CORRIGENDUM: EXPLICIT CONSTRUCTION OF A SMALL EPSILON-NET FOR LINEAR THRESHOLD FUNCTIONS

YUVAL RABANI* AND AMIR SHPILKA†

Abstract. The purpose of this note is to correct mistakes and inaccuracies in technical claims in [RS10]. These have no effect on the main results in the paper.

Key words. ϵ-net, hitting sets, derandomization, explicit construction

AMS subject classifications. 68Q99

1. Overview of the mistakes in [RS10].

We first give an overview of the mistakes and then give the proofs.

1.1. Lemma 2.7 in [RS10].

In a private communication, William Hoza pointed out a mistake in the proof of Lemma 2.7 in [RS10].

The flawed proof in the appendix of [RS10] constructs $\mathcal{H}$ in two steps. The error is that the first step uses known constructions of pairwise independent hash families mapping $[n]$ to $[s]$. However, the cardinality of such families is $2^{O(\log(sn))}$, which is too large to give the bounds stated in the lemma (in the paper we wrongly claimed that the size is much smaller). If we were to use this bound in the construction of the hitting set then we would get a hitting set of size $(n/\epsilon)^{O(\log\log 1/\epsilon)}$, which is larger than the claimed size (alternatively, in terms of seed length, we get seed of length $O(\log(n/\epsilon)\log\log 1/\epsilon)$ instead of the optimal $O(\log(n/\epsilon))$). In a footnote we also proposed to add a preliminary step that maps $[n]$ to $[s^2]$ and then to map $[s^2]$ to $[s]$. This, too, is a flawed construction as it does not guarantee that we get from this a family of pairwise independent hash functions. Our fix is to apply a different preliminary step that reduces the domain size, using the construction of lossless condensers of [GUV09]. This leads to a weaker statement, that nonetheless is good enough to derive the main result (Theorem 1.1).

It was then pointed out to us that there is a second mistake in the original statement of the lemma. We argued that there exists an h such that for all t ... but actually proved that for every t there is a good h.

The following is a revised version of Lemma 2.7.

LEMMA 1.1 (New Lemma 2.7). *There exists a universal constant A such that the following holds. For every integers s, n such that $s \le n$, there is an explicit family $\mathcal{H}$ of hash functions $h : [n] \to [8s]$ of cardinality $|\mathcal{H}| = 2^{(4+o(1))\cdot s + A\cdot\log 2s \log\log n + O(1)}$ such that the following holds for every unit vector $v \in \mathbb{S}^{n-1}$. Let $i_1, i_2, \dots, i_n$ be an enumeration of $[n]$ such that $|v_{i_1}| \ge |v_{i_2}| \ge \cdots \ge |v_{i_n}|$, and let I_t denote the set $\{i_1, i_2, \dots, i_t\}$. For every $t \in [s-1]$, there exists some $h \in \mathcal{H}$ such that*

1. *The map h is an injection on I_s.*
2. *If $v_{i_{t+1}}^2 \le \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then*

$$\sum_{r\in[8s]} \min\left\{ \|v_{h^{-1}(r)\setminus I_t}\|_2^2, \frac{2}{s} \cdot \|v_{[n]\setminus I_t}\|_2^2 \right\} \ge \frac{1}{2} \cdot \|v_{[n]\setminus I_t}\|_2^2. \tag{1.1}$$

*The Rachel and Selim Benin School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem 91904, Israel. Email: yrabani@cs.huji.ac.il. Research supported by Israel Science Foundation grant number 1109/07 and by US-Israel Binational Science Foundation grant number 2008059.

†Blavatnik School of Computer Science at Tel Aviv University, Tel Aviv, Israel. Email: shpilka@tauex.tau.ac.il. The research leading to these results received funding from the Israel Science Foundation (grant number 514/20) and from the Len Blavatnik and the Blavatnik Family foundation.

Furthermore, the $o(1)$ term in the exponent of $|\mathcal{H}|$ depends only on s.

The main differences between the statement of Lemma 2.7 in [RS10] and Lemma 1.1 are that we corrected the order of quantifiers and we now have the extra A factor in the exponent (earlier we took $A = 1$). A similar modification was needed in the statement of Corollary 2.8. Another modification to that corollary is explained next.

1.2. "Case 3" in the proof of Theorem 1.1 and Corollary 2.8. Another mistake in the proof of Theorem 1.1. in [RS10] is that in the analysis of "Case 3." we used t in two different meanings (in the fix below we replaced some of the appearances of t with q). This confusion caused us to use the "wrong" t in Equation (5). The problem is that when using the "correct" t the proof won't work as is as that t can be small and so we cannot argue that

$$\|v_{I_t}\|_1 + \sqrt{2\log(2/\epsilon)} \cdot \|v_{[n]\setminus I_t}\|_2 > \theta \ ,$$

at the end of the proof of Theorem 1.1.

To fix the issue we had do slightly strengthen the statement of Corollary 2.8 in [RS10] to guarantee that the "correct" t (which we now call q) is large enough.

COROLLARY 1.2. *[New Corollary 2.8] Let $24 \leq s \leq n$ be integers and $\mathcal{H}$ the hash family guaranteed by Lemma 1.1. There exists constants c_1 and c_2 such that one of the following conditions holds (using the same notation as in Lemma 1.1):*

1. *either $\sum_{q=\lceil 2s/3 \rceil}^{s-1} |v_{i_{q+1}}| \geq \frac{\sqrt{s}}{32} \|v_{[n]\setminus I_s}\|_2$;*
2. *or, there exists $\lceil 2s/3 \rceil \leq q \leq s-1$ and $h \in \mathcal{H}$ such that h is an injection on I_s and for at least $c_1 \cdot 8s$ buckets r it holds that $\|v_{h^{-1}(r)\setminus I_q}\|_2^2 \geq \frac{c_2}{s} \cdot \|v_{[n]\setminus I_q}\|_2^2$.*

The difference of the version above to Corollary 2.8 is that in the second option we guarantee that q is at least $\lceil 2s/3 \rceil$.

1.3. The size of N_ϵ. As a consequence of the modification in Lemma 2.7 (Lemma 1.1 here) the calculation of the size of $|N_\epsilon|$ in the proof of Theorem 1.1 slightly changed. The end result is the same (except for a different small correction that we explain next) but the justification is slightly different (due to the extra A factor from Lemma 1.1):

$$\begin{aligned} |N_\epsilon| = |\mathcal{H}| \cdot d^{8t-1} \cdot m &= O\left(2^{(4+o(1))\cdot t + A\log 2t \log\log n} \cdot d^7 \cdot (2/\epsilon)^{8c\log d} \cdot n^{k/2}\right) \\ &= O\left(n^a \cdot (1/\epsilon)^b\right), \end{aligned}$$

for any constants $a > k/2$ and $b > 4c + 8c\log d$.

Observe that another small change in the estimate above is that we replaced the $n^{o(1)}$ term with $(n/\epsilon)^{o(1)}$, which is required when ϵ is extremely small. This only affects the constants a and b that were earlier taken to be $a = k/2$ and $b = 4c + 8c\log d$.

1.4. Small corrections. Beside the major issues described above we slightly changed the text in a few places to either correct typos or explain some of the new claims. Below we give the details of the changes.

1.4.1. First paragraph of Section 2.3. The paragraph was changed to "A set H of functions $h : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, m\}$ such that for every $S \subset \{1, 2, \ldots, n\}$ with $|S| = s$ there exists $h \in H$ such that $|h(S)| = s$ is called an (n, m, s)-perfect hash family. For all $n, s \in \mathbb{N}$, $s \leq n$, there are explicit constructions of $(n, O(s), s)$-perfect hash families H with $|H| = 2^{O(s+\log\log n)}$ (see Theorem 6 in [SS90]). Lemma 2.7 is strengthening of the above requirement. Informally, the strengthened version says that we can construct H to have the following property. For every vector $v = (v_1, \ldots, v_n)$ there is $h \in H$ that maps its "heaviest" s coordinates (in absolute value) to different locations, and furthermore, if the remaining coordinates have sufficient L_2 mass, then it is distributed by h roughly evenly among the $O(s)$ locations."

Most of the modification is the text starting with "Lemma 2.7 is a strengthening ..." Another small change is that we give the relevant theorem number from the paper of [SS90].

1.4.2. Proof of Theorem 1.1, last line before "Case 1". We removed the sentence "We also define, for every $i \in [8t]$, $J'_{h,i} = h^{-1}(i) \setminus I_t$." as it uses the "wrong" t. J'_{h_i} is defined later in the analysis of "Case 3."

1.4.3. Acknowledgments. Naturally, we revised this section to thank William Hoza and the anonymous referees to which we are grateful for finding the mistakes in [RS10].

We thank Noga Alon and Avi Wigderson for helpful discussions and for bringing [SS90] to our attention. We also thank Noga for sharing his proof of Corollary 1.3 with us.

We are grateful to William Hoza who found a mistake in our original argument. We also wish to thank William Hoza and the anonymous reviewers for pointing out several other inaccuracies and for comments that helped us significantly improve the presentation of the proof.

2. Proof of Lemma 1.1.

2.1. Lossless condensers. An important ingredient in the proof is the following construction of Guruswami, Umans, and Vadhan [GUV09] (see also Chapter 6 in [Vad12]) of lossless condensers.

For completeness we first give some basic definitions and then discuss lossless condensers.

DEFINITION 2.1. *Let D be a distribution on $\{0,1\}^a$. We say that D is a k-source if every point in the hypercube has probability at most 2^{-k}. We use U_d to denote the uniform distribution on $\{0,1\}^d$. We say that two distributions D_1, D_2 are ϵ-close if their statistical distance (half their L_1 distance) is at most ϵ.*

DEFINITION 2.2. *A function Con : $\{0,1\}^a \times \{0,1\}^d \to \{0,1\}^b$ is a $k \to_\epsilon k'$ condenser iff for every k-source X on $\{0,1\}^a$, there exists a k'-source Z on $\{0,1\}^b$ such that $Con(X, U_d)$ is ϵ-close to Z. The function Con is lossless iff $k' = k + d$.*

The main result that we shall need is Theorem 1.7 of [GUV09].[1]

THEOREM 2.3 (Theorem 1.7 of [GUV09]). *There exists an absolute constant $\beta > 0$ such that: for all positive k, all $a \geq k$, where a is an integer, and all $\epsilon > 0$, there is an explicit $k \to_\epsilon k + d$ lossless condenser Con : $\{0,1\}^a \times \{0,1\}^d \to \{0,1\}^b$ with $d = \lceil 3(\log a + \log k + \log(1/\epsilon)) + \beta \rceil$ and $b \leq 2(k + d)$.*

2.2. Proof of Lemma 1.1 **(the new version of Lemma 2.7).** First, note that when $s = 1$ the statement of the claim is trivial and so we only consider the case $s \geq 2$ in the proof.

Our proof consists of three steps. In the first step we use the lossless condenser, of Theorem 2.3, to map $[n]$ to $[\log(n) \cdot \mathrm{poly}(s)]$. Then we use the oblivious implementation due to Schmidt and Siegel [SS90] of the Fredman, Komlós, and Szemerédi (FKS) adaptive hashing scheme [FKS84].[2] The Schmidt and Siegel implementation consists of two steps. In the first step (second step of our proof), they use a universal family of pairwise-independent hash function to reduce the domain size further to size $O(s)$. The last step repairs the few collisions that may exist.

To ease the readability of the proof we shall assume that n is a power of 2. This has no effect on the claim or the result.

[1] We consider the special case $\alpha = 1/2$ of Theorem 1.7 of [GUV09].

[2] For the construction of our hitting set we need the hash family to be fixed and to not depend on the input.

Step 1. Let $C : \{0,1\}^a \times \{0,1\}^d \to \{0,1\}^b$ be the condenser promised in Theorem 2.3 for parameters $a = \log n$, $k = \log(64s)$, $\epsilon = s^{-100}$, $d = \lceil 3(\log a + \log k + \log(1/\epsilon)) + \beta \rceil$ and $b \leq 2(k+d)$. (Notice that if $s > n/64$ and hence $k > a$, we can simply skip Step 1.) For each seed $y \in \{0,1\}^d$ denote $C_y(x) = C(x,y)$. We think of the family C_y as a family of hash functions from $[n]$ to $[2^b]$.

We next show that for a random seed $y \in \{0,1\}^d$, with high probability, C_y is one to one on the set I_s.

CLAIM 2.4. *Let $X \subseteq \{0,1\}^a$ be a k-source. Then for all but $\sqrt{\epsilon}$ of the seeds $y \in \{0,1\}^d$ it holds that $C_y(X)$ is $\sqrt{\epsilon}$-close to a k-source.*

Proof. The proof is an easy application of Markov's inequality. □

COROLLARY 2.5. *Let $I \subseteq [n]$ be a set of size $|I| \leq 2^k$. Then, except with probability $\sqrt{\epsilon}$ over $y \in \{0,1\}^d$, the map C_y is injective on I.*

Proof. Let X be a random variable that is uniformly distributed over a set of size exactly 2^k that contains I. Let y be such that $C_y(X)$ is $\sqrt{\epsilon}$-close to a k-source Z. Then,

$$\forall z \in \{0,1\}^b \ , \quad \Pr[C_y(X) = z] \leq \Pr[Z = z] + \sqrt{\epsilon} \leq 2^{-k} + \sqrt{\epsilon} < 2 \cdot 2^{-k} \ ,$$

where the third inequality follows from the choice of ϵ. In particular, no two elements of I were mapped to the same element z. □

We next show that, with high probability, C_y distributes the weight "nicely".

CLAIM 2.6. *For all but $\sqrt{\epsilon}$ fraction of $y \in \{0,1\}^d$ the following holds. If for $t \in [s-1]$, we have that $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then for every $z \in \{0,1\}^b$,*

$$\|v_{C_y{}^{-1}(z)\setminus I_t}\|^2 \leq \left(\frac{1}{64s} + \sqrt{\epsilon}\right) \cdot \|v_{[n]\setminus I_t}\|_2^2 \ .$$

Proof. Consider the following distribution on $[n]$:

$$\Pr[X = i_j] = \begin{cases} \frac{v_{i_j}^2}{\|v_{[n]\setminus I_t}\|_2^2} & \text{if } j > t \\ 0 & otherwise \end{cases} \ .$$

By our assumption, X is a $\log(64s)$-source (k-source). Claim 2.4 implies that except for a $\sqrt{\epsilon}$ fraction of the seeds y, $C_y(X)$ is $\sqrt{\epsilon}$-close to a k-source Z (note that Z may depend on y). For such a good y and for $z \in \{0,1\}^b$ we have that

$$\begin{aligned} 2^{-k} + \sqrt{\epsilon} \geq \Pr[C_y(X) = z] &= \sum_{j\in[n]\setminus[t]:C_y(i_j)=z} \Pr[X = i_j] \\ &= \sum_{j\in[n]\setminus[t]:C_y(i_j)=z} \frac{v_{i_j}^2}{\|v_{[n]\setminus I_t}\|_2^2} \\ &= \frac{\|v_{C_y^{-1}(z)\setminus I_t}\|_2^2}{\|v_{[n]\setminus I_t}\|_2^2} \end{aligned}$$

as claimed. □

COROLLARY 2.7. *With probability at least $1 - s\sqrt{\epsilon}$ a random seed y satisfies that:*

1. *C_y is one-to-one on I_s.*
2. *For every $t \in [s-1]$, if $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then for every $z \in \{0,1\}^b$,*

$$\|v_{C_y{}^{-1}(z)\setminus I_t}\|^2 \leq \left(\frac{1}{64s} + \sqrt{\epsilon}\right) \cdot \|v_{[n]\setminus I_t}\|_2^2 \ .$$

Proof. Follows immediately from applying the union bound to Corollary 2.5 and Claim 2.6. □

We say that a seed y is "good" if it is one of the $1-s\sqrt{\epsilon}$ fraction of seeds in the statement of Corollary 2.7.

Fix such a good y. We now have that C_y has reduced the domain size to $2^b \leq 2^{2d+2k} = (\log(n)\cdot s)^{O(1)}$.

Step 2. We now proceed as in the construction of perfect hash families of Schmidt and Siegel [SS90]. We first apply a map $f : [2^b] \to [s]$, taken from a pairwise independent family of hash functions $\mathcal{F}$. There are known explicit constructions of $\mathcal{F}$ with $|\mathcal{F}| = 2^{b+\log s+O(1)}$ (see Theorem 3.26 in [Vad12] and the historical discussion there).

A pairwise independent family of hash functions $\mathcal{F}$ has the following property. If f is chosen uniformly at random from $\mathcal{F}$, then for every $x, y \in [2^b]$, $x \neq y$, it holds that $f(x)$ is distributed uniformly in $[s]$, even when conditioned on $f(y)$. In particular, $\Pr[f(x) = f(y)] = \frac{1}{s}$.

Let $S \subseteq [2^b]$ be an arbitrary set of size $|S| \leq s$. Consider the following event.

$$\sum_{j=1}^{s} |f^{-1}(j) \cap S|^2 < 4s. \tag{2.1}$$

We now show that the probability of this event, when f is chosen uniformly at random from $\mathcal{F}$, is more than $\frac{1}{2}$. Indeed, denoting by χ_p the indicator of an event p, we have that

$$\begin{aligned}\mathbb{E}\left[\sum_{j=1}^{s} |f^{-1}(j)\cap S|^2\right] &= \mathbb{E}\left[\sum_{x,x'\in S} \chi_{f(x)=f(x')}\right] = \sum_{x,x'\in S} \mathbb{E}\left[\chi_{f(x)=f(x')}\right] \\ &= \sum_{x\neq x'\in S} \mathbb{E}\left[\chi_{f(x)=f(x')}\right] + s = s\cdot(s-1)\cdot(1/s) + s = 2s-1\ .\end{aligned}$$

By applying Markov's inequality we conclude that

$$\Pr\left[\sum_{j=1}^{s} |f^{-1}(j)\cap S|^2 \geq 4s\right] < \frac{1}{2}. \tag{2.2}$$

Thus, the average square of the number of pre-images of a bucket, is of size at most 4.

Step 3. The second phase of the Fredman, Komlós, and Szemerédi hashing scheme is adaptive, and depends on the hashed set S. The idea is the following. If c_i elements of S landed in bucket $i \in [s]$, then by mapping this bucket to c_i^2 buckets using a pairwise independent family of hash functions, it is likely that no collision between the elements of S occurs. As the first phase guarantees that $\sum_{i\in[s]} c_i^2 = O(s)$, we end up with a hash table of size $O(s)$. Note that for this construction to work, we need to know the values $\{c_i\}$ which is the reason for the adaptiveness. The Schmidt and Siegel implementation proceeds as follows. It uses a pairwise independent family of hash functions $\mathcal{G}$. Here it will be convenient to assume that $g \in \mathcal{G}$ maps $[2^b]$ to bit vectors. So every $g \in \mathcal{G}$ is a function $g : [2^b] \to \{0,1\}^{2+\log s}$. We can take $|\mathcal{G}| = 2^{\log s+b+O(1)}$. The second phase uses a selection of s (not necessarily distinct) hash functions from $\mathcal{G}$. The hash functions are selected and used as follows. Take a sequence of $\log s$ hash functions $g_1, g_2, \ldots, g_{\log s} \in \mathcal{G}$. Notice that there are at most $|\mathcal{G}|^{\log s} = 2^{\log^2 s+b\log s\ +O(\log s)}$ such sequences. In addition, take a sequence of s non-negative integers $\bar{c} = (c_1, c_2, \ldots, c_s)$ that satisfy $\sum_{j=1}^{s} c_j = s$ and $\sum_{j=1}^{s} c_j^2 < 4s$. There are at most 2^{2s} such sequences (easily bounded by writing the sequence elements in unary notation, separated by zeros). This sequence is our guess of the bucket loads due to S after the first phase. Finally, use an assignment $\rho : [s] \to [\log s]$, that assigns values from $[\log s]$ to elements of $[s]$ in the following way: $1 \in [\log s]$ is assigned to $\frac{s}{2}$ elements of $[s]$, $2 \in [\log s]$ is assigned to $\frac{s}{4}$ elements of $[s]$, and in general $i \in [\log s]$ is assigned to $\frac{s}{2^i}$ elements of $[s]$. Exceptionally, $\log s$ is assigned to 2 elements of $[s]$, in order to cover the entire set. The number of such assignments is at most $2^{s \cdot (1 + \sum_{i=1}^{\log s} 2^{-i})} < 2^{2s}$ (write the s assigned values in unary, separated by zeros). The assignment ρ is our guess as to which of the $\log s$ selected hash functions should be used for each bucket.

Each setting of y, f, $\bar{g}$, $\bar{c}$ and ρ defines a hash function $h \in \mathcal{H}$ as follows. For every $x \in [n]$,

$$h(x) = \left(\sum_{i < f(C_y(x))} 2^{\lceil 2 \log c_i \rceil} \right) + \bar{g}_{\rho(f(C_y(x)))}(C_y(x)),$$

where for $i = f(C_y(x))$, $\bar{g}_{\rho(i)}(C_y(x))$ is the first $\lceil 2 \log c_i \rceil$ bits of $g_{\rho(i)}(C_y(x))$. We shall also think of $\bar{g}_{\rho(i)}(C_y(x))$ as the binary expansion of an integer number. Notice that

$$|\mathcal{H}| \le 2^d \cdot |\mathcal{F}| \cdot |\mathcal{G}|^{\log s} \cdot \#\{\bar{c}\} \cdot \#\{\rho\} \le 2^{4s + O(\log^2 s) + O(\log 2s \log \log n) + O(\log s)} , \tag{2.3}$$

implying the claim in the lemma.[3] Also notice that each $h \in \mathcal{H}$ maps $[n]$ to

$$\sum_{i=1}^{s} 2^{\lceil 2 \log c_i \rceil} \le 2 \cdot \sum_{i=1}^{s} c_i^2 < 8s,$$

as required.

Wrapping up. Recall that we still work with a fixed "good" y.

CLAIM 2.8. *For every vector $v \in \mathbb{S}^{n-1}$, the probability that when we pick f at random there is a choice of $\bar{g}$ and ρ such that $h = h_{y,f,\bar{g},\rho}$ is injective on I_s is at least $\frac{1}{2}$.*

Proof. Let $S = I_s$. Since y is good, C_y is injective on S. Denote $S_y = C_y(S)$. For this set S_y, Equation (2.1) holds for at least half of the choices of f (by Equation (2.2)). Fix any such choice f. For $i = 1, 2, \ldots, s$, let $C_i = \{x \in S_y : \ f(x) = i\}$. Consider the choice of $c_i = |C_i|$, for $i = 1, 2, \ldots, s$. Fix i. For every $g \in \mathcal{G}$ and $x \in [2^b]$, let $\bar{g}(x)$ denote the first $\lceil 2 \log c_i \rceil$ bits of $g(x)$. Consider the "bad" event

$$A_i = A_i(g) = \exists x, x' \in C_i, \ x \ne x' : \ \bar{g}(x) = \bar{g}(x') \ .$$

As $\mathcal{G}$ is a pairwise independent family of hash functions, if g is chosen uniformly at random in $\mathcal{G}$, then $\Pr[A_i] \le \binom{c_i}{2} \cdot \frac{1}{c_i^2} < \frac{1}{2}$. Therefore, there exists a choice of g_1 that is good for a set $J_1 \subset [s]$ of buckets of cardinality $|J_1| = \frac{s}{2}$. Similarly, for $j = 2, 3, \ldots, \log s - 1$, there exists a choice of g_j that is good for a set $J_j \subset [s] \setminus \bigcup_{j' < j} J_{j'}$ of cardinality $|J_j| = \frac{s}{2^j}$. Similarly, there exists a choice of $g_{\log s}$ that is good for both elements in $[s] \setminus \bigcup_{j < \log s} J_j$. So, for every f that satisfies Equation (2.1), there is a choice of g, c, and ρ such that the resulting hash function h is an injection on I_s. □

CLAIM 2.9. *For every $t \in [s-1]$, if t satisfies that $v_{i_{t+1}}^2 \le \frac{1}{64s} \cdot \|v_{[n] \setminus I_t}\|_2^2$, then with probability at least $\frac{2}{3}$, f satisfies that*

$$\sum_{r \in [s]} \min \left\{ \|v_{(f \circ C_y)^{-1}(r) \setminus I_t}\|_2^2, \frac{2}{s} \cdot \|v_{[n] \setminus I_t}\|_2^2 \right\} \ge \frac{1}{2} \cdot \|v_{[n] \setminus I_t}\|_2^2. \tag{2.4}$$

[3] A careful calculation shows that the constant A in the statement of Lemma 1.1 is at most 5.

Observe that Equation (2.4) implies Equation (1.1), as the g_i-s only further split hash buckets.

The intuition behind this claim is simple: If $v_{i_{t+1}}^2 \leq \frac{1}{64s} \cdot \|v_{[n]\setminus I_t}\|_2^2$, then no $i \in [n] \setminus I_t$ has v_i^2 very large relative to $\|v_{[n]\setminus I_t}\|_2^2$, so $\|v_{[n]\setminus I_t}\|_2^2$ is spread roughly evenly on many coordinates. As $f \circ C_y$ is likely to map the coordinates of $v_{[n]\setminus I_t}$ roughly evenly, it also maps the weight $\|v_{[n]\setminus I_t}\|_2^2$ roughly evenly.

Proof. To ease the reading, let us use the following notation. Let $u \in \mathbb{R}^{2^b}$ be defined as follows. For $z \in \{0,1\}^b$,

$$u_z = \|v_{C_y^{-1}(z)\setminus I_t}\|_2^2 \ .$$

Let us also denote

$$W = \|u\|_1 = \|v_{[n]\setminus I_t}\|_2^2 \ .$$

With these notations, what we wish to prove is that with probability at least $\frac{2}{3}$, over the choice of f,

$$\sum_{r\in[s]} \min\left\{\|u_{f^{-1}(r)}\|_1, \frac{2}{s}W\right\} \geq \frac{1}{2}W \ . \tag{2.5}$$

Since y is good, Corollary 2.7 guarantees that for all z,

$$u_z \leq \left(\frac{1}{64s} + \sqrt{\epsilon}\right) \cdot \|v_{[n]\setminus I_t}\|_2^2 = \left(\frac{1}{64s} + \sqrt{\epsilon}\right) \cdot W < \frac{W}{7.99^2 s} \ .$$

Let X_z^i be the indicator random variable for the event that $f(z) = i$. As $\Pr[f(z) = i] = \frac{1}{s}$, we have that

$$\mathbb{E}\left[\|u_{f^{-1}(i)}\|_1\right] = \sum_{z=1}^{2^b} \mathbb{E}[X_z^i] \cdot u_z = \frac{1}{s}\|u\|_1 = \frac{1}{s}W \ .$$

Moreover, as f comes from a pairwise independent family of hash functions, for fixed i the random variables X_z^i are pairwise independent, so

$$\begin{aligned}\sigma^2\left[\|u_{f^{-1}(i)}\|_1\right] &= \mathrm{Var}\left[\|u_{f^{-1}(i)}\|_1\right] = \mathrm{Var}\left[\sum_{z=1}^{2^b} X_z^i \cdot u_z\right] \\ &= \sum_{z=1}^{2^b} \mathrm{Var}[X_z^i] \cdot u_z^2 = \left(1 - \frac{1}{s}\right) \cdot \frac{1}{s} \cdot \|u\|_2^2 \ .\end{aligned}$$

Thus, as $u_z \leq \frac{W}{7.99^2 s}$, we get that

$$\sigma\left[\|u_{f^{-1}(i)}\|_1\right] \leq \frac{1}{\sqrt{s}} \cdot \|u\|_2 \leq \frac{1}{\sqrt{s}} \cdot \sqrt{\frac{W}{7.99^2 s}} \cdot \sqrt{\|u\|_1} = \frac{W}{7.99s} \ .$$

By Chebyshev's inequality, for every $r > 1$,

$$\Pr\left[\|u_{f^{-1}(i)}\|_1 \geq \frac{r}{s}W\right] \leq \frac{1}{7.99^2(r-1)^2} \ . \tag{2.6}$$

For each value of r, consider all values λ in the interval $[2^r, 2^{r+1}]$ such that $\Pr\left[\|u_{f^{-1}(i)}\|_1 = \frac{\lambda}{s}W\right] \neq 0$. Clearly there are finitely many such values. From Equation (2.6) we get that

$$\begin{aligned}\sum_{\lambda\in[2^r,2^{r+1}]} \lambda \cdot \Pr\left[\|u_{f^{-1}(i)}\|_1 = \frac{\lambda}{s}W\right] &\leq 2^{r+1}\Pr\left[\|u_{f^{-1}(i)}\|_1 \geq \frac{2^r}{s}\cdot W\right] \\ &\leq \frac{2^{r+1}}{7.99^2(2^r-1)^2} \ .\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{E}\left[\max\left\{0, \|u_{f^{-1}(i)}\|_1 - \frac{2}{s}W\right\}\right] &\leq \frac{W}{7.99^2 s}\cdot\sum_{r=1}^{\infty}\frac{2^{r+1}}{(2^r-1)^2} \\ &= \frac{4W}{7.99^2 s}\cdot\sum_{r=1}^{\infty}\frac{2^{r-1}}{(2^r-1)^2} \\ &< \frac{4W}{7.99^2 s}\cdot\sum_{r=1}^{\infty}\frac{1}{2^{r-1}} \\ &= \frac{8W}{7.99^2 s} \ .\end{aligned}$$

Let $Y^i = \max\left\{0, \|u_{f^{-1}(i)}\|_1 - \frac{2}{s}W\right\}$. We just showed that $\mathbb{E}\left[\sum_{i\in[s]} Y^i\right] < \frac{8W}{7.99^2}$, so by Markov's Inequality, $\Pr\left[\sum_{i\in[s]} Y^i > \frac{1}{2}W\right] < \frac{1}{3}$.

We next show that when $\sum_{i\in[s]} Y^i \leq \frac{1}{2}W$, Equation (2.5) holds, and thus it holds with probability at least $\frac{2}{3}$ over the choice of $f \in \mathcal{F}$, which is what we wanted to prove.

Let m be the number of $i \in [s]$ such that $Y^i > 0$. We now get that

$$\frac{1}{2}W \geq \sum_{i\in[s]} Y^i = \sum_{i:Y^i>0}\left(\|u_{f^{-1}(i)}\|_1 - \frac{2}{s}W\right) = \left(\sum_{i:Y^i>0}\|u_{f^{-1}(i)}\|_1\right) - \frac{2m}{s}W \ .$$

Hence,

$$\begin{aligned}\sum_{i=1}^{s}\min\left\{\|u_{f^{-1}(i)}\|_1, \frac{2}{s}W\right\} &= \left(\sum_{i:Y^i=0}\|u_{f^{-1}(i)}\|_1\right) + \frac{2m}{s}W \\ &= \left(W - \sum_{i:Y^i>0}\|u_{f^{-1}(i)}\|_1\right) + \frac{2m}{s}W \\ &\geq W - \frac{1}{2}W = \frac{1}{2}W \ ,\end{aligned}$$

and Equation (2.5) holds. □

To conclude the proof of Lemma 1.1 we recall that y is good with probability at least $1 - s\sqrt{\epsilon} > 0$ and that for each good y, Claim 2.8 holds for a random choice of f with probability at least $\frac{1}{2}$. Furthermore, for a good y, we have that Equation (2.5) holds for at least $\frac{2}{3}$ of the choices of $f \in \mathcal{F}$. As $\frac{2}{3} + \frac{1}{2} > 1$, we get that for each good y, there is a good choice of f, so that both Equation (2.5) and the condition in the statement of Claim 2.8 hold. This is exactly what Lemma 1.1 claims.

3. Proof of Corollary 1.2 (the new version of Corollary 2.8). We consider two cases.

Case 1. There is some $\lceil 2s/3\rceil \leq q \leq s-1$ such that $v_{i_{q+1}}^2 \leq \frac{1}{64s}\cdot\|v_{[n]\setminus I_q}\|_2^2$.

Case 2. For every $\lceil 2s/3\rceil \leq q \leq s-1$ we have that $v_{i_{q+1}}^2 > \frac{1}{64s}\cdot\|v_{[n]\setminus I_q}\|_2^2$.

Consider Case 1. By the assumption in Case 1 we get from Lemma 1.1 that there exists $h\in\mathcal{H}$ such that Eq (1.1) is satisfied. We will show that for some constants c_1, c_2 at least $c_1\cdot 8s$ buckets satisfy that $\|v_{h^{-1}(r)\setminus I_q}\|_2^2 \geq \frac{c_2}{s}\cdot\|v_{[n]\setminus I_q}\|_2^2$. Assume for a contradiction that less than $c_1\cdot 8s$ buckets have high norm. Hence,

$$\frac{1}{2}\cdot\|v_{[n]\setminus I_q}\|_2^2 \leq \sum_{r\in[8s]}\min\left\{\|v_{h^{-1}(r)\setminus I_q}\|_2^2, \frac{2}{s}\cdot\|v_{[n]\setminus I_q}\|_2^2\right\}$$
$$\leq c_1\cdot 8s\cdot\frac{2}{s}\cdot\|v_{[n]\setminus I_q}\|_2^2 + 8s\cdot\frac{c_2}{s}\cdot\|v_{[n]\setminus I_q}\|_2^2 = (16c_1+8c_2)\cdot\|v_{[n]\setminus I_q}\|_2^2.$$

Therefore, for $c_1 = \frac{1}{48}$ and $c_2=\frac{1}{49}$ we get a contradiction, unless $\|v_{[n]\setminus I_q}\|_2^2 = 0$. However, the claim is trivial if this is the case.

Let us now assume that we are in Case 2. It follows that

$$\sum_{q=\lceil 2s/3\rceil}^{s-1}|v_{i_{q+1}}| \geq \sum_{q=\lceil 2s/3\rceil}^{s-1}\frac{1}{8\sqrt{s}}\cdot\|v_{[n]\setminus I_q}\|_2 \geq \sum_{q=\lceil 2s/3\rceil}^{s-1}\frac{1}{8\sqrt{s}}\cdot\|v_{[n]\setminus I_s}\|_2 \geq \frac{\sqrt{s}}{32}\|v_{[n]\setminus I_s}\|_2,$$

where in the last inequality we used the assumption that $s\geq 24$.

4. Analysis of "Case 3" from the proof of Theorem 1.1. *Case* 3. We now assume that $\sum_{r=\lceil 2t/3\rceil}^{t-1}|v_{i_{r+1}}| < \frac{\sqrt{t}}{32}\|v_{[n]\setminus I_t}\|_2$. Hence, Corollary 1.2 implies that there exist $\lceil 2t/3\rceil \leq q\leq t-1$ and some $h\in\mathcal{H}$ such that h is an injection on I_t, and for at least $c_1\cdot 8t$ buckets $r\in[8t]$ it holds that $\|v_{h^{-1}(r)\setminus I_q}\|_2^2 \geq \frac{c_2}{t}\cdot\|v_{[n]\setminus I_q}\|_2^2$ for two universal constants c_1 and c_2. Denote the set of $\geq c_1\cdot 8t$ "good" buckets r with $R\subset[8t]$. We also define, for every $i\in[8t]$, $J'_{h,i} = h^{-1}(i)\setminus I_q$. It follows that for every $i\in R$

$$\|v_{J'_{h,i}}\|_2 \geq \sqrt{\frac{c_2}{t}}\cdot\|v_{[n]\setminus I_q}\|_2.$$

By Lemma 2.5 (in [RS10]), specialized to $k=5$, we get that for every $i\in h(I_q)$

$$\Pr_{s\in S_{h,i}}\left[\langle s, v_{J_{h,i}}\rangle \geq \|v_{I_q\cap J_{h,i}}\|_1 + \frac{1}{7}\|v_{J'_{h,i}}\|_2\right] \geq \frac{4}{5}\cdot 2^{-5} = \frac{1}{40}, \tag{4.1}$$

where we recall that by our assumption on h we have that $|I_q\cap J_{h,i}| = 1$. In addition, Lemma 2.4 (in [RS10]) implies that for $i\notin h(I_q)$ (this actually holds for every i)

$$\Pr_{s\in S_{h,i}}\left[\langle s, v_{J_{h,i}}\rangle \geq \frac{\|v_{J_{h,i}}\|_2}{7}\right] \geq \frac{1}{20}. \tag{4.2}$$

For every $i\in[8t]$ denote with $A_i\subseteq S_{h,i}$ the set of $s\in S_{h,i}$ that belong to the "good" sets defined in (4.1), (4.2), namely, those elements from $S_{h,i}$ that have large inner product with $v_{J_{h,i}}$. Clearly, for every i we have that $|A_i|/|S_{h,i}| \geq \min(\frac{1}{40},\frac{1}{20}) = \frac{1}{40}$. We will now show that there exists a walk on G such that for every i, $w_i\in A_i$. Indeed, G is an $[m,d,\lambda]$-expander and so Theorem 2.6 (in [RS10]) guarantees that if $\frac{1}{40} > 6\lambda/d$, then there exists a walk that hits all the A_i's. As we picked a graph G with $\lambda\leq d/1000$ we have the required property.

Thus, there exists a walk $w = (w_1, \ldots, w_{8t})$ such that for every i, $w_i \in A_i$. Calculating, we get that

$$\begin{aligned}
\left\langle x^{h,w}, v \right\rangle &= \sum_{i=1}^{8t} \left\langle w_i, v_{J_{h,i}} \right\rangle \\
&= \sum_{i \in h(I_q)} \left\langle w_i, v_{J_{h,i}} \right\rangle + \sum_{i \notin h(I_q)} \left\langle w_i, v_{J_{h,i}} \right\rangle \\
&\geq \sum_{i \in h(I_q)} \left(\|v_{I_q \cap J_{h,i}}\|_1 + \frac{1}{7} \|v_{J'_{h,i}}\|_2 \right) + \sum_{i \notin h(I_q)} \frac{\|v_{J_{h,i}}\|_2}{7} \\
&= \|v_{I_q}\|_1 + \frac{1}{7} \sum_{i \in [8t]} \|v_{J'_{h,i}}\|_2 \geq \|v_{I_q}\|_1 + \frac{1}{7} \sum_{i \in R} \|v_{J'_{h,i}}\|_2 \\
&\geq \|v_{I_q}\|_1 + \frac{1}{7} \sum_{i \in R} \sqrt{\frac{c_2}{t}} \cdot \|v_{[n] \setminus I_q}\|_2 \\
&\geq^{\ddagger} \|v_{I_q}\|_1 + \frac{8c_1\sqrt{c_2}}{7} \cdot \sqrt{t} \cdot \|v_{[n] \setminus I_q}\|_2 \\
&\geq \|v_{I_q}\|_1 + \frac{8c_1\sqrt{c \cdot c_2}}{7} \cdot \sqrt{\log(2/\epsilon)} \cdot \|v_{[n] \setminus I_q}\|_2 \\
&\geq^{*} \|v_{I_q}\|_1 + \sqrt{2\log(2/\epsilon)} \cdot \|v_{[n] \setminus I_q}\|_2 >^{\dagger} \theta,
\end{aligned}$$

where inequality ($\ddagger$) follows from the fact that $|R| \geq c_1 \cdot 8t$, inequality ($*$) holds for a large enough universal constant c, and inequality ($\dagger$) holds from the same argument as in case 2 (for c large enough), recalling that $q \geq \lceil 2t/3 \rceil = \lceil \frac{2}{3} c \log 2/\epsilon \rceil$. Thus, $L_{v,\theta}(x^{h,w}) = 1$ as required.

Acknowledgments:. We are grateful to William Hoza who found a mistake in our original argument. We also wish to thank William Hoza and the anonymous reviewers for comments that helped us significantly improve the presentation of the proof.

REFERENCES

[FKS84] M. L. Fredman, J. Komlós, and E. Szemerédi. Storing a sparse table with 0(1) worst case access time. *J. ACM*, 31(3):538–544, 1984.

[GUV09] V. Guruswami, C. Umans, and S. Vadhan. Unbalanced expanders and randomness extractors from Parvaresh–Vardy codes. *J. ACM*, 56(4):1–34, 2009.

[RS10] Y. Rabani and A. Shpilka. Explicit construction of a small ϵ-net for linear threshold functions. *SIAM J. on Computing*, 39(8):3501–3520, 2010.

[SS90] J. P. Schmidt and A. Siegel. The analysis of closed hashing under limited randomness (extended abstract). In *Proceedings of the 22nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 224–234, 1990.

[Vad12] S. P. Vadhan. Pseudorandomness. *Foundations and Trends in Theoretical Computer Science*, 7(1-3):1–336, 2012.