

Static analysis of memory manipulations by abstract interpretation

Algorithmics of tropical polyhedra, and application to abstract interpretation

Xavier ALLAMIGEON

In this thesis, we define a static analysis by abstract interpretation of memory manipulations. It is based on a new numerical abstract domain, which is able to infer program invariants involving the operators \min and \max . This domain relies on tropical polyhedra, which are the analogues of convex polyhedra in tropical algebra. Tropical algebra refers to the set $\mathbb{R} \cup \{-\infty\}$ endowed with \max as addition and $+$ as multiplication.

This abstract domain is provided with sound abstract primitives, which allow to automatically compute over-approximations of semantics of programs by means of tropical polyhedra. Thanks to them, we develop and implement a sound static analysis inferring \min - and \max -invariants over the program variables, the length of the strings, and the size of the arrays in memory.

In order to improve the scalability of the abstract domain, we also study the algorithmics of tropical polyhedra. In particular, a tropical polyhedron can be represented in two different ways, either internally, in terms of extreme points and rays, or externally, in terms of tropically affine inequalities. Passing from the external description of a polyhedron to its internal description, or inversely, is a fundamental computational issue, comparable to the well-known vertex/facet enumeration or convex hull problems in classical algebra. It is also a crucial operation in our numerical abstract domain.

For this reason, we develop two original algorithms allowing to pass from an external description of tropical polyhedra to an internal description, and vice versa. They are based on a tropical analogue of the double description method introduced by Motzkin *et al.* We show that they outperform the other existing methods, both in theory and in practice. The cornerstone of these algorithms is a new combinatorial characterization of extreme elements in tropical polyhedra defined by means of inequalities: we establish that the extremality of an element amounts to the existence of a strongly connected component reachable from any node in a directed hypergraph. We also show that the latter property can be checked in almost linear time in the size of the hypergraph.

Moreover, in order to have a better understanding of the intrinsic complexity of tropical polyhedra, we study the problem of determining the maximal number of extreme points in a tropical polyhedron. In the classical case, this problem is addressed by McMullen upper bound theorem. We prove that the maximal number of extreme points in the tropical case is bounded by a similar result. We introduce a class of tropical polyhedra appearing as natural candidates to be maximizing instances. We establish lower and upper bounds on their number of extreme points, and show that the McMullen type bound is asymptotically tight when the dimension tends to infinity and the number of inequalities defining the polyhedra is fixed.

Finally, we experiment our tropical polyhedra based static analyzer on programs manipulating strings and arrays. These experimentations show that the analyzer successfully determines precise properties on memory manipulations, and that it scales up to highly disjunctive invariants which could not be computed by the existing methods.

The implementation of the algorithms and abstract domains on tropical polyhedra developed in this work is available in the library TPLib.¹

¹Tropical Polyhedra Library, <http://www.penjili.org/tplib.html>.

Analyse statique de manipulations de mémoire par interprétation abstraite

Algorithmique des polyèdres tropicaux, et application à l'interprétation abstraite

Xavier ALLAMIGEON

Dans cette thèse, nous introduisons une technique d'analyse statique de manipulations de mémoire par interprétation abstraite. Elle repose sur un nouveau domaine abstrait numérique capable de déterminer des invariants faisant intervenir les opérateurs min et max. Ce domaine se fonde sur les polyèdres tropicaux, qui sont les analogues des polyèdres convexes en algèbre tropicale. L'algèbre tropicale désigne l'ensemble $\mathbb{R} \cup \{-\infty\}$ muni des opérations max et + comme lois additive et multiplicative.

Le domaine abstrait est équipé de primitives abstraites sûres, ce qui permet de calculer automatiquement des sur-approximations de la sémantique de programmes par des polyèdres tropicaux. Grâce à cela, nous développons et implémentons un analyseur statique capable de déterminer des invariants à base de min et max et portant sur les variables de programmes, la longueur de chaînes de caractères, et la taille des tableaux en mémoire.

Afin d'améliorer les performances du domaine abstrait, nous étudions également l'algorithmique des polyèdres tropicaux. En particulier, un polyèdre tropical peut être représenté de deux manières possibles, soit de façon interne, à l'aide de points et rayons extrêmes, soit de manière externe, à partir d'inégalités affines au sens tropical. Passer d'une représentation à l'autre est un problème fondamental, comparable au problème bien connu en algèbre classique d'énumération des sommets ou des faces d'un polyèdre convexe. C'est également une opération cruciale dans notre domaine abstrait numérique.

C'est pourquoi nous développons deux nouveaux algorithmes calculant une représentation interne d'un polyèdre tropical à partir d'une description externe, et inversement. Ces algorithmes reposent sur un analogue tropical de la méthode de la double description introduite par Motzkin *et al.*. Nous montrons qu'ils dépassent tant en théorie qu'en pratique les méthodes existantes. La clé de voûte de ces algorithmes est une nouvelle caractérisation combinatoire des éléments extrêmes de polyèdres tropicaux définis par des inégalités. Celle-ci s'exprime comme l'existence d'un puit dans un certain hypergraphe orienté. Nous montrons également que cette dernière propriété peut être déterminée en un temps presque linéaire en la taille de l'hypergraphe.

En outre, afin de mieux comprendre la complexité intrinsèque des polyèdres tropicaux, nous étudions le nombre maximal de points extrêmes dans un polyèdre tropical. Dans le cas classique, la solution de ce problème est connue grâce à un résultat dû à McMullen (*upper bound theorem*). Nous montrons que le nombre de points extrêmes est borné par une quantité analogue dans le cas tropical. Nous introduisons une classe de polyèdres tropicaux apparaissant comme des candidats naturels pour être des instances maximisantes. Nous établissons des bornes inférieures et supérieures sur le nombre de leurs points extrêmes, et montrons que la borne à la McMullen est atteinte asymptotiquement lorsque la dimension tend vers $+\infty$ et que le nombre d'inégalités définissant les polyèdres est fixé.

Enfin, nous expérimentons notre analyseur statique sur des programmes manipulant des chaînes de caractères et des tableaux. Ces expérimentations montrent que l'analyseur statique parvient à déterminer des propriétés précises sur les manipulations de mémoire. Elles montrent aussi que l'analyseur est capable de passer à l'échelle, en calculant des invariants hautement disjonctifs et inaccessibles à partir des méthodes existantes.

Une implémentation des algorithmes et des domaines abstraits développés dans ces travaux est disponible dans la bibliothèque TPLib¹.

1. Tropical Polyhedra Library, <http://www.penjili.org/tplib.html>.

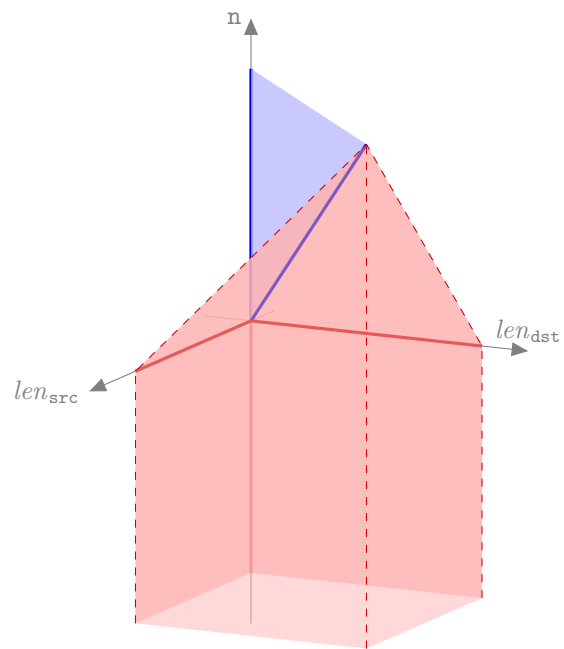


Figure 1: A tropical polyhedron representing the invariant of the C function memcpy