Interval Analysis on DAGs for Global Optimization II

. . . .

Hermann Schichl
Arnold Neumaier

Universität Wien

as part of the COCONUT project,
a European community initiative for the integration of different global optimization techniques

www.mat.univie.ac.at/~neum/glopt/coconut/

Motivation

Global optimization with guarantees

- constrained propagation
- interval Newton type methods
- outer approximation
- branch & bound

Outer approximation replaces "difficult" (= e.g. nonconvex) constraints by simpler, redundant constraints to create a tractable problem whose solution helps in the b&b process (pruning leaves in the search tree reducing box sizes)

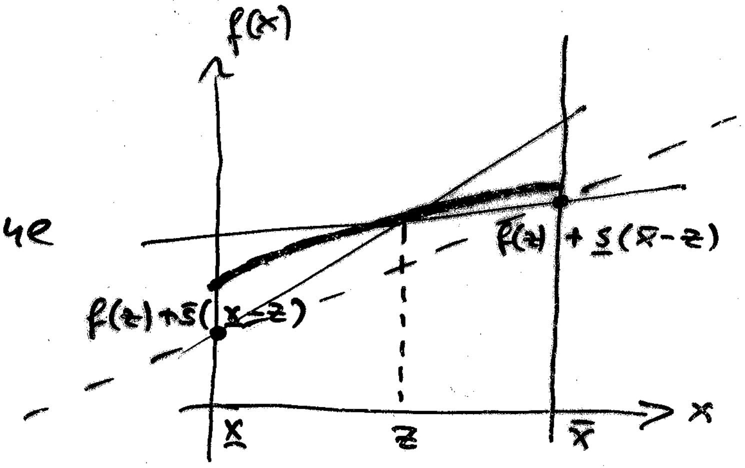
- linear underestimation ⇒ LP relaxation
  traditional; solve by Simplex/IP
- convex quadratic underestimation
  ⇒ SOC relaxations
  solve by interior point methods
- polynomial underestimation
  ⇒ polynomial relaxation
  solve by moment methods
  (Gloptipoly – Lasserre)

⇓

Q: How to find linear / quadratic underestimators?

Linear underestimation

(i) Using slopes



The slope defines a double cone containing the function

The linear interpolant at the cone's ends defines an underestimating line. Separable $\Rightarrow$ each component treated independently

$$f(x) \geq f(z) + \bar{s}^T(\underline{x} - z) + c^T(x - \underline{x})$$

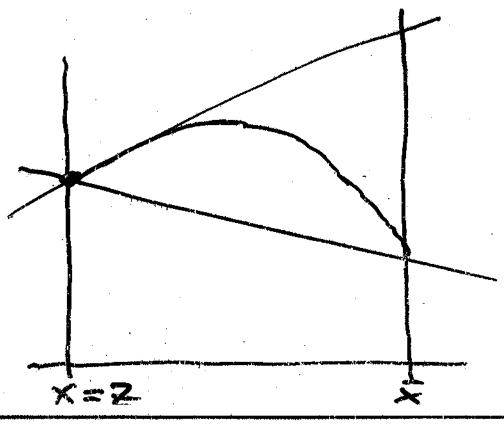
where $$c_i = \frac{\underline{s}_i(\bar{x}_i - z_i) - \bar{s}_i(\underline{x}_i - z_i)}{\bar{x}_i - \underline{x}_i}$$

⊕
- uses standard techniques
- nice geometric interpretation

⊖
- suboptimal even in simple cases (e.g. quadratic f)

<u>special case</u>: slope at a corner

⊕ optimal if slope optimal



Example. Consider the nonlinear constraint

$$f(x) = (4x_1 - x_2x_3)(x_1x_2 + x_3) \le -96$$

$$x_1 \in [1,2], \quad x_2 \in [3,4], \quad x_3 \in [3,4]$$

$$z_1 = 2, \quad z_2 = 3, \quad z_3 = 4$$

The center is in a corner of the box

$\Rightarrow$ underestimator using slopes is expected to be good

slope from DAG $\Rightarrow$ centered form

$$f(x) \in -96 + [9,24]\underbrace{(x_1-2)}_{\le 0} + [-64,-48]\underbrace{(x_2-4)}_{\le 0} + [-56,-32]\underbrace{(x_3-4)}_{\le 0}$$

$$\Rightarrow f(x) \ge -96 + 24(x_1-2) - 48(x_2-4) - 32(x_3-4)$$

$\Rightarrow$ linear relaxation of the constraint is

$$\underbrace{24(x_1-2)}_{\ge -24} \underbrace{-48(x_2-4)}_{\ge 0} \underbrace{-32(x_3-4)}_{\ge 0} \le 0$$

$$\Rightarrow x_2 \ge 3.5, \quad x_3 \ge 3.25$$

$\Rightarrow$ reduced box $\quad x_2 \in [3.5, 4], \quad x_3 \in [3.25, 4]$

Linear underestimation

(ii) Recursive underestimation

- backward on the DAG
- all intermediate results get a label x_i
- propagate inequality of the form

$$f(x) \geq f_c + s^T(x-z)$$

- at each node, eliminate the intermediate x_i using linear bound from the operation

$x_i = \varphi(x_k)$:

$$\left.\begin{array}{l} \varphi(x_k) \geq \alpha + \beta(x_k - z_k) \\ \varphi(x_k) \leq \alpha' + \beta'(x_k - z_k) \end{array}\right\} \text{ for all } x_k \in \mathbf{x}_k$$

$$\Rightarrow \quad s_i(x_i - z_i) \geq \begin{cases} s_i(\alpha - z_i) + s_i\beta(x_k - z_k) & \text{if } s_i > 0 \\ s_i(\alpha' - z_i) + s_i\beta'(x_k - z_k) & \text{if } s_i < 0 \end{cases}$$

$$x_i = x_j x_k \quad \begin{array}{l} \geq x_j \underline{x}_k + \underline{x}_j x_k - \underline{x}_j \underline{x}_k \\ \geq x_j \overline{x}_k + \overline{x}_j x_k - \overline{x}_j \overline{x}_k \\ \leq x_j \underline{x}_k + \overline{x}_j x_k - \overline{x}_j \underline{x}_k \\ \leq x_j \overline{x}_k + \underline{x}_j x_k - \underline{x}_j \overline{x}_k \end{array}$$

McCormick
Al-Khayyal
(optimal!)

but which choice to take?

$$x_i = x_j / x_k \quad \geq \underline{x}_k^{-1} x_j - \overline{x}_k^{-1} \underline{x}_i (x_k - \underline{x}_k)$$

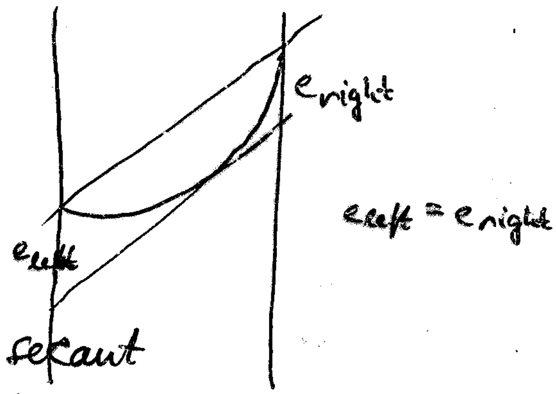
and 3 similar formulas $\geq, \leq, \leq$

Substitution $\Rightarrow$ new linear underestimate without x_i

at the end $\Rightarrow$ underestimate in the original variables only

optimal linear enclosure in 1D, φ convex:
(minimize maximal error)

$\Rightarrow$ secant & tangent parallel to secant.



φ concave $\Rightarrow$ same recipe

φ general $\Rightarrow$ two tangents parallel to secant is suboptimal but good

$\Rightarrow$ task reduced to find range of $e(x) = \varphi(x) - \varphi(\underline{x}) - \varphi[\underline{x}, \overline{x}](x - \underline{x})$

tangent equation: $e'(x) = 0$: $\varphi'(x) = \varphi[\underline{x}, \overline{x}]$

- easy for all elementary functions

- can be extended to univariate subexpressions $\varphi(x)$
  $\Rightarrow$ reduces to a 1D global optimization problem
  (only <u>one</u> of the directions is needed)

- optimal possible in general – uses convex envelopes
  (Rainer Haderer)

Linear enclosures

needed as relaxations of equality constraints

$$h(x) = 0$$

$$\alpha + a^T x \le h(x) \le \beta + b^T x$$

$$\Downarrow$$

$$\alpha + a^T x \le 0$$

$$\beta + b^T x \ge 0$$

computed either by underestimating $h(x)$ and $-h(x)$

or, with $a=b$, as

$$h(x) \in [\underline{c}, \overline{c}] + a^T x$$

recursively in backward mode

$$x_i = \varphi(x_k) \in [\underline{\alpha}, \overline{\alpha}] + \beta x_k \qquad \text{(secant parallels!)}$$

$$\Rightarrow [\underline{c}, \overline{c}] + a^T x \le [\underline{c}, \overline{c}] + a_i [\underline{\alpha}, \overline{\alpha}] + \text{linear}$$

similarly for binary operations

Caution: Rounding error control !

Quadratic underestimators

(i) from interval Hessians

$$f(x) = f(z) + g(z)^T(x-z) + \frac{1}{2}(x-z)^T G (x-z)$$

for some $G \in f''(\mathbf{x}) =: \mathbf{G}$

- Interval Hessians can be computed on the DAG as in automatic differentiation

$$\Rightarrow f(x) \geq f_0 + g_0^T(x-z) + \frac{1}{2}(x-z)^T G_0 (x-z)$$

where g_0, G_0 arbitrary

$$f_0 \leq \inf_{\substack{x \in \mathbf{x} \\ G \in \mathbf{G}}} \left(f(z) + (g(z)-g_0)^T(x-z) + \frac{1}{2}(x-z)^T(G-G_0)(x-z) \right)$$

Natural choice: $g_0 = \operatorname{mid} g(z)$

$$G_0 = \begin{pmatrix} \underline{G}_{11} & & \operatorname{mid}\mathbf{G}_{ik} \\ & \ddots & \\ \operatorname{mid}G_{ik} & & \underline{G}_{nn} \end{pmatrix}$$

$\Rightarrow f_0 = \underline{\mathbf{f}}_0$ where

$$\mathbf{f}_0 = f(z) + (g(z) - g_0 + \mathbf{R}(\mathbf{x}-z))^T(\mathbf{x}-z)$$

$$\mathbf{R} = \begin{pmatrix} 0 & & \mathbf{R}_{ik} \\ & \ddots & \\ 0 & & 0 \end{pmatrix}, \quad \mathbf{R}_{ik} = \mathbf{G}_{ik} - \operatorname{mid}\mathbf{G}_{ik}$$

or sharper but more expensive enclosures

Quadratic underestimators

(ii) from 2nd order slopes

$$f(x) = f(z) + g(z)^T (x-z) + (x-z)^T f[z,z,x]\,(x-z)$$

$$f''(x) = 2f[x,x,x]$$

$f(x) = x^3 \quad \Rightarrow \quad f[\mathbf{x},\mathbf{x},\mathbf{x}] = 3\mathbf{x}$ radius $3r$

$\qquad\qquad\qquad\qquad f[z,z,\mathbf{x}] = 2z + \mathbf{x}$ radius r

$\Rightarrow$ slopes are superior

- recursive 2nd order slopes on DAG in backward mode (modified Hessian computation)

- For 1D subexpressions: 2nd order <u>forward</u> slopes are cheap

    - For convex or concave elementary functions $\varphi(x)$:

      $\varphi[z,z,\mathbf{x}] = \square\{\varphi[z,z,\underline{x}], \varphi[z,z,\overline{x}]\}$ (Kolev)

Quadratic underestimators

(iii) direct backward underestimation

- propagate equations of the <u>semiseparable form</u>
$$f(x) = \sum f_k(x_k) + (x-z)^T B(x-z) \quad , B \in \mathbb{B}$$

- at each node, eliminate the intermediate x_i from the separable part only
  - $x_i = \varphi(x_k)$ simply changes $f_k(x_k) \leftarrow f_k(x_k) + f_i(\varphi(x_k))$ (<u>symbolical</u> substitution only)
  - $x_i = x_j \circ x_k$: Use quadratic underestimator for $f_i(x_i)$ needs 1D 2nd order slopes
$$f_i(x_i) = f_i(z_i) + f_i'(z_i)(x_i - z_i) + f_i[z_i, z_i, x_i](x_i - z_i)^2$$
    - move the quadratic coefficient into B
    - rewrite $x_i - z_i = x_j \circ x_k - z_i$ as quadratic form with interval coefficient

$x_i = \lambda x_j + \mu x_k \quad \Rightarrow \quad x_i - z_i$ is already linear in x_j, x_k

$x_i = x_j x_k \quad \Rightarrow \quad x_i - z_i = z_j x_k + x_j z_k - z_j z_k - z_i + (x_j - z_j)(x_k - z_k)$

$x_i = x_j / x_k \quad \Rightarrow \quad x_i - z_i = \dfrac{x_j - z_j}{z_k} + \dfrac{z_j}{z_k^2} x_k - z_i + \dfrac{(x_k - z_k)(z_j x_k - z_k x_j)}{z_k^2 x_k}$

- For each intermediate variable in the final quadratic term find a linear enclosure w.r. to the original variables

- Finally, substitute these linear enclosures, and move interval uncertainties into the constant term