From structures to heuristics to global solvers

Timo Berthold

Zuse Institute Berlin

DFG Research Center MATHEON
Mathematics for key technologies

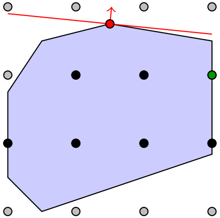
OR2013, 04/Sep/13, Rotterdam

From structures to heuristics to global solvers

1. Undercover: the largest sub-MIP [B., Gleixner. Math. Prog., 2013]
2. RENS: the optimal rounding [B. Math. Prog. C, under review]
3. Measuring the impact of primal heuristics [B. OR Letters, accepted]

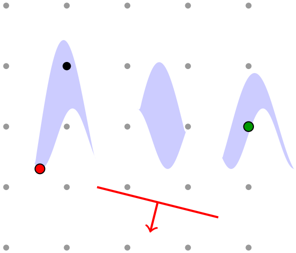
Mixed-Integer Linear Programming (MIP):

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & Ax \leq b \\ & x_i \in \mathbb{Z} \text{ for all } i \in \mathcal{I} \end{aligned}$$



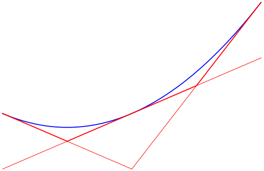
Mixed-Integer Nonlinear Programming (MINLP):

$$\begin{aligned} \min \quad & c^T x \\ s.t. \quad & g(x) \leqslant 0, \quad g \in \mathcal{C}^1(\mathbb{R}^n, \mathbb{R}^m) \\ & x_i \in \mathbb{Z} \text{ for all } i \in \mathcal{I} \end{aligned}$$

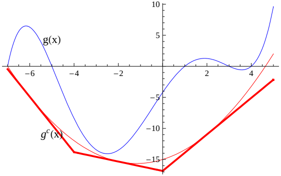


MINLP solving techniques

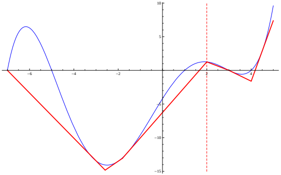
Gradient cuts



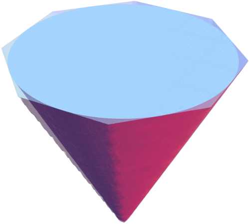
Underestimators



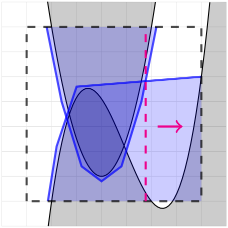
Spatial branching



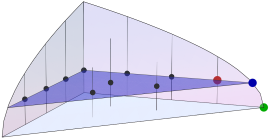
Reformulation



Bound tightening

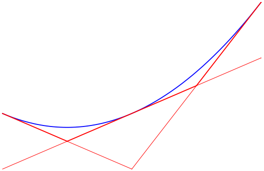


Primal heuristics

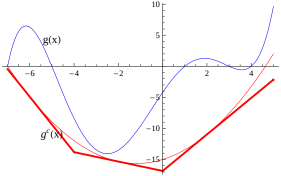


MINLP solving techniques

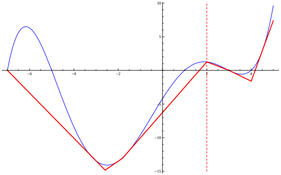
Gradient cuts



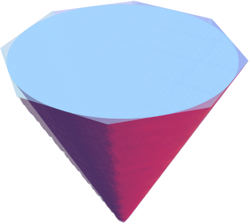
Underestimators



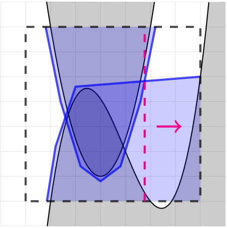
Spatial branching



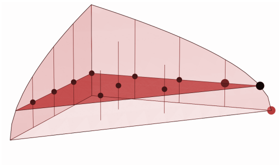
Reformulation



Bound tightening



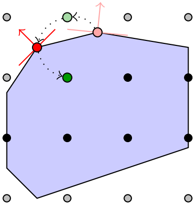
Primal heuristics



Primal heuristics inside a global solver

Primal heuristics . . .

- ▷ are incomplete methods which
- ▷ often find good solutions
- ▷ within a reasonable time
- ▷ without any warranty!



Why use primal heuristics inside a global solver?

- ▷ to prove feasibility of the model
- ▷ often nearly optimal solutions suffice in practice
- ▷ feasible solutions guide remaining search process

Improving the Feasibility Pump

DINS, a MIP Improvement Heuristic

OCTANE: A New Heuristic for Pure 0-1 Programs

RENS - the optimal rounding

An Evolutionary Algorithm for Polishing Mixed Integer Programming Solutions

Rapid Learning for Binary Programs

Proximity Search for 0-1 Mixed-Integer Convex Programming

A Generalized Wedelin Heuristic for Integer Programming

Shift-And-Propagate

Exploring relaxation induced neighborhoods to improve MIP solutions

Undercover: a primal MINLP heuristic exploring a largest sub-MIP

Local Branching

A Good Recipe for Solving MINLPs

ZI round, a MIP rounding heuristic

From structures to heuristics to global solvers

1. Undercover: the largest sub-MIP [B., Gleixner. Math. Prog., 2013]
2. RENS: the optimal rounding [B. Math. Prog. C, under review]
3. Measuring the impact of primal heuristics [B. OR Letters, accepted]

- Large Neighborhood Search: common paradigm in MIP heuristics

  fix a subset of variables ⇝ easy subproblem ⇝ solve

  MIP: "easy" = few integralities
  MINLP: "easy" = few nonlinearities

- observation: any MINLP can be reduced to a MIP by fixing (sufficiently many) variables.

- idea: fix a small subset of variables to obtain a linear subproblem
- use solution of an LP or NLP relaxation to determine fixing values

Definition (cover of an MINLP)

Let

- a domain box $[L, U] = \times_i [L_i, U_i]$,
- a function $g_j : [L, U] \to \mathbb{R}$, $x \mapsto g_j(x)$ on $[L, U]$, and
- a set $\mathcal{C} \subseteq \mathcal{N} := \{1, \ldots, n\}$ of variable indices be given.

We call $\mathcal{C}$ a *cover of g* if and only if for all $\bar{x} \in [L, U]$ the set

$$\{(x, g_j(x)) \mid x \in [L, U], x_k = \bar{x}_k \text{ for all } k \in \mathcal{C}\}$$

is an affine set intersected with $[L, U] \times \mathbb{R}$.

We call $\mathcal{C}$ a *cover of P* if and only if $\mathcal{C}$ is a cover for $g_1, \ldots, g_m$.

Input: MINLP P

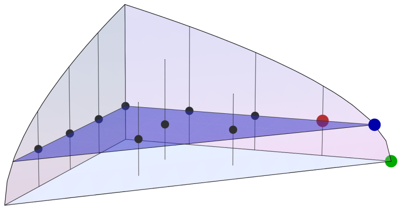
begin

compute a solution $\bar{x}$ of an approximation of P;

round $\bar{x}_k$ for all $k \in \mathcal{I}$;

determine a cover $\mathcal{C}$ of P;

solve the sub-MIP of P given by fixing $x_k = \bar{x}_k$ for all $k \in \mathcal{C}$;



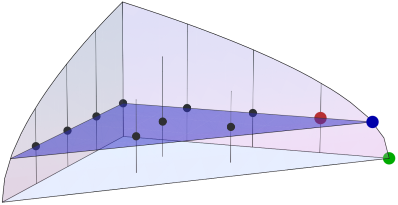
Input: MINLP P
begin

compute a solution $\bar{x}$ of an approximation of P;

round $\bar{x}_k$ for all $k \in \mathcal{I}$;

determine a cover $\mathcal{C}$ of P;

solve the sub-MIP of P given by fixing $x_k = \bar{x}_k$ for all $k \in \mathcal{C}$;



Input: MINLP P

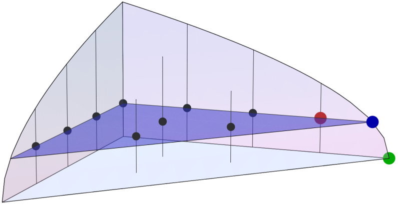
begin

compute a solution $\bar{x}$ of an approximation of P;

round $\bar{x}_k$ for all $k \in \mathcal{I}$;

determine a cover $\mathcal{C}$ of P;

solve the sub-MIP of P given by fixing $x_k = \bar{x}_k$ for all $k \in \mathcal{C}$;

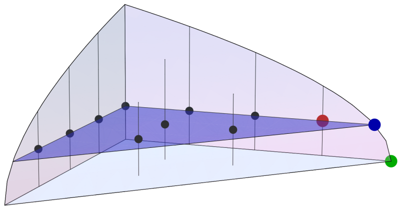


Input: MINLP P
begin
compute a solution $\bar{x}$ of an approximation of P;

round $\bar{x}_k$ for all $k \in \mathcal{I}$;

determine a cover $\mathcal{C}$ of P;

solve the sub-MIP of P given by fixing $x_k = \bar{x}_k$ for all $k \in \mathcal{C}$;



Input: MINLP P

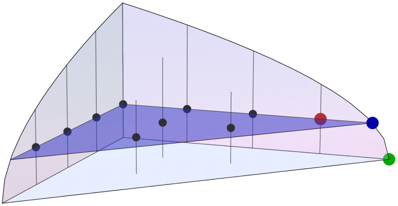
begin

compute a solution $\bar{x}$ of an approximation of P;

round $\bar{x}_k$ for all $k \in \mathcal{I}$;

determine a cover $\mathcal{C}$ of P;

solve the sub-MIP of P given by fixing $x_k = \bar{x}_k$ for all $k \in \mathcal{C}$;



Input: MINLP P

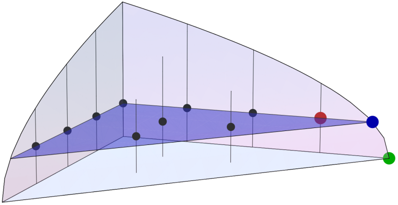
begin

compute a solution $\bar{x}$ of an approximation of P;

round $\bar{x}_k$ for all $k \in \mathcal{I}$;

determine a cover $\mathcal{C}$ of P;

solve the sub-MIP of P given by fixing $x_k = \bar{x}_k$ for all $k \in \mathcal{C}$;



Remark:

- MIP heuristics: trade-off fixing many vs. few variables

  here: eliminate nonlinearities by fixing as few as possible variables

Input: MINLP P

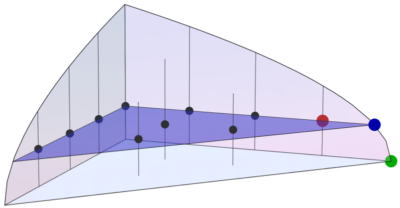
begin

compute a solution $\bar{x}$ of an approximation of P;

round $\bar{x}_k$ for all $k \in \mathcal{I}$;

determine a cover $\mathcal{C}$ of P;

solve the sub-MIP of P given by fixing $x_k = \bar{x}_k$ for all $k \in \mathcal{C}$;



Remark:

- MIP heuristics: trade-off fixing many vs. few variables

  here: eliminate nonlinearities by fixing as few as possible variables
- How to find minimum cover?

Let $g_j : \mathbb{R}^n \to \mathbb{R}, x \mapsto x^T Qx + qx + c,\ Q \in \mathbb{R}^{n\times n}$ symmetric for all j.

Auxiliary binary variables: $\alpha_k = 1 :\Leftrightarrow\ x_k$ is fixed in P

Set covering constraints:

$\mathcal{C}(\alpha) := \{k \mid \alpha_k = 1\}$ is a cover of P if and only if

$$\alpha_k = 1 \qquad \text{for all square nonzeros: } Q^i_{kk} \neq 0, \qquad \text{(SN)}$$

$$\alpha_k + \alpha_j \geqslant 1 \qquad \text{for all bilinear nonzeros: } Q^i_{kj} \neq 0, k \neq j. \qquad \text{(BN)}$$

Solve the covering problem

$$\min \left\{ \sum_{k=1}^{n} \alpha_k : \text{(SN)}, \text{(BN)}, \alpha \in \{0,1\}^n \right\}.$$

Let $g_j : \mathbb{R}^n \to \mathbb{R}, x \mapsto x^T Qx + qx + c,\ Q \in \mathbb{R}^{n\times n}$ symmetric for all j.

Auxiliary binary variables: $\alpha_k = 1 :\Leftrightarrow\ x_k$ is fixed in P

Set covering constraints:

$\mathcal{C}(\alpha) := \{k \mid \alpha_k = 1\}$ is a cover of P if and only if

$$\alpha_k = 1 \qquad \text{for all square nonzeros: } Q^i_{kk} \neq 0, \tag{SN}$$

$$\alpha_k + \alpha_j \geqslant 1 \qquad \text{for all bilinear nonzeros: } Q^i_{kj} \neq 0, k \neq j. \tag{BN}$$

Solve the covering problem

$$\min\left\{\sum_{k=1}^{n} \alpha_k : (\text{SN}), (\text{BN}), \alpha \in \{0,1\}^n\right\}.$$

Let $g_j : \mathbb{R}^n \to \mathbb{R}, x \mapsto x^T Qx + qx + c,\ Q \in \mathbb{R}^{n\times n}$ symmetric for all j.

Auxiliary binary variables: $\alpha_k = 1 :\Leftrightarrow\ x_k$ is fixed in P

Set covering constraints:

$\mathcal{C}(\alpha) := \{k \mid \alpha_k = 1\}$ is a cover of P if and only if

$$\alpha_k = 1 \qquad \text{for all square nonzeros: } Q^i_{kk} \neq 0, \qquad \text{(SN)}$$

$$\alpha_k + \alpha_j \geqslant 1 \qquad \text{for all bilinear nonzeros: } Q^i_{kj} \neq 0, k \neq j. \qquad \text{(BN)}$$

Solve the covering problem

$$\min \left\{ \sum_{k=1}^{n} \alpha_k : \text{(SN)}, \text{(BN)}, \alpha \in \{0,1\}^n \right\}.$$

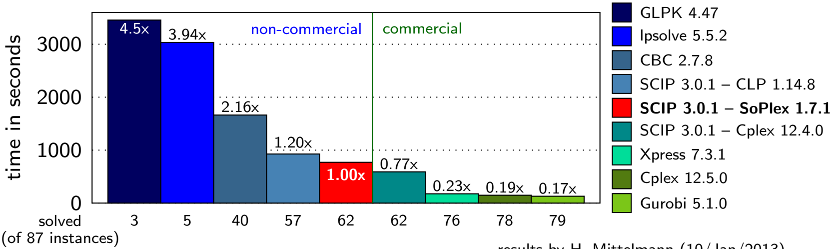
- Covering problem is $\mathcal{NP}$-hard, but branch-and-cut (empirically) very fast

- For general MINLPs, the covering problem becomes more difficult, e.g. for a global cover of a monomial $x_1^{p_1} \cdots x_n^{p_n}$, $p_1, \ldots, p_n \in \mathbb{N}_0$:

$$\alpha_k = 1 \qquad \text{for all } p_k \geqslant 2$$
$$\sum_{k:p_k=1} (1 - \alpha_k) \leqslant 1.$$

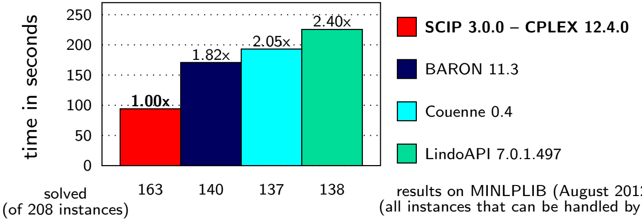
- For general MINLPs, global covers become larger and larger.

- one of the fastest non-commercial MIP solvers



results by H. Mittelmann (10/Jan/2013)

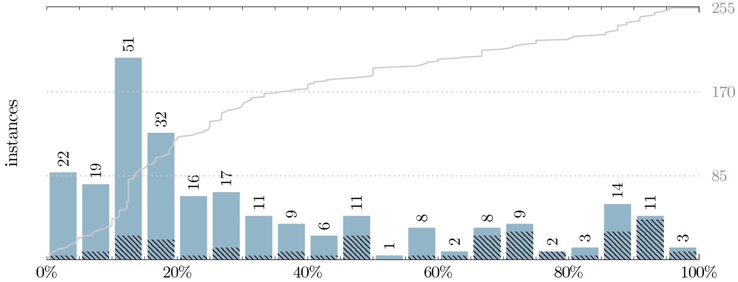
- MINLP solver benchmark



results on MINLPLIB (August 2012)
(all instances that can be handled by all solvers)

How big are minimum covers?

- ▷ SCIP 3.0 on 255 instances from MINLPLib



- ▷ $\leqslant$ 14% on 85, $\leqslant$ 36% on 170 instances
- ▷ similar on MIQCPs and general MINLPs

Test set

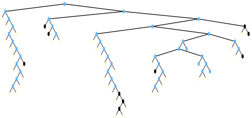
- 149 MIQCPs from GloMIQO test set

Comparison to other heuristics

- Undercover: solution for 76 instances (typically less than 0.1 sec)
- root heuristics: Baron 65, Couenne 55, SCIP 98
- lower success rate on general MINLPs

Extensions

- cover structure can be further exploited
- Undercover branching [B., Gleixner. Proc. of SEA 2013]



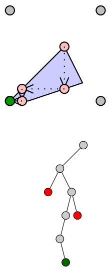
From structures to heuristics to global solvers

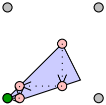
1. Undercover: the largest sub-MIP [B., Gleixner. Math. Prog., 2013]
2. RENS: the optimal rounding [B. Math. Prog. C, under review]
3. Measuring the impact of primal heuristics [B. OR Letters, accepted]

The motivation(s)

First motivation:

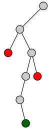
- many MIP heuristics based on rounding
- How likely to succeed?
- How good can they get?

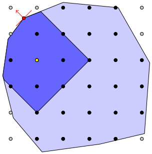




First motivation:

- many MIP heuristics based on rounding
- How likely to succeed?
- How good can they get?

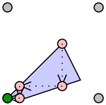


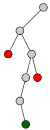


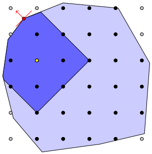
Second motivation:

- Large Neighborhood Search
  - e.g., RINS, Local Branching
- typically requires incumbent solution


BOINK!
BOINK!







Definition (rounding)

Let $\bar{x} \in [L, U]$. The set

$$\mathcal{R}(\bar{x}) := \{x \in [L, U] \mid x_j \in \{\lfloor \bar{x}_j \rfloor, \lceil \bar{x}_j \rceil\} \; \forall j \in \mathcal{I}\}$$

is called the set of *roundings* of $\bar{x}$.

Definition (optimal rounding)

Let $\bar{x} \in [L, U]$ and $\tilde{x} \in \mathcal{R}(\bar{x})$.

1. We call $\tilde{x}$ a *feasible rounding* of $\bar{x}$, if $g_i(\tilde{x}) \leqslant 0 \; \forall \, i \in \mathcal{M}$.
2. We call $\tilde{x}$ an *optimal rounding* of $\bar{x}$, if $\tilde{x} \in \operatorname{argmin}\{d^T x \mid x \in \mathcal{R}(x), g_i(x) \leqslant 0 \; \forall \, i \in \mathcal{M}\}$.
3. We call $\bar{x}$ *roundable* if it has a feasible rounding.

Input: MINLP P

begin

$\bar{x} \leftarrow$ relaxation optimum;

Fix all integral variables
$x_i := \bar{x}_i \quad$ for all $i : \bar{x}_i \in \mathbb{Z}$;

Reduce domain of fractional variables
$x_i \in \{\lfloor \bar{x}_i \rfloor; \lceil \bar{x}_i \rceil\}$;

Solve the resulting sub-MINLP;

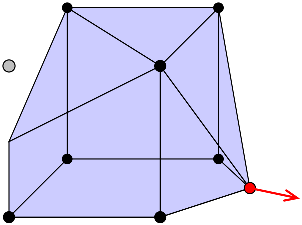
Input: MINLP P
begin

$\bar{x} \leftarrow$ relaxation optimum;

Fix all integral variables
$x_i := \bar{x}_i$ for all $i : \bar{x}_i \in \mathbb{Z}$;

Reduce domain of fractional variables
$x_i \in \{\lfloor \bar{x}_i \rfloor; \lceil \bar{x}_i \rceil\}$;

Solve the resulting sub-MINLP;



Input: MINLP P

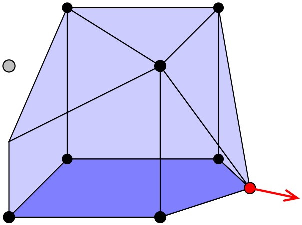
begin

$\bar{x} \leftarrow$ relaxation optimum;

Fix all integral variables
$x_i := \bar{x}_i$ for all $i : \bar{x}_i \in \mathbb{Z}$;

Reduce domain of fractional variables
$x_i \in \{\lfloor \bar{x}_i \rfloor; \lceil \bar{x}_i \rceil\}$;

Solve the resulting sub-MINLP;



Input: MINLP P

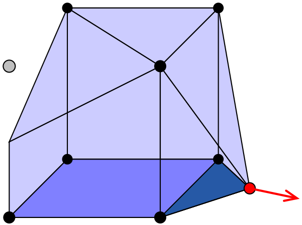
begin

$\bar{x} \leftarrow$ relaxation optimum;

Fix all integral variables
$x_i := \bar{x}_i$ for all $i : \bar{x}_i \in \mathbb{Z}$;

Reduce domain of fractional variables
$x_i \in \{\lfloor \bar{x}_i \rfloor; \lceil \bar{x}_i \rceil\}$;

Solve the resulting sub-MINLP;



Input: MINLP P

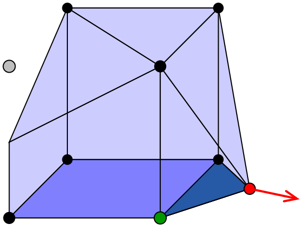
begin

$\bar{x} \leftarrow$ relaxation optimum;

Fix all integral variables
$x_i := \bar{x}_i \quad$ for all $i : \bar{x}_i \in \mathbb{Z}$;

Reduce domain of fractional variables
$x_i \in \{\lfloor \bar{x}_i \rfloor ; \lceil \bar{x}_i \rceil\}$;

Solve the resulting sub-MINLP;



Lemma

Let the starting point $\bar{x}$ be feasible for a relaxation.

1. If the sub-MINLP is infeasible, then no feasible rounding of $\bar{x}$ exists.
2. If not, the optimum of the sub-MINLP is the optimal rounding of $\bar{x}$.

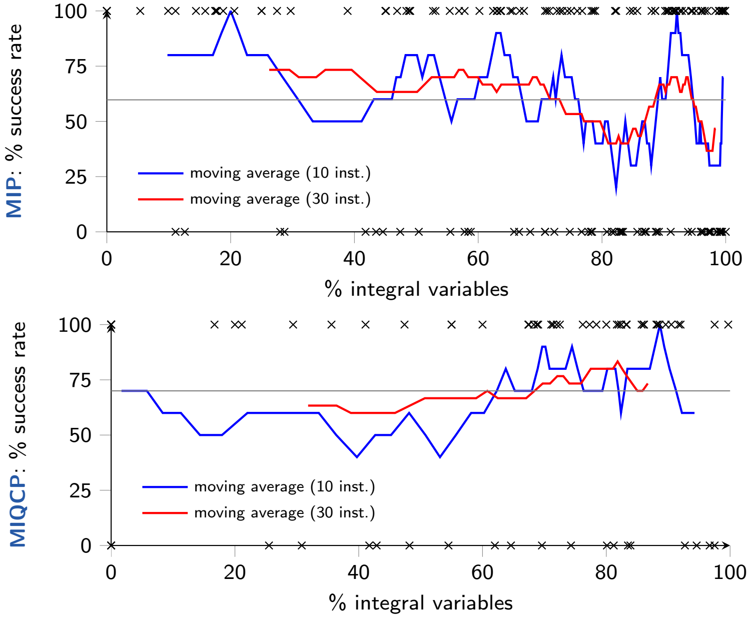
Corollary

$\bar{x}$ either has an optimal rounding or is not roundable.

Computational results

Results when using an LP relaxation:

	MIPLIB	MIQCP	MINLPLib
integrality (% vars)	71.7	59.9	63.5
roundability (% inst)	59.7	70.0	61.9


MIP: % success rate
100
75
50
25
0
moving average (10 inst.)
moving average (30 inst.)
0
20
40
60
80
100
% integral variables
MIQCP: % success rate
100
75
50
25
0
moving average (10 inst.)
moving average (30 inst.)
0
20
40
60
80
100
% integral variables

Results when using an LP relaxation:

	MIPLIB	MIQCP	MINLPLib
integrality (% vars)	71.7	59.9	63.5
roundability (% inst)	59.7	70.0	61.9

Further findings

- cuts: no effect on integrality, bad for roundability
- NLP solutions much less integral, similar roundability

Results when using an LP relaxation:

	MIPLIB	MIQCP	MINLPLib
integrality (% vars)	71.7	59.9	63.5
roundability (% inst)	59.7	70.0	61.9

Further findings

- cuts: no effect on integrality, bad for roundability
- NLP solutions much less integral, similar roundability

RENS as primal heuristic

- defensive version succeeds in $\approx 50\%$ (MIP and MIQCP)
- inferior to portfolio, superior to single best
- mixed runtimes: mostly ≤ 5 sec, for MIP sometimes ≈ 1 min

From structures to heuristics to global solvers

1. Undercover: the largest sub-MIP [B., Gleixner. Math. Prog., 2013]
2. RENS: the optimal rounding [B. Math. Prog. C, under review]
3. Measuring the impact of primal heuristics [B. OR Letters, accepted]

And what about the global solvers?

How important are primal heuristics in global solvers?

A MIP software vendor says:

Our advanced MIP heuristics for quickly finding feasible solutions often produce good quality solutions where other solvers fall flat, leading to some of our biggest wins vs. the competition.

And what about the global solvers?

How important are primal heuristics in global solvers?

A MIP software vendor says:

Our advanced MIP heuristics for quickly finding feasible solutions often produce good quality solutions where other solvers fall flat, leading to some of our biggest wins vs. the competition.

Which heuristics?

- trade secret
- few parameters to influence behavior of heuristics
- output only tells you that solution was found by **some** heuristic

And what about the global solvers?

How important are primal heuristics in global solvers?

A MIP software vendor says:

Our advanced MIP heuristics for quickly finding feasible solutions often produce good quality solutions where other solvers fall flat, leading to some of our biggest wins vs. the competition.

Which heuristics?

- trade secret
- few parameters to influence behavior of heuristics
- output only tells you that solution was found by **some** heuristic

$\rightsquigarrow$ **very important**

And what about the global solvers?

How important are primal heuristics?

Typical measure: running time to prove optimality

And what about the global solvers?

How important are primal heuristics?

Typical measure: running time to prove optimality

And what about the global solvers?

How important are primal heuristics?

Typical measure: running time to prove optimality

- one vendor: 6% improvement
- other vendor: 9% improvement
- non-commercial solver: 15 % improvement

$\rightsquigarrow$ **not important at all**

And what about the global solvers?

How important are primal heuristics?

Typical measure: running time to prove optimality

- ▷ one vendor: 6% improvement
- ▷ other vendor: 9% improvement
- ▷ non-commercial solver: 15 % improvement

$\rightsquigarrow$ **not important at all**

So, what is wrong here?

Comparing performance

How to measure the added value of a primal heuristic?

- time to optimality t_{solved}, number of branch-and-bound nodes
  - very much depends on dual bound
- time to best solution t_{opt}
  - nearly optimal solution might be found long before
- time to first solution t_1
  - disregards solution quality
- performance profiles
  - depend on t_{solved}, hence on dual bound
  - not an absolute number
- primal integral

Let $\tilde{x}$ be a solution, $\tilde{x}_{\text{opt}}$ be an optimum, $\mathrm{t}_{\max} \in \mathbb{R}_{\geq 0}$ be a timelimit.

Primal gap $\gamma \in [0, 1]$ of $\tilde{x}$:

$$\gamma(\tilde{x}) := \begin{cases} 0, & \text{if } \left|c^T \tilde{x}_{\text{opt}}\right| = \left|c^T \tilde{x}\right| = 0, \\ 1, & \text{if } c^T \tilde{x}_{\text{opt}} \cdot c^T \tilde{x} < 0, \\ \frac{\left|c^T \tilde{x}_{\text{opt}} - c^T \tilde{x}\right|}{\max\{\left|c^T \tilde{x}_{\text{opt}}\right|, \left|c^T \tilde{x}\right|\}}, & \text{else.} \end{cases}$$

Primal gap function $p\colon [0, \mathrm{t}_{\max}] \mapsto [0, 1]$:

$$p(t) := \begin{cases} 1, & \text{if no incumbent until point } t, \\ \gamma(\tilde{x}(t)), & \text{with } \tilde{x}(t) \text{ incumbent at point } t. \end{cases}$$

- step function, changes at points t_i when new incumbent found
- $p(0) = 1$, $p(t) = 0$ for all $t \geq t_{\text{opt}}$
- monotonically decreasing

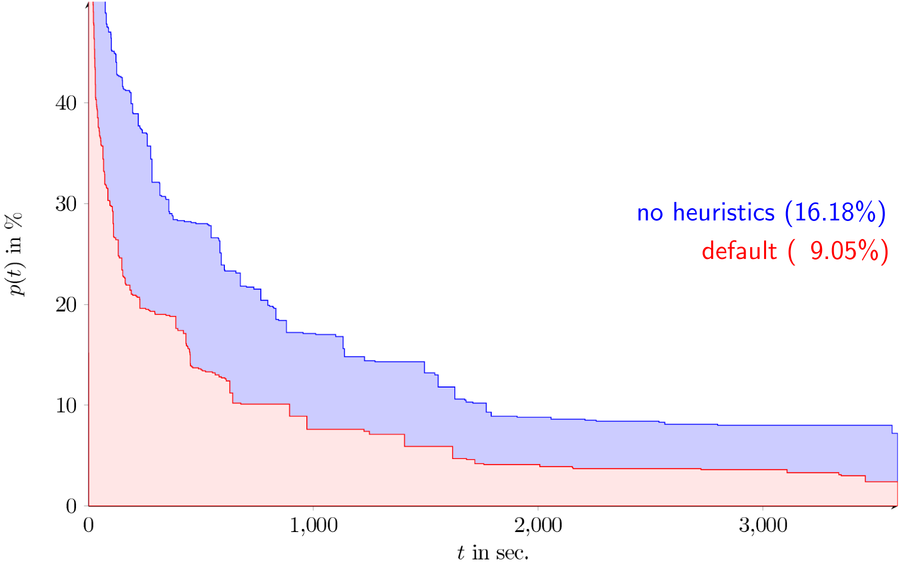
Primal integral $P(T)$ of $T \in [0, \mathrm{t}_{\max}]$:

$$P(T) := \int_{t=0}^{T} p(t)\, dt = \sum_{i=1}^{I} p(t_{i-1}) \cdot (t_i - t_{i-1}),$$

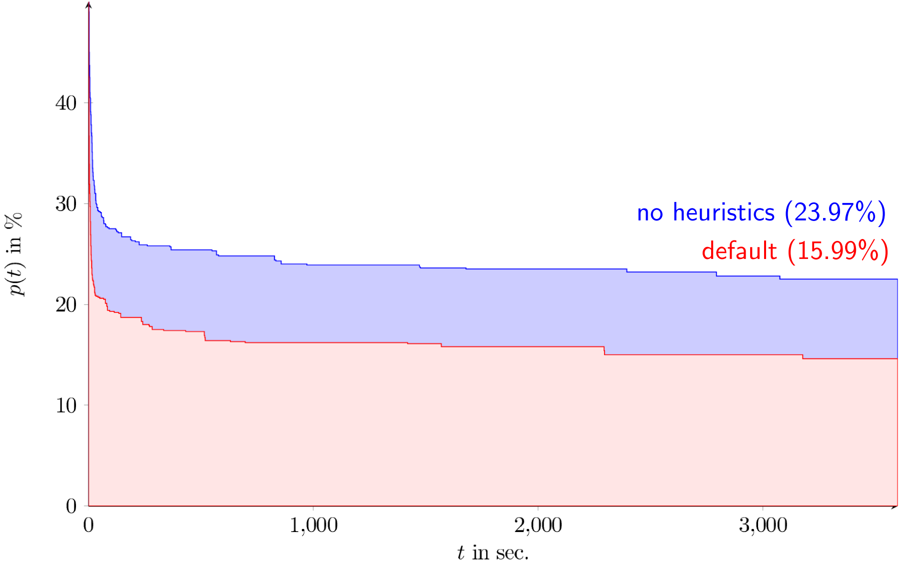
How to measure the added value of a primal heuristic?

- time to optimality t_{solved}, number of branch-and-bound nodes
  - very much depends on dual bound
- time to best solution t_{opt}
  - nearly optimal solution might be found long before
- time to first solution t_1
  - disregards solution quality
- performance profiles
  - depend on t_{solved}, hence on dual bound
  - not an absolute number
- primal integral $P(\mathrm{t}_{\max})$
  - favors finding good solutions early
  - considers each update of incumbent
  - $P(\mathrm{t}_{\max})/\mathrm{t}_{\max}$ “average solution quality”
  - expected quality of the incumbent, if stopped arbitrarily

Average primal integral: SCIP / MIP



Average primal integral: SCIP / MINLP



Undercover & RENS:

- LNS start heuristics for MINLP
- exploit generic problem structures
- successful application as root node heuristics

Primal integral:

- new performance measure, captures overall solution process
- more robust, though not immune, against randomness
- overall impact of primal heuristics (w.r.t. $P(\mathrm{t}_{\mathrm{max}})$) significant

slides and technical reports of presented papers: http://www.zib.de/berthold

SCIP Optimization Suite: http://scip.zib.de

From structures to heuristics to global solvers

Timo Berthold

Zuse Institute Berlin