On well-connected sets of strings

Peter Frankl*  János Pach†
Rényi Institute
POB 127 Budapest, 1364 Hungary
peter.frankl@gmail.com  pach@cims.nyu.edu

Submitted: Mar 8, 2021; Accepted: Feb 7, 2022; Published: Mar 25, 2022
© The authors. Released under the CC BY-ND license (International 4.0).

Abstract

Given \(n\) sets \(X_1, \ldots, X_n\), we call the elements of \(S = X_1 \times \cdots \times X_n\) strings. A nonempty set of strings \(W \subseteq S\) is said to be well-connected if for every \(v \in W\) and for every \(i (1 \leq i \leq n)\), there is another element \(v' \in W\) which differs from \(v\) only in its \(i\)th 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.

Mathematics Subject Classifications: 05C88, 05C89

1 Introduction

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

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

be the set of strings \(x = (x_1, \ldots, x_n)\), where \(x_i\) is called the \(i\)th 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 \(i\)th 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].

---

*Partially supported by the Ministry of Education and Science of the Russian Federation in the framework of MegaGrant 075-15-2019-1926.
†Partially supported by ERC Advanced Grant GeoScape, NKFIH (Hungarian National Research, Development and Innovation Office) grant K-131529, and by the Ministry of Education and Science of the Russian Federation in the framework of MegaGrant 075-15-2019-1926.
Theorem 1. Let $T$ be a subset of $S = X_1 \times \cdots \times X_n$ 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 nonempty 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

$$T_0 = (X_1 \times \cdots \times X_n) \setminus (X_1' \times \cdots \times X_n')$$

(1)
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 $x \not\in T_0$. 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 $k$th 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 $i \in [n]$, the number of elements $x' \in W$ which differ from $x$ only in its $i$th 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. We will assume throughout, without loss of generality, that the sets $X_i$ are pairwise disjoint.

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

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

Clearly, we have $|S_n| = |S| = \prod_{i=1}^{n} |X_i|$. With a slight abuse of notation, we identify $S_n$ with $S$. The set system $\cup_{k=0}^{n} S_k$ is an abstract simplicial complex, that is, for each of its elements $A$, every subset of $A$ also belongs to $\cup_{k=0}^{n} S_k$. This simplicial complex has a geometric realization in $\mathbb{R}^{2n-1}$, where every element $A$ is represented by an $(|A| - 1)$-dimensional simplex. (See [1], part II, Section 9 or [3], Section 1.5. Note that not all textbooks consider the empty set a $-1$-dimensional simplex, but we do.)

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 $\text{GF}(2)$ whose dimension satisfies
\begin{equation}
\dim V_k = |S_k| = \sum_{1 \leq j_1 < j_2 < \ldots < j_k \leq n} d_{j_1} d_{j_2} \cdots d_{j_k}. \tag{2}
\end{equation}

We use the standard definition of the boundary operations $\partial_k$. (See [2], Section 2.1.) Informally, the boundary of each $(k-1)$-dimensional simplex that corresponds to a member $A \in S_k$ consists of all $(k-2)$-dimensional simplices corresponding to $(k-1)$-element subsets $B \subset A$. This definition naturally extends to any collection (“chain”) of $(k-1)$-dimensional simplices that correspond to members of $S_k$, with multiplicities taken modulo 2.

**Definition 2.** Let $\partial_0 : V_0 \rightarrow 0$. For every $k \in [n]$ and every $A \in S_k$, let
\[ \partial_k(v_A) = \sum_{B \subset A \atop |B|=k-1} v_B. \]

Extend this map to a homomorphism $\partial_k : V_k \rightarrow V_{k-1}$ 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 $\text{GF}(2)$.

Let $\ker(\partial_k) \subseteq V_k$ and $\text{im}(\partial_k) \subseteq V_{k-1}$ denote the kernel and the image of $\partial_k$, respectively. Our proof is based on the following lemma.

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

**Proof.** Before proving the statement, we show that $\text{im}(\partial_k) \subseteq \ker(\partial_{k-1})$ for every $k \in [n]$. The statement is obviously true for $k = 1$. If $k \geq 2$, then for every $A \in S_k$, we have
\[ \partial_{k-1}\partial_k v_A = \sum_{B \subset A \atop |B|=k-1} \sum_{C \subset B \atop |C|=k-2} v_C = \sum_{C \subset A \atop |C|=k-2} 2v_C = 0. \]
Thus, $\partial_{k-1}\partial_k(v) = 0$ for every $v \in V_k$, as claimed. In fact, the containment $\text{im}(\partial_k) \subseteq \ker(\partial_{k-1})$ holds for every simplicial complex.

We prove that in our case, all the above containments hold with equality. For every $i \in [n]$, let $K_i$ denote the 0-dimensional abstract simplicial complex consisting of the 1-element subsets of $X_i$ and the empty set. Consider now their join $K = K_1 \ast \ldots \ast K_n$; see [2], Chapter 0. By definition, $K$ is the same as the simplicial complex $\cup_{i=0}^{n} S_i$.

Let $j \geq -1$ be an integer. We need three well-known properties of the notion of $j$-connectedness of complexes; see Proposition 4.4.3 in [3].

(i) A complex is $-1$-connected if and only if it contains a nonempty simplex.
(ii) If $K_1$ is $a$-connected and $K_2$ is $b$-connected, then their join $K_1 \ast K_2$ is $(a + b - 2)$-connected.

(iii) If a complex is $j$-connected, then $\text{im}(\partial_k) = \text{ker}(\partial_{k-1})$ holds for every $k$, $1 \leq k \leq j+2$.

In our case, each $X_i$ is nonempty, hence, by property (i), each $K_i$ is $-1$-connected. By repeated application of (ii), we obtain that $K = K_1 \ast \ldots \ast K_n$ is $(n - 2)$-connected. In view of (iii), this implies that $\text{im}(\partial_k) = \text{ker}(\partial_{k-1})$ for every $k \in [n]$, as required. □

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

Proof. By induction on $k$. According to the Rank Nullity Theorem, we have

$$\dim V_i = \dim \text{ker}(\partial_i) + \dim \text{im}(\partial_i),$$

(3)

for every $i \leq n$. Since $\dim V_0 = 1$ and $\dim \text{im}(\partial_0) = 0 = 0$, the corollary is true for $k = 0$.

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

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

Hence, using the induction hypothesis, we have

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

$$= \dim V_{k+1} - \sum_{i=0}^{k} (-1)^{k-i} \dim V_i = \sum_{i=0}^{k+1} (-1)^{k+1-i} \dim V_i,$$

as required. □

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

Corollary 5. We have

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

Proof. From Corollary 4, we get

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

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

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

□
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 $S = X_1 \times \cdots \times X_n$ 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 nonempty subset $W \subseteq T$ with the property that for every $x \in W$ and $i \in [n]$, the number of elements $x' \in W$ which differ from $x$ only in their $i$th 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

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

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 $V_n - \text{dim ker}(\partial_n)$ 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

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

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. □

### 3 Direct proof of Theorem 6

In this section, we prove Corollary 5 and, hence, Theorem 6 directly, without using Lemma 3.

As in the Introduction, fix an element $y_i \in X_i$ and let $X_i' = X_i \setminus \{ y_i \}$, for every $i \in [n]$. 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 $|T| > |T_0|$. To prove Corollary 5, 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 $S = X_1 \times \cdots \times X_n$ with $|X_i| > 1$ for every $i \in [n]$. 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.

It remains to prove that $\{\partial_n v_A : A \in T_0\}$ is a base of $\text{im}(\partial_n)$, that is, 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 over $\text{GF}(2)$.

To see this, consider any string $C = \{z_1, \ldots, z_n\} \in S \setminus T_0$. Since $C \not\in T_0$, 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 $\text{GF}(2)$. This completes the proof of the lemma and, hence, of Theorem 6. □

Acknowledgements

We thank Gábor Tardos and an anonymous referee for several helpful suggestions.

References

[1] A. Björner. Topological Methods. Handbook of Combinatorics, R. Graham, M. Grötschel, L. Lovász (eds.), Vol. 2, 1995, 1819–1872.
[2] A. Hatcher. Algebraic Topology. Cambridge University Press, Cambridge, 2001.
[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. Manuscript, 2020.