OSE SEMINAR 2011

A generalization of classical αBB underestimation to include bilinear terms

Anders Skjäl

CENTER OF EXCELLENCE IN
OPTIMIZATION AND SYSTEMS ENGINEERING
ÅBO AKADEMI UNIVERSITY

ÅBO, DECEMBER 8 2011

OSE OPTIMIZATION AND SYSTEMS ENGINEERING

Åbo Akademi University

The Big Picture

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) & \\ \text{s.t.} \quad & f_m(\boldsymbol{x}) \leq 0, & m \in \{1, 2, \ldots, M\} \\ & x_i^L \leq x_i \leq x_i^U, & i \in \{1, 2, \ldots, n\} \end{aligned}$$

The Big Picture

$$\begin{aligned} \min \quad & f_0(\boldsymbol{x}) & \\ \text{s.t.} \quad & f_m(\boldsymbol{x}) \leq 0, & m \in \{1, 2, \ldots, M\} \\ & x_i^L \leq x_i \leq x_i^U, & i \in \{1, 2, \ldots, n\} \end{aligned}$$

- The variables can be real and/or discrete
- To get a lower bound we replace the functions f_i with convex underestimators and solve the resulting convex problem.
- αBB is a well-known convexification method and this work generalizes that method

The Big Picture

$$\begin{array}{lll} \min & f_0(\boldsymbol{x}) & \\ \text{s.t.} & f_m(\boldsymbol{x}) \leq 0, & m \in \{1, 2, \ldots, M\} \\ & x_i^L \leq x_i \leq x_i^U, & i \in \{1, 2, \ldots, n\} \end{array}$$

- The variables can be real and/or discrete
- To get a lower bound we replace the functions f_i with convex underestimators and solve the resulting convex problem.
- αBB is a well-known convexification method and this work generalizes that method
- A joint work with Ruth Misener (PrincetonU), Prof. Christodoulos A. Floudas (PrincetonU), and Prof. Tapio westerlund (ÅAU)

Gerschgorin's Circle Theorem

Theorem

Let $A \in \mathbb{C}^{n \times n}$ with entries a_{ij} and define $R_i = \sum_{j \neq i} |a_{ij}|$. Every eigenvalue of A lies within at least one of the Gerschgorin disks

$$D(a_{ii}, R_i) = \{x : \ |x - a_{ii}| \leq R_i\}.$$

Gerschgorin's Circle Theorem

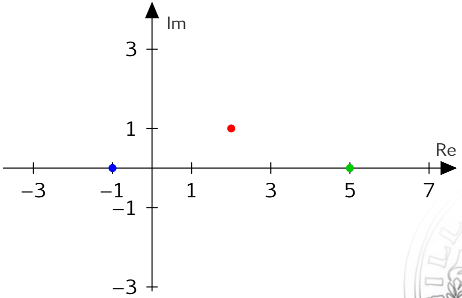
Theorem

Let $A \in \mathbb{C}^{n \times n}$ with entries a_{ij} and define $R_i = \sum_{j \neq i} |a_{ij}|$. Every eigenvalue of A lies within at least one of the Gerschgorin disks

$$D(a_{ii}, R_i) = \{x : |x - a_{ii}| \leq R_i\}.$$

Example

$$A = \begin{bmatrix} 2+i & 2 & -1 \\ 1 & 5 & i \\ 1 & -1 & -1 \end{bmatrix}$$



Gerschgorin's Circle Theorem

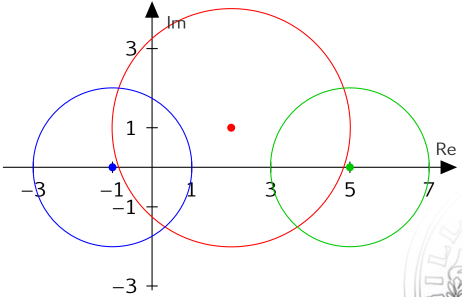
Theorem

Let $A \in \mathbb{C}^{n\times n}$ with entries a_{ij} and define $R_i = \sum_{j\neq i} |a_{ij}|$. Every eigenvalue of A lies within at least one of the Gerschgorin disks

$$D(a_{ii}, R_i) = \{x : |x - a_{ii}| \leq R_i\}.$$

Example

$$A = \begin{bmatrix} 2+i & 2 & -1 \\ 1 & 5 & i \\ 1 & -1 & -1 \end{bmatrix}$$



Gerschgorin's Circle Theorem

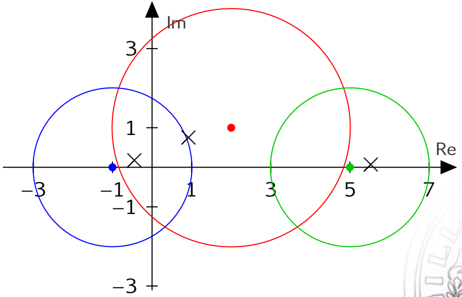
Theorem

Let $A \in \mathbb{C}^{n\times n}$ with entries a_{ij} and define $R_i = \sum_{j\neq i} |a_{ij}|$. Every eigenvalue of A lies within at least one of the Gerschgorin disks

$$D(a_{ii}, R_i) = \{x : \; |x - a_{ii}| \leq R_i\}.$$

Example

$$A = \begin{bmatrix} 2+i & 2 & -1 \\ 1 & 5 & i \\ 1 & -1 & -1 \end{bmatrix}$$

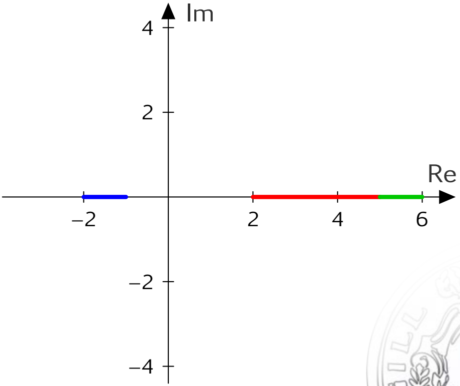


Gerschgorin's Circle Theorem

- The circle theorem can be extended to interval matrices by considering the worst case
- We want positive-semidefiniteness, therefore ″worst case“ should be interpreted as lowest eigenvalue

Example

$$H = \begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix}$$

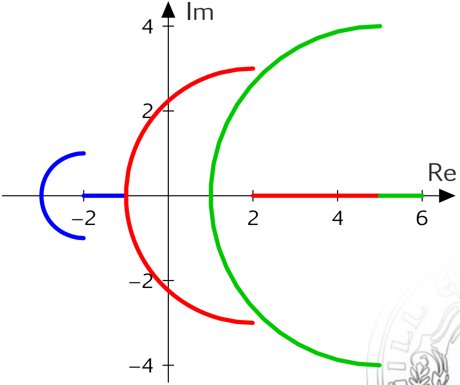


Gerschgorin's Circle Theorem

- The circle theorem can be extended to interval matrices by considering the worst case
- We want positive-semidefiniteness, therefore ″worst case“ should be interpreted as lowest eigenvalue

Example

$$H = \begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix}$$



Original αBB

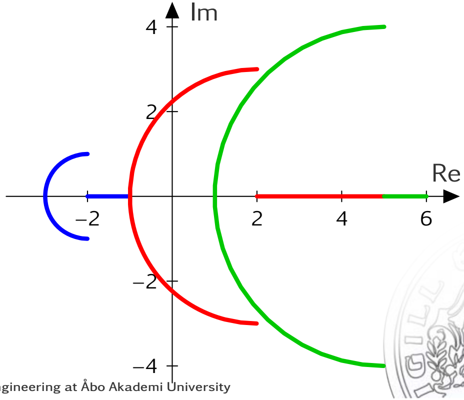
- The function f is underestimated by adding the perturbation
  $$-\sum_i \alpha_i (x_i^U - x_i)(x_i - x_i^L)$$
- To guarantee positive-semidefiniteness we set the constraints

$$\underline{h_{ii}} + 2\alpha_i - \sum_{j \neq i} \max(|\underline{h_{ij}}|, |\overline{h_{ij}}|) \geq 0, \quad i = 1, 2, \ldots, n$$

$$\begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & & \\ & 0 & \\ & & 3 \end{bmatrix}$$

$$= \begin{bmatrix} [3,6] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [1,4] \end{bmatrix}$$



Original αBB

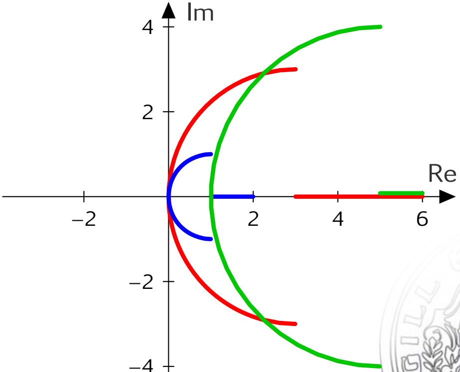
- The function f is underestimated by adding the perturbation
  $-\sum_i \alpha_i (x_i^U - x_i)(x_i - x_i^L)$
- To guarantee positive-semidefiniteness we set the constraints

$$\underline{h_{ii}} + 2\alpha_i - \sum_{j \neq i} \max(|\underline{h_{ij}}|, |\overline{h_{ij}}|) \geq 0, \quad i = 1, 2, \ldots, n$$

$$\begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix}$$

$$+ \begin{bmatrix} 1 & & \\ & 0 & \\ & & 3 \end{bmatrix}$$

$$= \begin{bmatrix} [3,6] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [1,4] \end{bmatrix}$$



Extended αBB

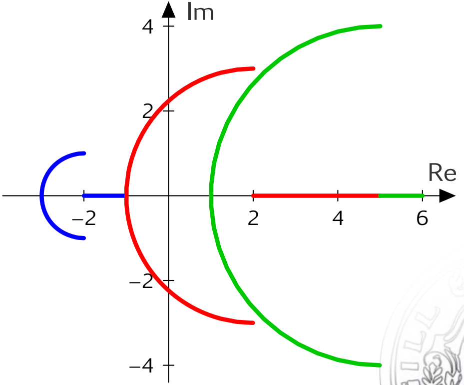
- From a Hessian perspective the bilinear extension is intuitive
- To guarantee positive-semidefiniteness we set the constraints

$$\underline{h_{ii}} + 2\alpha_i - \sum_{j \neq i} \max(|\underline{h_{ij}} + \beta_{ij}|, |\overline{h_{ij}} + \beta_{ij}|) \geq 0, \quad i = 1, 2, \ldots, n$$

$$\begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix}$$

$$+ \begin{bmatrix} & -1 & \\ -1 & & 0.5 \\ & 0.5 & 2.5 \end{bmatrix}$$

$$= \begin{bmatrix} \color{red}{[2,5]} & \color{red}{[-2,2]} & \color{red}{0} \\ \color{green}{[-2,2]} & \color{green}{[5,6]} & \color{green}{[-0.5,0.5]} \\ \color{blue}{0} & \color{blue}{[-0.5,0.5]} & \color{blue}{[0.5,1.5]} \end{bmatrix}$$



Extended αBB

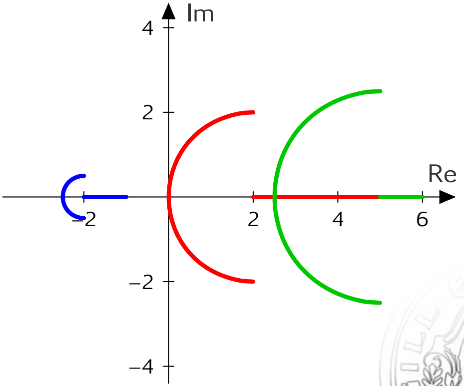
- From a Hessian perspective the bilinear extension is intuitive
- To guarantee positive-semidefiniteness we set the constraints

$$\underline{h_{ii}} + 2\alpha_i - \sum_{j \neq i} \max(|\underline{h_{ij}} + \beta_{ij}|, |\overline{h_{ij}} + \beta_{ij}|) \geq 0, \quad i = 1, 2, \dots, n$$

$$\begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix}$$

$$+ \begin{bmatrix} & -1 & \\ -1 & & 0.5 \\ & 0.5 & 2.5 \end{bmatrix}$$

$$= \begin{bmatrix} [2,5] & [-2,2] & 0 \\ [-2,2] & [5,6] & [-0.5,0.5] \\ 0 & [-0.5,0.5] & [0.5,1.5] \end{bmatrix}$$



Extended αBB

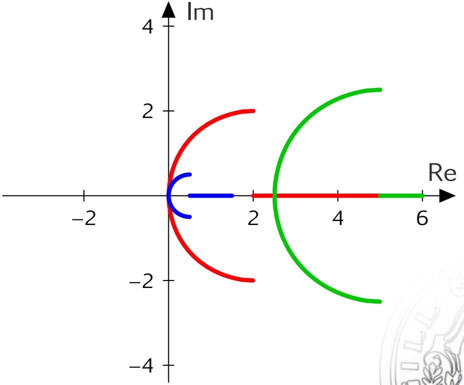
- From a Hessian perspective the bilinear extension is intuitive
- To guarantee positive-semidefiniteness we set the constraints

$$\underline{h_{ii}} + 2\alpha_i - \sum_{j \neq i} \max(|\underline{h_{ij}} + \beta_{ij}|, |\overline{h_{ij}} + \beta_{ij}|) \geq 0, \quad i = 1, 2, \dots, n$$

$$\begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix}$$

$$+ \begin{bmatrix} & -1 & \\ -1 & & 0.5 \\ & 0.5 & 2.5 \end{bmatrix}$$

$$= \begin{bmatrix} [2,5] & [-2,2] & 0 \\ [-2,2] & [5,6] & [-0.5,0.5] \\ 0 & [-0.5,0.5] & [0.5,1.5] \end{bmatrix}$$



Bilinear Perturbation Terms

$$\begin{bmatrix} [2,5] & [-1,3] & 0 \\ [-1,3] & [5,6] & [-1,0] \\ 0 & [-1,0] & [-2,-1] \end{bmatrix} + \begin{bmatrix} 0 & -1 & 0 \\ -1 & 0 & 0.5 \\ 0 & 0.5 & 2.5 \end{bmatrix} = \begin{bmatrix} [2,5] & [-2,2] & 0 \\ [-2,2] & [5,6] & [-0.5,0.5] \\ 0 & [-0.5,0.5] & [0.5,1.5] \end{bmatrix}$$

Two questions must be answered

- How can we interpret the off-diagonal adjustments of the Hessian as perturbation terms?
- Is the new underestimator an improvement?

Realizing the Perturbations

Let the perturbation Hessian be

$$H_P = \begin{bmatrix} 2\alpha_1 & \beta_{1,2} & \cdots & \beta_{1,n} \\ \beta_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \beta_{n-1,n} \\ \beta_{n,1} & \cdots & \beta_{n,n-1} & 2\alpha_n \end{bmatrix}$$

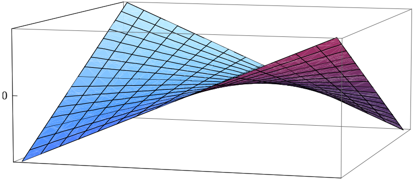
Realizing the Perturbations

Let the perturbation Hessian be

$$H_P = \begin{bmatrix} 2\alpha_1 & \beta_{1,2} & \cdots & \beta_{1,n} \\ \beta_{2,1} & \ddots & & \vdots \\ \vdots & & \ddots & \beta_{n-1,n} \\ \beta_{n,1} & \cdots & \beta_{n,n-1} & 2\alpha_n \end{bmatrix}$$

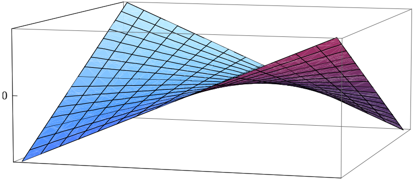
- The intuitive realization of β_{ij} is $\beta_{ij} x_i x_j$
- By adding linear and constant terms we get a symmetric perturbation, $\beta_{ij}(x_i - x_i^M)(x_j - x_j^M)$, where $x_i^M = \frac{x_i^L + x_i^U}{2}$

Realizing the Perturbations

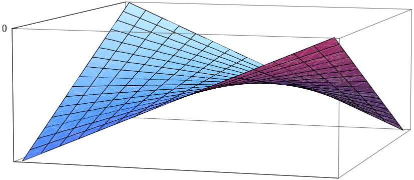


$\beta_{ij}(x_i - x_i^M)(x_j - x_j^M)$

Realizing the Perturbations



$\beta_{ij}(x_i - x_i^M)(x_j - x_j^M)$



$\beta_{ij}(x_i - x_i^M)(x_j - x_j^M) - \frac{x_i^L + x_i^U}{2} \cdot \frac{x_j^L + x_j^U}{2}$

- We can subtract a positive constant to ensure underestimation
- This works but restricts the potential of the new underestimator

Realizing the Perturbations

- How else can we adjust $\beta_{ij} x_i x_j$ to ensure the underestimation property?
- We can utilize the well-known concave envelope of a bilinear function (McCormick 1976)

Realizing the Perturbations

- How else can we adjust $\beta_{ij} x_i x_j$ to ensure the underestimation property?
- We can utilize the well-known concave envelope of a bilinear function (McCormick 1976)
- The suggested perturbation corresponding to β_{ij} is

$$\beta_{ij} x_i x_j - \widehat{\beta_{ij} x_i x_j}$$

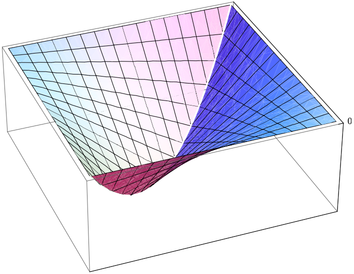
where $\widehat{\ }$ denotes the concave envelope

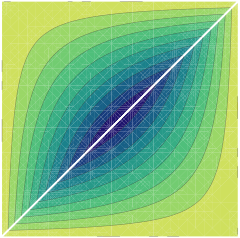
Realizing the Perturbations

- How else can we adjust $\beta_{ij} x_i x_j$ to ensure the underestimation property?
- We can utilize the well-known concave envelope of a bilinear function (McCormick 1976)
- The suggested perturbation corresponding to β_{ij} is

$$\beta_{ij} x_i x_j - \widehat{\beta_{ij} x_i x_j}$$

where $\widehat{\ }$ denotes the concave envelope





Is the New Underestimator Tighter?

- We measure tightness as the largest underestimation error
- The largest error obtained in the hyper-rectangular domain is

$$\sum_i \alpha_i \left(\frac{x_i^U - x_i^L}{2} \right)^2 + \sum_i \sum_{j>i} |\beta_{ij}| \frac{(x_i^U - x_i^L)(x_j^U - x_j^L)}{4}$$

Is the New Underestimator Tighter?

- We measure tightness as the largest underestimation error
- The largest error obtained in the hyper-rectangular domain is

$$\sum_i \alpha_i \left(\frac{x_i^U - x_i^L}{2} \right)^2 + \sum_i \sum_{j>i} |\beta_{ij}| \frac{(x_i^U - x_i^L)(x_j^U - x_j^L)}{4}$$

- We can optimize (α, β), minimizing the maximum error under the convexification constraints → a convex NLP

Is the New Underestimator Tighter?

- We measure tightness as the largest underestimation error
- The largest error obtained in the hyper-rectangular domain is

$$\sum_i \alpha_i \left(\frac{x_i^U - x_i^L}{2} \right)^2 + \sum_i \sum_{j>i} |\beta_{ij}| \frac{(x_i^U - x_i^L)(x_j^U - x_j^L)}{4}$$

- We can optimize (α, β), minimizing the maximum error under the convexification constraints $\rightarrow$ a convex NLP
- The minimization can be reformulated as a linear program

Choosing the Parameters

$$J_i^+ := \left\{ j : j \neq i,\ \underline{h_{ij}} + \overline{h_{ij}} \geq 0 \right\}, \qquad J_i^- := \left\{ j : j \neq i,\ \underline{h_{ij}} + \overline{h_{ij}} < 0 \right\}$$

$$\min_{\alpha,\beta} \quad \sum_i \frac{\alpha_i}{4}(x_i^U - x_i^L)^2 - \sum_i \sum_{\substack{j>i \\ j \in J_i^+}} \frac{\beta_{ij}}{4}(x_i^U - x_i^L)(x_j^U - x_j^L)$$

$$+ \sum_i \sum_{\substack{j>i \\ j \in J_i^-}} \frac{\beta_{ij}}{4}(x_i^U - x_i^L)(x_j^U - x_j^L)$$

$$\begin{aligned}
\text{s.t.} \quad & \underline{h_{ii}} + 2\alpha_i - \sum_{j \in J_i^+} (\overline{h_{ij}} + \beta_{ij}) + \sum_{j \in J_i^-} (\underline{h_{ij}} + \beta_{ij}) \geq 0, && \forall\, i \\
& \alpha_i \geq 0, && \forall\, i \\
& \beta_{ij} = \beta_{ji}, && \forall\, i, j : j > i \\
& \min\left(0, -(\underline{h_{ij}} + \overline{h_{ij}})/2\right) \leq \beta_{ij} \leq \max\left(0, -(\underline{h_{ij}} + \overline{h_{ij}})/2\right), && \forall\, i, j : j \neq i
\end{aligned}$$

Example

$$f(\boldsymbol{x}) = (1 + x_1 - e^{x_2})^2, \qquad \boldsymbol{x} \in [0,1] \times [0,2]$$

$$H(\boldsymbol{x}) = \begin{bmatrix} 2 & -2e^{x_2} \\ -2e^{x_2} & -2e^{x_2}(1 - 2e^{x_2} + x_1) \end{bmatrix} \in \begin{bmatrix} 2 & [-14.8, -2] \\ [-14.8, -2] & [-12.8, 203.6] \end{bmatrix}$$

Example

$$f(\boldsymbol{x}) = (1 + x_1 - e^{x_2})^2, \qquad \boldsymbol{x} \in [0,1] \times [0,2]$$

$$H(\boldsymbol{x}) = \begin{bmatrix} 2 & -2e^{x_2} \\ -2e^{x_2} & -2e^{x_2}(1 - 2e^{x_2} + x_1) \end{bmatrix} \in \begin{bmatrix} 2 & [-14.8, -2] \\ [-14.8, -2] & [-12.8, 203.6] \end{bmatrix}$$

Original αBB

$$\check{f}(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{12.8}{2}(1 - x_1)x_1 - \frac{27.6}{2}(2 - x_2)x_2$$

maximum error: 15.4

Example

$$f(\boldsymbol{x}) = (1 + x_1 - e^{x_2})^2, \qquad \boldsymbol{x} \in [0,1] \times [0,2]$$

$$H(\boldsymbol{x}) = \begin{bmatrix} 2 & -2e^{x_2} \\ -2e^{x_2} & -2e^{x_2}(1 - 2e^{x_2} + x_1) \end{bmatrix} \in \begin{bmatrix} 2 & [-14.8, -2] \\ [-14.8, -2] & [-12.8, 203.6] \end{bmatrix}$$

Original αBB

$$\check{f}(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{12.8}{2}(1 - x_1)x_1 - \frac{27.6}{2}(2 - x_2)x_2$$

maximum error: 15.4

Extended αBB

$$\check{f}(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{4.4}{2}(1 - x_1)x_1 - \frac{19.2}{2}(2 - x_2)x_2 + 8.4\, x_1 x_2 - 8.\widehat{4\, x_1} x_2$$

maximum error: 14.35

Example

$$f(\boldsymbol{x}) = (1 + x_1 - e^{x_2})^2, \qquad \boldsymbol{x} \in [0,1] \times [0,2]$$

$$H(\boldsymbol{x}) = \begin{bmatrix} 2 & -2e^{x_2} \\ -2e^{x_2} & -2e^{x_2}(1 - 2e^{x_2} + x_1) \end{bmatrix} \in \begin{bmatrix} 2 & [-14.8, -2] \\ [-14.8, -2] & [-12.8, 203.6] \end{bmatrix}$$

Original αBB

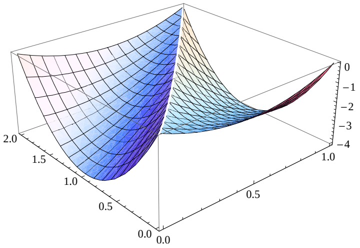
$$\check{f}(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{12.8}{2}(1 - x_1)x_1 - \frac{27.6}{2}(2 - x_2)x_2$$

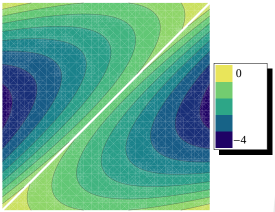
maximum error: 15.4

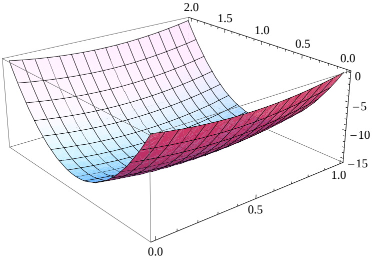
Extended αBB

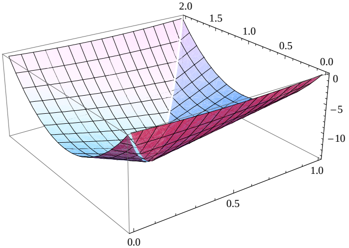
$$\check{f}(\boldsymbol{x}) = f(\boldsymbol{x}) - \frac{4.4}{2}(1 - x_1)x_1 - \frac{19.2}{2}(2 - x_2)x_2 + 8.4\, x_1 x_2 - 8.\widehat{4\, x_1} x_2$$

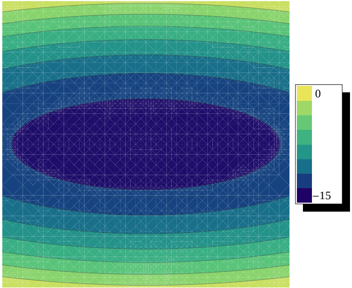
maximum error: 14.35

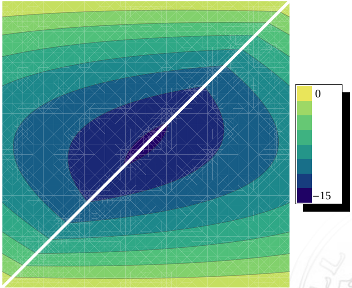





2.0
1.5
1.0
0.5
0.0
0
−5
−10
−15
1.0
0.5
0.0


2.0
1.5
1.0
0.5
0.0
0
−5
−10
1.0
0.5
0.0


0
−15


0
−15

References

C.S. Adjiman, S. Dallwig, C.A. Floudas, and A. Neumaier.
A global optimization method, αBB, for general twice-differentiable constrained NLPs – I. theoretical advances.
Computers & Chemical Engineering, 22(9):1137 – 1158, 1998.

A. Skjäl, R. Misener, T. Westerlund, and C.A. Floudas.
A generalization of classical αBB underestimation to include bilinear terms.
In *Proceedings of the 22nd European Symposium on Computer Aided Process Engineering*, 2012.
submitted for review.

A. Skjäl, T. Westerlund, R. Misener, and C.A. Floudas.
A generalization of the classical αBB convex underestimation via diagonal and non-diagonal quadratic terms.
2011.
submitted for review.

Thank you for listening!

Thank you for listening!

Questions?