Generalized Convexity Applied to Branch-and-Bound Algorithms for MINLPs

Ksenia Bestuzheva
bestuzheva@zib.de

33rd European Conference on Operational Research,
Copenhagen, Denmark

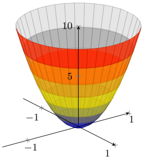
July 2, 2024

ZIB ZUSE INSTITUTE BERLIN
MODAL FORSCHUNGSCAMPUS
SPONSORED BY THE Federal Ministry of Education and Research

Problem

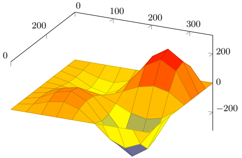
$$\begin{aligned} \min\ & c^\mathsf{T} x \\ \text{s.t. } & g_i(x) \le 0 \ \forall k \in \mathcal{C}, \\ & x \in [\underline{x}, \overline{x}], \\ & x_j \in \mathbb{R}^{|\mathcal{J}|} \ \forall j \in \mathcal{J}. \end{aligned} \tag{1}$$

- The functions $g_i : [\underline{x}, \overline{x}] \to \mathbb{R}$ can be



convex

or



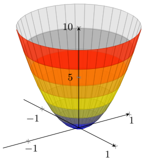
nonconvex

and are given in **algebraic form**.

- Bounds can be infinite
- Our approaches are aimed to be applied within a **spatial and integer branch & bound** algorithm.

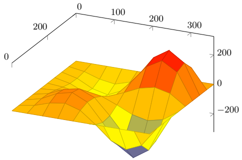
Generalized convexity: motivation

- Convex problems can be solved to global optimality by polynomial time algorithms
- Nonconvexity $\rightarrow$ local optima and non-optimal KKT points, optimality and feasibility difficult to prove



convex

or



nonconvex

- Some nonconvex problems behave in some ways similarly to convex problems
- Empirical example: for optimal power flow and pooling problems, solvers tend to converge to the global optimum
- Research on generalized convexity is concerned with leveraging such similarities

Generalized convex functions

Consider differentiable function $f\colon\ X \to \mathbb{R}$

Convex function:
$\nabla f(x)(y-x) \leq f(y) - f(x)$ for all $x, y \in X$ or
$f(\lambda x + (1-\lambda)y) \leq \lambda f(x) + (1-\lambda f(y))$ for all $\lambda \in [0,1]$ and $x, y \in X$

Generalizations based on various relaxations of this condition:

- Pseudoconvex function: $\nabla f(x)(y-x) \geq 0 \Rightarrow f(y) \geq f(x)$ for all $x, y \in X$ (Mangasarian, 1965)
- Quasiconvex function: $f(\lambda x + (1-\lambda)y) \leq \max\{f(x), f(y)\}$ for all $\lambda \in [0,1]$ and $x, y \in X$ (Fenchel, 1953)
- Invex function: $\eta(x,y)\nabla f(x) \leq f(y) - f(x)$ for all $x, y \in X$ and some vector valued function η (Hanson, 1981)

Generalized convexity and optimization

- Quasiconvex problems can be solved efficiently in theory, but the algorithms are slow in practice (Kiwiel (2001))
- KKT-invex optimization problem: a problem such that every KKT (Karush-Kuhn-Tucker) point is a global optimum (Hanson, 1981)
- Type-I invexity, a generalization of invex functions, is necessary and sufficient for KKT-invexity (Hanson, 1999)
- Further generalizations: B-preinvexity (Yang, 2002), E-invexity (Jaiswal, Panda, 2012), B-invexity and -preinvexity of interval-valued functions (Abdulaleem, 2023), etc...

Boundary-invexity

A new approach to KKT-invexity, based on a special subclass of stationary points (Bestuzheva, Hijazi, 2019):

For each $i \in \mathcal{C}$ consider:

$$\max_{x \in \mathbb{R}^n} c^T x \ \text{ s.t. } \ g_i(x) = 0, \tag{2}$$

Definition

(Boundary-invexity) Problem (1) is boundary-invex if, for every KKT point x^* of (2) for every $i \in \mathcal{C}$ that corresponds to a non-convex constraint, at least one of the following holds: a) x^* is infeasible for (1), b) the Lagrange multiplier for x^* in (2) is non-negative, c) x^* is a local maximum with respect to (1).

- Boundary-invexity is sufficient for KKT-invexity of problems in $\mathbb{R}^2$ and a weaker version of boundary-invexity is necessary for KKT-invexity of problems of an arbitrary finite dimension
- Proved KKT-invexity of optimal power problems for one line with realistic parameters

New approach: optima-invexity

A further shift of perspective: study different types of stationary points of (1) and their interplay

Definition

A problem is optima-invex if it has a unique connected locally optimal subset.

- Focus on distinct (sets of) local optima
- Optima-invexity is concerned with local optima (and not non-optimal KKT points)
- In practice, convergence to non-optimal KKT points tends to not be an issue (Lee et al., 2016)
- The notion of optima-invexity serves as a basis for studying the relation between non-optimal KKT points and local optima

Main theoretical result

Theorem

If problem (1) does not have non-optimal KKT points, then it is optima-invex.

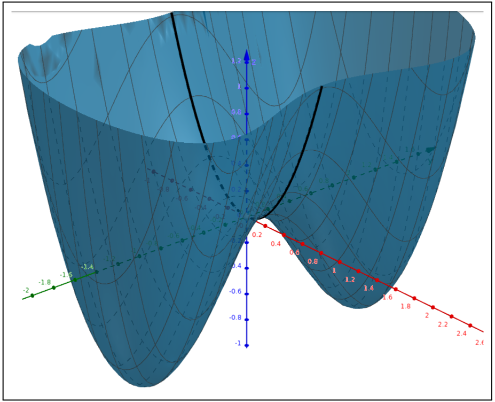
That is, the absence of non-optimal KKT points implies the absence of multiple distinct local optima.

Main theoretical result: proof outline

Step 1: Assume the contrary: there exist distinct local optima $\mathbf{x}^{1*}$, $\mathbf{x}^{2*}$.

Step 2: consider regions of attraction A_1, A_2 of $\mathbf{x}^{1*}$, $\mathbf{x}^{2*}$ respectively, and the shared boundary δA.
Prove that $\delta A \not\subset A_i$ for $i = 1, 2$.

Step 3: consider $x^B \in \delta A$. There are two possibilities:
- x^B is a non-optimal KKT point or
- x^B belongs to a region of attraction of a non-optimal KKT point.



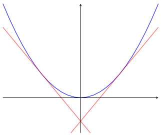
Practical uses

- Optimality proofs: showing the lack of non-optimal KKT points is enough (challenging in general, but may work in special cases)
- Optimality proofs of subproblems: for example a node subproblem in branch and bound, or a subproblem in a decomposition
- Heuristics within branch-and-bound approaches
  - Use invexity-breaking points to guide the search towards new distinct local optima
  - Can also be applied to classes of problems for which there is yet no formal proof of sufficiency

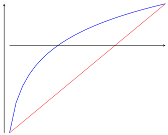
Spatial branch-and-bound

LP relaxation via convexification and linearization:

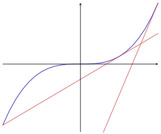
convex functions



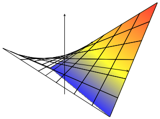
concave functions



$x^k \quad (k \in 2\mathbb{Z}+1)$



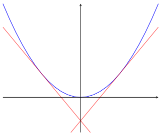
$x \cdot y$



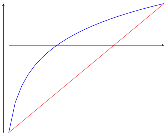
Spatial branch-and-bound

LP relaxation via convexification and linearization:

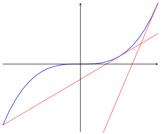
convex functions



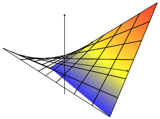
concave functions



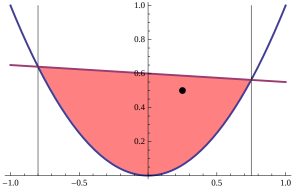
$x^k \quad (k \in 2\mathbb{Z}+1)$

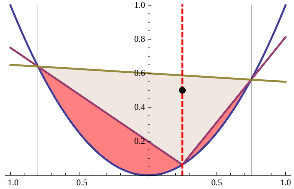


$x \cdot y$



Branching on variables in violated nonconvex constraints:

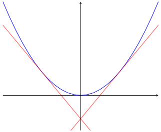




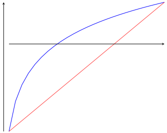
Spatial branch-and-bound

LP relaxation via convexification and linearization:

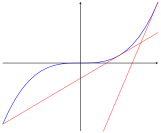
convex functions



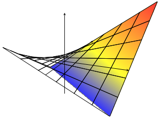
concave functions



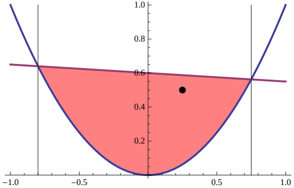
$x^k \quad (k \in 2\mathbb{Z}+1)$

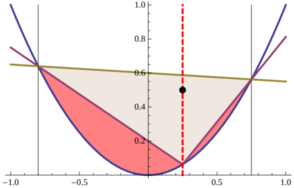


$x \cdot y$



Branching on variables in violated nonconvex constraints:





...and **bound tightening**, **primal heuristics** (e.g., sub-NLP/MIP/MINLP), other special techniques
When integer variables are present: combine with integer branching and other techniques

Invexity-guided diving heuristics

Idea

Use invexity-breaking points to guide the search towards new distinct local optima.

- Multi-start or diving heuristics for NLPs that divide the region at non-optimal KKT points
- Diving heuristics for MINLPs that prioritise fixing integer variables that participate in invexity-breaking constraints
- Challenges:
  - What is the relation between invexity-breaking points and regions of attraction?
  - Boundaries between regions of attractions may not be linear
  - Finding invexity-breaking points

Constraint-wise algorithm

- Initialize grid to one cell defined by variable domains (can be infinite)
- For each new cell in the grid:
  - Set a starting point within the cell
  - Solve the NLP and get the local optimum x^*
  - For each active nonconvex constraint g_i:
    - Choose a variable x_j such that g_i is nonconvex in x_j
    - Find non-optimal KKT points of $\min c^T \hat{x}^{(j)}$ s.t. $g_i(\hat{x}^{(j)})$, where $\hat{x}^{(j)}$ is obtained by fixing all components of x except for x_j; denote the points by $x^{*(ijk)}$
    - Add the hyperplanes $x_j = x^{*(ijk)}$ for each k to grid and mark all adjacent cells as new
    - Mark the combination of variable and constraint as processed

Considerations:

- The difficult part is finding the non-optimal KKT points
- However, it can be done efficiently for simple objectives and constraints
- In some cases this can be done in closed form
- In practice, solvers often decompose expressions into simpler expressions

Example problem

ex14_1_1 from MINLPLib (from Floudas et al., Handbook of Test Problems in Local and Global Optimization)

Formulation:

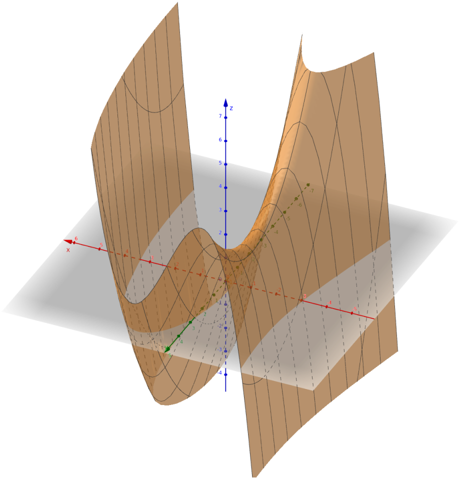
$$\min \ z \text{ s.t.}$$

$$2y^2 + 4xy - 42x + 4x^3 - 14 <= z$$

$$-2xy^2 - 4xy + 42x - 4x^3 + 14 <= z$$

$$2x^2 + 4xy - 26y + 4y^3 - 22 <= z$$

$$-2x^2 - 4xy + 26y - 4y^3 + 22 <= z$$



The problem has multiple local optima (with the same value)

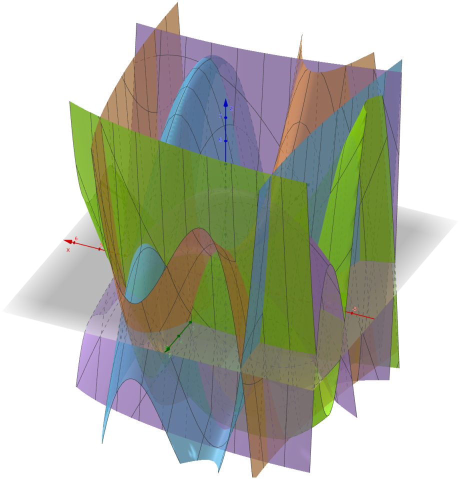
Example problem

ex14_1_1 from MINLPLib (from Floudas et al., Handbook of Test Problems in Local and Global Optimization)

Formulation:

$$
\begin{aligned}
&\min\ z \text{ s.t.}\\
&2y^2 + 4xy - 42x + 4x^3 - 14 <= z\\
&-2xy^2 - 4xy + 42x - 4x^3 + 14 <= z\\
&2x^2 + 4xy - 26y + 4y^3 - 22 <= z\\
&-2x^2 - 4xy + 26y - 4y^3 + 22 <= z
\end{aligned}
$$

The problem has multiple local optima (with the same value)



Example step 1

Grid: $(-5,5) \times (-5,5)$
Starting point: (0,0,5)
NLP solution: (-0.27,-0.92,0)
Active constraints: (1), (2), (3), (4)

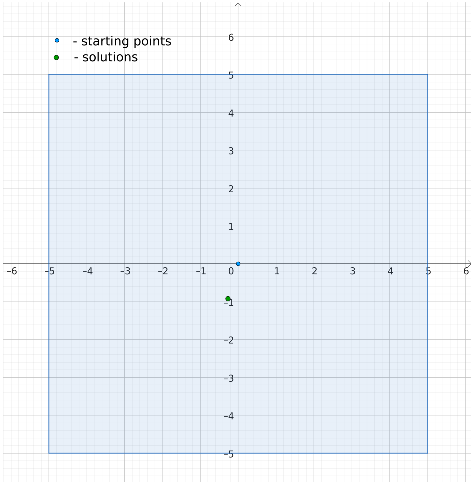
Constraint (1):

- Convex in y, nonconvex in x $\Rightarrow$ focus on x
- Non-optimal stationary point in x: $x* = -\sqrt{\frac{21-2\hat{y}}{6}}$
- Close to linear in $\hat{y}$, separation line: $x = 0.09y - 1.81$
- This works well for separable or near-separable functions

Constraint (2):

- Concave in y, nonconvex in x
- Nonconvexity of x is more pronounced $\Rightarrow$ focus on x
- Non-optimal stationary point in x: $x* = \sqrt{\frac{21-2\hat{y}}{6}}$
- Separation line: $x = -0.09y + 1.81$

Similarly find separation lines for constraints (3) and (4)



Example step 1

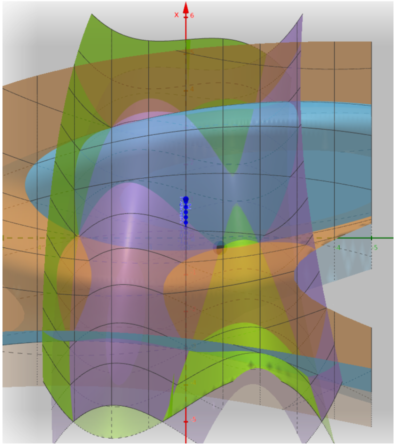
Grid: $(-5, 5) \times (-5, 5)$
Starting point: (0,0,5)
NLP solution: (-0.27,-0.92,0)
Active constraints: (1), (2), (3), (4)

Constraint (1):

- Convex in y, nonconvex in x $\Rightarrow$ focus on x
- Non-optimal stationary point in x: $x* = -\sqrt{\frac{21-2\hat{y}}{6}}$
- Close to linear in $\hat{y}$, separation line: $x = 0.09y - 1.81$
- This works well for separable or near-separable functions

Constraint (2):

- Concave in y, nonconvex in x
- Nonconvexity of x is more pronounced $\Rightarrow$ focus on x
- Non-optimal stationary point in x: $x* = \sqrt{\frac{21-2\hat{y}}{6}}$
- Separation line: $x = -0.09y + 1.81$



Similarly find separation lines for constraints (3) and (4)

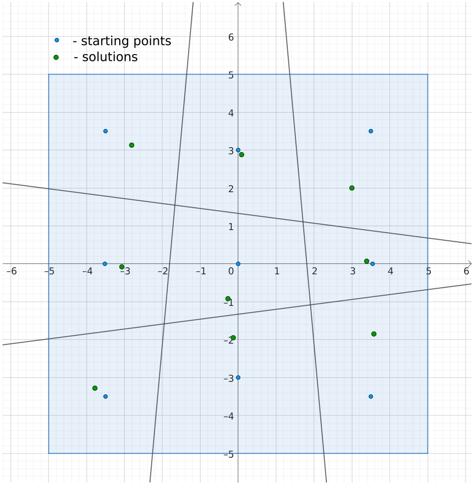
Example step 2

Grid: $(-5, 5) \times (-5, 5)$ divided by four separation lines:
$x = 0.09y - 1.81$
$x = -0.09y + 1.81$
$y = -0.13x + 1.33$
$y = 0.13x - 1.33$

Starting points: (-3.5,3.5,5), (0,3,5), (3.5,3.5,5), (-3.5,0,5), (3.5,0,5), (-3.5,-3.5,5), (0,-3,5), (3.5,-3.5,5)

NLP solutions: (-2.81,3.13,0), (0.09,2.88,0), (3,2,0), (-3.07,-0.08,0), (3.39,0.07,0), (-3.78,-3.28,0), (-0.13,-1.95,0), (3.58,-1.85,0)

Every cell of the grid yielded a new local minimum



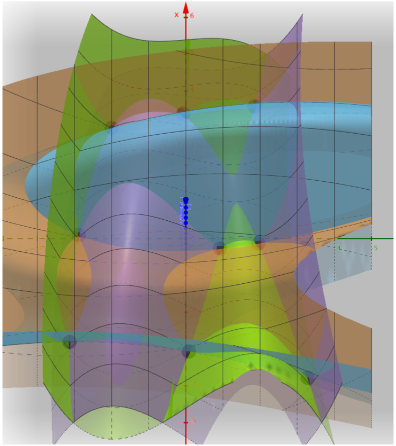
Example step 2

Grid: $(-5,5) \times (-5,5)$ divided by four separation lines:
$x = 0.09y - 1.81$
$x = -0.09y + 1.81$
$y = -0.13x + 1.33$
$y = 0.13x - 1.33$

Starting points: (-3.5,3.5,5), (0,3,5), (3.5,3.5,5), (-3.5,0,5), (3.5,0,5), (-3.5,-3.5,5), (0,-3,5), (3.5,-3.5,5)

NLP solutions: (-2.81,3.13,0), (0.09,2.88,0), (3,2,0), (-3.07,-0.08,0), (3.39,0.07,0), (-3.78,-3.28,0), (-0.13,-1.95,0), (3.58,-1.85,0)

Every cell of the grid yielded a new local minimum



Discussion of the example

- In this case the heuristic works well because the constraints have almost-linear ridges
- How to define and efficiently find such ridges?
- There is a connection between ridges and non-optimal KKT points
- In practice, we often deal with simple expressions (e.g. obtained through extended formulations) - leverage this

Future work

- Generalize the sufficiency proof to the constrained case
- Develop invex decompositions or apply invexity concepts to existing decompositions
- Invexity on graphs: what does invexity of a problem defined on a subgraph imply for the entire problem?
- Further study the interplay between invexity and integrality
- Implement and test these methods in SCIP

Thank you!