On well-connected sets of strings

Peter Frankl[*] *János Pach*[†]

Abstract

Given n pairwise disjoint sets $X_1, \ldots, X_n$, we call the elements of $S = X_1 \times \ldots \times X_n$ *strings*. A nonempty set of strings $W \subseteq S$ is said to be *well-connected* if for every $v \in W$ and for every $i\,(1 \le i \le n)$, there is another element $v' \in W$ which differs from v only in its ith coordinate. We prove a conjecture of Yaokun Wu and Yanzhen Xiong by showing that every set of more than $\prod_{i=1}^{n} |X_i| - \prod_{i=1}^{n}(|X_i| - 1)$ strings has a well-connected subset. This bound is tight.

1 Introduction

Let $X_1, \ldots, X_n$ be pairwise disjoint sets with $|X_i| = d_i > 1$ for $1 \le i \le n$. Let

$$S = X_1 \times \ldots \times X_n = \{(x_1, \ldots, x_n) : x_i \in X_i \text{ for every } i \in [n]\}$$

be the set of *strings* $x = (x_1, \ldots, x_n)$, where x_i is called the ith coordinate of x and $[n] = \{1, \ldots, n\}$.

A subset $W \subseteq S$ is called *well-connected* if for every $x \in W$ and for every $i \in [n]$, there is another element $x' \in W$ which differs from x only in its ith coordinate. That is, $x'_j \neq x_j$ if and only if $j = i$.

The following statement was conjectured by Yaokun Wu and Yanzhen Xiong [4].

Theorem 1. *Let T be a subset of $S = X_1 \times \ldots \times X_n$ with $|X_i| = d_i > 1$ for every $i \in [n]$. If*

$$|T| > \prod_{i=1}^{n} d_i - \prod_{i=1}^{n}(d_i - 1),$$

then T has a well-connected subset. This bound cannot be improved.

To see the tightness of the theorem, fix an element y_i in each X_i and let $X'_i = X_i \setminus \{y_i\}$. We claim that the set of strings

$$T_0 = (X_1 \times \ldots \times X_n) \setminus (X'_1 \times \ldots \times X'_n) \tag{1}$$

[*]Rényi Institute, P.O.Box 127 Budapest, 1364 Hungary and MIPT, Moscow, partially supported by the Ministry of Education and Science of the Russian Federation in the framework of MegaGrant 075-15-2019-1926. Email: peter.frankl@gmail.com.

[†]Rényi Institute, P.O.Box 127 Budapest, 1364 Hungary and MIPT, Moscow, partially supported by ERC Advanced Grant GeoScape and by the Ministry of Education and Science of the Russian Federation in the framework of MegaGrant 075-15-2019-1926. Email: pach@cims.nyu.edu

does not have any nonempty well-connected subset. Suppose for contradiction that there is such a subset $W \subseteq T_0$, and let $x = (x_1, \ldots, x_n)$ be an element of W with the minimum number of coordinates i for which $x_i = y_i$ holds. Obviously, this minimum is positive, otherwise $x \notin T_0$. Pick an integer k with $x_k = y_k$. Using the assumption that W is well-connected, we obtain that there exists $x' \in W$ that differs from x only in its kth coordinate. However, then x' would have one fewer coordinates with $x_i = y_i$ than x does, contradicting the minimality of x.

In the next section, we establish a result somewhat stronger than Theorem 1: we prove that under the conditions of Theorem 1, T also has a subset W such that for every $x \in W$ and $i \in [1, n]$, the number of elements $x' \in W$ which differ from x only in its ith coordinate is *odd* (see Theorem 6). In Section 3, we present a self-contained argument which proves this stronger statement.

Shortly after learning about our proof of the conjecture of Wu and Xiong, another proof was found by Chengyang Qian.

2 Exact sequence of maps

In this section, we introduce the necessary definitions and terminology, and we apply a basic topological property of simplicial complexes to establish Theorem 1.

For every $k\,(0 \le k \le n)$, let

$$S_k = \{A \subseteq X_1 \cup \ldots \cup X_n : |A| = k \text{ and } |A \cap X_i| \le 1 \text{ for every } i\}.$$

Clearly, we have $|S_n| = |S| = \prod_{i=1}^{n} |X_i|$. With a slight abuse of notation, we identify S_n with S.

Assign to each $A \in S_k$ a different symbol v_A, and define V_k as the family of all formal sums of these symbols with coefficients 0 or 1. Then

$$V_k = \{\sum_{A \in S_k} \lambda_A v_A : \lambda_A = 0 \text{ or } 1\}$$

can be regarded as a vector space over GF(2) whose dimension satisfies

$$\dim V_k = |S_k| = \sum_{1 \le j_1 < j_2 < \ldots < j_k \le n} d_{j_1} d_{j_2} \cdot \ldots \cdot d_{j_k}. \tag{2}$$

We use the standard definition of the *boundary operations* ∂_k.

Definition 2. Let $\partial_0 : V_0 \to 0$. For every $k \in [1, n]$ and every $A \in S_k$, let

$$\partial_k(v_A) = \sum_{\substack{B \subset A \\ |B| = k-1}} v_B.$$

Extend this map to a homomorphism $\partial_k : V_k \to V_{k-1}$ by setting

$$\partial_k(\sum_{A \in S_k} \lambda_A v_A) = \sum_{A \in S_k} \lambda_A \partial_k(v_A),$$

where the sum is taken over GF(2).

Let $\ker(\partial_k) \subseteq V_k$ and $\mathrm{im}(\partial_k) \subseteq V_{k-1}$ denote the *kernel* and the *image* of ∂_k, respectively.

Our proof is based on the following lemma.

Lemma 3. *The sequence of homomorphisms* $V_n \xrightarrow{\partial_n} V_{n-1} \xrightarrow{\partial_{n-1}} \dots \xrightarrow{\partial_1} V_0 \xrightarrow{\partial_0} 0$ *is an* exact sequence*, i.e.,* $\mathrm{im}(\partial_k) = \ker(\partial_{k-1})$ *holds for every* k*.*

Proof. First, we show that $\mathrm{im}(\partial_k) \subseteq \ker(\partial_{k-1})$ for every $k \in [1, n]$. The statement is obviously true for $k = 1$. If $k \geq 2$, then for every $A \in S_k$, we have

$$\partial_{k-1}\partial_k v_A = \sum_{\substack{B \subset A \\ |B|=k-1}} \sum_{\substack{C \subset B \\ |C|=k-2}} v_C = \sum_{\substack{C \subset A \\ |C|=k-2}} 2v_C = 0.$$

Thus, $\partial_{k-1}\partial_k(v) = 0$ for every $v \in V_k$, as claimed.

On the other hand, $\cup_{i=0}^n V_i$ can be regarded as an $(n-1)$-dimensional triangulated topological space which is the join of discrete topological spaces (sets), $X_1, \dots, X_n$, each of size at least 2 and, hence, -1-connected. Proposition 4.4.3 in [3] (see also [1]) states that if X is a-connected and Y is b-connected, then their join is $(a + b + 2)$-connected. By repeated application of this statement, we obtain that $\cup_{i=0}^n V_i$, as a topological space, is $(n-2)$-connected. This means that, up to dimension $n - 2$ (that is, up to V_{n-1}) their homotopy groups are trivial. This implies that the corresponding homology groups (the quotient groups $\ker(\partial_{k-1})/\mathrm{im}(\partial_k)$) are also trivial. □

Corollary 4. *For every* $k \in [1, n]$*, we have* $\dim \ker(\partial_k) = \sum_{i=0}^k (-1)^{k-i} \dim V_i$.

Proof. By induction on k. According to the Rank Nullity Theorem [2], we have

$$\dim V_i = \dim \ker(\partial_i) + \dim \mathrm{im}(\partial_i), \tag{3}$$

for every $i \in [1, n]$. Since $\dim \mathrm{im}(\partial_1) = \dim V_0 = 1$, the corollary is true for $k = 1$.

Assume we have already verified it for some $k < n$. To show that it is also true for $k+1$, we use that $\dim \mathrm{im}(\partial_{k+1}) = \dim \ker(\partial_k)$, by Lemma 3. Plugging this into (3) with $i = k + 1$, we obtain

$$\dim V_{k+1} = \dim \ker(\partial_{k+1}) + \dim \ker(\partial_k).$$

Hence, using the induction hypothesis, we have

$$\begin{aligned} \dim \ker(\partial_{k+1}) &= \dim V_{k+1} - \dim \ker(\partial_k) \\ &= \dim V_{k+1} - \sum_{i=0}^{k} (-1)^{k-i} \dim V_i = \sum_{i=0}^{k+1} (-1)^{k+1-i} \dim V_i, \end{aligned}$$

as required. □

By (2), we know the value of $\dim V_i$ for every i. Therefore, Corollary 4 enables us to compute $\dim \ker(\partial_n)$ and, hence, $\dim V_n - \dim \ker(\partial_n)$.

Corollary 5. *We have*

$$\dim V_n - \dim \ker(\partial_n) = \prod_{i=1}^{n} d_i - \prod_{i=1}^{n} (d_i - 1).$$

Proof. From Corollary 4, we get

$$\dim V_n - \dim \ker(\partial_n) = \sum_{i=0}^{n-1} (-1)^{n-1-i} \dim V_i.$$

Using (2) and the fact that $\dim V_0 = 1$, this is further equal to

$$\sum_{i=1}^{n-1} (-1)^{n-1-i} \sum_{1 \le j_1 < j_2 < \ldots < j_i \le n} d_{j_1} d_{j_2} \cdots d_{j_i} + (-1)^{n-1} = \prod_{i=1}^{n} d_i - \prod_{i=1}^{n} (d_i - 1). \qquad \square$$

Now we are in a position to establish the following statement, which is somewhat stronger than Theorem 1.

Theorem 6. *Let T be a subset of $S = X_1 \times \ldots \times X_n$ with $|X_i| = d_i > 1$ for every $i \in [n]$. If*

$$|T| > \prod_{i=1}^{n} d_i - \prod_{i=1}^{n} (d_i - 1),$$

then there is a subset $W \subseteq T$ with the property that for every $x \in W$ and $i \in [1, n]$, the number of elements $x' \in W$ which differ from x only in their ith coordinate is odd. *This bound cannot be improved.*

Proof. The tightness of the bound follows from the tightness of Theorem 1 shown at the end of the Introduction.

Let T be a system of strings of length n satisfying the conditions of the theorem. Using the notation introduced at the beginning of this section, let

$$V(T) = \{ \sum_{A \in T} \lambda_A v_A : \lambda_A = 0 \text{ or } 1 \}.$$

Then $V(T)$ can be regarded as a linear subspace of V_n with $\dim V(T) = |T|$. Comparing the size of T with the value of $\dim V_n - \dim \ker(\partial_n)$ given by Corollary 5, we obtain that there is a nonzero vector $v = \sum_{A \in T} \lambda_A v_A$ that belongs to $V(T) \cap \ker(\partial_n)$. Let $W = \{A \in T : \lambda_A = 1\}$. Then we have

$$0 = \partial_n(v) = \sum_{A \in W} \partial_n(v_A) = \sum_{A \in W} \sum_{\substack{B \subset A \\ |B| = n-1}} v_B = \sum_{\substack{B \subset [1,n] \\ |B| = n-1}} |\{A \in W : A \supseteq B\}| v_B.$$

Thus, for each B, the coefficient of v_B is *even*. This means that the set of strings $W \subset T$ meets the requirements of the theorem. $\square$

3 Direct proof of Theorem 6

In this section, we prove Theorem 6 directly, without using Lemma 5 and, hence, Corollary 5.

As in the Introduction, fix an element $y_i \in X_i$ and let $X_i' = X_i \setminus \{y_i\}$, for every $i \in [1, n]$. Defining T_0 as in (1), we have that $|T_0| = \prod_{i=1}^n d_i - \prod_{i=1}^n (d_i - 1)$.

Suppose that $|T| > |T_0|$. As in the first proof of Theorem 6, it is sufficient to show that there exists a nonzero vector $v = \sum_{A \in T} \lambda_A v_A$ with suitable coefficients $\lambda_A \in \{0, 1\}$ such that $v \in \ker(\partial_n)$, *i.e.*, we have $\partial_n v = \sum_{A \in T} \lambda_A (\partial_n v_A) = 0$. Thus, it is enough to establish the following statement.

Lemma 7. *Let T be a subset of $S = X_1 \times \ldots \times X_n$ with $|X_i| > 1$ for every $i \in [1, n]$.*

If $|T| > |T_0|$, then the set of vectors $\{\partial_n v_A : A \in T\}$ is linearly dependent over GF(2).

Proof. First, we show that the set of vectors $\{\partial_n v_A : A \in T_0\}$ is linearly independent. Suppose, for a contradiction, that there is a nonempty subset $W \subset T_0$ such that $\sum_{A \in W} \partial_n v_A = 0$. Pick an element $A = \{x_1, \ldots, x_n\}$ of W for which the number of coordinates i with $x_i = y_i$ is as small as possible. By the definition of T_0, there is at least one such coordinate $x_k = y_k$. In view of Definition 2, one of the terms of the formal sum $\partial_n v_A$ is v_B with $B = A \setminus \{y_k\}$, and this term cannot be canceled out by a term of $\partial_n v_{A'}$ for any other $A' \in W$, because in this case A' would have fewer coordinates that are equal to some y_i than A does. Hence, $\sum_{A \in W} \partial_n v_A \neq 0$, contradicting our assumption.

Next, we prove that there exists no set of strings $T \supset T_0$ with $|T| > |T_0|$ such that the set of vectors $\{\partial_n v_A : A \in T\}$ is linearly independent.

To see this, consider any string $C = \{z_1, \ldots, z_n\} \in S \setminus T_0$. Since $C \notin T_0$, we have $z_i \neq y_i$ for every i. Define $T(C)$ as the set of all strings $A = \{x_1, \ldots, x_n\} \in S$ whose every coordinate x_i is either y_i or z_i. Then we have $\sum_{A \in T(C)} \partial_n v_A = 0$. As we have $T(C) \subseteq T_0 \cup \{C\}$, this means that the set of vectors $\{\partial_n v_A : A \in T_0 \cup \{C\}\}$ is linearly dependent over GF(2).

Now we can prove the lemma in its full generality. Suppose that there is at least one set of strings T with $|T| > |T_0|$ contradicting Lemma 7. Choose a counterexample T for which $|T \cap T_0|$ is as large as possible. According to the previous paragraph, T cannot fully contain T_0. Pick a string $C \in T_0 \setminus T$. Then $T \cup \{C\}$ is no longer a counterexample, so there is a nonempty subset $W \subset T \cup \{C\}$ such that $\sum_{A \in W} \partial_n v_A = 0$. (Obviously, $C \in W$.) We claim that the choice of W is *unique*. In other words, there exists no other subset $W^* \subset T \cup \{C\}$ having the same property. Indeed, otherwise we would have

$$\sum_{A \in W \triangle W^*} \partial_n v_A = 0,$$

where $W \triangle W^*$ denotes the symmetric difference of W and W^*. Since $W \triangle W^* \subset T$, this would contradict our assumption that $\{\partial_n v_A : A \in T\}$ is a set of linearly independent vectors over GF(2).

Obviously, W must have at least one element B that does not belong to T_0. Define a new set of strings $T^* = (T \setminus \{B\}) \cup \{C\}$. We have $|T^* \cap T_0| > |T \cap T_0|$ and, because of the maximality of $|T \cap T_0|$, we know that T^* is not a counterexample to the lemma, *i.e.,* T^* is linearly dependent. Therefore, there exists a nonempty subset $W^* \subseteq T^* \subset T \cup \{C\}$ such that $\sum_{A \in W^*} \partial_n v_A = 0$. The sets W and W^* must be distinct, because $C \in W$, but $C \notin W^*$. This would contradict the uniqueness of W, completing the proof of the lemma and, hence, of Theorem 6. □

References

[1] A. Hatcher: *Algebraic Topology,* Cambridge University Press, Cambridge, 2001.

[2] P. D. Lax: *Linear Algebra and Its Applications (2nd edition),* Wiley and Sons, New York, 2013.

[3] J. Matoušek: *Using the Borsuk-Ulam Theorem,* Springer, Berlin, 2003.

[4] Y. Wu and Y. Xiong: Sparse $(0, 1)$ arrays and tree-like partition systems, submitted, 2020.