Nonconvex Mixed-Integer Nonlinear Optimization

Summer School on Optimization of Dynamical Systems

Sven Leyffer and Jeff Linderoth

Argonne National Laboratory

September 3-7, 2018

Outline

Challenges of Nonconvex MINLP

Mixed-Integer Nonlinear Program (MINLP)

$$\underset{x}{\text{minimiße}} \; f(x) \quad \text{subject to } c(x) \le 0, \; x \in X, \; x_i \in \mathbb{Z} \; \forall \, i \in I$$

... now drop assumption that $f(x)$ and $c(x)$ are convex

Challenges of nonconvex MINLP

- Objective function $f(x)$ can have many local minimißers
- Continuous relaxation of constraint set $\{x | c(x) \le 0, \; x \in X\}$
  ... can be disjoint, may have no interior

Nonconvexity arise naturally, e.g. AC power-flow equations

$$F(U_k, U_l, \theta_k, \theta_l) := b_{kl} U_k U_l \sin(\theta_k - \theta_l) + g_{kl} U_k^2 \\ - g_{kl} U_k U_l \cos(\theta_k - \theta_l)$$

Challenges of Nonconvex MINLP

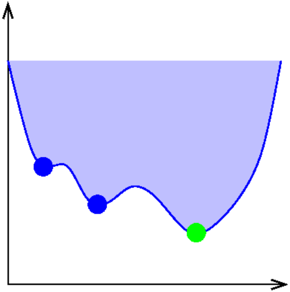
$$\underset{x}{\text{minimiße}} \; f(x) \quad \text{subject to } c(x) \leq 0,\; x \in X,\; x_i \in \mathbb{Z} \; \forall \, i \in I$$

Definition (Global Minimum)

A point x^* is a global minimum of

$$\underset{x}{\text{minimiße}} \; f(x) \quad \text{subject to } x \in \mathcal{F}$$

iff $f(x) \geq f(x^*)$ for all $x \in \mathcal{F}$



Remarks:

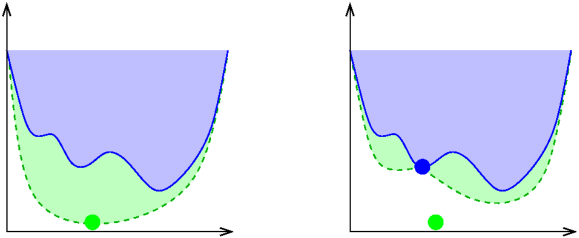
- NLP solvers are not guaranteed to find even local minima ... though they work remarkably well in practice!
- Finding a global min is difficult ... proving it is really hard

General Approach to Nonconvex MINLP

$$\underset{x}{\text{minimiße}} \ f(x) \quad \text{subject to } c(x) \leq 0,\ x \in X,\ x_i \in \mathbb{Z}\ \forall\ i \in I$$

Use old MIP trick: convex relaxation to get tractable NLP

- Relax integrality as before: $x_i \in \mathbb{R}\ \forall\ i \in I$
- Also need to relax $f(x)$ and constraints $c(x)$ ... new aspect



Need constraint enforcement to guarantee convergence

- Branching reduces area of relaxation
- Refinement tightens the relaxation over subdomain

Example: Relaxing Graph of Bilinear Product

Consider simple nonconvex (bounded) expression:

$$x_k = x_i x_j$$

where $l_i \le x_i \le u_i$ and $l_j \le x_j \le u_j$ are bounded

Derive McCormick outer approximations from mixing bounds:

1. $0 \le (x_i - l_i)(x_j - l_j)$ implies $x_k = x_i x_j \ge l_i x_j + l_j x_i - l_i l_j$

Example: Relaxing Graph of Bilinear Product

Consider simple nonconvex (bounded) expression:

$$x_k = x_i x_j$$

where $l_i \le x_i \le u_i$ and $l_j \le x_j \le u_j$ are bounded

Derive McCormick outer approximations from mixing bounds:

1. $0 \le (x_i - l_i)(x_j - l_j)$ implies $x_k = x_i x_j \ge l_i x_j + l_j x_i - l_i l_j$
2. $0 \le (x_i - u_i)(x_j - u_j)$ implies $x_k = x_i x_j \ge u_i x_j + u_j x_i - u_i u_j$
3. $0 \ge (x_i - l_i)(x_j - u_j)$ implies $x_k = x_i x_j \le l_i x_j + u_j x_i - l_i u_j$
4. $0 \ge (x_i - u_i)(x_j - l_j)$ implies $x_k = x_i x_j \le u_i x_j + l_j x_i - u_i l_j$

... tightest possible convex approximation of $x_k = x_i x_j$

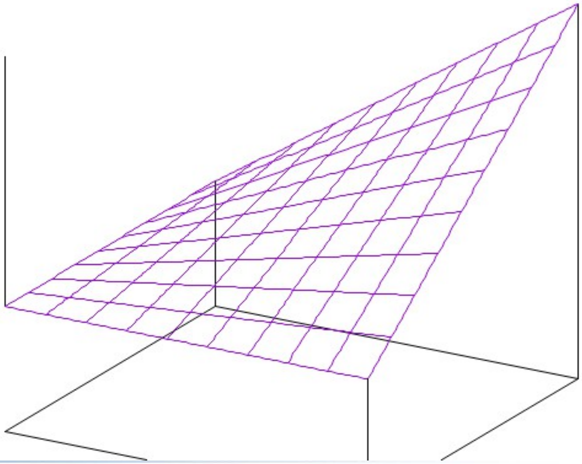
Special Case of Strong McCormick

$x_i, x_j \in \{0, 1\}$ then $x_k = x_i x_j \in \{0, 1\}$ if

$$x_k \in [0, 1] \quad \text{and} \quad x_k \ge x_j + x_i - 1, \; x_k \le x_i, \; x_k \le x_j$$

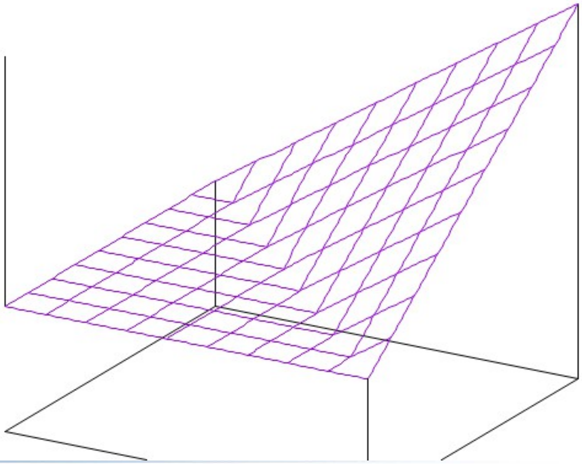
The set M: Worth 1000 Words?

$$X_{ij} = x_i x_j$$



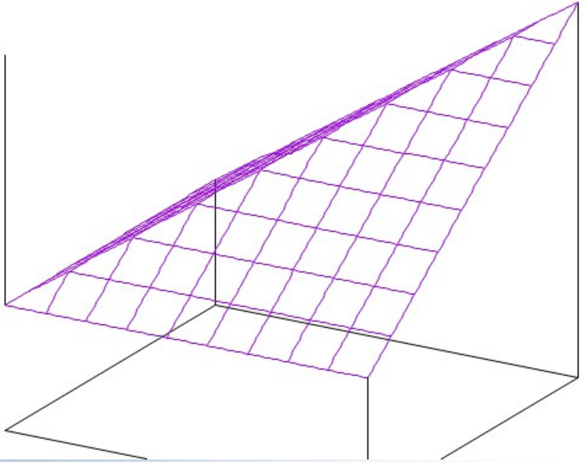
The set M: Worth 1000 Words?

$$X_{ij} \geq 0, \;\; X_{ij} \geq x_i + x_j - 1$$



The set M: Worth 1000 Words?

$$X_{ij} \leq x_i, X_{ij} \leq x_j$$



Outline

Generalization for Polynomial Functions

Reformulation-Linearization Technique (RLT) [Adams and Sherali, 1986]

- Generalizes McCormick underestimators to polynomials over $x \in \mathbb{R}^n$
  - Introduce multi-index $\boldsymbol{i}$ for $0 \leq i_j \leq q$

$$\boldsymbol{i} := (i_1, \ldots, i_n),$$

  - Define monomials $x^{\boldsymbol{i}} := x_1^{i_1} \cdot x_2^{i_2} \cdots x_n^{i_n}$ and constants $a_{\boldsymbol{i}} \in \mathbb{R}$

  Then, $p(x) = \sum_{\boldsymbol{i} \in I} a_{\boldsymbol{i}} x^{\boldsymbol{i}} = \sum_{\boldsymbol{i} \in I} a_{\boldsymbol{i}} x_1^{i_1} \cdots x_n^{i_n}$
- Define tensors (sparse) for higher-order terms (dropping zeroes)

$$x^{\boldsymbol{i}} \simeq X_{\boldsymbol{i}} = X_{i_1 \ldots i_n}$$

- Valid inequalities for higher-order by multiplying more bounds ... namely i_1 bounds on x_1 with ... i_n bounds on x_n

$\Rightarrow$ Valid linear underestimator for any polynomial

Example of RLT for General Polynomial Functions

Consider nonconvex polynomial optimization problem

$$\begin{aligned} \underset{x}{\text{minimiße}} \quad & f(x) := -x_1^2 x_2^2 + 2x_1 x_2^3 - x_2^4 \\ \text{subject to} \ & -1 \le x_i \le 1 \quad \text{for} \quad i = 1, 2 \end{aligned}$$

with three nonconvex terms ... use RLT ideas:

1. New linear variables: $X_{1122} \simeq x_1^2 x_2^2, X_{1222} \simeq x_1 x_2^3, X_{2222} \simeq x_2^4$
2. Multiply bounds to get polyhedral relaxation, e.g. for X_{1122}:

Valid inequality:

$$\begin{aligned} 0 &\le (1 - x_1)^2 (1 - x_2)^2 \\ 0 &\le 1 - 2x_1 - 2x_2 + 4X_{12} + X_{11} + X_{22} - 2X_{122} - 2X_{211} + X_{1122} \end{aligned}$$

$\Rightarrow$ need intermediate powers: $X_{12}, X_{11}, X_{22}, X_{122}, X_{211}$ etc

General Nonconvex Quadratic Functions

Strengthening RLT

1. Exploiting binary variables ... similar for integers
   - If $x_i \in \{0,1\}$ then $x_i^2 = x_i$ for all feasible points
   - Add linear constraints $X_{ii} = x_i$
2. Multiply linear constraints to improve RLT relaxation
   - Multiplying $x_i \geq 0$ and $b_t - \sum_{j=1}^n a_{tj} x_j \geq 0$ gives

$$b_t x_i - \sum_{j=1}^{n} a_{tj} x_i x_j \geq 0$$

   - Again linearize $X_{ij} = x_i x_j$ to get inequality

$$b_t x_i - \sum_{j=1}^{n} a_{tj} X_{ij} \geq 0$$

... generalizes to products between linear constraints

Snag: results in potentially huge LP relaxation!

Outline

Factorable Functions and MINLP

Consider MINLP with nonconvex, factorable $f(x)$ and $c(x)$

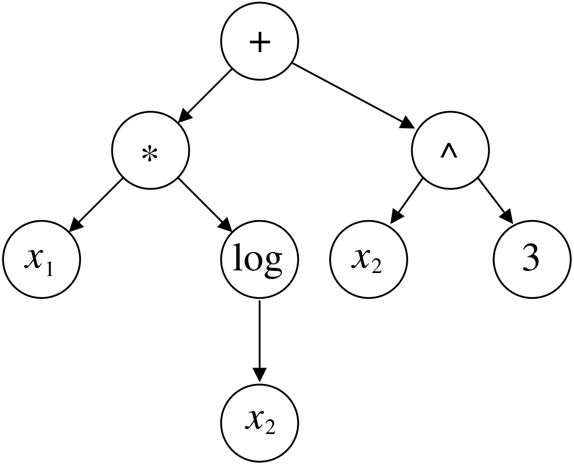
$$\underset{x}{\text{minimiße}} \;\; f(x) \quad \text{subject to } c(x) \leq 0,\; x \in X,\; x_i \in \mathbb{Z}\; \forall\, i \in I$$

Definition (Factorable Function)

$g(x)$ is factorable iff expressed as sum of products of unary functions of a finite set $\mathcal{O}_{\text{unary}} = \{\sin, \cos, \exp, \log, |\cdot|\}$ whose arguments are variables, constants, or other functions, which are factorable.

- Combination of functions from set of operators $\mathcal{O} = \{+, \times, /, \hat{}, \sin, \cos, \exp, \log, |\cdot|\}$.
- Excludes integrals $\int_{\xi=x_0}^{x} h(\xi) d\xi$ and black-box functions
- Represented as expression trees

Expression Tree Example



Expression tree of $f(x_1, x_2) = x_1 \log(x_2) + x_2^3$ ... also used in AD!

Relaxations of Factorable Functions

MINLP with nonconvex, factorable $f(x)$ and $c(x)$

$$\underset{x}{\text{minimiße}}\ f(x) \quad \text{subject to } c(x) \leq 0,\ x \in X,\ x_i \in \mathbb{Z}\ \forall\ i \in I$$

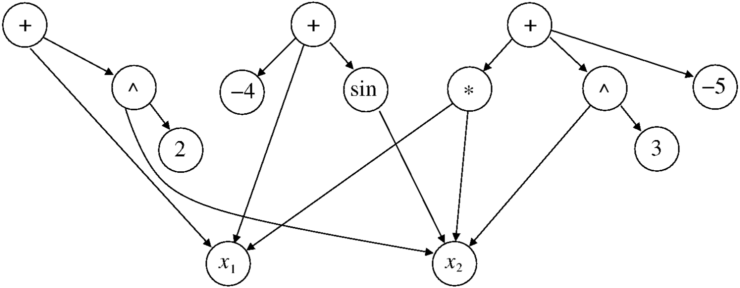
Combine expression trees of objective and constraints

- Root of each expression is $c_1(x), c_2(x), \ldots, c_m(x)$, or $f(x)$
- Associated bounds: $[-\infty, 0]$ for $c_i(x)$, and $[-\infty, \bar{\eta}]$ for $f(x)$
- Leaf nodes of all trees represent variables $x_1, x_2, \ldots, x_n$

$\Rightarrow$ gives directed acyclic graph (DAG)

Modeling languages (e.g. AMPL, GAMS) have DAG & "API"

Example of DAG of Optimization Problem



$$\begin{aligned} &\min\ x_1 + x_2^2 \\ &\text{s.t. }\ x_1 + \sin x_2 \leq 4, \quad x_1 x_2 + x_2^3 \leq 5 \\ &\qquad x_1 \in [-4, 4] \cap \mathbb{Z},\ x_2 \in [0, 10] \cap \mathbb{Z}. \end{aligned}$$

Three nodes without entering arcs for objective & constraints

Reformulation of Factorable MINLP

Reformulate factorable MINLP as

$$\begin{cases} \underset{x}{\text{minimiße}} & x_{n+q} \\ \text{subject to} & x_k = \vartheta_k(x) \qquad k = n+1, n+2, \ldots, n+q \\ & l_i \leq x_i \leq u_i \qquad i = 1, 2, \ldots, n+q \\ & x \in X, \\ & x_i \in \mathbb{Z}, \ \forall i \in I, \end{cases}$$

see e.g. [Smith and Pantelides, 1997]

- q new auxiliary variables, $x_{n+1}, \ldots, x_{n+q}$
- ϑ_k is operator from $\mathcal{O}\{+, \times, /, \hat{}, \sin, \cos, \exp, \log\}$
- Bounds on variables written explicitly

Example of Reformulation of Factorable MINLP

$$\begin{aligned} \min\ & x_1 + x_2^2 \\ \text{s.t.}\ & x_1 + \sin x_2 \le 4, \quad x_1 x_2 + x_2^3 \le 5 \\ & x_1 \in [-4,4] \cap \mathbb{Z},\ x_2 \in [0,10] \cap \mathbb{Z}. \end{aligned}$$

Reformulation with uni-/bi-variate nonconvex terms

$$\begin{array}{l|l|l|l|l} \min & x_9 & & & \\ \text{s.t.} & x_3 = \sin x_2 & x_7 = x_5 + x_6 - 5 & 0 \le x_2 \le 10 & 0 \le x_6 \le 1000 \\ & x_4 = x_1 + x_3 - 4 & x_8 = x_2^2 & -1 \le x_3 \le 1 & -45 \le x_7 \le 0 \\ & x_5 = x_1 x_2 & x_9 = x_1 + x_8 & -9 \le x_4 \le 0 & 0 \le x_8 \le 100 \\ & x_6 = x_2^3 & -4 \le x_1 \le 4 & -40 \le x_5 \le 40 & -4 \le x_9 \le 104 \\ & x_1, x_2, x_5, x_6, x_7, x_8, x_9 \in \mathbb{Z}. & & & \end{array}$$

- Integrality inherited from function
- Bounds inherited from function ... decomposition non-unique!

Reformulation of Factorable MINLP

Theorem (Equivalence of Factorable Formulation)

MINLP and factorable MINLP are equivalent, i.e. optimal solutions to one can be transformed into optimal solution of the other.

Factorable form makes it easier to get convex relaxation:

- Nonconvex sets, $k = n+1, n+2, \ldots, n+q$

  $$\Theta_k = \{x \in \mathbb{R}^{n+q} : x_k = \vartheta_k(x), x \in X, l \leq x \leq u, x_i \in \mathbb{Z}, i \in I\}$$

  ... nonconvex due to nonlinear equality
- Let $\breve{\Theta}_k \supset \Theta_k$ convex relaxation of $= \vartheta_k(x)$

$$\begin{cases} \underset{x}{\text{minimize}} & x_{n+q} \\ \text{subject to} & x_k \in \breve{\Theta}_k \quad k = n+1, n+2, \ldots, n+q \\ & l_i \leq x_i \leq u_i \; i = 1, 2, \ldots, n+q \\ & x \in X. \end{cases}$$

... convex relaxation ... only look at simple sets!

Reformulation of Factorable MINLP

General convex relaxation with polyhedral sets $\breve{\Theta}_k$:

$$\begin{cases} \underset{x}{\text{minimize}} & x_{n+q} \\ \text{subject to } x \in \breve{\Theta}_k & k = n+1, n+2, \ldots, n+q \\ \qquad l_i \le x_i \le u_i & i = 1, 2, \ldots, n+q \\ \qquad x \in X. \end{cases}$$

Polyhedral set $\breve{\Theta}_k$ defined by $a^k \in \mathbb{R}^{m_k}$, $B^k \in \mathbb{R}^{m_k \times (n+q)}$, and $d^k \in \mathbb{R}^{m_k}$:

$$\breve{\Theta}_k = \{x \in \mathbb{R}^{n+q} : a^k x_k + B^k x \ge d^k, x \in X, l \le x \le u\},$$

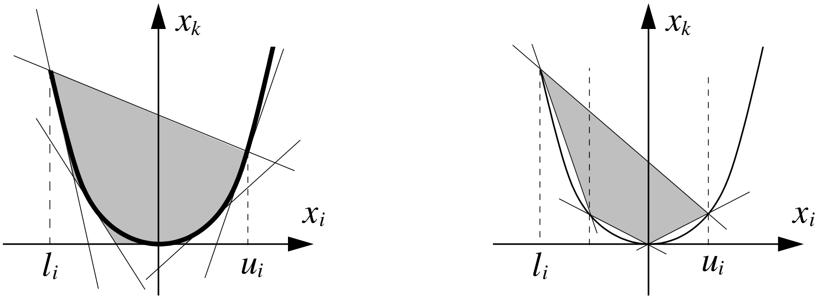
Gives lower bounding LP relaxation for MINLP solvers:

$$\begin{cases} \underset{x}{\text{minimize}} & x_{n+q} \\ \text{subject to } a^k x_k + B^k x \ge d^k & k = n+1, n+2, \ldots, n+q \\ \qquad l_i \le x_i \le u_i & i = 1, 2, \ldots, n+q \\ \qquad x \in X. \end{cases}$$

Now just need to construct polyhedral sets ... e.g. McCormick

Examples of Polyhedral Relaxations

Polyhedral relaxation, $\breve{\Theta}_k$, of $x_k = x_i^2$ with x_i continuous/integer



... if $x_i \in \mathbb{Z}$ then add inequalities violated at $x_i' \notin \mathbb{Z}$

Outline

Spatial Branch-and-Bound (BnB)

To separate solution of relaxation use spatial BnB

- Implicit enumeration technique like integer BnB
- Recursively define partitions of feasible set into two sets
- Use reformulation outlined above
- Solve LP relaxations ($\Rightarrow$ lower bounds)
  ... and nonconvex NLPs ($\Rightarrow$ upper bound if feasible)

Classic references & Solvers:

- [Tawarmalani and Sahinidis, 2002] BARON solver
- [Smith and Pantelides, 1997]
- [Belotti et al., 2009] Couenne solver ... open-source

Spatial Branch-and-Bound (BnB)

Key ingredients of spatial BnB

1. Procedure to compute lower bound for subproblem
2. Procedure for partitioning feasible set of subproblem: $\text{NLP}(l^-, u^-)$ and $\text{NLP}(l^+, u^+)$

... generates tree almost like integer BnB

NLP node is subproblem: $\text{NLP}(l, u)$

$$\begin{cases} \underset{x}{\text{minimize}} & f(x), \\ \text{subject to} & c(x) \leq 0, \\ & x \in X \\ & l_i \leq x_i \leq u_i \ \forall i = 1, 2, \ldots, n \\ & x_i \in \mathbb{Z}, \ \forall i \in I \end{cases}$$

... restriction of original MINLP

Spatial Branch-and-Bound (BnB)

Lower bounding problem at NLP(l, u), e.g. LP(l, u)

$$\begin{cases} \underset{x}{\text{minimize}} & x_{n+q} \\ \text{subject to} & a^k x_k + B^k x \geq d^k \quad k = n+1, n+2, \ldots, n+q \\ & l_i \leq x_i \leq u_i \qquad i = 1, 2, \ldots, n+q \\ & x \in X. \end{cases}$$

If LP(l, u) infeasible, then prune node.

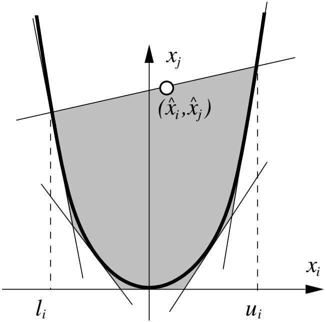
Otherwise, $\hat{x}$ optimal solution of LP(l, u):

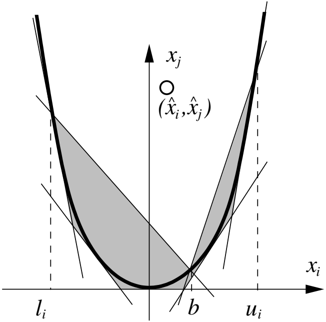
- If $\hat{x}$ feasible in NLP(l, u) (hence MINLP), then fathom node (new incumbent)
- If $\hat{x}$ not feasible in NLP(l, u) then ... branch ...
  1. $\hat{x}$ not integral, i.e., $\exists i \in I : \hat{x}_i \notin \mathbb{Z}$
  2. Nonconvex constraint is violated, i.e.

$$\exists k \in \{n+1, n+2, \ldots, n+q\} : \hat{x}_k \neq \vartheta_k(\hat{x}).$$

Illustration of Branching for Spatial Branch-and-Bound

Branching example $x_k = \vartheta_k(x_i) = (x_i)^2$ violated





Outline

Tighting Bounds and Relaxations

Bound tightening to reduce range of bounds $x_i \in [l_i, u_i]$

- Variable bounds are very important in Global Optimization
- Tighter Bounds $\Rightarrow$ tighter relaxations $\Rightarrow$ smaller trees

$$\begin{aligned} \text{minimiße}_x \quad & x_3 + x_1 x_5 + x_2 x_5 + x_3 x_5 \\ \text{subject to} \quad & x_5 - x_1 x_4 = 0 \\ & x_6 - x_2 x_3 = 0 \\ & x_1^2 + x_2^2 + x_3^2 + x_4^2 = 40 \\ & x_5 x_6 \geq 25 \\ & 1 \leq x_k \leq K \qquad k = 1, 2, 3, 4 \end{aligned}$$

Tightening Bounds and Relaxations

Bound tightening to reduce range of bounds $x_i \in [l_i, u_i]$

- Variable bounds are very important in Global Optimization
- Tighter Bounds $\Rightarrow$ tighter relaxations $\Rightarrow$ smaller trees

$$
\begin{aligned}
\text{minimiße}_x \quad & x_3 + x_1x_5 + x_2x_5 + x_3x_5 \\
\text{subject to} \quad & x_5 - x_1x_4 = 0 \\
& x_6 - x_2x_3 = 0 \\
& x_1^2 + x_2^2 + x_3^2 + x_4^2 = 40 \\
& x_5x_6 \geq 25 \\
& 1 \leq x_k \leq K \qquad k = 1,2,3,4
\end{aligned}
$$

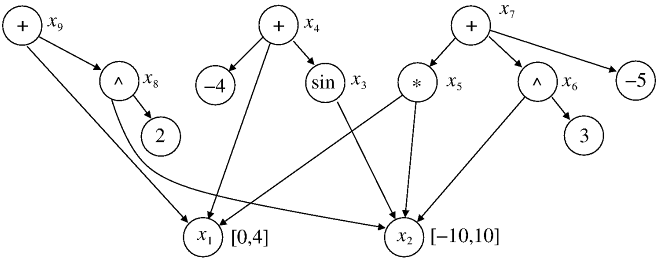
K	**Nodes**
5	210
10	788
25	6834
100	> 70000

Two relaxation-based bound-tightening techniques:

1. FBBT: feasibility-based bound tightening
2. OBBT: optimality-based bound tightening

FBBT: Feasibility-Based Bound Tightening

$$\min_x \; x_1 + x_2^2 \;\text{ s.t. }\; x_1 + \sin x_2 \le 4,\; x_1 x_2 + x_2^3 \le 5,\; x_1 \in [-4,4],\; x_2 \in [0,10]$$



Assume solution $\hat{x}$ found with $f(\hat{x}) = 10$:

1. $10 \ge x_9 := x_1 + x_8$ and $x_1 \ge -4$ imply $x_8 \le 14 < 100$ tighter
2. Propagate to $x_8 = x_2^2$ implies $-\sqrt{14} \le x_2 \le \sqrt{14}$ tightens x_2

Propagate bounds until improvement tails off ...

FBBT: Feasibility-Based Bound Tightening

Properties of FBBT

- Efficient and fast implementation for large-scale MINLP
- Can exhibit poor convergence, e.g. for $\alpha > 1$ consider: $\min x_1$ s.t. $x_1 = \alpha x_2,\ x_2 = \alpha x_1,\ x_1 \in [-1, 1]\}$
  - Solution is $(0, 0)$
  - FBBT does not terminate in finite number of steps
  - Sequence of tighter bounds for $l = 1, 2, \ldots$ with $\{[-\frac{1}{\alpha^l}, \frac{1}{\alpha^l}]\}_l \to (0, 0)$

... hence combine with other techniques

OBBT: Optimality-Based Bound Tightening

Solving min / max x_i s.t. $x \in \mathcal{F}$ (nonconvex MINLP) not practical
Instead, define (linear) relaxation

$$\mathcal{F}(l,u) = \left\{ x \in \mathbb{R}^{n+q} : \begin{array}{ll} a^k x_k + B^k x \geq d^k & k = n+1, n+2, \ldots, n+q \\ l_i \leq x_i \leq u_i & i = 1, 2, \ldots, n+q \\ x \in X & \end{array} \right\}.$$

Now get bounds on x_i for $i = 1, \ldots, n$ by solving $2n$ LPs:

$$\begin{aligned} l_i' &= \min\{x_i : x \in \mathcal{F}(l,u)\} \\ u_i' &= \max\{x_i : x \in \mathcal{F}(l,u)\} \end{aligned} \tag{1}$$

... only apply at root node, or small number of nodes

Teaching Points: Nonconvex MINLPs

Nonconvex MINLP

$$\underset{x}{\text{minimiße}} \; f(x) \quad \text{subject to } c(x) \le 0, \; x \in X, \; x_i \in \mathbb{Z} \; \forall \, i \in I$$

- Functions $f(x), c(x)$ no longer convex ... many applications
- Many local minimißers and disconnected feasible set

Combine Relaxation and Branching

- Exploit factorable description of $f(x), c(x)$
- Construct (linear) relaxations of elements,
  e.g. McCormick of $x_k = x_i x_j$ by multiplying bounds
- Apply spatial branch-and-bound to tighten relaxations
- Tight bounds are critical $\Rightarrow$ apply
  1. Feasibility-based bound tightening ... computational graph
  2. Optimality-based bound tightening ... solve LPs

Adams, W. and Sherali, H. (1986).
A tight linearization and an algorithm for zero-one quadratic programming problems.
Management Science, 32(10):1274–1290.

Belotti, P., Lee, J., Liberti, L., Margot, F., and Wächter, A. (2009).
Branching and bounds tightening techniques for non-convex MINLP.
Optimization Methods and Software, 24(4-5):597–634.

Smith, E. M. B. and Pantelides, C. C. (1997).
Global optimization of nonconvex MINLPs.
Computers & Chemical Engineering, 21:S791–S796.

Tawarmalani, M. and Sahinidis, N. V. (2002).
Convexification and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Algorithms, Software, and Applications.
Kluwer Academic Publishers, Boston MA.