

The Complexity of Maximum Matroid-Greedoid Intersection^{*}

Taneli Mielikäinen and Esko Ukkonen

Department of Computer Science, P.O. Box 26
FIN-00014 University of Helsinki, Finland

Abstract. The maximum intersection problem for a matroid and a greedoid, given by polynomial-time oracles, is shown *NP*-hard by expressing the satisfiability of boolean formulas in 3-conjunctive normal form as such an intersection. Also the corresponding approximation problem is shown *NP*-hard for certain approximation performance bounds. This is in contrast with the maximum matroid-matroid intersection which is solvable in polynomial time by an old result of Edmonds.

1 Introduction

A set system (S, F) where S is a finite set (the *domain* of the system) and F is a collection of subsets of S is a *matroid* if

- (M1) $\emptyset \in F$;
- (M2) If $Y \subseteq X \in F$ then $Y \in F$;
- (M3) If $X, Y \in F$ and $|X| > |Y|$ then there is an $x \in X \setminus Y$ such that $Y \cup \{x\} \in F$.

A *greedoid* is a set system (S, F) that satisfies (M1) and (M3).

In applications a matroid or a greedoid is given by an oracle, i.e., by a deterministic algorithm that answers the question whether X belongs to F for any $X \subseteq S$.

Many combinatorial problems can be formulated using matroids or greedoids (see e.g. [5, 6]). The seminal example is the maximum matching problem in bipartite graphs. Each instance of the problem can be represented as the intersection of two matroids. The maximum matching corresponds to the largest set in the intersection.

To be able to consider in the matroid-greedoid framework general combinatorial problems which have infinitely many instances we introduce families of matroids and greedoids that have uniform polynomial-time representation. More formally, let $\mathcal{F} = \{(S_h, F_h)_{h \in H}\}$ be a possibly infinite set of matroids or greedoids. Then \mathcal{F} is said to be given by a *uniform polynomial-time oracle* if there is an algorithm O , that when given h and some $X \subseteq S_h$ answers whether or not $X \in F_h$. The algorithm O runs in polynomial time in $|S_h|$.

^{*} A work supported by the Academy of Finland.

Let $\mathcal{F} = \{(S_h, F_h)_{h \in H}\}$ and $\mathcal{G} = \{(S_h, G_h)_{h \in H}\}$ be two such families given by uniform polynomial-time oracles. Note that the index set H is the same for both and for given h , both have the same domain S_h .

The *maximum intersection problem* for \mathcal{F} and \mathcal{G} is to find, given an index $h \in H$, a set $X \in F_h \cap G_h$ such that $|X|$ is maximum.

Edmonds [3] gave the first polynomial-time solution for the intersection problem in the case that both \mathcal{F} and \mathcal{G} are families of matroids. In this paper we consider the obvious next step, namely the intersection of families of matroids and greedoids.

In Section 2 we show, by reduction from the 3SAT, that the maximum intersection problem for a matroid family and a greedoid family, given by uniform polynomial-time oracles, is *NP*-hard. In Section 3 we modify the above construction to show that the maximum matroid-greedoid intersection is not approximable within a factor $|S_h|^{1-\epsilon}$ for any fixed $\epsilon > 0$ and its weighted version, the maximum weight matroid-greedoid intersection, is not approximable within $2^{|S_h|^k}$ for any fixed $k > 0$, unless $P = NP$.

2 NP-hardness

Recall that the *NP*-complete problem *3-satisfiability* (3SAT) is, given a boolean formula h in 3-conjunctive normal form (3CNF), to decide whether or not there is a truth assignment Z for the variables of h such that $h(Z) = \text{true}$.

We construct the instance $(S_h, F_h), (S_h, G_h)$ of matroid-greedoid intersection that corresponds to h as follows. Let h contain n different boolean variables. Then S_h contains symbols $t_1, f_1, \dots, t_n, f_n$. The symbols $t_1, f_1, \dots, t_n, f_n$ will be used to encode truth assignments: t_i encodes that the i th variable is *true* and f_i that it is *false*.

The subset collection F_h consists of all subsets of S_h that contain at most one of symbols t_i, f_i for $i = 1, \dots, n$. It is immediate, that (S_h, F_h) satisfies the matroid properties (M1), (M2) and (M3).

The subset collection G_h consists of two groups. The first group A consists of all subsets X of S_h such that $|X| \leq n$ and $X \cap \{t_n, f_n\} = \emptyset$. The second group B consists of the sets that represent a truth assignment that satisfies h . Such a set is of size n and contains one element from each t_i, f_i .

To verify that (S_h, G_h) is a greedoid, first note that (M1) is obviously true. To verify (M3), let $X, Y \in G_h$ such that $|X| > |Y|$.

1. If $|X| < n$ then X and Y must belong to group A. Hence for any element $x \in X \setminus Y$, set $Y \cup \{x\}$ belongs to group A and hence to G_h .
2. If $|X| = n$ and $|X \setminus Y| = 1$ then $Y \cup (X \setminus Y) = X$, i.e., property (M3) holds.
3. In the remaining case $|X| = n$ and $|X \setminus Y| > 1$. As $X \setminus Y$ contains at least two elements and no set of G_h contains both t_n and f_n , at least one element $x \in X \setminus Y$ must be different from t_n, f_n . Then $Y \cup \{x\}$ belongs to group A.

The matroid-greedoid intersection $F_h \cap G_h$ contains a set X such that $|X| = n$ if and only if the group B in the definition of G_h is non-empty, that is, if and

only if h is satisfiable. As such a set X is also the largest in $F_h \cap G_h$, we have shown:

Lemma 1. *Boolean formula h is satisfiable if and only if the maximum element in $F_h \cap G_h$ for matroid (S_h, F_h) and greedoid (S_h, G_h) is of size n where n is the number of variables of h .*

The above construction yields a matroid family $\mathcal{F} = \{(S_h, F_h)_{h \in 3\text{CNF}}\}$ and a greedoid family $\mathcal{G} = \{(S_h, G_h)_{h \in 3\text{CNF}}\}$. Both have a uniform polynomial-time oracle for checking membership in F_h and G_h : The only nontrivial task of the oracle is to verify when a truth assignment satisfies a given formula h , but this is doable in polynomial time in $|h|$ using well-known techniques. As $|h| = O(n^3)$ for a 3CNF formula h and $|S_h| = 2n$, the running time of the oracle is in fact polynomial in $|S_h|$, as required.

It follows from Lemma 1 that our construction is a polynomial-time reduction of 3SAT to the maximum matroid-greedoid intersection problem. Therefore we have the following.

Theorem 1. *The maximum intersection problem for a matroid family and a greedoid family that are given by uniform polynomial-time oracles is NP-hard.*

Also the *maximum weight* matroid-greedoid intersection is NP-hard since maximum matroid-greedoid intersection is its special case. In this problem one should find, given integer weights $w(x)$ for $x \in S_h$, a set $X \in F_h \cap G_h$ such that $\sum_{x \in X} w(x)$ is maximum.

3 Inapproximability

As the maximum matroid-greedoid intersection problem is a maximization problem whose exact solution turned out to be NP-hard, it is of interest to see whether or not an *approximation algorithm* with a performance guarantee is possible. An approximation algorithm would find an element in the intersection of the matroid and the greedoid which is not necessarily the largest one.

Following the standard approach (see e.g. [1, 2]), we say that maximization problem is *polynomial-time approximable within r* where r is a function from \mathbb{N} to \mathbb{N} if there is a polynomial-time algorithm that finds for each instance x of the problem a feasible solution with value $c(x)$ such that

$$\frac{c_{Max}(x)}{c(x)} \leq r(|x|)$$

where $c_{Max}(x)$ is the largest possible value (the optimal value) of a feasible solution of x . The performance ratio of such an approximation algorithm is bounded by the performance guarantee r .

Theorem 2. *The maximum intersection problem for a matroid family and a greedoid family with domains $\{S_h : h \in H\}$, given by uniform polynomial-time oracles, is not polynomial-time approximable within $|S_h|^{1-\epsilon}$ for any fixed $\epsilon > 0$, unless $P = NP$.*

Proof. Assume that for some $\epsilon > 0$, the maximum matroid-greedoid intersection is polynomial-time approximable within $|S_h|^{1-\epsilon}$. We show that then we can solve 3SAT in polynomial time.

As in the proof of Lemma 1, let h again be a boolean formula with n variables in 3-conjunctive normal form. Now set S_h contains in addition to the truth assignment symbols $t_1, f_1, \dots, t_n, f_n$ also some indicator elements $p_i, 1 \leq i \leq I(\epsilon)$. Here the number of indicators, $I(\epsilon)$, depends on ϵ as will be shown below. The indicators are needed for padding the elements of the matroid and the greedoid such that the maximum intersection becomes for a satisfiable h sufficiently larger than for a non-satisfiable h .

We again construct a matroid (S_h, F_h) and (S_h, G_h) as follows.

The subset collection F_h contains all subsets of S_h that do not contain both t_i and f_i for any $1 \leq i \leq n$. It is again clear, that (S_h, F_h) satisfies properties (M1) and (M2). As regards (M3), let $X, Y \in F_h$ such that $|X| > |Y|$. If there is some indicator x in $X \setminus Y$, then $Y \cup \{x\} \in F_h$. Otherwise X must contain more truth assignment symbols than Y . Then there must be index i such that either t_i or f_i , call it x , belongs to X but neither of t_i and f_i belongs to Y . Then $Y \cup \{x\} \in F_h$. Thus (S_h, F_h) is a matroid.

The subset collection G_h consists of three groups. Group A and B are exactly same as in the construction of Lemma 1. Hence the sets in groups A and B do not contain any indicator elements. Group C consists of the sets of size n in groups A and B, padded with indicators in all possible ways. That is, if $X \in A$ or $X \in B$ such that $|X| = n$ and Q is a non-empty subset of $\{p_1, \dots, p_{I(\epsilon)}\}$, then $X \cup Q$ belongs to group C.

To verify that (S_h, G_h) is a greedoid, property (M1) clearly holds. To verify (M3), let $X, Y \in G_h, |X| > |Y|$ and consider the following cases.

1. If $|Y| < n$ then there is a truth assignment symbol $x \in X \setminus Y$ such that $Y \cup \{x\}$ belongs to group A or to group B as shown in the proof of Lemma 1.
2. If $|Y| \geq n$ then there is there is an indicator $x \in X \setminus Y$ and thus $Y \cup \{x\}$ belongs to group C.

By our construction, the boolean formula h is satisfiable if and only if the largest element in $F_h \cap G_h$ is of size $|S_h| - n = I(\epsilon) + n$. Otherwise the size of the largest element is at most $n - 1$.

Let now $I(\epsilon) = (2n)^{1/\epsilon} - 2n$. Thus $|S_h| = (2n)^{1/\epsilon}$. To test the satisfiability of h we use the approximation algorithm to find a approximately largest element of $F_h \cap G_h$. Let c be the size of this element. If h is not satisfiable then certainly $c < n$. On the other hand, if h is satisfiable, then the largest element of $F_h \cap G_h$ is of size $|S_h| - n$. Therefore

$$\frac{|S_h| - n}{c} \leq |S_h|^{1-\epsilon}.$$

But then

$$c \geq \frac{|S_h| - n}{|S_h|^{1-\epsilon}} \geq \frac{|S_h|}{2|S_h|^{1-\epsilon}} = \frac{|S_h|^\epsilon}{2} = n$$

where the second inequality follows from that $|S_h| \geq 2n$. Hence $c \geq n$ if h is satisfiable and $c < n$ if it is not. We have a polynomial-time satisfiability test because $I(\epsilon)$ is a polynomial in n and hence in $|h|$ when ϵ is fixed, and therefore the matroid family $\{(S_h, F_h)_{h \in 3\text{CNF}}\}$ and the greedoid family $\{(S_h, G_h)_{h \in 3\text{CNF}}\}$ can be represented by uniform oracles whose run times are polynomial in $|S_h|$, hence in $|h|$. \square

Theorem 3. *The maximum weight intersection problem for a matroid family and a greedoid family with domains $\{S_h : h \in H\}$, given by uniform polynomial-time oracles, is not polynomial-time approximable within $2^{|S_h|^k}$ for any fixed $k > 0$, unless $P = NP$.*

Proof. The construction is similar to the proof of Theorem 2.

Domain S_h consists of the truth assignment symbols and only one indicator element. The indicator has weight $n2^{|S_h|^k} - n$ while all the truth assignment symbols have weight equal to 1. Then the formula h is satisfiable if and only if the heaviest element in $F_h \cap G_h$ has weight $n2^{|S_h|^k}$. Otherwise the heaviest element weights less than n .

Let c be the weight obtained by an approximation algorithm that is assumed to approximate the optimal solution within $2^{|S_h|^k}$. If h is not satisfiable then $c < n$. Otherwise

$$\frac{n2^{|S_h|^k}}{c} \leq 2^{|S_h|^k}$$

which means that $c \geq n$.

We again have a polynomial-time satisfiability test because $|S_h| = 2n + 1$ and therefore the weights have polynomial-size (binary) representations in $|h|$ and hence the weighted set S_h can be constructed in polynomial time in $|h|$. Moreover, the uniform oracles for the families of (S_h, F_h) and (S_h, G_h) can be made to run in polynomial time in $|h|$ as before. \square

References

1. G. Ausiello, P. Crescenzi, G. Gambosi, V. Kann, A. Marchetti-Spaccamela, and M. Protasi. *Complexity and Approximation: Combinatorial Optimization Problems and Their Approximability Properties*. Springer-Verlag, 1999.
2. P. Crescenzi, V. Kann, R. Silvestri, and L. Trevisan. Structure in approximation classes. *SIAM Journal on Computing* 28 (1999) 1759–1782.
3. J. Edmonds. Minimum partition of a matroid into independent subsets. *Journal of Research of the National Bureau of Standards* 69B (1965) 67–72.
4. M. R. Garey, and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman and Company, 1979.
5. B. Korte, and J. Vygen. *Combinatorial Optimization: Theory and Algorithms*. Springer-Verlag, 2000.
6. C. H. Papadimitriou, and K. Steiglitz. *Combinatorial Optimization: Algorithms and Complexity*. Dover Publications, 1998.