Quadratic Underestimators of Differentiable McCormick Relaxations for Deterministic Global Optimization

Matthew Wilhelm, Matthew Stuber

2018 AIChE Annual Meeting
Pittsburgh, PA, October 30th

Process Systems and Operations Research Laboratory

UCONN SCHOOL OF ENGINEERING

Outline

Optimizing Simulations

Process Systems and
Operations Research
Laboratory

We commonly encounter problems that can be described by simulations. These simulations often haven a greatly reduced problem dimension compared to problems represented explicitly as closed-form equations since intermediate variables must be introduced in the latter approach, $n_p << n_x < n_y$. Examples:

- Regressions with embedded ODE (**Chemical Kinetics**) [1]
- Yield optimization of flowsheets (**Process Design**)

Full-Space

$$f^* = \min_{\mathbf{y} \in Y \subset \mathbb{R}^{n_y}} f(\mathbf{y})$$
$$\text{s.t. } \mathbf{h}(\mathbf{y}) = \mathbf{0}$$
$$\mathbf{g}(\mathbf{y}) \leq \mathbf{0}$$

Reduced-Space

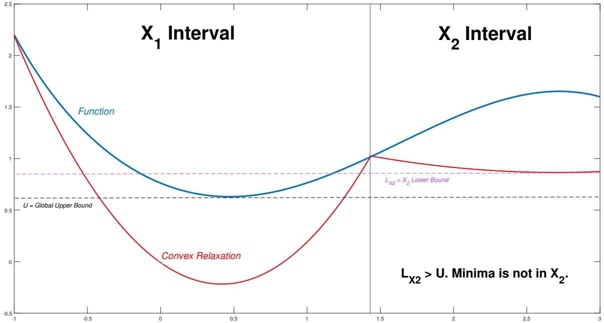
$$f^* = \min_{\mathbf{p} \in P} f(\mathbf{x}(\mathbf{p}), \mathbf{p})$$
$$\text{s.t. } \mathbf{g}(\mathbf{x}(\mathbf{p}), \mathbf{p}) \leq \mathbf{0}$$

1 Stuber, M. et al. Optimization Methods and Software, 2015, 30, 424-460

Dealing with Nonconvexity

Process Systems and
Operations Research
Laboratory

- Many simulations exhibit significant nonconvexity.
- NP-hard and solved via branch-and-bound variations [2].



2 Horst, R. & H. Tuy. Global Optimization: Deterministic Approaches, 2013

Relaxing Algorithms

- One approach to generating these lower bounds is via the use of set-valued arithmetics.
- Using these approaches an enclosure of the image of a function is defined along with operators that take these objects as inputs and output a new enclosure (method overloading).
- Approaches include are interval arithmetic [3], affine arithmetic [4], and McCormick operators [5].

3 Moore, R.E. Introduction to Interval Analysis, 2009

4 De Figueiredo, L.H. et al. Numerical Algorithms, 2004, 37, 147-158

5 Mitsos, A. et al. SIAM Journal of Optimization, 2009, 20, 573-601

McCormick Operators

McCormick Composition Rule [5]:

Let $Z \subset \mathbb{R}^n$, $X \subset \mathbb{R}$ be nonempty convex. The composite function $g = \phi \circ f$ s.t. $f : Z \rightarrow \mathbb{R}$ is continuous, $F : X \rightarrow \mathbb{R}$, $f(Z) \subset X$. Let $f^{cv}, f^{cc} : Z \rightarrow \mathbb{R}$ be relaxations of f on Z. Let $\phi^{cv}, \phi^{cc} : X \rightarrow \mathbb{R}$ be relaxations of ϕ on X. Let $\xi^*_{\text{min}}/\xi^*_{\text{max}}$ be a min/max of ϕ^{cv}/ϕ^{cc} on X.

$$g^{cv} : Z \rightarrow \mathbb{R} : z \mapsto \phi^{cv}(\text{mid}(f^{cv}, f^{cc}, \xi^*_{\text{min}}))$$
$$g^{cc} : Z \rightarrow \mathbb{R} : z \mapsto \phi^{cc}(\text{mid}(f^{cv}, f^{cc}, \xi^*_{\text{max}}))$$

- Usually second-order convergent and tighter than interval bounds [5,6].
- Desirable to minimize clustering about optima in branch and bound algorithm [7].
- Rules for propagating differentiable relaxations have been introduced [8].

5 Mitsos, A. et al. SIAM Journal of Optimization, 2009, 20, 573-601

6 Bompadre, A. et al. Journal of Global Optimization, 2012, 52, 1-28

7 Kannan, R. et al. Journal of Global Optimization, 2017, 69, 629-676

8 Khan, K. et al. Journal of Global Optimization, 2017, 67(4), 687-729

Lower Bounds from Subproblems

Process Systems and Operations Research Laboratory

No agreement exists in the literature on the best optimization problem to construct with these relaxations [8,9,10]. Affine relaxations may be weaker but the linear solvers are more robust and faster which may justify evaluating more nodes.

- **Standard McCormick Operators**
  - Nonsmooth NLP [1] ⇒ Nonsmooth NLP solver (e.g. Proximal Methods[9])
  - Relax Further [5] ⇒ Linear Program (e.g. CPLEX [10])
- **Differentiable NLP [5]**
  - Solve with Interior point method (e.g. Ipopt [11])
  - Further relax to QCQP ⇒ Interior point method

1 Stuber, M. et al. Optimization Methods and Software, 2015, 30, 424-460

5 Mitsos, A. et al. SIAM Journal of Optimization, 2009, 20, 573-601

8 Khan, K. et al. Journal of Global Optimization, 2017, 67(4), 687-729

9 L. Luksan et al. ACM Transactions on Mathematical Software 27 (2001), 193-213

10 IBM ILOG CPLEX Optimizer, 2017

11 Wächter, A. et al. Mathematical Programming, 2006, 106(1), 25-57

Outline

m-Convex Function [12]

Let $f : Z \subset \mathbb{R}^n \rightarrow \mathbb{R}$ be a proper, closed, m-convex, Whitney-1 differentiable, locally Lipschitz continuous function. At every point $x \in \text{int}(Z)$ there is a second-order quadratic expansion in the form

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) + \frac{m}{2}||\mathbf{y} - \mathbf{x}||_2^2 \quad (1)$$

- In many cases, m-convexity is required for superlinear convergence of optimization methods [12].

12 Vial, J.P. et al. Mathematics of Operations Research,8(2), 231-259

QCQP Relaxations

The quadratically-constrained quadratic programming (QCQP) relaxation of a nonlinear program is given below:

$$
\begin{aligned}
&\min_{\mathbf{y},\eta} \eta \\
\text{s.t.}\quad & f^{cv}(\mathbf{y}_0) + (\mathbf{y}-\mathbf{y}_0)^T \nabla f^{cv}(\mathbf{y}_0) + \frac{m_{f^{cv}}}{2}||\mathbf{y}-\mathbf{y_0}||_2^2 \leq \eta \\
& \mathbf{h}^{cc}(\mathbf{y}_0) + (\mathbf{y}-\mathbf{y}_0)^T \nabla \mathbf{h}^{cc}(\mathbf{y}_0) + \frac{m_{\mathbf{h}^{cc}}}{2}||\mathbf{y}-\mathbf{y_0}||_2^2 \geq \mathbf{0} \\
& \mathbf{h}^{cv}(\mathbf{y}_0) + (\mathbf{y}-\mathbf{y}_0)^T \nabla \mathbf{h}^{cv}(\mathbf{y}_0) + \frac{m_{\mathbf{h}^{cv}}}{2}||\mathbf{y}-\mathbf{y_0}||_2^2 \leq \mathbf{0} \\
& \mathbf{g}^{cv}(\mathbf{y}_0) + (\mathbf{y}-\mathbf{y}_0)^T \nabla \mathbf{g}^{cv}(\mathbf{y}_0) + \frac{m_{\mathbf{g}^{cv}}}{2}||\mathbf{y}-\mathbf{y_0}||_2^2 \leq \mathbf{0}
\end{aligned}
$$

Quadratic Bounds Algebra

Process Systems and Operations Research Laboratory

Addition of m-Convex Function [12]

Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a m-convex and $g : Z \subset \mathbb{R}^n \to \mathbb{R}$ be convex on Z then $f + g$ is p-convex on Z with $p \geq m$.

Linearity of m-Convex Function [12]

Let $f_1, f_2 : Z \subset \mathbb{R}^n \to \mathbb{R}$ be m_1-convex and m_2-convex, respectively. Let α_1, α_2 be positive real numbers then $\alpha_1 f_1 + \alpha_2 f_2$ is $(\alpha_1 m_1 + \alpha_2 m_2)$-convex.

Additive Inverse of m-Convex Function [12]

The function $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ is m-concave on Z if and only if $-f$ is m-convex on Z.

12 Vial, J.P. et al. Mathematics of Operations Research,8(2), 231-259

Propagating m-Convexity Bounds [12]

Process Systems and
Operations Research
Laboratory

Composition of m-Convex Function

Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a m-convex and $g : Z \subset R \to \mathbb{R}$ be a monotone convex increasing function on Z. Suppose g' is bounded below by β then $g \circ f$ is $m\beta$-convex.

Basic McCormick Scheme Fails

We know that $x \to x$ isn't m-convex. The composition rule fails to imply m-convexity.

Need to Track Linearity Properties to Start

For $z_j = f(z_i)$ such that z_i is affine, calculate m by rule for f then propagate m values using previously defined rules.

13 Vial, J.P. et al. Mathematics of Operations Research,8(2), 231-259

Generating m-values

Composition with Affine Functions

Let $f : Z \subset \mathbb{R}^n \to \mathbb{R}$ be a m-convex and $g : Z \subset \mathbb{R}^n \to \mathbb{R}$ is affine then $f \circ g$ is m-convex on Z.

- Define point, gradient, monotonicity flag, convexity flag, and interval bounds for variable.
- Define ruleset for computing m for each operator based on convexity flag and monotonicity.
- Propagate further bounds by composition rules.

13 Vial, J.P. et al. Mathematics of Operations Research,8(2), 231-259

Convergence Order: Spoilers!

Process Systems and Operations Research Laboratory

Theorem: Second-Order Pointwise Convergence

Consider a nonempty open set $Z \subset \mathbb{R}^n$, a nonempty compact set $Q \subset Z$, and a $C^{1,1}$ function $f : Z \to \mathbb{R}$. For each interval $\mathbf{w} \in \mathbb{IR}^n \cup \mathbb{Q} = \mathbb{IQ}$, a convex underestimator $f_w^C : \mathbf{w} \to R$ of f on w, suppose that there exists a scalar $\tau^C > 0$ for which

$$\sup_{z \in \mathbf{w}} \left(f(z) - f_{\mathbf{w}}^C(z) \right) \leq \tau^C wid(\mathbf{w})^2, \qquad \forall \mathbf{w} \in \mathbb{IQ}$$

Then, for each $\alpha \in [0, 1)$, there exists $\tau_\alpha > 0$ for which

$$\sup_{z \in \mathbf{w}} f(z) - (f_{\mathbf{w}}^C(\epsilon) + \langle \nabla f(z), z - \epsilon \rangle) + \langle A(z - \epsilon), z - \epsilon \rangle \leq \tau^C wid(\mathbf{w})^2, \quad \forall \mathbf{w} \in \mathbb{IQ}, \quad \forall \epsilon \in s_\alpha(\mathbf{w})) \tag{3}$$

That is to say, the quadratic underestimator inherits second-order point-wise convergence from the second-order point-wise convergence of the subdifferential.

Convergence Order of Subdifferential [13]

Consider a nonempty open set $Z \subset \mathbb{R}^n$, a nonempty compact set $Q \subset Z$, and a $C^{1,1}$ function $f : Z \to \mathbb{R}$. For each interval $\mathbf{w} \in \mathbb{IR}^n \cup \mathbb{Q} = \mathbb{IQ}$, a convex underestimator $f^C_w : \mathbf{w} \to R$ of f on w, suppose that there exists a scalar $\tau^C > 0$ for which

$$\sup_{z \in \mathbf{w}} \left(f(z) - f^C_{\mathbf{w}}(z) \right) \leq \tau^C wid(\mathbf{w})^2, \qquad \forall \mathbf{w} \in \mathbb{IQ}$$

Then, for each $\alpha \in [0, 1)$, there exists $\tau_\alpha > 0$ for which

$$\sup_{z \in \mathbf{w}} \left(f(z) - (f^C_{\mathbf{w}}(\epsilon) + \langle s, z - \epsilon \rangle) \right) \leq \tau^C wid(\mathbf{w})^2,$$

$$\forall \mathbf{w} \in \mathbb{IQ}, \qquad \forall \epsilon \in s_\alpha(\mathbf{w}), \qquad \forall s \in \partial f^C_{\mathbf{w}}(\epsilon)$$

Proof.

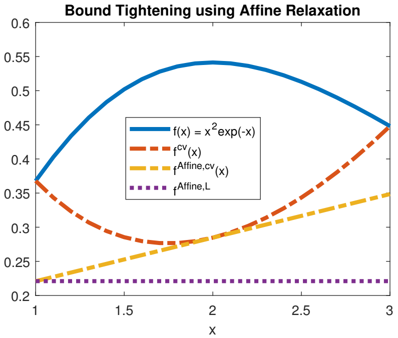
Note that $\nabla f(\mathbf{x}) \in \partial f^C_{\mathbf{w}}(\epsilon)$ and $\langle A(z - \epsilon), z - \epsilon \rangle \geq 0$ since A is positive semidefinite. Then $f(z) - (f^C_{\mathbf{w}}(\epsilon) + \langle s, z - \epsilon \rangle + \langle A(z - \epsilon), z - \epsilon \rangle) \leq f(z) - (f^C_{\mathbf{w}}(\epsilon) + \langle s, z - \epsilon \rangle)$ and the quadratic underestimator inherits second-order pointwise convergence. □

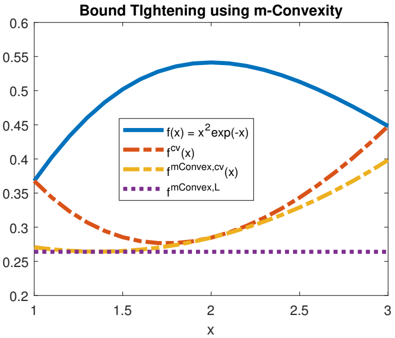
13 K. Khan, Subtangent-Based Approaches for Optimization of Parametric Process Systems, AIChE Annual Meeting, October 30, 2018

Tightening Interval Bounds

Process Systems and Operations Research Laboratory

- Subgradients may be used to contract interval bounds [14].
- We know closed form envelopes for univariate and bivariate quadratics [15,16].
- For univariate and bivariate functions these hulls can tighten interval bounds.





14 Najman, J et al. arXiv preprint arXiv:1710.09188, 2017

15 S. Vigerske, Ph.D. diss., Humboldt-UniversitÃďt zu Berlin, 2013

16 F. Domes and A. Neumaier, Constraints 15 (2010), pp. 404-429

Outline

Numerical Results

- We selected twelve problems from the GLOBAL library and literature examples and each subproblem relaxation was compared. The standard EAGO settings were used for all other parameters.
- An absolute tolerance of 10^{-4} was selected as the termination criteria. Ran single threaded on a 3.60GHz Intel Xeon E3-1270 v5 processor with 32GB in Ubuntu 16.04LTS and Julia v1.0. Ipopt v3.12 [11] was used to solve the NLP upper bound problem.
  - Linear lower-problem solved using CPLEX 12.8.0 [10].
  - Quadratic lower-problem solved using Ipopt v3.12.
  - Smooth NLP lower-problem solved using Ipopt v3.12.

10 IBM ILOG CPLEX Optimizer, 2017

11 Wächter, A. et al. Mathematical Programming, 2006, 106(1), 25-57

Numerical Results

Process Systems and Operations Research Laboratory

Problem	Variables	Inequalities	Equalities	CPU[s] (Affine)	CPU[s] (Affine + QBT)	CPU[s] (Quadratic)	CPU[s] (Convex NLP)
ex4_1_7	1	0	0	1.0	0.7	0.6	3.5
ex6_2_10	6	0	3	95.2	54.3	81.3	253.2
growthls	3	0	0	5.1	1.2	1.01	15.2
filter	2	0	1	0.6	0.5	2.9	3.1
hydro	30	0	25	0.9	0.4	3.2	6.4
hs62	3	0	1	4.5	4.1	4.7	16.1
st_ph1	6	5	0	0.1	0.1	1.2	2.3
tre	2	0	0	0.15	0.09	0.45	4.4
kinetic[5]	3	0	0	95.1%	96.1%	95.5%	89.2%
heat[5]	1	0	0	1.2	1.01	1.01	15.2
CS I [12]	2	0	9	0.7	0.6	1.6	8.6
CS II [12]	5	12	1	60.7	42.1	28.6	90.4
CS III [12]	8	1	22	71.8%	78.6%	81.3%	51.2%

5 Mitsos, A. et al. SIAM Journal of Optimization, 2009, 20, 573-601

12 Bongartz, D. et al. Journal of Global Optimization, 2017, 20, 761-796

Numerical Results - Trends

- For simulations with an extremely large number of intermediate terms, the m-convexity of the objective and constraints tends to vanish (kinetic/heat) models. M-convexity based bound tighten yields a small improvement in solution times in these cases.

- For smaller problems, with a significant number of quadratic constraints the NLP-subproblem form provides faster solution times.

- For mid-range problems, and simulations with constraints arising from simple intermediate terms the M-convexity problem formulation provides fast solution times.

Conclusions

- We can construct tighter than linear relaxations by propagating strong convexity information.
- Tighter than linear relaxations inherit second-order convergence properties from the McCormick relaxation.
- In general, relaxations that minimize the number of simulation evaluations tend to reduce computational burden for McCormick operator-based optimization.

Future Work

- **Evaluate full incorporation into global algorithms**
  - Develop the notion of numerically safe inequalities
  - Evaluate rules for selecting between nonlinear, quadratic and linear outer-estimators
- **Further theoretical developments**
  - Multiplication operator that propagates m-convexity.
  - Composition operator that propagates m-convexity generally.
  - Explore second-order nonsmooth methods for generating quadratic underestimators.

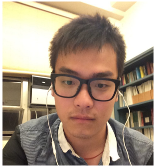
Acknowledgements

Process Systems and Operations Research Laboratory

UCONN SCHOOL OF ENGINEERING

- University of Connecticut for provide funding.
- Thanks to the PSOR lab for valuable discussions.

Prof. Matthew Stuber



Chenyu Wang

William Hale

Connor Dion

Jacob Chicano

Abiha Jafri

Questions?