Latest Progress in Optimization Software

Hans D Mittelmann

School of Mathematical and Statistical Sciences
Arizona State University

INFORMS Annual Meeting
Atlanta, GA
28 October 2025

Outline

Outline

Introduction

Recent Developments

Selected Benchmarks

Conclusions

HOME SOFTWARE BENCHMARKS TESTCASES BOOKS MORE

DECISION TREE FOR OPTIMIZATION SOFTWARE

BENCHMARKS FOR OPTIMIZATION SOFTWARE

By Hans Mittelmann (mittelmann at asu.edu)

END OF A BENCHMARKING ERA

For many years our benchmarking effort had included the solvers CPLEX, Gurobi, and XPRESS. Through an action by Gurobi at the 2018 INFORMS Annual Meeting this has come to an end. IBM and FICO demanded that results for their solvers be removed. See here for more details. The resulting void was filled by other developers.

In August 2024 Gurobi decided to withdraw from the benchmarks as well and their results have been removed. See the note at the bottom of the MIPLIB benchmark.

What is on top of the Benchmark Page?

- Links to history and older benchmarks
- Especially details on the elimination of CPLEX, XPRESS, and Gurobi
- Hint that full solver logfiles are provided
- Explanation why performance profiles are not used
- Link to Matt Miltenberger's visualization tool

History and Scope of the Benchmarks

- Start about 1998
- Summary presented at INFORMS Annual starting 2002
- 22 benchmarks in these areas
  - TSP, (MI)LP, (MI)SOCP, SDP, convex/nonconvex (MI)QCP, QUBO, (MI)NLP, MPEC
- Codes evaluated: 46, of those 25 actively developed, 14 commercial
- Total separate runs: 138, needing up to 3 weeks

Outline

Introduction

Recent Developments

Selected Benchmarks

Conclusions

The Benchmarks at a Turning Point

We are in the age of AI

- Potential use of over-tuning/machine learning
- benchmarks prone to AI except those with unknown datasets (LP, MIPLIB)
- Optverse (the AI code) is unavailable and cannot be compared to other codes; its numbers will be removed after the conference
- MIPLIB2024 to be published soon; will it be AI-proof?
- utilization of GPUs (accelerated computing), so far in LP, QP, SDP
- more solvers can handle nonconvexity (globally)

Outline

Selected Benchmarks

These seven benchmarks follow next:

- convex: LPfeas, MILP, SDP
- nonconvex: NLP, MIQCP(2), MINLP
- remember: "the great watershed in optimization isn't between linearity and nonlinearity, but convexity and nonconvexity"
- Talk will report some progress with accelerated computing; more substantial progress may be possible in the future

LPfeas Benchmark (find a primal-dual feasible point)

- Same own instance selection as in LPopt benchmark
- 49 published instances 16 undisclosed instances
- Intel i7-11700K, 3.6GHz, 64GB, time limit 15000 secs, 1000 for GPU
- shifted geometric mean of runtimes

	Intel CPU							NVIDIA H100 GPU				
65 pr	COPT	MOSEK	HiGHS	KNTRO	PDLP	XOPT	OPTV	CUPDL	CUOPT	CUPDLX	COPTG	HPRLP
unscd	25.3	83.3	556	431	421	146	44.1	47.1	22.4	41.1	21.2	64.3
scald	1.19	3.93	26.2	20.3	20.0	6.87	2.08	2.22	1.06	1.94	1	3.03
solvd	65	59	49	49	50	59	64	56	62	61	64	59

LPfeas Benchmark (additional instances with GPU)

- selected problems from Oliver Hinder's collection (github, julia)
- two of five problems in various dimensions; time limit 15000 secds

problem	cuPDLP	cuOpt	COPT-G	cuPDLPx	HPR-LP	13b1GPU
heat_250_10_500_200	7288	m	m	t	t	t
heat_250_10_500_300	3696	m	m	t	t	t
heat_250_10_500_400	3014	m	m	t	t	t
mcf_2500_100_500	1166	m	763	1103	t	t
mcf_5000_50_500	2260	m	795	6660	t	t
mcf_5000_100_250	1461	m	842	5360	t	t

	constraints	variables	nonzeros
heat*	15625000	31628008	125000000
mcf*	1.5m	126m	253m

for larger dimensions there are memory issues

Crossover

- The unscaled geomean ratio LPfeas/LPopt allows to assess the crossover effort.

Code	Ratio	Crossover
COPT	25.3/38.5	52%
HiGHS	556/659	19%
MOSEK	83.3/288	246%
OPTV	38.9/67.8	74%
XOPT	146/255	75%

Performance on known and unknown datasets

- This comparison attempts to sense any tuning effort.

LPFEAS-Benchmark on the 49 public or all 65 instances

probs	COPT	MOSEK	HiGHS	XOPT	OPTV	CUPDL	CUOPT	CUPDLX	COPTG	HPRLP
PUBLIC	1.25	3.63	25.7	4.87	1.74	4.13	1.32	2.34	1	3.78
ALL	1.19	3.93	26.2	6.87	2.08	2.22	1.06	1.94	1	3.03

- LPOPT-Benchmark on the 49 public or all 65 instances

probs	CLP	COPT	OPTV	MOSEK	HiGHS	GLOP	SPLX	XOPT
PUBLIC	27.3	1	1.45	5.95	17.0	77.9	116	4.15
ALL	27.4	1	1.76	7.49	17.1	59.7	94.1	6.63

The final MIPLIB2017 Benchmark

- Total of 240 instances
- Sizes up to 1.5m/1m/43m rows/cols/nonzeros
- AMD Ryzen 9 5900X (12 cores, 128GB), 7,200 secs wall clock
- modification through random row and column reordering
- NOTE! OPTV received on 10/15, COPT on 9/21

	COPT	SCIP	SCIPC	HiGHS	XOPT	OPTV	XSMOO
unscal	133	1175	935	880	677	[117]	551
scaled	1.14	10.1	8.02	7.55	5.81	[1]	4.73
solved	218	128	145	157	160	[218]	172

MIPLIB redone; will be posted after conference

- Since the previous reordering of the MIPLIB dataset was done a while ago, the random seed was changed and the benchmark rerun

	COPT	SCIP	SCIPC	HiGHS	XOPT	OPTV	XSMOO	13b1GPU
unscal	103					139	515	77
scaled	1					1.35	5.00	
solved	217					211	171	228

- to be completed

Several SDP-codes on sparse and other SDP problems

- Total of 75 instances, own instance selection
- Sizes up to 100k/100k/450m variables/constraints/nonzeros
- AMD Ryzen 9 5900X, 12 cores, 128GB , 40,000 secs wall clock

problem	COPT	CSDP	MOSEK	SDPA	SDPT3	SeDuMi	cuLoRAD&
	1	5.38	2.67	7.96	5.30	29.8	2.99
count of "a"	6	5	1	18	13	2	14
solved of 75	75	70	73	61	69	62	71

"a": reduced accuracy

AMPL-NLP Benchmark

- Total of 47 instances
- Sizes up to 140k/260k constraints/variables
- AMD Ryzen 9 5900X (12 cores, 128GB), 7,200 secs wall clock

47 problems	IPOPT$	KNIT	SNOPT	CONPT	WORHP	MATLB	COPT	UNO	[OPTV]
scaled geomean	15.3	2.56	135	164	15.6	48.8	1.32	364	[1]
solved	46	46	30	26	45	36	47	22	[47]

$ Higher performance of IPOPT possible for different LA routines

Discrete Non-Convex QPLIB Benchmark (non-binary)

- Total of 160 instances
- Sizes up to 36k/17k/116k/48k variables/binaries/constraints/quads
- Instances from QPLib
- Intel Xeon Gold 6226R 2.90GHz (32 cores, 512GB), 3 hrs, 8 threads

	BARON	SHOT	SCIP	ANTIGONE	COPT
unscaled	513	74.6	2109	4798	155
scaled	6.87	1	29.3	64.3	2.07
solved	69	95	40	28	88

Continuous Non-Convex QPLIB Benchmark

- Total of 102 instances
- Sizes up to 200k/140k/140k variables/constraints/quads
- Intel Xeon Gold 6226R 2.90GHz (32 cores, 512GB), 3 hrs, 8 threads
- Instances from QPLib

	ANTIGONE	BARON	MINOTAUR	SCIP	COPT
unscaled	2169	428	1868	4649	1121
scaled	5.23	1	4.49	11.4	2.62
solved	27	35	25	14	25

Mixed Integer Nonlinear Programming Benchmark

- Total of 200 instances; run through GAMS
- Sizes up to 100k/100k/5k variables/lin constraints/nonlin constraints
- AMD Ryzen 9 5900X (12 cores, 128GB), 7,200 secs wall clock
- Instances from MINLPLib

	BARON	GUROBI*	LINDO	SCIP	SHOT#	XPRESS*
unscaled	73.6		391.4	126.0	414.3	
scaled	1.0		5.3	1.7	5.6	
solved	161		116	153	96	

*: publication suppressed

#: SHOT can only handle quadratic instances

Outline

Conclusions

Some Takeaways

- COPT leads in both CPU and GPU versions for LP, in MILP, SDP, SOCP and NLP
- COPT is second to SHOT-Gurobi in nonconvex MIQCP and to BARON in QCP
- cuOpt and COPT-G lead among GPU codes; COPT leads among CPU codes
- cuPDLP-C only GPU code to solve all large problems
- BARON continues to lead in MINLP; SCIP is a very strong second

THANK YOU

Questions?

Slides of talk at
https://plato.asu.edu/talks/informs2025.pdf