Some Methods of Problem Solving in Elementary Geometry

Thomas C. Hales

Math Department, University of Pittsburgh
E-mail: hales@pitt.edu

Abstract

Many elementary problems in geometry arise as part of the proof of the Kepler conjecture on sphere packings. In the original proof, most of these problems were solved by hand. This article investigates the methods that were used in the original proof and describes a number of other methods that might be used to automate the proofs of these problems. A companion article presents the collection of elementary problems in geometry for which automated proofs are sought. This article is a contribution to the Flyspeck project, which aims to give a complete formal proof of the Kepler conjecture.

1. Introduction

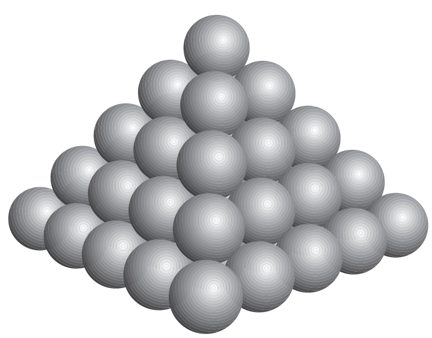

Figure 1. The face-centered cubic packing

In 1998, Sam Ferguson and I gave a proof of the Kepler Conjecture, which asserts that no packing of congruent balls has density greater than the face-centered cubic packing (Fig. 1). For a history of the problem see [17]. The proof relies on a significant number of computer calculations. The published text of the proof is nearly 300 pages [6].

As a result of the difficulties encountered in checking the correctness of the Kepler conjecture, I have become interested in formal theorem proving, as a way of checking complex computer proofs. In 2003, I announced a project called Flyspeck, whose purpose is to give a completely formal proof of the Kepler Conjecture [11]. The name '**F**lys**p**ec**k**' comes as an expansion of the acronym 'F*P*K', which stands for the '**F**ormal **P**roof of **K**epler.' The name is apt, because 'Flyspeck' can mean to scrutinize or examine in detail.

Since then, a number of researchers have become involved in this project. One major component of this formalization project has been completed by G. Bauer and T. Nipkow [1]. Significant progress in toward a verification of the computer code used in the proof of the Kepler conjecture has been reported by S. Obua and R. Zumkeller [15], [19].

In a companion paper, I present a large collection of elementary problems in geometry [9]. These problems have been extracted from the proof of the Kepler conjecture [6]. These problems share a number of features. Each problem can be expressed as a problem in the elementary theory of the real numbers. Each problem involves a small number of variables. Each problem deals with a configuration of points in $\mathbb{R}^3$.

We hope that the proofs of these theorems will be mechanized. This collection is reminiscent of a number of other collection of problems in geometry whose proofs have been successfully mechanized, such as [3], [18], [8].

In this article, we explore some of the methods that might be used to mechanize the proofs of the problems in this collection. We do not limit ourselves to traditional methods of automated theorem proving such as Gröbner basis methods and Wu's algorithm. Instead, we explore a variety of other methods that seem particularly promising. To describe the methods, we apply them to a couple of simple test problems. We begin with a description of a method in discrete geometry that was used in the original proof of the Kepler conjecture. It seems that this method, like many of the other methods described in this article, is open to automation.

This research has been supported by NSF grant 0503447.

2. A Sample Problem

To describe the wide variety of methods that can be used to solve a typical problem in elementary geometry, we introduce a simple test problem. This test problem is in fact rather trivial. We keep the problem simple so that a wide variety of different solutions can be used. We are not so much interested in the problem as in understanding the methods of solution, how generally the methods can be applied to related problems, and how fully they can be automated.

Let $|x| = \sqrt{x_1^2 + x_2^2 + x_3^2}$ be the Euclidean norm of $x = (x_1, x_2, x_3)$ in $\mathbb{R}^3$.

Problem 1. *Let p, p', q, q' be points in $\mathbb{R}^3$. Suppose that the following inequalities hold.*

- A_1*:* $|p - p'| < \sqrt{8}$, $|q - q'| < \sqrt{8}$.
- A_2*:* $2 \le |p - q|, |p - q'|, |p' - q|, |p' - q'|$.

Under these assumptions on distances, show that the line segment extending from p to p' does not meet the line segment extending from q to q'.

If the line segments cross, the two line segments can be visualized as the diagonals of a quadrilateral. The assumptions of the problem put lower bound constraints on the edges of the quadrilateral and upper bound constraints on the diagonals of the quadrilaterals. See Figure 2.

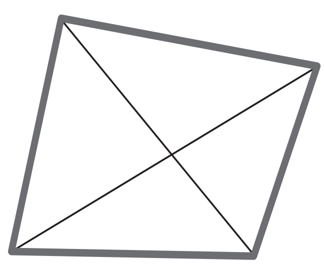

Figure 2. A quadrilateral with long sides cannot have short diagonals (Problem 1).

In the Kepler conjecture, the four points p, p', q, q' represent the centers of balls of radius 1. The lower bounds in Assumption A_2 express the constraint that the balls in the packing do not overlap one another. The upper bounds in Assumption A_1 come from a truncation that is used to limit the number of balls under consideration. The number $\sqrt{8}$ arises as the shortest distance between the centers of two non-touching balls in an optimal sphere packing in three dimensions.

If we relax the bounds in Assumption A_1 to make them weak inequalities, then a square of side 2 satisfies the weakened constraints, and yet the two diagonals meet. Thus, at least one of the inequalities of Assumption A_1 must be strict, in order for the conclusion to hold. A positive solution to this problem implies that a square cannot be deformed in the plane in a way that decreases the diagonals and is non-decreasing in the side lengths.

Some of the methods are better illustrated with the following dual problem. It comes by reversing the signs of the inequalities in the original problem. This particular dual problem does not arise in the solution of the Kepler conjecture, but many closely related problems do.

Problem 2. *Show that there are no points p, p', q, q' in $\mathbb{R}^3$ that satisfy the following inequalities:*

- B_1*:* $|p - p'| \ge \sqrt{8}$, $|q - q'| \ge \sqrt{8}$.
- B_2*:* $2 > |p - q|, |p - q'|, |p' - q|, |p' - q'|$.

If we allow the inequalities in Assumption B_2 to be weak, then a square with side length 2 satisfies the constraints.

3. Solution Methods

3.1. Discrete Geometry

In the original proof of the Kepler conjecture, problems such as Problem 1 were generally regarded as being too trivial to merit a proof. Perhaps they would get merit one sentence, saying that the extreme case is a square, which does not quite satisfy the constraints, because the inequalities are strict in Assumption A_1. Of course, in a formal proof, everything merits a proof.

Here is a solution of Problem 1, based on the methods and style of the original proof of the Kepler conjecture.

Proof. Assume for a contradiction, that Assumptions A_1 and A_2 hold, as well as the negation C of the conclusion: The segment from p to p' meets the segment from q to q' at point $v \in \mathbb{R}^3$. The intersection at v forces the points p, p', q, q' to lie in a plane. Without loss of generality we may assume that the points lie in $\mathbb{R}^2$. Note that v is not equal to any of the points p, p', q, q'. If for example, $v = p$, then we get the contradiction

$$\begin{aligned} 2\sqrt{2} \quad & > |q - q'| \\ & = |q - v| + |v - q'| = |q - p| + |p - q'| \\ & \ge 2 + 2 > 2\sqrt{2}. \end{aligned}$$

We will continuously deform the planar figure p, p', q, q' along a path in such a way that the constraints A_1 and A_2 continue to hold. As the points p, p', q, q' are deformed, the point of intersection v of the two lines, pp' and qq', continues to lie on the segment from p to p' and on the segment from q to q', because of the argument of the previous paragraph (which keeps v from "sliding off the end" of either segment). Thus, constraint C is automatically satisfied as we deform the planar figure.

Our deformations will be non-increasing in $|p - p'|$ and $|q - q'|$. This insures that the constraints A_1 hold under the deformations. To deform, assume for example that $|p - q| > 2$, then we move q slightly along the circle at fixed distance from q' until $|p-q| > 2$ and $|q-p'| > 2$. Then we move q slightly directly toward q' to decrease $|q-q'|$, while preserving the constraints A_1 and A_2. By compactness, we may continue in this way until all four constraints of A_2 are equalities:

$$|p-q| = |p'-q| = |p-q'| = |p'-q'| = 2.$$

This figure is a rhombus. Thus, there exist vectors v_1 and v_2 of length 2 such that

$$q = p + v_1, \quad q' = p + v_2, \quad p' = p + v_1 + v_2.$$

We obtain the desired contradiction:

$$\begin{aligned} 8+8 \quad &> |p-p'|^2 + |q-q'|^2 \\ &= |v_1+v_2|^2 + |v_1 - v_2|^2 \\ &= 2(|v_1|^2 + |v_2|^2) \\ &= 16. \end{aligned}$$

□

3.2. Quantifier Elimination

Problem 1 can be formulated as a problem in the elementary theory of the real numbers as follows:

$$\begin{aligned} \nexists \quad & p, p', q, q', s, t : \\ & 0 \le s \le 1 \ \wedge \quad 0 \le t \le 1 \ \wedge \\ & sp + (1-s)p' = tq + (1-t)q' \ \wedge \\ & |p-p'|^2 < 8 \ \wedge |q-q'|^2 < 8 \ \wedge \\ & |p-q|^2 \ge 4 \ \wedge |p-q'|^2 \ge 4 \ \wedge \\ & |p'-q|^2 \ge 4 \ \wedge |p'-q'|^2 \ge 4. \end{aligned}$$

There are quantifier elimination algorithms that allow us to decide the truth of any statement in the elementary theory of the reals. In particular, Problem 1 can be decided by such a decision procedure. There is a large mathematical literature on quantifier elimination over the reals. For a survey, see [13]. The critical point method presented in [2] seems most closely related to the other methods presented in this paper. There are various software implementations of quantifier elimination, such as QEPCAD. A proof-producing implementation is found in [12].

This particular problem has a single block of 14 quantifiers. It can be reduced to 10 quantifiers if the points are viewed in $\mathbb{R}^2$, rather than $\mathbb{R}^3$. If we apply rigid motions of the plane to make $p = 0$ and $p' = (x, 0)$, we reduce the number of quantifiers to 7. A more aggressive elimination of variables can bring the number of quantifiers down to 5.

All of the problems in our collection of problems in elementary geometry are of this nature. They can be expresses in the elementary theory of the real numbers. All of the problems in our elementary geometry collection involve a small number of variables (meaning at most twenty variables or so). Even if current software implementations of quantifier elimination are unable to solve problems of this size directly, these problems are certainly not hopelessly out of reach of automated solvers.

By way of contrast, the Kepler conjecture, when formulated as a statement in the elementary theory of the reals, involves hundreds of variables. See [5] for such a formulation. A direct assault on the Kepler conjecture, viewed as a statement in the elementary theory of the reals, is utterly hopeless.

3.3. Cayley-Menger Determinants

The points $p, p', q, q' \in \mathbb{R}^3$ are the vertices of a tetrahedron. The volume V of this tetrahedron is given by a classical formula, due to Cayley and Menger [10]. This formula gives the volume of a simplex in n dimensions. The formula in three dimensions takes the following simple form.

$$V = \frac{1}{12}\sqrt{\Delta(x_1, \ldots, x_6)},$$

where

$$\begin{aligned} &\Delta(x_1, \ldots, x_6) = \\ &\quad -x_2x_3x_4 - x_1x_3x_5 - x_1x_2x_6 - x_4x_5x_6 + \\ &\quad x_3(x_1 + x_2 - x_3 + x_4 + x_5 - x_6)x_6 + \\ &\quad x_2x_5(x_1 - x_2 + x_3 + x_4 - x_5 + x_6) + \\ &\quad x_1x_4(-x_1 + x_2 + x_3 - x_4 + x_5 + x_6), \end{aligned}$$

and where $x_1, \ldots, x_6$ are the squares of the lengths of the edges of the tetrahedron, arranged so that x_i and x_{i+3} correspond to opposite edges.

We show how to apply the method of Cayley-Menger determinants on Problem 2. Here is the formulation of Problem 2 in terms of edge lengths.

Lemma 3. *If $x_1, x_4 \ge 8$ and $x_2, x_3, x_5, x_6 < 4$, then*

$$\Delta(x_1, \ldots, x_6) < 0.$$

Hence there does not exist a tetrahedron with these squared edge lengths.

Proof. We calculate $\partial\Delta(x_1, \ldots, x_6)/\partial x_1$ to be

$$\begin{array}{lllll} -x_1x_4 & +x_2x_5 & +x_3x_6 & -x_3x_5 & -x_2x_6 \\ & & +x_4(-x_1+x_2+x_3) & +x_4(-x_4+x_5+x_6) & \\ < -64 & +16 & +16 & +0 & \\ < 0. & & & & \end{array}$$

Similarly, the x_4 partial is negative. Thus, Δ is bounded above by its value at $x_1 = x_4 = 8$. Similarly, we find that $\partial\Delta(8, x_2, x_3, 8, x_4, x_5, x_6)/\partial x_2$ is

$$\begin{aligned} x_5(4 - x_5) \quad & +x_5(8 - 2x_2) + 8(8 - x_3 - x_6) \\ & +4x_5 + x_3x_5 + x_3x_6 + x_5x_6 \\ & > 0. \end{aligned}$$

Thus,

$$\begin{aligned} \Delta(x_1, \ldots, x_6) \quad & \leq \Delta(8, x_2, x_3, 8, x_5, x_6) \\ & < \Delta(8, 4, 4, 8, 4, 4) \\ & = 0. \end{aligned}$$

□

The function Δ appears throughout the proof of the Kepler conjecture in arguments such as this one. The Cayley-Menger determinant for the five points in $\mathbb{R}^4$ is also quite useful in the proof of the Kepler conjecture. If the five vertices lie in $\mathbb{R}^3$, then the four-dimensional volume is zero. This gives a relationship among the ten pairs of distances between five points in $\mathbb{R}^3$. This relationship leads to a formula (which we call $\mathcal{E}$ in [6]) for the tenth distance as a function of the other nine. The function $\mathcal{E}$ appears throughout the proof.

3.4. Energy of Tensegrities

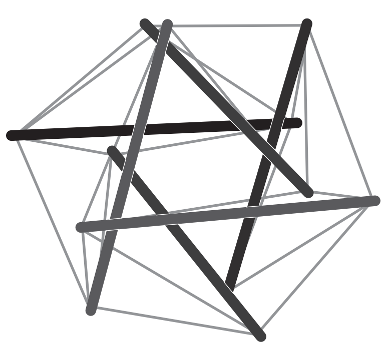

Figure 3. A tensegrity is a system of cables and struts.

The term tensegrity was coined by Buckminster Fuller to describe tension and compression in architecture [7]. R. Connelly pointed out to me several years ago that many of the geometrical lemmas in the proof of the Kepler conjecture can be formulated and solved with tensegrities. In the language of tensegrities, a *cable* is an upper bound constraint between two points in $\mathbb{R}^3$ such as the upper bound constraints B_2 in Problem 2. A *strut* is a lower bound constraint between two points in $\mathbb{R}^3$ such as the lower bounds constraints B_1 in Problem 2. Figure 3 shows the cables and struts of a tensegrity. Expressed in these terms, Problem 2 asserts that there does not exist a quadrilateral with the given cables running along the sides of the quadrilateral and struts across the two diagonals.

We describe one method to solve Problem 2.

Proof. A *stress* on a graph with straight edges and with vertices in the plane is a function ω from edges to real numbers satisfying the equilibrium condition:

$$\sum_{e \in E(v)} \omega(e)(v_e^* - v) = 0, \quad \text{for all vertices } v,$$

where $E(v)$ is the set of edges at v, and v_e^* is the endpoint of e opposite v. The stress is *proper* if $\omega(e) \geq 0$ on edges e with cables and $\omega(e) \leq 0$ on edges with struts. The graph with six edges, consisting of the edges of a square and its two diagonals, has a proper stress ω given by $+1$ on each side and -1 along each diagonal. See Figure 4.

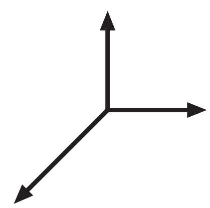

Figure 4. To be a stress in equilibrium, the weighted vector sum around each vertex must be zero.

From the proper stress we form an energy function. For any four points in the plane p, p', q, q' we define its energy to be

$$\begin{aligned} E(p, p', q, q') = & \\ & -|p - p'|^2 - |q - q'|^2 \\ & +|p - q|^2 + |q - p'|^2 + |p' - q'|^2 + |q' - p|^2. \end{aligned}$$

The coefficient of each square is the value of the stress on the corresponding edge. The energy is a quadratic form in the eight coordinates (p_1, p_2), $(p_1', p_2'), \ldots$ of p, p', q, q'. The energy can be written in the form ${}^tXMX$ for some 8×8 symmetric matrix M. This matrix is positive semidefinite (see [4, Theorem 5]). We can see this by a direct calculation of its eigenvalues: it has eigenvalue 0 with multiplicity 6 and eigenvalue 4 with multiplicity 2. Thus, the energy is always nonnegative.

On the other hand, if the figure satisfies the constraints B_1 and B_2 of Problem 2, then the energy is negative.

$$E(p, p', q, q') < -8 - 8 + 4 + 4 + 4 + 4 \leq 0.$$

This contradiction proves that there does not exist a figure satisfying the constraints of Problem 2. □

3.5. Sums of Squares and Semidefinite Programming

As I have learned from J. Harrison, P. Parrilo has developed powerful methods of proving polynomial inequalities over the real numbers [16]. The basic idea is that to prove that a polynomial f is nonnegative, it is enough to find polynomials $g_1, \ldots, g_k$ such that

$$f = g_1^2 + \ldots + g_k^2,$$

because the right-hand side is evidently nonnegative. There are many variations on the theme. For example, to prove an implication of polynomial inequalities

$$f_1 \geq 0 \wedge \cdots f_r \geq 0 \Rightarrow f \geq 0$$

it is enough to find an identity of the form

$$f = h_1 s_1 + \ldots + h_m s_m \tag{1}$$

where each h_i is a monomial $f_1^{a_{i1}} \cdots f_r^{a_{ir}}$ and each s_i is a sum of squares

$$s_i = g_{i1}^2 + \cdots g_{ik_i}^2.$$

The right-hand side of Equation 1 is evidently nonnegative if each polynomial f_i is nonnegative.

The procedures to search for a representation of f in the form of Equation 1 is related to positive semidefinite matrices and semidefinite programming. In fact, the term ${}^tXMX$, for any symmetric semidefinite matrix M, has a sum-of-squares decomposition

$${}^tXMX = \sum_i (\sqrt{\lambda_i}\,(v_i \cdot X))^2$$

for a complete set of eigenvalues and orthonormal eigenvectors (λ_i, v_i) of M.

As an example, we return to Connelly's solution of Problem 2 by energy minimization of a tensegrity. Since the energy is positive semidefinite, we may compute the eigenvectors of the corresponding symmetric matrix to find a sum-of-squares representation of the energy. This leads to the following solution of Problem 2.

Proof. Let $e_1 = (1,0)$, $e_2 = (0,1)$, and $g_i = (p + p' - q - q') \cdot e_i$. We have the polynomial identity

$$\begin{aligned} &(|p-q|^2 - 4) && +(|q-p'|^2 - 4) \\ & && +(|p'-q'|^2 - 4) + (|q'-p|^2 - 4) \\ =\ &(|p-p'|^2 - 8) && +(|q-q'|^2 - 8) + g_1^2 + g_2^2. \end{aligned}$$

Each term on the left is negative by the constraints of Problem 2. Each term on the right is a square or a nonnegative constraint of Problem 2. Thus the two sides cannot be equal and no configuration exists. □

3.6. Global Optimization

We may eliminate the existential variable v in Problem 1 by requiring $v = 0$, which can always be arranged by translating the figure. We write

$$p = av, \quad p' = -cv, \quad q = bw, \quad q' = -dw,$$

for some unit vectors v, w and nonnegative scalars a, b, c, d. Then we reformulate Problem 1 as a constrained global minimization problem.

Lemma 4. *The minimum of the function* $(a+c)^2 + (b+d)^2$ *is* 16*, when subject to the following constraints*

$$\begin{aligned} &|v| = |w| = 1, \\ &|av - bw| \geq 2, \quad |bw + cv| \geq 2, \\ &|cv - dw| \geq 2, \quad |dw + av| \geq 2, \\ &a, b, c, d \geq 0. \end{aligned}$$

This minimization clearly gives the solution to Problem 1 (in the negated form used in the discrete geometry solution).

To solve, we can formulate the constraints as Lagrange multipliers and write down the Kuhn-Tucker conditions. (I tried this in Mathematica and it did not terminate. However, the calculation can be done without trouble by hand: the analysis of Section 3.1 is essentially this method, viewed from a geometric perspective.)

There is no need to restrict ourselves to Lagrange multipliers in solving global optimization problems. Many of the problems in our collection of elementary geometry problems were solved in [6] by methods of constrained global optimization, including interval arithmetic, Taylor approximations with explicit error bounds, linear relaxation and linear programming methods, and so forth. Many of these methods are well-suited for formalization. We refer the reader to the survey article [14], for an overview of some of the methods that are available.

3.7. Other Methods

There are a number of methods for theorem proving in geometry, such as Wu's method and Gröbner basis methods. Some of the problems in our collection may be amenable to these methods. However, bear in mind that most of our problems contain inequality constraints.

4. The Collection of Elementary Geometry Problems

A collection of elementary problems in geometry appears in [9]. The hope is that by some combination of methods presented here, the proofs of these problems might be automated. None of the problems are hopelessly out of reach. In fact, they have all been solved by 'manual procedures' in [6].

We have used the following selection criteria for these problems. Each problem can be expressed in the elementary theory of the reals. The problems involve a small number of quantifiers and free variables. They can be described as configurations of points in $\mathbb{R}^3$. The problems do not require any of the technical definitions from [6] for their formulation. They do not require calculations of area or volume. Each problem is required as part of the published proof of the Kepler conjecture.

References

[1] G. Bauer, T. Nipkow, P. Schultz, Flyspeck I: Tame Graphs, in Automated Reasoning (IJCAR 2006), ed. U. Furbach and N. Shankar, Springer,LNCS,vol. 4130, 21–35, 2006.

[2] S. Basu, R. Pollack, M.-F. Roy, On the Combinatorial and Algebraic Complexity of Quantifier Elimination, J. ACM, 43 (6), 1002–1045, 1996.

[3] S.-C. Chou, X.-S. Gao, and J.-Z. Zhang, *Machine Proofs in Geometry: Automated Production of Readable Proofs for Geometry Theorems*, World Scientific, 1994.

[4] R. Connelly, Rigidity and Energy, Invent. math. 66, 11-33 (1982).

[5] L. Fejes Tóth, *Lagerungen in der Ebene auf der Kugel und im Raum*, second edition, Springer-Verlag, Berlin New York, 1972.

[6] S. Ferguson and T. Hales, Kepler conjecture, Disc. Comp. Geom, 36 (1), 1–265, 2006.

[7] B. Fuller, Portfolio and Art News Annual, No.4, 1961.

[8] H.-G. Gräbe, The SymbolicData GEO Records – A Public Repository of Geometry Theorem Proof Schemes, *Automated Deduction in Geometry*, LNCS 2930, Springer, Berlin, 2004.

[9] T. Hales, A Collection of Elementary Problems in Geometry; http://www.math.pitt.edu/˜thales/papers/collection_geom.pdf.

[10] T. Hales, Easy Pieces in Geometry; preliminary version, 2007, http://www.math.pitt.edu/˜thales/papers/easy.pdf.

[11] T. Hales, The Flyspeck Fact Sheet, http://www.math.pitt.edu/˜thales/flyspeck/index.html.

[12] J. Harrison, S. McLaughlin, A Proof-Producing Decision Procedure for Real Arithmetic, *Automated Deduction - CADE-20*, LNCS 3632, Springer, 2005.

[13] B. Mishra, Computational Real Algebraic Geometry, Handbook of Discrete and Computational Geometry, CRC Press, 1997.

[14] A. Neumaier, Complete Search in Continuous Global Optimization and Constraint Satisfaction, *Acta Numerica 2004*, ed. A. Iserles, Cambridge University Press, 2004.

[15] S. Obua, Proving Bounds for Real Linear Programs in Isabelle/HOL, Dagstuhl Seminar Proceedings, ed. T. Coquand et al., Dagstuhl, Germany, 2006.

[16] P. A. Parrilo, Semidefinite programming relaxations for semialgebraic problems, Mathematical Programming Ser. B, Vol. 96, No.2, pp. 293-320, 2003.

[17] G. G. Szpiro, Kepler's Conjecture, Wiley, 2002.

[18] W. S. Wu, X. F. Jin, and D. M. Wang, *Mechanical Theorem Proving in Geometries*, Springer, 1994.

[19] R. Zumkeller, Formal Global Optimisation with Taylor Models, Int. Joint Conf. Automated Reasoning — IJCAR 2006, ed. U. Furbach and N. Shankar, Springer, LNCS, 2006.