Kees Roos
Delft Institute of Applied Mathematics
TU Delft
c.roos@tudelft.nl


Kees Roos

Research

On the Union-Closed Conjecture of Péter Frankl

In the March issue of NAW the editor-in-chief (Robbert Fokkink) recalled that the well-known Union-Closed Conjecture (UCC) for families of sets is still an open problem. Two months earlier there was an article on the UCC in the scientific supplement of the *NRC* with the title 'Het probleem van de kleurige knikkers lijkt bijna opgelost'. The reason for this article was the renewed interest in the UCC in 2022. In this paper Kees Roos, professor emeritus of TU Delft, describes the conjecture and some observations related to the UCC, especially for the cases where the UCC is tight. A nice byproduct is a Sierpinski-type fractal. A clever use of the structure of this fractal might eventually path the way to the solution of the UCC.

Frankl's conjecture

Let $\mathcal{F}$ be a family of distinct subsets of a finite set X such that if $S \in \mathcal{F}$ and $T \in \mathcal{F}$ then $S \cup T \in \mathcal{F}$. We then say that $\mathcal{F}$ is *union-closed* or shortly that $\mathcal{F}$ is a *UC*-family. In this note we deal with the following conjecture, which is attributed to Péter Frankl (1979).

Conjecture 1. *There exists an element* $x \in X$ *that belongs to at least half of the sets in* $\mathcal{F}$.

For a quite complete historical survey, see [2]. Despite many attempts the UCC is still open. There exist many partial results. One of these is that the UCC holds if $|\mathcal{F}| \geq 2^{n-1}$, where $n = |X|$ [6]. More recent papers are [4, 5, 8, 9, 10]. Following many authors we call an element $x \in X$ as in the conjecture an *abundant* element.

Remark 1. According to [5] Conjecture 1 exists since 1979. But I could find no paper of Frankl from that year. The first result of him related to the subject seems to be [4, Conjecture 2.1]. It deals with intersection closed (or IC-) families $\mathcal{F}'$. It states that in that case there exists an element $x \in X$ that belongs to at most half of the sets in $\mathcal{F}'$. By taking for $\mathcal{F}'$ the complements of the sets in $\mathcal{F}$ the two conjectures turn out to be equivalent. This is a consequence of the well-known theorem of de Morgan: the complement of the union of two sets is equal to the intersection of their complements. The relation between UC- and IC-families will become more clear in the last section.

Without loss of generality we assume that the elements of X are represented by the numbers 1 to n. In other words, $X = \{1, \ldots, n\}$, which is also denoted as $[n]$, and $\mathcal{F} \subseteq 2^X$, where 2^X denotes the so-called power set of X, which consists of all subsets of X. We represent each subset S of X by its ordered support vector, i.e., the $\{0,1\}$-vector

$$s = (s_n, s_{n-1}, \ldots, s_1), \tag{1}$$

with $s_i = 1$ if $i \in S$ and $s_i = 0$ otherwise. The matrix whose rows are the support vectors of the sets in $\mathcal{F}$ is denoted as $A_{\mathcal{F}}$. In this note we focus on this matrix representation of $\mathcal{F}$. It is clear that there exists an element x that belongs to at least half of the sets in $\mathcal{F}$ if and only if $A_{\mathcal{F}}$ has a column with at least $m/2$ ones, where $m = |\mathcal{F}|$. We can therefore reformulate the UCC as follows:

Conjecture 2. *If* $\mathcal{F}$ *is a UC-family then* $A_{\mathcal{F}}$ *has a column with at least* $m/2$ *ones.*

Note that $A_{\mathcal{F}}$ is unique up to the ordering of its rows. Without loss of generality one may assume that the rows are ordered lexicographically, but (for the moment) this is not required. Also without loss of generality we assume that $A_{\mathcal{F}}$ has no zero column, because such a column does not affect the status of the conjecture. It means that every element of $[n]$ belongs to at least one set in $\mathcal{F}$. As a consequence we always have $[n] \in \mathcal{F}$. We denote the empty set as $[0]$ and as usual, we allow $[0] \in \mathcal{F}$ in the conjecture only if $m \geq 2$, because otherwise (i.e. if $\mathcal{F} = \{[0]\}$) the conjecture does not hold.

Subsets as nonnegative integers

The binary vector s in (1) uniquely defines the nonnegative integer

$$z(s) = \sum_{s_i = 1} 2^{i-1}.$$

The extreme values of this number are 0 (for the empty set $[0]$) and $2^{n-1} + \cdots + 2^0 = 2^n - 1$ (for the whole set $[n]$). We call s the binary representation, or binary value, of the set S and $z(s)$ its decimal value. In this way we have two ways for representing the family $\mathcal{F}$ of subsets of $[n]$. Using binary values we obtain the binary matrix $A_{\mathcal{F}}$ of size $m \times n$. Otherwise, when using decimal values, $\mathcal{F}$ simply becomes a subset of m (distinct) integers in the interval $[0,2^n-1]$. In the sequel we feel free to denote $z(s)$ also as $z(S)$.

Before proceeding, we consider a small example. With $n=5$, let $\mathcal{F}$ be determined by the subset $\{7,14,21\}$ of integers in the interval $[0,31]$. We then have

$$\mathcal{F}=\{7,14,21\}, \quad A_{\mathcal{F}}=\begin{bmatrix}0&0&1&1&1\\0&1&1&1&0\\1&0&1&0&1\end{bmatrix}.$$

In other words, the subsets in $\mathcal{F}$ are $U=\{3,2,1\}$, $V=\{4,3,2\}$ and $W=\{5,3,1\}$. Obviously, $\mathcal{F}$ is not union-closed. For example, the set $U\cup V=\{4,3,2,1\}$ does not belong to $\mathcal{F}$. Denoting the binary representations of U and V as u and v, respectively, the binary representation of $U\cup V$ is simply

$$\max(u,v)=\begin{bmatrix}0&1&1&1&1\end{bmatrix}.$$

So $z(U\cup V)=15$, whence the decimal value 15 must be added to make $\mathcal{F}$ a UC-family. In a similar way we obtain $z(U\cup W)=23$ and $z(V\cup W)=31$, which forces us to also add the decimal values 23 and 31 to $\mathcal{F}$. Thus we arrive at

$$\mathcal{F}=\{7,14,15,21,23,31\},$$
$$A_{\mathcal{F}}=\begin{bmatrix}0&0&1&1&1\\0&1&1&1&0\\0&1&1&1&1\\1&0&1&0&1\\1&0&1&1&1\\1&1&1&1&1\end{bmatrix}.$$

The family $\mathcal{F}$ is now union-closed as one may easily check. Unfortunately the decimal values do not immediately reveal this fact. For this we really need the binary values of the subsets.

Definition of the 'closure matrix'

In order to facilitate the process of checking closedness of a given family $\mathcal{F}$ we construct the so-called *closure matrix* C_n. This is a matrix of size $2^n\times 2^n$. Its rows and columns are labeled by the integers from 0 to 2^n-1. Thereby we consider each such label as the decimal value of a subset of $[n]$. For any pair of integers $i,j\in[0,2^n-1]$ the (i,j)-entry of C_n is defined as follows. With the subsets S and T such that $z(S)=i$ and $z(T)=j$ we define

$$C_n(i,j)=\begin{cases}0 & \text{if } S\cup T=S \text{ or } S\cup T=T,\\ z(S\cup T) & \text{otherwise.}\end{cases}$$

The underlying idea is that if $S\cup T=S$ (which holds if and only if $T\subseteq S$), the pair (S,T) will certainly not prevent $\mathcal{F}$ to be a *UC*-family. Similarly if $S\cup T=T$. In all other cases $C_n(i,j)$ is the (positive!) decimal value representing the subset $S\cup T$.

By way of example we present in Figure 1 the closure matrix for $n=4$, with its row and column numbering.

The zeros at the boundary of C_4 are due to the fact that the empty set (which has decimal value 0, the label of the first row and the first column) is contained in every subset of the set $X=[4]$; moreover X itself (whose decimal value is the label of the last row and last column) contains every subset. Due to the construction of C_n we may state the following lemma without further proof.

Lemma 1. *Let $\mathcal{F}\subseteq[0,2^n-1]$. Then $\mathcal{F}$ is union-closed if and only if every (nonzero) entry of $C_n(\mathcal{F},\mathcal{F})$ belongs to $\mathcal{F}$.*

If $[0]\in\mathcal{F}$, the word 'nonzero' in the lemma can be omitted. Then we feel free to reformulate Lemma 1 shortly as follows:

$$\mathcal{F} \text{ is a UC-family} \iff C_n(\mathcal{F},\mathcal{F})\subseteq\mathcal{F}.$$

By way of example consider the case where $n=4$ and $\mathcal{F}=\{0,2,3,5,12\}$. Then

$$A_{\mathcal{F}}=\begin{bmatrix}0&0&0&0\\0&0&1&0\\0&0&1&1\\0&1&0&1\\1&1&0&0\end{bmatrix}$$

and

$$C_4(\mathcal{F},\mathcal{F})=\begin{bmatrix}0&0&0&0&0\\0&0&0&7&14\\0&0&0&7&15\\0&7&7&0&13\\0&14&15&13&0\end{bmatrix}.$$

Observe that this submatrix of C_4 contains entries that do not belong to $\mathcal{F}$. Obviously, to get a UC-family these entries must be added to $\mathcal{F}$. We thus obtain $\mathcal{F}=\{0,2,3,5,7,12,13,14,15\}$. For this $\mathcal{F}$ we get

$$C_n(\mathcal{F},\mathcal{F})=\begin{bmatrix}0&0&0&0&0&0&0&0&0\\0&0&0&7&0&14&15&0&0\\0&0&0&7&0&15&15&15&0\\0&7&7&0&0&13&0&15&0\\0&0&0&0&0&15&15&15&0\\0&14&15&13&15&0&0&0&0\\0&15&15&0&15&0&0&15&0\\0&0&15&15&15&0&15&0&0\\0&0&0&0&0&0&0&0&0\end{bmatrix}.$$

At this stage there are no new decimal values in $C_n(\mathcal{F},\mathcal{F})$, which implies that the current family $\mathcal{F}$ is a UC-family. The binary representation of $\mathcal{F}$ is

$$A_{\mathcal{F}}=\begin{bmatrix}0&0&0&0\\0&0&1&0\\0&0&1&1\\0&1&0&1\\0&1&1&1\\1&1&0&0\\1&1&0&1\\1&1&1&0\\1&1&1&1\end{bmatrix}.$$

We conclude this section with an interesting observation. It concerns the nonzero pattern of C_n. For $n=8$ this is depicted in Figure 2. Readers who are familiar with

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	3	0	5	0	7	0	9	0	11	0	13	0	15	0
2	0	3	0	0	6	7	0	0	10	11	0	0	14	15	0	0
3	0	0	0	0	7	7	7	0	11	11	11	0	15	15	15	0
4	0	5	6	7	0	0	0	0	12	13	14	15	0	0	0	0
5	0	0	7	7	0	0	7	0	13	13	15	15	13	0	15	0
6	0	7	0	7	0	7	0	0	14	15	14	15	14	15	0	0
7	0	0	0	0	0	0	0	0	15	15	15	15	15	15	15	0
8	0	9	10	11	12	13	14	15	0	0	0	0	0	0	0	0
9	0	0	11	11	13	13	15	15	0	0	11	0	13	0	15	0
10	0	11	0	11	14	15	14	15	0	11	0	0	14	15	0	0
11	0	0	0	0	15	15	15	15	0	0	0	0	15	15	15	0
12	0	13	14	15	0	13	14	15	0	13	14	15	0	0	0	0
13	0	0	15	15	0	0	15	15	0	0	15	15	0	0	15	0
14	0	15	0	15	0	15	0	15	0	15	0	15	0	15	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Figure 1 The matrix C_4, with its row and column numbering.

fractals will recognize the well-known so-called Sierpinski fractal and mirrored versions of this fractal.[1] It came as a surprise, reminding us of Isaac Newton's 'smoother pebble or prettier shell than ordinary'.[2] At this stage it is not clear how to use this beautiful picture for the proof of the UCC. But as we demonstrate in the next section, the matrix underlying this picture turns out to be a useful tool when performing computational experiments.

Remark 2. The matrix C_n can also be obtained in a different way. Let $\tilde{C}_0$ denote the 1×1 zero matrix: $\tilde{C}_0 = 0$. For $n \geq 0$ we define

$$\tilde{C}_{n+1} = \begin{bmatrix} \tilde{C}_n & \tilde{C}_n + 2^n \\ \tilde{C}_n + 2^n & \tilde{C}_n + 2^n \end{bmatrix},$$

where $\tilde{C}_n + 2^n$ denotes the matrix arising from $\tilde{C}_n$ by adding 2^n to each of its entries. Then we obtain C_n from $\tilde{C}_n$ by replacing $\tilde{C}_n(i,j)$ by 0 if $\tilde{C}_n(i,j) = i$ or $\tilde{C}_n(i,j) = j$. The closure matrix G_n for IC-families can be obtained in a similar way. With $\tilde{G}_0 = \tilde{C}_0$, we define for $n \geq 0$,

$$\tilde{G}_{n+1} = \begin{bmatrix} \tilde{G}_n & \tilde{G}_n \\ \tilde{G}_n & \tilde{G}_n + 2^n \end{bmatrix},$$

Then we obtain the closure matrix G_n for IC-families from $\tilde{G}_n$ by replacing $\tilde{G}_n(i,j)$ by 0 if $\tilde{G}_n(i,j) = i$ or $\tilde{G}_n(i,j) = j$.

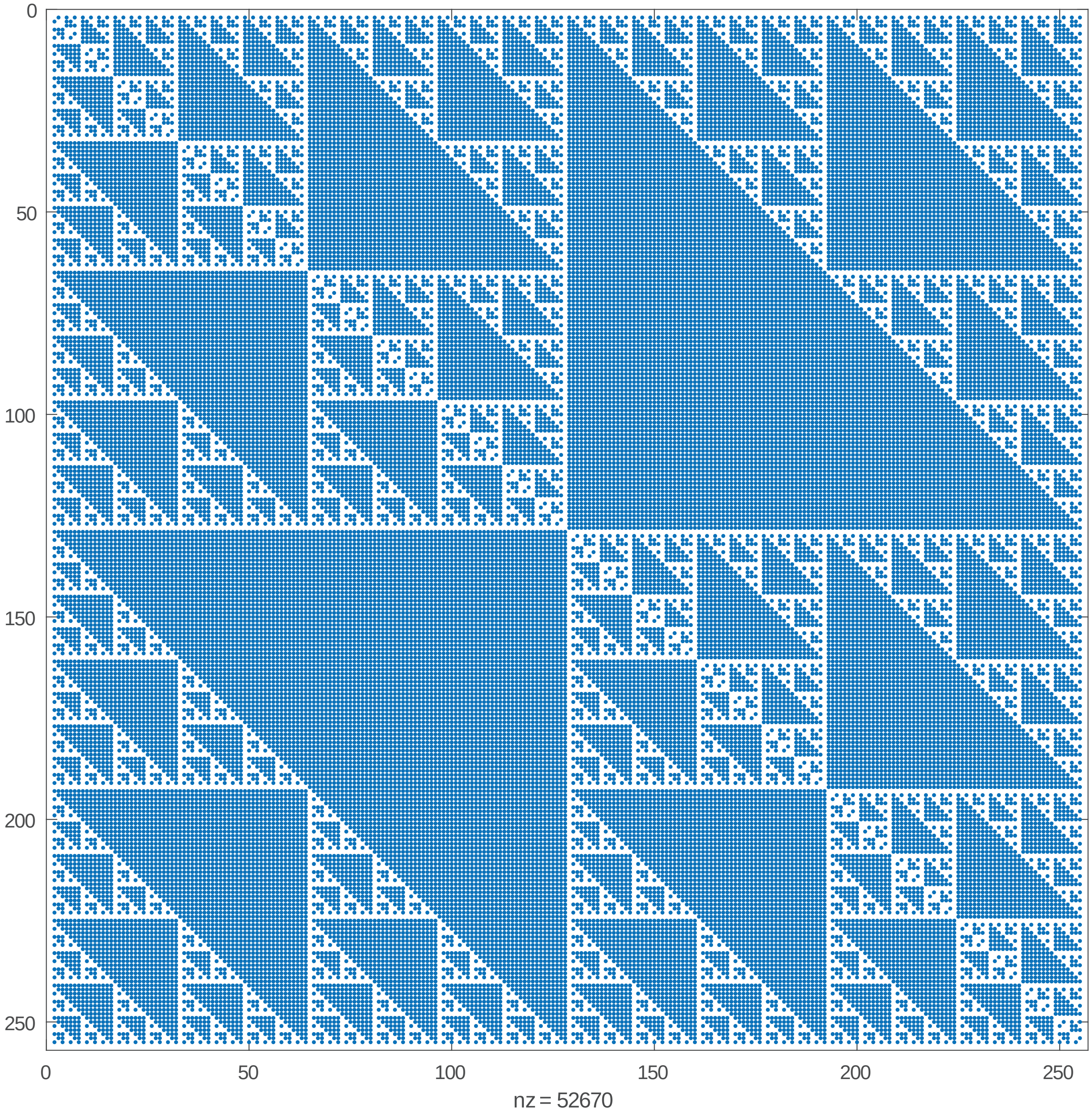

Figure 2 Nonzero pattern of C_n for $n = 8$.

Optimization model with binary variables

Obviously the matrix C_n contains all the information we need for deciding whether a family $\mathcal{F}$ is union-closed or not. In this section we give some examples of its usefullness for performing numerical experiments.[3]

In this section A denotes a fixed matrix whose (distinct!) rows are all binary vectors of length n. So A is the binary representation of the power set of $[n]$. We assume that the rows are ordered according to their decimal values, in increasing order. So the first row of A contains only zeros and the last row only ones. Obviously, A has size $N \times n$, with $N = 2^n$, and A is certainly union-closed. Moreover, the number of ones in each column equals $N/2$, so A satisfies the UCC tightly.

We consider submatrices of A that consist of a subset of its rows. Such a submatrix equals $A_{\mathcal{F}}$ for the family $\mathcal{F}$ whose decimal representation consists of the decimal values of the rows in the submatrix. Below we represent $\mathcal{F}$ by its support vector x. So x is the binary (column) vector of length N defined by

$$x_i = 1 \;\Leftrightarrow\; i \in \mathcal{F}.$$

With $m = |\mathcal{F}|$, we then have $x^T \mathbf{1}_N = m$, where $\mathbf{1}_N$ denotes the all-one vector of length N. Obviously the entries in the vector $x^T A$ count the number of ones in the successive columns of $A_{\mathcal{F}}$. Since we assumed that $A_{\mathcal{F}}$ has no zero column, each of these numbers must be at least one, which holds if and only if

$$x^T A \geq \mathbf{1}_n.$$

Assuming that $\mathcal{F}$ is union-closed, it will now be clear that the UCC requires that

$$\max(x^T A) \geq \tfrac{1}{2} \mathbf{1}_N^T x.$$

Using the closure matrix C_n we may formulate the UC condition in terms of x as follows. With $k = C_n(i,j)$, we must have

$$x_i = 1 \text{ and } x_j = 1 \;\Rightarrow\; x_k = 1, \qquad \forall\, i,j \in \mathcal{F}.$$

Since x is a $\{0,1\}$-vector, this implication can be modelled by the linear inequality

$$x_k \geq x_i + x_j - 1. \tag{2}$$

We conclude that every triple (i,j,k) such that $k = C_n(i,j) > 0$ gives rise to an inequality for the $\{0,1\}$-vector x.

To simplify the notation in the sequel, the symbol $\mathbf{1}$ will denote an all-one vector of suitable length if the meaning of 'suitable' is clear from the context. Then, by rewriting (2) in the form $x_i + x_j - x_k \leq 1$, these inequalities altogether form a system of constraints

$$D_n x \leq \mathbf{1}$$

for some suitable matrix D_n. Thus we may conclude that x defines a UC-family if and only if

$$D_n x \leq \mathbf{1}, \quad x^T A \geq \mathbf{1}, \\ x_i \in \{0,1\},\ 0 \leq i \leq N-1.$$

Now we can state Frankl's conjecture as follows:

$$D_n x \leq \mathbf{1}, \quad x^T A \geq \mathbf{1}, \\ x \in \{0,1\}^N \;\Rightarrow\; \max(x^T A) \geq \tfrac{1}{2} \mathbf{1}^T x.$$

It may be noted that this holds if and only if the system

$$D_n x \leq \mathbf{1}, \quad x^T A \geq \mathbf{1}, \\ x \in \{0,1\}^N, \quad \max(x^T A) < \tfrac{1}{2} \mathbf{1}^T x.$$

is infeasible.

Some computational results

The first question that we consider is whether there exist UC-families for every value of m. To investigate this we consider the optimization problem

$$\min\{\tau : \max(x^T A) \leq \tau, x^T A \geq \mathbf{1}, \\ D_n x \leq \mathbf{1}, \mathbf{1}^T x = m, x \in \{0,1\}^N\}, \tag{3}$$

for $m = 2$ to N. Any feasible solution x of this minimization problem corresponds to a UC-family, due to the definition of the matrix D_n. Since we are minimizing τ, at optimality we have $\tau = \max(x^T A)$. Therefore, if the pair (x,τ) solves problem (3) then τ is equal to the lowest possible number of ones in an abundant column. According to the conjecture of Frankl we should therefore have $2\tau \geq \mathbf{1}^T x$.

For $n = 4$ Table 1 shows the optimal values of τ, τ/m and $x^T A$ for every value of m with $2 \leq m \leq N$. The table makes clear (at least if $n = 4$) that a UC-family exists for every value of m. Moreover, in all cases we have $2\tau \geq m$, in agreement with the UCC. Note, however, that the conjecture is tight only when m is a power of two and, moreover, in that case every column has the same number of ones. The latter means that in the tight case all elements are abundant. See also Remark 3 below. Table 1 makes also clear that if we add the constraint $2\tau \leq m$ to (3) then only the cases where m is a power of 2 survive; in all other cases problem (3) becomes infeasible.

We also considered the case where the condition $\mathbf{1}^T x = m$ in problem (3) is replaced by $\mathbf{1}^T x \geq m$, while including the constraint $2\tau \leq \mathbf{1}^T x$, in order to guarantee tightness with respect to Frankl's conjecture. Table 2 shows the results for the corresponding problem, which is given by

$$\min\{\tau : \max(x^T A) \leq \tau \leq \tfrac{1}{2} \mathbf{1}^T x, x^T A \geq \mathbf{1}, \\ D_n x \leq \mathbf{1}, \mathbf{1}^T x \geq m, x \in \{0,1\}^N\}. \tag{4}$$

The table confirms that tightness occurs if and only if $|\mathcal{F}|$ a power of 2.

Remark 3. It is worth mentioning that in our experiments the solution of the optimization problems (3) and (4) had a surprisingly simple structure in the cases where $x^T \mathbf{1}$ is a power of 2. Suppose $m = 2^k$, with $1 \leq k \leq n$. In order to describe the special structure we use for the moment the notation A_k for the $2^k \times k$ matrix whose rows are all binary vectors of length k. Then the solution always contained A_k as a submatrix, and the remaining $n-k$ columns were copies of columns in A_k. It means that there are exactly k different columns and these are the columns of A_k. In terms of Frankl's 'subset model' this means that there are one or more pairs of elements in X such that both elements belong to exactly the same subsets in $\mathcal{F}$. Moreover, by removing the $n-k$ columns outside the columns of A_k, we are left with the power set of $[k]$.

Based on these experiments and similar experiments for other small values of n we state the following conjecture:

Conjecture 3. *If Frankl's conjecture is tight for some UC-family $\mathcal{F}$ then $|\mathcal{F}| = 2^k$ for some k and $A_{\mathcal{F}}$ contains A_k as a submatrix.*

It looks as if the proof of Conjecture 3 will need a clever use of the structure of the matrix C_n. In fact, a simple observation already leads to a special case where this conjecture is true. This is the topic of the next section.

A small step forward

Before dealing with the theorem below, we suggest the reader to verify the reasoning in the proof by using the matrix C_4 in Figure 1 for the case where $n = 4$, $k = 3$ and

m	$\mathbf{1}^Tx$	τ	τ/m	x^TA
2	2	1	0.5000	[1 1 1 1]
3	3	2	0.6667	[2 1 1 1]
4	4	2	0.5000	[2 2 2 2]
5	5	3	0.6000	[3 3 3 1]
6	6	4	0.6667	[4 4 3 1]
7	7	4	0.5714	[4 4 4 3]
8	8	4	0.5000	[4 4 4 4]
9	9	5	0.5556	[5 5 5 1]
10	10	6	0.6000	[6 6 5 4]
11	11	7	0.6364	[7 6 6 5]
12	12	7	0.5833	[7 7 6 6]
13	13	8	0.6154	[8 7 7 6]
14	14	8	0.5714	[8 8 7 7]
15	15	8	0.5333	[8 8 8 8]
16	16	8	0.5000	[8 8 8 8]

Table 1 Solutions of (3) for $n=4$ and $2 \le m \le N$.

m	$\mathbf{1}^Tx$	τ	$\tau/\mathbf{1}^Tx$	x^TA
2	2	1	0.5000	[1 1 1 1]
3	4	2	0.5000	[2 2 2 2]
4	4	2	0.5000	[2 2 2 2]
5	8	4	0.5000	[4 4 4 4]
6	8	4	0.5000	[4 4 4 4]
7	8	4	0.5000	[4 4 4 4]
8	8	4	0.5000	[4 4 4 4]
9	16	8	0.5000	[8 8 8 8]
10	16	8	0.5000	[8 8 8 8]
11	16	8	0.5000	[8 8 8 8]
12	16	8	0.5000	[8 8 8 8]
13	16	8	0.5000	[8 8 8 8]
14	16	8	0.5000	[8 8 8 8]
15	16	8	0.5000	[8 8 8 8]
16	16	8	0.5000	[8 8 8 8]

Table 2 Solutions of (4), where we require $\mathbf{1}^Tx \ge m$ and $\tau \le \mathbf{1}^Tx/2$.

$m = 2^k$. It makes the statement in the following theorem almost trivial.

Theorem 1. *With* $N = 2^n$, *let* $2 \le m < N$ *and*

$$\mathcal{F} := \{i: 0 \le i < m\}. \tag{5}$$

Then $\mathcal{F}$ is a UC-family if and only if m is a power of 2 and then the UCC holds tightly.

Proof. The decimal representation of $\mathcal{F}$ is the set $\{0,\dots,m-1\}$. Let k be such that

$$2^k \le m < 2^{k+1}. \tag{6}$$

Since $m \ge 2^k$, $\mathcal{F}$ contains all subsets whose decimal values are strictly less than 2^k. The support vectors of these subsets are the binary vectors $s=(s_n,\dots,s_1)$ with $s_i = 0$ for each $i > k$. These are precisely the 2^k subsets of $[k] = \{1,\dots,k\}$.

If $m = 2^k$ these are all subsets in $\mathcal{F}$. So, then $\mathcal{F}$ is the power set $[k]$. Therefore, $\mathcal{F}$ is union-closed and each element $i \in [k]$ belongs to precisely $2^k/2 = m/2$ of the subsets in $\mathcal{F}$. This proves that Frankl's conjecture is tight if m is a power of 2.

Next we consider the case where (6) holds and m is not a power of 2. Then

$$2^k < m \le 2^{k+1} - 1. \tag{7}$$

As we already established, $\mathcal{F}$ contains all subsets of $[k]$. Let j be the decimal value of a subset of $S \subseteq [k]$. The subset represented by 2^k is the singleton $\{k+1\}$. Obviously, the set S and $\{k+1\}$ are disjoint. Therefore, $z(\{k+1\} \cup S) = 2^k + j$. Taking $j = 2^k - 1$, we get $C_n(2^k,j) = 2^k + 2^k - 1 = 2^{k+1} - 1 \ge m$, where the inequality is due to (7). By the definition of $\mathcal{F}$ in (5) this implies $C_n(2^k,j) \notin \mathcal{F}$. By Lemma 3.1 this implies that $\mathcal{F}$ is not union-closed. This completes the proof. □

We conclude this section with a similar result. Its proof is much simpler than that of the previous theorem — straightforward, without using C_n — and is therefore omitted.

Theorem 2. *Let N be as before and*

$$\mathcal{F}_m := \{i: N+1-m \le i \le N\}, \quad 1 \le m \le N. \tag{8}$$

Then $\mathcal{F}_m$ is a UC-family of size m, for each m. Moreover, the largest number of ones per column equals $\min(m, N/2)$, *which is tight for the UCC if and only if* $m = N$.

UC- versus IC-families

Let $\mathcal{F}$ be a family of m subsets of the set X $(=[n])$ and let the family $\mathcal{F}'$ consist of the complements of the subsets in $\mathcal{F}$. Then $|\mathcal{F}'| = |\mathcal{F}| = m$. As already mentioned in Remark 1, the theorem of de Morgan implies that $\mathcal{F}$ is a UC-family if and only if $\mathcal{F}'$ is a IC-family. Below it will be assumed that $\mathcal{F}$ is a UC-family.

As before, let A be the matrix whose rows are the support vectors of all subsets of X in the 'natural' order (i.e., ordered with respect to their decimal values). The submatrix of A determined by $\mathcal{F}$ is denoted as A_x, where x is the support vector of $\mathcal{F}$. Similarly A_y denotes the submatrix of A determined by $\mathcal{F}'$, with y denoting the support vector of $\mathcal{F}'$.

If S and T are two subsets of X then they are each others complement if and only if the sum of their binary values equals the all-one vector of length n, and this holds if and only if the sum of their decimal values equals $N-1$, where $N = 2^n$. Thus we may conclude that the support vectors x and y are related as follows:

$$y_i = 1 \Leftrightarrow x_{N-1-i} = 1, \quad 0 \le i \le N-1.$$

Hence, the vector y arises from x by inverting the order of the entries of x. Stated otherwise,

$$y = Qx, \tag{9}$$

where Q is the $N \times N$ $\{0,1\}$-matrix with ones on its main anti-diagonal. We therefore call x and y each others *opposite vector*.

Now let $B := QA$. So B is the matrix whose rows are the support vectors of all subsets of X in the 'opposite' order. Since Q is symmetric, we then may write

$$y^TA = (Qx)^TA = x^TQ^TA = x^TQA = x^TB. \tag{10}$$

Since each row of B is the complement of the corresponding row of A, $A+B$ is the all-one matrix of size $N\times n$. In other words,

$$A+B=\mathbf{1}_N\mathbf{1}_n^T. \tag{11}$$

It follows that

$$\begin{aligned} x^TA+y^TA &= x^TA+x^TB \\ &= x^T(A+B) \\ &= x^T\mathbf{1}_N\mathbf{1}_n^T \\ &= (\mathbf{1}_N^Tx)\mathbf{1}_n^T = m\mathbf{1}_n^T. \end{aligned} \tag{12}$$

The first equality holds because of (10), the third equality because of (11) and the last equality because $x^T\mathbf{1}_N=m$. We may rewrite (12) as follows

$$(x^TA)_k+(y^TA)_k=m,\quad 1\le k\le n, \tag{13}$$

which shows that the k-th column sum for A_x plus the k-th column sum for A_y is equal to m, for each k. An immediate consequence is that for each k we have either

$$(x^TA)_k\ge \tfrac{1}{2}m\ge (y^TA)_k \tag{14}$$

or

$$(y^TA)_k\ge \tfrac{1}{2}m\ge (x^TA)_k. \tag{15}$$

Let us emphasize that until now the results in this section hold for every family $\mathcal{F}$ and its 'opposite' family $\mathcal{F}'$. Moreover, the UCC simply means that if $\mathcal{F}$ is union-closed then (14) holds for at least one k.

Finally, we consider the case where x is *self-opposite*, i.e., when $Qx=x$. Then $y^TA=(Qx)^TA=x^TA$ and hence, by (12), all column sums in A_x (and in A_y) are equal to $\frac{1}{2}m$. So, if x is self-opposite this implies tightness. We observed in our computations that if x is the support vector of a tight UC-family the converse is also true. This is confirmed by Figure 3. This figure shows the nonzero patterns of the vectors x for the 52 UC-families $\mathcal{F}$ (with $n=5$) for which the UCC is tight. The symmetry in this figure reflects the fact that these vectors are self-opposite. Also note that the number of ones in these vectors x is always a power of 2, in agreement with Conjecture 3.[4]

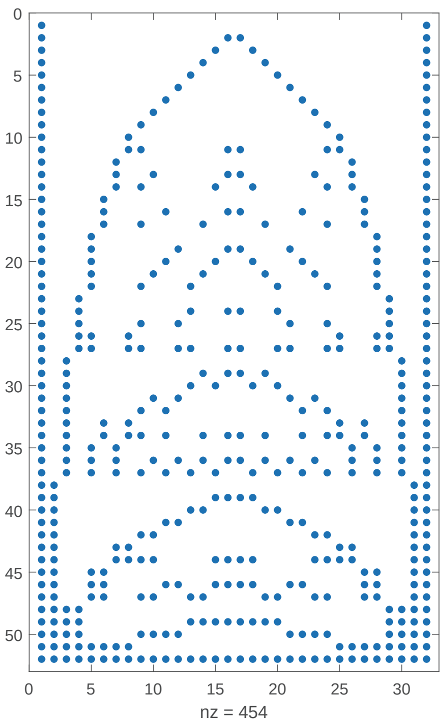

Figure 3 Nonzero patterns of tight vectors x for $n=5$.

Conclusion

As may be clear, in essence this short note leaves the conjecture of Frankl open. But hopefully it brings us somehow closer to the solution.

Acknowledgement

Thanks are due to our neighbor ir. C.H. (Kees) Boer; he developed the habit to cut articles related to mathematics from his daily newspaper (*NRC*) and to put them in our mailbox. Thanks are also due to my friend dr. J.J. (Hans) Visser; he sends me almost daily links to such and other interesting articles. Due to them I learned about Frankl's conjecture, earlier this year on January 23. I dedicate this paper to them by citing the Dutch mathematician Ludolph van Ceulen (1540–1610), who became famous for his computation of 35 decimals of the number π.[5] I gladly concur with what he wrote (in old Dutch) to prince Maurits in the preface to his book [3]: "Alsoo heeftet den almachtighen Godt believt / oock mijn de Schuppe inde handt te gheven / om de selve te ghebruycken tot mijn beroup / d'welck is (Hoogh-gheboren Vorst ...) de konste van Meten ende Tellen (sonder wiens behulp meest alle andere soo Liberale als Mechanische scientien ende konsten creupel ende inperfect bevonden werden) in welcke ick het mijne ghedaen hebbe / ende daer inne soo vele gevordert / als den almachtighen Godt believt / ende syn ghenade verleent heeft/."

Notes

1 A natural question is which fractal is obtained when dealing with IC-families. We leave it to the interested reader to verify that it differs from the fractal in Figure 2 only in the way the Sierpinski fractals are arranged.

2 "I do not know what I may appear to the world; but to myself I seem to have been only like a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell than ordinary, whilst the great ocean of truth lay all undiscovered before me", Isaac Newton, 1642–1724 [1].

3 In our experiments we used the software package Mosek for solving optimization problems together with the popular Matlab modelling package CVX. Both Mosek and CVX are freely available for academic purposes.

4 If $n=5$ there are $2^{(2^5)}/4=1\,073\,741\,824$ families of subsets that contain both the empty set and the full set. Only $1\,373\,701$ of these are UC-families, and — as the figure shows — only 52 are tight. On my PC it required about 45 minutes computational time to obtain Figure 3. For $n=6$ the total number of families becomes $2^{(2^6)}/4=4(1\,073\,741\,824)^2$. A rough estimate of the computational time required for a similar figure for $n=6$ is $4\times 1\,073\,741\,824\times 45$ minutes, i.e., about $367\,720$ years.

5 These decimals can be seen on a plaque affixed to one of the pillars of the Pieterskerk in Leiden [7].

References

1 David Brewster, Newton, Sir Isaac, in *Encyclopaedia Britannica*, 1842, pp. 175–181, https://www.newtonproject.ox.ac.uk/catalogue/record/OTHE00010.

2 Henning Bruhn and Oliver Schaudt, The journey of the union-closed sets conjecture, *Graphs Combin.* 31(6) (2015), 2043–2074.

3 Ludolph van Ceulen, *Van den Circkel*, Jan Andrieszoon, Boeckvercooper, Delft, 1596. I quote from the transcript of Marloes Bazeliet et al., which can be found at http://www.ludolphvanceulen.nl/site/vandencirckel.php.

4 Peter Frankl, Extremal set systems, in *Handbook of Combinatorics, Vol. 1, 2*, Elsevier Science, 1995, pp. 1293–1329.

5 Justin Gilmer, A constant lower bound for the union-closed sets conjecture (2022), arXiv:2211.09055.

6 Ilan Karpas, Two results on union-closed families (2017), arXiv:1708.01434.

7 R.M.Th.E. Oomes, J.J.T.M. Tersteeg and J. Top, Het grafschrift van Ludolph van Ceulen, *Nieuw Archief voor Wiskunde* 5/1(2) (2000), 156–161.

8 Luke Pebody, Extension of a method of Gilmer (2022). arXiv:2211.13139.

9 Will Sawin, An improved lower bound for the union-closed set conjecture (2022), arXiv:2211.11504.

10 Lei Yu, Dimension-free bounds for the union-closed sets conjecture (2022), arXiv:2212.00658.