\mathcal{L}$ that satisfy all the edges. Define the OR, $h^*(\mathbf{x}) = \bigvee_{v\in V} x_{v,\rho(v)}$. Let u be the choice in Step 2 above, and assume that $\rho(u) \in J$ as chosen in Step 4. Then, we know that for all $v \in V_u^x \cup V_u^z$ (as chosen in Step 3), $\pi_{vu}(\rho(v)) = \rho(u) \in J$. By construction of $\mathbf{x}$ and $\mathbf{z}$ therefore, $h^*(\mathbf{x}) = 1$ and $h^*(\mathbf{z}) = 0$, and thus B is satisfied by h^*. Similarly, when $\rho(u) \in \overline{J}$, we obtain that $h^*(\mathbf{x}) = 0$ and $h^*(\mathbf{z}) = 1$.

3.2 NO Case

Assume for a contradiction an ℓ-clause CNF formula h' s.t. $\Pr_{B \leftarrow \mathcal{D}_{\mathcal{B}}}[h' \text{ satisfies } B] \geq \delta$. From the bi-regularity of $\mathcal{L}$, if any clause of h' contains a negated literal then with probability at least $1 - 2t/|V|$ the literal's coordinate is not from those of the vertices in $V_u^x \cup V_u^z$ and the clause evaluates to 1 on both points of a random bag $B \leftarrow \mathcal{D}_{\mathcal{B}}$. Removing all such clauses we obtain r-clause CNF $h = C_1 \wedge \cdots \wedge C_r$ $(r \leq \ell)$ that satisfies at least $\delta - 2t\ell/|V| \geq \delta/2$ fraction of the bags (since we can take $|V| \gg 2^{T+2}\ell/\delta$ as $|V|$ is super-constant). We have the following lemma.

▶ **Lemma 10.** *For any constant $\zeta > 0$, there is a choice of $T = T(\zeta)$ s.t. for any C_i $(i \in [r])$,* $\Pr_{B=(\mathbf{x},\mathbf{z}) \leftarrow \mathcal{D}_{\mathcal{B}}}[C_i(\mathbf{x}) \neq C_i(\mathbf{z})] \leq \zeta$.

Using the above lemma, the NO case proof can be completed by taking $\zeta = \delta/(6\ell)$. By a union bound, the probability that any one of $C_1, \ldots, C_r$ evaluates differently on $\mathbf{x}$ and $\mathbf{y}$ is at most $\delta/6$. This also upper bounds the probability of satisfying the bags of $\mathcal{D}_{\mathcal{B}}$, contradicting our assumption.

Proof of Lemma 10. Fix any clause $C \in \{C_1, \ldots, C_r\}$, and from our construction above C has no negated literals. Call a vertex which has at least one variable from C as *non-empty*, otherwise call it *empty*. For each non-empty vertex v arbitrarily choose i_v such that the (v, i_v)-th variable is in C. For each $u \in U$ let $\mu(u)$ denote the fraction of its neighbors which are non-empty. In our analysis below we shall be collecting the *error* probabilities using

$$\Pr[A] \leq \Pr[A \cap B] + \Pr\left[\bar{B}\right], \text{ and } \Pr[A \cap B] \leq \min\{\Pr[A], \Pr[A \mid B]\} \tag{4}$$

for any two events A and B, where $\bar{B}$ denotes the complement of B.

We will first bound γ which we define to be the probability over the choice of t, u, V_u^x and V_u^z that there is a pair of non-empty vertices $v, v' \in V_u^x \cup V_u^z$ s.t. $\pi_{vu}(i_v) = \pi_{v'u}(i_{v'})$. We call this event Ψ. In this case we can construct a randomized partial labeling ρ for the vertices of $\mathcal{L}$ as follows: for each $v \in V$ which is non-empty, assign it the label i_v defined above. For each u, select a random neighbor v_u and assign u the label $\pi_{v_u u}(i_{v_u})$ if v_u is non-empty. Since $t \leq 2^T$ we obtain that this randomized labeling satisfies in expectation at least $\max_t \gamma/(2t)^2 = \gamma/\left(2^{2(T+1)}\right)$ fraction of the edges. Choosing the soundness of $\mathcal{L}$ to be small enough one can ensure that $\gamma \leq \zeta/100$.

We consider two the cases for t and u below, and in each of them (setting $B = \Psi$ in (4)) we can assume that Ψ does not occur, while incurring at most $\zeta/100$ probability error.

Case I: $t \in L(1/\mu(u))$. In this case the probability that there is non-empty vertex in $V_u^x \cup V_u^z$ is at most $2t\mu(u) \leq 2\left(1/\sqrt{T}\right)(1/\mu(u))\,\mu(u) \leq 2/\sqrt{T}$. Therefore, except with probability at most $2/\sqrt{T}$, C evaluates to zero on $\mathbf{x}$ and $\mathbf{z}$.

Case II: $t \in R(1/\mu(u))$. We shall first show that w.h.p. V_u^x contains a significant number of non-empty vertices, and given that happens, w.h.p. C evaluates to 1 on $\mathbf{x}$. The expected number of non-empty vertices in V_u^x is $t\mu(u) \geq \left(\sqrt{T}/\mu(u)\right)\mu(u) \geq \sqrt{T}$. Therefore, by the Chernoff bound (see Appendix A), except with probability $\exp\left(-\sqrt{T}/8\right)$, the number of non-empty vertices is at least $\sqrt{T}/2$. From the assumption that Ψ does not occur, for all pairs of non-empty vertices $v, v' \in V_u^x$, $\pi_{vu}(i_v) \neq \pi_{v'u}(i_{v'})$. Thus, each $\{x_{v,i_v} \mid v \in V_u^x, v \text{ is non-empty }\}$ is independently set to 1 with probability 1/2. In particular, except with probability $(1/2)^{\sqrt{T}/2}$, at least one of $\{x_{v,i_v} \mid v \in V_u^x, v \text{ is non-empty }\}$ is set to 1 and thus C evaluates to 1 on $\mathbf{x}$. The same argument as above also works for V_u^z and $\mathbf{z}$.

From the analysis of the above two cases, using Lemma 9 and repeated applications of (4) to add up the error probabilities above (for V_u^z and $\mathbf{z}$ as well in Case II) we obtain that:

$$\Pr\left[C(\mathbf{x}) \neq C(\mathbf{z})\right] \leq 2/\sqrt{T} + \zeta/100 + 2\left(\frac{\log T}{T} + \exp\left(-\sqrt{T}/8\right) + 2^{-\sqrt{T}/2}\right) \tag{5}$$

Choosing T to be $100/\zeta^2$ we can ensure that the above probability is at most ζ. ◀

4 Hardness of LLP Learning OR using ℓ-DNF

This section proves the following hardness reduction which implies Theorem 2.

▶ **Theorem 11.** *For any constants $\delta > 0$ and $\ell \in \mathbb{Z}^+$, there is a polynomial time reduction from an instance $\mathcal{L}$ of* Label-Cover *to an* LLP-OR[2] *instance $\mathcal{B}$ s.t.*
YES Case: *If $\mathcal{L}$ is YES instance then there is an* OR *consistent with all the bags of $\mathcal{B}$.*
NO Case: *If $\mathcal{L}$ is a NO instance then there is no ℓ-*DNF *formula satisfying at least $(1/2+\delta)$-fraction of the bags.*

4.1 Hardness Reduction

The setup is the same as in the previous section. The reduction outputs a distribution $\mathcal{D}_{\mathcal{B}}$ over two types of bags with equal probability: (i) the first type has size 1 with label proportion 1, (ii) the second type has size 2 and label proportion 1/2. With T being a large enough constant to be chosen later, the following steps define a random bag of $\mathcal{D}_{\mathcal{B}}$.

1. U.a.r. sample a vertex $u \in U$.
2. With probability 1/2 do the following:
   a. U.a.r. sample vertex $v \in N(u)$ and create the point $\overline{\mathbf{x}}^{(v)}$ as follows: set all coordinates $\{\overline{x}_{v,i}^{(v)}\}_{i=1}^{M}$ to 1, and set all the other coordinates to 0.
   b. Output $\left(B = \{\overline{\mathbf{x}}^{(v)}\}, \sigma_B = 1\right)$.
3. With the remaining probability 1/2 do the following:
   a. Independently and u.a.r. sample vertices $V_u^x = \{\hat{v}_1, \ldots, \hat{v}_T\}$ and $V_u^z = \{\tilde{v}_1, \ldots, \tilde{v}_T\}$ from the neighborhood $N(u)$ of u in V.
   b. Randomly sample $J \subseteq [N]$, and let $\overline{J} = [N] \setminus J$.
   c. Define a point $\mathbf{x} \in \mathcal{X}$ as follows. For each $i \in [M]$ and $v \in V$ set:

$$x_{v,i} = \begin{cases} 1 & \text{if } v \in V_u^x \text{ and } \pi_{vu}(i) \in J \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

   d. Define another point $\mathbf{z} \in \mathcal{X}$ as follows. For each $i \in [M]$ and $v \in V$ set:

$$z_{v,i} = \begin{cases} 1 & \text{if } v \in V_u^z \text{ and } \pi_{vu}(i) \in \overline{J} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

   e. Output $(B = \{\mathbf{x}, \mathbf{z}\}, \sigma_B = 1/2)$.

YES Case. This is easy to see using the same OR formula h^* defined in the previous section. Using the same arguments h^* satisfies all the bags of size 2 from $\mathcal{D}_{\mathcal{B}}$. Further, h^* has exactly one (positive) literal from the coordinates corresponding to each $v \in V$ so that $h^*(\overline{\mathbf{x}}^{(v)}) = 1$ where $\overline{\mathbf{x}}^{(v)}$ is as defined in Step 2a. of $\mathcal{D}_{\mathcal{B}}$. Thus, h^* satisfies all the bags of size 1 as well.

4.2 NO Case

Let us assume that there is an ℓ-DNF h that satisfies $1/2+\delta$ fraction of the bags of $\mathcal{D}_{\mathcal{B}}$. First, observe that if h has a term consisting only of negated literals, then from the bi-regularity of $\mathcal{L}$ and a union bound that term will not have any coordinate from among the vertices chosen in Step 3a with probability at least $1-2T\ell/|V|$. Thus, h will have label proportion 1 on at least $1-2T\ell/|V|$ fraction of the 2-sized bags i.e., h will not satisfy them, implying that the maximum fraction of bags satisfied by h is $1/2+T\ell/|V|$. This is a contradiction since $|V|=\omega(T\ell)$ and can be taken to be large enough. Thus, we may assume that h does not have a term of only negated literals.

Before proceeding, let us call a $v \in V$ as *non-empty* if h has a term in which all the positive literals correspond to v and none of the negated literals correspond to v. Let $\Gamma(v)$ be an arbitrary map from each non-empty v to one such term corresponding to it. Note that $\Gamma(v)$ is injective. Further, define $\Delta(v)$ to be the set of all indices $i \in [M]$ such that the positive literal corresponding to (v,i) occurs in $\Gamma(v)$.

Define for each $u \in U$, κ_u to be the probability that the choice of V_u^x and V_u^z in Step 3a satisfies that there exists a pair $v, v' \in V_u^x \cup V_u^z$ of non-empty vertices such that $\pi_{vu}(\Delta(v)) \cap \pi_{v'u}(\Delta(v')) \neq \emptyset$ (call these *intersecting non-empty* pair of vertices). Using this, let us define a randomized labeling ρ for the vertices in $\mathcal{L}$ as follows: for each non-empty $v \in V$, $\rho(v)$ is chosen u.a.r from $\Delta(v)$, and for each $u \in U$, v_u is chosen u.a.r. from $N(u)$ and then if v_u is non-empty $\rho(u)$ set to i which is chosen u.a.r. from $\Delta(v_u)$. By standard arguments, this labeling satisfies in expectation at least $\kappa_u/(4\ell^2T^2)$ fraction of the edges incident on u, and thus overall in expectation $\mathbb{E}_u[\kappa_u]/(4\ell^2T^2)$ fraction of edges of $\mathcal{L}$. Choosing the soundness of $\mathcal{L}$ to be small enough (and since T is a constant), we can assume that $\mathbb{E}_u[\kappa_u] \leq \delta^2/200$.

By averaging, for at least $\delta/2$ fraction of $u \in U$ (call them *good*) h satisfies at least $1/2+\delta/2$ fraction of the bags from $\mathcal{D}_{\mathcal{B}}|_u$ i.e., given the choice of u in Step 1. In particular from the above,

$$\mathbb{E}_u[\kappa_u | u \text{ good }] \leq \delta/100. \tag{8}$$

For any such good u, h must satisfy at least δ fraction of the bags of size 1 and label proportion 1 (as they constitute exactly half of the bags), implying that for at least δ-fraction of $v \in N(v)$, $h\left(\overline{\mathbf{x}}^{(v)}\right)=1$. Clearly, these vertices are non-empty for h to evaluate to 1 on them, and thus any good u has at least δ-fraction non-empty v in $N(u)$.

Let us fix on one such good u. First, from above we can assume by adding an error probability of κ_u that in Step 3a, $V_u^x \cup V_u^z$ does not contain an intersecting non-empty pair. Further, the probability that any term in $\{\Gamma(v) \mid v \in V_u^x \cup V_u^z \text{ and non-empty}\}$ contains a negated literal corresponding to vertices in $V_u^x \cup V_u^z$ is at most $2T\ell/d_U$ where d_U is the uniform degree on U. Since d_U can be taken to be an arbitrarily large constant, possibly by replicating V, we can assume by adding an error probability of $\delta/100$ that this event does not occur.

Given the above, we will show that w.h.p over the choice of a 2-sized bag, h evaluates to 1 on both $\mathbf{x}$ and $\mathbf{z}$. We shall prove this for $\mathbf{x}$, the argument for $\mathbf{z}$ is analogous and the conjunction is obtained by a union bound. Since u is good, the expected number of non-empty vertices in V_u^x is at least δT. By Chernoff bound, by adding an error probability of $\exp(-\delta T/8)$ we can assume there are at least $\delta T/2$ non-empty vertices in V_u^x. By our assumptions above, each term in $\{\Gamma(v) \mid v \in V_u^x \cup V_u^z \text{ and non-empty}\}$ independently evaluates to 1 w.p. at least $(1/2)^\ell$ over the choice of $\mathbf{x}$. Thus, the probability that none of them evaluate to 1 is at most $(1-(1/2)^\ell)^{\delta T/2}$. This analysis can be repeated for V_u^z and $\mathbf{z}$.

Summing up the error probabilities (using repeated applications of (4)), we obtain that:

$$\Pr\left[h(\mathbf{x}) \neq h(\mathbf{z})\right] \leq \kappa_u + \delta/100 + 2\left(\exp(-\delta T/8) + (1-(1/2)^{\ell})^{\delta T/2}\right) \leq \kappa_u + \delta/50 \qquad (9)$$

for a good u, using an appropriate choice of $T = O\left(2^{\ell}\log(1/\delta)/\delta\right)$. By (8), the average of the LHS of (9) over all good u is at most $3\delta/100$, which means that h satisfies on an average at most $1/2 + 3\delta/100$ bags corresponding to a random choice of good u. This is a contradiction to the definition of good u thus completes the NO case analysis.

5 Hardness of LLP Learning Parities

This section is devoted to proving the following hardness reduction which, along with the inapproximability of Smooth-Label-Cover (Th. 7) proves Theorem 3.

▶ **Theorem 12.** *For any constants $\delta > 0$ and $q \in \mathbb{Z}^+$, $q \geq 2$, there is a polynomial time reduction from an instance $\mathcal{L}$ of* Smooth-Label-Cover *to an* LLP-PARITY$[q]$ *instance $\mathcal{B}$ s.t.*
YES Case: *If $\mathcal{L}$ is YES instance then there is a parity consistent with all the bags of $\mathcal{B}$.*
NO Case: *If $\mathcal{L}$ is a NO instance then there is no parity satisfying at least $(q/2^{q-1}+\delta)$-fraction of the bags.*

We begin with the dictatorship test below using which the hardness reduction is described and analyzed in Sec. 5.2.

5.1 Dictatorship Test

Consider a large $M \in \mathbb{Z}^+$, and the space of vectors $\mathbb{F}_2^M$. For some integer $q > 1$, let he dictatorship test distribution $\mathcal{D}^{\mathsf{dict}}_{M,q}$ on (B, σ) be as follows:

1. Choose $B = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(q)}\right\}$ where for each $i \in [M]$, $\left(x_i^{(1)}, \ldots, x_i^{(q)}\right)$ is sampled u.a.r. from $\left\{\mathbf{e}^{(j)}\right\}_{j=1}^{q}$ where $\mathbf{e}^{(j)} \in \mathbb{F}_2^q$ is the jth coordinate vector.
2. Output $(B, \sigma = 1/q)$.

We prove the following lemma summarizing the completeness and soundness of t $\mathcal{D}^{\mathsf{dict}}_{M,q}$.

▶ **Lemma 13.** *The distribution $\mathcal{D}^{\mathsf{dict}}_{M,q}$ satisfies the following properties:*
Completeness: *For any $i \in [M]$ the parity function $h_i(\mathbf{x}) := x_i$ has the property that $\sigma(B, h_i) = 1/q$ for any B in the support of $\mathcal{D}^{\mathsf{dict}}_{M,q}$ i.e., h_i satisfies all bags of $\mathcal{D}^{\mathsf{dict}}_{M,q}$.*
Soundness: *Let $h(\mathbf{x}) := c_0 \oplus \bigoplus_{i=1}^{M} c_i x_i$ be such that $|\{i \in [M] \mid c_i = 1\}| = K \leq M$. Then, h satisfies a random $(B, \sigma) \sim \mathcal{D}^{\mathsf{dict}}_{M,q}$ with probability at most $q/2^{q-1} + \exp\left(-2K/q + q/2\right)$.*

Proof. The completeness follows from construction, since for any $(B, \sigma) \sim \mathcal{D}^{\mathsf{dict}}_{M,q}$, where $B = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(q)}\right\}$, $\left(h_i\left(x_i^{(j)}\right)\right)_{j=1}^{q} = \left(x_i^{(j)}\right)_{j=1}^{q} \in \left\{\mathbf{e}^{(j)}\right\}_{j=1}^{q}$ for all $i \in [M]$. Therefore, $\sigma(B, h_i) = 1/q = \sigma$, and therefore h_i satisfies all the bags, for all $i \in [M]$.

The proof of the soundness is given next in Sec. 5.1.1. ◀

5.1.1 Soundness of $\mathcal{D}^{\mathsf{dict}}_{M,q}$

Let $I_h = \{i \in [M] \mid c_i = 1\}$ so that $|I_h| = K$. Letting $B = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(q)}\right\}$ be a random bag from $\mathcal{D}^{\mathsf{dict}}_{M,q}$, for convenience define the random variable $Z_j := h\left(\mathbf{x}^{(j)}\right)$ for $j \in [q]$ and let $\mathcal{D}_Z$ be the distribution on $(Z_1, \ldots, Z_q)$. Since any (B, σ) in the support of $\mathcal{D}^{\mathsf{dict}}_{M,q}$ has $\sigma = 1/q$, using the construction of $B = \left\{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(q)}\right\}$ we obtain

$$\bigoplus_{j=1}^{q} Z_j = \bigoplus_{j=1}^{q} h\left(\mathbf{x}^{(j)}\right) = \bigoplus_{j=1}^{q}\left(c_0 \oplus \bigoplus_{i=1}^{M} c_i x_i^{(j)}\right) = \bigoplus_{j=1}^{q} c_0 \oplus \bigoplus_{i\in I_h}\bigoplus_{j=1}^{q} x_i^{(j)}$$
$$= \bigoplus_{j=1}^{q} c_0 \oplus \bigoplus_{i\in I_h} 1 = \bigoplus_{j=1}^{q} c_0 \oplus \bigoplus_{i=1}^{K} 1 = \psi^* \in \mathbb{F}_2 \tag{10}$$

Our goal is to show that $\mathcal{D}_Z$ is close to $\overline{\mathcal{D}}_q$ which we define to be the uniform distribution over the elements of $\mathbb{F}_2^q$ with parity ψ^*. Towards this we prove the following lemma, which shows that the distribution $\mathcal{D}_Z$ has low-bias.

▶ **Lemma 14.** *Consider any strict non-empty subset $S \subsetneq [q]$, s.t. $1 \le |S| = s \le q$. Then,*

$$\left|\Pr\left[\bigoplus_{j\in S} Z_j = 0\right] - \frac{1}{2}\right| \le \frac{1}{2}\cdot\exp\left(-2K/q\right).$$

Proof. First, we may assume that $s \le q/2$, otherwise we can use $[q]\setminus S$ along with (10) to complete the argument. Analogous to (10) we have that

$$\bigoplus_{j\in S} Z_j = \bigoplus_{j\in S} h\left(\mathbf{x}^{(j)}\right) = \bigoplus_{j=1}^{|S|} c_0 \oplus \bigoplus_{i\in I_h}\bigoplus_{j\in S} x_i^{(j)} = \psi_{S,c_0} \oplus \bigoplus_{i\in I_h} r_i \tag{11}$$

where $\psi_{S,c_0} = \bigoplus_{j=1}^{|S|} c_0$ is a constant and $r_i := \bigoplus_{j\in S} x_i^{(j)}$. From the construction of the random bag B, we have that $\{r_i\}_{i\in I_h}$ are iid $\mathbb{F}_2$-valued random variables such that $\Pr[r_i = 1] = s/q, \forall i \in I_h$. In other words, the RHS of (11) denotes the parity of K such iid random variables. To analyze this, let us consider an alternate way of sampling $\{r_i\}_{i=1}^{M}$:

1. Sample $T \subseteq I_h$ by including each $i \in I_h$ into T independently with probability $2s/q \le 1$.
2. For each $i \in I_h \setminus T$, set $r_i = 0$. Independently for each $i \in T$, set $r_i = 1$ w.p. $1/2$ and to 0 otherwise.

It is easy to see that conditioned on $T \neq \emptyset$, $\oplus_{i\in I_h} r_i$ is unbiased. This, along with (11) leads us to,

$$\Pr\left[\bigoplus_{j\in S} Z_j = \psi_{S,c_0}\right] = \frac{1}{2}\cdot(1 - \Pr[T=\emptyset]) + p\cdot\Pr[T=\emptyset] = \frac{1}{2} + \Pr[T=\emptyset]\left(p - \frac{1}{2}\right) \tag{12}$$

where $p \in [0,1]$ is some probability. Further,

$$\Pr[T=\emptyset] = \left(1 - \frac{2s}{q}\right)^K \le \exp(-2Ks/q)$$

Since $s \ge 1$ and $|p - 1/2| \le 1/2$, the above along with (12) completes the proof. ◀

The rest of the argument is similar to the Vazirani XOR Lemma except we need to show closeness to $\overline{\mathcal{D}}_q$ rather than the uniform distribution. We now transition to Fourier analysis of $\{0,1\}$-valued functions over $\{-1,1\}^q$. For this purpose, we shall map $\mathbb{F}_2$ to $\{-1,1\}$ via $b \mapsto (-1)^b$ and think of $\mathcal{D}_Z$ and $\overline{\mathcal{D}}_q$ as distributions over $\{-1,1\}^q$. First, from the definitions of $\mathcal{D}_Z$ and $\overline{\mathcal{D}}_q$, $\chi_{[q]}(\mathbf{z}) = (-1)^{\psi^*}$ for any $\mathbf{z}$ in the support of $\mathcal{D}_Z$ or $\overline{\mathcal{D}}_q$. Thus,

$$\mathbb{E}_{\mathbf{z}\leftarrow\overline{\mathcal{D}}_q}\left[\chi_{[q]}(\mathbf{z})\right] = \mathbb{E}_{\mathbf{z}\leftarrow\mathcal{D}_Z}\left[\chi_{[q]}(\mathbf{z})\right] = (-1)^{\psi^*} \tag{13}$$

and it is also easy to observe that for any $S \subsetneq [q]$,

$$\mathbb{E}_{\mathbf{z} \leftarrow \overline{\mathcal{D}}_q}[\chi_S(\mathbf{z})] = 0 \quad \text{and} \quad |\mathbb{E}_{\mathbf{z} \leftarrow \mathcal{D}_Z}[\chi_S(\mathbf{z})]| \leq \exp(-2K/q), \tag{14}$$

where the upper bound follows from Lemma 14. Consider any function $f : \{-1,1\}^q \to [0,1]$ having Fourier expansion $\sum_{S \subseteq [q]} \hat{f}_S \chi_S$. Using (13), and (14) we obtain

$$\begin{aligned}\left|\mathbb{E}_{\overline{\mathcal{D}}_q}[f] - \mathbb{E}_{\mathcal{D}_Z}[f]\right| &\leq \exp(-2K/q) \sum_{S \subsetneq [q]} \left|\hat{f}_S\right| \\ &\leq 2^{q/2} \cdot \exp(-2K/q) \sqrt{\sum_{S \subsetneq [q]} \hat{f}_S^2} \leq 2^{q/2} \cdot \exp(-2K/q),\end{aligned}$$

where we use Cauchy-Schwarz and Parseval's bound. We can take f to be indicator function of the event that exactly one of the coordinates is -1. This function evaluates to 1 on $\overline{\mathcal{D}}_q$ with probability exactly $q/2^{q-1}$. Using this along with the above bound completes the proof.

5.2 Hardness Reduction

Our hardness reduction is from an instance $\mathcal{L}$ of Smooth-Label-Cover given in Theorem 7.

Points, bags and label proportions. The initial set of points is defined in the space $\mathbb{F}_2^{V \times [M]}$. For a point $\hat{\mathbf{x}} \in \mathbb{F}_2^{V \times [M]}$ let $\hat{\mathbf{x}}[v] = (\hat{x}_{v,1}, \ldots, \hat{x}_{v,M})$ be the vector of M coordinates corresponding to $v \in V$. Let $\mathcal{D}_{\mathcal{B}}$ be the distribution on bags and label proportions given by the following process.

1. Sample $v \in V$ u.a.r.
2. Sample $(B = \{\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(q)}\}, 1/q) \leftarrow \mathcal{D}^{\mathsf{dict}}_{M,q}$.
3. For $j \in [q]$: define $\hat{\mathbf{x}}^{(j)} \in \mathbb{F}_2^{V \times [M]}$ by letting $\hat{\mathbf{x}}^{(j)}[v] = \mathbf{x}^{(j)}[v]$ and for all $v' \neq v$, $\hat{\mathbf{x}}^{(j)}[v'] = \mathbf{0}$.
4. Output $(\hat{B} = \{\hat{\mathbf{x}}^{(1)}, \ldots, \hat{\mathbf{x}}^{(q)}\}, 1/q)$

Folding and projected point-set. For each $e = (v_1, v_2) \in E$ and $j \in [N]$ define the linear constraint $C[e,j]$ over point $\hat{\mathbf{x}} \in \mathbb{F}_2^{V \times [M]}$ as

$$C[e,j] \Leftrightarrow \bigoplus_{i \in \pi^{-1}_{ev_1}(j)} \hat{x}_{v_1 i} = \bigoplus_{i \in \pi^{-1}_{ev_2}(j)} \hat{x}_{v_2 i}. \tag{15}$$

Let $H \subset \mathbb{F}_2^{V \times [M]}$ be the subspace of all the points which satisfy the set of homogeneous linear constraints $\mathcal{C} := \{C[e,j] \mid e \in E, j \in [N]\}$. We let H be the space in which our final instance resides by linearly projecting all points $\hat{\mathbf{x}}$ created in the support of $\mathcal{D}_{\mathcal{B}}$ into points $\overline{\mathbf{x}} \in H$. Since our final instance is represented in a coordinate system corresponding to a linear basis for H, this also forces any solution h to be represented in a basis for H. In particular, h represented in the original space by $h(\hat{\mathbf{x}}) := c_0 \oplus \langle \mathbf{c}, \hat{\mathbf{x}} \rangle$ (where the inner product is over $\mathbb{F}_2$) must obey $\mathbf{c} \in H$. Let $\overline{\mathcal{D}}_{\mathcal{B}}$ be the new distribution on the bags $(\overline{B}, 1/q)$ given by the linear projection of all the points in the bags of $\mathcal{D}_{\mathcal{B}}$ on to H.

5.2.1 YES Case

In this case, there is a labeling $\rho : V \to [M]$ which satisfies all the edges of $\mathcal{L}$. Consider over $\mathbb{F}_2^{V \times [M]}$ the parity $h^*(\hat{\mathbf{x}}) = \bigoplus_{v \in V} \hat{x}_{v,\rho(v)} =: \langle \mathbf{c}^*, \hat{\mathbf{x}} \rangle$. Since, for any edge $e = (v_1, v_2)$, $\pi_{ev_1}(\rho(v_1)) = \pi_{ev_2}(\rho(v_2))$, $\mathbf{c}^* \in H$. Now fix a choice of v in Step 1 of the distribution $\mathcal{D}_{\mathcal{B}}$. Restricted to the coordinates corresponding to v (since the others are set to 0), h^* is simply

$x_{v,\rho(v)}$. We can now directly apply the completeness property of $\mathcal{D}^{\mathsf{dict}}_{M,q}$ in Lemma 13 to obtain that h^* satisfies all the bags given the choice v. Since this holds for all choices of v, h^* satisfies all the bags of $\mathcal{D}_{\mathcal{B}}$.

5.2.2 NO Case

Assume for a contradiction that there is a parity in the space H that satisfies $(q/2^{q-1}+\delta)$-fraction of the bags of $\mathcal{D}_{\mathcal{B}}$. This parity can be written as

$$h(\hat{\mathbf{x}}) = c_0 \oplus \bigoplus_{v\in V}\bigoplus_{i=1}^{M} c_i\hat{x}_i = c_0 \oplus \bigoplus_{v\in V}\langle \mathbf{c}[v], \hat{\mathbf{x}}[v]\rangle, \tag{16}$$

where $\mathbf{c}$ satisfies the constraints $\mathcal{C}$. By averaging there are $\delta/2$ fraction of *good* $v \in V$ such that h satisfies $(q/2^{q-1}+\delta/2)$-fraction of the bags of $\mathcal{D}_{\mathcal{B}}\mid_v$ i.e., $\mathcal{D}_{\mathcal{B}}$ given v is chosen in Step 1. By the weak-expansion property of $\mathcal{L}$ in Theorem 7, the subset of edges E' induced by the good vertices satisfies $|E'| \geq (\delta/2)^2|E|$. Let $S_v := \{i \in [M] \mid c_{v,i} = 1\}$. From the soundness of $\mathcal{D}^{\mathsf{dict}}_{M,q}$ (Lemma 13), we obtain that all good v satisfy

$$|S_v| \leq \Delta := q(\log(2/\delta) + q/2)/2\ . \tag{17}$$

The smoothness of $\mathcal{L}$ implies that for any good $v \in V$, $\Pr_{e\sim v}\left[\pi_{ev}\left(S_v\right) = |S_v|\right] \geq 1 - |S_v|^2/(2Q) \geq 1 - \Delta^2/(2Q)$. Let $E^* = \{e = (v_1, v_2) \in E' \mid \pi_{ev_r}\left(S_{v_r}\right) = |S_{v_r}|, r = 1, 2\}$. Then

$$\frac{|E^*|}{|E|} \geq \zeta := \frac{\delta^2}{4} - \frac{\Delta^2}{Q}. \tag{18}$$

We have the following lemma.

▶ **Lemma 15.** *For any $e = (v_1, v_2) \in E^*$, $\pi_{ev_1}\left(S_{v_1}\right) \cap \pi_{ev_2}\left(S_{v_2}\right) \neq \emptyset$.*

Proof. Since h satisfies at least one bag of $\mathcal{D}\mid_{v_1}$, $\mathbf{c}[v] \neq \mathbf{0}$, and thus $S_{v_1} \neq \emptyset$. Consider any $j \in \pi_{ev_1}\left(S_{v_1}\right)$. From the definition of E^*, $\left|\pi^{-1}_{ev_1}(j) \cap S_{v_1}\right| = 1$. Thus, $\bigoplus_{i\in\pi^{-1}_{ev_1}(j)} \hat{x}_{v_1 i} = 1$ and from (15) and the fact that $\hat{\mathbf{c}}$ satisfies $C[e, j]$ we obtain that $\pi^{-1}_{ev_2}(j) \cap S_{v_2} \neq \emptyset$. ◀

Let ρ be the randomized labeling to the good vertices given by randomly assigning each good $v \in V$ a label chosen u.a.r. from S_v. From Lem. 15, (17) and (18) we obtain that ρ satisfies in expectation at least, $\nu := \zeta/\Delta^2$ fraction of the edges of $\mathcal{L}$. By choosing the parameter Q in Theorem 7 to be large enough we can take $\zeta \geq \delta^2/8$ and then taking taking the parameter R to be be large enough we obtain a contradiction.

6 Approximately LLP Learning Parities

Here we may assume that $q \geq 2$, otherwise all bags are of size 1 and one can use Gaussian elimination to solve for the satisfying parity. Let $\mathcal{B} := \{(B_k, \sigma_k)\}_{k=1}^m$ be an instance of LLP-PARITY$[q]$ over $\mathbb{F}_2^d$ such that there is an (unknown) parity that satisfies all the bags of $\mathcal{B}$. We therefore need to solve for the coefficients of the parity.

Let us first define the subsets of bags $\mathcal{B}_a := \{(B, \sigma) \in \mathcal{B} \mid \sigma = a\}$ for $a \in \{0, 1\}$. We call the bags in $\mathcal{B}_0 \cup \mathcal{B}_1$ as *monochromatic* since we know that the vectors in any such bag are either all labeled 0 or all labeled 1. Therefore, one can write a (possibly non-homogeneous) $\mathbb{F}_2$-linear constraint (in the coefficients of the parity) for each vector in any monochromatic

bag. Further, since the label proportion of each bag is given, the parity of labels in each bag is also determined. Thus, we can add these $\mathbb{F}_2$-linear constraints capturing the parity of the labels for each bag.

Since this system of linear equations is feasible (due to the existence of the satisfying parity) one can do Gaussian elimination to obtain a reduced instance $\mathcal{B}'$ of LLP-PARITY$[q]$ which satisfies the following properties:

1. $\mathcal{B}'$ has no monochromatic bags.
2. A subset of the coefficients may be eliminated or assigned a fixed value $\in \mathbb{F}_2$, and the rest are *free*.
3. For any bag $(B, \sigma) \in \mathcal{B}'$, any assignment to the *free* coefficients yields a labeling which satisfies the parity constraint of that bag. In particular, the set of such assignments yields a (possibly affine) subspace of labelings of size at most 2^{t-1} where $t = |B|$. Let us call this subspace of labelings as F_B.

The algorithm outputs a random parity given by a random assignment to the free coefficients. To analyze its performance, let us consider a bag $(B, \sigma) \in \mathcal{B}'$ where $|B| = t \leq q$ and $t\sigma \in \{1, \ldots, t-1\}$ since $\mathcal{B}'$ has no monochromatic bags. By feasibility there exists a vector in F_B which has Hamming weight $t\sigma$. The probability that the bag will be satisfied by a random parity is precisely the probability that a random point in F_B has Hamming weight $t\sigma$. There are two cases:

1. *F_B contains all vectors of Hamming weight $t\sigma$.* Since $t\sigma \in \{1, \ldots, t-1\}$ the number of such vectors is at least t. Since $|F| \leq 2^{t-1}$, the probability that the bag is satisfied is at least $t/2^{t-1} \geq 1/2^{q-2}$ for any positive integers $t \leq q$ and $q \geq 2$.
2. *F_B does not contains all vectors of Hamming weight $t\sigma$.* In this case, F_B is at most $(t-2)$-dimensional and thus $|F_B| \leq 2^{t-2}$. Since F_B does contain one vector of Hamming weight $t\sigma$, the probability that the bag is satisfied is at least $1/2^{t-2} \geq 1/2^{q-2}$.

References

1 S. Arora, L. Babai, J. Stern, and Z. Sweedyk. The hardness of approximate optima in lattices, codes, and systems of linear equations. *J. Comput. Syst. Sci.*, 54(2):317–331, 1997.

2 S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.

3 S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998.

4 D. Barucic and J. Kybic. Fast learning from label proportions with small bags. *CoRR*, abs/2110.03426, 2021. `arXiv:2110.03426`.

5 G. Bortsova, F. Dubost, S. N. Ørting, I. Katramados, L. Hogeweg, L. H. Thomsen, M. M. W. Wille, and M. de Bruijne. Deep learning from label proportions for emphysema quantification. In *MICCAI*, volume 11071 of *Lecture Notes in Computer Science*, pages 768–776. Springer, 2018. `arXiv:1807.08601`.

6 R. I. Busa-Fekete, H. Choi, T. Dick, C. Gentile, and A. M. Medina. Easy learning from label proportions. *arXiv*, 2023. `arXiv:2302.03115`.

7 L. Chen, T. Fu, A. Karbasi, and V. Mirrokni. Learning from aggregated data: Curated bags versus random bags. *arXiv*, 2023. `arXiv:2305.09557`.

8 L. Chen, Z. Huang, and R. Ramakrishnan. Cost-based labeling of groups of mass spectra. In *Proc. ACM SIGMOD International Conference on Management of Data*, pages 167–178, 2004.

9 L. M. Dery, B. Nachman, F. Rubbo, and A. Schwartzman. Weakly supervised classification in high energy physics. *Journal of High Energy Physics*, 2017(5):1–11, 2017.

10 V. Feldman, V. Guruswami, P. Raghavendra, and Y. Wu. Agnostic learning of monomials by halfspaces is hard. *SIAM J. Comput.*, 41(6):1558–1590, 2012.

11 S. Ghoshal and R. Saket. Hardness of learning DNFs using halfspaces. In *Proc. STOC*, pages 467–480, 2021.
12 V. Guruswami, P. Raghavendra, R. Saket, and Y. Wu. Bypassing UGC from some optimal geometric inapproximability results. *ACM Trans. Algorithms*, 12(1):6:1–6:25, 2016. URL: `http://eccc.hpi-web.de/report/2010/177`.
13 J. Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.
14 J. Hernández-González, I. Inza, L. Crisol-Ortíz, M. A. Guembe, M. J. Iñarra, and J. A. Lozano. Fitting the data from embryo implantation prediction: Learning from label proportions. *Statistical methods in medical research*, 27(4):1056–1066, 2018.
15 S. Khot and R. Saket. Hardness of minimizing and learning DNF expressions. In *Proc. FOCS*, pages 231–240, 2008.
16 C. O'Brien, A. Thiagarajan, S. Das, R. Barreto, C. Verma, T. Hsu, J. Neufeld, and J. J. Hunt. Challenges and approaches to privacy preserving post-click conversion prediction. *CoRR*, abs/2201.12666, 2022. `arXiv:2201.12666`.
17 S. N. Ørting, J. Petersen, M. Wille, L. Thomsen, and M. de Bruijne. Quantifying emphysema extent from weakly labeled ct scans of the lungs using label proportions learning. In *The Sixth International Workshop on Pulmonary Image Analysis*, pages 31–42, 2016.
18 R. O'Donnell. *Analysis of boolean functions*. Cambridge University Press, 2014.
19 R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.
20 S. Rueping. SVM classifier estimation from group probabilities. In *Proc. ICML*, pages 911–918, 2010.
21 R. Saket. Learnability of linear thresholds from label proportions. In *Proc. NeurIPS*, 2021. URL: `https://openreview.net/forum?id=5BnaKeEwuYk`.
22 R. Saket. Algorithms and hardness for learning linear thresholds from label proportions. In *Proc. NeurIPS*, 2022. URL: `https://openreview.net/forum?id=4LZo68TuF-4`.
23 L. G. Valiant. A theory of the learnable. *Commun. ACM*, 27(11):1134–1142, 1984.
24 J. Wojtusiak, K. Irvin, A. Birerdinc, and A. V. Baranova. Using published medical results and non-homogenous data in rule learning. In *Proc. International Conference on Machine Learning and Applications and Workshops*, volume 2, pages 84–89. IEEE, 2011.
25 F. X. Yu, K. Choromanski, S. Kumar, T. Jebara, and S. F. Chang. On learning from label proportions. *CoRR*, abs/1402.5902, 2014. `arXiv:1402.5902`.

A Useful Bounds

We use the Chernoff bound stated as follow.

▶ **Theorem 16** (Chernoff Bound). *Suppose $X_1, \ldots, X_n$ and independent $\{0,1\}$-valued random variables with $S = \sum_{i=1}^{n} X_i$ and $\mu = \mathbb{E}[S]$. Then, $\Pr[S \leq (1-\delta)\mu] \leq \exp(-\delta^2\mu/2)$ for any $\delta > 0$.*