Combinatorics, Probability and Computing (2018) 27, 411–426. © Cambridge University Press 2017
doi:10.1017/S0963548317000578

On Lipschitz Bijections Between Boolean Functions

SHRAVAS RAO[1†] and IGOR SHINKAR[2]

[1]Courant Institute of Mathematical Sciences, New York University, NY 10012, USA
(e-mail: rao@cims.nyu.edu)
[2]Department of Electrical Engineering and Computer Sciences, UC Berkeley, CA 94720, USA
(e-mail: igors@berkeley.edu)

Received 15 February 2015; revised 24 October 2017; first published online 19 December 2017

Given two functions $f, g : \{0,1\}^n \to \{0,1\}$, a mapping $\psi : \{0,1\}^n \to \{0,1\}^n$ is said to be a *mapping from f to g* if it is a bijection and $f(z) = g(\psi(z))$ for every $z \in \{0,1\}^n$. In this paper we study Lipschitz mappings between Boolean functions.

Our first result gives a construction of a C-Lipschitz mapping from the Majority function to the Dictator function for some universal constant C. On the other hand, there is no $n/2$-Lipschitz mapping in the other direction, namely from the Dictator function to the Majority function. This answers an open problem posed by Daniel Varga in the paper of Benjamini, Cohen and Shinkar (*FOCS 2014* [1]).

We also show a mapping from Dictator to XOR that is 3-local, 2-Lipschitz, and its inverse is $O(\log(n))$-Lipschitz, where by L-local mapping we mean that each output bit of the mapping depends on at most L input bits.

Next, we consider the problem of finding functions such that any mapping between them must have large *average stretch*, where the average stretch of a mapping ϕ is defined as

$$\mathsf{avgStretch}(\phi) = \mathbb{E}_{x,i}[\mathsf{dist}(\phi(x), \phi(x + e_i)].$$

We show that any mapping ϕ from XOR to Majority must satisfy $\mathsf{avgStretch}(\phi) \geqslant c\sqrt{n}$ for some absolute constant $c > 0$. In some sense, this gives a 'function analogue' to the question of Benjamini, Cohen and Shinkar (*FOCS 2014* [1]), who asked whether there exists a set $A \subseteq \{0,1\}^n$ of density 0.5 such that any bijection from $\{0,1\}^{n-1}$ to A has large average stretch.

Finally, we show that for a random balanced function $f : \{0,1\}^n \to \{0,1\}^n$, with high probability there is a mapping ϕ from Dictator to f such that both ϕ and ϕ^{-1} have constant average stretch. In particular, this implies that one cannot obtain lower bounds on average stretch by taking uniformly random functions.

2010 *Mathematics subject classification*: Primary 68R05

† This material is based upon work supported by the National Science Foundation Graduate Research Fellowship Program under grant DGE-1342536. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

1. Introduction

Given two functions $f, g : \{0,1\}^n \to \{0,1\}$, a mapping $\psi : \{0,1\}^n \to \{0,1\}^n$ is said to be a *mapping from f to g* if ψ is a bijection and for every $z \in \{0,1\}^n$ it holds that $f(z) = g(\psi(z))$. Of course, if $\mathbb{E}[f] = \mathbb{E}[g]$, then there are many mappings from f to g, and we can further ask whether there are 'simple' mappings from f to g, where 'simple' can mean, for example, that ψ is computable by a small circuit, or has some other nice structure. In this paper we ask about the existence of Lipschitz mappings between some well-studied Boolean functions, including the functions Dictator, Majority, XOR, and a uniformly random balanced function.

As a first example suppose that f is obtained from g by renaming the coordinates. Then, trivially, there is a 1-Lipschitz mapping from f to g, which simply permutes the coordinates, and in particular, each output bit of the mapping depends on exactly one input of its input bits.

In some sense existence of a Lipschitz mapping between f and g implies some similarity between them because such a mapping induces

- a Lipschitz bijection from $f^{-1}(0)$ to $g^{-1}(0)$,
- a Lipschitz bijection from $f^{-1}(1)$ to $g^{-1}(1)$, and
- a Lipschitz mapping from the cut in $\{0,1\}^n$ defined by f to the cut defined by g.

Below we summarize the results shown in this paper.

Bijections between Dictator and Majority. It is a recurring theme in the analysis of Boolean functions that the Dictator function and the Majority function are in some senses the opposite of one another. For example, the 'Majority is Stablest' theorem [7] states that if the noise stability of a function significantly deviates from the noise stability of Majority, the function must have an influential coordinate, and hence is non-trivially correlated with the corresponding Dictator function. Another example is the theorem of Bourgain [2] (see also a recent improvement by Kindler and O'Donnell [5]) saying that if the Fourier transform of a function deviates in an appropriate sense from that of Majority, then the function can be approximated by a *junta*, that is, it essentially depends on a small number of coordinates, which also implies some correlation with a Dictator function.

Lovett and Viola [6], motivated by questions related to lower bounds on sampling by low-level complexity classes [8], suggested further exploring the differences between the two functions, and asked whether it is true that any bijective mapping $\phi : \mathsf{Dictator}^{-1}(1) \to \mathsf{Majority}^{-1}(1)$ must have a large average stretch, where by average stretch we refer to the quantity

$$\mathsf{avgStretch}(\phi) = \mathbb{E}_{x \sim y \in \mathsf{Dictator}^{-1}(1)}[\mathsf{dist}(\phi(x), \phi(y))],$$

with $x \sim y \in \mathsf{Dictator}^{-1}(1)$ denoting a random edge in $\{0,1\}^n$ such that $x_1 = y_1 = 1$.

Benjamini, Cohen and Shinkar [1] answered the question negatively in a stronger sense, by showing an explicit bi-Lipschitz bijection that maps $\{0,1\}^{n-1}$ to the upper half of $\{0,1\}^n$ (or equivalently a mapping from $\mathsf{Dictator}^{-1}(1)$ to $\mathsf{Majority}^{-1}(1)$). In the same paper the following problem was raised.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

Problem 1.1. *Let n be odd. Is there a bi-Lipschitz bijection $f : \{0,1\}^n \to \{0,1\}^n$ that maps the half-cube to the Hamming ball? In other words, is there a bi-Lipschitz bijection from* Majority *to* Dictator*?*

It is easy to see that there is no C-Lipschitz bijection from Dictator to Majority for $C < n/2$. Indeed for any bijection ϕ from $\mathsf{Dictator}(x) = x_1$ to Majority, consider $x \in \{0,1\}^n$ such that $\phi(x) = (1,1,\ldots,1)$, and let $y = x - e_1$. Then, the weight of $\phi(y)$ must be at most $n/2$ since $\mathsf{Majority}(\phi(y)) = \mathsf{Dictator}(y) = 0$, and thus $\mathsf{dist}(\phi(x), \phi(y)) \geqslant n/2$.

In the other direction the answer is not as obvious, and we resolve this question in this paper. Specifically, we prove the following theorem in Section 2.

Theorem 1.2. *For all odd integers $n \in \mathbb{N}$ there exists a C-Lipschitz bijection $\psi : \{0,1\}^n \to \{0,1\}^n$ from* Majority *to* Dictator*, where $C \in \mathbb{N}$ is some absolute constant.*

As mentioned above, Benjamini, Cohen and Shinkar [1] showed that there exists a bi-Lipschitz bijection that maps the upper half of $\{0,1\}^n$ to $\{0,1\}^{n-1}$. Therefore, there exists a bijection from $\{0,1\}^n$ to $\{0,1\}^n$ that maps the upper half of $\{0,1\}^n$ to $\{x \in \{0,1\}^n : x_1 = 1\}$, and maps the lower half of $\{0,1\}^n$ to $\{x \in \{0,1\}^n : x_1 = 0\}$, that is Lipschitz on the upper half of the hypercube, and on the lower half of the hypercube. However, it was not clear how to 'stitch' these two bijections so that the endpoints of the edges in the middle layer will also be mapped close to each other. Theorem 1.2 says that this is indeed possible.

Bijections between Dictator **and** XOR**.** We further study the notion of mappings between Boolean functions by studying mappings between the Dictator function and the XOR function. For this question the well-known mapping $\phi(x_1, x_2, \ldots, x_n) = (x_1 + x_2, x_2 + x_3, \ldots, x_{n-1} + x_n, x_n)$ is clearly a bijection from $\mathsf{Dictator}(x) = x_1$ to XOR. Note, however, that it is not bi-Lipschitz, as flipping the kth bit in the output changes its pre-image in the first k coordinates. That is, if $y = \phi(x)$, then $\phi^{-1}(y + e_k) = x + \sum_{i=1}^{k} e_i$.

In fact, it is not difficult to come up with a bi-Lipschitz mapping from Dictator to XOR. Indeed, define $\psi(x_1, x_2, \ldots, x_n) = (\mathsf{XOR}(x), x_2, \ldots, x_n)$. It is easy to check that ψ is indeed a 2-bi-Lipschitz mapping from $\mathsf{Dictator}(x) = x_1$ to XOR. It makes sense, however, to ask for more, namely, whether there exists a bi-Lipschitz mapping from Dictator to XOR that is in $\mathbf{NC^0}$, that is, each of its output bits depends only on a constant number of input bits.[1]

We prove the following theorem in Section 3.

Theorem 1.3. *There exists a Lipschitz mapping ϕ from* Dictator *to* XOR *such that each of its output bits depends on at most three input bits, ϕ is 2-Lipschitz, and its inverse ϕ^{-1} is $O(\log(n))$-Lipschitz. Furthermore, the mapping ϕ is a linear operator over $GF(2)$.*

Bijections between Majority **and** XOR**.** Benjamini, Cohen and Shinkar [1] asked whether there exists a subset $A \subset \{0,1\}^{n+1}$ of density $1/2$ such that any bijection from $\{0,1\}^n$

[1] Note that the inverse mapping, namely a bijection from XOR to Dictator, cannot be local since the dictating output coordinate must be the parity of all input bits, and thus depend on all of them.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

to A must map endpoints of many edges of the hypercube far apart. Specifically, for a mapping $\phi : \{0,1\}^n \to A$ they define the *average stretch of* ϕ as $\mathsf{avgStretch}(\phi) = \mathbb{E}_{x\in\{0,1\}^n, i\in[n]}[\mathsf{dist}(\phi(x), \phi(x+e_i))]$ and pose the following problem.

Problem 1.4. *Is there* $A \subset \{0,1\}^{n+1}$ *of density* $1/2$ *such that any bijection* $\phi : \{0,1\}^n \to A$ *has* $\mathsf{avgStretch}(\phi) = \omega(1)$.

We remark that we are not aware of the existence of a subset $A \subset \{0,1\}^{n+1}$ of density $1/2$ such that any bijection $f : \{0,1\}^n \to A$ has $\mathsf{avgStretch}(f) > 2.1$, and we find this open problem very interesting. It also makes sense to relax Problem 1.4 to an appropriate 2-set version, where we ask for two sets $A, B \in \{0,1\}^n$ of density $1/2$ such that any bijection $f : A \to B$ has large average stretch in the appropriate sense.

Below we give a positive answer to the 'function analogue' of this question. Specifically, we show that any bijection from XOR to Majority must have large average stretch.

Theorem 1.5. *Any mapping* $\phi : \{0,1\}^n \to \{0,1\}^n$ *from* XOR *to* Majority *must satisfy*

$$\mathsf{avgStretch}(\phi) \geqslant c\sqrt{n}$$

for some absolute constant $c > 0$.

On the other hand, there exists a C*-Lipschitz mapping* $\psi : \{0,1\}^n \to \{0,1\}^n$ *from* Majority *to* XOR *for some absolute constant* C.

We prove Theorem 1.5 in Section 4.

Bijections between Dictator **and a random balanced function.** We also show that for a random balanced function $f : \{0,1\}^n \to \{0,1\}$ with high probability there is a mapping ϕ from the Dictator function to f such that both ϕ and ϕ^{-1} have constant average stretch.

Theorem 1.6. *Let* $f : \{0,1\}^n \to \{0,1\}$ *be a uniformly random balanced Boolean function. Then, with probability* $1 - 2^{-2^{\Omega(n)}}$ *there exists a mapping* $\phi : \{0,1\}^n \to \{0,1\}^n$ *from* Dictator *to* f *such that for* $1 - O(1/n)$ *of the* $x \in \{0,1\}^n$ *it holds that* $\mathsf{dist}(x, \phi(x)) \leqslant 2$. *In particular,* ϕ *satisfies* $\mathsf{avgStretch}(\phi) = O(1)$ *and* $\mathsf{avgStretch}(\phi^{-1}) = O(1)$.

This implies that, for two random balanced functions, with high probability there is a bijective mapping between them such that both the mapping and its inverse have constant average stretch.

Corollary 1.7. *Let* $f, g : \{0,1\}^n \to \{0,1\}$ *be two uniformly random balanced Boolean functions. Then, with probability* $1 - 2^{-2^{\Omega(n)}}$ *there exists a bijection* $\phi : \{0,1\}^n \to \{0,1\}^n$ *from* f *to* g *that satisfies* $\mathsf{avgStretch}(\phi) = O(1)$ *and* $\mathsf{avgStretch}(\phi^{-1}) = O(1)$.

Proof. Let ϕ_f, ϕ_g be bijections given by Theorem 1.6 when applied on f and g respectively. Then, it is easy to see that the composition of ϕ_g with the inverse of ϕ_f gives us the desired mapping $\phi = \phi_g \circ \phi_f^{-1}$. Indeed, since ϕ_f is a bijection, by Theorem 1.6 it

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

satisfies

$$\mathbb{P}_{x\in\{0,1\}^n}[\mathsf{dist}(\phi_f(x),x)\leqslant 2]=1-O(1/n),$$

and hence

$$\begin{aligned}\mathbb{P}_x[\mathsf{dist}(\phi_g(\phi_f^{-1}(x)),x)\geqslant 4]&\leqslant \mathbb{P}_x[\mathsf{dist}(\phi_g(\phi_f^{-1}(x)),\phi_f^{-1}(x))\geqslant 2]+\mathbb{P}_x[\mathsf{dist}(\phi_f^{-1}(x),x)\geqslant 2]\\&=O(1/n).\end{aligned}$$

Therefore, for $1-O(1/n)$ of the edges it holds that $\mathsf{dist}(\phi(x),\phi(x+e_i))\leqslant 9$. For the remaining $O(1/n)$ of the edges, their endpoints are trivially mapped to distance at most n and so the average stretch of ϕ is $O(1)$, as required. □

The proof of Theorem 1.6 relies on the work of Håstad, Leighton and Newman [4]. In this work they show that if A is a random subset of $\{0,1\}^n$, where each element is added to S with probability $p>1/2$ independently, then with high probability there is an injective mapping from $\{0,1\}^{n-1}$ to A that is 3-Lipschitz. In the setting of Theorem 1.6 we have a random set of size exactly 2^{n-1}, which is more restricting, but we require the mapping to be $O(1)$-Lipschitz only on average. It turns out that the proof from [4] can also be adapted to our setting, and we do it in Section 5.

Motivation. The main motivation of this paper is to obtain a better understanding of mappings between subsets of the hypercube that have small average stretch. Specifically, Benjamini, Cohen and Shinkar [1] asked the following question.

Problem 1.8. *Is there $A\subset\{0,1\}^n$ of density $1/2$ such that for any bijection $f:\{0,1\}^{n-1}\to A$ it holds that* $\mathsf{avgStretch}(f)$ *is large?*

It is easy to see that if A is the set of all strings x such that $\sum x_i$ is even, then $\mathsf{avgStretch}(f)\geqslant 2$ for any bijection $f:\{0,1\}^{n-1}\to A$. However, we cannot prove the existence of a set A that requires $\mathsf{avgStretch}(f)$ to be larger than 2.1. Since we do not know how to approach this problem, we change the setting and ask the analogous question about the 'function version' of the problem.

1.1. Notation

Below we define the Boolean functions used in this paper. The function $\mathsf{Majority}:\{0,1\}^n\to\{0,1\}$ is defined as

$$\mathsf{Majority}(x)=\begin{cases}1 & \text{if } \sum_{i=1}^n x_i>n/2,\\ 0 & \text{otherwise.}\end{cases}$$

The function $\mathsf{Dictator}:\{0,1\}^n\to\{0,1\}$ is defined as $\mathsf{Dictator}(x)=x_1$, that is, its value is dictated by the first coordinate. The function $\mathsf{XOR}:\{0,1\}^n\to\{0,1\}$ is defined as $\mathsf{XOR}(x)=\sum_{i=1}^n x_i \pmod 2$.

A mapping $\phi:\{0,1\}^n\to\{0,1\}^n$ is said to be *C-Lipschitz* if for every $x,y\in\{0,1\}^n$ it holds that $\mathsf{dist}(\phi(x),\phi(y))\leqslant C\,\mathsf{dist}(x,y)$, where $\mathsf{dist}(\cdot,\cdot)$ denotes the Hamming distance between the strings.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

Note that in order to prove that a mapping ϕ is C-Lipschitz it is enough to show that for every edge of the hypercube $(x, x+e_i)$ it holds that $\mathsf{dist}(\phi(x), \phi(x+e_i)) \leqslant C$.

As a relaxation of the notion of being C-Lipschitz, define the average stretch of ϕ as

$$\mathsf{avgStretch}(\phi) = \mathbb{E}_{x,i}[\mathsf{dist}(\phi(x), \phi(x+e_i)].$$

Informally, this means that if $\mathsf{avgStretch}(\phi)$ is large then many edges of the hypercube are sent by ϕ far apart, while if it is small, then the endpoints of an average edge are mapped by ϕ close to each other.

2. A bijection from Majority to Dictator

In this section we prove Theorem 1.2. The proof is based on the idea from [1], which relies on a classical partition of the vertices of $\{0,1\}^n$ into symmetric chains, due to de Bruijn, Tengbergen and Kruyswijk [3], where a symmetric chain is a path $(c_k, c_{k+1}, \ldots, c_{n-k})$ in $\{0,1\}^n$, such that each c_i has Hamming weight i.

De Bruijn, Tengbergen and Kruyswijk [3] suggested a partition of $\{0,1\}^n$ into symmetric chains, which we will call the *BTK partition*. We describe the partition by specifying for each $x \in \{0,1\}^n$ the chain C_x that contains x. In order to describe C_x consider x as a sequence of parentheses, where each 1 is an opening parenthesis and each 0 is a closing parenthesis. For every maximal subsequence of legal parentheses mark the corresponding coordinates with hats, and replace all unmarked symbols with $\sqcup$. We denote the obtained string by $\mathsf{signature}(x)$. Define C_x to be the collection of all strings whose signature is equal to $\mathsf{signature}(x)$, that is, all strings y that agree with x in the marked coordinates of x, and so that the remaining coordinates of y are of the form $00\ldots01\ldots11$.

For example, the string $x = 0110100110$ corresponds to the sequence $')(()())(()'$. It is easy to check that the maximal legal subsequences of parentheses in x is that marked with hats in $)\hat{(}\hat{(}\hat{)}\hat{)}\hat{)}(\hat{(}\hat{)}$. Therefore, we have

$$\mathsf{signature}(x) = \sqcup\hat{1}\hat{1}\hat{0}\hat{1}\hat{0}\hat{0} \sqcup \hat{1}\hat{0},$$

and

$$C_x = \{\underline{0}110100\underline{0}10, \underline{0}110100\underline{1}10, \underline{1}110100\underline{1}10\},$$

where the underlines specify the coordinates where strings in C_x differ.

The key part of the proof is the following lemma. It is essentially proved in [1]. We provide the proof of the lemma in the Appendix for completeness.

Lemma 2.1. *Let $n \in \mathbb{N}$, and let $x, y \in \{0,1\}^n$ be such that $\mathsf{dist}(x,y) = 1$. Let*

$$C_x = (c_k, c_{k+1} \ldots, c_{n-k}), \quad C_y = (c'_{k'}, c_{k'+1} \ldots, c'_{n-k'})$$

be the chains of the BTK partition that contain x and y respectively. Then,

$$\mathsf{dist}(\mathsf{signature}(x), \mathsf{signature}(y)) \leqslant 3,$$

where $\mathsf{dist}(\cdot,\cdot)$ denotes the Hamming distance between two strings, that is, the number of coordinates where the two strings differ. In particular, this implies the following.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

(1) $|k - k'| \leqslant 1$.
(2) *If* $c_j \in C_x$ *and* $c'_{j'} \in C_y$ *for some* $j \in [k, n-k]$ *and* $j' \in [k', n-k']$, *then* $\mathsf{dist}(c_j, c'_{j'}) \leqslant |j - j'| + 6$.
(3) *For any* $c \in C_x$ *there exists* $c' \in C_y$ *such that* $\mathsf{dist}(c, c') \leqslant 7$.[2]

2.1. The mapping

Below we prove Theorem 1.2. In order to prove the theorem it will be convenient to partition the hypercube as follows. For each BTK chain C of the $(n-1)$-dimensional hypercube define $P_C = \{c \circ b \in \{0,1\}^n : c \in C, b \in \{0,1\}\}$. There is a clear one-to-one correspondence between the $(n-1)$-dimensional BTK chains and our partition of $\{0,1\}^n$. For example, the block P_C corresponding to the chain $C = \{00, 01, 11\}$ consists of the six elements $P_C = \{000, 001, 010, 011, 110, 111\}$.

Proof. We define the mapping ψ as follows. Let $n \in \mathbb{N}$ be an odd integer. For $x \in \{0,1\}^n$, write it as $x = x' \circ x_n$, where $x' \in \{0,1\}^{n-1}$ represents the first $n-1$ bits of x, and $x_n \in \{0,1\}$ is the last bit of x. Let $C = \{c_k, c_{k+1}, \ldots, c_{n-1-k}\}$ be a symmetric chain in the BTK partition that contains x', and let j be the index such that $x' = c_j$. Define

$$\psi(x = x' \circ x_n) \stackrel{\text{def}}{=} \begin{cases} 1 \circ c_{2j-(n-k)+x_n} & \text{if } |x| \geqslant (n+1)/2, \\ 0 \circ c_{(n+k)-2j-1-x_n} & \text{if } |x| \leqslant (n-1)/2. \end{cases} \tag{2.1}$$

In order to illustrate the mapping, let us consider as an example the case of $n = 3$ and the block P_C that corresponds to the chain $C = \{00, 01, 11\}$:

$$\begin{aligned} \psi(11 \circ 1) &= 1 \circ 11, \\ \psi(11 \circ 0) &= 1 \circ 01, \\ \psi(01 \circ 1) &= 1 \circ 00, \\ \text{---} \quad &\quad \text{-----} \\ \psi(01 \circ 0) &= 0 \circ 00, \\ \psi(00 \circ 1) &= 0 \circ 01, \\ \psi(00 \circ 0) &= 0 \circ 11. \end{aligned}$$

Note that ψ maps the upper half of $\{0,1\}^n$ to points whose first coordinate is 1, and maps the lower half of $\{0,1\}^n$ to points whose first coordinate is 0. It should be mentioned that the mapping ψ restricted to the upper half of $\{0,1\}^n$ is exactly the mapping used in [1], where it was shown that the restriction of ψ to the upper half of $\{0,1\}^n$ is a bi-Lipschitz bijection. Similarly, the restriction of ψ to the lower half of $\{0,1\}^n$ is also a bi-Lipschitz bijection, and so, as mentioned above, the main difficulty in this construction was to 'stitch' these two bijections so that the endpoints of the edges in the middle layer are mapped by ψ close to each other.

We will show that the mapping ψ is indeed an 11-Lipschitz bijection from Majority to Dictator. Note that by the triangle inequality it is enough to show that for all edges $(x, y = x + e_i) \in \{0,1\}^n$ it holds that $\mathsf{dist}(\psi(x), \psi(y)) \leqslant 11$.

[2] That is, the Hausdorff distance between C_x and C_y is at most 7.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

Write $x = x' \circ x_n$, where $x' \in \{0,1\}^{n-1}$ represents the first $n-1$ bits of x, and $x_n \in \{0,1\}$ is the last bit of x. Analogously write $y = y' \circ y_n$. Let

$$C_{x'} = \{c_k, c_{k+1} \ldots, c_{n-k}\}, \quad C_{y'} = \{c'_{k'}, c_{k'+1} \ldots, c'_{n-k'}\}$$

be the BTK chains that contain $x' = c_j$ and $y' = c'_{j'}$ respectively, where $j = |x'|$ and $j' = |y'|$. Recall that $|j - j'| \leqslant 1$ since $|x'| = j$ and $|y'| = j'$ and $x \sim y$.

Our goal is to show that $\mathsf{dist}(\psi(x), \psi(y)) \leqslant 11$. We now consider several cases.

Case 1. $|x| = (n-1)/2$ and $|y| = (n+1)/2$.
In this case we have

$$\psi(x) = 0 \circ c_{(n+k)-(n-1)-1-x_n} = 0 \circ c_{k-x_n}$$

and

$$\psi(y) = 1 \circ c_{n+1-(n-k')+y_n} = 1 \circ c'_{k'+1+y_n}.$$

By Lemma 2.1 we have $|k - k'| \leqslant 1$ and hence

$$\mathsf{dist}(\psi(x), \psi(y)) \leqslant 1 + |(k - x_n) - (k' + 1 + y_n)| + 6 \leqslant 11,$$

as required.

Case 2. $|x| \geqslant (n+1)/2$ and $|y| \geqslant (n+1)/2$.
In this case we have $|(2j - (n-k) + x_n) - (2j' - (n-k') + y_n)| \leqslant 5$, and so by Lemma 2.1 the distance between $\psi(x)$ and $\psi(y)$ is at most $\mathsf{dist}(\psi(x), \psi(y)) = \mathsf{dist}(c_{2j-n+k+x_n}, c'_{2j'-n+k'+y_n})$ $\leqslant 5 + 6 = 11$.

Case 3. $|x| \leqslant (n-1)/2$ and $|y| \leqslant (n-1)/2$.
This is handled similarly to Case 2.

This completes the proof of Theorem 1.2. □

3. A linear bijection from Dictator to XOR

In this section we prove Theorem 1.3.

Proof of Theorem 1.3. We give an explicit mapping from Dictator to XOR that is a linear transformation over $GF(2)$. Let A be the matrix representing the linear transformation. We first show that the mapping satisfies the conditions in the lemma if A has the following properties.

(1) A is invertible.
(2) The first column of A has odd weight, and all other columns have even weight.
(3) All rows of A have weight 3.
(4) All columns of A have weight 2.
(5) All columns of A^{-1} have weight $O(\log(n))$.

The first condition implies that A is a bijection. The second condition implies that A maps from Dictator to XOR. To see this, consider $\mathsf{XOR}(Av)$ for any vector $v \in \{0,1\}^n$.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

This is the sum over $GF(2)$ of the weights of all columns j for which $v_j = 1$. Because the weights of all columns but the first are 0, they can be ignored, and therefore $\mathsf{XOR}(Av) = 1$ if and only if $v_1 = 1$.

The third condition implies that each output bit is local, as the ith output bit depends only on the ith row of A. The fourth condition implies that A is 2-Lipschitz. To see this, note that $\mathsf{dist}(x, y)$ is the weight of $x - y$, and $\mathsf{dist}(Ax, Ay)$ is the weight of $A(x - y)$. If the weight of each column of A is at most C, then the weight of $A(x - y)$ is at most C times the weight of $(x - y)$. The same argument applied to A^{-1} implies that A^{-1} is $O(\log(n))$-Lipschitz.

Note that under the assumption that the mapping is a linear transformation, the above is necessary for a mapping to satisfy the conditions of the lemma.

We now construct the mapping A. Let G be the complete binary tree on n vertices, with directed edges so that each points to the child. We uniquely label each vertex with a label from 1 to n, with the root labelled 1. Let A be $I_n + M$, where M is the adjacency matrix of G. That is, $A_{i,j} = 1$ if i is the parent of j in G or $i = j$, and $A_{i,j} = 0$ otherwise.

We claim that the inverse of A is the matrix defined as $B_{i,j} = 1$ if j is a descendant of i, and $B_{i,j} = 0$ otherwise. Indeed, consider the (i, j)th entry of the product $A \cdot B$. Then

$$\begin{aligned}(A \cdot B)_{i,j} &= \sum_{k=1}^{n} A_{i,k} B_{k,j} \\ &= |\{k : i = k \text{ or } i \text{ is the parent of } k\} \cap \{k : k = j \text{ or } j \text{ is a descendant of } k\}|\end{aligned}$$

If $i = j$, these sets have exactly one element in common, namely i, and therefore $(A \cdot B)_{i,j} = 1$. If vertex j is not in the subtree rooted at vertex i, these sets have no vertices in common, and therefore $(A \cdot B)_{i,j} = 0$. Finally, if vertex j is in the subtree rooted at i but is not i, these sets have exactly two vertices in common, and therefore $(A \cdot B)_{i,j} = 0$ over $GF(2)$.

Because vertex 1 is the only vertex without a parent, the first column of A has weight 1. All other vertices have exactly one parent, and hence all other columns of A have weight 2. The weight of each row of A is at most 3 since this is equal to the number of children of the corresponding vertex plus 1. The ith column of A^{-1} is the indicator vector of the set of ancestors of i in G including i itself. The size of this set is bounded above by $\log(n) + 1$. Therefore, A satisfies the conditions of the lemma. □

Note that the above proof can be generalized to obtain a mapping ϕ that is L-local and 2-Lipschitz such that ϕ^{-1} is C-Lipschitz, for any L and C that satisfy $(L-1)^C \geqslant n$, by replacing the tree G in the proof with a complete $(L-1)$-ary tree on n vertices. Such a tree will have height less than C.

4. Any bijection from XOR to Majority has large average stretch

In this section we prove Theorem 1.5. We start with the second part of the theorem.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

Proposition 4.1. *There exists a C-Lipschitz bijection* $\psi : \{0,1\}^n \to \{0,1\}^n$ *from* Majority *to* XOR *for some absolute constant* C.

Proof. Take the C-Lipschitz bijection from Majority to Dictator from Theorem 1.2, and compose it with the 2-Lipschitz bijection $\phi(x_1, x_2, \ldots, x_n) = (x_1 + x_2, x_2 + x_3, \ldots, x_{n-1} + x_n, x_n)$ from Dictator to XOR. The resulting mapping is clearly a $2C$-Lipschitz bijection from Majority to XOR. □

Next we prove the first part of Theorem 1.5 showing that any bijection from XOR to Majority must have large average stretch. In fact we prove a stronger statement, saying that in every direction $i \in [n]$ it holds that the average stretch in the direction e_i must be $\Omega(\sqrt{n})$.

Proposition 4.2. *Let* ϕ *be a bijection from* XOR *to* Majority*, and let* $i \in [n]$. *Then*

$$\mathbb{E}_x[\mathsf{dist}(\phi(x), \phi(x + e_i)] \geqslant c\sqrt{n}$$

for some absolute constant $c > 0$.

Proof. Fix $i \in [n]$ and consider all edges in the direction e_i, that is, the edges of the form $\{(x, x + e_i)\}_{x \in \{0,1\}^n}$. Since $\mathsf{XOR}(x) \neq \mathsf{XOR}(x + e_i)$ for all $x \in \{0,1\}^n$, it follows that for every such edge one of its endpoints must be mapped to the upper half of the hypercube and the other endpoint to the bottom half of the hypercube. On the other hand, since each level of the hypercube contains at most $\binom{n}{n/2} < 2^n/\sqrt{n}$ vertices, it follows that for 0.9 of the points $z \in \{0,1\}^n$ their weight differs from $n/2$ by more than $0.01\sqrt{n}$. Let us call such z *typical*. Therefore, for 0.8 of inputs x it holds that at least one of the endpoints of the edge $(x, x + e_i)$ is mapped to a typical point. Let us say for concreteness that x is typical and is mapped above level $n/2 + 0.01\sqrt{n}$. Then, since $\phi(x) \neq \phi(x + e_i)$ it follows that $\phi(x + e_i)$ must belong to the lower half of the hypercube, and thus $\mathsf{dist}(\phi(x), \phi(x + e_i)) \geqslant \Omega(\sqrt{n})$. Therefore, for at least 0.8 of the edges in the direction e_i it holds that $\mathsf{dist}(\phi(x), \phi(x + e_i)) \geqslant 0.02\sqrt{n}$. □

This clearly implies Theorem 1.5.

5. Bijection from Dictator to a random function

In this section we prove Theorem 1.6. We start with the following claim.

Lemma 5.1. *Let* $A \subseteq \{0,1\}^n$ *be a random set chosen by picking each* $x \in \{0,1\}^n$ *to be in* A *independently with probability* 0.5. *Then with probability* $1 - 2^{-2^{\Omega(n)}}$ *there exists an injective mapping* $\phi_A : \{0,1\}^{n-1} \to \{0,1\}^n$ *such that the following holds.*

(1) *For all* $x \in \{0,1\}^{n-1}$ *it holds that* $\mathsf{dist}(x, \phi_A(x)_{[1,\ldots,n-1]}) \leqslant 1$, *where* $\phi_A(x)_{[1,\ldots,n-1]}$ *denotes the restriction of* $\phi_A(x)$ *to the first* $n-1$ *coordinates.*

(2) $\mathbb{P}_{x \in \{0,1\}^{n-1}}[\phi_A(x) \in A] = 1 - O(1/n)$.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

We postpone the proof of Lemma 5.1 and show how it implies Theorem 1.6.

Proof of Theorem 1.6. We start with the following simple claim, which is immediate from Lemma 5.1.

Claim 5.2. *Let $A_1 \subset \{0,1\}^n$ be a uniformly random subset of size 2^{n-1}. Then, with probability $1-2^{-2^{\Omega(n)}}$ there exists a bijection $\phi : \{0,1\}^{n-1} \to A_1$ such that for $1-O(1/n)$ of the inputs it holds that $\mathsf{dist}(x, \phi(x)_{[1,\ldots,n-1]}) \leqslant 1$.*

Proof. Let us sample a subset $A_1 \subset \{0,1\}^n$ of size exactly 2^{n-1} in the following manner. Pick a random subset $A \subseteq \{0,1\}^n$ by choosing each $x \in \{0,1\}^n$ to be in A independently with probability 0.5. Then, if $|A| < 2^{n-1}$ we add to A uniformly random elements from $\{0,1\}^n \setminus A$ one by one until the size of A becomes 2^{n-1}. Similarly, if $|A| > 2^{n-1}$ we remove random elements from A one by one until the size of A becomes 2^{n-1}. Let A_1 be the obtained set. Clearly, A_1 is a uniformly random subset of $\{0,1\}^n$ of size 2^{n-1}.

By Lemma 5.1, with probability $1-2^{-2^{\Omega(n)}}$ there is an injective mapping $\phi : \{0,1\}^{n-1} \to A$ such that $\mathsf{dist}(x, \phi(x)_{[1,\ldots,n-1]}) \leqslant 1$ for all $x \in \{0,1\}^{n-1}$, and for all but $e_1 = O(1/n)$ of the inputs it holds that $\phi(x) \in A$. By the Chernoff bound, with probability $1-2^{-2^{\Omega(n)}}$ we have $||A| - 2^{n-1}| < 2^{0.51n}$. Therefore, we can modify ϕ in at most $e_1 + 2^{0.51n}/2^n = O(1/n)$ of the inputs so that the obtained mapping is a bijection from $\{0,1\}^{n-1}$ to A_1 that satisfies the requirements of the claim. □

In order to prove Theorem 1.6 we sample a uniformly random balanced Boolean function $f : \{0,1\}^n \to \{0,1\}$ as follows. Pick $A_1 \subset \{0,1\}^n$ of size 2^{n-1} uniformly at random, and let $A_0 = \{0,1\}^n \setminus A_1$, Define f to be the indicator function of A_1, that is, $A_1 = f^{-1}(1)$ and $A_0 = f^{-1}(0)$.

Since A_1 is a uniformly random set of size 2^{n-1}, by Claim 5.2 with probability $1-2^{-2^{\Omega(n)}}$ there is a bijection $\phi_1 : \{x \in \{0,1\}^{n-1} : x_1 = 1\} \to A_1$ such that $\mathsf{dist}(x, \phi_1(x)) \leqslant 2$ for all but $O(1/n)$ of the domain of ϕ_1. Similarly, A_0 is also a uniformly random subset of $\{0,1\}^n$ of size 2^{n-1}, and hence with probability $1-2^{-2^{\Omega(n)}}$ there is a bijection $\phi_0 : \{x \in \{0,1\}^{n-1} : x_1 = 0\} \to A_0$ such that $\mathsf{dist}(x, \phi_0(x)) \leqslant 2$ for all but $O(1/n)$ of the inputs

By the union bound, with probability $1-2^{-2^{\Omega(n)}}$, both ϕ_0 and ϕ_1 exist. Since ϕ_0 and ϕ_1 are defined on disjoint domains, whose union is the entire hypercube, we can define $\phi : \{0,1\}^n \to \{0,1\}^n$ to be

$$\phi(x) = \begin{cases} \phi_1(x) & \text{if } x_1 = 1, \\ \phi_0(x) & \text{if } x_1 = 0. \end{cases}$$

Clearly ϕ is a bijection from $\mathsf{Dictator}$ to f and it satisfies

$$\mathbb{P}_{x \in \{0,1\}^n}[\mathsf{dist}(x, \phi(x)) \leqslant 2] = 1 - O(1/n), \tag{5.1}$$

as required. The 'in particular' part of Theorem 1.6 follows immediately from (5.1). Indeed, by (5.1) it follows that $1-O(1/n)$ of the edges $(x, x+e_i)$ satisfy

$$\mathsf{dist}(\phi(x), \phi(x+e_i)) \leqslant \mathsf{dist}(\phi(x), x) + \mathsf{dist}(x, x+e_i) + \mathsf{dist}(x+e_i, \phi(x+e_i)) \leqslant 5,$$

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

and therefore

$$\mathsf{avgStretch}(\phi) = \mathbb{E}_{x\in\{0,1\}^n, i\in n}[\mathsf{dist}(\phi(x), \phi(x+e_i)] \leqslant 5\cdot(1-O(1/n)) + n\cdot O(1/n) = O(1).$$

In order to see that $\mathsf{avgStretch}(\phi^{-1}) = O(1)$, note that ϕ is a bijection, and so by (5.1) we have

$$\mathbb{P}_{x\in\{0,1\}^n}[\mathsf{dist}(x, \phi^{-1}(x)) \leqslant 2] = 1 - O(1/n).$$

This completes the proof of Theorem 1.6. □

We now return to the proof of Lemma 5.1. The main idea of the proof is based on [4]. In particular, the idea of extending $x \in \{0,1\}^{n-1}$ by adding the nth coordinate and the terminology *rich*/*poor* are borrowed from [4].

Proof of Lemma 5.1. In order to describe the algorithm, let A be a random subset of $\{0,1\}^n$. For each $x \in \{0,1\}^{n-1}$, say that x is *rich* if both $x\circ 0$ and $x\circ 1$ belong to A, and say that x is *poor* if neither $x\circ 0$ nor $x\circ 1$ belong to A. If there were no poor vertices in $\{0,1\}^{n-1}$, then we could define ϕ_A by extending its input x to either $x\circ 0$ or $x\circ 1$. However, since the subset A is uniformly random, roughly $1/4$ of the vertices in $\{0,1\}^{n-1}$ will be poor, and we will match all but $O(1/n)$ of the poor vertices with a neighbouring rich vertex. Then, we will define a mapping ϕ_A in the following way:

(1) if x is neither rich nor poor, then define $\phi_A(x) = x\circ b$, where $b\in\{0,1\}$ is such that $x\circ b \in A$;
(2) if x is rich, then define $\phi_A(x) = x\circ 1$;
(3) if x is poor and is matched with a rich vertex y, then define $\phi_A(x) = y\circ 1$;
(4) otherwise x is poor and is not matched with a rich vertex, in which case we define $\phi_A(x) = x\circ 0$.

Clearly such a mapping ϕ_A satisfies the condition that $\mathsf{dist}(x, \phi_A(x)_{[1,\ldots,n-1]}) \leqslant 1$ for all $x\in\{0,1\}^{n-1}$. We will define the matching so that only $O(1/n)$ of the poor vertices will be of type (4), that is, will be poor and not matched to a neighbouring rich vertex, and hence only those vertices x will be such that $\phi_A(x)\notin A$.

The algorithm for finding such a matching is given in Algorithm 1.

Algorithm 1 Matching poor vertices to rich vertices

```
for i = 1..., n/2 do
    if x is poor and not matched and x + e_i is rich and not matched then
        Match x with x + e_i
    end if
end for
```

Remark. We remark that we could allow the loop to run until n. However for the analysis it will be more convenient to stop after $n/2$ steps.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

The following two claims from [4] are the key steps in the analysis of the algorithm above.

Claim 5.3 ([4, Lemma 1]). *For every $k \leqslant n/2$, the status of x in the kth iteration of the algorithm is independent of all vertices that differ from x in some coordinate larger than k.*

In particular, for any $z \in \{0,1\}^k$ let $A_z = \{x \in \{0,1\}^{n-1} : x_{[1,\ldots,k]} = z\}$. Then, in the kth iteration of the algorithm the status of each vertex in A_z is independent of the others.

Proof. At each iteration i the vertices that affect each other are matched according to the edges in the direction e_i. Therefore, any two vertices that differ in some coordinate larger than k had no interaction between them, and so are independent of each other. □

Claim 5.4. *For every $x \in \{0,1\}^{n-1}$ and $i = 1, \ldots, n/2$, let p_i be the probability that x is poor and unmatched after iteration i. Then*

(1) $p_{i+1} = p_i(1 - p_i)$ *for all* $i < n/2$,

(2) $p_{n/2} < 2/n$.

Proof. Let q_i be the probability that x is rich but unmatched after round i, analogous to p_i. In the $(i+1)$st round, x is poor and unmatched if x was poor and unmatched after the ith round, and $x + e_{i+1}$ is rich and unmatched after the ith round. By Claim 5.3, these two events are independent, and therefore $p_{i+1} = p_i(1 - q_i)$. Similarly, q_{i+1} can be expressed as $q_{i+1} = q_i(1 - p_i)$. Subtracting these two equations, we see that $q_{i+1} - p_{i+1} = q_i - p_i = q_0 - p_0$ for all i. This is natural as the difference between the number of rich unmatched vertices and poor unmatched vertices stays constant throughout the rounds. Because $p_0 = 1/4$ and $q_0 = 1/4$, this difference is 0 and therefore $q_i = p_i$ for all i. Substituting q_i in the expression for p_{i+1} yields $p_{i+1} = p_i(1 - p_i)$. This proves the first part of the claim.

To prove the second part of the claim, we show by induction that $p_i \leqslant 1/i$ for all $i \geqslant 1$. Indeed, since $p_0 = 1/4$, the claim holds for $i \leqslant 4$. For the induction step for $i \geqslant 4$, if $p_i < 1/i \leqslant 1/4$, then $p_{i+1} = p_i(1 - p_i) \leqslant (1/i)(1 - 1/i) < 1/(i+1)$, as required. □

We are now ready to complete the proof of Lemma 5.1. For each $z \in \{0,1\}^{n/2}$ consider the set

$$A_z = \{x \in \{0,1\}^{n-1} : x_{[1,\ldots,n/2]} = z\}.$$

By Claims 5.3 and 5.4, each $x \in A_z$ is poor and unmatched with probability $p_{n/2} < 2/n$ independently of all other vertices in A_z. Therefore, since $|A_z| = 2^{n/2}$, by the Chernoff bound the probability that A_z contains more than $4/n$ of the poor and unmatched vertices is at most $2^{-2^{\Omega(n)}}$. By taking the union bound over all $z \in \{0,1\}^{n/2}$, we conclude that with probability $1 - 2^{-2^{\Omega(n)}}$ there exists a matching that matches all but $O(1/n)$ of the proof vertices with a rich neighbouring vertex. This completes the proof of Lemma 5.1. □

6. Open problems

Below we list several open problems.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

Question 6.1. *In Theorem 1.3 we constructed a linear mapping from* Dictator *to* XOR *that is* 3*-local,* 2*-Lipschitz, such that its inverse is* $O(\log(n))$*-Lipschitz. Is there a mapping from* Dictator *to* XOR *that is* $O(1)$*-local and* $O(1)$*-bi-Lipschitz? In particular, it would be interesting to find such a mapping that is non-linear.*

Question 6.2. *We proved in Theorem 1.6 that, for a random balanced function* f*, with high probability there is a mapping* ϕ_f *from* Dictator *to* f *such that* $\mathsf{avgStretch}(\phi) = O(1)$ *and* $\mathsf{avgStretch}(\phi^{-1}) = O(1)$*. Is it true that with high probability there is a bi-Lipschitz mapping from* Dictator *to a random function, that is, a mapping with bounded worst case stretch?*

Question 6.3. *We proved in Theorem 1.5 that any mapping from* XOR *to* Majority *must have average stretch larger than* $\Omega(\sqrt{n})$*. Is this bound tight? Is there a mapping* ϕ *from* XOR *to* Majority *such that* $\mathsf{avgStretch}(\phi) = o(n)$*?*

In this paper we have only considered mappings between functions with the same domain. If we allow one of the functions to have a larger domain, we may relax the requirement that a mapping between function must be a bijection, and only require the mapping to be one-to-one. Given Theorem 1.5 we ask the following question.

Question 6.4. *Is there a Lipschitz embedding* $\phi : \{0,1\}^n \to \{0,1\}^{\mathrm{poly}(n)}$ *such that* $\mathsf{XOR}(z) = \mathsf{Majority}(\phi(z))$ *for all* $z \in \{0,1\}^n$*?*

Acknowledgement

We are grateful to the anonymous referee for carefully reading the paper and providing very helpful comments, which significantly improved the presentation of the results.

Appendix: Proof of Lemma 2.1

Proof. Fix $x, y \in \{0,1\}^n$ such that they differ only in the ith coordinate and $x_i = 0$ and $y_i = 1$. Perform the following steps on x and y.

(1) Mark all the subsequences of legal parentheses in the prefix of the string of length $i-1$.
(2) Mark all the subsequences of legal parentheses in the suffix of the string of length $n-i$.
(3) Ignore the marked coordinates, and mark all the subsequences of legal parentheses in the obtained string.

Since x and y agree on all but the ith coordinate, the running of the marking stage on x and y in steps (1) and (2) yield the same marking, and so $\mathsf{signature}(x)$ agrees with $\mathsf{signature}(y)$ in these coordinates. Therefore, from now on we may ignore the coordinates marked in the first two steps.

Next we analyse the difference between the markings after step (3). Let $s \in \{0,1,\hat{0},\hat{1}\}^{i-1}$ and $t \in \{0,1,\hat{0},\hat{1}\}^{n-i}$ denote the two partially marked strings such that the resulting strings

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

after step (2) on inputs x and y are $s \circ 0 \circ t$ and $s \circ 1 \circ t$ respectively. Let us suppose for concreteness that the string s contains a unmarked zeros and b unmarked ones, and the string t contains c unmarked zeros and d unmarked ones. Recall that at the end of the marking stage, all unmarked zeros are to the left of all unmarked ones in both s and t. Therefore, after ignoring the coordinates marked in the first two steps, we have

$$x = 0^a 1^b \circ 0 \circ 0^c 1^d \quad \text{and} \quad y = 0^a 1^b \circ 1 \circ 0^c 1^d.$$

At this point, it is fairly easy to be convinced that $\mathsf{dist}(\mathsf{signature}(x), \mathsf{signature}(y))$ is bounded by *some* constant. Proving that the constant is 3 is done by a somewhat tedious case analysis, according to the relations between a, b, c and d.

Claim A.1. *For every $a, b, c, d \in \mathbb{N}$, we have*

$$\mathsf{dist}(\mathsf{signature}(0^a 1^b \circ 0 \circ 0^c 1^d), \mathsf{signature}(0^a 1^b \circ 1 \circ 0^c 1^d)) \leqslant 3.$$

We postpone the proof of the claim for now, and move to the 'in particular' part. Let $U_x = \{i \in [n] : \mathsf{signature}(x)_i = \sqcup\}$ denote the unmarked coordinates of x. The first item follows from the observation that $|U_x| = n - 2k$ and $|U_y| = n - 2k'$. Therefore, if $\mathsf{dist}(\mathsf{signature}(x), \mathsf{signature}(y)) \leqslant 3$, then $2|k - k'| \leqslant |U_x \Delta U_y| \leqslant 3$, and since k and k' are integers it follows that $|k - k'| \leqslant 1$.

For the second item take $c_j \in C_x$ and $c'_{j'} \in C_y$. Then c_j and $c'_{j'}$ differ in at most three coordinates outside $U_x \cap U_y$, since x and y have almost the same signature except in at most three coordinates. Inside the set U_x, the string c_j is just a sequence of zeros followed by a sequence of ones such that its weight is j, and similarly $c'_{j'}$ restricted to U_y is a sequence of zeros followed by a sequence of ones such that its weight is j'. Hence, inside the set $U_x \cap U_y$ the strings c_j and $c'_{j'}$ differ in at most $|j - j'| + 3$ coordinates. Therefore, $\mathsf{dist}(c_j, c'_{j'}) \leqslant |j - j'| + 6$, which completes the proof of Lemma 2.1. □

We now return to the proof of Claim A.1.

Proof of Claim A.1. Let $x = 0^a 1^b \circ 0 \circ 0^c 1^d$ and $y = 0^a 1^b \circ 1 \circ 0^c 1^d$. Our goal is to show that $\mathsf{dist}(\mathsf{signature}(x), \mathsf{signature}(y)) \leqslant 3$. The proof uses the following case analysis.

Case 1. $b = c$.
In this case we have

$$x = 0^a \circ 1^b 0^b \circ 01^d \quad \text{and} \quad y = 0^a 1 \circ 1^b 0^b \circ 1^d.$$

Their signatures are

$$\mathsf{signature}(x) = \sqcup^a \circ 1^b 0^b \circ \sqcup^{d+1} \quad \text{and} \quad \mathsf{signature}(y) = \sqcup^{a+1} \circ 1^b 0^b \circ \sqcup^d.$$

It is easy to verify that in this case the distance

$$\mathsf{dist}(\mathsf{signature}(0^a 1^b \circ 0 \circ 0^c 1^d), \mathsf{signature}(0^a 1^b \circ 1 \circ 0^c 1^d)) \leqslant 3.$$

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press

Case 2. $b > c$.
In this case we have

$$x = 0^a 1^{b-c-1} \circ 1^{c+1} 0^{c+1} \circ 1^d \quad \text{and} \quad y = 0^a 1^{b+1-c} \circ 1^c 0^c \circ 1^d.$$

Their signatures are

$$\mathsf{signature}(x) = \sqcup^{a+b-c-1} \circ 1^{c+1} 0^{c+1} \circ \sqcup^d \quad \text{and} \quad \mathsf{signature}(y) = \sqcup^{a+b-c+1} \circ 1^c 0^c \circ \sqcup^d.$$

It is again easy to verify that

$$\mathsf{dist}(\mathsf{signature}(0^a 1^b \circ 0 \circ 0^c 1^d), \mathsf{signature}(0^a 1^b \circ 1 \circ 0^c 1^d)) \leqslant 3.$$

Case 3. $b < c$.
In this case we have

$$x = 0^a \circ 1^b 0^b \circ 0^{c-b+1} 1^d \quad \text{and} \quad y = 0^a \circ 1^{b+1} 0^{b+1} \circ 0^{c-b-1} 1^d.$$

Their signatures are

$$\mathsf{signature}(x) = \sqcup^a \circ 1^b 0^b \circ \sqcup^{d+c-b+1} \quad \text{and} \quad \mathsf{signature}(y) = \sqcup^a \circ 1^{b+1} 0^{b+1} \circ \sqcup^{d+c-b-1}$$

Here it is also easy to verify that

$$\mathsf{dist}(\mathsf{signature}(0^a 1^b \circ 0 \circ 0^c 1^d), \mathsf{signature}(0^a 1^b \circ 1 \circ 0^c 1^d)) \leqslant 3.$$

□

References

[1] Benjamini, I., Cohen, G. and Shinkar, I. (2016) Bi-Lipschitz bijection between the Boolean cube and the Hamming ball. *Israel J. Math.* **212** 677–703. Also in *FOCS 2014: IEEE 55th Annual Symposium on Foundations of Computer Science*.

[2] Bourgain, J. (2002) On the distributions of the Fourier spectrum of Boolean functions. *Israel J. Math.* **131** 269–276.

[3] de Bruijn, N. G., van Ebbenhorst Tengbergen, C. and Kruyswijk, D. (1951) On the set of divisors of a number. *Nieuw Arch. Wiskunde* (2) **23** 191–193.

[4] Håstad, J., Leighton, T. and Newman, M. (1987) Reconfiguring a hypercube in the presence of faults. In *STOC 1987: 19th Annual ACM Symposium on Theory of Computing*, pp. 274–284.

[5] Kindler, G. and O'Donnell, R. (2012) Gaussian noise sensitivity and Fourier tails. In *CCC 2012: 27th IEEE Conference on Computational Complexity*, IEEE Computer Society, pp. 137–147.

[6] Lovett, S. and Viola, E. (2012) Bounded-depth circuits cannot sample good codes. *Comput. Complexity* **21** 245–266.

[7] Mossel, E., O'Donnell, R. and Oleszkiewicz, K. (2010) Noise stability of functions with low influences: Invariance and optimality. *Ann. of Math.* (2) **171** 295–341.

[8] Viola, E. (2012) The complexity of distributions. *SIAM J. Comput.* **41** 191–218.

https://doi.org/10.1017/S0963548317000578 Published online by Cambridge University Press