Mixed-Integer Nonlinear Optimisation: Convex Underestimators

Benoit Chachuat
b.chachuat@imperial.ac.uk

Ruth Misener
r.misener@imperial.ac.uk

Computational Optimisation Group
Centre for Process Systems Engineering
Department of Computing

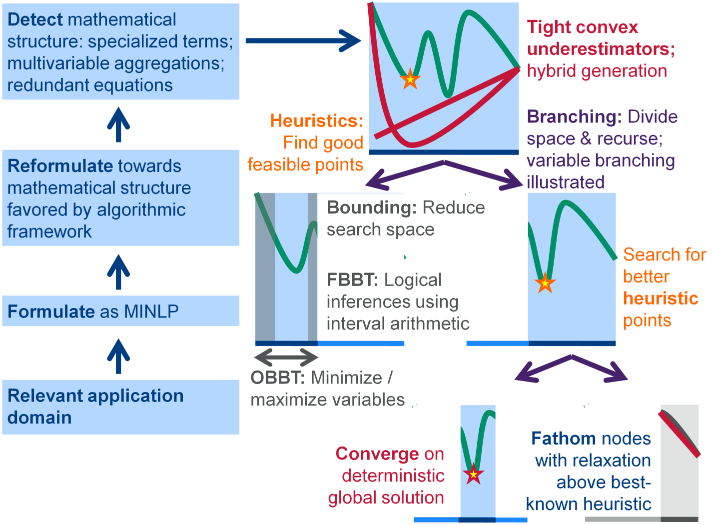
Imperial College London

computational optimisation group

Centre for Process Systems Engineering

09 May 2017

Last Time[1]



[1]Boukouvala, Misener, & Floudas, *Eur J Oper Res*, 2016

Convex Underestimators

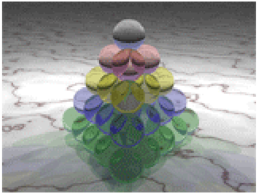
1. Interval Analysis
2. αBB Relaxations
3. Factorable Programming Relaxations
4. Trade-Offs with Decomposition Approaches

Introduction to Interval Analysis

- Methods for computing with intervals of real numbers, $[x^{\mathrm{L}}, x^{\mathrm{U}}] := \{x : x^{\mathrm{L}} \leq x \leq x^{\mathrm{U}}\}$
  - Combine set operations on intervals with interval function evaluations
  - New arithmetic needed!
- Enclosure the set of solutions to computational problems
  - E.g., range of functions; enclosure of integral value; enclosure of ODE solutions; enclosure of LE/NLE solutions; etc.
  - Applications in global optimization; asteroid orbits; beam physics; economics; etc.

Kepler's Conjecture: Proved by T. Hales & S. Ferguson using Interval Analysis in 1998![2]

No arrangement of equally sized spheres filling space has a greater average density than that of the cubic or hexagonal close packing ($\sim$ 74.048%)



[2]Hales, T.C. (2005), *Annals of Mathematics*, Second Series **162**(3): 1065-1185

Introduction to Interval Analysis [cont'd]

- Outward Rounding: Guarantee of rigorous enclosure despite round-off errors inherent to finite machine arithmetic

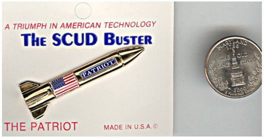
Workshop.

Consider sequence: $x_{n+1} = (x_n)^2$, with $x_0 = 1 - 10^{-20}$. What is $\lim_{n\to\infty} x_n$?

1. in infinite precision arithmetic?
2. in 16-place finite precision arithmetic?

Patriot Missile Failure[3]

Accumulation of round-off error led to a time delay of 0.34 seconds after 100 hours – enough time for the incoming scud to be outside the range gate and kill 28 soldiers...

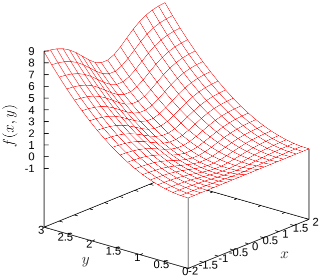


- But, inherently conservative! (over-approximation)
- Available Interval Packages: INTLAB (MATLAB), Profil/Bias (C++), FILIB++ (C++), etc.

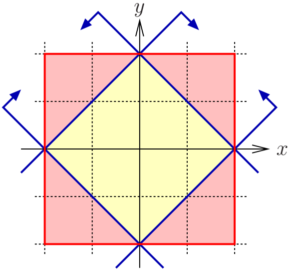
[3] http://www.ima.umn.edu/~arnold/455.f96/disasters.html

Interval Analysis: Two Examples

Workshop. Calculate the range of the function $f(x, y) = y^2 - y\exp(-x^2)$ for $-2 \leq x \leq 2$ and $0 \leq y \leq 3$



- $-4 \leq -x^2 \leq 0$, $e^{-4} \leq \exp(-x^2) \leq 1$, $0 \leq y\exp(-x^2) \leq 3$, $0 \leq y^2 \leq 9$, $-3 \leq y^2 - y\exp(-x^2) \leq 9$
- Actual range: $[-1/4, \sim 8.945]$
- Dependency problem of IA



Workshop. Find the tightest interval vector (interval hull) enclosing all points satisfying $-2 \leq x + y \leq 2$, and $-2 \leq x - y \leq 2$ (graphically)

- Interval hull: $-2 \leq x \leq 2$, $-2 \leq y \leq 2$
- Wrapping effect of IA

The Class of Factorable Functions

Factorable Function: Defined by a finite recursive composition of

1. binary sums
2. binary products
3. a given library of univariate functions

$$f(\mathbf{x}) = \left(\exp(x_1) - x_2^2\right) x_1 x_2 \quad \xrightarrow[\text{form}]{\text{factored}} \quad \begin{cases} v_1(\mathbf{x}) = x_1 \\ v_2(\mathbf{x}) = x_2 \\ v_3(\mathbf{x}) = \exp(v_1(\mathbf{x})) \\ v_4(\mathbf{x}) = -v_2(\mathbf{x})^2 \\ v_5(\mathbf{x}) = v_3(\mathbf{x}) + v_4(\mathbf{x}) \\ v_6(\mathbf{x}) = v_4(\mathbf{x}) v_1(\mathbf{x}) \\ f(\mathbf{x}) = v_6(\mathbf{x}) v_2(\mathbf{x}) \end{cases}$$

- Extremely inclusive class of functions – a.k.a. FC class[1]
- Nearly every function that can be represented finitely on a computer

[1] Moore, Kearfott, Cloud, *Introduction to Interval Analysis*, SIAM, 2009

Interval Analysis: Usual Binary Operations

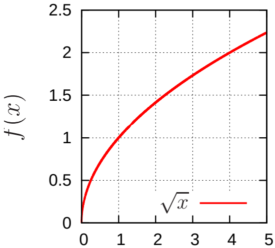
$$X \odot Y := \{x \odot y : x \in X, y \in Y\}, \quad \odot \in \{+, -, \times, \div\}$$

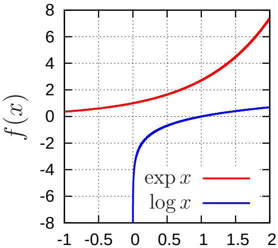
- Let $X := [x^{\mathrm{L}}, x^{\mathrm{U}}]$, $Y := [y^{\mathrm{L}}, y^{\mathrm{U}}]$
- Addition:
$$X + Y = [x^{\mathrm{L}} + y^{\mathrm{L}}, x^{\mathrm{U}} + y^{\mathrm{U}}]$$
- Subtraction:
$$X - Y = [x^{\mathrm{L}} - y^{\mathrm{U}}, x^{\mathrm{U}} - y^{\mathrm{L}}]$$
  - Do we have $X - X = 0$ in general?
- Multiplication:
$$X \times Y = [\min P, \max P], \quad P := \{x^{\mathrm{L}}y^{\mathrm{L}}, x^{\mathrm{L}}y^{\mathrm{U}}, x^{\mathrm{U}}y^{\mathrm{L}}, x^{\mathrm{U}}y^{\mathrm{U}}\}$$
  - Subdistributivity: $X(Y + Z) \subseteq XY + XZ$ (equality only if $YZ > 0$)
- Division:
$$X/Y = X \times (1/Y), \quad 1/Y = [1/y^{\mathrm{U}}, 1/y^{\mathrm{L}}],\ 0 \notin Y$$

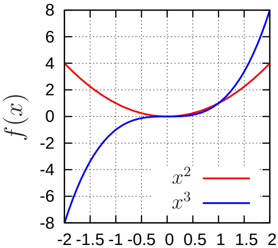
Interval Analysis: Usual Unary Operations

$$f(X) := \{f(x) : x \in X\}$$

- $x \mapsto \exp(x)$:
  $$\exp(X) = [\exp(x^{\mathrm{L}}), \exp(x^{\mathrm{U}})]$$
- $x \mapsto \log(x)$:
  $$\log(X) = [\log(x^{\mathrm{L}}), \log(x^{\mathrm{U}})],\ x^{\mathrm{L}} > 0$$
- $x \mapsto \sqrt{x}$:
  $$\sqrt{X} = [\sqrt{x^{\mathrm{L}}}, \sqrt{x^{\mathrm{U}}}],\ x^{\mathrm{L}} \geq 0$$
- $x \mapsto x^3$:
  $$X^3 = [(x^{\mathrm{L}})^3, (x^{\mathrm{U}})^3]$$
- $x \mapsto x^2$:
  Why not use $X^2 = X \times X$?
  $$X^2 = [\,\mathrm{mid}\{0, x^{\mathrm{L}}, x^{\mathrm{U}}\}, \max\{(x^{\mathrm{L}})^2, (x^{\mathrm{U}})^2\}]$$







Interval Analysis: General Expressions

Question. How to compute bounds for general, factorable functions?

- Computing the exact range $f(X_1, \ldots, X_n)$ boils down to solving a (nonconvex) optimization problem in general!
- Over-approximations need to be considered...

- Interval Extension of a Real-Valued Function f on $X_1 \times \cdots \times X_n$:
  Any interval-valued function F such that
  $F(x_1, \ldots, x_n) = f(x_1, \ldots, x_n), \quad \forall (x_1, \ldots, x_n) \in X_1 \times \ldots \times X_n$
  - e.g., $f(x) = 1 - x^2$ has interval extensions as $F(X) = 1 - X^2$ and $F(X) = 1 - X \times X$
- Inclusion Isotonicity of an Interval Extension F:
  $Y_i \subseteq X_i,\ i = 1, \ldots, n \quad \Rightarrow \quad F(Y_1, \ldots, Y_n) \subseteq F(X_1, \ldots, X_n)$
  - e.g., the previous unary and binary operations are inclusion isotonic

Fundamental Theorem. If F is an inclusion isotonic, interval extension of f on $X_1 \times \cdots \times X_n$, then $F(X_1, \ldots, X_n) \supseteq f(X_1, \ldots, X_n)$

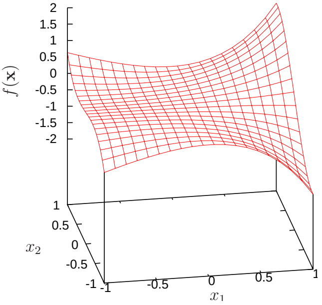
Interval Analysis: General Expressions [cont'd]

Question. How to construct inclusion isotonic, interval extensions of factorable functions?

Natural Interval Extensions:

1. Rewrite the function in factored form
2. Compute the range for each intermediate factor
   - Apply the inclusion isotonic, interval extension for the corresponding unary/binary operations

Workshop. Calculate the natural interval extension of $f(\mathbf{x}) = (\exp(x_1) - x_2^2)x_1x_2$ on $\mathbf{X} = [-1, 1]^2$

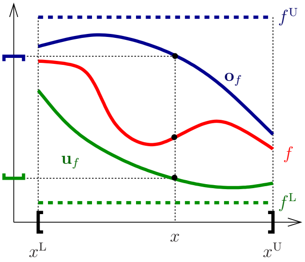


- $v_1(\mathbf{X}) = [-1, 1]$, $v_2(\mathbf{X}) = [-1, 1]$, $v_3(\mathbf{X}) = [\mathrm{e}^{-1}, \mathrm{e}]$, $v_4(\mathbf{X}) = [-1, 0]$, $v_5(\mathbf{X}) = [\mathrm{e}^{-1} - 1, \mathrm{e}]$, $v_6(\mathbf{X}) = [-\mathrm{e}, \mathrm{e}]$, $f(\mathbf{X}) = [-\mathrm{e}, \mathrm{e}]$
- Actual range: $[-\mathrm{e} + 1, \mathrm{e} - 1]$

αBB Convex & Concave Relaxations

Given a function f, construct functions $f^{\rm cv}$, $f^{\rm cc}$ such that:

- $f^{\rm cv}(\mathbf{x}) \leq f(\mathbf{x}) \leq f^{\rm cc}(\mathbf{x}), \forall \mathbf{x} \in \mathbf{X}$
- $f^{\rm cv} : \mathbf{X} \to \mathbb{R}$ convex
- $f^{\rm cc} : \mathbf{X} \to \mathbb{R}$ concave



Use of Convex/Concave Relaxations:

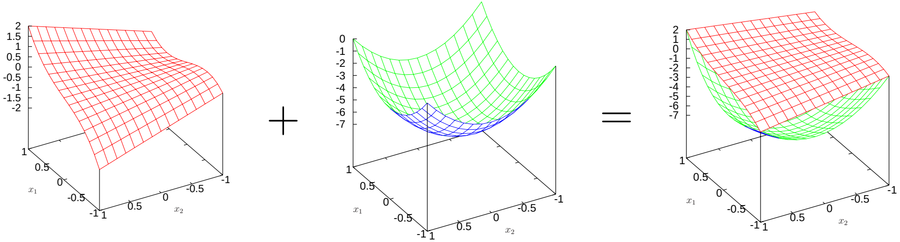
Nonconvex NLP	Relaxed Convex NLP
$\min_{\mathbf{x}} f(\mathbf{x})$	$\min_{\mathbf{x}} f^{\rm cv}(\mathbf{x})$
s.t. $\mathbf{g}_1(\mathbf{x}) \leq \mathbf{0},\ \mathbf{g}_2(\mathbf{x}) \geq \mathbf{0}$	s.t. $\mathbf{g}_1^{\rm cv}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{g}_2^{\rm cc}(\mathbf{x}) \geq \mathbf{0}$
$\mathbf{h}(\mathbf{x}) = \mathbf{0}$	$\mathbf{h}^{\rm cv}(\mathbf{x}) \leq \mathbf{0},\ \mathbf{h}^{\rm cc}(\mathbf{x}) \geq \mathbf{0}$

- No extra variable or extra constraint needed
- Faster B&B convergence with relaxations than interval bounds

αBB Relaxations: Principle [1/2]

Idea. Construct a convex underestimator f^{cv} of f on $\mathbf{X} := [\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}]$ as

$$f^{\mathrm{cv}}(\mathbf{x}) = f(\mathbf{x}) + \alpha \left[\mathbf{x} - \mathbf{x}^{\mathrm{L}}\right]^{\mathsf{T}} \left[\mathbf{x} - \mathbf{x}^{\mathrm{U}}\right], \quad \alpha\text{: shift parameter}$$



- Condition for f^{cv} to be an underestimator of f on $\mathbf{X}$? $\alpha \geq 0$
- Condition for f^{cv} to be convex on $\mathbf{X}$? $\mathbf{H}(\mathbf{x}) + 2\alpha\,\mathrm{I} \succeq 0, \ \forall \mathbf{x} \in \mathbf{X}$
- Non-uniform diagonal shift:

$$f^{\mathrm{cv}}(\mathbf{x}) = f(\mathbf{x}) + \sum_{i=1}^{n} \alpha_i \left[x_i - x_i^{\mathrm{L}}\right] \left[x_i - x_i^{\mathrm{U}}\right]$$

αBB Relaxations: Principle [2/2]

Challenge. How to automatically identify a valid shift parameter α?

$$\alpha \geq \max\left\{0; -\frac{1}{2} \underbrace{\min_{i=1,\ldots n} \left(\lambda_i[\mathbf{H}(\mathbf{x})] : \mathbf{x} \in [\mathbf{x}^{\mathrm{L}}, \mathbf{x}^{\mathrm{U}}]\right)}_{\lambda_{\min}[\mathbf{H}(\mathbf{X})]}\right\}$$

- Exact estimation of α intractable in general
- Variety of $O(n^2)$ and $O(n^3)$ over-approximation methods for factorable, twice continuously-differentiable functions
- $O(n^2)$ method based on Gershgorin's theorem:

$$\lambda_{\min}[\mathbf{H}(\mathbf{X})] \geq \min_{i=1,\ldots,n}\left(h^{\mathrm{L}}_{i,i} - \sum_{j\neq i} \max\{|h^{\mathrm{L}}_{i,j}|, |h^{\mathrm{U}}_{i,j}|\}\right)$$

- Tightness of αBB Relaxations (quadratic convergence):

$$\max_{\mathbf{x}^{\mathrm{L}}\leq\mathbf{x}\leq x^{\mathrm{U}}} \left(f(\mathbf{x}) - f^{\mathrm{cv}}(\mathbf{x})\right) = \frac{\alpha}{4}\left[\mathbf{x}^{\mathrm{U}} - \mathbf{x}^{\mathrm{L}}\right]^{\mathsf{T}}\left[\mathbf{x}^{\mathrm{U}} - \mathbf{x}^{\mathrm{L}}\right]$$

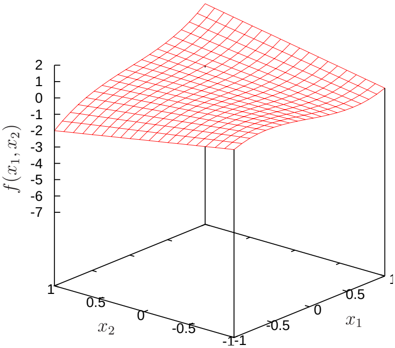
αBB Relaxations: Example

Workshop. Consider the function $f(\mathbf{x}) = x_1(x_2 + x_1^2)$, $\mathbf{x} \in \mathbf{X} := [-1, 1]^2$

1. Construct an αBB convex underestimator using Gershgorin's method
2. Construct the tightest possible αBB convex underestimator

Gershgorin's Method:

- $\mathbf{H}(\mathbf{X}) = \begin{pmatrix} [-6, 6] & 1 \\ 1 & 0 \end{pmatrix}$
- $\lambda_{\min}[\mathbf{H}(\mathbf{X})] \geq -7$
- $\alpha \geq \min\{0, \frac{7}{2}\} = \frac{7}{2}$



Tightest αBB Convex Underestimator:

- $\lambda_i[\mathbf{H}(\mathbf{x})] = 3x_1 \pm \sqrt{9x_1 + 1}$
- $\lambda_{\min}[\mathbf{H}(\mathbf{X})] = -3 - \sqrt{10}$
- $\alpha_{\min} = \frac{3+\sqrt{10}}{2} \approx 3.08$

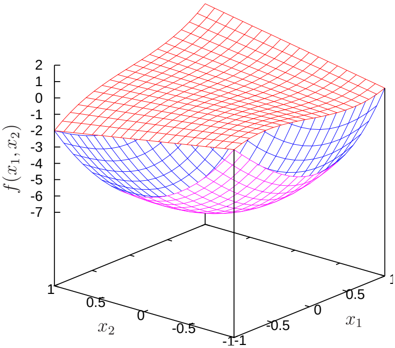
αBB Relaxations: Example

Workshop. Consider the function $f(\mathbf{x}) = x_1(x_2 + x_1^2)$, $\mathbf{x} \in \mathbf{X} := [-1, 1]^2$

1. Construct an αBB convex underestimator using Gershgorin's method
2. Construct the tightest possible αBB convex underestimator

Gershgorin's Method:

Tightest αBB Convex Underestimator:



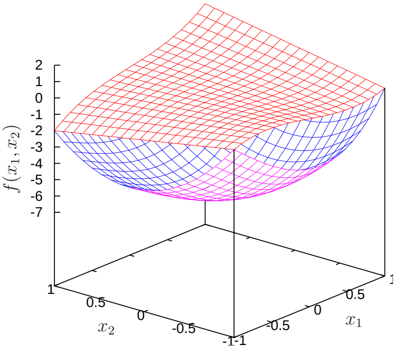
αBB Relaxations: Example

Workshop. Consider the function $f(\mathbf{x}) = x_1(x_2 + x_1^2)$, $\mathbf{x} \in \mathbf{X} := [-1, 1]^2$

1. Construct an αBB convex underestimator using Gershgorin's method
2. Construct the tightest possible αBB convex underestimator

Gershgorin's Method:

Tightest αBB Convex Underestimator:



Convex Underestimators

Factorable Programming: Principle

Factorable Function: Define a finite recursive composition of: binary sums; binary products; a given library of univariate functions.

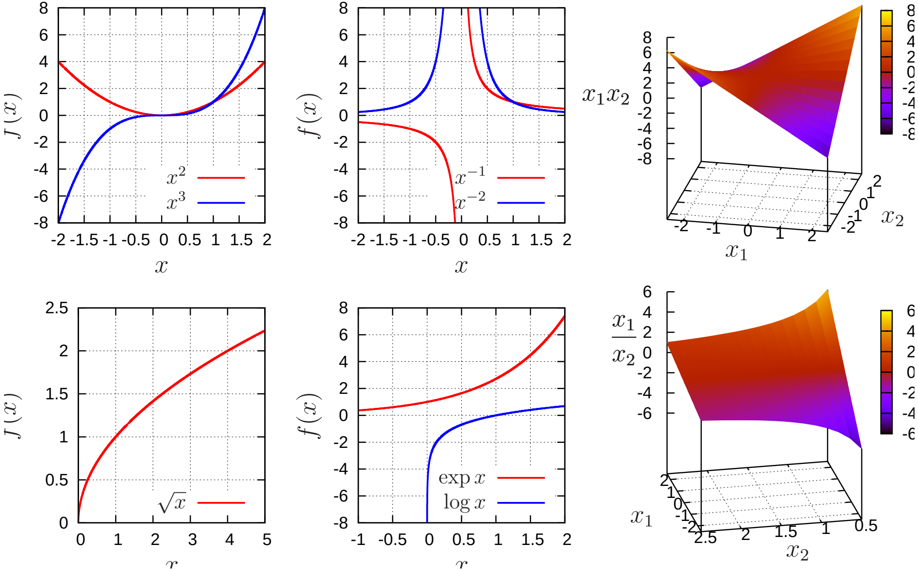
$$\mathrm{LMTD} \approx f(\mathbf{x}) = \left(\frac{x_1^2 x_2 + x_1 x_2^2}{2}\right)^{1/3} \xrightarrow[\text{form}]{\text{factored}} \begin{cases} v_1(\mathbf{x}) = x_1^2 \\ v_2(\mathbf{x}) = x_2 \; v_1(\mathbf{x}) \\ v_3(\mathbf{x}) = x_2^2 \\ v_4(\mathbf{x}) = x_1 \; v_3(\mathbf{x}) \\ v_5(\mathbf{x}) = v_2(\mathbf{x}) + v_4(\mathbf{x}) \\ f(\mathbf{x}) = (v_5(\mathbf{x})/2)^{1/3} \end{cases}$$

- Extremely inclusive class of functions – a.k.a. FC class[2]
- Nearly every function that can be represented finitely on a computer

Idea. Apply arithmetic rules for relaxing the: binary sums, binary products and library of univariate functions appearing in the factored form.

[2] Moore, Kearfott, Cloud, *Introduction to Interval Analysis*, SIAM, 2009

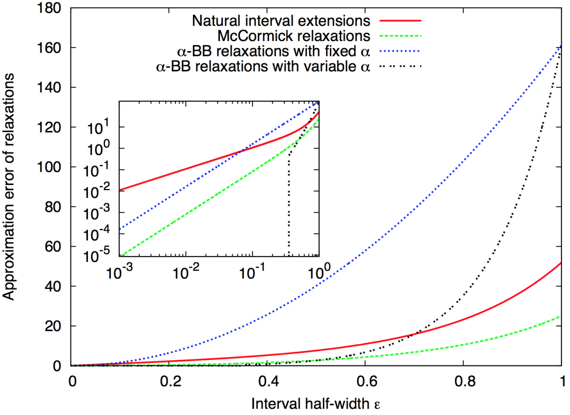
Common Functions in Modeling



Convex Underestimators

Trade-Offs Between Relaxation Schemes[4] [1/2]

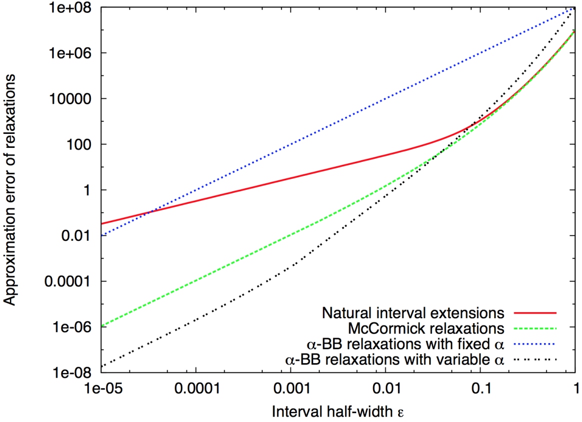
$$(z - z^2)(\log(z) + \exp(-z))$$



[4] Bompadre & Mitsos, *J Glob Optim*, 2012

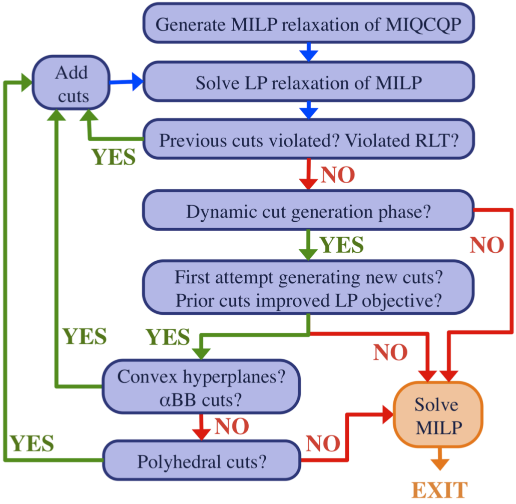
Trade-Offs Between Relaxation Schemes[5] [2/2]

$$\exp(1 - z^2)$$



[5] Bompadre & Mitsos, *J Glob Optim*, 2012

Idea: Let's Use Trade-Offs to our Advantage![6]



[6] Misener, Smadbeck, & Floudas, *Optim Met Softw*, 2015