Convex Mixed-Integer Nonlinear Optimization II

Summer School on Optimization of Dynamical Systems

Sven Leyffer and Jeff Linderoth

Argonne National Laboratory

September 3-7, 2018

Outline

Mixed-Integer Nonlinear Optimization

Mixed-Integer Nonlinear Program (MINLP)

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & c(x) \le 0 \\ & x \in \mathcal{X} \\ & x_i \in \mathbb{Z} \text{ for all } i \in \mathcal{I} \end{aligned}$$

Basic Assumptions for Convex MINLP

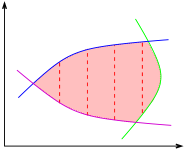
- A1 $\mathcal{X}$ is a bounded polyhedral set.
- A2 f and c twice continuously differentiable convex
- A3 MINLP satisfies a constraint qualification.

A2 (convexity) most restrictive (show how to relax later)
A3 is technical (MFCQ would have been sufficient)

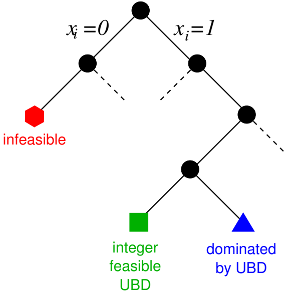
Outline

Recall: Methods for MINLP

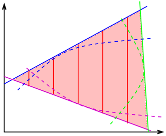
Branch-and-Bound



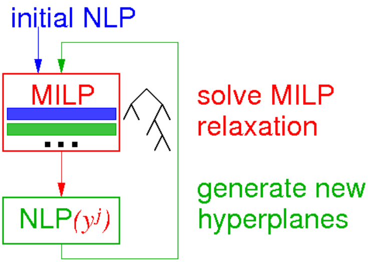
... tree-search method



Outer Approximation



... alternate MILP & NLP

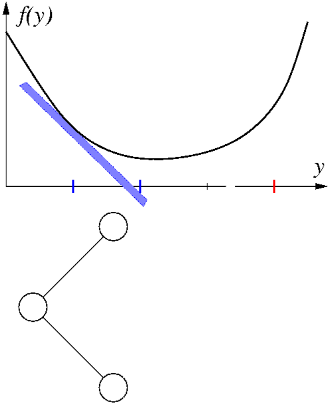


... now build hybrid method!

LP/NLP-Based Branch-and-Bound

Aim: avoid solving expensive MILPs

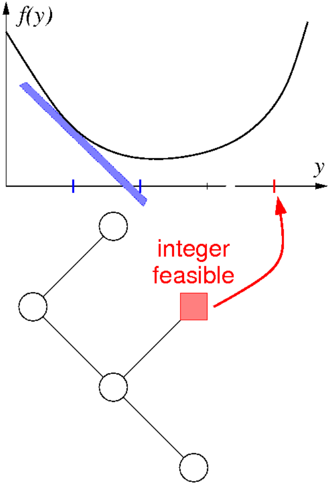
- Start solving master MILP ... using MILP branch-and-cut



LP/NLP-Based Branch-and-Bound

Aim: avoid solving expensive MILPs

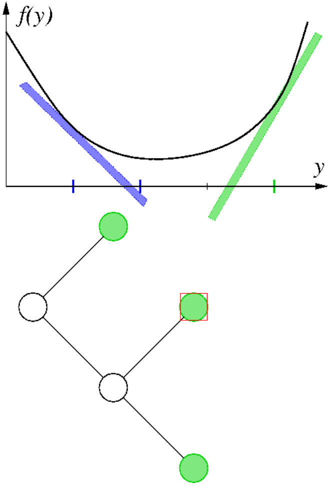
- Start solving master MILP ... using MILP branch-and-cut
- If $x_I^{(j)}$ integral, then interrupt MILP; solve NLP($x_I^{(j)}$) get $x^{(j)}$



LP/NLP-Based Branch-and-Bound

Aim: avoid solving expensive MILPs

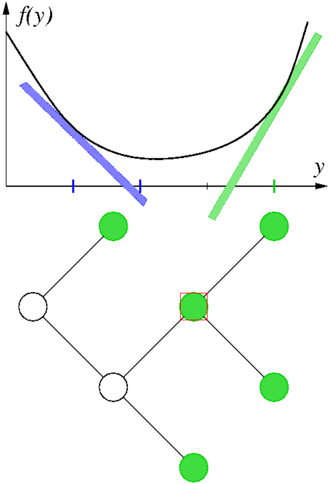
- Start solving master MILP ... using MILP branch-and-cut
- If $x_I^{(j)}$ integral, then interrupt MILP; solve NLP($x_I^{(j)}$) get $x^{(j)}$
- Linearize f, c about $x^{(j)}$
  $\Rightarrow$ add linearization to tree



LP/NLP-Based Branch-and-Bound

Aim: avoid solving expensive MILPs

- Start solving master MILP ... using MILP branch-and-cut
- If $x_I^{(j)}$ integral, then interrupt MILP; solve NLP($x_I^{(j)}$) get $x^{(j)}$
- Linearize f, c about $x^{(j)}$ $\Rightarrow$ add linearization to tree
- Continue MILP tree-search

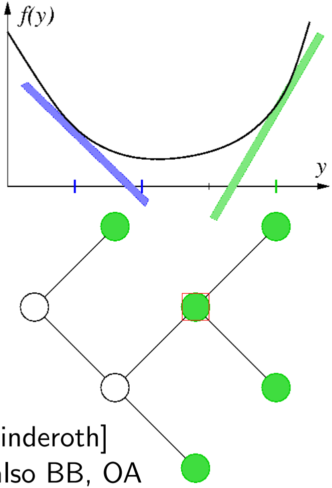


LP/NLP-Based Branch-and-Bound

Aim: avoid solving expensive MILPs

- Start solving master MILP ... using MILP branch-and-cut
- If $x_I^{(j)}$ integral, then interrupt MILP; solve NLP($x_I^{(j)}$) get $x^{(j)}$
- Linearize f, c about $x^{(j)}$
  $\Rightarrow$ add linearization to tree
- Continue MILP tree-search

... until lower bound $\geq$ upper bound

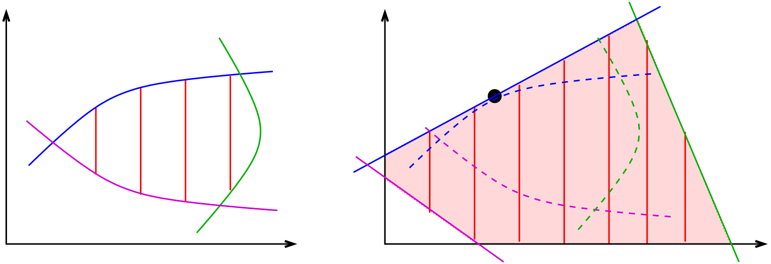


Software:
`FilMINT`: FilterSQP + MINTO [L & Linderoth]
`BONMIN`: IPOPT + CBC [IBM/CMU] also BB, OA

LP/NLP Branch and Bound

LP/NLP-based branch-and-bound

- Branch-and-cut algorithm with cuts from NLP solves
- Create MILP relaxation of MINLP



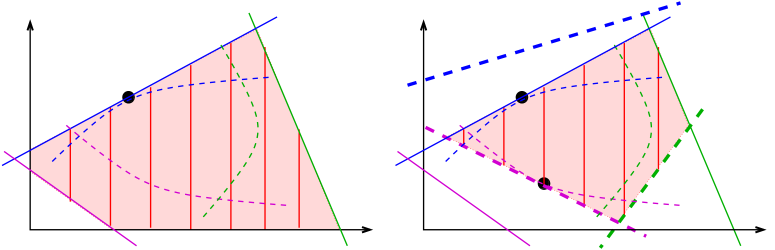
$0 \geq c(x)$ $\qquad$ $0 \geq c^{(k)} + \nabla c^{(k)^T}(x - x^{(k)})$

- Search MILP-tree $\Rightarrow$ faster re-solves
- Interrupt MILP tree-search to create new linearizations

LP/NLP Branch and Bound

LP/NLP-based branch-and-bound

- Branch-and-cut algorithm with cuts from NLP solves
- Create MILP relaxation of MINLP & refine linearizations



$$0 \geq c(x) \qquad\qquad 0 \geq c^{(k)} + \nabla c^{(k)^T}(x - x^{(k)})$$

- Search MILP-tree $\Rightarrow$ faster re-solves
- Interrupt MILP tree-search to create new linearizations

LP/NLP-Based Branch-and-Bound

Algorithmic refinements, e.g. [Abhishek et al., 2010]

- Advanced MILP search and cut management techniques
  ... remove "old" OA cuts from LP relaxation ⇒ faster LP
- Generate cuts at non-integer points: ECP cuts are cheap
  ... generate cuts early (near root) of tree
- Strong branching, adaptive node selection & cut management
  - Fewer nodes, if we add more cuts (e.g. ECP cuts)
  - More cuts make LP harder to solve
    ⇒ remove outdated/inactive cuts from LP relaxation

  ... balance OA accuracy with LP solvability
- Compress OA cuts into Benders cuts can be OK

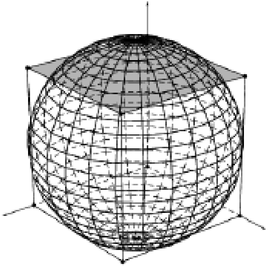
Interpret as hybrid algorithm, [Bonami et al., 2008]

Benders and ECP versions are also possible.

Worst Case Example of Outer Approximation

[Hijazi et al., 2010] construct infeasible MINLP:

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & 0 \\ \text{subject to} \quad & \sum_{i=1}^{n} \left(y_i - \frac{1}{2} \right)^2 \leq \frac{n-1}{4} \\ & y \in \{0,1\}^n \end{aligned}$$



Intersection of ball of radius $\frac{\sqrt{n-1}}{2}$ with unit hypercube.

Lemma

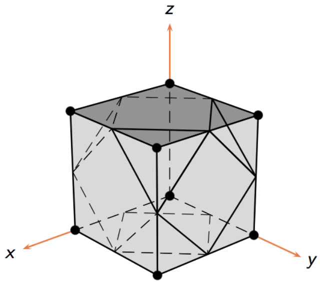
OA cannot cut more than one vertex of the hypercube
MILP master problem feasible for any $k < 2^n$ OA cuts

Theorem

OA visits all 2^n vertices

Worst-Case Example for Linearizationm-Based Methods

- No OA constraint can cut 2 vertices of the hypercube.
  - If an inequality cuts two vertices, it cuts the segment joining them. This can not be: the ball has a non-empty intersection with any such segment.



A "basic" implementation of a linearization-based method for solving this problem enumerates at least 2^n nodes

How Bad Is It?

n	2^n	CPLEX 12.4 nodes	SCIP 2.1 nodes	Bonmin (Hyb) nodes
10	1024	2047	720	11156
15	32,768	65535	31993	947014
20	1,048,576	2,097,151	1,216,354	

- One "trick": The problem is simple for CPLEX/SCIP if variables are 0-1.
- If $x \in \{0, 1\}$, then replace x_i^2 by x_i and the contradiction $n/4 < (n-1)/4$ follows immediately.

Outline

Separability and Constraint Disaggregation

Nonlinear disaggregation [Tawarmalani and Sahinidis, 2005]

$$S := \{x \in \mathbb{R}^n : c(x) = h(g(x)) \le 0\},$$

$g : \mathbb{R}^n \to \mathbb{R}^p$ smooth convex;
$h : \mathbb{R}^p \to \mathbb{R}$ smooth, convex, and nondecreasing
$\Rightarrow c(x)$ smooth convex

Disaggregated formulation: introduce $y = g(x) \in \mathbb{R}^p$

$$S_d := \left\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^p : h(y) \le 0,\ y \ge g(x)\right\}.$$

Lemma

S is projection of S_d onto x.

Separability and Constraint Disaggregation

Consider

$$S := \{x \in \mathbb{R}^n : c(x) = h(g(x)) \le 0\},$$

and

$$S_d := \left\{(x, y) \in \mathbb{R}^n \times \mathbb{R}^p : h(y) \le 0,\ y \ge g(x)\right\}.$$

Theorem

Any outer approximation of S_d is stronger than OA of S

Given $\mathcal{X}^k := \left\{x^{(1)}, \ldots, x^{(k)}\right\}$ construct OA for S, S_d:

$$S^{oa} := \left\{x : c^{(l)} + \nabla c^{(l)^T}(x - x^{(l)}) \le 0,\ \forall x^{(l)} \in \mathcal{X}^k\right\}$$

$$S_d^{oa} := \left\{(x, y) : h^{(l)} + \nabla h^{(l)^T}(y - g(x^{(l)})) \le 0,\right.$$
$$\left. y \ge g^{(l)} + \nabla g^{(l)^T}(x - x^{(l)}),\ \forall x^{(l)} \in \mathcal{X}^k\right\},$$

[Tawarmalani and Sahinidis, 2005] show S_d^{oa} stronger than S^{oa}

Separability and Constraint Disaggregation

[Hijazi et al., 2010] study

$$\left\{ x : c(x) := \sum_{j=1}^{q} h_j(a_j^T x + b_j) \le 0 \right\}$$

where $h_j : \mathbb{R} \to \mathbb{R}$ are smooth and convex

Disaggregated formulation: introduce $y \in \mathbb{R}^q$

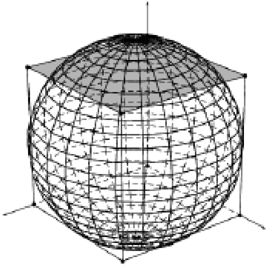
$$\left\{ (x, y) : \sum_{j=1}^{q} y_j \le 0, \quad \text{and} \quad y_j \ge h_j(a_j^T x + b_j) \right\}$$

can be shown to be tighter

Recall: Worst Case Example of Outer Approximation

Apply disaggregation to [Hijazi et al., 2010] example:

$$\begin{aligned} \underset{y}{\text{minimize}} \quad & 0 \\ \text{subject to} \quad & \sum_{i=1}^{n} \left(x_i - \frac{1}{2} \right)^2 \leq \frac{n-1}{4} \\ & x \in \{0,1\}^n \end{aligned}$$



Intersection of ball of radius $\frac{\sqrt{n-1}}{2}$ with unit hypercube.

Disaggregate $\sum \left(x_i - \frac{1}{2}\right)^2 \leq \frac{n-1}{4}$ as

$$\sum_{i=1}^{n} y_i \leq 0 \quad \text{and} \quad \left(x_i - \frac{1}{2} \right)^2 \leq y_i$$

Separability and Constraint Disaggregation

[Hijazi et al., 2010] disaggregation on worst-case example of OA

- Linearize around $x^{(1)} \in \{0,1\}^n$ and complement $x^{(2)} := e - x^{(1)}$, where $e = (1, \dots, 1)$
- OA of disaggregated constraint is

$$\sum_{i=1}^{n} y_i, \quad \text{and} \quad x_i - \tfrac{3}{4} \le y_i, \quad \text{and} \quad \tfrac{1}{4} - x_i \le y_i,$$

- Using $x_i \in \{0,1\}$ implies $z_i \geq 0$, implies $\sum z_i \geq \frac{n}{4} > \frac{n-1}{4}$

$\Rightarrow$ OA-MILP master of $x^{(1)}$ and $x^{(2)}$ is infeasible.

... terminate in two iterations

Extension: Group-Partial Separability

Definition (Group Partially Separability [Conn et al., 1992])

A nonlinear function $f(x)$ is group partially separable, iff

$$f(x) = \sum_{j=1}^{q} g_j \left(a_j^T x + b_j + \sum_{i \in \mathcal{E}_j} f_i(x_{[i]}) \right)$$

where $f_i : \mathbb{R}^{n_i} \to \mathbb{R}$ depends on subvector $x_{[i]}$ of x where $n_i \ll n$, and $g_j : \mathbb{R} \to \mathbb{R}$ are univariate functions.

- Extends partial separability ($\sum_i f_i(x_{[i]})$) and sparsity
- Structured quasi-Newton updates for large-scale optimization
- Most complex functional form for which Hessian of augmented Lagrangian can be computed from element Hessians $\nabla^2 f_i(x_{[i]})$

Outline

Perspective Formulations

MINLPs use binary indicator variables, x_b, to model nonpositivity of $x_c \in \mathbb{R}$

Model as variable upper bound

$$0 \leq x_c \leq u_c x_b, \quad x_b \in \{0,1\}$$

$\Rightarrow$ if $x_c > 0$, then $x_b = 1$

Perspective reformulation applies, if x_b also in convex $c(x) \leq 0$

- Significantly improve reformulation
- Pioneered by [Frangioni and Gentile, 2006];
  ... strengthen relaxation using perspective cuts

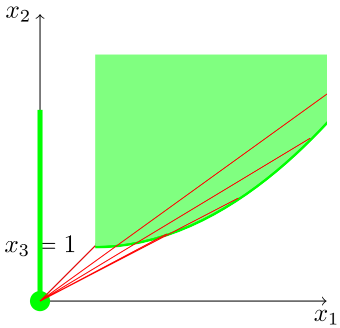
Example of Perspective Formulation

Consider MINLP set with three variables:

$$S = \left\{(x_1, x_2, x_3) \in \mathbb{R}^2 \times \{0,1\} \ : \ x_2 \geq x_1^2, \quad ux_3 \geq x_1 \geq 0\right\}.$$

Can show that $S = S^0 \cup S^1$, where

$$S^0 = \left\{(0, x_2, 0) \in \mathbb{R}^3 \ : \ x_2 \geq 0\right\},$$
$$S^1 = \left\{(x_1, x_2, 1) \in \mathbb{R}^3 \ : \ x_2 \geq x_1^2, \ u \geq x_1 \geq 0\right\}.$$



Example of Perspective Formulation

Geometry of convex hull of S:
Lines connecting origin ($x_3 = 0$) to parabola $x_2 = x_1^2$ at $x_3 = 1$

Define convex hull of S as conv(S)

$$:= \left\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_2 x_3 \geq x_1^2,\ u x_3 \geq x_1 \geq 0, 1 \geq x_3 \geq 0, x_2 \geq 0\right\}$$

where $x_2 x_3 \geq x_1^2$is defined in terms of perspective function

$$\mathcal{P}_f(x, z) := \begin{cases} 0 & \text{if } z = 0, \\ z f(x/z) & \text{if } z > 0. \end{cases}$$

Epigraph of $\mathcal{P}_f(x, z)$: cone pointed at origin with lower shape $f(x)$

$x_b \in \{0, 1\}$ indicator forces $x_c = 0$, or $c(x_c) \leq 0$ if $x_b = 1$ write

$x_b c(x_c / x_b)$ ...is tighter convex formulation

Generalization of Perspective Cuts

[Günlük and Linderoth, 2012] consider more general problem

$$(P) \qquad \min_{(x,z,\eta)\in\mathbb{R}^n\times\{0,1\}\times\mathbb{R}} \Big\{\eta \mid \eta \geq f(x) + cz,\, Ax \leq bz\Big\}.$$

where

1. $X = \{x \mid Ax \leq b\}$ is bounded
2. $f(x)$ is convex and finite on X, and $f(0) = 0$

Theorem (Perspective Cut)

For any $\bar{x} \in X$ and subgradient $s \in \partial f(\bar{x})$, the inequality

$$\eta \geq f(\bar{x}) + c + s^T(x - \bar{x}) + (c + f(\bar{x}) - s^T\bar{x}))(z - 1)$$

is valid cut for (P)

Stronger Relaxations [Günlük and Linderoth, 2012]

- z_R: Value of NLP relaxation
- z_{GLW}: Value of NLP relaxation after GLW cuts
- z_P: Value of perspective relaxation
- z^*: Optimal solution value

Separable Quadratic Facility Location Problems

$\|M\|$	$\|N\|$	z_R	z_{GLW}	z_P	z^*
10	30	140.6	326.4	346.5	348.7
15	50	141.3	312.2	380.0	384.1
20	65	122.5	248.7	288.9	289.3
25	80	121.3	260.1	314.8	315.8
30	100	128.0	327.0	391.7	393.2

$\Rightarrow$ Tighter relaxation gives faster solves!

Mini-Quiz on Nonlinear Perspective of the Perspective

Perspective of our uncapacitated facility location problem:

$$\begin{aligned} &\underset{x,y,z}{\text{minimize}} \;\; z + y \\ &\text{subject to } (x/z)^2 - (y/z) \leq 0 \quad 0 \leq x \leq z,\; z \in \{0,1\},\; y \geq 0 \end{aligned}$$

which clearly is not defined at $z = 0$.

Why not multiply through by z^2?

$$\begin{aligned} &\underset{x,y,z}{\text{minimize}} \;\; z + y \\ &\text{subject to } x^2 - zy \leq 0 \quad 0 \leq x \leq z,\; z \in \{0,1\},\; y \geq 0 \end{aligned}$$

... isn't this a much nicer NLP?

Numerical Experience with the Bad the Perspective

Bad perspective of uncapacitated facility location problem:

$$\begin{aligned} & \underset{x,y,z}{\text{minimize}} \quad z + y \\ & \text{subject to } x^2 - zy \le 0 \quad 0 \le x \le z,\ z \in \{0,1\},\ y \ge 0 \end{aligned}$$

```
Major Minor TrustRad RegParam StepNorm Constrnts  Objective Optimal Phase Step
------------------------------------------------------------------------------
    0     0       10       10        0       0.5       1.01       0     2
    1     1       10       10    0.625         0      0.385       0     2  SQP
    2     1       10       10    0.188         0     0.1875       0     2  SQP
[ ... ]
   28     1       10       10 2.79e-09         0  2.794e-09       0     2  SQP
   29     1       10       10  1.4e-09         0  1.397e-09       0     2  SQP
   30     1       10       10 6.98e-10         0  6.985e-10       2     2  SQP

ASTROS Version 2.0.2 (20100913): Solution Summary
================================================
  Major iters     =         30 ; Minor iters     =        30 ;
  KKT-residual    =     0.4286 ; Complementarity =  1.996e-10 ;
  Final step-norm =  6.985e-10 ; Final TR-radius =        10 ;
---------------------------------------------------------------
ASTROS  Version 2.0.2 (20100913): Step got too small
```

Linear rate of convergence ... similar for MINOS, FilterSQP, ...

Mini-Quiz on Nonlinear Perspective of the Perspective

Perspective of our uncapacitated facility location problem:

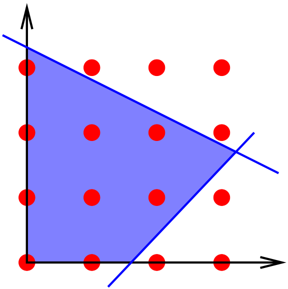
$$\begin{aligned} \underset{x,y,z}{\text{minimize}} \quad & z + y \\ \text{subject to} \quad & (x/z)^2 - (y/z) \le 0 \quad 0 \le x \le z, \; z \in \{0,1\}, \; y \ge 0 \end{aligned}$$

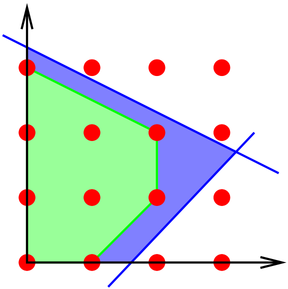
which clearly is not defined at $z = 0$ ...

... yet ... this may be OK, after all:

1. Show that the perspective of QFL is well-defined at $z = 0$.
2. Can you compute the gradients, $\nabla c(x, y, z)$, are they defined?
3. What about $\nabla^2 c$?

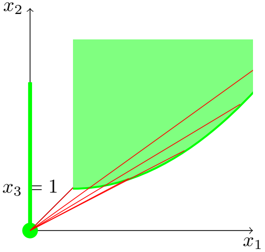
Summary and Teaching Points

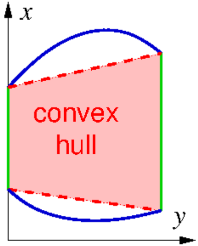




Classes of Cuts

1. Perspective cuts
2. Disjunctive cuts (not discussed)





Abhishek, K., Leyffer, S., and Linderoth, J. T. (2010).
FilMINT: An outer-approximation-based solver for nonlinear mixed integer programs.
INFORMS Journal on Computing, 22:555–567.
DOI:10.1287/ijoc.1090.0373.

Bonami, P., Biegler, L., Conn, A., Cornuéjols, G., Grossmann, I., Laird, C., Lee, J., Lodi, A., Margot, F., Sawaya, N., and Wächter, A. (2008).
An algorithmic framework for convex mixed integer nonlinear programs.
Discrete Optimization, 5(2):186–204.

Conn, A. R., Gould, N. I. M., and Toint, P. L. (1992).
LANCELOT: A Fortran package for large-scale nonlinear optimization (Release A).
Springer Verlag, Heidelberg.

Frangioni, A. and Gentile, C. (2006).
Perspective cuts for a class of convex 0-1 mixed integer programs.
Mathematical Programming, 106:225–236.

Günlük, O. and Linderoth, J. T. (2012).
Perspective reformulation and applications.
In *IMA Volumes*, volume 154, pages 61–92.

Hijazi, H., Bonami, P., and Ouorou, A. (2010).
An outer-inner approximation for separable MINLPs.
Technical report, LIF, Faculté des Sciences de Luminy, Université de Marseille.

Tawarmalani, M. and Sahinidis, N. V. (2005).

A polyhedral branch-and-cut approach to global optimization.
Mathematical Programming, 103(2):225–249.