

Improved Monotonicity Testers via Hypercube Embeddings

Mark Braverman ✉

Department of Computer Science, Princeton University, NJ, USA

Subhash Khot ✉

Courant Institute of Mathematical Sciences, New York University, NY, USA

Guy Kindler ✉

Engineering and Computer Science Department, The Hebrew University, Jerusalem, USA

Dor Minzer ✉

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

We show improved monotonicity testers for the Boolean hypercube under the p -biased measure, as well as over the hypergrid $[m]^n$. Our results are:

1. For any $p \in (0, 1)$, for the p -biased hypercube we show a non-adaptive tester that makes $\tilde{O}(\sqrt{n}/\varepsilon^2)$ queries, accepts monotone functions with probability 1 and rejects functions that are ε -far from monotone with probability at least $2/3$.
2. For all $m \in \mathbb{N}$, we show an $\tilde{O}(\sqrt{nm^3}/\varepsilon^2)$ query monotonicity tester over $[m]^n$.

We also establish corresponding directed isoperimetric inequalities in these domains, analogous to the isoperimetric inequality in [15]. Previously, the best known tester due to Black, Chakrabarty and Seshadhri [2] had $\Omega(n^{5/6})$ query complexity. Our results are optimal up to poly-logarithmic factors and the dependency on m .

Our proof uses a notion of monotone embeddings of measures into the Boolean hypercube that can be used to reduce the problem of monotonicity testing over an arbitrary product domains to the Boolean cube. The embedding maps a function over a product domain of dimension n into a function over a Boolean cube of a larger dimension n' , while preserving its distance from being monotone; an embedding is considered efficient if n' is not much larger than n , and we show how to construct efficient embeddings in the above mentioned settings.

2012 ACM Subject Classification Theory of computation $\rightarrow$ Streaming, sublinear and near linear time algorithms; Mathematics of computing $\rightarrow$ Combinatorics

Keywords and phrases Property Testing, Monotonicity Testing, Isoperimetric Inequalities

Digital Object Identifier 10.4230/LIPIcs.ITCS.2023.25

Funding *Mark Braverman*: Research supported in part by the NSF Alan T. Waterman Award, Grant No. 1933331, a Packard Fellowship in Science and Engineering, and the Simons Collaboration on Algorithms and Geometry.

Subhash Khot: Supported by the NSF Award CCF-1422159, NSF Award CCF-2130816, and the Simons Investigator Award.

Guy Kindler: supported by Israel Science Foundation grant no. 2635/19.

Dor Minzer: Supported by a Sloan Research Fellowship.

1 Introduction

Let $[m] = \{0, 1, \dots, m-1\}$ be thought of as an ordered set, and consider the partial ordering induced by it on $[m]^n$: two points $x, y \in [m]^n$ satisfy $x \leq y$ if and only if $x_i \leq y_i$ for all $i = 1, \dots, n$. A function $f: [m]^n \rightarrow \{0, 1\}$ is called *monotone* if for every $x, y \in [m]^n$ such that $x \leq y$ we have $f(x) \leq f(y)$. Given a function $f: [m]^n \rightarrow \{0, 1\}$, we measure its distance from being monotone, with respect to a probability measure μ over $[m]^n$, by



© Mark Braverman, Subhash Khot, Guy Kindler, and Dor Minzer;
licensed under Creative Commons License CC-BY 4.0

14th Innovations in Theoretical Computer Science Conference (ITCS 2023).

Editor: Yael Tauman Kalai; Article No. 25; pp. 25:1–25:24



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

$$\varepsilon(f; \mu) = \min_{g: [m]^n \rightarrow \{0,1\} \text{ monotone}} \Delta(f, g; \mu), \text{ where } \Delta(f, g; \mu) = \mu(\{x \in [m]^n \mid f(x) \neq g(x)\}).$$

In this paper we present monotonicity testers for functions over $[m]^n$ (under the uniform measure), as well as over the Boolean hypercube $\{0, 1\}^n$ with the p -biased measure, defined as $\mu_p^{\otimes n}(x) = p^{|x|}(1-p)^{n-|x|}$. That is, we construct a randomized algorithm that makes oracle queries to an unknown function f over the domain, which accepts with probability 1 if f is monotone, and rejects with probability $\geq \frac{2}{3}$ if f is ε -far from monotone.

1.1 Prior Works

The monotonicity testing problem has received significant attention over the years, as we shall now review. For simplicity, below we think of ε as being a small constant. The problem was originally studied over Boolean hypercube with the uniform measure [13], where a non-adaptive algorithm that makes $O(n)$ queries was shown. This bound was improved by [8], who showed an $\tilde{O}(n^{7/8})$ query tester by proving a directed version of Margulis' isoperimetric inequality [17] and using it towards developing improved monotonicity testers. Following it, Chen, Servedio and Tan [9] modified the algorithm and the analysis of [8] and established a $\tilde{O}(n^{5/6})$ query tester. Finally, an $\tilde{O}(\sqrt{n})$ query tester was given in [15], who proved a directed version of an isoperimetric inequality due to Talagrand [19]. The tester of [15] is tight up to poly-logarithmic factors, by a bound for non-adaptive testers due to [12]. The best known lower bound for adaptive testers is not too far off [10], and currently stands at $\tilde{\Omega}(n^{1/3})$.

Following the investigation of complexity of monotonicity testing over the hypercube, variants of the problem were also considered in the literature, in which either the domain or the range of the function are different [11, 7, 2, 3, 5]. Most relevant to us is the monotonicity testing problem over different measures on the Boolean hypercube as well as the closely related hypergrid $[m]^n$, wherein the state of the art result is an $O(n^{5/6} \text{poly}(\log m))$ query tester due to [2]. To prove their result, the authors of [2] established an analog of the directed isoperimetric inequality of [8] for the hypergrid.

1.2 Parallel Works

Following initial submission of this paper, we have learned that Black, Chakrabarty and Seshadhri have independently obtained results similar to ours [4]. They use a different method, first proving analogous directed isoperimetric inequalities over the hypergrid, and then using these to construct and analyze a monotonicity tester for the hypergrid.

1.3 Main Results

Our first result is an essentially-optimal monotonicity tester for the p -biased cube:

► **Theorem 1.** *For every $p \in (0, 1)$, there is a non-adaptive monotonicity tester over $(\{0, 1\}^n, \mu_p^{\otimes n})$ that makes $\tilde{O}(\sqrt{n}/\varepsilon^2)$ queries.*

Second, we focus on the hypergrid $[m]^n$. Here and throughout, U_m refers to the uniform distribution over $[m]$, and we often drop the subscript m when it is clear from the context.

► **Theorem 2.** *For all $m, n \in \mathbb{N}$, there is a non-adaptive monotonicity tester over $([m]^n, U^{\otimes n})$ that makes $\tilde{O}(\sqrt{nm}^3/\varepsilon^2)$ queries.*

Our techniques also imply analogs of the directed isoperimetric inequality of [15] for the hypergrid as well as for the p -biased cube. For simplicity we state the result for the hypergrid, and defer the statement for the p -biased cube to Theorem 13.

Let $f: [m]^n \rightarrow \{0, 1\}$ be a function and fix an input $x \in [m]^n$. The negative sensitivity of f at x , denoted by $s_f^-(x)$, is defined to be the number of coordinates i such that there is an input y differing from x only on the i^{th} coordinate, such that the pair (x, y) violates monotonicity. Namely, it is the number of coordinates i such that for the point y which differs from x only on its i^{th} coordinate, we have that $x < y$ and $f(x) > f(y)$ (if $f(x) = 1$) or $x > y$ and $f(x) < f(y)$ (if $f(x) = 0$).

► **Theorem 3.** *If $f: [m]^n \rightarrow \{0, 1\}$ is ε -far from monotone with respect to $U^{\otimes n}$, then*

$$\mathbb{E}_{x \in [m]^n} \left[\sqrt{s_f^-(x)} \right] \geq \Omega \left(\frac{\varepsilon}{m^3 \log(mn/\varepsilon)^2} \right).$$

1.4 Our Technique

Our proofs rely on the following elementary notion of an embedding of a domain (that we wish to test monotonicity over) into a hypercube of not too-large dimension.

► **Definition 4.** *We say that a probability distribution $([m], \mu_1)$ can be r -locally embedded if there is a Boolean hypercube $\{0, 1\}^r$, a map $\phi: \{0, 1\}^r \rightarrow [m]$, a collection $\Psi = \{\Psi_\omega\}_{\omega \in \Omega}$ of maps $\Psi_\omega: [m] \rightarrow \{0, 1\}^r$, and a probability distribution P over Ω such that:*

1. *Each one of ϕ and Ψ_ω is monotone.*
2. *Sampling $x \sim U(\{0, 1\}^r)$, the distribution of $\phi(x)$ is μ_1 .*
3. *Sampling $y \sim \mu_1$ and $\omega \sim P$, the distribution of $\Psi_\omega(y)$ is uniform over $\{0, 1\}^r$.*
4. *For each $\omega \in \Omega$, the composition $\phi \circ \Psi_\omega$ is the identity on $[m]$.*

The usefulness of Definition 4 comes from the fact that given a local embedding of $[m]$, we can reduce the problem of testing monotonicity over $([m]^n, \mu_1^{\otimes n})$ to that of testing it over Boolean hypercubes of dimension rn , which we already know how to solve. Towards showing the reduction we note that if $([m], \mu_1)$ can be r -locally embedded, then given a function $f: [m]^n \rightarrow \{0, 1\}$ we may define $g: \{0, 1\}^{r \times n} \rightarrow \{0, 1\}$ by

$$g(x(1), \dots, x(n)) = f(\phi(x(1)), \dots, \phi(x(n))).$$

The following lemma asserts that if f is monotone then g is also monotone, and if f is ε -far from monotone, then g is ε -far from monotone.

► **Lemma 5.** *If f is monotone, then g is monotone. Moreover, $\varepsilon(g; U^{\otimes n}) \geq \varepsilon(f; \mu_1^{\otimes n})$.*

Proof. Assume f is monotone. Then taking any $(x(1), \dots, x(n)) \leq (x(1)', \dots, x(n)')$ in $\{0, 1\}^{r \times n}$, by the monotonicity of ϕ we have $(\phi(x(1)), \dots, \phi(x(n))) \leq (\phi(x(1)'), \dots, \phi(x(n)'))$, and using the monotonicity of f we get that $g(x(1), \dots, x(n)) \leq g(x(1)', \dots, x(n)').$

For the other direction, let g' be the closest monotone function to g , and choose $\vec{\omega} = (\omega_1, \dots, \omega_n) \sim P^{\otimes n}$. Define

$$f_{\vec{\omega}}(x_1, \dots, x_n) = g'(\Psi_{\omega_1}(x_1), \dots, \Psi_{\omega_n}(x_n)).$$

Since each Ψ_{ω_i} is monotone we have that $f_{\vec{\omega}}$ is monotone as well. Also,

$$\begin{aligned} \mathbb{E}_{\vec{\omega}} [\Delta(f, f_{\vec{\omega}}; \mu_1^{\otimes n})] &= \mathbb{E}_{\vec{\omega}} \left[\mathbb{E}_{x \sim \mu_1^{\otimes n}} [1_{f(x) \neq f_{\vec{\omega}}(x)}] \right] = \mathbb{E}_{\vec{\omega}} \left[\mathbb{E}_{x \sim \mu_1^{\otimes n}} [1_{f(x) \neq g'(\Psi_{\omega_1}(x_1), \dots, \Psi_{\omega_n}(x_n))}] \right] \\ &= \mathbb{E}_{\vec{\omega}} \left[\mathbb{E}_{x \sim \mu_1^{\otimes n}} [1_{g(\Psi_{\omega_1}(x_1), \dots, \Psi_{\omega_n}(x_n)) \neq g'(\Psi_{\omega_1}(x_1), \dots, \Psi_{\omega_n}(x_n))}] \right], \end{aligned}$$

where in the last equality we used the fact that $\phi \circ \Psi_\omega$ is the identity. Note that by property 3 of an embedding, given the distribution of $\vec{\omega}$ and x , the distribution of $(\Psi_{\omega_1}(x_1), \dots, \Psi_{\omega_n}(x_n))$ is uniform over $\{0, 1\}^{r \times n}$, so the last expression is equal to $\Delta(g, g'; U^{\otimes n}) = \varepsilon(g; U^{\otimes n})$. It follows that there is an $\vec{\omega}$ such that $\Delta(f, f_{\vec{\omega}}; \mu_1^{\otimes n}) \leq \varepsilon(g; U)$, and the proof is concluded. $\blacktriangleleft$

For the Boolean hypercube with the uniform measure, a 2-query path tester is constructed in [15] which always accepts monotone functions, and rejects functions that are ε -far from monotone with probability at least

$$R(n, \varepsilon) = \frac{\varepsilon^2}{\sqrt{n} \text{poly}(\log n)}. \quad (1)$$

Combining that tester with Lemma 5 we get the following conclusion:

► **Lemma 6.** *Suppose that $([m], \mu_1)$ can be r -locally embedded; then there is a 2-query monotonicity testing algorithm for functions over $([m]^n, \mu_1^{\otimes n})$ that always accepts monotone functions, and rejects functions that are ε -far from monotone with probability at least $R(rn, \varepsilon)$.*

Proof. Given $f: ([m]^n, \mu_1^{\otimes n}) \rightarrow \{0, 1\}$, define g as above, then run the monotonicity tester of the hypercube on g , and accept/reject accordingly. Note that a single query to g can be answered by making a single query to f . By Lemma 5, if f is monotone then g is monotone, hence the tester always accepts. If f is ε -far from monotone, then by Lemma 5 g is also ε -far from monotone, hence the tester rejects with probability at least $R(rn, \varepsilon)$. $\blacktriangleleft$

Thus, Theorems 1 and 2 follow from Lemma 6 once we show the existence of sufficiently good local embeddings. In Section 2 we show constructions of such embeddings for the p -biased measure on $\{0, 1\}$, as well as basic embeddings for $[m]^n$ which are not good enough for our purpose (but gives some intuition). To construct efficient embeddings for $[m]^n$ we have to work harder, and for divisibility reasons we only know how to construct such embeddings for m 's that are power of 2. For other m 's, we have to consider a slightly relaxed notion of embeddings, asserting that there are distributions μ'_1 and μ'_2 that are extremely close to the distributions $([m], \mu_1)$ and $(\{0, 1\}^r, U)$ such that one can embed $([m], \mu'_1)$ into $(\{0, 1\}^r, \mu'_2)$; see Sections 2.3, 2.4 for the formal definition. This relaxed notion has the same monotonicity testing and directed isoperimetric implications. The construction of embeddings for the hypergrid is more involved than our construction of embeddings for the p -biased cube, and can be found in Section 3.

As for the directed isoperimetric inequalities, we recall the isoperimetric result from [15]

► **Theorem 7.** *If $f: (\{0, 1\}^n, U^{\otimes n}) \rightarrow \{0, 1\}$ is ε -far from monotone, then $\mathbb{E}_x \left[\sqrt{s_f^-(x)} \right] \geq \Omega\left(\frac{\varepsilon}{\log(n/\varepsilon)}\right)$.*

Combining Theorem 7 with Lemma 5 we get:

► **Lemma 8.** *Suppose that $([m], \mu_1)$ can be r -locally embedded; then for any $f: ([m]^n, \mu_1^{\otimes n}) \rightarrow \{0, 1\}$ that is ε -far from monotone it holds that $\mathbb{E}_{y \sim \mu_1^{\otimes n}} \left[\sqrt{s_f^-(y)} \right] \geq \Omega\left(\frac{\varepsilon(f)}{\sqrt{r} \log(rn/\varepsilon(f))}\right)$.*

Proof. Define $g: \{0, 1\}^{r \times n} \rightarrow \{0, 1\}$ as above, and note that $\frac{1}{r} s_g^-(x) \leq s_f^-(\phi(x))$ for all $x \in \{0, 1\}^{r \times n}$. Indeed, letting $k = \frac{1}{r} s_g^-(x)$ and viewing $x = (x(1), \dots, x(n))$ where $x(i) \in \{0, 1\}^r$, there are at least k i 's such that there is x' such that $x'(j) = x(j)$ for all $j \neq i$ and the pair x, x' violates monotonicity of g . In that case, we see that the pair $y = \phi(x) = (\phi(x(1)), \dots, \phi(x(n)))$ and $y' = \phi(x') = (\phi(x'(1)), \dots, \phi(x'(n)))$ only differ in their i^{th} coordinate and violate monotonicity of f , hence $s_f^-(\phi(x)) \geq k$. It follows that

$$\begin{aligned}
\mathbb{E}_{y \sim \mu_1^{\otimes n}} \left[\sqrt{s_f^-(y)} \right] &= \mathbb{E}_{x \in \{0,1\}^{r \times n}} \left[\sqrt{s_f^-(\phi(x))} \right] \geq \frac{1}{\sqrt{r}} \mathbb{E}_{x \in \{0,1\}^{r \times n}} \left[\sqrt{s_g^-(x)} \right] \\
&\geq \frac{1}{\sqrt{r}} \Omega \left(\frac{\varepsilon(g)}{\log(rn/\varepsilon(g))} \right),
\end{aligned}$$

and the proof is concluded by Lemma 5. $\blacktriangleleft$

We note that Theorem 3 follows from Lemma 8 (or rather, a slight variant of it which is suitable for slightly relaxed embeddings) by showing that $([m], U)$ can be r -locally embedded for $r = O(m^6)$ (under the aforementioned slightly relaxed notion of embeddings).

2 Elementary Constructions of Embeddings

In this section we present several ideas for constructing local embeddings and prove Theorem 1.

2.1 Embedding p -biased Cubes

We begin by constructing some basic embeddings for p -biased distributions over $\{0,1\}$, and then combining them to prove Theorem 1. First, we show that the measure μ_p can be locally embedded when p is a powers of 2.

► Lemma 9. *Let $p = 2^{-r}$, and consider the distribution μ_p over $\{0,1\}$ where $\mu_p(1) = p$. Then $(\{0,1\}, \mu_p)$ can be r -locally embedded.*

Proof. We define $\phi(x_1, \dots, x_r) = x_1 \wedge \dots \wedge x_r$. As for Ψ , we take the distribution (Ω, P) to be uniform over $\{0,1\}^r \setminus \{\vec{1}\}$, and define $\Psi_\omega(1) = (1, \dots, 1)$ and $\Psi_\omega(0) = \omega$. $\blacktriangleleft$

Secondly, we show that if μ_p can be locally embedded, then so can μ_{1-p} .

► Lemma 10. *Let $p \in (0,1)$, and suppose $(\{0,1\}, \mu_p)$ can be r -locally embedded. Then $(\{0,1\}, \mu_{1-p})$ can be r -locally embedded.*

Proof. Let $(\phi, \{\Psi_\omega\}_{\omega \in \Omega}, P)$ be an r -local embedding of μ_p . Define $\phi'(x) = 1 - \phi(1-x)$ and $\Psi'_\omega(a) = \vec{1} - \Psi_\omega(1-a)$. First, note that ϕ' and Ψ'_ω are monotone. Second, sampling $x \sim U$, $\phi'(x)$ is distributed according to μ_{1-p} . Also,

$$\phi'(\Psi'_\omega(a)) = 1 - \phi(1 - \Psi'_\omega(a)) = 1 - \phi(\Psi_\omega(1-a)) = 1 - (1-a) = a.$$

Finally, if $a \sim \mu_{1-p}$, then $1-a \sim \mu_p$, hence $\Psi_\omega(1-a) \sim U$ and so $\Psi'_\omega(a) \sim U$. $\blacktriangleleft$

Third, we show how $\mu_{p_1 p_2}$ can be locally embedded given local embeddings for μ_{p_1} and μ_{p_2} .

► Lemma 11. *Suppose that μ_{p_1} can be r_1 locally embedded, and μ_{p_2} can be r_2 locally embedded. Then $\mu_{p_1 p_2}$ can be $r_1 + r_2$ locally embedded.*

Proof. Let $(\phi_1, \{\Psi_{1,\omega}\}_{\omega \in \Omega_1}, P_1)$ and $(\phi_2, \{\Psi_{2,\omega}\}_{\omega \in \Omega_2}, P_2)$ be the local embeddings of μ_{p_1} and μ_{p_2} , respectively. We define $\phi: \{0,1\}^{r_1+r_2} \rightarrow \{0,1\}$ by $\phi(x,y) = \phi_1(x) \wedge \phi_2(y)$.

Now let $\Omega' = \{0,1\}^2 \setminus \{(1,1)\}$, and define P' to be the distribution obtained by taking $(a,b) \sim \mu_{p_1} \times \mu_{p_2}$ and conditioning on the event $[(a,b) \neq (1,1)]$. We take $\Omega = \Omega_1 \times \Omega_2 \times \Omega'$ and $P = P_1 \times P_2 \times P'$. For $w = (\omega_1, \omega_2, \omega') \in \Omega$ where $w' = (a,b)$, we finally define Ψ_ω as follows:

$$\Psi_{(\omega_1, \omega_2, \omega')}(1) = (\Psi_{1, \omega_1}(1), \Psi_{2, \omega_2}(1)),$$

and

$$\Psi_{(\omega_1, \omega_2, \omega')}(0) = (\Psi_{1, \omega_1}(a), \Psi_{2, \omega_2}(b)).$$

It is clear that ϕ and Ψ are monotone, that $\phi \circ \Psi_{\vec{\omega}} = \text{identity}$, and that the distributions are correct. $\blacktriangleleft$

Next, by an easy approximation argument we conclude that for all values of p there is some p' close to p such that $\mu_{p'}$ can be locally embedded.

► **Corollary 12.** *For all $\delta > 0$ and $p \in (0, 1)$, there exists a $p' \in (0, 1)$ such that $|p - p'| \leq \frac{\delta}{n^{10}}$ and that $\mu_{p'}$ can be $O(\log^2(n/\delta))$ -locally embedded.*

Proof. By Lemma 10, we may assume that $p \leq 1/2$. Set $s = \lceil 10 \log(n/\delta) \rceil$, and for $a \in \mathbb{Z}^s$, define $q(\vec{a}) = \prod_{i=1}^s (1 - 2^{-i})^{a_i}$.

Below, we show that there exists a vector $a \in \mathbb{Z}^s$ such that $p' = q(a)$ satisfies

$$p \leq p' \leq p + \frac{\delta}{n^{10}}, \quad (2)$$

and where $a_1 \in \{1, \dots, s\}$ and $a_i \in \{0, 1, 2, 3\}$ for any $i > 1$. Note that this implies, by Lemmas 9, 10, and 11, that p' can be r -locally embedded for $r \leq s + O(s^2) = O(\log^2(n/\delta))$, finishing the proof.

To find the required vector a , we begin by taking k to be the maximal number that satisfies $\frac{1}{2^k} \geq p$. If $k \geq s$ we set $a_1 = s$, and note that we are done since the vector $a = (a_1, 0, \dots, 0)$ satisfies (2) as required. Otherwise if $k < s$, we continue to set $a_1 = k$, and define $a^1 = (a_1, 0, \dots, 0)$. We then go over $i = 2, \dots, s$, finding at each step the largest number k that satisfies $q(a^{i-1} + k \cdot e_i) \geq p$, and then taking $a_i = k$ and $a^i = (a_1, \dots, a_i, 0, \dots, 0)$ (here e_i is the i^{th} unit vector).

We set our final vector to be $a = a^s$. It follows immediately from the definition of a^s that $q(a) \geq p$ and that $q(a) \cdot (1 - 2^{-s}) < p$, which implies that $q(a) \leq p \cdot (1 - 2^{-s})^{-1} \leq p \cdot (1 + 2 \cdot 2^{-s}) \leq p + 2^{-s} \leq p + \frac{\delta}{n^{10}}$. We therefore have that a satisfies (2). It is also clear from the definition that $a_1 \in \{1, \dots, s\}$. To show that $a_i \in \{0, 1, 2, 3\}$ for all $i > 1$, we first observe that it is clear from the definition of the a_i 's that for all i , $p \leq q(a^i) \leq p \cdot (1 - 2^{-i})^{-1}$. It then follows for each $i > 1$ that

$$q(a^{i-1} + 4 \cdot e_i) \leq p \cdot (1 - 2^{-(i-1)})^{-1} \cdot (1 - 2^{-i})^4 < p,$$

as can be verified by a simple application of the binomial expansion to $(1 - 2^{-i})^4$. The definition of a_i therefore dictates that $a_i < 4$, as desired. $\blacktriangleleft$

Proof of Theorem 1

Notice that if $|p' - p| \leq \frac{\varepsilon}{2n}$, a function $f : \{0, 1\}^n \rightarrow \{0, 1\}$ which is ε far from monotone with respect to μ_p^n is $\frac{\varepsilon}{2}$ far from monotone with respect to the measure $\mu_{p'}^n$. Hence it is enough to apply a monotonicity testing algorithm to f with respect to $\mu_{p'}^n$. We thus use Corollary 12 to find a p' that is sufficiently close to p and that is r -locally embeddable for $r = O(\log^2(n/\varepsilon))$, and then apply the tester from Lemma 6 with respect to the measure $\mu_{p'}^n$ and the error $\varepsilon/2$. To obtain Theorem 1, we independently repeat this tester $\frac{10}{R(rn, \varepsilon/2)} = \tilde{O}(\sqrt{n}/\varepsilon^2)$ times.

A directed isoperimetric inequality over the p -biased hypercube

By Corollary 12 and Lemma 8, we get an analog of Theorem 7 for the p -biased cube, stated below.

► **Theorem 13.** *For all $p \in (0, 1)$, if $f: (\{0, 1\}^n, \mu_p^{\otimes n}) \rightarrow \{0, 1\}$ is ε -far from monotone, then*

$$\mathbb{E}_{x \sim \mu_p^{\otimes n}} \left[\sqrt{s_f^-(x)} \right] \geq \Omega \left(\frac{\varepsilon}{\log(n/\varepsilon)^2} \right).$$

Proof. Let $r \in \mathbb{N}$ and $p' \in (0, 1)$ be from Corollary 12 for $\delta = \varepsilon^3$. Note that the distributions $\mu_p^{\otimes n}$ and $\mu_{p'}^{\otimes n}$ are δ/n^9 close, hence f is at least $\varepsilon/2$ far from monotone over $\mu_{p'}^{\otimes n}$ and

$$\mathbb{E}_{x \sim \mu_p^{\otimes n}} \left[\sqrt{s_f^-(x)} \right] \geq \mathbb{E}_{x \sim \mu_{p'}^{\otimes n}} \left[\sqrt{s_f^-(x)} \right] - \sqrt{n} \frac{\delta}{n^9} \geq \frac{\varepsilon}{\sqrt{r} \log(nr/\varepsilon)} - \frac{\delta}{n^8},$$

where the last inequality is by Lemma 8. The theorem follows as $r = O(\log^2(n/\varepsilon))$. ◀

2.2 Monotone Symmetric Embeddings

A function $T: \{0, 1\}^r \rightarrow [m]$ is monotone and symmetric, if and only if for each $i \in [m]$, $T^{-1}(i)$ contains all elements x with hamming weights in some segment, and the segment that corresponds to i is 'below' that which corresponds to $i + 1$ for each i . Next, we show that if a function $T: \{0, 1\}^r \rightarrow [m]$ is monotone and symmetric, then the distribution $T(U)$ is r -locally embedded. Here by $T(U)$ we mean the distribution over $[m]$ resulting from choosing x uniformly from $\{0, 1\}^r$, and outputting $T(x)$.

► **Lemma 14.** *Suppose $T: \{0, 1\}^r \rightarrow [m]$ is monotone and symmetric. Then the distribution $T(U)$ is r -locally embedded.*

Proof. Denote $\nu = T(U)$. Defining $\phi: \{0, 1\}^r \rightarrow [m]$ by $\phi(x) = T(x)$, it is clear that ϕ is monotone and that the distribution of $\phi(U)$ is the same as ν , and we next discuss the construction of Ψ_ω .

A monotone path in $\{0, 1\}^r$ is a sequence of vertices $v_0 = \vec{0}, v_1, \dots, v_r = \vec{1}$ wherein $v_0 < v_1 < \dots < v_r$ and any two consecutive vertices differ in exactly one coordinate. The probability space (Ω, P) indexes a uniform choice of a monotone path in $\{0, 1\}^r$ and additional auxiliary randomness. One way to generate such path is by choosing a random permutation π in S_r , considering the path going through $\vec{0}, e_1, e_1 + e_2, \dots, e_1 + \dots + e_i, \dots, \vec{1}$, and applying the permutation to re-label the indices $\{1, \dots, r\}$. We remark that taking a random path $\ell = (v_0, \dots, v_r)$, the marginal distribution of v_t is uniform in $\{0, 1\}^r$ among all vertices of Hamming weight t .

To define Ψ_ω , we look at ω which specifies a path $\ell = (v_0, \dots, v_r)$ and additional randomness ω' . We use the additional randomness to generate, for each $a \in [m]$, a Hamming weight t_a according to the distribution of $|z|$ where we sample $z \sim T^{-1}(a)$ uniformly. We then define $\Psi_\omega(a) = v_{t_a}$.

Note that $\phi \circ \Psi_\omega = \text{identity}$, and that for a specific choice of ω , $\Psi_\omega(0) \leq \dots \leq \Psi_\omega(m-1)$ since these are vertices from a monotone path. Finally, fixing a , the distribution of $\Psi_\omega(a)$ over the randomness of ω is v_{t_a} where $t_a = |z|$ and $z \sim T^{-1}(a)$, so v_{t_a} is the t_a vertex on a random monotone path in $\{0, 1\}^r$. In other words, $\Psi_\omega(a)$ is a uniformly chosen vertex from layer t_a , where t_a is distributed as above, hence it is uniform in $T^{-1}(a)$. Hence, the distribution of $\Psi_\omega(a)$ over $\omega \sim P$ and $a \sim \nu$ is uniform over $\{0, 1\}^r$. ◀

Lemma 14 can be used to show that distributions that are close to uniform over $[3]$ can be locally embedded. For example, one can choose two thresholds $t_1 < t_2$ and consider the function $T_{t_1, t_2}: \{0, 1\}^r \rightarrow [3]$ defined as $T(x) = 0$ if $|x| \leq t_1$, $T(x) = 2$ if $|x| \geq t_2$, and otherwise $T(x) = 1$. A straightforward argument shows that for any r , one can choose t_1, t_2 so that the distribution $T_{t_1, t_2}(U)$ is $O(1/\sqrt{r})$ close to uniform over $[3]$. This implies, in particular that as long as $r \geq \frac{n}{\delta^2}$, the distributions $T_{t_1, t_2}(U)^{\otimes n}$ and $U^{\otimes n}$ over $[3]$ are $O(\delta)$ -close to each other,¹ hence for $\delta < \frac{\varepsilon}{1000}$, if $f: ([3]^n, U^{\otimes n}) \rightarrow \{0, 1\}$ is ε -far from monotone, then $f: ([3]^n, T_{t_1, t_2}(U)^{\otimes n}) \rightarrow \{0, 1\}$ is $\varepsilon/2$ -far from monotone, and using Lemma 6 we get a 2-query monotonicity tester with rejection probability at least $R(rn, \varepsilon/2)$. A closer inspection shows that the resulting rejection probability is $\Omega(\varepsilon^2/\sqrt{rn}) = \Omega(\varepsilon^3/n)$ hence worse than known results.

Having said that, the above argument also highlights that if we can design an approximate embedding T such that $T(U)$ is ξ -close to uniform over $[3]$ for $\xi = o(1/\sqrt{r})$, then we will get a non-trivial monotonicity tester over $[3]^n$. Using elementary arguments, one can show that for any r , there is $r' = \Theta(r)$ and thresholds t_1, t_2 such that $T_{t_1, t_2}(U)$ is $O(1/r)$ -close to uniform, which allows one to take $r = \Theta(\sqrt{n}/\delta^2)$ and thus get a tester with rejection probability $\Omega(\varepsilon^3/n^{3/4})$, which already improves upon the state of the art result.

Using threshold as embedding strategy though has its limits. Indeed, it seems that using thresholds we will never be able to get $T(U)$ to be ξ -close to uniform over $[3]$ for $\xi = o(1/r^{3/2})$. For each $r' \in [r, 100r]$ consider the threshold function $T = T_{t_1, t_2}: \{0, 1\}^{r'} \rightarrow \{0, 1, 2\}$ that minimizes the distance between $T(U)$ and U_3 . Heuristically, one can think of this distance as a random number in the interval $[0, \Theta(1)/\sqrt{r}]$, hence we would expect the minimum of these to be of the order $1/r^{3/2}$. Thus, to get near optimal monotonicity testers we have to venture beyond threshold functions. In the next section we facilitate this by formulating embeddings in the language of monotone perfect matchings (or almost perfect matchings), and show that slight perturbations of thresholds can be used for embeddings.

2.3 Embeddings from Monotone Perfect Matchings

In this section, we present a combinatorial method of constructing embeddings using monotone matchings on the hypercube. For simplicity, we tailor our presentation for uniform measures, however one may consider analogs for other distributions.

We will think of the hypercube $G = (\{0, 1\}^r, E)$ as a directed graph, wherein (x, y) is an edge if $x < y$. We may thus view any $\phi: \{0, 1\}^r \rightarrow [m]$ as defining a partitioning of the vertices into sets $A_0, \dots, A_{m-1}$ where $A_i = \{x \mid \phi(x) = i\}$.

► **Definition 15.** For $\delta > 0$, we say a function $\phi: \{0, 1\}^r \rightarrow [m]$ admits a δ -almost perfect matching if there are matchings $E_0, \dots, E_{m-2}$ in G , wherein E_i is a matching between A_i and A_{i+1} , such that for each i , E_i covers all but δ fraction of the vertices of A_i and A_{i+1} .

If ϕ admits a δ -almost perfect matching for $\delta = 0$, we simply say that ϕ admits a perfect matching.

The following lemma asserts that a monotone function ϕ that admits a perfect matching can be used toward constructing an embedding of $([m], U)$.

¹ This can be observed by computing either the KL-divergence or the Hellinger distance between $T_{t_1, t_2}(U)$ and U , which by sub-additivity implies a bound on that measure between $T_{t_1, t_2}(U)^{\otimes n}$ and $U^{\otimes n}$. One may then conclude a bound on the statistical distance between $T_{t_1, t_2}(U)^{\otimes n}$ and $U^{\otimes n}$ by the relation between KL-divergence and statistical distance (via Pinsker's inequality) or by an analogous result for the Hellinger distance.

► **Lemma 16.** *Let $m \in \mathbb{N}$ and let $\phi: \{0,1\}^r \rightarrow [m]$ be a monotone function. If ϕ admits a perfect matching, then $([m], U)$ can be r -locally embedded.*

Proof. Let $E_0, \dots, E_{m-2}$ be monotone matchings in G that cover all vertices for ϕ , and consider the collection $\mathcal{P}$ of vertex disjoint paths of length $m-1$ they form. I.e., starting from a vertex $x \in A_0$ we use the matching edge of x from E_0 to go to a vertex A_1 , then use the edge of E_1 to go upwards and so on, until we reach A_{m-1} . We construct an embedding $(\phi, \Psi = (\Psi_\omega)_{\omega \in \Omega}, \Omega, P)$, where the probability space Ω is $\mathcal{P}$ and the measure P is uniform over Ω . We define $\Psi_\omega(i) = \omega_i$, where ω_i is the vertex from A_i on the path ω .

The monotonicity of ϕ is clear by assumption and the monotonicity of Ψ_ω follows because ω is a monotone path. Finally, it is clear that $\phi \circ \Psi_\omega = \text{identity}$ and that the distribution of $\Psi_\omega(i)$ when choosing $i \in [m]$ uniformly and $\omega \sim P$ is uniform over $\{0,1\}^r$, as $\mathcal{P}$ is a collection of vertex disjoint paths that covers all of $\{0,1\}^r$. ◀

In light of Lemma 16, it makes sense it should be possible to locally embed $([m], U)$ with good parameters. Indeed, for $m = 4$ we found an 9-local embedding of $[4]$ using computer search [1], which immediately gives near optimal monotonicity testers and directed isoperimetric inequalities. For divisibility reasons though, to have a perfect matching m must be a power of 2, however as we show in subsequent sections, this is the only limitation that exists. To address the divisibility issues, we need to state an analog of approximate embeddings and prove analogs of Lemmas 6, 8 and 16.

2.4 Monotonicity Testers and Isoperimetric Inequalities from Almost Perfect Matchings

To circumvent the divisibility issues we consider a more general version of embeddings, which is nevertheless sufficient for the purposes of monotonicity testing as well as for proving isoperimetric inequalities:

► **Definition 17.** *We say that a probability distribution $([m], \mu_1)$ can be r -locally embedded in $(\{0,1\}^r, \mu_2)$ if there are a map $\phi: \{0,1\}^r \rightarrow [m]$, a collection of maps $\Psi = \{\Psi_\omega: [m] \rightarrow \{0,1\}^r\}_{\omega \in \Omega}$ and a probability distribution P over Ω such that:*

1. *Each one of ϕ and Ψ_ω are monotone.*
2. *Sampling $x \sim \mu_2$, the distribution of $\phi(x)$ is μ_1 .*
3. *Sampling $y \sim \mu_1$ and $\omega \sim P$, the distribution of $\Psi_\omega(y)$ is μ_2 .*
4. *For each $\omega \in \Omega$, $\phi \circ \Psi_\omega$ is the identity on $[m]$.*

Definition 17 generalizes Definition 4 in the sense that now we allow the distribution over the hypercube $\{0,1\}^r$ to not be uniform. In all consequent applications of Definition 17 the distribution μ_2 will be very close to uniform, though. We now prove analogs of Lemmas 6, 8 and 16 for relaxed embeddings. We begin by showing that almost perfect matchings imply local embeddings as per Definition 17:

► **Lemma 18.** *Let $m \in \mathbb{N}$ and $\delta > 0$, and let $\phi: \{0,1\}^r \rightarrow [m]$ be a monotone function. If ϕ admits a δ -almost perfect matching, then there are distributions μ_1 over $[m]$ and μ_2 over $\{0,1\}^r$, such that μ_1 is $m\delta$ -close to uniform over $[m]$, μ_2 is $m\delta$ -close to uniform² over $\{0,1\}^r$ and $([m], \mu_1)$ can be r -locally embedded in $(\{0,1\}^r, \mu_2)$.*

² In fact, μ_1 is the uniform distribution over a subset of $\{0,1\}^r$ of fractional size at least $1 - 2m\delta$.

Proof. We repeat the same construction in Lemma 16, except that now the collection $\mathcal{P}$ may include paths of length less than $m - 1$. We take $\mathcal{P}' \subseteq \mathcal{P}$ to be the collection of paths of length $m - 1$. We argue that $\mathcal{P}'$ covers at least $1 - m\delta$ fraction of vertices of G . To see that, note that each path in $\mathcal{P}$ whose length is shorter than $m - 1$ can be uniquely associated with some $i = 0, \dots, m - 2$ and a vertex x either from A_i or A_{i+1} that is not matched in E_i . Thus, the number of paths in $\mathcal{P}$ shorter than $m - 1$ is at most the total number of (i, x) such that $x \in A_i$ is unmatched in E_i plus the number of (i, x) such that $x \in A_{i+1}$ is unmatched in E_i , which is at most 2δ fraction of $\{0, 1\}^r$. Since each such path contains at most $m - 1$ vertices, it follows that $\mathcal{P}'$ covers all but $1 - 2(m - 1)\delta$ fraction of $\{0, 1\}^r$.

With this in mind, we define the distribution μ_2 over $\{0, 1\}^r$ by picking $\ell \in \mathcal{P}'$ uniformly, $j \in [m]$ uniformly and outputting the vertex at the j^{th} spot of the path ℓ , i.e. ℓ_j . The distribution μ_1 over $[m]$ is defined by sampling $x \sim \mu_2$ and outputting $\phi(x)$. We also define (Ω, P) by taking $\Omega = \mathcal{P}'$ and P to be the uniform distribution over Ω , and take as before $\Psi = (\Psi_\omega)_{\omega \in \Omega}$ defined as $\Psi(j) = \omega_j$.

By definition, the distribution over $\Psi_\omega(j)$ where $j \sim \mu_1$ and $\omega \sim \mathcal{P}'$ is μ_1 , and the distribution of $\phi(x)$ where $x \sim \mu_2$ is μ_1 . The monotonicity of ϕ, Ψ_ω is clear as before, as well as the fact that $\phi \circ \Psi_\omega = \text{identity}$. $\blacktriangleleft$

We now turn to the analog of Lemmas 6, 8.

► **Lemma 19.** *There is an absolute constant $c > 0$ such that the following holds. Let $r, m, n \in \mathbb{N}$, $\varepsilon, \delta > 0$ and suppose that $0 < \delta < \frac{c\varepsilon}{mr^2n^2}$. If there is a monotone function $\phi: \{0, 1\}^r \rightarrow [m]$ that admits a δ -almost perfect matching, then:*

1. *there is a 2-query monotonicity testing algorithm for functions over $([m]^n, U^{\otimes n})$ that always accepts monotone functions, and rejects functions that are ε -far from monotone with probability at least $R(rn, \varepsilon/4)$ (recall (1) for the definition of R).*
2. *If $f: [m]^n \rightarrow \{0, 1\}$ is ε -far from monotone with respect to $U^{\otimes n}$, then*

$$\mathbb{E}_{x \in [m]^n} \left[\sqrt{s_f(x)} \right] \geq \Omega \left(\frac{\varepsilon}{\sqrt{r} \log(rn/\varepsilon)} \right).$$

Proof. Let μ_1 and μ_2 be the distributions from Lemma 18 from ϕ , and let $(\phi, (\Psi_\omega)_{\omega \in \Omega}, P)$ be an r -local embedding of $([m], \mu_1)$ in $(\{0, 1\}^r, \mu_2)$. Given $f: ([m]^n, U^{\otimes n}) \rightarrow \{0, 1\}$, define $g: \{0, 1\}^{rn} \rightarrow \{0, 1\}$ by

$$g(x(1), \dots, x(n)) = f(\phi(x(1)), \dots, \phi(x(n))).$$

To prove the first item, run the monotonicity tester of the hypercube on g , and accept/reject accordingly. Note that a single query to g can be answered by making a single query to f , and that if f is monotone then g is monotone, hence the tester always accepts in this case. If $\varepsilon(f; U^{\otimes n}) \geq \varepsilon$, then $\varepsilon(f; \mu_1^{\otimes n}) \geq \varepsilon(f) - \Delta(\mu_1^{\otimes n}, U^{\otimes n}) \geq \varepsilon - 2mn\delta > \varepsilon/2$. By the same argument as in Lemma 5, it follows that $g: (\{0, 1\}^{rn}, \mu_2^{\otimes n}) \rightarrow \{0, 1\}$ is $\varepsilon/2$ -far from monotone, and so $\varepsilon(g; U^{rn}) \geq \varepsilon(g; \mu_2^{\otimes n}) - \Delta(U^{rn}, \mu_2^{\otimes n}) \geq \varepsilon/2 - 2mrn\delta \geq \varepsilon/4$, hence the tester rejects with probability at least $R(rn, \varepsilon/4)$.

To prove the second item, we note that

$$\begin{aligned}
\mathbb{E}_{x \in [m]^n} \left[\sqrt{s_f^-(x)} \right] &\geq \mathbb{E}_{x \sim \mu_1^{\otimes n}} \left[\sqrt{s_f^-(x)} \right] - \sqrt{n} \Delta(\mu_1^{\otimes n}, U^{\otimes n}) \\
&\geq \mathbb{E}_{y \sim \mu_2^{\otimes n}} \left[\sqrt{\frac{s_g^-(y)}{r}} \right] - \sqrt{n} \Delta(\mu_1^{\otimes n}, U^{\otimes n}) \\
&= \frac{1}{\sqrt{r}} \mathbb{E}_{y \sim \mu_2^{\otimes n}} \left[\sqrt{s_g^-(y)} \right] - \sqrt{n} \Delta(\mu_1^{\otimes n}, U^{\otimes n}) \\
&\geq \frac{1}{\sqrt{r}} \mathbb{E}_{y \in \{0,1\}^{rn}} \left[\sqrt{s_g^-(y)} \right] - \sqrt{rn} \Delta(\mu_2^{\otimes n}, U^{\otimes rn}) - \sqrt{n} \Delta(\mu_1^{\otimes n}, U^{\otimes n}) \\
&\geq \frac{c}{\sqrt{r}} \frac{\varepsilon}{4 \log(rn)} - \sqrt{rn} \Delta(\mu_2^{\otimes n}, U^{\otimes rn}) - \sqrt{n} \Delta(\mu_1^{\otimes n}, U^{\otimes n}),
\end{aligned}$$

where $c > 0$ is an absolute constant; in the last transition, we used Theorem 7. Bounding $\Delta(\mu_2^{\otimes n}, U^{\otimes rn}) \leq 2rmn\delta$ and $\Delta(\mu_1^{\otimes n}, U^{\otimes n}) \leq 2mn\delta$ and using the upper bound on δ shows that the second and third terms are negligible compared to the first, hence we get that $\mathbb{E}_{x \in [m]^n} \left[\sqrt{s_f^-(x)} \right] \geq \frac{c\varepsilon}{8\sqrt{r} \log(rn)}$ as required. $\blacktriangleleft$

With these lemmas in hand, to prove Theorems 2, 3 it now suffices to construct good enough almost perfect matchings for some monotone function $\phi: \{0, 1\}^r \rightarrow [m]$. The following result asserts that such almost perfect monotone matchings exists:

► **Theorem 20.** *There is an absolute constant $C > 0$ such that for all $m \in \mathbb{N}$, for any $r \geq C \cdot m^6$ there is a monotone function $\phi: \{0, 1\}^r \rightarrow [m]$ that admits a δ -almost perfect matching for $\delta \leq m2^{-r}$.*

The proof of Theorem 20 is deferred to Section 3. Before embarking on this proof, we quickly show how it implies several results stated in the introduction.

► **Lemma 21.** *Theorem 20 implies Theorems 2, 3.*

Proof. Take $r = Cm^6 \log(mn/\varepsilon)$ for sufficiently large absolute constant $C > 0$. By Theorem 20 we get that there is $\phi: \{0, 1\}^r \rightarrow [m]$ that admits a δ -almost perfect matching for $\delta \leq m2^{-r}$, and the result is concluded by appealing to Lemma 19. $\blacktriangleleft$

3 Constructing Efficient Monotone Matchings on the Hypercube

3.1 Theorem 20: Proof Overview

We start from a threshold embedding as in Section 2.2, that is $T = T_{t_1, \dots, t_{m-1}}: \{0, 1\}^r \rightarrow [m]$ defined as $T(x) = i$ if $t_i \leq |x| < t_{i+1}$. Using it, we can make sure that the partition it defines, $A_i = \{x \mid T(x) = i\}$ is δ -almost perfect matching for $\delta = O(1/\sqrt{r})$. The reason for this δ is that A_i 's may have sizes which differ by $2^r \delta$, as this is the number of points in each slice. Therefore, to improve upon this construction a natural idea is shift elements around by adding to some A_i 's elements either from the bottom level of A_{i+1} or from top level of A_{i-1} , so that eventually the sizes of all A_i 's are equal up to 1. We do not know though how to carry out this adjustment so that the embedding construction from Section 2.2 still works. Instead, we vary the sets A_i in a randomized way, and show that with high probability there is an almost perfect monotone matching between each A_i and A_{i+1} for all i 's.

In more details, consider a random ordering π of $\{0, 1\}^n$ which starts with some ordering of $\{0, 1\}^n$ according to Hamming weight (that is, the vertices of Hamming weight i appear in a chunk before the vertices of Hamming weight $i + 1$, for all i), and within each Hamming weight chunk applies a random ordering. We think of π as $\pi: \{0, 1\}^n \rightarrow [2^n]$, wherein $\pi^{-1}(i)$ is the i th point in the order. We then define, for each $i = 1, \dots, m$, the set A_i as the chunk of $\lfloor \frac{2^n}{m} \rfloor$ next elements in π , namely

$$P_i = \pi^{-1} \left(\left\{ \left\lfloor \frac{2^n}{m} \right\rfloor (i-1) + j \mid j = 1, \dots, \left\lfloor \frac{2^n}{m} \right\rfloor \right\} \right),$$

and show that, with high probability, there is a monotone matching between each P_i and P_{i+1} .

To show that, we first develop a bit of machinery. First, we generalize the notion of perfect matching to that of fractional perfect matching (see Definition 26): a fractional perfect matching can be defined between two sets of equal size, in which case it is promised that it can be replaced by a true matching. But it can also be defined over two sets that are each endowed with a arbitrary measure, as long as the total measure of each set is the same. Another important property is that the existence of a fractional perfect matching is transitive, namely if there is a perfect matching between μ and ν , and also between ν and τ (where μ , ν and τ are sets endowed with measures), then there exists a perfect matching between μ and τ .

Then, we view P_i as a collection of t slices, $s_1(i), \dots, s_t(i)$ and two random subsets $S_0(i)$ and $S_{t+1}(i)$ of the slices $s_0(i)$ and $s_{t+1}(i)$. We show that there is a perfect fractional matching between the vertices of $\{x \mid x \in S_0(i) \text{ or } x \text{ has Hamming weight } s_1(i), \dots, s_{t/2}(i)\}$, and vertices of Hamming weight $s_t(i)$ (when weighted appropriately). In words, this says that we can find a fractional matching between a union of layers with a random subset of another subsequent layer, and a layer that is a bit above them. Using the same arguments, we prove that there is a perfect fractional matching between the vertices of $\{x \mid x \in S_{t+1}(i) \text{ or } x \text{ has Hamming weight } s_{t/2+1}(i), \dots, s_t(i)\}$ and Hamming weight $s_{t/4}(i+1)$ vertices. Thus, in effect we are reduced to matching complete slices again; indeed, to show the matching between P_i and P_{i+1} we break them into “lower half” and “upper half” and use the above statements to find matchings of these with slices a bit above them and a bit below them. Using transitivity now and the fact there are perfect fractional matchings between $s_{t/2}(i)$ and $s_0(i+1)$ (which exists as we make sure that $s_{t/2}(i) < s_0(i+1)$), and $s_{t/4}(i+1)$ and $s_{t/2}(i+1)$ (which again exists as we make sure that $s_{t/4}(i+1) < s_{t/2}(i+1)$), one can then stitch these matchings to get a perfect fractional matching between P_i and P_{i+1} , and thus conclude the existence of perfect matching.

The proof of statements such that “there is a perfect fractional matching between $S_0(i)$ and the slices $s_1(i), \dots, s_{t/2}(i)$ and $s_{t/2}(i)$ (when weighted appropriately)” consists the bulk of the work, and to do that we show that with high probability Hall’s condition holds. To do that, we use the notion of upper shadows (which, roughly speaking, counts the number of neighbours a set of vertices S has in the directed hypercube graph) as well as the Kruskal-Katona theorem which gives us a lower bound on it. We show that only sets of vertices T which have very smaller upper shadow can violate Hall’s condition, and for them we show by a careful application of Chernoff’s bound that, with high probability, Hall’s condition still holds. The main difficulty in the last step is that the number of such sets T is quite large, however we show that these sets admit an efficient “ ε -net” type approximations. This reduces the number of sets T we need to union bound over enough so that Chernoff’s bound works.

3.2 Shadows, Kruskal-Katona and Approximating Collections with Small Shadow

3.2.1 The Kruskal-Katona Theorem

Throughout this section, we consider slices of the Boolean hypercube,

$$\binom{[n]}{k} = \{x \in \{0, 1\}^n \mid |x| = k\},$$

and denote by μ_k the uniform measure on $\binom{[n]}{k}$. Our proof uses the Kruskal-Katona Theorem [14, 16, 6], which we present next. We will use a more convenient form of it as stated in [18, Section 1.2].

► **Definition 22.** For a collection $\mathcal{A} \subseteq \binom{[n]}{k}$, define the upper shadow $\partial^u \mathcal{A}$ and lower shadow $\partial^d \mathcal{A}$ of $\mathcal{A}$ as

$$\partial^u \mathcal{A} = \left\{ y \in \binom{[n]}{k+1} \mid \exists x \in \mathcal{A}, x < y \right\}, \quad \partial^d \mathcal{A} = \left\{ y \in \binom{[n]}{k-1} \mid \exists x \in \mathcal{A}, y < x \right\}.$$

The Kruskal-Katona Theorem states:

► **Lemma 23.** For all $\mathcal{A} \subseteq \binom{[n]}{k}$ we have that

$$\mu_{k+1}(\partial^u \mathcal{A}) \geq \mu_k(\mathcal{A})^{1-\frac{1}{n}}, \quad \mu_{k-1}(\partial^d \mathcal{A}) \geq \mu_k(\mathcal{A})^{1-\frac{1}{n}}.$$

In words, Lemma 23 asserts that if $\mathcal{A}$ is a small sub-set of a slice, then the upper shadow (as well as the lower shadow) have considerably larger densities. Typically, we will apply the upper shadow/ lower shadow operators more than once; given $\mathcal{A} \subseteq \binom{[n]}{k}$, we will look at $\mu_{k+t}(\partial^u \dots \partial^u \mathcal{A})$ where we applied the upper shadow operator t -times. To simplify notations, we denote this by $\mu_{k+t}(\partial^{t \cdot u} \mathcal{A})$,

3.2.2 Approximating a Collection with a Small Shadow

In general, the conclusion of Lemma 23 is tight, as can be evidenced by collections of the type

$$\mathcal{A} = \left\{ x \in \binom{[n]}{k} \mid x_1 = \dots = x_\ell = 1 \right\}.$$

Intuitively, the reason that $\mathcal{A}$ above is tight for Kruskal-Katona is that for any element in $y \in \binom{[n]}{k+1}$, we either have that almost all of the $x < y$ of Hamming weight k are in $\mathcal{A}$ – in which case $y \in \partial^u \mathcal{A}$, or else none of these x 's are in $\mathcal{A}$. Hence, many of the x 's “vote” for the same set of y 's to be included in the upper shadow, leading to only a moderate increase in density. We show that in general, collections $\mathcal{A}$ with small shadow exhibit such behaviour, and use it to show that this collection of families admits a small ε -net:

► **Lemma 24.** Let $s, t, n \in \mathbb{N}$ such that $0 \leq t \leq n - s$, and let $0 < \varepsilon \leq \frac{1}{100}$. If $\mathcal{A} \subseteq \binom{[n]}{s}$ satisfies $\mu_{s+t}(\partial^{t \cdot u} \mathcal{A}) \leq (1 + \varepsilon)\mu_s(\mathcal{A})$, then there is a collection $\mathcal{M} \subseteq \mathcal{A}$ and $\mathcal{B}_{\mathcal{M}} \subseteq \binom{[n]}{s}$, $\mathcal{B}'_{\mathcal{M}} \subseteq \binom{[n]}{s+t}$ (defined only by $\mathcal{M}$) such that

1. $|\mathcal{M}| \leq 100 \frac{\ln(1/\varepsilon)}{\binom{s+t}{t}} \cdot |\mathcal{A}|$.
2. $\mathcal{B}'_{\mathcal{M}} = \partial^{t \cdot u} \mathcal{M}$ and $\mu_{s+t}(\mathcal{B}'_{\mathcal{M}} \Delta \partial^{t \cdot u} \mathcal{A}) \leq 6\varepsilon \cdot \mu_{s+t}(\partial^{t \cdot u} \mathcal{A})$.
3. $\mathcal{B}_{\mathcal{M}} = \left\{ x \in \binom{[n]}{s} \mid \Pr_{y > x, |y|=s+t} [y \in \mathcal{B}'_{\mathcal{M}}] \geq \frac{1}{2} \right\}$ and $\mu_s(\mathcal{B}_{\mathcal{M}} \Delta \mathcal{A}) \leq 18\varepsilon \mu_s(\mathcal{A})$.

25:14 Improved Monotonicity Testers via Hypercube Embeddings

Proof. We show that taking $\mathcal{M} \subseteq \mathcal{A}$ randomly of size $M = 100 |\mathcal{A}| \frac{\ln(1/\varepsilon)}{\binom{s+t}{t}}$, the collections $\mathcal{B}_{\mathcal{M}}$ and $\mathcal{B}'_{\mathcal{M}}$ as defined in the statement work with positive probability.

Consider the bi-partite graph $G = (V \cup U, E)$ where the sides are $V = \mathcal{A}$ and $U = \partial^{t,u} \mathcal{A}$, and $(x, y) \in E$ is an edge if $x \in \mathcal{A}$, $y \in \partial^{t,u} \mathcal{A}$ and $x < y$. Then G is left-regular with degree $h_L = \binom{n-s}{t}$, and so

$$|E| = |V| \cdot \binom{n-s}{t} = \mu_s(\mathcal{A}) \binom{n-s}{t} \binom{n}{s} = \mu_s(\mathcal{A}) \frac{n!}{s! \cdot t! \cdot (n-s-t)!}.$$

As for the right side, the degree of each vertex is at most $h_R = \binom{s+t}{t}$ and the average degree of a vertex is

$$\frac{|E|}{|U|} = \frac{\mu_s(\mathcal{A}) \frac{n!}{s! \cdot t! \cdot (n-s-t)!}}{\frac{n!}{s! \cdot t! \cdot (n-s-t)!}} = \frac{\mu_s(\mathcal{A}) \binom{n}{s+t} h_R}{|U|} = h \frac{\mu_s(\mathcal{A})}{\mu_{s+t}(\mathcal{A})} \geq \frac{h_R}{1+\varepsilon}.$$

Thus, choosing $y \in U$ uniformly, the expected value of $h_R - d(y)$ is at most εh_R , and by Markov's inequality it follows that $h_R - d(y) \leq h_R/2$ except with probability 2ε . Thus, denoting by δ the fraction of $y \in U$ such that $d(y) < h_R/2$, we get that $\delta \leq 2\varepsilon$. Thus,

$$\mathbb{E}_{\mathcal{M}} [\mu_{s+t}(\mathcal{B}'_{\mathcal{M}} \Delta \partial^{t,u} \mathcal{A})] \leq \delta \mu_{s+t}(\partial^{t,u} \mathcal{A}) + (1-\delta) \mu_{s+t}(\partial^{t,u} \mathcal{A}) \left(1 - \frac{h_R/2}{|U|}\right)^M \leq 3\varepsilon \cdot \mu_{s+t}(\partial^{t,u} \mathcal{A}),$$

where in the last inequality we used the fact that $\left(1 - \frac{h_R/2}{|U|}\right)^M \leq e^{-\frac{M h_R}{2|U|}} \leq e^{-50 \ln(1/\varepsilon)} \leq \varepsilon$.

The third item follows similarly, and we first upper bound $\mathbb{E}_{\mathcal{M}} [\mu_s(\mathcal{A} \setminus \mathcal{B}_{\mathcal{M}})]$. For each $x \in \mathcal{A} \setminus \mathcal{B}_{\mathcal{M}}$ we have that x has at most $h_L/2$ of its neighbours in $\mathcal{B}'_{\mathcal{M}}$, hence at least $h_L/2$ of its neighbours in $U \setminus \mathcal{B}'_{\mathcal{M}}$. It follows that

$$\begin{aligned} \mathbb{E}_{\mathcal{M}} [\mu_s(\mathcal{A} \setminus \mathcal{B}_{\mathcal{M}})] &\leq \frac{1}{\binom{n}{s}} \mathbb{E}_{\mathcal{M}} \left[\frac{|U \setminus \mathcal{B}'_{\mathcal{M}}| h_R}{h_L/2} \right] = \frac{2h_R \binom{n}{s+t}}{h_L \binom{n}{s}} \mathbb{E}_{\mathcal{M}} [\mu_{s+t}(\mathcal{B}'_{\mathcal{M}} \Delta \partial^{t,u} \mathcal{A})] \\ &\leq 6\varepsilon \cdot \mu_{s+t}(\partial^{t,u} \mathcal{A}), \end{aligned}$$

which is at most $7\varepsilon \mu_s(\mathcal{A})$. To upper bound $\mathbb{E}_{\mathcal{M}} [\mu_s(\mathcal{B}_{\mathcal{M}} \setminus \mathcal{A})]$, we note that any $x \in \mathcal{B}_{\mathcal{M}} \setminus \mathcal{A}$ has at least $h_L/2$ of the y of Hamming weight $s+t$ for which $x < y$ in $\mathcal{B}'_{\mathcal{M}}$, and in particular in U . The total number of pairs (x, y) such that $x < y$ and $x \notin \mathcal{A}$, $y \in U$ is at most $h_R |U| - |E|$ (as these are all non-edges in G), so we get that

$$\mathbb{E}_{\mathcal{M}} [\mu_s(\mathcal{B}_{\mathcal{M}} \setminus \mathcal{A})] \leq \frac{1}{\binom{n}{s}} \frac{h_R |U| - |E|}{h_L/2} \leq \frac{1}{\binom{n}{s}} \frac{2h_R |U| \varepsilon}{h_L (1+\varepsilon)} \leq 2\varepsilon \frac{\mu_{s+t}(U)}{1+\varepsilon} \leq 2\varepsilon \mu_s(\mathcal{A}).$$

In conclusion, we get that

$$\mathbb{E}_{\mathcal{M}} [\mu_s(\mathcal{B}_{\mathcal{M}} \Delta \mathcal{A})] \leq 9\varepsilon \mu_s(\mathcal{A}), \quad \mathbb{E}_{\mathcal{M}} [\mu_{s+t}(\mathcal{B}'_{\mathcal{M}} \Delta \partial^{t,u} \mathcal{A})] \leq 3\varepsilon \mu_{s+t}(\partial^{t,u} \mathcal{A}),$$

so by Markov's inequality there is a choice for $\mathcal{M}$ satisfying the conclusion of the claim. $\blacktriangleleft$

For future reference, we state a version of Lemma 24 for the operator $\partial^{t,d}$:

► **Lemma 25.** *Let $s, t, n \in \mathbb{N}$ such that $0 \leq t \leq n-s$, and let $0 < \varepsilon \leq \frac{1}{100}$. If $\mathcal{A} \subseteq \binom{[n]}{s}$ satisfies $\mu_{s-t}(\partial^{t,d} \mathcal{A}) \leq (1+\varepsilon) \mu_s(\mathcal{A})$, then there is a collection $\mathcal{M} \subseteq \mathcal{A}$ and $\mathcal{B}'_{\mathcal{M}} \subseteq \binom{[n]}{s-t}$, $\mathcal{B}_{\mathcal{M}} \subseteq \binom{[n]}{s}$ such that*

1. $|\mathcal{M}| \leq 100 \frac{\ln(1/\varepsilon)}{\binom{s}{s-t}} \cdot |\mathcal{A}|$.
2. $\mathcal{B}'_{\mathcal{M}} = \partial^{t,d} \mathcal{M}$ and $\mu_{s-t}(\mathcal{B}'_{\mathcal{M}} \Delta \partial^{t,d} \mathcal{A}) \leq 6\varepsilon \cdot \mu_{s-t}(\partial^{t,d} \mathcal{A})$.
3. $\mathcal{B}_{\mathcal{M}} = \left\{ x \in \binom{[n]}{s} \mid \Pr_{y < x, |y|=s-t} [y \in \mathcal{B}'_{\mathcal{M}}] \geq \frac{1}{2} \right\}$ and $\mu_s(\mathcal{B}_{\mathcal{M}} \Delta \mathcal{A}) \leq 18\varepsilon \mu_s(\mathcal{A})$.

Proof. The proof is essentially the same as the proof of Lemma 24 and we omit the details. $\blacktriangleleft$

3.3 Fractional Monotone Matchings

We now formally define the concept of a monotone fractional matching, which is central to the proof of Theorem 20:

► **Definition 26.** Let $w_U: \{0, 1\}^n \rightarrow [0, \infty)$ and $w_V: \{0, 1\}^n \rightarrow [0, \infty)$ be weight functions such that $\sum_{x \in \{0, 1\}^n} w_U(x) = \sum_{x \in \{0, 1\}^n} w_V(x)$. We say there is a monotone fractional matching from w_U to w_V , and denote $w_U \lesssim w_V$ if, letting U be the support of w_U and V be the support of w_V , there is a weight function $w: U \times V \rightarrow [0, \infty)$ such that $w(u, v) > 0$ only when $u \leq v$, and for every $x \in U$ $y \in V$ it holds that

$$\sum_{z \in V} w(x, z) = w_U(x), \quad \sum_{z \in U} w(z, y) = w_V(y).$$

In this section, we establish several basic properties of fractional monotone matchings. The first of which is a fractional version of Hall's Theorem for monotone matchings. For completeness, we include the (straight-forward) deduction of it from the usual formulation of Hall's Theorem.

► **Lemma 27.** Suppose that w_U and w_V are as in Definition 26, let U and V their supports, and suppose that for all $S \subseteq U$, defining $N(S) = \{v \in V \mid \exists u \in S, u \leq v\}$ we have that

$$\sum_{v \in N(S)} w_V(v) \geq \sum_{u \in S} w_U(u).$$

Then there is a monotone fractional matching between w_U and w_V .

Proof. By approximation, it suffices to show that statement for weight functions w_V and w_U that assign rational values. Let M be a number such that all values of Mw_V and Mw_U are integers, and define the bi-partite graph G whose sides are U' and V' , where each $u \in U$ has $Mw_U(u)$ copies in U' and $v \in V$ has $Mw_V(v)$ copies in V' . We connect (u', v') by an edge if they are copies of $u \in U$, $v \in V$ respectively where $u \leq v$. Our assumption then implies that G satisfies Hall's condition, so we may find a perfect matching $M \subseteq U' \times V'$. Define

$$w(u, v) = \sum_{\substack{u' \text{ copy of } u \\ v' \text{ copy of } v}} 1_{(u', v') \in M},$$

and note that then w forms a fractional monotone matching showing $w_U \lesssim w_V$. ◀

Secondly, we have the following basic properties of $\lesssim$:

► **Lemma 28.** Suppose that $w_U, w_V, w_R, w_L: \{0, 1\}^n \rightarrow [0, \infty)$ are weight functions.

1. *Transitivity:* if $w_U \lesssim w_V$ and $w_V \lesssim w_R$, then $w_U \lesssim w_R$.
2. *Linearity:* if $w_U \lesssim w_R$ and $w_V \lesssim w_L$, then for all $p, q \geq 0$, $pw_U + qw_V \lesssim pw_R + qw_L$.

Proof. For the first item, let U, V, R be the supports of w_U, w_V and w_R respectively, and let $w_1: U \times V \rightarrow [0, \infty)$ and $w_2: V \times R \rightarrow [0, \infty)$ be the weight functions demonstrating that $w_U \lesssim w_V$ and $w_V \lesssim w_R$, respectively. Define $w: U \times R \rightarrow [0, \infty)$ by

$$w(u, r) = \sum_{v \in V} \frac{1}{w_V(v)} w_1(u, v) w_2(v, r).$$

25:16 Improved Monotonicity Testers via Hypercube Embeddings

First, if $w(u, r) > 0$ then there is $v \in V$ such that $w_1(u, v), w_2(v, r) > 0$ and so $u \leq v \leq r$, hence $u \leq r$. Secondly, note that for all u ,

$$\sum_{r \in R} w(u, r) = \sum_{v \in V} \frac{1}{w_V(v)} w_1(u, v) \sum_{r \in R} w_2(v, r) = \sum_{v \in V} \frac{1}{w_V(v)} w_1(u, v) w_V(v) = \sum_{v \in V} w_1(u, v)$$

which is equal to $w_U(u)$. Similarly, for all $r \in R$ we have $\sum_{u \in U} w(u, r) = w_R(r)$. It follows that w is a monotone matching between showing that $w_U \lesssim w_R$.

For the second item, let U, R, V, L be the supports of w_U, w_R, w_V and w_L respectively and let $w_1: U \times R \rightarrow [0, \infty)$ and $w_2: V \times L \rightarrow [0, \infty)$ be weight functions demonstrating that $w_U \lesssim w_R$ and $w_V \lesssim w_L$. Then $w(x, y) = pw_1(x, y) + qw_2(x, y)$ is a weight function showing that $pw_U + qw_V \lesssim pw_R + qw_L$. ◀

Third, we show that if $k \leq k'$, then $\mu_k \lesssim \mu_{k'}$.

► **Lemma 29.** *If $k \leq k'$, then $\mu_k \lesssim \mu_{k'}$.*

Proof. Let $P = (v_1, \dots, v_n)$ be a uniformly chosen monotone path in $\{0, 1\}^n$ starting at $(0, \dots, 0)$ and ending at $(1, \dots, 1)$, and define $w(x, y)$ to be the probability that $v_k = x$ and $v_{k'} = y$. Then it is easily seen that $w(x, y) > 0$ only if $x < y$, and also for every x of Hamming weight k , $\sum_y w(x, y)$ is equal to the probability a uniformly chosen vertex of Hamming weight k is equal to x , hence is $\mu_k(x)$. Similarly, $\sum_x w(x, y) = \mu_{k'}(y)$. ◀

The last statement is a standard connection between fractional matchings and perfect matchings.

► **Lemma 30.** *Suppose that μ, μ' are distributions which are uniform over $A, A' \subseteq \{0, 1\}^n$ respectively, where $|A| = |A'|$. If $\mu \lesssim \mu'$, then there is a monotone perfect matching between A and A' .*

Proof. Consider the bipartite graph $G = (A \cup A', E)$ whose sides are A and A' and the edge set is $E = \{(a, a') \mid a \in A, a' \in A', a \leq a'\}$. As $\mu \lesssim \mu'$, we get that there is $w: A \times A' \rightarrow [0, \infty)$ supported only on E satisfying the properties of a monotone fractional matching. Define $w' = |A|w$, and note that for all $a \in A$ we have that $\sum_{a' \in A'} w'(a, a') = 1$ and also $\sum_{a \in A} w'(a, a') = 1$ for all $a' \in A'$. Thus, the fractional matching number of G is at least

$$\sum_{a \in A, a' \in A'} w'(a, a') = |A|.$$

We argue that the smallest vertex cover in G has size $|A|$. Indeed, if $W \subseteq A \cup A'$ is a vertex cover then

$$|A| = \sum_{e \in E} w'(e) \leq \sum_{z \in W} \sum_{e \ni z} w'(e) = \sum_{z \in W} 1 = |W|.$$

It now follows from König's theorem that G has a perfect matching, and we are done. ◀

3.4 Monotone Matchings on Random Subsets of the Slice

The next lemma is the heart of the proof that our construction admits a good monotone almost perfect matching. For a collection $\mathcal{S} \subseteq \binom{[n]}{k}$, we denote $\mu_{\mathcal{S}}(x) = \mu_k(x)1_{x \in \mathcal{S}}$.

► **Lemma 31.** *For all $C > 0$ there is $n_0 \in \mathbb{N}$ such that the following holds. Let $n, k, s, t \in \mathbb{N}$ and assume that $\frac{n}{2} - C\sqrt{n \log n} \leq k \leq \frac{n}{2} + C\sqrt{n \log n}$ and $10n^{1/3} \leq t \leq C\sqrt{n \log n}$. Then for every $0 \leq s \leq \binom{n}{k}$, setting $\rho = \frac{s}{\binom{n}{k}}$ we have*

$$\Pr_{\substack{\mathcal{S} \subseteq \binom{[n]}{k} \\ |\mathcal{S}|=s}} [(t + \rho)\mu_{k-t} \lesssim t\mu_k + \mu_{\mathcal{S}}] \geq 1 - 2^{-\Omega(2^{n/2})}.$$

Proof. We will use Lemma 27. Denoting $\mu = (t + \rho)\mu_{k-t}$ and $\nu = \nu_{\mathcal{S}} = t\mu_k + \mu_{\mathcal{S}}$, our goal is to show that with high probability over the choice of $\mathcal{S}$, for all $\mathcal{T} \subseteq \binom{[n]}{k-t}$ it holds that $\nu(\partial^{t,u}\mathcal{T}) \geq \mu(\mathcal{T})$. Equivalently, we will upper bound the probability that there is $\mathcal{T}$ that violates it, and we present two arguments depending on the fractional size of $\mathcal{T}$.

The case that $\mu_{k-t}(\mathcal{T}) \leq 1/2$. Let $\mathcal{T}$ be such that $\mu_{k-t}(\mathcal{T}) \leq 1/2$, and suppose that $\nu(\partial^{t,u}\mathcal{T}) < \mu(\mathcal{T})$. We denote by N the size of $\mathcal{T}$. Then we have that

$$\mu_k(\partial^{t,u}\mathcal{T}) \leq \frac{1}{t}\nu(\partial^{t,u}\mathcal{T}) \leq \frac{1}{t}\mu(\mathcal{T}) \leq \frac{t+1}{t}\mu_{k-t}(\mathcal{T}), \quad (3)$$

On the other hand, using Lemma 23 we can deduce a lower bound on the measure of the upper shadow of $\mathcal{T}$, namely that

$$\mu_k(\partial^{t,u}\mathcal{T}) \geq \mu_{k-t}(\mathcal{T})^{(1-\frac{1}{n})^t}. \quad (4)$$

First, this implies a lower bound on the measure of $\mathcal{T}$, as we get that $\frac{t+1}{t}\mu_{k-t}(\mathcal{T}) \geq \mu_{k-t}(\mathcal{T})^{(1-\frac{1}{n})^t}$, and standard manipulations now imply that $\mu_{k-t}(\mathcal{T}) \geq (1 - 1/(t+1))^{n/t}$ and so $\mu_{k-t}(\mathcal{T}) \geq e^{-O(n/t^2)}$, which implies in particular that $N \geq e^{-O(n/t^2)}\binom{n}{k-t} \geq 2^{0.8n}$. Secondly, from (4) and the fact that $\mu_{k-t}(\mathcal{T}) \leq 1/2$ we also get that

$$\mu_k(\partial^{t,u}\mathcal{T}) \geq \mu_{k-t}(\mathcal{T})(1/2)^{(1-\frac{1}{n})^t-1} \geq \mu_{k-t}(\mathcal{T})2^{t/n}. \quad (5)$$

Combining our assumption on $\mathcal{T}$ and (5) yields

$$\begin{aligned} 0 \leq \mu(\mathcal{T}) - \nu(\partial^{t,u}\mathcal{T}) &= (t + \rho)\mu_{k-t}(\mathcal{T}) - t\mu_k(\partial^{t,u}\mathcal{T}) - \mu_k(\partial^{t,u}\mathcal{T} \cap \mathcal{S}) \\ &\leq (t + \rho)\mu_{k-t}(\mathcal{T}) - t\mu_{k-t}(\mathcal{T})2^{t/n} - \mu_k(\partial^{t,u}\mathcal{T} \cap \mathcal{S}) \\ &\leq \left(\rho - \frac{t^2}{2n}\right)\mu_{k-t}(\mathcal{T}) - \mu_k(\partial^{t,u}\mathcal{T} \cap \mathcal{S}), \end{aligned}$$

and in particular we get that

$$\mu_k(\partial^{t,u}\mathcal{T} \cap \mathcal{S}) \leq \left(\rho - \frac{t^2}{2n}\right)\mu_{k-t}(\mathcal{T}) \leq \left(\rho - \frac{t^2}{2n}\right)\mu_k(\partial^{t,u}\mathcal{T}). \quad (6)$$

Noting that the expectation of the left hand side, over the choice of $\mathcal{S}$, is $\rho\mu_k(\partial^{t,u}\mathcal{T})$, this inequality suggests that the probability for this for a specific $\mathcal{T}$ is small. A naive application of Chernoff's bound is not good enough since we would need to union bound over too many choices for $\mathcal{T}$. To cut down on the number of events we union bound over, we observe that as (3) holds we may move to a sparse approximator of $\mathcal{T}$ and thus handle much less sets.

More precisely, using Lemma 24 and the guarantee from (3) we get that there is $\mathcal{M}$ of size at most αN for $\alpha = \frac{100 \log t}{\binom{k}{t}}$ satisfying the conclusion of the lemma for $\mathcal{B}_{\mathcal{M}}$ and $\mathcal{B}'_{\mathcal{M}}$ as therein. It follows that

$$\begin{aligned} \mu_k(\mathcal{B}'_{\mathcal{M}} \cap \mathcal{S}) &\leq \mu_k(\partial^{t,u}\mathcal{T} \cap \mathcal{S}) + \mu_k(\partial^{t,u}\mathcal{T} \Delta \mathcal{B}'_{\mathcal{M}}) \leq \left(\rho - \frac{t^2}{2n} + \frac{6}{t}\right)\mu_k(\partial^{t,u}\mathcal{T}) \\ &\leq \left(\rho - \frac{t^2}{2n} + \frac{7}{t}\right)\mu_k(\mathcal{B}'_{\mathcal{M}}), \end{aligned}$$

25:18 Improved Monotonicity Testers via Hypercube Embeddings

where in the last inequality we used the fact that $\mu_k(\partial^{t,u}\mathcal{T}) \leq (1 + \frac{1}{t})\mu_k(\mathcal{B}'_{\mathcal{M}})$ by the conclusion of Lemma 24. By the condition on t , $\rho - \frac{t^2}{2n} + \frac{7}{t} \leq \rho - \frac{t^2}{3n}$ and hence we conclude that

$$\mu_k(\mathcal{B}'_{\mathcal{M}} \cap \mathcal{S}) \leq \left(\rho - \frac{t^2}{3n}\right) \mu_k(\mathcal{B}'_{\mathcal{M}}).$$

We may assume that $\rho \geq t^2/3n$, otherwise the last inequality is impossible. We also note that from Lemma 24 we have $\mu_k(\mathcal{B}'_{\mathcal{M}}) \geq \frac{1}{2}\mu_k(\partial^{t,u}\mathcal{T}) \geq \frac{1}{2}\mu_{k-t}(\mathcal{T}) = \frac{N}{2\binom{n}{k-t}}$. From everything claimed so far we conclude that

$$\begin{aligned} & \Pr_S [\exists \mathcal{T} \text{ with } |\mathcal{T}| = N \text{ such that } \nu(\partial^{t,u}\mathcal{T}) < \mu(\mathcal{T})] \\ & \leq \Pr_S \left[\exists |\mathcal{M}| \leq \alpha N, \mu_k(\mathcal{B}'_{\mathcal{M}}) \geq \frac{N}{2\binom{n}{k-t}}, \mu_k(\mathcal{B}'_{\mathcal{M}} \cap \mathcal{S}) \leq \left(\rho - \frac{t^2}{3n}\right) \mu_k(\mathcal{B}'_{\mathcal{M}}) \right] \\ & \leq \sum_{\substack{\mathcal{M} \subseteq \binom{n}{k} \\ |\mathcal{M}| \leq \alpha N \\ \mu_k(\mathcal{B}'_{\mathcal{M}}) \geq \frac{N}{2\binom{n}{k-t}}}} \Pr_S \left[\mu_k(\mathcal{B}'_{\mathcal{M}} \cap \mathcal{S}) \leq \left(\rho - \frac{t^2}{3n}\right) \mu_k(\mathcal{B}'_{\mathcal{M}}) \right]. \end{aligned} \quad (7)$$

For each $\mathcal{M}$ such that $|\mathcal{M}| \leq \alpha N$ and $\mu_k(\mathcal{B}'_{\mathcal{M}}) \geq \frac{N}{2\binom{n}{k-t}}$, let $E_{\mathcal{M}}$ be the event that $\mu_k(\mathcal{B}'_{\mathcal{M}} \cap \mathcal{S}) \leq \left(\rho - \frac{t^2}{3n}\right) \mu_k(\mathcal{B}'_{\mathcal{M}})$; we upper bound the probability of each $E_{\mathcal{M}}$ separately, and for that we use Chernoff's bound. There is a slight technical issue in applying Chernoff's bound, namely that $\mathcal{S}$ is selected to be of fixed size, and to circumvent it we consider $\mathcal{S}' \subseteq \binom{[n]}{k}$ chosen randomly by including each set from $\binom{[n]}{k}$ in it with probability $\rho' = \rho - t^2/6n$. Then we get that

$$\Pr[|\mathcal{S}'| > s] = \Pr\left[|\mathcal{S}'| > \left(\rho' + \frac{t^2}{6n}\right) \binom{n}{k}\right] \leq e^{-\Omega\left(\frac{t^2}{n^4} \rho \binom{n}{k}\right)} \leq 0.5,$$

where we used Chernoff's bound and $\rho \geq t^2/3n$. Note that $\Pr_S[\mathcal{E}_{\mathcal{M}}] \leq \Pr_{\mathcal{S}'}[\mathcal{E}_{\mathcal{M}} \mid |\mathcal{S}'| \leq s]$, and combining with the above bound on the probability that $|\mathcal{S}'| > s$ we get that

$$\begin{aligned} \Pr_S[\mathcal{E}_{\mathcal{M}}] & \leq 2\Pr_{\mathcal{S}'}[\mathcal{E}_{\mathcal{M}}] = 2\Pr_{\mathcal{S}'}\left[\mu_k(\mathcal{B}'_{\mathcal{M}} \cap \mathcal{S}') \leq \left(\rho - \frac{t^2}{3n}\right) \mu_k(\mathcal{B}'_{\mathcal{M}})\right] \\ & = 2\Pr_{\mathcal{S}'}\left[\mu_k(\mathcal{B}'_{\mathcal{M}} \cap \mathcal{S}') \leq \left(\rho' - \frac{t^2}{6n}\right) \mu_k(\mathcal{B}'_{\mathcal{M}})\right] \\ & \leq 2e^{-\Omega\left(\frac{t^4}{n^2} \rho' \mu_k(\mathcal{B}'_{\mathcal{M}}) \binom{n}{k}\right)} \\ & \leq 2e^{-\Omega\left(\frac{t^4}{n^2} \rho' \frac{N}{2\binom{n}{k-t}} \binom{n}{k}\right)} \\ & \leq e^{-\Omega\left(\frac{t^6}{n^3} \frac{N \binom{n}{k}}{\binom{n}{k-t}}\right)}. \end{aligned}$$

Thus, using (7) we get that the left hand side therein is upper bounded by

$$\sum_{\substack{\mathcal{M} \subseteq \binom{n}{k} \\ |\mathcal{M}| \leq \alpha N}} e^{-\Omega\left(\frac{t^6}{n^3} \frac{N \binom{n}{k}}{\binom{n}{k-t}}\right)} \leq 2^{n\alpha N - \Omega\left(\frac{t^6}{n^3} \frac{N \binom{n}{k}}{\binom{n}{k-t}}\right)} \leq 2^N \left(n\alpha - \Omega\left(\frac{t^6}{n^3} \frac{\binom{n}{k}}{\binom{n}{k-t}}\right) \right).$$

Estimating, we get that $\frac{\binom{n}{k}}{\binom{n}{k-t}} \geq 2^{-O(t)}$ and $\alpha \leq \frac{100 \log n}{(k/t)^t} \leq n^{-\Omega(t)}$, hence $n\alpha - \Omega\left(\frac{t^6}{n^3} \frac{\binom{n}{k}}{\binom{n}{k-t}}\right) \leq -2^{-O(t)}$ and plugging this above yields that the left hand side of (7) is upper bounded by $2^{-2^{-O(t)}N}$. Thus, we conclude that

$$\Pr_S [\exists \mathcal{T} \text{ with } \mu_{k-t}(\mathcal{T}) \leq 1/2 \text{ such that } \nu(\partial^{t,u}\mathcal{T}) < \mu(\mathcal{T})] \leq \sum_{N \geq 2^{0.8n}} 2^{-2^{-O(t)}N},$$

which is at most $2^{-2^{-O(t)}2^{0.8n}} \leq 2^{-2^{n/2}}$.

The case that $\mu_{k-t}(\mathcal{T}) > 1/2$. Let $\mathcal{T}$ be such that $\mu_{k-t}(\mathcal{T}) > 1/2$, and suppose that $\nu(\partial^{t,u}\mathcal{T}) < \mu(\mathcal{T})$. The analysis is similar to before, except that we look at $\mathcal{R} = \partial^{t,u}\overline{\mathcal{T}}$ instead of $\mathcal{T}$. Thus, we get that $\nu(\mathcal{R}) > \mu(\overline{\mathcal{T}})$, and we argue that $\partial^{t,d}\mathcal{R} \subseteq \overline{\mathcal{T}}$. Indeed, if $x \in \partial^{t,d}\mathcal{R}$, then there is $y \in \mathcal{R}$ such that $x < y$, and as $y \in \mathcal{R}$ it follows that $y \notin \partial^{t,u}\mathcal{T}$ so for all $x' < y$ of Hamming weight $(k-t)$ – and in particular for $x' = x$ – we have that $x' \notin \mathcal{T}$, so $x \in \overline{\mathcal{T}}$.

Thus, it follows that $\mu(\partial^{t,d}\mathcal{R}) < \nu(\mathcal{R})$ and now $\mu_k(\mathcal{R}) = \mu_k(\partial^{t,u}\overline{\mathcal{T}}) = 1 - \mu_k(\partial^{t,u}\mathcal{T}) \leq 1 - \mu_{k-t}(\mathcal{T}) \leq 1/2$, and the rest of the argument is analogous to the previous argument. Let $N = |\mathcal{R}|$. First, we have

$$\mu_{k-t}(\partial^{t,d}\mathcal{R}) \leq \frac{1}{t} \mu(\partial^{t,d}\mathcal{R}) \leq \frac{1}{t} \nu(\mathcal{R}) \leq \frac{t+1}{t} \mu_k(\mathcal{R}). \quad (8)$$

On the other hand, using Lemma 23 we have $\mu_{k-t}(\partial^{t,d}\mathcal{R}) \geq \mu_k(\mathcal{R})^{(1-\frac{1}{n})^t}$. As before, this implies $\mu_k(\mathcal{R}) \geq e^{-O(n/t^2)}$ and so $N \geq e^{-O(n/t^2)} \binom{n}{k} \geq 2^{0.8n}$. Also, it implies

$$\mu_{k-t}(\partial^{t,d}\mathcal{R}) \geq \mu_k(\mathcal{R})(1/2)^{(1-\frac{1}{n})^t-1} \geq \mu_k(\mathcal{R})2^{t/n}. \quad (9)$$

We now conclude from (9) that

$$\begin{aligned} 0 < \nu(\mathcal{R}) - \mu(\partial^{t,d}\mathcal{R}) &= t\mu_k(\mathcal{R}) + \mu_k(\mathcal{R} \cap \mathcal{S}) - (t+\rho)\mu_k(\partial^{t,d}\mathcal{R}) \\ &\leq \mu_k(\mathcal{R} \cap \mathcal{S}) - (t+\rho-t2^{-t/n})\mu_{k-t}(\partial^{t,d}\mathcal{R}) \\ &\leq \mu_k(\mathcal{R} \cap \mathcal{S}) - (\rho+t^2/2n)\mu_{k-t}(\partial^{t,d}\mathcal{R}), \end{aligned}$$

so analogously to (6) we get that

$$\mu_k(\mathcal{R} \cap \mathcal{S}) \geq (\rho+t^2/2n)\mu_{k-t}(\partial^{t,d}\mathcal{R}) \geq (\rho+t^2/2n)\mu_k(\mathcal{R}). \quad (10)$$

As (8) holds, using Lemma 25, we get that there is $\mathcal{M}$ of size at most αN for $\alpha = \frac{100 \log t}{\binom{n}{k}}$ satisfying the conclusion of the lemma for $\mathcal{B}_{\mathcal{M}}$ and $\mathcal{B}'_{\mathcal{M}}$ as therein. It follows that

$$\begin{aligned} \mu_k(\mathcal{B}_{\mathcal{M}} \cap \mathcal{S}) &\geq \mu_k(\mathcal{R} \cap \mathcal{S}) - \mu_k(\mathcal{R} \Delta \mathcal{B}_{\mathcal{M}}) \geq \left(\rho + \frac{t^2}{2n} - \frac{18}{t}\right) \mu_k(\mathcal{R}) \\ &\geq \left(\rho + \frac{t^2}{2n} - \frac{36}{t}\right) \mu_k(\mathcal{B}_{\mathcal{M}}) \\ &\geq \left(\rho + \frac{t^2}{3n}\right) \mu_k(\mathcal{B}_{\mathcal{M}}), \end{aligned}$$

where we used the fact that $t \geq 10n^{1/3}$. We may assume $\rho < 1 - \frac{t^2}{3n}$, otherwise this is impossible. We denote $\rho' = \rho + \frac{t^2}{6n}$, so that now we are guaranteed that $t^2/6n \leq \rho' \leq 1 - \frac{t^2}{6n}$. We also get that

$$\mu_k(\mathcal{B}_{\mathcal{M}}) \geq \frac{1}{2} \mu_k(\mathcal{R}) \geq \frac{N}{2 \binom{n}{k}}.$$

Denote by $E_{\mathcal{M}}$ the event that $\mu_k(\mathcal{B}_{\mathcal{M}} \cap \mathcal{S}) \geq \left(\rho + \frac{t^2}{3n}\right) \mu_k(\mathcal{B}_{\mathcal{M}})$. We now apply the Chernoff argument again; letting $\mathcal{S}' \subseteq \binom{[n]}{k}$ be chosen randomly by including each set with probability $\rho'' = \rho' + \frac{t^2}{12n}$, we get by Chernoff's bound that $|\mathcal{S}'| \geq s$ except with probability at most $1/2$ and so

$$\Pr_{\mathcal{S}}[\mathcal{E}_{\mathcal{M}}] \leq 2\Pr_{\mathcal{S}'}[\mathcal{E}_{\mathcal{M}}] \leq 2e^{-\Omega\left(\frac{t^4}{n^2} \rho' \mu_k(\mathcal{B}_{\mathcal{M}}) \binom{n}{k}\right)} \leq e^{-\Omega\left(\frac{t^6}{n^3} \frac{\binom{n}{k} N}{\binom{n}{k}}\right)}.$$

Thus, by the union bound

$$\Pr_{\mathcal{S}}[\exists \mathcal{R} \text{ of size } N \text{ such that } \mu(\partial^{t,d}\mathcal{R}) < \nu(\mathcal{R})] \leq \sum_{|\mathcal{M}| \leq \alpha N} \Pr_{\mathcal{S}}[\mathcal{E}_{\mathcal{M}}] \leq 2^{n\alpha N - \Omega\left(\frac{t^6}{n^3} N\right)},$$

and by a direct computation the last expression is at most $2^{-2^{-O(t)}N}$. Summing over $N \geq 2^{0.8n}$ yields that

$$\begin{aligned} & \Pr_{\mathcal{S}}[\exists \mathcal{T} \text{ such that } \mu_{k-t}(\mathcal{T}) \geq 1/2, \mu(\partial^{t,u}\mathcal{T}) > \nu(\mathcal{T})] \\ & \leq \Pr_{\mathcal{S}}[\exists \mathcal{R} \text{ such that } \mu_k(\mathcal{R}) \leq 1/2, \mu(\partial^{t,d}\mathcal{R}) < \nu(\mathcal{R})] \\ & \leq \sum_{N \geq 2^{0.8n}} 2^{-2^{-O(t)}N}, \end{aligned}$$

which is at most $2^{-2^{n/2}}$ provided that n_0 is large enough. $\blacktriangleleft$

We will also need a version of Lemma 31 that works the other way around – namely one that matches a slice and a random subset of it with a slice above it, and we state it separately below.

► **Lemma 32.** *For all $C > 0$ there is $n_0 \in \mathbb{N}$ such that the following holds. Let $n, k, s, t \in \mathbb{N}$ and assume that $\frac{n}{2} - C\sqrt{n \log n} \leq k \leq \frac{n}{2} + C\sqrt{n \log n}$ and $10n^{1/3} \leq t \leq C\sqrt{n \log n}$. Then for every $0 \leq s \leq \binom{n}{k}$, setting $\rho = \frac{s}{\binom{n}{k}}$ we have*

$$\Pr_{\substack{\mathcal{S} \subseteq \binom{[n]}{k} \\ |\mathcal{S}|=s}}[t\mu_k + \mu_{\mathcal{S}} \lesssim (t + \rho)\mu_{k+t}] \geq 1 - 2^{-\Omega(2^{n/2})}.$$

Proof. Let $\mathcal{S}' = \{[n] \setminus A \mid A \in \mathcal{S}\}$ and note that it is a random subset of $\binom{[n]}{n-k}$ of size s , so applying Lemma 31 on $n - k$ instead of k we get that with probability at least $1 - 2^{-\Omega(2^{n/2})}$ there is a monotone fractional matching $w(x, y)$ from $(t + \rho)\mu_{n-k-t}$ to $t\mu_{n-k} + \mu_{\mathcal{S}'}$. Define $w'(A, B) = w(\overline{B}, \overline{A})$, and note that it is a monotone fractional matching from $t\mu_k + \mu_{\mathcal{S}}$ to $(t + \rho)\mu_{k+t}$. $\blacktriangleleft$

3.5 Matching Union of Slices and a Random Subset to a Slice

Next, we use Lemma 31 to show that given a union of consecutive slices and a random subset of the topmost one, one can find a monotone fractional matching with each of the following: (1) a slice which is a bit above them, and (2) a slice which is a bit below them.

► **Corollary 33.** *For all $C > 0$ there is $n_0 \in \mathbb{N}$, such that the following holds for all $n \geq n_0$. Let n, t, k, s be as in Lemma 31, let $\mathcal{S}$ be random subset of $\binom{[n]}{k}$ of size s and let $t \leq d \leq k/2$ be a parameter such that $\sum_{i=k-d}^{k-1} \binom{n}{i} \geq t \binom{n}{k}$. Denote $\mathcal{T} = \mathcal{S} \cup \bigcup_{i=k-d}^{k-1} \binom{[n]}{i}$, and let $\nu_{\mathcal{T}}$ be the uniform distribution over $\mathcal{T}$. Then*

$$\Pr_{\mathcal{S}}[\mu_{k-2d} \lesssim \nu_{\mathcal{T}} \lesssim \mu_{k+d}] \geq 1 - 2^{-\Omega(2^{n/2})}.$$

Proof. We show that with probability $1 - 2^{-\Omega(2^{n/2})}$ we have that $\nu_{\mathcal{T}} \lesssim \mu_{k+d}$, and also that with probability $1 - 2^{-\Omega(2^{n/2})}$ we have that $\mu_{k-2d} \lesssim \nu_{\mathcal{T}}$. The statement then follows from the union bound.

For the first statement, let $\rho = s/\binom{n}{k}$. By Lemma 29 we have that $\mu_i \lesssim \mu_k$ for $i \leq k$, so using Lemma 28 we get that

$$\begin{aligned} \nu_{\mathcal{T}} &= \frac{\binom{n}{k}}{|\mathcal{T}|} \left(\rho \mu_S + \sum_{i=k-d}^{k-1} \frac{\binom{n}{i}}{\binom{n}{k}} \mu_i \right) \lesssim \frac{\binom{n}{k}}{|\mathcal{T}|} \left(\rho \mu_S + \sum_{i=k-d}^{k-1} \frac{\binom{n}{i}}{\binom{n}{k}} \mu_k \right) \\ &= \frac{\binom{n}{k}}{|\mathcal{T}|} \left(\rho \mu_S + \frac{|\mathcal{T}| - \rho \binom{n}{k}}{\binom{n}{k}} \mu_k \right). \end{aligned}$$

By Lemma 32 we have that $t\mu_k + \rho\mu_S \lesssim (t+\rho)\mu_{k+t}$ with probability $1 - 2^{-\Omega(2^{n/2})}$, in which case we get that

$$\nu_{\mathcal{T}} \lesssim \frac{\binom{n}{k}}{|\mathcal{T}|} \left((t+\rho)\mu_{k+t} + \frac{|\mathcal{T}| - \rho \binom{n}{k} - t \binom{n}{k}}{\binom{n}{k}} \mu_k \right),$$

where we used the fact that $|\mathcal{T}| - \rho \binom{n}{k} = \sum_{i=k-d}^{k-1} \binom{n}{i} \geq t \binom{n}{k}$. Using $\mu_k \lesssim \mu_{k+t}$ and Lemma 28 again and then simplifying, we conclude that $\nu_{\mathcal{T}} \lesssim \mu_{k+t}$.

For the second statement, we note that by Lemmas 29, 31 we have that $(t+\rho)\mu_{k-t} \lesssim t\mu_k + \rho\mu_S$ with probability at least $1 - 2^{-\Omega(2^{n/2})}$.

▷ **Claim 34.** If $(t+\rho)\mu_{k-t} \lesssim t\mu_k + \rho\mu_S$, then $(t+\rho)\mu_{k-t-d} \lesssim t\mu_{k-d} + \rho\mu_S$.

Proof. Let $w(x, y)$ be a weight function showing that $(t+\rho)\mu_{k-t} \lesssim t\mu_k + \rho\mu_S$. We consider the probability distribution $p(x, y) = \frac{1}{t+\rho} w(x, y)$, and define a probability distribution p' over (x', y') as follows:

1. Sample $(x, y) \sim p$ and independently a random permutation π on $[n]$.
2. Let $J \subseteq [n]$ be the set of first d coordinates according to π wherein $x_j = 1$. We define $x'_i = x_i$ on $i \notin J$ and $x'_i = 0$ on $i \in J$.
3. If $y \in \mathcal{S}$, with probability $1/2$ take $y' = y$. Otherwise, let $J' \subseteq [n]$ be the set of first d coordinates according to π wherein $y_j = 1$, and take y' to be the vector where $y'_i = y_i$ on $i \notin J'$ and $y'_i = 0$ on $i \in J'$.

We argue that $w'(x, y) = (t+\rho)p'(x, y)$ shows that the fractional monotone matching as stated in the claim exists. For $y \in \mathcal{S}$ we have

$$\sum_{x'} w'(x', y) = \frac{1}{2} \sum_x w(x, y) = \rho = (t\mu_{k-d} + \rho\mu_S)(y),$$

and for $y' \notin \mathcal{S}$ we have that $\sum_{x'} p'(x', y')$ is the probability that we pick y according to μ_k , turn from 1 to 0 a random set of d coordinates and reach y' , which is the $\mu_{k-d}(y')$. Thus, $\sum_{x'} w'(x', y') = (t\mu_{k-d} + \rho\mu_S)(y')$.

For x' , $\sum_{y'} p'(x', y')$ is the probability we take $x \sim \mu_{k-t}$, turn from 1 to 0 a random set of d coordinates and reach x' , which is equal to $\mu_{k-t-d}(x')$, hence $\sum_{y'} w'(x', y') = (t+\rho)\mu_{k-t-d}(x')$.

◁

Using Claim 34 we get by Lemmas 28, 29

$$\begin{aligned}
\mu_{k-2d} &\lesssim \frac{\binom{n}{k}}{|\mathcal{T}|} \left(t\mu_{k-d} + \rho\mu_S + \frac{\sum_{i=k-d}^{k-1} \binom{n}{i} - t\binom{n}{k}}{\binom{n}{k}} \mu_{k-2d} \right) \\
&\lesssim \frac{\binom{n}{k}}{|\mathcal{T}|} \left(t\mu_{k-d} + \rho\mu_S + \frac{\sum_{i=k-d}^{k-1} \binom{n}{i} - t\binom{n}{k}}{\binom{n}{k}} \mu_{k-d} \right) \\
&= \frac{\binom{n}{k}}{|\mathcal{T}|} \left(\rho\mu_S + \frac{\sum_{i=k-d}^{k-1} \binom{n}{i}}{\binom{n}{k}} \mu_{k-d} \right) \\
&\lesssim \frac{\binom{n}{k}}{|\mathcal{T}|} \left(\rho\mu_S + \frac{\sum_{i=k-d}^{k-1} \binom{n}{i} \mu_i}{\binom{n}{k}} \right),
\end{aligned}$$

which is equal to $\nu_{\mathcal{T}}$. ◀

3.6 Proof of Theorem 20

In this section, we prove Theorem 20 which by Lemma 21 implies Theorems 2, 3. Theorem 20 is a direct consequence of the following more precise statement:

► **Theorem 35.** *There exists $C > 0$ such that for all $m \in \mathbb{N}$ and all $n \geq C \cdot m^6$, there are m sets $P_1, \dots, P_m \subseteq \{0, 1\}^n$ satisfying the following properties:*

1. $P_i \cap P_j = \emptyset$ for all $i \neq j$.
2. $|P_i| = \lfloor \frac{2^n}{m} \rfloor$ for all i .
3. For each i , $x \in P_i$ and $y \in P_{i+1}$ we have that $|x| \leq |y|$. For all $x \in P_1 \cup \dots \cup P_m$ and $y \notin P_1 \cup \dots \cup P_m$ we have that $|x| \leq |y|$.
4. For each i there is a monotone matching from P_i to P_{i+1} .

In particular, the following function ϕ is monotone and admits an $m2^{-n}$ -almost perfect matching: $\phi(x) = i - 1$ if $x \in P_i$, and otherwise $\phi(x) = m - 1$.

Proof. We present a randomized construction and show that it works with probability $1 - o(1)$. Consider a random $\pi: \{0, 1\}^n \rightarrow [2^n]$ such that $\pi(x) \leq \pi(y)$ whenever $|x| \leq |y|$; in other words, we first think of an ordering of $\{0, 1\}^n$ as $x_0, \dots, x_{2^n-1}$, where we first enumerate according to Hamming weight and within each layer we order randomly. Thus, we may take $\pi(x_i) = i$.

Define the $P_1, \dots, P_m$ as

$$P_i = \pi^{-1} \left(\left\{ \left\lfloor \frac{2^n}{m} \right\rfloor (i-1) + j \mid j = 1, \dots, \left\lfloor \frac{2^n}{m} \right\rfloor \right\} \right),$$

so that the first three items holds trivially. In the rest of the proof, we argue that the fourth item holds with probability $1 - o(1)$. Denote by ℓ_i and u_i the smallest and largest Hamming weight of vectors from P_i , and by ν_i the uniform distribution over P_i . It suffices to prove that with probability $1 - 2^{-\Omega(2^{n/2})}$ we have that $\nu_i \lesssim \nu_{i+1}$ for all i . Indeed, then we get by Lemma 30 that there is a monotone matching from P_i and P_{i+1} , and the fourth item follows.

We now show that for each i , $\nu_i \lesssim \nu_{i+1}$ with probability $1 - 2^{-\Omega(2^{n/2})}$, and then the claim follows by the union bound. We intend to use Corollary 33 to show that and therefore we break each one of ν_i, ν_{i+1} into lower and upper part. Let m_i be the median Hamming weight of ν_i , namely such that $\nu_i(\{x \mid |x| < m_i\}) < 0.5$ but $\nu_i(\{x \mid |x| \leq m_i\}) \geq 0.5$. Define $p_i = \nu_i(\{x \mid |x| < m_i\})$, $q_i = \nu_i(\{x \mid |x| > m_i\})$, and let

$$\nu_i^-(x) = \nu_i(x)1_{|x| < m_i} + \left(\frac{1}{2} - p_i\right) \mu_{m_i}(x), \quad \nu_i^+(x) = \nu_i(x)1_{|x| > m_i} + \left(\frac{1}{2} - q_i\right) \mu_{m_i}(x).$$

Let $t = \lceil 10n^{1/3} + 1 \rceil$. Our goal is to show that with probability $1 - 2^{-\Omega(2^{n/2})}$, for all i we have

$$\frac{1}{2} \mu_{\ell_i - 2t} \lesssim \nu_i^-, \quad \nu_i^- \lesssim \frac{1}{2} \mu_{m_i}, \quad \nu_i^+ \lesssim \frac{1}{2} \mu_{u_i + 2t}, \quad \frac{1}{2} \mu_{m_i} \lesssim \nu_i^+, \quad (11)$$

in which case we get, using Lemma 28, that

$$\nu_i = \nu_i^- + \nu_i^+ \lesssim \frac{1}{2} \mu_{m_i} + \frac{1}{2} \mu_{u_i + 2t} \lesssim \frac{1}{2} \mu_{\ell_{i+1} - 2t} + \frac{1}{2} \mu_{m_{i+1}} \lesssim \nu_{i+1}^- + \nu_{i+1}^+ = \nu_{i+1}.$$

Here, we also use the facts that $m_i \leq \ell_{i+1} - 2t$, $u_i + 2t \leq m_{i+1}$ and Lemma 29. This follows since the probability mass of each layer in the hypercube is at most $O(1/\sqrt{n})$, hence each P_i must intersect at least $\Omega(\sqrt{n}/m)$ distinct layers and so $u_i - m_i \geq \Omega(\sqrt{n}/m) \geq \Omega(C^{1/6}n^{1/3}) > 100t$, and in the same way $m_i - \ell_i > 100t$ and $\ell_{i+1} - m_i \geq 100t$. We also note that all of the u_i, m_i and ℓ_i 's are all in the range $[\frac{n}{2} - \sqrt{100n \log n}, \frac{n}{2} + \sqrt{100n \log n}]$ since the total probability mass outside this range is at most $e^{-\frac{1}{2}100 \log n} \leq n^{-10} < 1/m$.

We finish by arguing that (11) holds for each i with probability $1 - 2^{-\Omega(2^{n/2})}$, and for that we apply Corollary 33. Set $d = 2t$; we argue that for each $\ell_1 \leq k \leq u_m$ it holds that

$$\sum_{i=k-d}^{k-1} \binom{n}{i} \geq t \binom{n}{k}.$$

Indeed, this follows since the ratio between any two consecutive binomial coefficients $\binom{n}{i}$ for $i = k-d, \dots, k$ is $1 + O(\sqrt{\log n/n})$, so the ratio between any two (not necessarily consecutive) binomial coefficients in that range is at most $\left(1 + O(\sqrt{\log n/n})\right)^d = 1 + o(1)$,

so $\sum_{i=k-d}^{k-1} \binom{n}{i} \geq (1 + o(1))d \binom{n}{k} = (2 + o(1))t \binom{n}{k} > t \binom{n}{k}$. Thus, the conditions of Corollary 33

hold, and applying it for various k 's we get that (11) holds with probability $1 - 2^{-\Omega(2^{n/2})}$. Below, we explain in details how to deduce that $\frac{1}{2} \mu_{\ell_i - 2t} \lesssim \nu_i^-$, and the other arguments are similar.

We view $\nu_i(x)1_{|x| < m_i}$ as a uniform weight function over the part of P_i of Hamming weight less than m_i , which is a union of slices and a random subset of the appropriate size of the slice ℓ_i . Thus by Corollary 33 we get that $p_i \mu_{\ell_i - 2t} \lesssim \nu_i(x)1_{|x| < m_i}$ with probability $1 - 2^{-\Omega(2^{n/2})}$, and as $\mu_{\ell_i - 2t} \lesssim \mu_{m_i}$ we get from Lemma 28 that $\frac{1}{2} \mu_{\ell_i - 2t} \lesssim \nu_i^-$. ◀

References

- 1 A matching for $[4]^n$. <https://www.cs.huji.ac.il/w~gkindler/matchings/index.html>. Accessed: 2022-08-06.
- 2 Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. A $o(d) \cdot \text{polylog } n$ monotonicity tester for boolean functions over the hypergrid $[n]^d$. In *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2018, New Orleans, LA, USA, January 7-10, 2018*, pages 2133–2151, 2018. doi:10.1137/1.9781611975031.139.

- 3 Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. Domain reduction for monotonicity testing: A $o(d)$ tester for boolean functions in d -dimensions. In *Proceedings of the 2020 ACM-SIAM Symposium on Discrete Algorithms, SODA 2020, Salt Lake City, UT, USA, January 5-8, 2020*, pages 1975–1994, 2020. doi:10.1137/1.9781611975994.122.
- 4 Hadley Black, Deeparnab Chakrabarty, and C. Seshadhri. Directed isoperimetric theorems for boolean functions on the hypergrid and an $\tilde{O}(n\sqrt{d})$ monotonicity tester, 2022. doi:10.48550/arXiv.2211.05281.
- 5 Hadley Black, Iden Kalemaj, and Sofya Raskhodnikova. Isoperimetric inequalities for real-valued functions with applications to monotonicity testing. *arXiv preprint*, 2020. arXiv:2011.09441.
- 6 Béla Bollobás and Arthur G Thomason. Threshold functions. *Combinatorica*, 7(1):35–38, 1987.
- 7 Deeparnab Chakrabarty and C. Seshadhri. An optimal lower bound for monotonicity testing over hypergrids. *Theory Comput.*, 10:453–464, 2014. doi:10.4086/toc.2014.v010a017.
- 8 Deeparnab Chakrabarty and C. Seshadhri. An $o(n)$ monotonicity tester for boolean functions over the hypercube. *SIAM J. Comput.*, 45(2):461–472, 2016. doi:10.1137/13092770X.
- 9 Xi Chen, Rocco A. Servedio, and Li-Yang Tan. New algorithms and lower bounds for monotonicity testing. In *55th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2014, Philadelphia, PA, USA, October 18-21, 2014*, pages 286–295, 2014. doi:10.1109/FOCS.2014.38.
- 10 Xi Chen, Erik Waingarten, and Jinyu Xie. Beyond talagrand functions: new lower bounds for testing monotonicity and unateness. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, pages 523–536, 2017. doi:10.1145/3055399.3055461.
- 11 Yevgeniy Dodis, Oded Goldreich, Eric Lehman, Sofya Raskhodnikova, Dana Ron, and Alex Samorodnitsky. Improved testing algorithms for monotonicity. In *Randomization, Approximation, and Combinatorial Optimization. Algorithms and Techniques*, pages 97–108. Springer, 1999.
- 12 Eldar Fischer, Eric Lehman, Ilan Newman, Sofya Raskhodnikova, Ronitt Rubinfeld, and Alex Samorodnitsky. Monotonicity testing over general poset domains. In *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pages 474–483, 2002. doi:10.1145/509907.509977.
- 13 Oded Goldreich, Shafi Goldwasser, Eric Lehman, Dana Ron, and Alex Samorodnitsky. Testing monotonicity. *Comb.*, 20(3):301–337, 2000. doi:10.1007/s004930070011.
- 14 Gyula Katona. A theorem of finite sets. In *Classic Papers in Combinatorics*, pages 381–401. Springer, 2009.
- 15 Subhash Khot, Dor Minzer, and Muli Safra. On monotonicity testing and boolean isoperimetric-type theorems. *SIAM J. Comput.*, 47(6):2238–2276, 2018. doi:10.1137/16M1065872.
- 16 Joseph B Kruskal. The number of simplices in a complex. *Mathematical optimization techniques*, 10:251–278, 1963.
- 17 G. A. Margulis. Probabilistic characteristics of graphs with large connectivity. *Problemy Peredači Informacii*, 10(2):101–108, 1974.
- 18 Ryan O’Donnell and Karl Wimmer. KKL, Kruskal–Katona, and monotone nets. *SIAM Journal on Computing*, 42(6):2375–2399, 2013.
- 19 M. Talagrand. Isoperimetry, logarithmic Sobolev inequalities on the discrete cube, and Margulis’ graph connectivity theorem. *Geom. Funct. Anal.*, 3(3):295–314, 1993. doi:10.1007/BF01895691.