swMATH
The Publication-based Approach to Software

Wolfgang Dalitz
Zuse Institute Berlin (ZIB)

Open Science Days 2019
5./6. February 2019
The Zuse Institute Berlin (ZIB)

is an interdisciplinary research institute for applied mathematics and data-intensive high-performance computing. Its research focuses on modeling, simulation and optimization with scientific cooperation partners from academia and industry.
Scientific Software in Mathematics

- Increasing role of scientific software
- Less credit for software developer
- An established infrastructure is missing
  - Standards
  - Information Services
  - „Reinvent the wheel“
  - Verification
  - Reproducibility
Famous Example: Four Color Theorem

„Every planar map can be colored with at most four colors“
Verifiability and Reproducibility

Have you ever tried to verify the four color theorem?

„Every planar map can be colored with at most four colors“

en.wikipedia.org/wiki/Four_color_theorem:

• proved in 1976 by K. Appel & W. Haken using a computer
• their proof were not accepted at all (infeasible for a human to check by hand)
• in 1997 was published a simpler proof by Robertson, Sanders, Seymour, and Thomas
• in 2005 the theorem was proved by Benjamin Werner and Georges Gonthier with Coq, a theorem-proving software
2. The computer programs

D-reduction was done dynamically (see [2]) in the sense that once a coloration was proved good, its goodness was immediately available for use in the testing of other colorations. The programs for D-reduction are extensions and modifications of those in [20].

The computer programs were greatly influenced by the facilities available. We had access to IBM computers (a 360-75 at Urbana-Champaign, a 370-158 at the University’s Chicago Circle Campus, and later a 370-168 of the University of Illinois administrative data processing unit). For this reason the programs were written in IBM assembler language to attempt to maximize efficiency. When we inquired, the operations staff suggested that we use less computer time at the expense of larger amounts of core storage. Therefore, to save steps we chose to use large tables. The core storage requirements were as follows: for twelve-rings, 220,000 bytes; for thirteen-rings, 600,000 bytes; for fourteen-rings, 1,700,000 bytes.
Scientific Workflow

Scientist

Software Developer
Scientific Software in Mathematics

Scientific Software

• plays an important role within the scientific workflow
• produces new scientific results
• is (sometimes/often) the base of a proof
  • e.g. four color problem/Vierfarbenproblem

Scientific Software Developer

• receives little or no scientific recognition
• gets no or less academic reputation
• gets no credit points in his academic career
Bridging the Gap:  www.swmath.org

- makes important software visible
- findable (unique and persistent identifier)
- accessible

- main idea: publication-based approach
- general: machine-based analysis of the content of publications
- cooperation with Zentralblatt MATH (zbMATH)
What is zbMATH (zbMATH.org)?

zbMATH is an abstracting and reviewing service in pure and applied mathematics.

1868-1942

since 1931
Mathematical Reviews® (MathSciNet®)

Since 1940, Mathematical Reviews® (MR) has served researchers and scholars in the mathematical sciences by providing timely information on peer-reviewed articles and books. MathSciNet®, the electronic version of MR, presents a fully searchable database with many tools designed to help navigate the mathematical sciences literature, including:

- reviews written by a community of experts
- bibliographic listings dating back to the early 1800s
- links to articles, journals, and publishers
- linked reference lists
- citation information on articles, books, and journals
zbMATH in Numbers

- zbMATH covers all available published and peer-reviewed articles, books, conference proceedings as well as other publication formats.

- zbMATH database contains:
  - 4 million bibliographic entries with reviews and abstracts
  - drawn from about 3,000 journals and serials and from
  - 180,000 books

- About 7,000 active reviewers from all over the world contribute reviews to zbMATH.

- All entries are classified according to the Mathematics Subject Classification Scheme (MSC2010).
Achterberg, Tobias

SCIP: solving constraint integer programs. (English) Zbl 1171.90476


Summary: Constraint integer programming (CIP) is a novel paradigm which integrates constraint programming (CP), mixed integer programming (MIP), and satisfiability (SAT) modeling and solving techniques. In this paper we discuss the software framework and solver SCIP (Solving Constraint Integer Programs), which is free for academic and non-commercial use and can be downloaded in source code. This paper gives an overview of the main design concepts of SCIP and how it can be used to solve constraint integer programs. To illustrate the performance and flexibility of SCIP, we apply it to two different problem classes. First, we consider mixed integer programming and show by computational experiments that SCIP is almost competitive to specialized commercial MIP solvers, even though SCIP supports the more general constraint integer programming paradigm. We develop new ingredients that improve current MIP solving technology. As a second application, we employ SCIP to solve chip design verification problems as they arise in the logic design of integrated circuits. This application goes far beyond traditional MIP solving, as it includes several highly non-linear constraints, which can be handled nicely within the constraint integer programming framework. We show anecdotally how the different solving techniques from MIP, CP, and SAT work together inside SCIP to deal with such constraint classes. Finally, experimental results show that our approach outperforms current state-of-the-art techniques for proving the validity of properties on circuits containing arithmetic.

MSC:
- 90C11 Mixed integer programming
- 68T20 AI problem solving (heuristics, search strategies, etc.)
- 90C27 Combinatorial optimization
- 90-04 Machine computation, programs (optimization)
- 90-08 Computational methods (optimization)

Keywords:
constraint programming; integer programming; SAT

Software:
Decision tree for optimization software; MIPLIB; Chaff; Valse-XT; SoPlex; Mesek; SCIL; SIMPL; CPLEX; MiniSat; Siege; ABACUS; MIPLIB2003; CLP; Zimpl; FEASPUMP; SCIP; Benchmarks for Optimization Software; COIN-OR; Tabu search; XPRESS
Software:
Decision tree for optimization software; MIPLIB; Chaff; Valse-XT; SoPlex; Mosek; SCIL; SIMPL; CPLEX; MiniSat; Siege; ABACUS; MIPLIB2003; CLP; Zimpi; FEASPUMP; SCIP; Benchmarks for Optimization Software; COIN-OR; Tabu search: XPRESS

References:
Main Idea: Publication-based Approach

- identification and analysis of software references in publications
  - which articles refer/review software in zbMATH?
  - store the result (publication/software) into a database

- swMATH offers a list of all publications and articles listed in Zentralblatt MATH (zbMATH) that refer to software
- swMATH includes also articles which describe the background and technical details of a program, as well as those publications in which a piece of software is applied or used for research
Building swMATH

Articles → zbMATH DB → swMATH DB → arXiv.org, JSS, JORS

Software Archives, github, Homepages
Quick Overview:  www.swmath.org

swMATH
an information service for mathematical software

Search  Advanced search  Browse

24928 software packages with 330103 references in 182904 zbMATH-articles
Example: Searching for SCIP

SCIP
SCIP is currently one of the fastest non-commercial solvers for mixed integer programming (MIP) and mixed integer nonlinear programming (MINLP). It is also a framework for constraint integer programming and branch-cut-and-price. It allows for total control of the solution process and the access of detailed information down to the guts of the solver. SCIP is part of the SCIP Optimization Suite, which also contains the LP solver SoPlex, the modelling language ZIMPL, the parallelization framework UG and the generic column generation solver GCG.

This software is also peer reviewed by journal MPG.

Keywords for this software
- integer programming
- constraint programming
- quadratic programming
- combinatorial optimization
- non-convex optimization
- mixed integer nonlinear programming
- polyhedral combinatorics
- integer linear programming
- mixed-integer programming

Related software:
- CPLEX
- MIPLIB
- MIPLIB2003
- SoPlex
- Gurobi
- XPRESS
- MINPLib
- LINDO
- Benchmarks for Optimization...
Example cont.

References in zbMATH (referenced in 237 articles, 4 standard articles)

Showing results 1 to 20 of 237.

1 2 3 ... 10 11 12 next

2. Assar, Benjamin; Gawrilow, Ewgenij; Herr, Katrin; Joswig, Michael; Lorenz, Benjamin; Paffenholz, Andreas; Rehn, Thomas: Computing convex hulls and counting integer points with polymake (2017)
4. Brinkmann, Philip; Ziegler, Günter M.: A flag vector of a 3-sphere that is not the flag vector of a 4-polytope (2017)
6. Gleixner, Ambros M.; Berthold, Timo; Müller, Benjamin; Weltge, Stefan: Three enhancements for optimization-based bound tightening (2017)
7. Göttl, Simone; Potschka, Andreas; Ziegler, Ute: Partial outer convexification for traffic light optimization in road networks (2017)
8. Haws, David; Cussens, James; Studeny, Milan: Polyhedral approaches to learning Bayesian networks (2017)
10. Ichim, Bogdan; Káli, Lukas; Moyano-Fernández, Julio Josant: How to compute the Stanley depth of a module (2017)
13. Modaresi, Sina; Vielma, Juan Pablo: Convex hull of two quadratic or a conic quadratic and a quadratic inequality (2017)
17. Puranik, Yash; Sahinidis, Nikolaos V.: Bounds tightening based on optimality conditions for nonconvex box-constrained optimization (2017)
18. Witzig, Jakob; Berthold, Timo; Heinz, Stefan: Experiments with conflict analysis in mixed integer programming (2017)

1 2 3 ... 10 11 12 next
Features: Browse by Name/MSC/Types/Keyword

Browse software by Mathematics Subject Classification (MSC 2010)

- 00 General mathematics
- 01 History; biography
- 02 Mathematical logic
- 05 Combinatorics
- 06 Order (MSC 2000)
- 08 General algebraic systems
- 11 Number theory
- 12 Field theory and polynomials
- 13 Commutative algebra
- 14 Algebraic geometry
- 15 Linear and multilinear algebra; matrix theory
- 16 Associative rings and algebras
- 17 Nonassociative rings and algebras
- 18 Category theory; homological algebra
- 19 K-theory
- 20 Group theory and generalizations
- 22 Topological groups, Lie groups
- 26 Real functions
- 28 Measure and integration
- 30 Functions of a complex variable
- 46 Functional analysis
- 47 Operator theory
- 49 Calculus of variations and optimal control; optimization
- 51 Geometry
- 52 Convex and discrete geometry
- 53 Differential geometry
- 54 General topology
- 55 Algebraic topology
- 57 Manifolds and cell complexes
- 58 Global analysis, analysis on manifolds
- 60 Probability theory and stochastic processes
- 62 Statistics
- 65 Numerical analysis
- 66 Computer science
- 70 Mechanics of particles and systems
- 74 Mechanics of deformable solids
- 76 Fluid mechanics
- 78 Optics, electromagnetic theory
- 80 Classical thermodynamics, heat transfer

Browse software by keywords

A
- a posteriori error estimation
- accuracy
- adaptive mesh refinement
- adaptivity
- algebraic geometry
- algebraic multigrid
- algebraic specification
- algebraic topology
- algorithms
- analysis of variance
- answer set programming
- applications
- approximation
- Arnoldi method
- artificial intelligence
- astrophysics
- asymptotic expansions
- asymptotic stability
- automated reasoning
- automated theorem proving
- automatic differentiation
- automorphism group
- harmonic analysis

B
- Galerkin method
- game theory
- Gaussian elimination
- geometry
- geophysics
- general relativity
- generalized eigenvalue problem
- genetic programming
- genetic algorithms
- global analysis
- global convergence
- global optimization
- GMRES
- graph theory
- graphics
- grid computing
- Gröbner bases
- group theory
- parallel algorithms
- parallel computing
- parallel processing
- partial differential equations
- periodic orbits
- periodic solutions
- perturbation
- Petri nets
- planning
- Poisson equation
- polynomial systems
- porous media
- preconditioning
- prediction
- preprocessing
- principal component analysis
Feature: Link to InternetArchive ...

SCIP
SCIP is currently one of the fastest non-commercial solvers for mixed integer programming (MIP) and mixed integer nonlinear programming (MINLP). It is also a framework for constraint integer programming and branch-cut-and-price. It allows for total control of the solution process and the access of detailed information down to the gate of the solver. SCIP is part of the SCIP Optimization Suite, which also contains the LP solver SoPlex, the modelling language ZIMPL, the parallelization framework UG and the generic column generation solver OCG.

This software is also peer reviewed by journal MPC.

Keywords for this software:
- mixed integer nonlinear programming
- combinational optimization
- analytic branch-and-bound
- branch and cut
- heuristics
- mixed integer programming
- branch-and-cut
- integer linear programming
- cutting planes
- Bayesian network
- integer programming
- network design
- quadratic programming
- decomposition
- linear programming
- local search
- network design
- global optimization

References in zbMATH (referenced in 339 articles, 4 standard articles)
Showing results 1 to 20 of 339.

2. Elliott, Sououd; Lambert, Alexandre; Amélie: Global solution of non-convex quadratically constrained quadratic programs (2019)
3. Aflahr, Lena G.; Dörig, Bastian; Ederer, Thorsten; Pelz, Peter F.; Pfetsch, Marco E.; Wolf, Jan: A mixed-integer nonlinear program for the design of gears (2016)
4. Balbiani-Lugojan, Radu; Misener, Ruth: Piecewise parametric structure in the pooling problem: from sparse strongly...
... leads to the WayBackMachine
Feature: Link to Version History ...
... leads to Tempas TimePortal (L3S)
Agrep

AGREP - approximate GREP for fast fuzzy string searching. Files are searched for a string or regular expression, with approximate matching capabilities and user-definable records. Developed 1989-1991 by Udi Manber, Sun Wu et al. at the University of Arizona. ISC open source license since Sept. 2014.

URL: github.com/Wikinaut/sgrep
Code
InternetArchive
Versions: Info
Authors: S. Wu, U. Manber

Add information on this software.

Related software:
PSI-BLAST
BLAST
GLIMPSE
Software Heritage

Browse archived visits for origin https://github.com/Wikinaut/agrep

Overview
Total number of visits: 10  Last full visit: 07 March 2018, 19:03 UTC  First full visit: 06 August 2015, 13:46 UTC  Last visit: 07 March 2018, 19:03 UTC

History

Calendar

Show full visits with different snapshots  Show all visits  Show all visits

2016

January
Su Mo Tu We Th Fr Sa
1 2 3 4 5 6
7 8 9 10 11 12 13
14 15 16 17 18 19 20
21 22 23 24 25 26 27
28 29 30 31

2017

February
Su Mo Tu We Th Fr Sa
1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28

2018

March
Su Mo Tu We Th Fr Sa
1 2 3
4 5 6 7 8 9 10
11 12 13 14 15 16 17
18 19 20 21 22 23 24
25 26 27 28 29 30 31

April
Su Mo Tu We Th Fr Sa
1 2 3 4 5 6 7
8 9 10 11 12 13 14
15 16 17 18 19 20 21
22 23 24 25 26 27 28
29 30
Some Statistics

- swMATH has been started in 2011, a joined project of Research Institute Oberwolfach (MFO) and FIZ Karlsruhe
- currently a project of the BMBF research campus MODAL with FIZ Karlsruhe/zbMATH and Zuse Institute Berlin (ZIB)
- ~ 25,000 Software Packages
- ~ 330,000 Software References in
- ~ 182,000 zbMATH and other Scientific Articles
Usage of swMATH

Usage Statistics for swmath.org (Apache-Logfile, Webalizer-pages, with robots)
01.2015 - 01.2019

0  750.000  1.500.000  2.250.000  3.000.000
2015 07 2016 07 2017 07 2018 07 2019

pages
Software and zbMATH References

swMATH figures for software and zbMATH references 01.2015 - 01.2019

- Software (25,000)
- zbMATH references (180,000)
Some Results (I)

- swMATH provides general information about software, especially about software products (or containers): information about all software artifacts which are under a common name.

- We have extended the swMATH pages by linking to further Internet information resources if existing:
  - Websites of the software - problem: not permanent
  - Internet Archive - problem: archive not complete
  - Software Heritage - problem: SWH provides artifacts (revisions/commits, releases), but not direct versions

- Verification of scientific results needs more information about the version of the software used.
swMATH makes the software container citable:

- each swMATH page gets a unique and persistent identifier
- swMATH page are so-called „landing pages“ in the terminology of the Software Citation Principles
Some Results (III)

- the numbers of citations listed in swMATH is an indicator of acceptance, spread and quality of the software
- software developer
  - receives more scientific recognition
  - gets more academic reputation
  - gets more credit points in his academic career
New Scientific Recognition

Scientist

Software Developer
The overall aim: developing an efficient infrastructure for (scientific) software based on the FAIR principles
  ➔ Findable, Accessible, Interoperable, Reusable

Definition and dissemination of a software citation standard: under work, especially discussed in Software Citation Implementation working group of the FORCE11 initiative
  ➔ The definition of a software citation standard is not trivial: Software is a complex and structured object and cannot be reduced to software code

Developing of efficient machine-based methods for analysis of software information

Compatibility, interoperability, and interaction of different software information services and archives
Thanks for your attention