The Effect of Hessian Evaluations in the Global Optimization αBB Method

Milan Hladík

Department of Applied Mathematics,
Faculty of Mathematics and Physics,
Charles University in Prague, Czech Republic
http://kam.mff.cuni.cz/~hladik/

6th International Conference on
High Performance Scientific Computing
Hanoi, Vietnam
March 16–20, 2015

Next Section

Interval computations

Principle of interval computations

Reliable results, covering outputs for all admissible input data.

Notation

An interval matrix

$$\mathbf{A} := [\underline{A}, \overline{A}] = \{A \in \mathbb{R}^{m\times n} \mid \underline{A} \le A \le \overline{A}\}.$$

The center and radius matrices

$$A_c := \frac{1}{2}(\overline{A} + \underline{A}), \quad A_\Delta := \frac{1}{2}(\overline{A} - \underline{A}).$$

The set of all $m \times n$ interval matrices: $\mathbb{IR}^{m\times n}$.

Main Problem

Let $f : \mathbb{R}^n \mapsto \mathbb{R}^m$ and $\mathbf{x} \in \mathbb{IR}^n$. Determine the image

$$f(\mathbf{x}) = \{f(x) : x \in \mathbf{x}\}.$$

Interval arithmetic

Interval Arithmetic

$$\mathbf{a}+\mathbf{b} = [\underline{a}+\underline{b}, \overline{a}+\overline{b}],$$
$$\mathbf{a}-\mathbf{b} = [\underline{a}-\overline{b}, \overline{a}-\underline{b}],$$
$$\mathbf{a}\cdot\mathbf{b} = [\min(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b}), \max(\underline{a}\underline{b}, \underline{a}\overline{b}, \overline{a}\underline{b}, \overline{a}\overline{b})],$$
$$\mathbf{a}/\mathbf{b} = [\min(\underline{a}/\underline{b}, \underline{a}/\overline{b}, \overline{a}/\underline{b}, \overline{a}/\overline{b}), \max(\underline{a}/\underline{b}, \underline{a}/\overline{b}, \overline{a}/\underline{b}, \overline{a}/\overline{b})], \quad 0 \notin \mathbf{b}.$$

Theorem (Basic properties of interval arithmetic)

- *Interval addition and multiplication is commutative and associative.*
- *It is not distributive in general, but sub-distributive instead,*

$$\forall \mathbf{a}, \mathbf{b}, \mathbf{c} \in \mathbb{IR} : \mathbf{a}(\mathbf{b}+\mathbf{c}) \subseteq \mathbf{ab}+\mathbf{ac}.$$

Example ($\mathbf{a} = [1,2]$, $\mathbf{b} = 1$, $\mathbf{c} = -1$)

$$\mathbf{a}(\mathbf{b}+\mathbf{c}) = [1,2]\cdot(1-1) = [1,2]\cdot 0 = 0,$$
$$\mathbf{ab}+\mathbf{ac} = [1,2]\cdot 1 + [1,2]\cdot(-1) = [1,2]-[1,2] = [-1,1].$$

Other functions

Basic functions

$\exp(\mathbf{x}) = [\exp(\underline{x}), \exp(\overline{x})]$, $\mathbf{x}^2 = \ldots$

General functions

Mean value / slope form: $f(\mathbf{x}) \subseteq f(a) + S(\mathbf{x}, a)(\mathbf{x} - a)$.

Other improvements

Monotonicity checking, interval refinements, ...

Next Section

Problem formulation

Convex underestimators

Let

1. $f : \mathbb{R}^n \mapsto \mathbb{R}$ be a twice-differentiable objective function,
2. $x_i \in \mathbf{x}_i = [\underline{x}_i, \overline{x}_i]$, $i = 1, \ldots, n$, interval domains for the variables.

Construct a function $g : \mathbb{R}^n \mapsto \mathbb{R}$ satisfying:

1. $f(x) \geq g(x)$ for every $x \in \mathbf{x}$,
2. $g(x)$ is convex on $x \in \mathbf{x}$.

Remark

- Deterministic global optimization methods based on branch & bound scheme.
- Rigorous enclosures of global minima and optimal value.
- One has has to bound the optimal value on $\mathbf{x}$ from above and below.

Forms of underestimators

Forms of underestimators

1. αBB method (Floudas et al., 1995–2013)

$$g(x) := f(x) - \sum_{i=1}^{n} \alpha_i(\overline{x}_i - x_i)(x_i - \underline{x}_i), \qquad (*)$$

2. non-diagonal αBB method (Akrotirianakis et al., 2004, Skjäl et al., 2012)

$$g(x) := f(x) - (\overline{x} - x)^T P(x - \underline{x}) + q(x),$$

where $q(x)$ is a piecewise linear convex function.

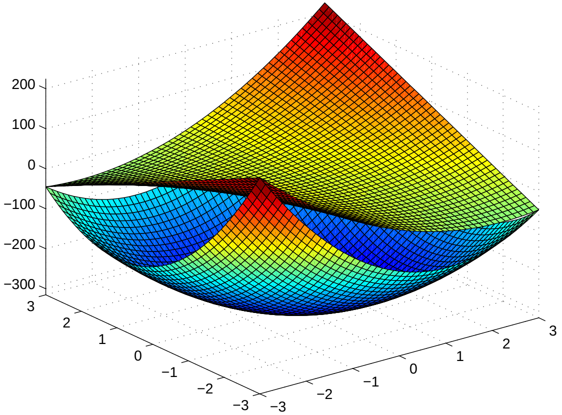
3. γBB method (Akrotirianakis and Floudas, 2004)

$$g(x) := f(x) - \sum_{i=1}^{n}(1 - e^{\gamma_i(\overline{x}_i - x_i)})(1 - e^{\gamma_i(x_i - \underline{x}_i)}) \qquad (\dagger)$$

Theorem (Floudas and Kreinovich, 2007)

Forms $(*)$ and $(\dagger)$ are the only shift-invariant forms of the gen. scheme

$$g(x) := f(x) - \sum_{i=1}^{n} \alpha_i(g_i(\overline{x}_i) - g_i(x_i))(g_i(x_i) - g_i(\underline{x}_i)).$$



Function $f(x)$ and its convex underestimator $g(x)$.

Computation of α

Idea

Choose α large enough to ensure positive semidefinitness of the Hessian of

$$g(x) := f(x) - \sum_{i=1}^{n} \alpha_i(\overline{x}_i - x_i)(x_i - \underline{x}_i),$$

The Hessian of $g(x)$ reads

$$\nabla^2 g(x) = \nabla^2 f(x) + 2\,\mathrm{diag}(\alpha).$$

Interval Hessian matrix

Let $\mathbf{H}$ be an interval matrix enclosing the image of $\nabla^2 f(x)$ over $x \in \mathbf{x}$:

$$\frac{\partial^2}{\partial x_i \partial x_j} f(x) \in \mathbf{h}_{ij} = [\underline{h}_{ij}, \overline{h}_{ij}], \quad \forall x \in \mathbf{x}.$$

Remarks

- Checking strong positive semidefiniteness of $\mathbf{H}$ is co-NP-hard.
- Various enclosures for eigenvalues of $H \in \mathbf{H}$.
- Scaled Gerschgorin method enables to express α_i-s.

Computation of α

Theorem (Scaled Gerschgorin inclusion)

Let $d \in \mathbb{R}^n$, $d > 0$, and $\mathbf{H} \in \mathbb{IR}^{n\times n}$. Then any real eigenvalue of any $H \in \mathbf{H}$ lies in one of the intervals

$$[\underline{h}_{ii} - \textstyle\sum_{j\neq i} |\mathbf{h}_{ij}| d_j/d_i,\ \overline{h}_{ii} + \textstyle\sum_{j\neq i} |\mathbf{h}_{ij}| d_j/d_i], \quad i = 1, \dots, n,$$

where $|\mathbf{h}_{ij}| = \max\left\{|\underline{h}_{ij}|, |\overline{h}_{ij}|\right\}$.

Scaled Gerschgorin method for α

$$\alpha_i := \max\left\{0, -\tfrac{1}{2}\left(\underline{h}_{ii} - \textstyle\sum_{j\neq i} |\mathbf{h}_{ij}| d_j/d_i\right)\right\}, \quad i = 1, \dots, n,$$

- To reflect the range of the variable domains, use $d := \overline{x} - \underline{x}$.

Theorem (H., 2014)

The choice $d := \overline{x} - \underline{x}$ is optimal (i.e., it minimizes the maximum separation distance between $f(x)$ and $g(x)$) if

$$\underline{h}_{ii} d_i - \textstyle\sum_{j\neq i} |\mathbf{h}_{ij}| d_j \le 0, \quad \forall i = 1, \dots, n.$$

Symbolic computation of the Hessian

Direct computation of α

Define

$$h_i(x) := \frac{\partial^2}{\partial x_i^2} f(x) - \sum_{j \neq i} \left| \frac{\partial^2}{\partial x_i \partial x_j} f(x) \right| d_j / d_i, \quad i = 1, \ldots, n.$$

The entries of α then follows

$$\alpha_i := \max \left\{ 0, -\frac{1}{2} \underline{h_i(\mathbf{x})} \right\}, \quad i = 1, \ldots, n.$$

Comments

- BEFORE: compute $\mathbf{H}$, and after α,
  NOW: compute α directly.
- If $0 \notin \frac{\partial^2}{\partial x_i \partial x_j} f(\mathbf{x})$, remove the absolute value.
- Apply symbolic simplifications to the expression for $h_i(x)$.

Next Section

Example I.

Example (Gounaris and Floudas, 2008)

Let

$$f(x) = (x_1 + 10x_2)^2 + 5(x_3 - x_4)^2 + (x_2 - 2x_3)^4 + 10(x_1 - x_4)^4,$$

where $x \in \mathbf{x} = [0, 1]^4$. It is known that the global minimum is $f^* = 0$.

First, we compute the interval Hessian

$$\nabla^2 f(\mathbf{x}) \subseteq \mathbf{H} = \begin{pmatrix} [-118, 122] & [20, 20] & [0, 0] & [-120, 120] \\ [20, 20] & [176, 248] & [-96, 48] & [0, 0] \\ [0, 0] & [-96, 48] & [-86, 202] & [-10, -10] \\ [-120, 120] & [0, 0] & [-10, -10] & [-110, 130] \end{pmatrix}.$$

and calculate

$$\alpha = (129, 0, 96, 120), \quad f^* \geq -85.1312.$$

Example I.

Example (cont'd)

Let us compute the Hessian matrix symbolically

$$\nabla^2 f(x) = \begin{pmatrix} 2+120(x_1-x_4)^2 & 20 & 0 & -120(x_1-x_4)^2 \\ 20 & 200+12(x_2-2x_3)^2 & -24(x_2-2x_3)^2 & 0 \\ 0 & -24(x_2-2x_3)^2 & 10+48(x_2-2x_3)^2 & -10 \\ -120(x_1-x_4)^2 & 0 & -10 & 10+120(x_1-x_4)^2 \end{pmatrix}.$$

Since all off-diagonal entries are sign stable, we calculate

$$\alpha = (69, 0, 48, 60), \quad f^* \geq -43.2171.$$

Symbolically simplifying

$$h_1(x) = 2 + 120(x_1 - x_4)^2 - 20 - 0 - 120(x_1 - x_4)^2$$

to $h_1(x) = -18$, we get

$$\alpha = (18, 0, 0, 0), \quad f^* \geq -1.9768.$$

Example II.

Example (Adjiman et al., 1998)

Let

$$f(x_1, x_2) = \cos(x_1)\sin(x_2) - \frac{x_1}{x_2^2 + 1},$$

where $x_1 \in [-1, 2]$ and $x_2 \in [-1, 1]$. The optimal value is known to be $f^* = -2.02181$.

By the classical αBB method, we compute

$$\nabla^2 f(\mathbf{x}) \subseteq \mathbf{H} = \begin{pmatrix} [-0.8415, 0.8415] & [-5.0000, 4.8415] \\ [-5.0000, 4.8415] & [-18.8415, 20.8415] \end{pmatrix},$$

and get

$$\alpha = (2.0874, 13.1707), \quad f^* \geq -18.4970.$$

Example II.

Example (cont'd)

Using the symbolical approach,

$$\nabla^2 f(x) = \begin{pmatrix} -\cos(x_1)\sin(x_2) & -\sin(x_1)\cos(x_2) + \dfrac{2x_2}{(x_2^2+1)^2} \\ -\sin(x_1)\cos(x_2) + \dfrac{2x_2}{(x_2^2+1)^2} & -\cos(x_1)\sin(x_2) + \dfrac{2x_1(x_2^2+1)^2 - 8x_1x_2^2(x_2^2+1)}{(x_2^2+1)^4} \end{pmatrix}.$$

If

$$h_2(x) = -\cos(x_1)\sin(x_2) + \frac{2x_1(x_2^2+1)^2 - 8x_1x_2^2(x_2^2+1)}{(x_2^2+1)^4} - \frac{2}{3}\left| -\sin(x_1)\cos(x_2) + \frac{2x_2}{(x_2^2+1)^2} \right|$$

is simplified to

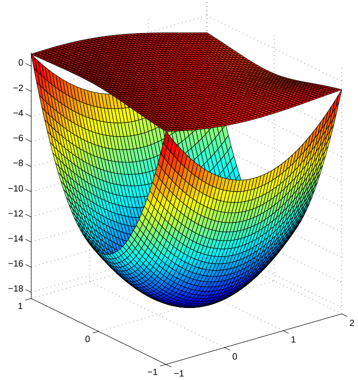
$$h_2(x) = -\cos(x_1)\sin(x_2) + \frac{2x_1(2-6x_2^2)}{(x_2^2+1)^3} - \frac{2}{3}\left| -\sin(x_1)\cos(x_2) + \frac{2x_2}{(x_2^2+1)^2} \right|.$$

we calculate

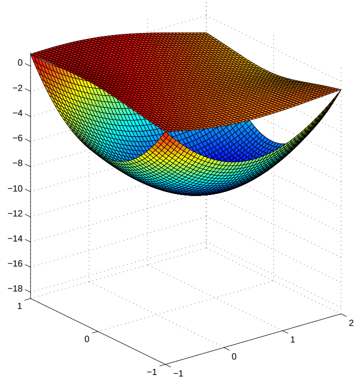
$$\alpha = (1.4208, 5.4208), \quad f^* \geq -9.3110.$$

Example II.

Example (cont'd)



Traditional approach.



Symbolic approach.

Example III.

Example (Skjäl et al., 2012)

Let

$$f(x_1, x_2) = (1 + x_1 - e^{x_2})^2,$$

where $x_1 \in [0, 1]$ and $x_2 \in [0, 2]$. The optimal value is $f^* = 0$.

By the classical αBB method, we compute

$$f^* \geq -14.46.$$

In the symbolic approach, by factoring $h_2(x)$ to

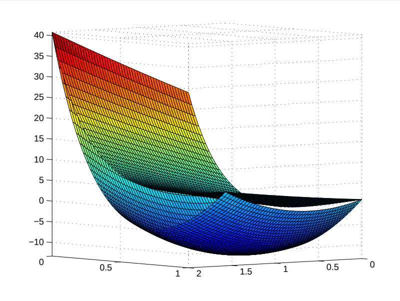
$$h_2(x) = (-3 - 2x_1 + 4e^{x_2})e^{x_2}.$$

we get

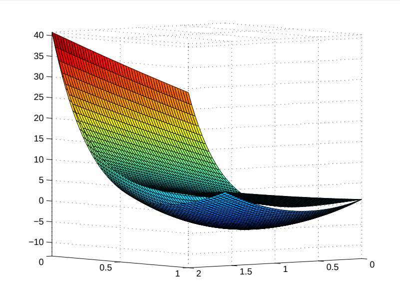
$$f^* \geq -6.5629.$$

Example III.

Example (cont'd)



Traditional approach.



Symbolic approach.

Further improvements

Recall that

$$\alpha_i := \max\left\{0, -\tfrac{1}{2}\,\underline{h_i(\mathbf{x})}\right\}, \quad i = 1, \ldots, n,$$

where

$$h_i(x) := \tfrac{\partial^2}{\partial x_i^2} f(x) - \textstyle\sum_{j \neq i} \left| \tfrac{\partial^2}{\partial x_i \partial x_j} f(x) \right| d_j / d_i, \quad i = 1, \ldots, n.$$

Idea: Linearize the absolute values from above.

Proposition

For every $y \in \mathbf{y} \subset \mathbb{R}$ with $\underline{y} < \overline{y}$ one has

$$|y| \leq \alpha y + \beta, \tag{$*$}$$

where

$$\alpha = \frac{|\overline{y}| - |\underline{y}|}{\overline{y} - \underline{y}} \quad \text{and} \quad \beta = \frac{\overline{y}|\underline{y}| - \underline{y}|\overline{y}|}{\overline{y} - \underline{y}}.$$

Moreover, if $\underline{y} \geq 0$ or $\overline{y} \leq 0$ then $(*)$ holds as equation.

Conclusion and future work

The main message

- Symbolic evaluation may much more efficient than the automatic one.
- Promising field of research.

Challenge

- Symbolic simplifications of expressions.

Future work

- Handling the absolute value: linearization and numerical testing.

References

I. G. Akrotirianakis and C. A. Floudas.
Computational experience with a new class of convex underestimators: Box-constrained NLP problems.
J. Glob. Optim., 29(3):249–264, 2004.

C. A. Floudas.
Deterministic global optimization. Theory, methods and applications .
Kluwer, Dordrecht, 2000.

C. A. Floudas and C. E. Gounaris.
A review of recent advances in global optimization.
J. Glob. Optim., 45(1):3–38, 2009.

M. Hladík.
On the efficient Gerschgorin inclusion usage in the global optimization αBB method.
J. Glob. Optim., 61(2):235–253, 2015.