COLLOQUIA MATHEMATICA SOCIETATIS JÁNOS BOLYAI
52. COMBINATORICS, EGER (HUNGARY), 1987

How to Tidy up Your Set-system?

C. A. J. HURKENS,[1] L. LOVÁSZ, A. SCHRIJVER and É. TARDOS

Suppose that you have a finite set S of n elements and a very "messy" set $\mathcal{F}$ of m of its subsets. You want to "tidy up" your system by taking any two incomparable sets X and Y in $\mathcal{F}$ (i.e., such that $X \not\subseteq Y$ and $Y \not\subseteq X$), and replace them by $X \cap Y$ and $X \cup Y$. The sets X and Y are thrown in the garbage. (We take no multiplicity into account; if, say, $X \cup Y$ had been a member of $\mathcal{F}$ before the operation, it simply remains a member.)

It is easy to see that repeating this at most $\binom{m}{2}$ times, we obtain a totally ordered family (the number of pairs of incomparable sets in F decreases at each step).

But suppose that your little daughter is also present and any time you throw X and Y in the garbage, she "salvages" one of them and puts it back in $\mathcal{F}$. (As many of us know, she is a great adversary to tidying up (GATU), so her objective is to prolong the procedure as much as possible.) Can you win? And if so, can you win in polynomial time?

We are going to show that — contrary to everyday experience — you can win in polynomial time. However, one has to be careful: choosing just any incomparable pair to uncross at each step may result in cycling. We shall show a simple rule which guarantees finite termination, and a more complicated rule that guarantees polynomial time termination. We do not know whether or not the first rule terminates in polynomial time.

The background of the problem is the following. Many proofs and algorithms in combinatorics (in particular in connection with graph connectivity and submod-

[1] Research supported by the Netherlands Organization for the Advancement of Pure Research Z.W.O. through the Mathematical Centre

ularity) use the "uncrossing" procedure, i.e., replacing two sets by their intersection and union. In many combinatorial optimization problems, however, optimum solutions consist of a set-system $(S, \mathcal{F})$ together with positive weights associated with its members, and the optimality of the solution is preserved by the following step: if set X occurs with weight λ and set Y, with weight μ, and, say, $\lambda \geq \mu$, then we can add μ to the weights of $X \cup Y$ and $X \cap Y$ (adding them to the system if necessary), and subtract μ from the weights of X and Y. Thereby Y (and possibly X) disappears from the system. If we manage to transform the optimum solution into a chain, then we have a more tractable situation (see, e.g., Grötschel, Lovász and Schrijver (1988)). The number of steps in this uncrossing procedure depends on the choice of the pairs X, Y; we want to find a rule which will make the number of steps polynomial in $|S|$ and $\mathcal{F}$. If we want a selection rule independent of the weights, we get the problem described above.

Let us start with an example showing that we do need some additional rule to guarantee finite termination.

Example 1. Let $S = \{1, 2, 3\}$ and

$$\mathcal{F} = \{\emptyset, \{1\}, \{2\}, \{1,3\}, \{2,3\}, \{1,2,3\}\}.$$

In the first step remove $\{1,3\}$ and $\{2,3\}$ to uncross; we put back $\{3\}, \{1,2,3\}$ (which is already there) and GATU puts back, say, $\{2,3\}$. Next, we select $\{3\}$ and $\{1\}$; we put back $\emptyset$ (which is already there), $\{1,3\}$ and GATU puts back $\{1\}$. We are back to our original system!

The rules we are going to study are the following.

Rule 1. At each step, we choose an incomparable pair X, Y of sets such that $|X \cup Y|$ is maximum. If there are ties, we choose $|X| + |Y|$ maximum.

Rule 2. We maintain an "active" part $\mathcal{F}_0$ of $\mathcal{F}$. To begin with, we take $\mathcal{F}_0 = \emptyset$.

(2.1) If $\mathcal{F}_0$ is a chain, and $\mathcal{F}_0 \neq \mathcal{F}$ then we select any $X \in \mathcal{F} - \mathcal{F}_0$ and add it to $\mathcal{F}_0$.

(2.2) If $\mathcal{F}_0$ is a chain and $\mathcal{F}_0 = \mathcal{F}$ then we (of course) stop.

(2.3) If $\mathcal{F}_0$ is not a chain then we apply Rule 1 to $\mathcal{F}_0$.

Let $\mathcal{L}$ denote the lattice generated by the members of $\mathcal{F}$ (i.e., the set of all subsets of S that can be obtained as unions of intersections of members of $\mathcal{F}$). Our first result implies that Rule 1 is sufficient for finite termination.

Theorem 1. *Rule 1 guarantees termination in $O(|\mathcal{L}|^2)$ steps.*

Unfortunately, the bound given in this theorem is typically exponential in $|S| + |\mathcal{F}|$ (although at one point it will be important that it is polynomial in some special cases). To get polynomial time convergence we can use Rule 2:

Theorem 2. *Rule 2 guarantees termination in $O(|S| \cdot |\mathcal{F}|)$ steps.*

Let us remark that some special instruction concerning breaking ties in Rule 1 is needed, as the following example shows:

Example 2. Let S be the set of edges of K_4, the complete graph on four points a, b, c, d, and let $\mathcal{F}$ consist of the triangles in K_4, the pair $\{ab, bc\}$, as well as of S and $\emptyset$. First uncross the triangle $\{ac, cd, ad\}$ and the pair $\{ab, bc\}$, salvaging the triangle; then uncross the quintuple $\{ab, bc, ac, cd, ad\}$ and the triangle $\{bc, cd, bd\}$, salvaging the triangle again. We are left with a set-system isomorphic with the original.

To prove these theorems, we need some lemmas about the procedure. Let $\mathcal{F}_k$ be the set-sytem after the k-th step. We say that a set $D \in \mathcal{F}_k$ is *dominant* if the following hold:

(a) for every $X \in \mathcal{F}_k$ such that $|X| \geq |D|$, we have $X \supseteq D$;

(b) for every $X, Y \in \mathcal{F}_k$ such that $|X \cup Y| > |D|$, we have either $X \supseteq D$ or $Y \supseteq D$.

Let $\mathcal{D}_k$ denote the set of dominant elements of $\mathcal{F}_k$. It is clear from the definition that $\mathcal{D}_k$ is a chain. Moreover, every element of $\mathcal{F}$ comparable with every other element of $\mathcal{F}$ is trivially dominant. We shall call these elements of $\mathcal{F}$ *trivial.*

Lemma 1. *If we obtain $\mathcal{F}_{k+1}$ from $\mathcal{F}_k$ by Rule 1, uncrossing X and Y, then both X and Y are maximal non-trivial sets in $\mathcal{F}_k$ and $X \cup Y$ is dominant in $\mathcal{F}_{k+1}$.*

Proof. I. Suppose that X is properly contained in a non-trivial member W of $\mathcal{F}_k$. If $Y \not\subseteq W$ then the pair $\{W, Y\}$ contradicts the choice of X and Y. On the other hand, if $Y \subseteq W$, then we can pick any set T non-comparable with W, and the pair $\{W, T\}$ will contradict the choice of X and Y.

II. Suppose that $X \cup Y$ is non-dominant in $\mathcal{F}_{k+1}$. This can happen in two ways.

(a) There exists a set $U \in \mathcal{F}_{k+1}$ such that $|U| \geq |X \cup Y|$, and yet $U \not\supseteq X \cup Y$. Without loss of generality we may assume that $U \not\supseteq X$. But then clearly X and U are incomparable and $|X \cup U| > |X \cup Y|$, contradicting Rule 1.

(b) There exist sets $U, V \in \mathcal{F}_{k+1}$ such that $|U \cup V| > |X \cup Y|$, and yet $U \not\supseteq X \cup Y$ and $V \not\supseteq X \cup Y$. Rule 1 would have forced us to pick U and V instead of X and Y unless U and V are comparable, say $U \subseteq V$. But then V contains $X \cup Y$ by the preceding argument. ∎

Let $\mathcal{A}_k$ denote the order ideal in $\mathcal{L}$ generated by $\mathcal{F}_k - \mathcal{D}_k$, i.e., the set of all members of $\mathcal{L}$ contained in some non-dominant member of $\mathcal{F}_k$.

Lemma 2. *If $\mathcal{A}_k = \{\emptyset\}$ then $\mathcal{F}_k$ is a chain.*

Proof. If $\mathcal{F}_k$ is not a chain then it contains two incomparable sets X and Y. If, say $|X| \leq |Y|$ then certainly X is non-dominant and so it belongs to $\mathcal{A}_k$. ∎

Lemma 3. *If Rule 1 is used, then* $\mathcal{A}_{k+1} \subseteq \mathcal{A}_k$.

Proof. Assume that X and Y are uncrossed to get $\mathcal{F}_{k+1}$ from $\mathcal{F}_k$. Let $V \in \mathcal{A}_{k+1}$. Then $V \in \mathcal{L}$ and moreover, there exists a set $U \in \mathcal{F}_{k+1} - \mathcal{D}_{k+1}$ such that $V \subseteq U$. If $U \in \mathcal{F}_k - \mathcal{D}_k$ then the conclusion is trivial. So suppose that this is not the case. Then there are two possibilities:

(1) $U \notin \mathcal{F}_k$. In this case U is either $X \cup Y$ or $X \cap Y$. But in the first case U is dominant in $\mathcal{F}_{k+1}$ by Lemma 1. In the second case, V is contained in both X and Y, and at least one of them is a non-dominant member of $\mathcal{F}_k$, since dominant sets in $\mathcal{F}_k$ form a chain.

(2) $U \in \mathcal{D}_k$. If U is comparable with every element of $\mathcal{F}_k$ then it is also comparable with every element of $\mathcal{F}_{k+1}$ and hence also dominant in $\mathcal{F}_{k+1}$, contradicting its definition. So there is a set $W \in \mathcal{F}_k$ not comparable with U. Rule 1 implies that $|X \cup Y| \geq |U \cup W| > |U|$. Since U is dominant in $\mathcal{F}_k$, this implies that either $X \supseteq U$ or $Y \supseteq U$. Say, $X \supseteq U$.

Now U is non-dominant in $\mathcal{F}_{k+1}$, which could mean two things:

(2.1) There exists a $T \in \mathcal{F}_{k+1}$ such that $|T| \geq |U|$ but T does not contain U. Since this situation cannot occur in $\mathcal{F}_k$, it follows that $T \notin \mathcal{F}_k$, i.e., either $T = X \cup Y$ or $T = X \cap Y$. The first equality is impossible as then $U \subseteq X \subset T$. The second equality is impossible since $U \not\subseteq T = X \cap Y$ implies $U \not\subseteq Y$; but $|Y| > |T| \geq |U|$, which contradicts the assumption that U is dominant in $\mathcal{F}_k$.

(2.2) There exist $T, R \in \mathcal{F}_{k+1}$ such that $|T \cup R| > |U|$ but neither T nor R contains U. Again, one of T and R, say T, must be a "new" set in $\mathcal{F}_{k+1}$, i.e., either $T = X \cup Y$ or $T = X \cap Y$. The first equality is impossible as then $U \subseteq X \subset T$. The second is impossible since then $U \not\subseteq Y$ but $|Y \cup R| \geq |T \cup R| > |U|$, which contradicts the assumption that U is dominant in $\mathcal{F}_k$. ■

It is easy to describe the case of equality:

Lemma 4. *Let* $\mathcal{F}_{k+1}$ *arise from* $\mathcal{F}_k$ *by uncrossing* X *and* Y, *where GATU salvages* X. *Assume that* $\mathcal{A}_{k+1} = \mathcal{A}_k$. *Then* Y *is a maximal non-trivial dominant set and* X *is a maximal non-dominant set in* $\mathcal{F}_k$.

(Note that Y is thus uniquely determined. Moreover, what happens to the system of dominant sets is that Y gets replaced by $X \cup Y$. This latter set may be trivial or it may be the largest non-trivial dominant set in $\mathcal{F}_{k+1}$.)

Proof. At least one of X and Y is non-dominant, since dominant sets form a chain. Let U denote this set temporarily, and let V be the other set in $\{X, Y\}$.

The set U must be a maximal non-dominant set, by Lemma 1. So U is removed from $\mathcal{A}_k$ by the uncrossing, and so $\mathcal{A}_{k+1} = \mathcal{A}_k$ can only hold if U is "salvaged" by GATU. Hence $U = X$. It also follows that we could not take Y as U, so Y must be a non-trivial dominant set. By Lemma 1, Y must be a maximal non-trivial dominant set. ■

Proof of Theorem 1. We study the sets $\mathcal{A}_k$ and $\mathcal{D}_k - \mathcal{A}_k$. By Lemma 3, the set $\mathcal{A}_k$ is monotone decreasing (not necessarily strictly), so it suffices to prove that it cannot stay constant for more that $O(n^2)$ steps. But in each such step, the largest non-trivial dominant set is either increased (which can happen at most $O(n)$ times consecutively), or it becomes trivial (which can happen only $O(n)$ times since the non-trivial dominant sets form a chain and so their number is $O(n)$). ∎

Proof of Theorem 2. Clearly, step (2.1) can be applied at most $|\mathcal{F}|$ times. Between two applications of this step, Rule 1 is applied to a system which (at least at the beginning) consists of a chain $\mathcal{C}$ and a single additional member W. It is not difficult to analyse what Rule 1 does in such a simple situation. (The lattice generated by such a family consists of all sets of the form $X \cup (S \cap Y)$ where $X, Y \in \mathcal{C}$, and so its cardinality is $O(n^2)$. So by Theorem 1, the number of steps between two consecutive applications of step (2.1) is $O(n^4)$. But we can do better.)

Claim At any step, the active part $\mathcal{F}_0$ will consist of three chains: $\mathcal{C}_1, \mathcal{C}_2$ and $\mathcal{C}_3$ with the following properties. Every member of $\mathcal{C}_1$ contains every member of $\mathcal{C}_2$ and $\mathcal{C}_3$. For every $T \in \mathcal{C}_2$ and $U, V \in \mathcal{C}_3, U \cap T = V \cap T$.

Proof. To begin with, we can take $\mathcal{C}_1 = \emptyset, \mathcal{C}_2 = \mathcal{C}$ and $\mathcal{C}_3 = \{W\}$. Let X and Y be the largest members of $\mathcal{C}_2$ and $\mathcal{C}_3$, respectively. We may assume that X and Y are incomparable, or else we can count the larger one of them in $\mathcal{C}_1$. Then Rule 1 clearly selects X and Y to uncross. Now $X \cup Y$ can go on the "bottom" of chain $\mathcal{C}_1$, while $X \cap Y$ can go on the "bottom" of chain $\mathcal{C}_3$. It is easy to verify that the resulting partition into (at most) three chains has the properties in the Claim. ∎

It is clear also that at each step, either X or Y disappears from the system (using the notation of the proof of the claim). So the sum of the sizes of the largest members of $\mathcal{C}_2$ and $\mathcal{C}_3$ decreases in each step. This gives a bound of $O(n)$ on the number of steps. ∎

Reference

M. Grötschel, L. Lovász, A. Schrijver, *Geometric Algorithms and Combinatorial Optimization,* Springer-Verlag, Berlin-Heidelberg-New York-Tokyo 1988.

C. A. J. Hurkens

Department of Econometrics
Tilburg University
Tilburg, The Netherlands

László Lovász

Department of Computer Science
L. Eötvös University
Budapest, Hungary H-1088

and *Department of Computer Science*
Princeton University
Princeton, NJ 08544.

Alex Schrijver

Department of Econometrics
Tilburg University
Tilburg, The Netherlands

and *Mathematical Centre*
Amsterdam, The Netherlands

Éva Tardos

Department of Computer Science
L. Eötvös University
Budapest, Hungary H-1088

and *Department of Mathematics*
Massachusetts Institute of Technology
Cambridge, MA 02139, USA