

SMT-Solving Induction Proofs of Inequalities

Ali K. Uncu^{1,2}, James H. Davenport² and Matthew England³

¹Austrian Academy of Science, Johann Radon Institute for Computational and Applied Mathematics, Linz, Austria

²University of Bath, Faculty of Science, Department of Computer Science, Bath, UK

³Coventry University, Research Centre for Computational Science and Mathematical Modelling, Coventry, UK

Abstract

This paper accompanies a new dataset of non-linear real arithmetic problems for the SMT-LIB benchmark collection. The problems come from an automated proof procedure of Gerhold–Kauers, which is well suited for solution by SMT. The problems of this type have not been tackled by SMT-solvers before. We describe the proof technique and give one new such proof to illustrate it. We then describe the dataset and the results of benchmarking. The benchmarks on the new dataset are quite different to the existing ones. The benchmarking also brings forward some interesting debate on the use/inclusion of rational functions and algebraic numbers in the SMT-LIB.

Keywords

Inequalities, Induction Proofs, Satisfiability Modulo Theories, Computer Algebra, Rational Functions

1. Introduction

Satisfiability Modulo Theories (SMT) fuses powerful modern SAT solvers with software from specialised theory domains to tackle satisfiability problems where the logical atoms are statements in that domain. The SMT-LIB [1] defines a common language for SMT-solvers to use and maintains a set of benchmarks organised according to the various theory domains.

In many cases, the algorithms for those domains have been traditionally implemented in computer algebra systems (although as described in [2], such algorithms require adaptation before they can be used efficiently within SMT). There is continuing progress in algorithms for such domains, driven in part by the connections built between symbolic computation and the satisfiability checking communities, by the SC-Square project [2] and others.

One of the SMT theory domains most closely aligned with symbolic computation, and the domain we consider, is QF_NRA (Quantifier-Free Nonlinear Real Arithmetic). In this case the solver seeks to answer a question on the existence of real variables x_1, \dots, x_k to solve a logical formula in which each atom is a (potentially non-linear) polynomial constraint. There is a significant number of benchmarks in the SMT-LIB for this domain, however, there are relatively few sources of these examples, and vast majority come from a single theorem-proving application.

⁷th International Workshop on Satisfiability Checking and Symbolic Computation, 12 Aug 2022, Haifa Israel


✉ aku21@bath.ac.uk (A. K. Uncu); masjhd@bath.ac.uk (J. H. Davenport); Matthew.England@coventry.ac.uk (M. England)

🌐 <http://akuncu.com> (A. K. Uncu); <https://people.bath.ac.uk/masjhd> (J. H. Davenport);

<https://matthewengland.coventry.domains> (M. England)

🆔 0000-0001-5631-6424 (A. K. Uncu); 0000-0002-3982-7545 (J. H. Davenport); 0000-0001-5729-3420 (M. England)

© 2021 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

In this paper we describe a new collection of examples which we have contributed, originating from inductive proof of some inequalities. We seek to (a) broaden the QF_NRA benchmark set to allow for better development of solvers; and (b) encourage further additions from other new application domains by demonstrating how well solvers can do on such problems.

1.1. SMT for QF_NRA

In the QF_NRA domain solvers tackle satisfiability problems whose atoms are of the form $p\sigma 0$ where $p := p(x_1, x_2, \dots, x_k) \in \mathbb{Q}[x_1, x_2, \dots, x_k]$ is a polynomial in variables x_1, \dots, x_k with rational coefficients, and $\sigma \in \{>, <, \geq, \leq, =, \neq\}$.

Such problems are usually tackled in the Lazy SMT paradigm where a SAT-solver proposes solutions to the logical structure which are then checked for validity in the theory domain: deciding whether the corresponding set of polynomial constraints can be satisfied together. In other words, we check the following, where p_i and σ_i are defined as above:

$$\exists x_1, x_2, \dots, x_k (p_1\sigma_1 0 \wedge p_2\sigma_2 0 \wedge \dots \wedge p_n\sigma_n 0). \quad (1)$$

Note that this is fully conjunctive and will involve only a subset of the atoms in the original formula. The answer will either be a set of assignments for the variables (x_1, \dots, x_k) to *witness* the existence, or a confirmation that this is *unsatisfiable*. The unsatisfiable confirmation confirms that there is no single point in \mathbb{R}^k space that could satisfy the relations.

There are many ways of tackling this conjunction. One expensive but well established method is to calculate the Cylindrical Algebraic Decomposition (CAD) [3] of the variable space to be sign-invariant for the polynomials p_i , and then check the regions for existence of such a point. CAD was developed for the more general problem of Quantifier Elimination (QE) over the reals. Such a heavy procedure is clearly far more work than required, at least when the problem is satisfiable and we need find only a single point. An adaptation of CAD for this purpose was presented in [4] to allow for early termination and repeated calls.

Another approach is to re-purpose the theory of CAD so that it better aligns to the satisfiability methodology, of searching for a model and learning from conflict. This is the approach taken in the Cylindrical Algebraic Covering method of [5] and the NLSAT algorithm of [6]. Both build model solutions gradually variable by variable: the former learns by identifying that a model cannot be extended and using CAD projection to rule out an interval around the top dimension of the current model; the latter learns by building an entire CAD cell whose description defines new atoms of the logical formula.

In addition to these CAD based methods there is a variety of incomplete methods implemented which may will not always give a solution, but when they do are often far faster, e.g. Virtual Term Substitution [7] or incremental linearization [8]. Although these techniques all root from formal arguments, there is not a clear answer as to which method is best in general or on a particular instance, and so there is a need for careful benchmarking. Also, different problems sets have their own *flavour*, and so benchmarking on them individually is valuable too. This set, rooting from mathematical inequalities, is internally diverse. Although the number of the variables do not go beyond 6 at any time, the natural appearance of rational functions in some problems and the degree of a single variable spiking in some other problems can be considered as some different characteristics within this sample set.

1.2. Problems from inequalities

We will discuss a family of QF_NRA problems new to the SMT-LIB, and to the best of the authors' knowledge, not tackled before using SMT.

In 2005, Gerhold and Kauers presented an algorithm that attempts induction proofs with great success [9]. Their original formulation, and later Kauers' `ProveInequality` function in the Mathematica package `SumCracker` [10], uses CAD to make these proofs. This method and the implementation has been used successfully applied in many works to automatically prove combinatorics and special function related inequalities [9, 10, 11, 12, 13, 14, 15, 16]. These applications utilised computer algebra, but the underlying algorithm is actually asking a sequence of satisfiability questions that terminates with a positive answer if it can be shown that a logical structure of the form (1) is unsatisfiable.

Later in the paper we will sketch the main ideas behind this procedure by proving the following result. Let k and n be positive integers, and let $x_1, x_2, \dots, x_n \in \mathbb{R}^+$. Then if $x_1 + x_2 + \dots + x_n = n$, we have that

$$\sum_{i=1}^n x_i^k \frac{x_i^4 + x_i^2 + 1}{x_i^2 + x_i + 1} \geq n \prod_{i=1}^n x_i. \quad (2)$$

This problem appeared as a generalization of a Monthly Problem in the American Mathematical Monthly [17]. Until now the inequality (2) only had a human proof. That proof required using an inequality on carrying positive exponents inside a finite sum, followed by an inequality of Chebyshev, and then an inequality between arithmetic-geometric means. However, by following the Gerhold–Kauers method, we will prove a stronger version of this inequality without any prior knowledge and with minimal human interference.

1.3. New dataset

We have put together a dataset of 300 problems in the SMT-LIB language, derived from the application of the Gerhold–Kauers method to examples given in [9, 10, 11, 13] and (2). This dataset and other files used in our benchmarks are available at [18]. These have been submitted for inclusion in the 2022 release of the SMT-LIB.

Unlike other problem sets in the SMT-LIB, a quarter of these problems have constraints that involve rational functions instead of purely polynomial constraints. This new characteristic also calls into question how best to pre-process such objects. There are at least two ways of clearing any non-constant denominators to get an equivalent expression with polynomial constraints, and then there is the handling of zeros of any denominators. For every problem involving a rational function, we generated two equivalent problems in polynomials, where we handled the denominators in a different way. We will observe different solver behaviour depending on the conversion method used.

1.4. Plan of the paper

The organization of this paper is as follows. In Section 2, we will briefly introduce the Gerhold–Kauers method by using it on the new example. Then in Section 3 we will discuss different

ways of clearing denominators before presenting our dataset and some benchmarking results for it in Sections 4 and 5. We finish with some conclusions.

2. The Gerhold–Kauers method to use CAD for Induction Proofs

2.1. General idea

A mathematical induction proof — at its core — is a finite set of initial conditions together with the logical structure of the problem implying the correctness of the next step. We require a discrete parameter, say n , for the indexing of the initial conditions. Let, our general claim, ϕ be a logical formula (or a collection of logical formulae) in n (without loss of generality $n \in \mathbb{Z}^+$) and possibly other variables. We would like to prove the correctness of ϕ by complete induction of n . The construction of ϕ is done through a difference field construction: we will brush over that and invite any interested readers to visit Gerhold and Kauers’ original paper [9].

It is not clear from which starting point and with how many (if any) initial conditions one can gather satisfactory knowledge to prove the induction step. Hence one needs to start with the selection of a t and r (both being most likely 1) and attempt to show

$$\psi \wedge (\phi \wedge s(\phi) \wedge \cdots \wedge s^{r-1}(\phi)) \Rightarrow s^r(\phi), \quad (3)$$

where ψ is a conjunction of all known assumptions on the parameters, and $s^k(\phi)$ is the k -th shift (in n) of the original statement ϕ . Let $[\phi]_k$ be the explicit evaluation of ϕ_n at the instance $k \in \mathbb{Z}_{\geq 0}$. If we can also confirm that each initial condition $[\phi]_k$ for $k = t, \dots, t + r - 1$ holds together with (3) then we get an induction proof for all $n \geq t$.

In their paper [9], Gerhold and Kauers decide to attempt refuting (3) by instead attempting to deduce that

$$\psi \wedge (\phi \wedge s(\phi) \wedge \cdots \wedge s^{r-1}(\phi)) \wedge \neg s^r(\phi) \quad (4)$$

would be false. Moreover, they do it in an efficient and iterative way by checking $[\phi]_n$ ’s at each step and extending r if (4) is still satisfiable for some selection of variables. A possible variable selection might be far away from the original problem, however, such an instance triggers the algorithm to iterate (pick a larger r) and repeat the process.

2.2. A New Proof of (2)

As an initial step towards the proof of (2), let us start with the claim that, for $x_i > 0, i = 1, \dots, n$, if $\sum_{i=1}^n x_i = n$, then

$$\prod_{i=1}^n x_i \leq 1. \quad (5)$$

The case $n = 1$ is obvious. The difference ring construction would define x in the place of x_n . Then we define another three variables X, Y , and Z and their shifts in n : $s(X) = X + 1$, $s(Y) = Y + s(x)$, and $s(Z) = Zs(x)$, where $s(\cdot)$ is the shift of the variable inside (the next element in the sequence) and $s(x)$ is kept as a new variable added to the problem. Here X

simulates n , Y simulates the sum $\sum_{i=1}^n x_i$ and Z simulates the product $\prod_{i=1}^n x_i$. Assuming $t = r = 1$, the logical statement we are trying to refute is

$$\begin{aligned} & (x > 0 \wedge s(x) > 0 \wedge X = Y \wedge s(X) = s(Y)) \wedge (Z \leq 1) \wedge \neg(s(Z) \leq 1) \\ & = (x > 0 \wedge s(x) > 0 \wedge X = Y \wedge X + 1 = Y + s(x)) \wedge (Z \leq 1) \wedge (Zs(x) > 1), \end{aligned} \quad (6)$$

together with the initial condition check $[Z]_1 = 1 \leq 1$. It is very easy to see that the first logical sentence implies $s(x) = 1$ and that together with the last two clauses yields a contradiction. Therefore, confirming the claim for the initial conditions $[X]_1, [Y]_1$ and $[Z]_1$, the induction step holds and our claim is true for generic $n \geq 1$.

Similarly, we can prove

$$\sum_{i=1}^n \frac{x_i^4 + x_i^2 + 1}{x_i^2 + x_i + 1} \geq n, \quad (7)$$

under the assumptions $x_i > 0$ for $i = 1, \dots, n$ and $\sum_{i=1}^n x_i = n$. To simulate the sum on the left-hand side of (7) and its iterations, is given as $[\hat{Z}]_1 = x_1^2 - x_1 + 1$ and $s(\hat{Z}) = \hat{Z} + (s(x)^4 + s(x)^2 + 1)/(s(x)^2 + s(x) + 1)$. For the proof, one can put together the logical formula to be refuted, similar to (6), with X, Y and the new variable \hat{Z} and easily show that to be contradiction.

In the same vein, we can prove

$$\sum_{i=1}^n (x_i - 1) \frac{x_i^4 + x_i^2 + 1}{x_i^2 + x_i + 1} \geq 0, \quad (8)$$

following similar steps as above with the new variable \tilde{Z} , where $s(\tilde{Z}) = \tilde{Z} + (s(x) - 1)(s(x)^4 + s(x)^2 + 1)/(s(x)^2 + s(x) + 1)$.

The next step needed to prove (2) is to show

$$\sum_{i=1}^n x_i^{j-1} (x_i - 1) \frac{x_i^4 + x_i^2 + 1}{x_i^2 + x_i + 1} \geq 0, \quad (9)$$

for any $j \geq 1$. For any fixed positive integer j this can be done with a logical solver for QF_NRA. In this example the logical formula to evaluate becomes

$$x > 0 \wedge s(x) > 0 \wedge X = Y \wedge X + 1 = Y + s(x) \wedge \bar{Z} \geq 0 \wedge \bar{Z} > s(x)^j (s(x)^3 - 2s(x)^2 + 2s(x) + 1),$$

where \bar{Z} simulates the sum on the left-hand side of (9). From the logical structure, we can deduce that $s(x) = 1$ and later conclude that $(\bar{Z} \geq 0) \wedge (\bar{Z} < 0)$ would yield a contradiction and prove (9). However, this is only possible to achieve on a computer for explicitly chosen positive integers j . Otherwise, since the input would not be a collection of polynomials/rational functions, we could not apply CAD (or other QE methods).

This is where we need a human touch to prove (9) for arbitrary j using (8). The case of all $x_i = 1$ is trivially true. Otherwise, since $\sum_{i=1}^n x_i = n$, there exists at least one $a \in \{1, 2, \dots, n\}$ such that $x_a > 1$ and at least one $b \in \{1, 2, \dots, n\}$ such that $x_b < 1$. Let A be the set of all such indices between 1 and n such that $x_a > 1$. Similarly, let B be the set of all indices of all x_b such

that $0 < x_b < 1$. A and B are both finite sets since all the indices are chosen from 1 to n . For non-empty A and B , notice that $x_a^{j-1} \geq x_a > 1$ and $0 < x_b^{j-1} \leq x_b$ for any $a \in A$ and $b \in B$. So by multiplying the i -th summand of (8) with x_i^{j-1} , we either keep the summand the same (if $j = 1$) or increase the contribution of the positive terms if $i \in A$. Similarly, by multiplying the i -th summand of (8) with x_i^{j-1} , we either keep the summand the same (if $j = 1$) or shrink the contribution of the negative terms if $i \in B$. Since (8) is assumed to hold and this modification to the summands increases the positive contribution while decreasing the negative contribution of the summands, the inequality (9) is holds for any positive integer j as well.

If we sum (9) over $j = 1, \dots, k$ we get

$$\sum_{i=1}^n (x_i^k - 1) \frac{x_i^4 + x_i^2 + 1}{x_i^2 + x_i + 1} \geq 0.$$

Adding (7) to this yields

$$\sum_{i=1}^n x_i^k \frac{x_i^4 + x_i^2 + 1}{x_i^2 + x_i + 1} \geq n, \quad (10)$$

under the same assumptions of the original problem (2): $k, n \in \mathbb{Z}^+$, $x_i > 0$ and $x_1 + \dots + x_n = n$.

Finally, using inequality (5) on the right-hand side of (10), we prove (2). One highlight is that we proved these inequalities without any prior knowledge of any mathematical inequalities. Moreover, note that (10) is a sharper inequality than (2).

2.3. Implementation

An implementation of this method has been completed by Kauers: the `ProveInequality` procedure in the `SumCracker` Mathematica package [10] does the identification of the variables to be included and their shifts automatically, and ships the statement to be refuted directly to the CAD implementation of Mathematica. The proof of (8), in Mathematica, then turns into a single command:

```
ProveInequality[SUM[(x[k]-1)(1+x[k]^2+(x[k])^4)/(1+x[k]+(x[k])^2),
{k, 1, n}]>=0, Using->{x[n]>0, SUM[x[k], {k, 1, n}]==n}, Free->{x}, Variable->n]
```

which terminates with an answer in milliseconds.

2.4. Suitability for SMT

A key-point to stress is that Gerhold-Kauers method actually generates and answers satisfiability problems, in the form (1), with all the existential quantifiers hidden but there. At each attempt, the Gerhold-Kauers method checks the initial conditions and it looks to see if the refuted induction-step (4) is unsatisfiable. Furthermore, any known information about the pieces of ϕ can be tagged alongside of (4) and get fed to the CAD machinery to further restrict the search space and get the desired unsatisfiable answer. Their implementation simply used the CAD implementation of Mathematica to see in which regions (4) can be satisfied. However, neither where this formulae is satisfied, nor the cylindrical structure of the decomposition to

refute satisfiability, is essential to the problem. Thus CAD could be safely replaced with any SMT solver that can tackle QF_NRA and may benefit from the incremental data-structures their internal machinery usually possesses.

3. Appearance and Handling of Rational Functions

In the automated proof sketches of (7) and (8), we already saw the possibility of rational functions arising. The shifts of the variables \hat{Z} and \tilde{Z} , which were used to simulate the sum and their shifts, introduced a rational function in the induction hypothesis clauses. In those examples above, the rational function could be simplified to a polynomial expression, but this is not true in general. We see that the satisfiability problems coming from this method naturally introduces rational functions.

Rational functions inclusion and handling in satisfiability problems seems to be a somewhat sensitive topic in the SMT community and there are discussions about whether and how best to allow SMT-LIB to include rational functions in its language. While we leave that discussion for later, we will mention some possible pre-processing ways that can help us remedy the situation in a mathematically consistent way.

Assume that we are given a satisfiability problem where one of the clauses includes multi-variate rational functions after simplifications. For example

$$\frac{P(\mathbf{x})}{Q(\mathbf{x})} \sigma \frac{F(\mathbf{x})}{G(\mathbf{x})}$$

where $P, Q, F, G \in \mathbb{Q}[\mathbf{x}]^1$, $\gcd(P, Q) = \gcd(F, G) = 1$, and $\sigma \in \{>, <, \geq, \leq, =, \neq\}$. We can simplify this problem to 0, by subtraction followed by any simplifications which handle a problem of the form

$$\frac{f(\mathbf{x})}{g(\mathbf{x})} := \frac{P(\mathbf{x})G(\mathbf{x}) - F(\mathbf{x})Q(\mathbf{x})}{Q(\mathbf{x})G(\mathbf{x})} \sigma 0, \quad (11)$$

with $\gcd(f(\mathbf{x}), g(\mathbf{x})) = 1$.

Handling rational functions in a mathematically consistent way is straightforward when the relation is an equation or an inequation. If σ is $=$ or \neq we can simplify (11) as

$$f(x) \sigma 0 \wedge g(x) \neq 0.$$

There are two equivalent formulations of (11) in the polynomial language when σ is an inequality. One way is to avoid any sign considerations for the denominator polynomial $g(\mathbf{x})$ and multiply both sides of the relation (11) with its square. However, the poles of the original rational function should not be forgotten and be reflected in the outcome. This way the equivalent formulation of (11) is

$$f(\mathbf{x})g(\mathbf{x}) \sigma 0 \wedge g(\mathbf{x}) \neq 0. \quad (12)$$

The disadvantage of this method is the likely rise in the degrees of the variables. When $f(\mathbf{x})$ and $g(\mathbf{x})$ are multiplied together some variables can get out of reach of the degree dependent QE techniques, such as virtual term substitutions [7].

¹In this discussion, the rational field \mathbb{Q} can be replaced by the reals \mathbb{R} , but here we restrict ourselves to stay within the limits of the SMT-LIB language.

Another possibility is to consider the sign of $g(\mathbf{x})$ and split the problem into two pieces driven by the guards $g(\mathbf{x}) > 0$ and $g(\mathbf{x}) < 0$. The statement we get using this approach is

$$(g(\mathbf{x}) > 0 \wedge f(\mathbf{x}) \sigma 0) \vee (g(\mathbf{x}) < 0 \wedge 0 \sigma f(\mathbf{x})). \quad (13)$$

Although this time the degrees of the variables stay lower, the size of the logical problem has grown. If the satisfiability problem starts with n clauses including rational functions this problem would split it to a disjunction of 2^n statements.

We suggest that the handling of rational functions be left to the SMT solvers. If users make this choice they may inadvertently disadvantage a solver. We elaborate on this later in §5.4.

4. Dataset and Benchmarking

4.1. Dataset

We went through most examples given in [9, 10, 11, 13] and equations (5), (7), and (8) (the parts of the proof of (2) which can be proven automatically) to describe them as non-linear arithmetic satisfiability problems in the SMT-LIB language, creating a dataset of 300 new SMT-LIB benchmarks. This was done by translating the original CAD calls of the `ProveInequality` procedure to SMT-LIB using the `SMTLIB` package in Maple [19]. This package already identifies the existence of a rational function in a clause and adds the `denominator-is-nonzero` clause to the problem.

When problems involved a rational function in these calls then we also created two additional equivalent formulations of the problem, by clearing out the denominators in the basic way (12) and in the disjunctive way (13) as demonstrated above. The original examples with only polynomials and these two later polynomial-made examples were submitted in the call for new benchmarks for the 2022 SMT Competition².

In one group of our problems (the `SignPattern` problems from [9]) the original problems contains a $\sqrt{5}$. Mathematica’s CAD implementation could handle these, but algebraic numbers are not permitted within the definition of `QF_NRA` in general. Thus we introduced the clauses $y^2 = 5 \wedge y > 0$ to bring the problem into `QF_NRA`. We note that the iteration of these clauses created high exponents for the pseudo-variable y : this was left for the solvers to handle.

4.2. Solvers

The SMT solvers used in this benchmarking are `Z3` (v 4.8.8) [20] and `Yices` (v 2.6.4) [21], which both utilise the `NLSAT` algorithm [6] for `QF_NRA`; and `CVC5-Linux` [22] (v 1.0.0) which uses the `Cylindrical Algebraic Coverings` algorithm for `QF_NRA` [5]. These three were selected as the strongest performers on `QF_NRA` in recent years.

We also evaluated some of the tools in Computer Algebra Systems, Maple and Mathematica: the versions used are Maple 2022 and Mathematica 12.0.0. In Maple, we used the `RegularChains:-QuantifierElimination` command [23] to eliminate the calls in (1) format. We also used the soon to be released Maple package `QuantifierElimination` [24].

²<https://smt-comp.github.io/2022/>

The former utilises CAD constructed via triangular sets technology and the latter CAD with Lazard projection interlaced with cubic virtual term substitutions. In Mathematica, we used the CAD command [25], as was used by Kauers' `ProveInequality` originally; the QE function `Resolve` which utilises also other QE methods such as virtual term substitution; and the meta-solver `Reduce` which makes use of Mathematica's other solving tools in addition.

Besides Maple's `RegularChains` implementation, all the other functions and solvers accepted inputs with rational functions.

4.3. Benchmarking Methodology

In general we followed the methodology explained in [26]. All benchmarks were undertaken on a computer running openSUSE Leap 15.3 with 16GB of RAM and an Intel Xeon CPU E5-1650 v3 running at 3.50 GHz. All functions were given 20 minutes to attempt each of these problems.

We display our results visually using survival plots. To produce these we first solve each problem q_i , noting the time t_i (up to our chosen threshold of 1200 seconds). Then for each solver we sort the t_i into increasing order, discard the timed-out problems, and plot points $(k, \sum_{i=1}^k t_i)$. This approach does not guarantee that the same problems are returned with an answer in the chosen threshold from different implementations. However, for the cumulative problem set survival plots effectively encapsulate a lot of information about the success rate and the total time taken to solve for the successful answers.

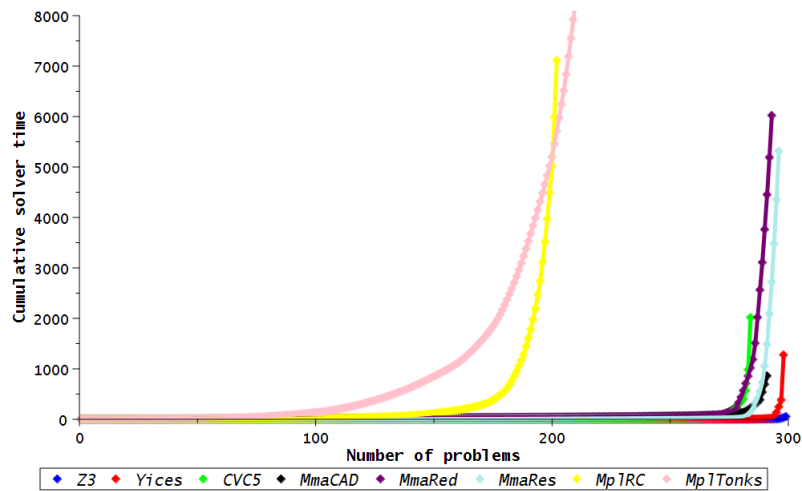


Figure 1: Survival plot of benchmarks with the time scale up to 8000 seconds.

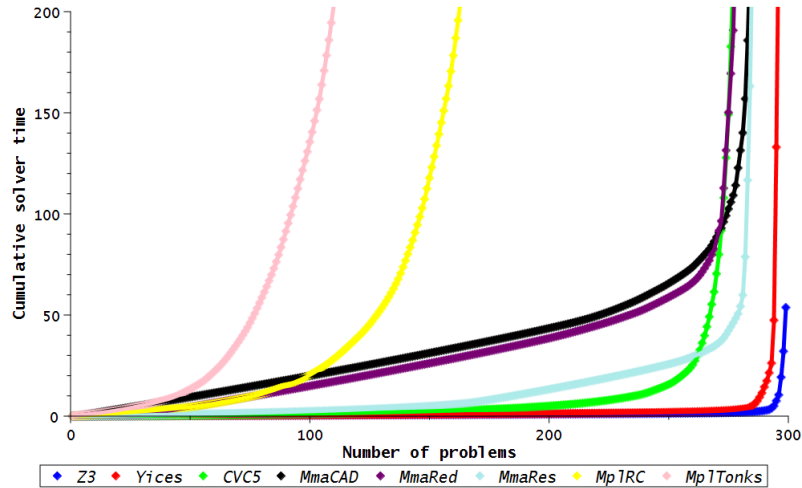


Figure 2: Survival plot of benchmarks with the time scale up to 200 seconds.

5. Analysis

5.1. Overall performance

Figures 1 and 2 show the survival plots on different scales of the solver time. It is clear that, for this dataset, Z3 is superior: it timed-out only in one example. It is then followed by Yices (timed out on 2 examples), then the various implementations in Mathematica (4, 7 and 9 time outs for Resolve, Reduce and CAD respectively) followed by CVC5 (16 timeouts). The two Maple functions performed far less well: RegularChains did not accept rational functions at all, and both functions took far longer amounts of time to reach their conclusions. We do note that Maple has also available direct calls to Z3 via `SMTLIB:-Satisfiable`.

It is not surprising that the SMT-solvers excel on satisfiability problems compared to full QE implementations using CAD. Satisfiability is a sub-problem of QE with lower complexity. What is surprising is that Mathematica's QE is competitive with the SMT-solvers. The local projections used [27] may offer similar benefits to the model based SMT searches of [6], [5]; and we also note Mathematica has access to sophisticated logical simplification routines [28]. The DEWCAD project is working now to address the shortcomings in Maple by building in Maple dedicated algorithms for satisfiability problems similar to those implemented in SMT [29].

5.2. Algebraic Number Substitutions

We suspect some timeout problems are due to a failure to substitute for algebraic numbers in the problems described in the final paragraph of Section 4.1. In the SignPattern problems discussed there, the exponent of a variable y (introduced to do bookkeeping of $\sqrt{5}$) in polynomials gets very high. The difficulty of this problem lowers immensely if a system can identify and at least utilize the second degree equational constraint $y^2 = 5$. We believe Z3

does this substitution and lowers the cost of calculations immensely. We also believe that most implementations would have been able to answer these questions in a matter of seconds if they were to do this preprocessing before asking for the satisfiability. For example, the CAD implementation of Mathematica can answer `SignPattern_Lemma4a-f` examples from the dataset in about half a minute to a minute each, but when the $y^2 = 5$ is used and the degree of y is reduced to a only linear powers, these numbers drop to under 10 seconds each. We note that CVC5 performs particularly badly on these problems.

5.3. Curiosities

On our examples, CVC5 is outperformed by Z3 and Yices overall, and it is outperformed by Mathematica for large compute times (see Figure 1). This is somehow at odds with the SMT Competition results of 2021³. In addition to the algebraic numbers issue above, the poor performance may also be down to the presence of rational functions. Although CVC5 accepts rational functions in its input, we do not think that it does much preprocessing. The SMTLIB Maple package that was used to translate these examples to SMT-LIB language adds clauses to keep the denominators non-zero. Therefore, we never experience CVC5 encountering a division by zero and quitting or throwing an exception. However, we observe that it takes a much longer time, and even times-out on occasion for the problems with rationals where others do not.

Mathematica's `Resolve` and `Reduce` solve slightly more problems than its CAD procedure. But at one point CAD overtakes `Reduce`. I.e. sometimes the cost of the extra considerations `Reduce` does hinders its success (see Figure 3a). This indicates that there is scope for a better meta-algorithm to decide when `Reduce` resorts to CAD.

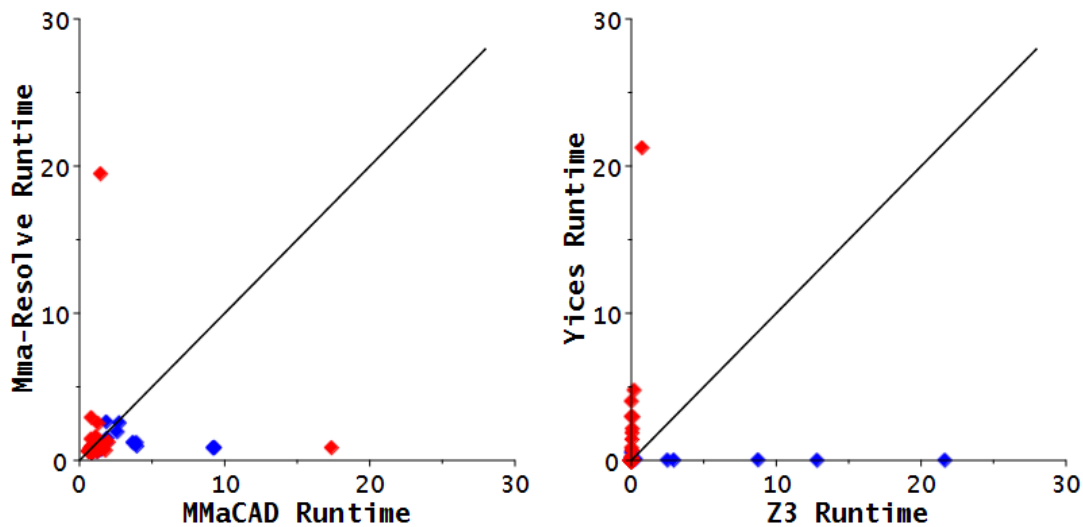
Another curious observation between front runners Z3 and Yices is that Z3 is actually slower than Yices on the SAT problems. But since Z3 was much faster to identify that a problem is unsatisfiable and there were more UNSAT problems in the dataset, it gained victory overall. See Figure 3b. To the best of our knowledge Z3 and Yices both rely on NLSAT as the underlying theory algorithms. So it suggests the difference is in either the heuristics inside that, or the other incomplete methods tried first.

One can also observe from Figure 1 that `QuantifierElimination` Maple package is cumulatively slower than `RegularChains:-QuantifierElimination` on this problem set, overtaking eventually it only due to its handling of rational functions. Nevertheless, even when only considering examples with polynomial entries, `QuantifierElimination` is faster on 14% of the examples. These examples might be where the virtual term substitution can be applied to make a significant difference.

5.4. Effects of Denominator Clearance

Finally, we observe that *how* we clear denominators affects conclusions about the best solver. We focus our attention to the rational function calls with their denominators cleared using (12) or (13). For polynomial calls acquired by (12), Z3 is still the best solver but the second best solver changes hands from Yices to Mathematica's `Resolve` function both in time and in the number

³<https://smt-comp.github.io/2021/results.html>



(a) Mathematica’s CAD in comparison to its `Resolve`

(b) Z3 in comparison to Yices

Figure 3: Scatter plots of noticeable time differences. Runtimes measured in seconds. Blue data points are SAT examples and red are UNSAT.

of problems solved. However, when we focus on rational call images under (13) denominator clearance, we see that Mathematica `Resolve` solves one extra problem than Z3 and Yices.

6. Conclusions

Our first conclusion is that SMT solvers do very well on most of the examples in this problem set, outperforming computer algebra systems designed to tackle broader QE problems (Section 5.1). We also observe that the solvers perform differently on this new dataset than they did on the `QF_NRA` section of the SMT-LIB overall in the most recent competition. This shows us that it offers some new characteristics, and they continue the much needed diversification of the `QF_NRA` benchmarks. They also exposes some interesting strengths and weaknesses of solvers that the developers may find interesting to study (Section 5.3).

Our second conclusion is that is a need for further work on the SMT-LIB language for `QF_NRA` to decide how best to deal with rational functions. We observed that the choice of how we clear denominators effects conclusions over the best solver (Section 5.4). At the moment, the SMT-LIB seems to suggest the user should make this choice, but would it not be more appropriate for the solver to do it? It clearly introduces a scope for new heuristics that researchers can explore. The authors support that rational function calls be included in the SMT-LIB language, with a semantics that implies the denominator be non-zero. But this must be defined so that the meaning is mathematically consistent and avoid getting conflicting results from solvers.

Our third conclusion is in a similar vein: we suggest the SMT-LIB considers allowing the use of algebraic numbers in the input (Section 5.2). There are 21 examples under the `SignPattern` header, where we replaced $\sqrt{5}$ with a variable y and two added clauses that $y^2 = 5 \wedge y > 0$.

Not only that, we let the iterations to grow the degree of ys and left the preprocessing to the solvers. On this dataset the usually competitive CVC5 performed poorly. But if we exclude this 21 problem subset, then among the polynomial calls CVC5 beats Mathematica methods in cumulative time. Allowing algebraic numbers in the problem statement stretches the definition of polynomial (usually assumed to have rational coefficients). But many of the theory algorithms such as CAD can handle these and they can be encoded into actual polynomials. Having the user do the encoding can make the problems artificially harder for solvers.

Acknowledgements

The authors would like to thank Manuel Kauers for providing a modified version of his `ProveInequality` function which exposed the CAD calls, easing the creation of the dataset.

All three authors are supported by the EPSRC DEWCAD Project (*Pushing Back the Doubly-Exponential Wall of Cylindrical Algebraic Decomposition*): JHD and AU by grant number EP/T015713/1 and ME by grant number EP/T015748/1. AU also acknowledges partial support from FWF grant P-34501N.

References

- [1] C. Barrett, P. Fontaine, C. Tinelli, The Satisfiability Modulo Theories Library (SMT-LIB), Online Resource, URL: <http://smtlib.cs.uiowa.edu/>, 2016.
- [2] E. Ábrahám, J. Abbott, B. Becker, A. Bigatti, M. Brain, B. Buchberger, A. Cimatti, J. Davenport, M. England, P. Fontaine, S. Forrest, A. Griggio, D. Kroening, W. Seiler, T. Sturm, SC²: Satisfiability checking meets symbolic computation, in: M. Kohlhase, M. Johansson, B. Miller, L. de Moura, F. Tompa (Eds.), *Intelligent Computer Mathematics: Proceedings C1CM 2016*, volume 9791 of *Lecture Notes in Computer Science*, Springer International Publishing, 2016, pp. 28–43. URL: https://doi.org/10.1007/978-3-319-42547-4_3.
- [3] G. Collins, Quantifier elimination for real closed fields by cylindrical algebraic decomposition, in: *Proceedings of the 2nd GI Conference on Automata Theory and Formal Languages*, Springer-Verlag (reprinted in the collection [30]), 1975, pp. 134–183. URL: https://doi.org/10.1007/3-540-07407-4_17.
- [4] G. Kremer, E. Ábrahám, Fully incremental CAD, *Journal of Symbolic Computation* 100 (2020) 11–37. URL: <https://doi.org/10.1016/j.jsc.2019.07.018>.
- [5] E. Ábrahám, J. Davenport, M.England, G. Kremer, Deciding the consistency of non-linear real arithmetic constraints with a conflict driven search using cylindrical algebraic coverings, *Journal of Logical and Algebraic Methods in Programming* 119 (2021) 100633. URL: <https://doi.org/10.1016/j.jlamp.2020.100633>.
- [6] D. Jovanovic, L. de Moura, Solving non-linear arithmetic, in: B. Gramlich, D. Miller, U. Sattler (Eds.), *Automated Reasoning: 6th International Joint Conference (IJCAR)*, volume 7364 of *Lecture Notes in Computer Science*, Springer, 2012, pp. 339–354. URL: https://doi.org/10.1007/978-3-642-31365-3_27.
- [7] T. Sturm, Thirty years of virtual substitution: Foundations, techniques, applications, in: *Proceedings of the 2018 ACM International Symposium on Symbolic and Algebraic*

- Computation, ISSAC '18, ACM, 2018, pp. 11–16. URL: <http://doi.org/10.1145/3208976.3209030>.
- [8] A. Cimatti, A. Griggio, A. Irfan, M. Roveri, R. Sebastiani, Incremental linearization for satisfiability and verification modulo nonlinear arithmetic and transcendental functions, *ACM Transactions on Computational Logic* 19 (2018) 19:1–19:52. URL: <http://doi.acm.org/10.1145/3230639>.
- [9] S. Gerhold, M. Kauers, A procedure for proving special function inequalities involving a discrete parameter, in: *Proceedings of the 2005 International Symposium on Symbolic and Algebraic Computation, ISSAC '05*, ACM, 2005, pp. 156–162. URL: <https://doi.org/10.1145/1073884.1073907>.
- [10] M. Kauers, SumCracker: A package for manipulating symbolic sums and related objects, *Journal of Symbolic Computation* 41 (2006) 1039–1057. URL: <https://doi.org/10.1016/j.jsc.2006.06.005>.
- [11] M. Kauers, Algorithms for Nonlinear Higher Order Difference Equations, Ph.D. thesis, Johannes-Kepler University, Linz (RISC), 2005. URL: <http://www.algebra.uni-linz.ac.at/people/mkauers/publications/kauers05c.pdf>.
- [12] A. Dixit, V. Moll, V. Pillwein, A Hypergeometric Inequality, *Annals of Combinatorics* 20 (2016) 65–72. URL: <https://doi.org/10.1007/s00026-015-0294-5>.
- [13] M. Kauers, V. Pillwein, When Can We Detect that a P-Finite Sequence is Positive?, in: S. Watt (Ed.), *Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation (ISSAC '10)*, 2010, pp. 195–202. URL: <https://doi.org/10.1145/1837934.1837974>.
- [14] M. Kauers, P. Paule, A computer proof of Moll's log-concavity conjecture, *Proceedings of the AMS* 135 (2007) 3847–3856. URL: <https://doi.org/10.1090/S0002-9939-07-08912-5>.
- [15] S. Gerhold, M. Kauers, A computer proof of Turán's inequality, *Journal of Inequalities in Pure and Applied Mathematics* 7 (2006). URL: <https://www.emis.de/journals/JIPAM/article659.html?sid=659>.
- [16] M. Kauers, How to use cylindrical algebraic decomposition, *Seminaire Lotharingien de Combinatoire* 65 (2011) B65a. URL: <https://www.emis.de/journals/SLC/>.
- [17] C. Zheng, Z. Zhou, A generalization of a monthly problem, *The American Mathematical Monthly* 125 (2018) 922–922. URL: <https://doi.org/10.1080/00029890.2018.1460976>.
- [18] A. K. Uncu, J. H. Davenport, M. England, SMT-Solving Induction Proofs of Inequalities Benchmarking Repository, 2023. URL: <https://doi.org/10.5281/zenodo.7777297>. doi:10.5281/zenodo.7777297.
- [19] S. Forrest, Integration of smt-lib support into maple, in: M. England, V. Ganesh (Eds.), *Proceedings of the 2nd International Workshop on Satisfiability Checking and Symbolic Computation (SC² 2017)*, number 1974 in *CEUR Workshop Proceedings*, 2017. URL: <http://ceur-ws.org/Vol-1974/>.
- [20] N. Bjørner, L. de Moura, L. Nachmanson, C. Wintersteiger, Programming Z3 (tutorial lecture notes), in: J. Bowen, Z. Liu, Z. Zhang (Eds.), *Engineering Trustworthy Software Systems: 4th International School (SETSS 2018)*, volume 11430 of *Lecture Notes in Computer Science*, Springer International Publishing, 2019, pp. 148–201. URL: https://doi.org/10.1007/978-3-030-17601-3_4.
- [21] B. Dutertre, Yices 2.2, in: A. Biere, R. Bloem (Eds.), *Computer Aided Verification*, Springer International Publishing, 2014, pp. 737–744. URL: <https://doi.org/10.1007/>

978-3-319-08867-9_49.

- [22] H. Barbosa, C. Barrett, M. Brain, G. Kremer, H. Lachnitt, M. Mann, A. Mohamed, M. Mohamed, A. Niemetz, A. Nötzli, A. Ozdemir, M. Preiner, A. Reynolds, Y. Sheng, C. Tinelli, Y. Zohar, cvc5: A versatile and industrial-strength smt solver, in: D. Fisman, G. Rosu (Eds.), *Tools and Algorithms for the Construction and Analysis of Systems*, Springer International Publishing, Cham, 2022, pp. 415–442. URL: https://doi.org/10.1007/978-3-030-99524-9_24.
- [23] C. Chen, M. Moreno Maza, Quantifier elimination by cylindrical algebraic decomposition based on regular chains, *Journal of Symbolic Computation* 75 (2016) 74–93. URL: <https://doi.org/10.1016/j.jsc.2015.11.008>.
- [24] Z. Tonks, A poly-algorithmic quantifier elimination package in Maple, in: J. Gerhard, I. Kotsireas (Eds.), *Maple in Mathematics Education and Research*, volume 1125 of *Communications in Computer and Information Science*, Springer International Publishing, 2020, pp. 330–333. URL: https://doi.org/10.1007/978-3-030-41258-6_13.
- [25] A. Strzeboński, Cylindrical algebraic decomposition using validated numerics, *Journal of Symbolic Computation* 41 (2006) 1021–1038. URL: <https://doi.org/10.1016/j.jsc.2006.06.004>.
- [26] M. Brain, J. Davenport, A. Griggio, Benchmarking solvers, SAT-style, in: M. England, V. Ganesh (Eds.), *Proceedings of the 2nd International Workshop on Satisfiability Checking and Symbolic Computation (SC² 2017)*, number 1974 in *CEUR Workshop Proceedings*, 2017. URL: <http://ceur-ws.org/Vol-1974/>.
- [27] A. Strzeboński, Cylindrical algebraic decomposition using local projections, *Journal of Symbolic Computation* 76 (2016) 36–64. URL: <https://doi.org/10.1016/j.jsc.2015.11.018>.
- [28] C. Brown, A. Strzeboński, Black-box / white-box simplification and applications to quantifier elimination, in: *Proceedings of the 2010 International Symposium on Symbolic and Algebraic Computation, ISSAC '10*, ACM, 2010, pp. 69–76. URL: <https://doi.org/10.1145/1837934.1837953>.
- [29] R. Bradford, J. H. Davenport, M. England, A. Sadeghimanesh, A. Uncu, The DEWCAD project: Pushing back the doubly exponential wall of cylindrical algebraic decomposition, *ACM Commun. Comput. Algebra* 55 (2022) 107–111. URL: <https://doi.org/10.1145/3511528.3511538>.
- [30] B. Caviness, J. Johnson, *Quantifier Elimination and Cylindrical Algebraic Decomposition*, Texts & Monographs in Symbolic Computation, Springer-Verlag, 1998. URL: <https://doi.org/10.1007/978-3-7091-9459-1>.