Evaluating Optimization Software

Hans D Mittelmann

School of Mathematical and Statistical Sciences
Arizona State University

Shenzhen Research Institute of Big Data
Shenzhen, PR China
5 June 2023

Outline

The History of our Benchmarking

- Introduction
- Early History
- Intermediate History
- Recent History
- Latest History

Epilogue

Outline

The History of our Benchmarking
- Introduction
- Early History
- Intermediate History
- Recent History
- Latest History

Epilogue

Introduction

Overview

- For more than twenty years we are benchmarking optimization software.
- In late 2018 an event had a major impact on this service which had gained considerable notoriety.
- The history up to this point is sketched and the developments to late 2019 and beyond are described.
- For additional information reference is made to the webpage Benchmarks for Optimization Software ▸ Link and all that is accessible through it.

Introduction

The Beginning

- About 1996 we started the optimization software guide Link together with Peter Spellucci. It was intended to be used by everyone looking for an **available tool** to solve various optimization problems. It contained not only structured lists of codes but additional information such as related literature.

- In order to assist users the mere listing of software was soon seen as insufficient and a year or two later a **subpage was started** on benchmarks. Some of the listed programs were tested on a selection of suitable instances and the results published.

- Initially, only non-commercial programs were included, as were nearly exclusively only such programs listed in the guide. Later, selected, popular, and powerful commercial codes were added to the benchmarks.

Introduction

Guide, Benchmarks, NEOS

- The benchmarks lead to **more transparency** about the capabilities of the codes tested and thus, not surprisingly, also to more competition between the developers.
- This benefitted the users who are the main target audience of these efforts. It also was appreciated by the well-performing programs because it provided free advertising which was heavily exploited.
- After these two service components, guide and benchmarks, a third one was started, namely the installation of various optimization programs on our computers which were accessible through the **NEOS gateway** Link, comprising about **one third** Link of the available solvers.
- Over time a collection of test instances, archived programs, and other related items accumulated and was made available. All this is done without any personnel or financial support.

Introduction

The Rationale

- The following rationale for our service may help explain our motivation. Optimization is **ubiquitous** and most number-crunching computation is done in optimization.
- While mathematically most optimization is **not hard**, the writing of efficient and robust computer programs is.
- Users of optimization software are well-advised to try not one but **several codes** that can handle the problem they are trying to solve.
- For this purpose, NEOS is a **perfect tool**, providing a number of programs in each category and with the same modeling language as input, or numerical input in the same format for problems requiring only numerical data.

Outline

The History of our Benchmarking
- Introduction
- Early History
- Intermediate History
- Recent History
- Latest History

Epilogue

The History of our Benchmarking

Early History, 1998-2009

- We leave the early years off and divide the history into four parts, the years 1998-2009, 2010-2017, 2018-2019, 2020-2023.
- For actual benchmark tables dated 1998-2005 we refer to a selection of **Older Benchmarks** Link.
- We start in 2006 and from then on in light of events in late 2018 we concentrate on three commercial products
  CPLEX, Gurobi, and XPRESS.
- For details of the following reference is made to the talk given at the EURO 2019 conference and linked to on Link

Early History, 1998-2009

going parallel

- In 2006 our benchmarks Link start listing the first results for (shared-memory) **parallel optimization**.
- The first AMD Opteron and Intel multi-core computers became available and were used to run CPLEX on mixed-integer linear programs (MILPs).
- The Opteron is initially quite competitive but is gradually surpassed by the Intel processors.
- The tables also show first results for two different forms of parallelization, namely **deterministic and opportunistic**, a feature CPLEX maintains up to this day, but which, for example, Gurobi decided not to offer.

Early History, 1998-2009

From CPLEX to Gurobi

- It can be said that **multi-core computing** became the standard, especially for the commercial codes. In addition to CPLEX we had tested XPRESS, but in 2007 the company DASH asked us to omit these results.

- Then, in 2008 at the INFORMS Annual Meeting Robert Bixby, the founder of CPLEX, which had first been acquired by ILOG and then together with ILOG by IBM, presented **first results for the software Gurobi**.

- Since its founding with the former CPLEX developers Zonghao Gu and Ed Rothberg eighteen months had passed and nine of those had been used for the coding.

Early History, 1998-2009

Role of XPRESS

- At the 2008 INFORMS meeting we gave a talk on our benchmarks, of which there were nineteen in seven categories, see ▸ Link, but not yet on Gurobi.
- This happened soon after the meeting and, when a new product manager was appointed at XPRESS, which meanwhile had been **acquired by FICO**, we were asked to include the solver again.
- This leads us to the next era, the intermediate history.

Outline

The History of our Benchmarking
- Introduction
- Early History
- Intermediate History
- Recent History
- Latest History

Epilogue

Intermediate History, 2010-2017

Emerging Libraries

- In order to test codes we typically had collected and partly generated, also as part of our research, see e.g. Link, suitable test instances.

- This had gone on for several years and in addition **several libraries** were curated by teams of experts and stakeholders. In particular, in 2010 the library MIPLIB2010 Link was created and made available in 2011.

- In July 2010 at the EURO meeting we showed benchmarks Link on two instance collections we had generated (p. 15-18), one meant to be a general benchmark set and one of instances that appeared to be **difficult for some codes**. Both included CPLEX, Gurobi, several other codes but not XPRESS. Remarkable is that in the latter benchmark the best times per instance were split between CPLEX, Gurobi and SCIP Link.

Intermediate History, 2010-2017

Big Three Performance Varies

- At INFORMS 2011 the benchmark subset of MIPLIB2010 was the basis for a comparison Link of nine codes, including XPRESS, on one, four, and twelve threads.

- In parallel execution CPLEX and Gurobi performed very similary while XPRESS was about twenty percent slower. Here, instead of performance profiles which are not suitable to compare more than two codes at a time Link, the **shifted geometric mean** of run times was used. A shift (of ten seconds) is added, the geometric mean of all times is formed and then the shift subtracted.

- In the above benchmark from about summer 2012 on Gurobi became consistently slightly faster than its main commercial competitors. For more detailed performance results of various solvers on MIPLIB2010 see Link

Intermediate History, 2010-2017

Expanding Instance Sets

- It became apparent that the dedicated benchmark set had become too **easy** and we reacted to this in two ways. On the one hand, we enlarged the set of **difficult** instances, which as requested by Bixby were renamed as **pathological**.

- In the 2013 such benchmark ▸ Link CPLEX clearly outperformed (p. 24) the others. But the number of instances in both this set and in the MIPLIB2010 benchmark set was rather small, so the major change we introduced was to run the codes on all MIPLIB instances which were classified as **easy** because they could be solved by at least one solver in an hour.

- In 2015 this set had 205 members and CPLEX and Gurobi performed (p. 21) **equally well** ▸ Link.

Intermediate History, 2010-2017

Gurobi Moves Ahead

- Again, a request was made to rename these instances as **solvable** and in 2016 on a slightly larger set and with both 12 and 48 threads, Gurobi was fastest with CPLEX ten to twenty percent slower ▸ Link.
- All codes were upgraded. In 2017 Gurobi **maintained its lead** ▸ Link.
- Also, a quite strong non-commercial code that had been developed unnoticed was included, MIPCL. (was meanwhile bought by Huawei)
- Another noteworthy event in the intermediate history period was the publication of another instance set, CBLIB ▸ Link. It formed the basis of new benchmarks in the categories SOCP and MISOCP.

Outline

The History of our Benchmarking
- Introduction
- Early History
- Intermediate History
- Recent History
- Latest History

Epilogue

Recent History, 2018-2019

Major step - MIPLIB2017

- In the summer of 2018 at ISMP Link the "solvable" benchmark on meanwhile 220 instances (omitting numerically unstable ones), confirmed Gurobi's lead: Scaled and shifted geometric means: Gurobi 1, CPLEX about 1.2, XPRESS about 1.5.

- While we listed only partial information about the benchmarks, here is a statistic about the overall situation in 2018: CPLEX is in 15/22 benchmarks, while Gurobi and XPRESS are in 13. For many years Gurobi had used our benchmarks in their **public relations effort**.

- Several major developments culminated at the time of the 2018 INFORMS meeting. A committee had worked on MIPLIB2017 Link. The difference to earlier libraries was that the selection of the benchmark set, it ended up having 240/1065 instances Link in contrast to 87/461 for MIPLIB2010, was done with a **sophisticated algorithm implemented in a computer program**.

Recent History, 2018-2019

INFORMS 2018 in Phoenix

- As the previous library it also was published with a one year delay, namely during INFORMS 2018. As member of the MIPLIB committee we had early access and prepared a benchmark for the meeting. Major stakeholders were represented on the committee as well including Gurobi.
- At INFORMS2018 an incident happened for which we refer to our benchmark webpage. Subsequently, CPLEX and XPRESS asked to **not be included** any longer.
- The benchmarking effort continued and became more tedious because the non-commercial codes on increasingly challenging test sets needed much more compute time.
- In the mean time another major problem library had been published, QPLIB ▸ Link. At INFORMS 2018 we had already included three parts of the 453 instance library in the benchmarks.

Recent History, 2018-2019

Including QP

- In the summer of 2019 we learnt that Gurobi was developing a **global solver** for (non-convex) bilinear/quadratic MIPs and likely by INFORMS 2019 would be the only one of the **big three** capable of solving these problems to global optimality.

- We decided to include the corresponding numbers in the QPLIB benchmarks when the new version 9.0 would become available. This was the case in late November of 2019, about one year after INFORMS 2018.

- The results came too late for INFORMS 2019 but are on the web. The **entire QPLIB** is split into five pieces and suitable commercial and non-commercial solvers are compared ▸ Link.

Outline

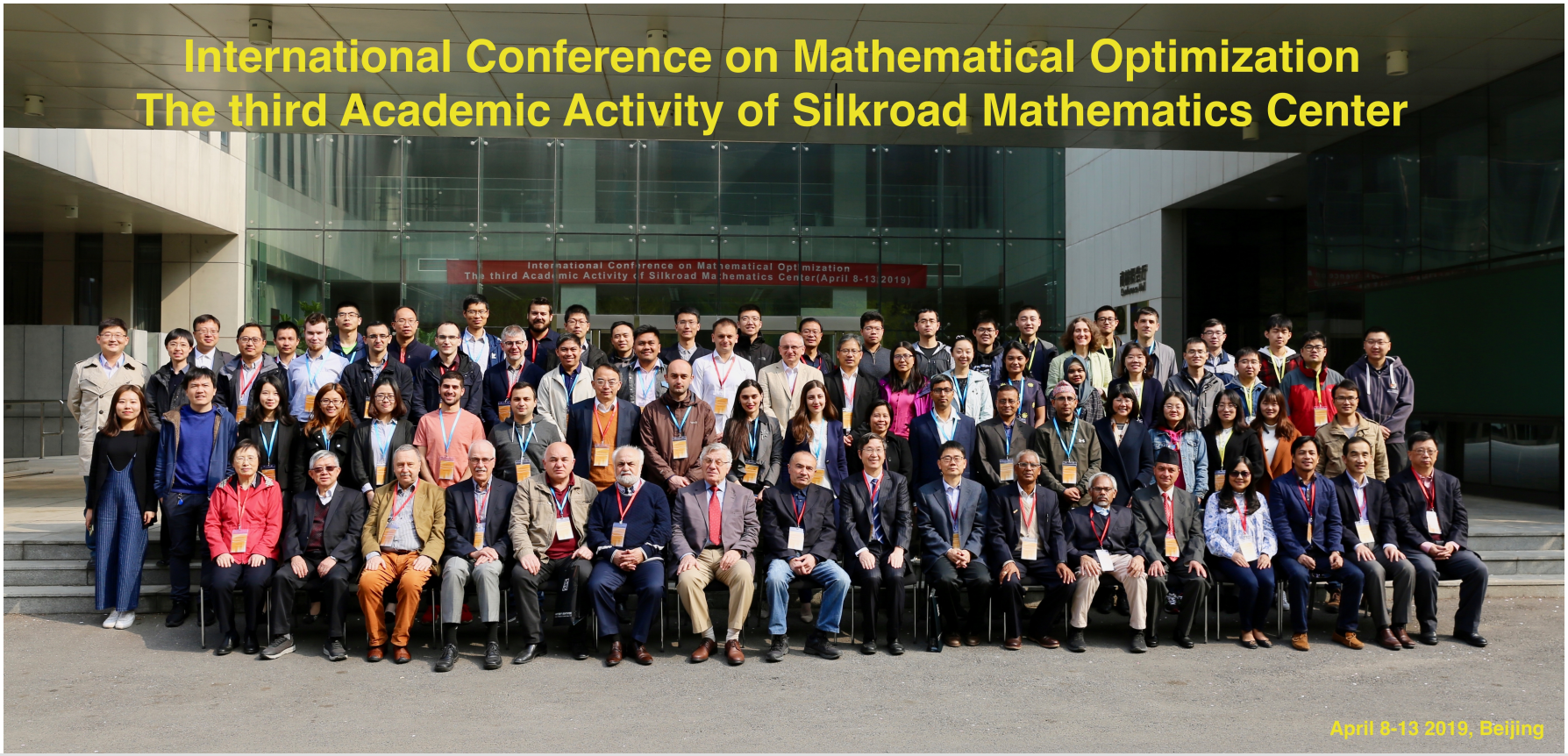
The History of our Benchmarking
- Introduction
- Early History
- Intermediate History
- Recent History
- Latest History

Epilogue

Latest History, 2020-2023

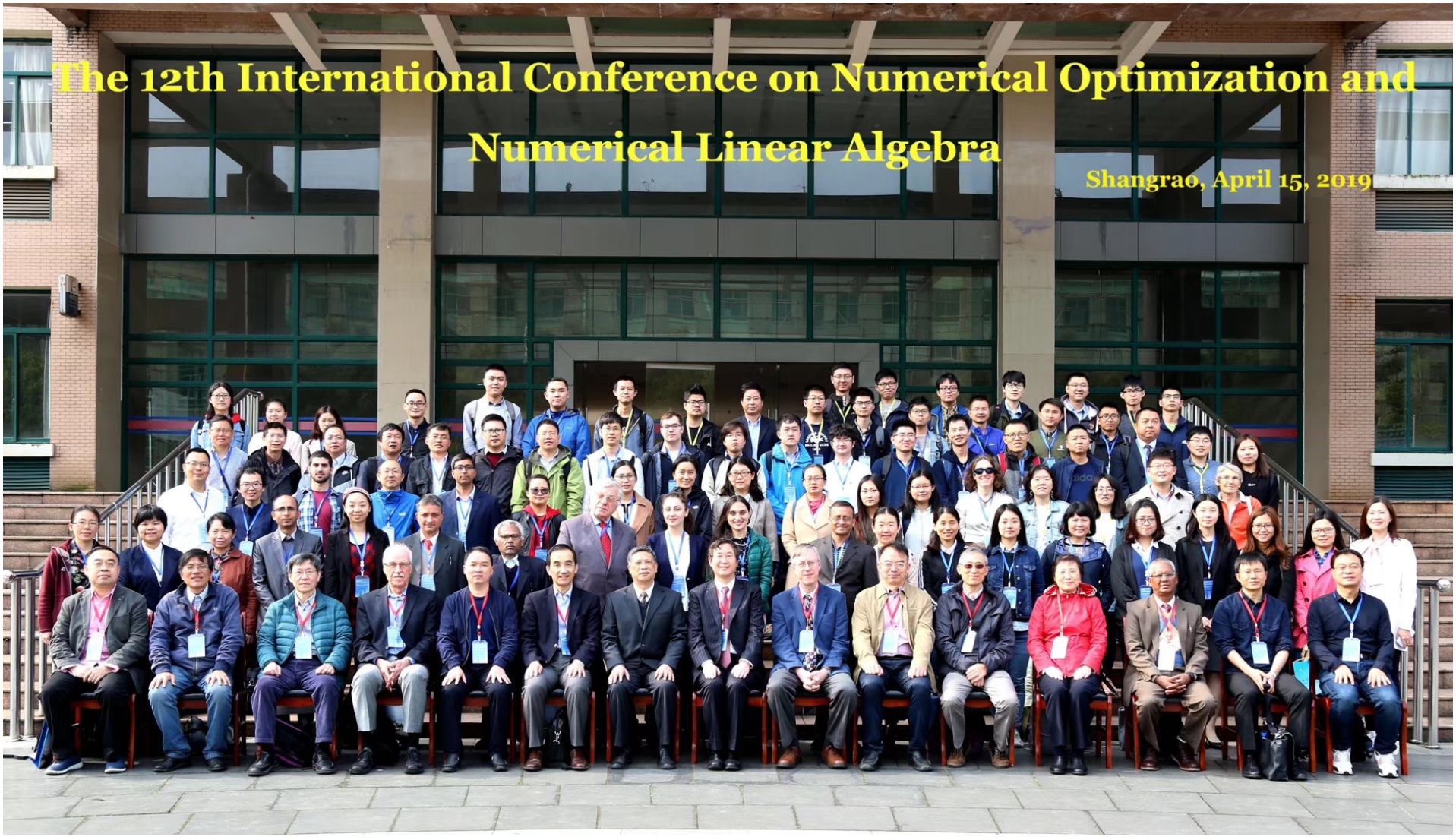
2019 marks a turning point

- So far **no Chinese codes** have been listed
- Just one Chinese company competed: **Cardinal Operations**
- two more started: **Alibaba-US**(MindOpt), **Huawei-CA**(OptVerse)
- Coincidentally, in 2019 we were invited to **three** conferences
-



Latest History, 2020-2023

2019 marks a turning point



Latest History, 2020-2023

Annual meeting of the OR section of the Chinese Mathematical Society

In Nanjing

About 900 participants

Two US speakers

Latest History, 2020-2023

China Rising

- To track the development 2020-2023 there are the following sources
  - Talk at INFORMS 2020 Link and INFORMS 2021 Link
  - Talk at INFORMS 2022 Link
  - Matt Miltenberger's visualization of our benchmarks Link
- Here are some geomeans for the Simplex benchmark
  - 2020: Gurobi 1.0, COPT 1.19, MindOpt 1.68
  - 2021: MindOpt 1.0, COPT 1.25, OptVerse 1.38, Gurobi 2.26
  - 2022: MindOpt 1.12, COPT 1.0, OptVerse 2.56, Gurobi 1.67
- and here are some geomeans for the Barrier benchmark
  - 2020: Gurobi 1.0, COPT 2.39
  - 2021: COPT 1.0, Gurobi 1.56, MindOpt 2.33
  - 2022: COPT 1.0, Gurobi 1.31, MindOpt 2.39

Latest History, 2020-2023

Leveling the Playing Field

- What had happened?
  - We had replaced several LPs by harder ones
  - Some competitors improved their code
  - Some competitors tuned their code to the instances
  - Measure taken: added 20 **undisclosed LPs** in late March 22
- At INFORMS 2022 Gurobi argues that the **traditional split** into a Simplex and a Barrier benchmark is obsolete because the two methods are **"mixed"** in actual codes.
- Suggestion is to have a benchmark to find **(somehow)** a primal-dual feasible point and another to find an optimal basic solution.
- Suggestion makes sense from the **user's viewpoint**

Latest History, 2020-2023

NEW LP Benchmarks

- In December 2022 these two new benchmarks were started
  - LPfeas ▸ Link
  - LPopt ▸ Link
- Gurobi and COPT introduced new parameters for LPfeas
  - Gurobi uses **SolutionTarget = 1**
  - COPT uses **LPMethod = 5**
  - MindOpt/Alibaba and OptVerse/Huawei **still working**
  - **CPLEX-barrier** w/o crossover reaches geomean of 1.36, but will not appear in benchmark

Latest History, 2020-2023

Other Recent Progress

- Other remarkable improvements (geomean and problems solved)
  - **OptVerse** has strong network algorithm and jumps to first place
    - CLP 5.66, COPT 1.99, Gurobi 1.60, **OptVerse 1**, HiGHS 11.9, Mindopt 1.20
  - **Octeract** asks to utilize CPLEX in comparison with other solvers
    - Nonbinary QP: Baron 29.3(31), **Octeract 3.21(79)**, Gurobi 1(81)
    - Nonconv cont QP: Baron 3.81(28), **Octeract 1.42(46)**, Gurobi 1(43)
  - **COPT** adds convex QP algorithm and jumps to first place
    - MOSEK 1.10(32), KNITRO 1.50(32), Gurobi 1.91(30), **COPT 1(32)**

Epilogue

- While the events at INFORMS 2018 clearly affected the benchmarks and did **shed a light** on the commercial side of optimization software, life goes on and a lot needs to be done.
- In no way can a claim be made that in the categories of optimization important in applications a rather satisfactory state has been reached. There are **still challenging** linear MIPs, but especially for non-convex quadratic and nonlinear MIPs there is a need for more efficient and robust solvers.
- Admittedly, a good **first step** was done in the quadratic case by Gurobi. A MINLP solver that is competing with BARON is OCTERACT. It is being constantly improved. An ambitious open source solver under active development is **HiGHS**. SCIP remains the most powerful research code. Not much happens at COIN-OR.

THANK YOU

Questions?

Slides of talk at
http://plato.asu.edu/talks/shenzhen23.pdf