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

1. Problem Definition and Assumptions
2. Single-Tree Methods
3. Separability and Constraint Disaggregation
4. Cutting Planes for MINLP
Mixed-Integer Nonlinear Optimization

Mixed-Integer Nonlinear Program (MINLP)

\[
\begin{align*}
\text{minimize} & \quad f(x) \\
\text{subject to} & \quad c(x) \leq 0 \\
& \quad x \in \mathcal{X} \\
& \quad x_i \in \mathbb{Z} \text{ for all } i \in \mathcal{I}
\end{align*}
\]

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

1. Problem Definition and Assumptions
2. Single-Tree Methods
3. Separability and Constraint Disaggregation
4. Cutting Planes for MINLP
Recall: Methods for MINLP

Branch-and-Bound

... tree-search method

Outer Approximation

... alternate MILP & NLP

initial NLP

solve MILP relaxation

generate new hyperplanes

... now build hybrid method!
LP/NLP-Based Branch-and-Bound

Aim: avoid solving expensive MILPs

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

\[ f(y) \]

\[ y \]

Software:
- FilMINT: FilterSQP + MINTO [L & Linderoth]
- BONMIN: IPOPT + CBC [IBM/CMU] also BB, OA
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 $x^{(j)}$
  \[ \Rightarrow \text{add linearization to tree} \]
LP/NLP-Based Branch-and-Bound

Aim: avoid solving expensive MILPs

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

\[ 0 \geq c(x) \]

\[ 0 \geq c^{(k)} + \nabla c^{(k)^T} (x - x^{(k)}) \]

- 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

\[
0 \geq c(x) \\
0 \geq c^{(k)} + \nabla c^{(k)^T} (x - x^{(k)})
\]

- Search MILP-tree ⇒ 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:

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

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 Linearization-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?

<table>
<thead>
<tr>
<th>$n$</th>
<th>$2^n$</th>
<th>CPLEX 12.4 nodes</th>
<th>SCIP 2.1 nodes</th>
<th>Bonmin (Hyb) nodes</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>1024</td>
<td>2047</td>
<td>720</td>
<td>11156</td>
</tr>
<tr>
<td>15</td>
<td>32,768</td>
<td>65535</td>
<td>31993</td>
<td>947014</td>
</tr>
<tr>
<td>20</td>
<td>1,048,576</td>
<td>2,097,151</td>
<td>1,216,354</td>
<td></td>
</tr>
</tbody>
</table>

- **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

1. Problem Definition and Assumptions
2. Single-Tree Methods
3. Separability and Constraint Disaggregation
4. Cutting Planes for MINLP
Separability and Constraint Disaggregation

Nonlinear disaggregation [Tawarmalani and Sahinidis, 2005]

\[ S := \{ x \in \mathbb{R}^n : c(x) = h(g(x)) \leq 0 \}, \]

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

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

\[ S_d := \{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^p : h(y) \leq 0, \ y \geq g(x) \}. \]

**Lemma**

\( S \) is projection of \( S_d \) onto \( x \).
Separability and Constraint Disaggregation

Consider

\[ S := \{ x \in \mathbb{R}^n : c(x) = h(g(x)) \leq 0 \} , \]

and

\[ S_d := \{ (x, y) \in \mathbb{R}^n \times \mathbb{R}^p : h(y) \leq 0, \ y \geq g(x) \} . \]

**Theorem**

Any outer approximation of \( S_d \) is stronger than OA of \( S \)

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

\[ S^{oa} := \{ x : c^{(l)} + \nabla c^{(l)\top} (x - x^{(l)}) \leq 0, \ \forall x^{(l)} \in X^k \} \]

\[ S_d^{oa} := \{ (x, y) : h^{(l)} + \nabla h^{(l)\top} (y - g(x^{(l)})) \leq 0, \]

\[ y \geq g^{(l)} + \nabla g^{(l)\top} (x - x^{(l)}), \ \forall x^{(l)} \in X^k \} , \]

[Tawarmalani and Sahinidis, 2005] show \( S_d^{oa} \) stronger than \( S^{oa} \)
Separability and Constraint Disaggregation

[Hijazi et al., 2010] study

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

where \( h_j : \mathbb{R} \rightarrow \mathbb{R} \) are smooth and convex

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

\[
\left\{ (x, y) : \sum_{j=1}^{q} y_j \leq 0, \text{ and } y_j \geq h_j(a_j^T x + b_j) \right\}
\]

can be shown to be tighter
Recall: Worst Case Example of Outer Approximation

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

\[
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{n} y_i \\
\text{subject to} & \quad \sum_{i=1}^{n} \left( x_i - \frac{1}{2} \right)^2 \leq \frac{n-1}{4} \\
& \quad x \in \{0, 1\}^n
\end{align*}
\]

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

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

\[
\sum_{i=1}^{n} y_i \leq 0 \quad \text{and} \quad \left( x_i - \frac{1}{2} \right)^2 \leq y_i
\]
[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 $e = (1, \ldots, 1)$
- OA of disaggregated constraint is

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

- 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

1. Problem Definition and Assumptions
2. Single-Tree Methods
3. Separability and Constraint Disaggregation
4. Cutting Planes for MINLP
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:

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

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 $\text{conv}(S)$

$$\{(x_1, x_2, x_3) \in \mathbb{R}^3 : x_2x_3 \geq x_1^2, ux_3 \geq x_1 \geq 0, 1 \geq x_3 \geq 0, x_2 \geq 0\}$$

where $x_2x_3 \geq x_1^2$ is defined in terms of perspective function

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

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

$$x_b c(x_c/x_b) \quad \text{...is tighter convex formulation}$$
Generalization of Perspective Cuts

[Günlük and Linderoth, 2012] consider more general problem

\[
(P) \quad \min_{(x,z,\eta) \in \mathbb{R}^n \times \{0,1\} \times \mathbb{R}} \left\{ \eta \mid \eta \geq f(x) + cz, Ax \leq bz \right\}.
\]

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

<table>
<thead>
<tr>
<th>$M$</th>
<th>$N$</th>
<th>$z_R$</th>
<th>$z_{GLW}$</th>
<th>$z_P$</th>
<th>$z^*$</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>30</td>
<td>140.6</td>
<td>326.4</td>
<td>346.5</td>
<td>348.7</td>
</tr>
<tr>
<td>15</td>
<td>50</td>
<td>141.3</td>
<td>312.2</td>
<td>380.0</td>
<td>384.1</td>
</tr>
<tr>
<td>20</td>
<td>65</td>
<td>122.5</td>
<td>248.7</td>
<td>288.9</td>
<td>289.3</td>
</tr>
<tr>
<td>25</td>
<td>80</td>
<td>121.3</td>
<td>260.1</td>
<td>314.8</td>
<td>315.8</td>
</tr>
<tr>
<td>30</td>
<td>100</td>
<td>128.0</td>
<td>327.0</td>
<td>391.7</td>
<td>393.2</td>
</tr>
</tbody>
</table>

$\Rightarrow$ Tighter relaxation gives faster solves!
Mini-Quiz on Nonlinear Perspective of the Perspective

Perspective of our uncapacitated facility location problem:

\[
\begin{align*}
\text{minimize} & \quad z + y \\
\text{subject to} & \quad (x/z)^2 - (y/z) \leq 0 \quad 0 \leq x \leq z, \quad z \in \{0, 1\}, \quad y \geq 0
\end{align*}
\]

which clearly is not defined at \( z = 0 \).

Why not multiply through by \( z^2 \)?

\[
\begin{align*}
\text{minimize} & \quad z + y \\
\text{subject to} & \quad x^2 - zy \leq 0 \quad 0 \leq x \leq z, \quad z \in \{0, 1\}, \quad y \geq 0
\end{align*}
\]

... isn’t this a much nicer NLP?
Numerical Experience with the Bad the Perspective

**Bad perspective** of uncapacitated facility location problem:

\[
\text{minimize } \quad z + y \\
\text{subject to } \quad x^2 - zy \leq 0 \quad 0 \leq x \leq z, \ z \in \{0, 1\}, \ y \geq 0
\]

<table>
<thead>
<tr>
<th>Major</th>
<th>Minor</th>
<th>TrustRad</th>
<th>RegParam</th>
<th>StepNorm</th>
<th>Constrnts</th>
<th>Objective</th>
<th>Optimal</th>
<th>Phase</th>
<th>Step</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>0</td>
<td>0.5</td>
<td>1.01</td>
<td>0</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>0.625</td>
<td>0</td>
<td>0.385</td>
<td>0</td>
<td>2</td>
<td>SQP</td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>0.188</td>
<td>0</td>
<td>0.1875</td>
<td>0</td>
<td>2</td>
<td>SQP</td>
</tr>
<tr>
<td></td>
<td>...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>2.79e-09</td>
<td>0</td>
<td>2.794e-09</td>
<td>0</td>
<td>2</td>
<td>SQP</td>
</tr>
<tr>
<td>29</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>1.4e-09</td>
<td>0</td>
<td>1.397e-09</td>
<td>0</td>
<td>2</td>
<td>SQP</td>
</tr>
<tr>
<td>30</td>
<td>1</td>
<td>10</td>
<td>10</td>
<td>6.98e-10</td>
<td>0</td>
<td>6.985e-10</td>
<td>2</td>
<td>2</td>
<td>SQP</td>
</tr>
</tbody>
</table>

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:

\[
\begin{align*}
\text{minimize} & \quad z + y \\
\text{subject to} & \quad (x/z)^2 - (y/z) \leq 0 \quad 0 \leq x \leq z, \ z \in \{0, 1\}, \ y \geq 0 \\
\end{align*}
\]

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)


A polyhedral branch-and-cut approach to global optimization. 