Evolutionary Bilevel Optimization

Kalyanmoy Deb
Koenig Endowed Chair Professor
Dept. of Electrical and Computer Engineering
Michigan State University, East Lansing, USA
kdeb@egr.msu.edu

Ankur Sinha
Post-Doctoral Researcher
Aalto University School of Business
Helsinki, Finland
ankur.sinha@aalto.fi

GECCO'14, July 12–16, 2014, Vancouver, BC, Canada.
ACM 978-1-4503-2662-9/14/07.

Instructors

- **Kalyanmoy Deb** is Koenig Endowed Chair Professor of Electrical and Computer Engineering at Michigan State University, East Lansing. His main research interest is in Evolutionary Multi-Criterion Optimization. His NSGA-II algorithm has more than 10,000 Google Scholar citations. His Computational Optimization and Innovation (COIN) laboratory develops original research and applications in Evolutionary Optimization. He has published 365 papers, 2 text books and 18 edited books. More of research papers can be found in http://www.egr.msu.edu/~kdeb.

- **Ankur Sinha** is a researcher at Aalto University School of Business, Helsinki, Finland. His research interests include Evolutionary Multi-Objective Optimization, Multi-Criteria Decision Making, Bilevel optimization and Statistical Language Processing. He completed his dissertation at the Aalto University School of Business in the year 2011, where he received the dissertation of the year award. More information about his research can be found at https://people.aalto.fi/ankur_sinha. He also maintains a website on Evolutionary Bilevel Optimization that can be accessed at http://www.bilevel.org.

Outline of Tutorial

- What is Bilevel Optimization?
- Difference from Single and Multi-objective optimization
- How practical are they?
- Evolutionary Bilevel Optimization (EBO)
- Past EBOs
- Recent advancements in EBO
- Conclusions and EBO Repository

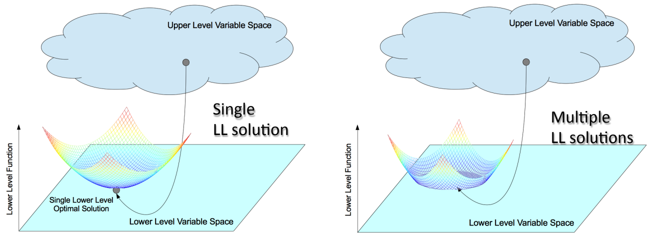
What is Bilevel Optimization?

- Two levels of optimization tasks
  - Upper level: (x_u, x_l)
  - Lower level: (x_l)
- An upper level feasible solution must be an optimal lower level solution

$$\begin{array}{rl} \mathrm{Min}_{(\mathbf{x}_u,\mathbf{x}_l)} & F(\mathbf{x}_u,\mathbf{x}_l), \\ \mathrm{st} & \mathbf{x}_l \in \mathrm{argmin}_{(\mathbf{x}_l)} \left\{ \begin{array}{c} f(\mathbf{x}_u,\mathbf{x}_l) \\ \mathbf{g}(\mathbf{x}_u,\mathbf{x}_l) \geq \mathbf{0}, \mathbf{h}(\mathbf{x}_u,\mathbf{x}_l) = \mathbf{0} \end{array} \right\}, \\ & \mathbf{G}(\mathbf{x}_u,\mathbf{x}_l) \geq \mathbf{0}, \mathbf{H}(\mathbf{x}_u,\mathbf{x}_l) = \mathbf{0}, \\ & (\mathbf{x}_u)_{min} \leq \mathbf{x}_u \leq (\mathbf{x}_u)_{max}, (\mathbf{x}_l)_{min} \leq \mathbf{x}_l \leq (\mathbf{x}_l)_{max} \end{array}$$

An Illustration

- Lower level solution x_l can be a singleton or multi-valued
- The best combination is preferred at upper level based on upper level function value



Multi-level Optimization: A Generic Optimization Problem

- Multi-level (L levels) optimization
  - Two (L=2) or more levels of optimization
  - Ideally, nested optimization
- Usual single, multi- and many-objective optimization problems
  - Special cases (L=1) of L-level optimization
- Bilevel: A more generic optimization concept than single-level optimization

---

Similarities with Constrained Single-Objective Optimization

- A single-objective optimization problem:

$$\begin{array}{lll} \text{Minimize} & f(\mathbf{x}) & \\ \text{Subject to} & h_k(\mathbf{x}) = 0, & \forall k \\ & g_j(\mathbf{x}) \geq 0, & \forall j \end{array}$$

- Equality constraint: $x_l = \Psi(\mathbf{x} \backslash x_l)$
  - Usually, a root-finding problem
  - A solution **x** is feasible, only if it satisfies all constraints
- In EBO, LL problem is an optimization problem
  - A solution (x_u, x_l) is not feasible, unless x_l is a solution to the LL optimization problem

---

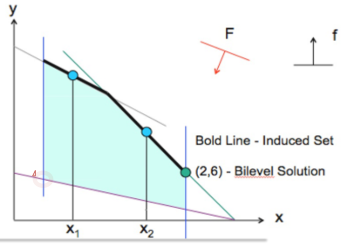
A Mathematical Bilevel Optimization Problem

$\text{Min}_{(x,y)}\ 3y + x$

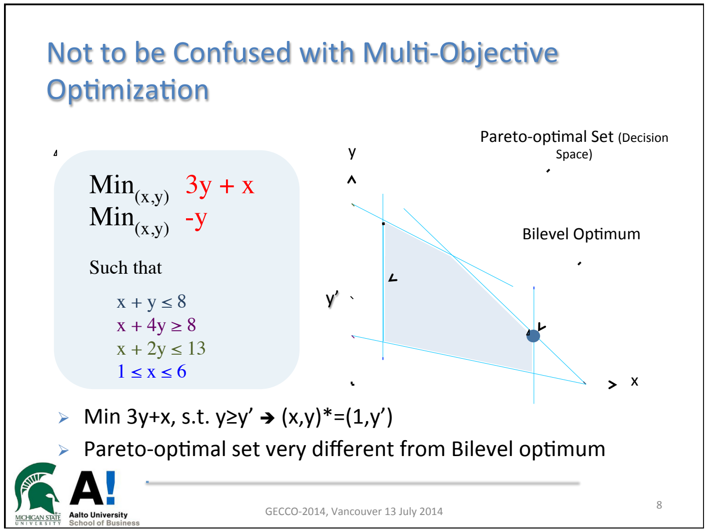
Such that

$y \in \text{argmin}_{(y)} \left\{ \begin{array}{l} -y \\ \text{Such that} \\ \quad x + y \leq 8,\ x + 4y \geq 8, \\ \quad x + 2y \leq 13,\ 1 \leq x \leq 6 \end{array} \right\}$

- x_u=x, x_l=y
- At LL, for x, maximize y
- Bold line is solution set for LL
- Min 3y+x for bold line
- The solution is (2,6)



Single-obj optimum



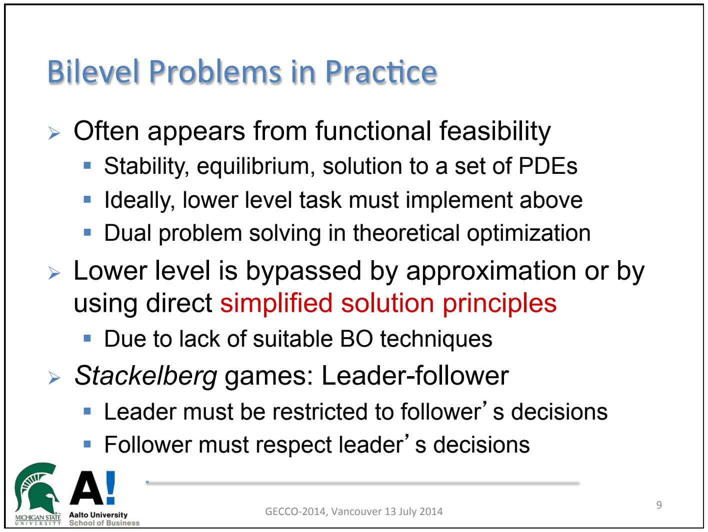
Not to be Confused with Multi-Objective Optimization

$\text{Min}_{(x,y)}$ 3y + x
$\text{Min}_{(x,y)}$ -y

Such that

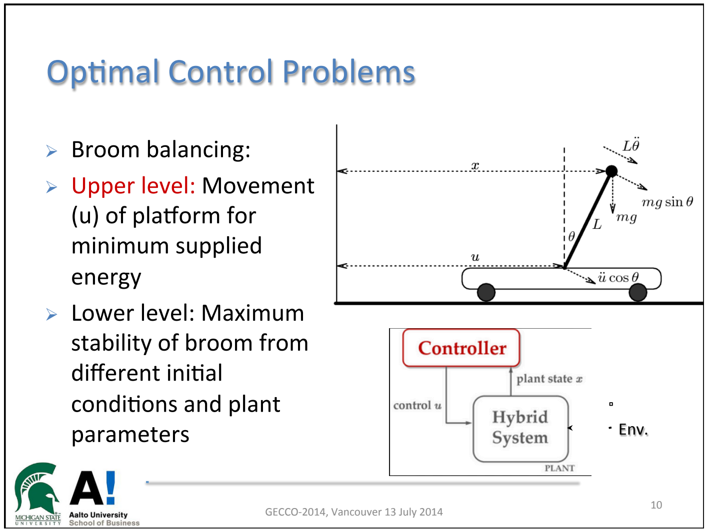
$x + y \leq 8$
$x + 4y \geq 8$
$x + 2y \leq 13$
$1 \leq x \leq 6$

- Min 3y+x, s.t. y≥y' ➔ (x,y)*=(1,y')
- Pareto-optimal set very different from Bilevel optimum



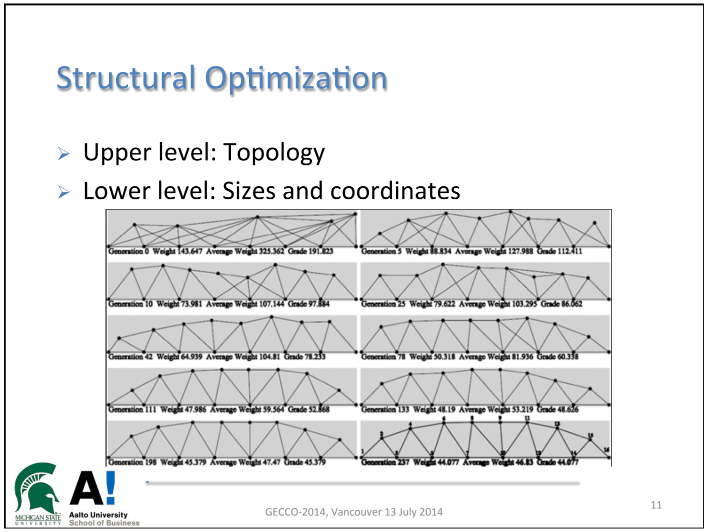
Bilevel Problems in Practice

- Often appears from functional feasibility
  - Stability, equilibrium, solution to a set of PDEs
  - Ideally, lower level task must implement above
  - Dual problem solving in theoretical optimization
- Lower level is bypassed by approximation or by using direct simplified solution principles
  - Due to lack of suitable BO techniques
- *Stackelberg* games: Leader-follower
  - Leader must be restricted to follower' s decisions
  - Follower must respect leader' s decisions



Optimal Control Problems

- Broom balancing:
- Upper level: Movement (u) of platform for minimum supplied energy
- Lower level: Maximum stability of broom from different initial conditions and plant parameters



Structural Optimization

- Upper level: Topology
- Lower level: Sizes and coordinates

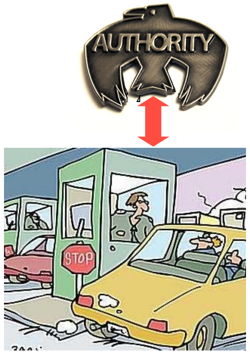
Toll Setting Problem

Authority's problem:

- Authority responsible for highway system wants to maximize its revenues earned from toll
- The authority has to solve the highway users optimization problem for all the possible tolls

Highway users' problem:

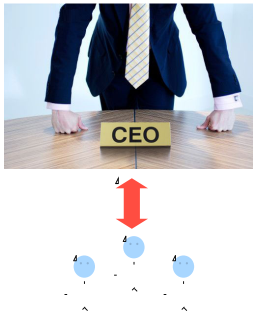
- For any toll chosen by the authority, highway users try to minimize their own travel costs
- A high toll will deter users to take the highway, lowering the revenues



Brotcorne et al. (2001)

Seller-Buyer Strategies

- An owner of a company dictates the selling price and supply. She/He wants to maximize profit.
- The buyers look at the product quality, pricing and various other options available to maximize their utility
- Mixed integer programs on similar lines have been formulated by Heliporn et al. (2010)



Heilporn et al. (2010)

Taxation Strategy

- Recently, there has been a controversy in Finland for gold mining in the Kuusamo region in Finland
- The region is a famous tourist resort endowed with immense natural beauty
- For any taxation strategy by the government, the mining company optimizes its own profits

Leader: Government Maximize revenue from taxes, Minimize Pollution

Follower: Mining Company Maximize Profit

Sinha, et al. (2013)

Properties of Bilevel Problems

- Bilevel problems are typically non-convex, disconnected and strongly NP-hard
- Solving an optimization problem produces a one or more feasible solutions
- Multiple global solutions at lower level can induce additional challenges
- Two levels can be cooperating or conflicting

Solution Methodologies

- KKT conditions of the lower level problem are used as constraints (Herskovits et al. 2000)
  - Lagrange multipliers increase the number of decision variables
  - Constrained search space
  - Applicable to differentiable problems only
- Another common approach: Nested optimization
  - For every x_u, lower level problem is solved completely
  - Computationally very expensive
- Discretization of the lower level problem
  - The best solution obtained from discrete set for a given x_u is used as a feasible member at upper level

---

Solution Methodologies (cont.)

- Penalty based approaches
  - Special forms of penalty functions have been used
  - Lower level is usually required to be convex
  - Penalty function may require differentiability
- Branch and Bound techniques (Bard et al. 1982)
  - Used KKT conditions
  - Handled linear problems
  - Converted the problems into variable separable form
  - Utilized the branch and bound approach
- Taking an approximation of the lower level optimization problem such that its optimum is readily available
  - The optimal solutions from lower level might not be accurate

---

Solution Methodologies (cont.)

- Evolutionary algorithms have also been used for bilevel optimization
- Most of the methods are nested strategies
- Mathieu et al. (1994): LP for lower level and GA for upper level
- Yin (2000): Frank Wolfe Algorithm for lower level
- Oduguwa and Roy (2005): Proposed a co-evolutionary approach
- Wang et al. (2005):
  - Solved bilevel problems using a constrained handling scheme in EA
  - Method is computationally expensive, but successfully handles a number of test problems
- Li et al. (2006): Nested strategy using PSO
- EA researchers have also tried replacing the lower level problems using KKT (Wang et al. (2008), Li et al. (2007))

Why Use Evolutionary Algorithms?

- First, no *implementable* mathematical optimality conditions exist (Dempe, Dutta, Mordokhovich, 2007)
  - LL problem is replaced with KKT conditions and constraint qualification (CQ) conditions of LL
  - UL problem requires KKT of LL-KKT conditions, but handling LL-CQ conditions in UL-KKT becomes difficult
  - Involves second-order differentials
- Moreover, classical numerical optimization methods require various simplifying assumptions like continuity, differentiability and convexity
- Most real-world applications do not follow these assumptions
- EA's flexible operators, direct use of objectives and population approach help solve BO problems better

Niche of Evolutionary Methods

- Usually, LL solutions are multi-modal
- Usually, BO problems are multi-objective BO
    - Both problems require to find and maintain multiple optimal solutions
    - EAs are known to be good for these cases
- Computationally faster methods possible through meta-modeling etc.
- Other complexities (robustness, parallel implementation, fixed budget) can be handled efficiently

---

Bilevel Test Problems

- Controlled difficulty in convergence at upper and lower levels
- Controlled difficulty caused by interaction of two levels
- Multiple global solutions at the lower level for any given set of upper level variables
- Clear identification of relationships between lower level optimal solutions and upper level variables
- Scalability to any number of decision variables at upper and lower levels
- Constraints (preferably scalable) at upper and lower levels
- Possibility to have conflict or cooperation at the two levels
- The optimal solution of the bilevel optimization is known

SMD Test Problem Framework
(Sinha, Malo & Deb, 2014)

The objectives and variables on both levels are decomposed as follows:

$$F(\mathbf{x}_u, \mathbf{x}_l) = F_1(\mathbf{x}_{u1}) + F_2(\mathbf{x}_{l1}) + F_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})$$
$$f(\mathbf{x}_u, \mathbf{x}_l) = f_1(\mathbf{x}_{u1}, \mathbf{x}_{u2}) + f_2(\mathbf{x}_{l1}) + f_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})$$
where
$$\mathbf{x}_u = (\mathbf{x}_{u1}, \mathbf{x}_{u2}) \quad \text{and} \quad \mathbf{x}_l = (\mathbf{x}_{l1}, \mathbf{x}_{l2})$$

Roles of Variables

Panel A: Decomposition of decision variables

Upper-level variables		Lower-level variables	
Vector	Purpose	Vector	Purpose
$\mathbf{x}_{u1}$	Complexity on upper-level	$\mathbf{x}_{l1}$	Complexity on lower-level
$\mathbf{x}_{u2}$	Interaction with lower-level	$\mathbf{x}_{l2}$	Interaction with upper-level

Panel B: Decomposition of objective functions

Upper-level objective function		Lower-level objective function	
Component	Purpose	Component	Purpose
$F_1(\mathbf{x}_{u1})$	Difficulty in convergence	$f_1(\mathbf{x}_{u1}, \mathbf{x}_{u2})$	Functional dependence
$F_2(\mathbf{x}_{l1})$	Conflict / co-operation	$f_2(\mathbf{x}_{l1})$	Difficulty in convergence
$F_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})$	Difficulty in interaction	$f_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})$	Difficulty in interaction

Controlling Difficulty for Convergence

- Convergence difficulties can be induced via following routes:
- Dedicated components: F_1 (upper) and f_2 (lower)
- Example:

$$F(\mathbf{x}_u, \mathbf{x}_l) = \boxed{F_1(\mathbf{x}_{u1})} + F_2(\mathbf{x}_{l1}) + F_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})$$

Quadratic

$$f(\mathbf{x}_u, \mathbf{x}_l) = f_1(\mathbf{x}_{u1}, \mathbf{x}_{u2}) + \boxed{f_2(\mathbf{x}_{l1})} + f_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})$$

Multi-modal

Controlling Difficulty in Interactions

- Interaction between variables x_{u2} and x_{l2} could be chosen as follows:
  - Dedicated components: F_3 and f_3
- Example:

$$F(\mathbf{x}_u, \mathbf{x}_l) = F_1(\mathbf{x}_{u1}) + F_2(\mathbf{x}_{l1}) + \boxed{F_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})}$$

$$\sum_{i=1}^{r}(x_{u2}^i)^2 + \sum_{i=1}^{r}((x_{u2}^i)^2 - \tan x_{l2}^i)^2$$

$$f(\mathbf{x}_u, \mathbf{x}_l) = f_1(\mathbf{x}_{u1}, \mathbf{x}_{u2}) + f_2(\mathbf{x}_{l1}) + \boxed{f_3(\mathbf{x}_{u2}, \mathbf{x}_{l2})}$$

$$\sum_{i=1}^{r}((x_{u2}^i)^2 - \tan x_{l2}^i)^2$$

Difficulty due to Conflict/Co-operation

- Dedicated components: F_2 and f_2 or F_3 and f_3 may be used to induce conflict or cooperation
- Examples:
  - Cooperative interaction = Improvement in lower-level improves upper-level (e.g. $F_2 = f_2$)
  - Conflicting interaction = Improvement in lower-level worsens upper-level (e.g. $F_2 = -f_2$)
  - Mixed interaction = Both cooperation and conflict (e.g. $F_2 = f_2$ and $F_3 = \Sigma_i (x_{u2}^i)^2 - f_3$

Controlled Multimodality

- Obtain multiple lower-level optima for every upper level solution:
  - Component used: f_2
- Example: Multimodality at lower-level

$$f_1(\mathbf{x}_{u1}, \mathbf{x}_{u2}) = (x_{u1}^1)^2 + (x_{u1}^1)^2 + (x_{u2}^1)^2 + (x_{u2}^2)^2$$
$$f_2(\mathbf{x}_{l1}) = \boxed{(x_{l1}^1 - x_{l1}^2)^2}$$
$$f_3(\mathbf{x}_{u2}, \mathbf{x}_{l2}) = (x_{u2}^1 - x_{l2}^1)^2 + (x_{u2}^2 - x_{l2}^2)^2$$

Induces multiple solutions: $x^1_{l1} = x^2_{l1}$

$$F_1(\mathbf{x}_{u1}) = (x_{u1}^1)^2 + (x_{u1}^1)^2$$
$$F_2(\mathbf{x}_{l1}) = \boxed{(x_{l1}^1)^2 + (x_{l1}^2)^2}$$
$$F_3(\mathbf{x}_{u2}, \mathbf{x}_{l2}) = (x_{u2}^1 - x_{l2}^2)^2 + (x_{u2}^2 - x_{l2}^2)^2$$

Gives best UL solution: $x^1_{l1} = x^2_{l1}=0$

Difficulty due to Constraints

Constraints are included at both levels with one or more of the following properties:

- Constraints exist, but are not active at the optimum
- A subset of constraints, or all the constraints are active at the optimum
- Upper level constraints are functions of only upper level variables, and lower level constraints are functions of only lower level variables
- Upper level constraints are functions of upper as well as lower level variables, and lower level constraints are also functions of upper as well as lower level variables
- Lower level constraints lead to multiple global solutions at the lower level
- Constraints are scalable at both levels

Test Problems: SMD1 – SMD3

- SMD1:
  - Interaction: Cooperative
  - Lower level: Convex (w.r.t. lower-level variables)
  - Upper level: Convex (induced space)
- SMD 2:
  - Interaction: Conflict
  - Lower level: Convex (w.r.t. lower-level variables)
  - Upper level: Convex (induced space)
- SMD 3:
  - Interaction: Cooperative
  - Lower level: Multimodality using Rastrigin's function
  - Upper level: Convex (induced space)

Test Problems: SMD4 – SMD6

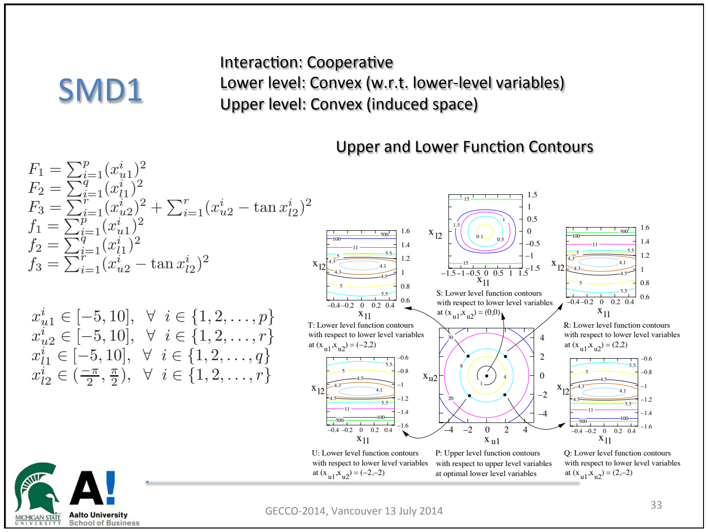
- SMD 4:
  - Interaction: Conflict
  - Lower level: Multimodality using Rastrigin's function
  - Upper level: Convex (Induced Space)
- SMD 5:
  - Interaction: Conflict
  - Lower level: Complexity with Rosenbrock's function
  - Upper level: Convex (Induced Space)
- SMD 6:
  - Interaction: Conflict
  - Lower level: Infinitely many global solutions for any given x_u
  - Upper level: Convex (Induced Space)

Test Problems: SMD7 – SMD9

- SMD 7:
  - Interaction: Conflict
  - Lower level: Convex (w.r.t. lower-level variables)
  - Upper level: Multimodality
- SMD 8:
  - Interaction: Conflict
  - Lower level: Complexity with banana function
  - Upper level: Multimodality
- SMD 9:
  - Interaction: Conflict
  - Lower level: Non-scalable constraints
  - Upper level: Non-scalable constraints

Test Problems: SMD10 – SMD12

- **SMD 10:**
  - Interaction: Conflict
  - Lower level: Scalable constraints
  - Upper level: Scalable constraints
- **SMD 11:**
  - Interaction: Conflict
  - Lower level: Non-scalable constraints, multiple global solutions
  - Upper level: Scalable constraints
- **SMD 12:**
  - Interaction: Conflict
  - Lower level: Scalable constraints, multiple global solutions
  - Upper level: Scalable constraints



SMD1

Interaction: Cooperative
Lower level: Convex (w.r.t. lower-level variables)
Upper level: Convex (induced space)

$$
\begin{aligned}
F_1 &= \textstyle\sum_{i=1}^{p}(x_{u1}^i)^2\\
F_2 &= \textstyle\sum_{i=1}^{q}(x_{l1}^i)^2\\
F_3 &= \textstyle\sum_{i=1}^{r}(x_{u2}^i)^2 + \sum_{i=1}^{r}(x_{u2}^i - \tan x_{l2}^i)^2\\
f_1 &= \textstyle\sum_{i=1}^{p}(x_{u1}^i)^2\\
f_2 &= \textstyle\sum_{i=1}^{q}(x_{l1}^i)^2\\
f_3 &= \textstyle\sum_{i=1}^{r}(x_{u2}^i - \tan x_{l2}^i)^2
\end{aligned}
$$

$$
\begin{aligned}
&x_{u1}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, p\}\\
&x_{u2}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, r\}\\
&x_{l1}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, q\}\\
&x_{l2}^i \in (\tfrac{-\pi}{2}, \tfrac{\pi}{2}), \quad \forall\ i \in \{1, 2, \dots, r\}
\end{aligned}
$$

Upper and Lower Function Contours

S: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (0,0)$

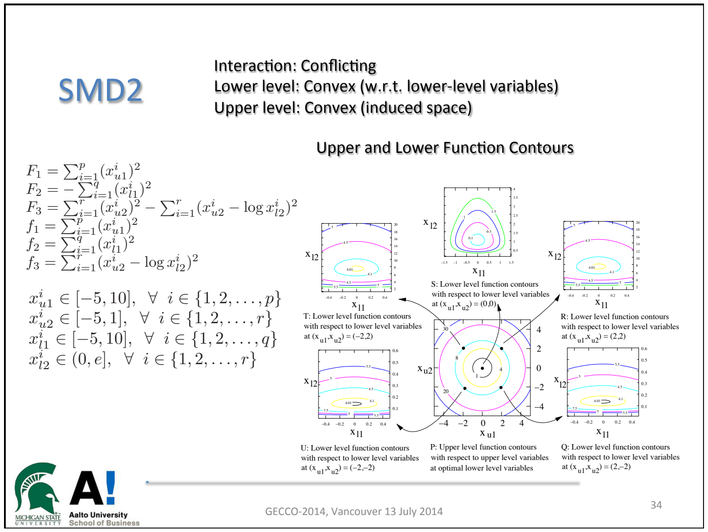
T: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,2)$

R: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,2)$

U: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,-2)$

P: Upper level function contours with respect to upper level variables at optimal lower level variables

Q: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,-2)$



SMD2

Interaction: Conflicting
Lower level: Convex (w.r.t. lower-level variables)
Upper level: Convex (induced space)

$$
\begin{aligned}
F_1 &= \textstyle\sum_{i=1}^{p}(x_{u1}^i)^2\\
F_2 &= -\textstyle\sum_{i=1}^{q}(x_{l1}^i)^2\\
F_3 &= \textstyle\sum_{i=1}^{r}(x_{u2}^i)^2 - \sum_{i=1}^{r}(x_{u2}^i - \log x_{l2}^i)^2\\
f_1 &= \textstyle\sum_{i=1}^{p}(x_{u1}^i)^2\\
f_2 &= \textstyle\sum_{i=1}^{q}(x_{l1}^i)^2\\
f_3 &= \textstyle\sum_{i=1}^{r}(x_{u2}^i - \log x_{l2}^i)^2
\end{aligned}
$$

$$
\begin{aligned}
&x_{u1}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, p\}\\
&x_{u2}^i \in [-5, 1], \quad \forall\ i \in \{1, 2, \dots, r\}\\
&x_{l1}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, q\}\\
&x_{l2}^i \in (0, e], \quad \forall\ i \in \{1, 2, \dots, r\}
\end{aligned}
$$

Upper and Lower Function Contours

S: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (0,0)$

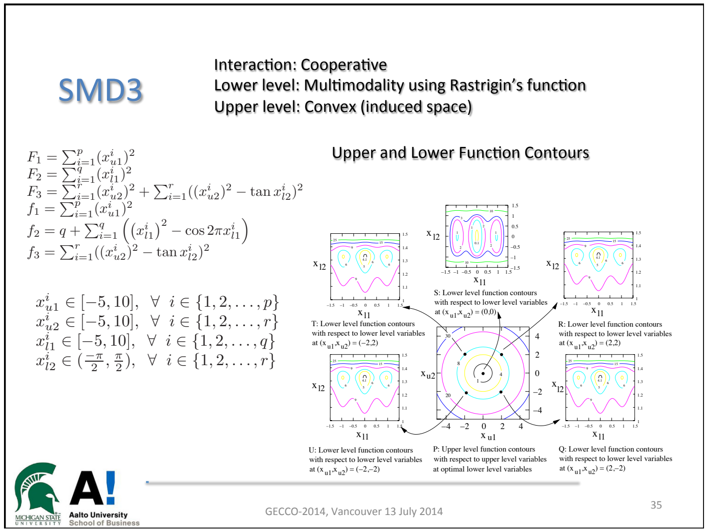
T: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,2)$

R: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,2)$

U: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,-2)$

P: Upper level function contours with respect to upper level variables at optimal lower level variables

Q: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,-2)$



SMD3

Interaction: Cooperative
Lower level: Multimodality using Rastrigin's function
Upper level: Convex (induced space)

$$
\begin{aligned}
F_1 &= \textstyle\sum_{i=1}^{p}(x_{u1}^i)^2\\
F_2 &= \textstyle\sum_{i=1}^{q}(x_{l1}^i)^2\\
F_3 &= \textstyle\sum_{i=1}^{r}(x_{u2}^i)^2 + \sum_{i=1}^{r}((x_{u2}^i)^2 - \tan x_{l2}^i)^2\\
f_1 &= \textstyle\sum_{i=1}^{p}(x_{u1}^i)^2\\
f_2 &= q + \textstyle\sum_{i=1}^{q}\left((x_{l1}^i)^2 - \cos 2\pi x_{l1}^i\right)\\
f_3 &= \textstyle\sum_{i=1}^{r}((x_{u2}^i)^2 - \tan x_{l2}^i)^2
\end{aligned}
$$

$$
\begin{aligned}
&x_{u1}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, p\}\\
&x_{u2}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, r\}\\
&x_{l1}^i \in [-5, 10], \quad \forall\ i \in \{1, 2, \dots, q\}\\
&x_{l2}^i \in (\tfrac{-\pi}{2}, \tfrac{\pi}{2}), \quad \forall\ i \in \{1, 2, \dots, r\}
\end{aligned}
$$

Upper and Lower Function Contours

S: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (0,0)$

T: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,2)$

R: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,2)$

U: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,-2)$

P: Upper level function contours with respect to upper level variables at optimal lower level variables

Q: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,-2)$

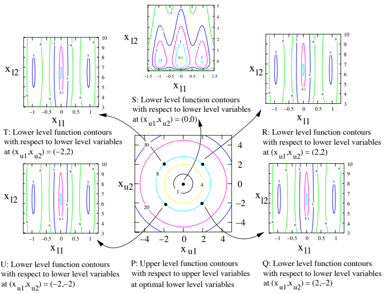
SMD4

Interaction: Conflicting
Lower level: Multimodality using Rastrigin's function
Upper level: Convex (Induced Space)

$$F_1 = \sum_{i=1}^{p}(x_{u1}^i)^2$$
$$F_2 = -\sum_{i=1}^{q}(x_{l1}^i)^2$$
$$F_3 = \sum_{i=1}^{r}(x_{u2}^i)^2 - \sum_{i=1}^{r}(|x_{u2}^i| - \log(1+x_{l2}^i))^2$$
$$f_1 = \sum_{i=1}^{p}(x_{u1}^i)^2$$
$$f_2 = q + \sum_{i=1}^{q}\left(\left(x_{l1}^i\right)^2 - \cos 2\pi x_{l1}^i\right)$$
$$f_3 = \sum_{i=1}^{r}(|x_{u2}^i| - \log(1+x_{l2}^i))^2$$

$$x_{u1}^i \in [-5,10],\ \ \forall\ \ i \in \{1,2,\ldots,p\}$$
$$x_{u2}^i \in [-1,1],\ \ \forall\ \ i \in \{1,2,\ldots,r\}$$
$$x_{l1}^i \in [-5,10],\ \ \forall\ \ i \in \{1,2,\ldots,q\}$$
$$x_{l2}^i \in [0,e],\ \ \forall\ \ i \in \{1,2,\ldots,r\}$$

Upper and Lower Function Contours



S: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (0,0)$

T: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,2)$

R: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,2)$

U: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (-2,-2)$

P: Upper level function contours with respect to upper level variables at optimal lower level variables

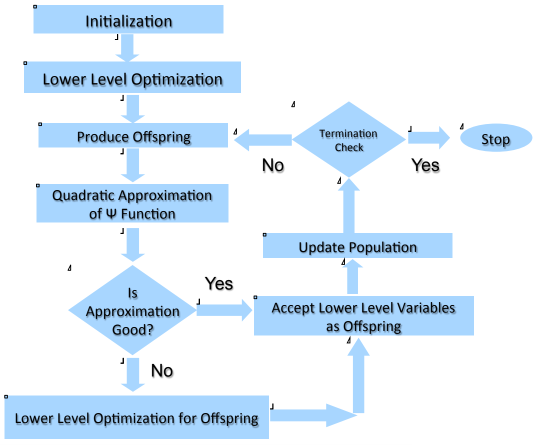
Q: Lower level function contours with respect to lower level variables at $(x_{u1},x_{u2}) = (2,-2)$

Efficient Evolutionary Bilevel Optimization Algorithm

- Nested algorithm is expensive
- Train a meta-model for optimal lower level variable vector and upper level variable vector mapping
- Quadratic approximation of the inducible region
  - BLEAQ (Sinha, Malo and Deb, 2013)
- Use meta-model until possible, else solve LL optimization problem

---

BLEAQ


Initialization

Lower Level Optimization

Produce Offspring

Quadratic Approximation of Ψ Function

Is Approximation Good?

Yes

No

Accept Lower Level Variables as Offspring

Lower Level Optimization for Offspring

Update Population

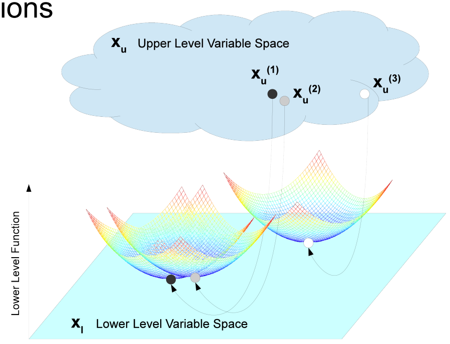
Termination Check

No

Yes

Stop

Assumption: Close upper-level members are expected to have close lower-level optimal solutions



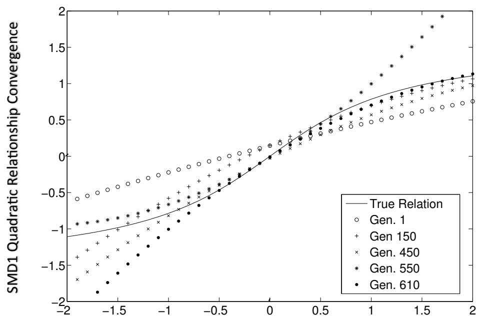
BLEAQ Results

- 10-variable SMD test problems using BLEAQ
- Comparison performed against nested evolutionary approach
  - Number of Runs: 21
  - Savings: Ratio of FE required by nested approach against BLEAQ

Pr. No.	Best Func. Evals.		Median Func. Evals.		Worst Func. Evals.	
	LL	UL	LL (Savings)	UL (Savings)	LL	UL
SDM1	99315	610	110716 (14.71)	740 (3.34)	170808	1490
SDM2	70032	376	91023 (16.49)	614 (3.65)	125851	1182
SDM3	110701	620	125546 (11.25)	900 (2.48)	137128	1094
SDM4	61326	410	81434 (13.59)	720 (2.27)	101438	1050
SDM5	102868	330	126371 (15.41)	632 (4.55)	168401	1050
SDM6	95687	734	118456 (14.12)	952 (3.25)	150124	1410

Aalto University
School of Business

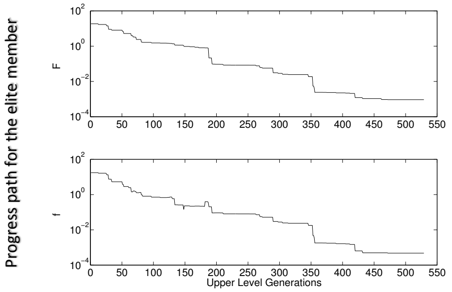
Results on SMD1



Quadratic approximation at optima (0,0) improves with increasing generations

Aalto University
School of Business

Convergence Plots on SMD1



Upper Level Function Value

Lower Level Function Value

Convergence Plot for SMD1

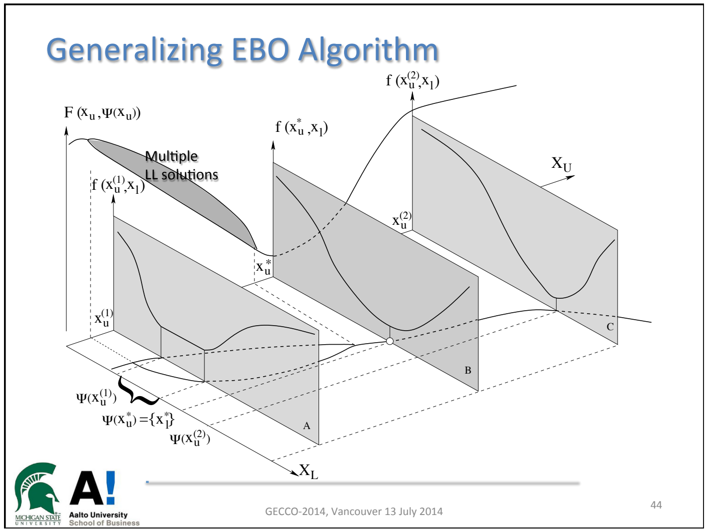
Aalto University
School of Business

Results on Ten Standard Test Problems

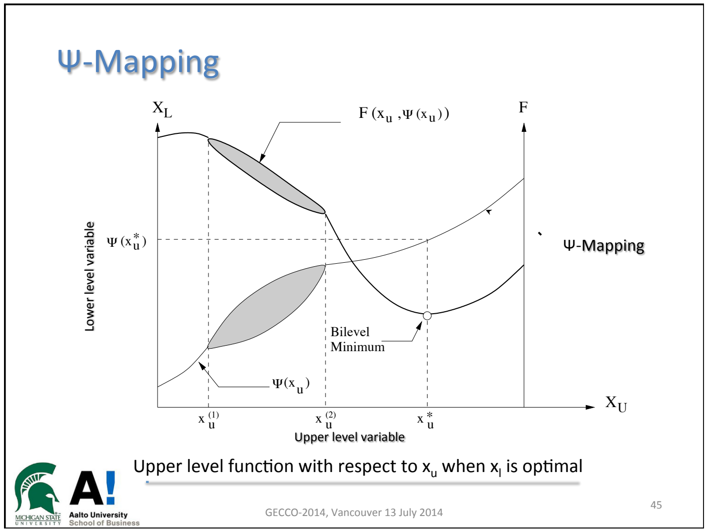
- Mean (total function evaluations) results for ten bilevel test problems
- Comparison against the evolutionary algorithm of Wang et al. (2005,2011)
- BLEAQ is almost an order of magnitude better

Pr. No.				
	BLEAQ	WJL (BLEAQ Savings)	WLD (BLEAQ Savings)	Nested (BLEAQ Savings)
TP1	14810	85499 (5.77)	86067 (5.81)	161204 (10.88)
TP2	14771	256227 (17.35)	171346 (11.60)	242624 (16.43)
TP3	4376	92526 (21.15)	95851 (21.91)	120728 (27.59)
TP4	15285	291817 (19.09)	211937 (13.87)	272843 (17.85)
TP5	15403	77302 (5.02)	69471 (4.51)	148148 (9.62)
TP6	17218	163701 (9.51)	65942 (3.83)	181271 (10.53)
TP7	272971	1074742 (3.94)	944105 (3.46)	864474 (3.17)
TP8	12065	213522 (17.70)	182121 (15.09)	318575 (26.40)
TP9	93517	- -	352883 (3.77)	665244 (7.11)
TP10	100357	- -	463752 (4.62)	599434 (5.97)

Aalto University
School of Business

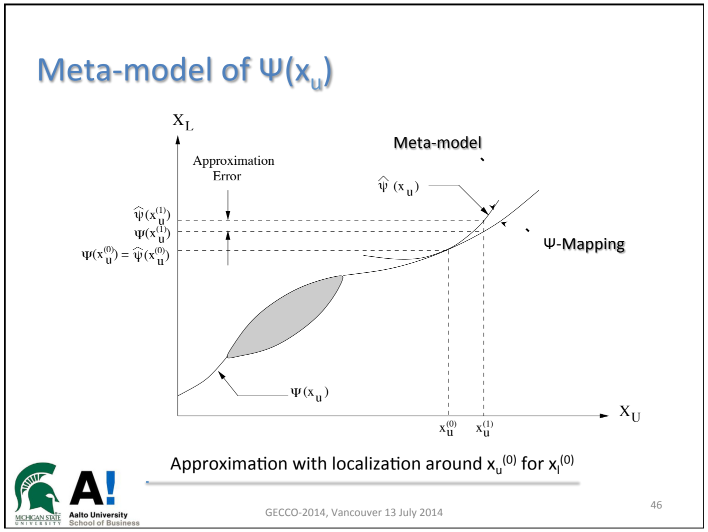


Generalizing EBO Algorithm



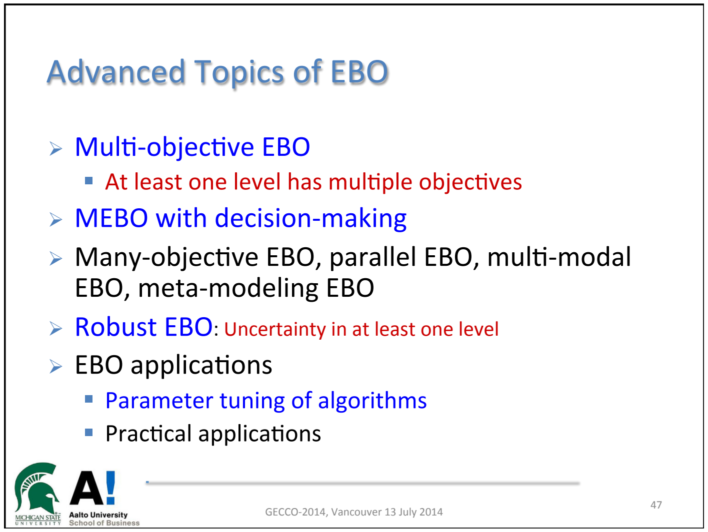
Ψ-Mapping

Upper level function with respect to x_u when x_l is optimal



Meta-model of $\Psi(x_u)$

Approximation with localization around $x_u^{(0)}$ for $x_l^{(0)}$



Advanced Topics of EBO

- Multi-objective EBO
  - At least one level has multiple objectives
- MEBO with decision-making
- Many-objective EBO, parallel EBO, multi-modal EBO, meta-modeling EBO
- Robust EBO: Uncertainty in at least one level
- EBO applications
  - Parameter tuning of algorithms
  - Practical applications

Advanced EBO Ideas (cont.)

- Highly constrained EBO
- Mixed-integer EBO
- EBO with a fixed budget at LL and UL
- EBO versus EO for *F=f*
- Error propagation from lower level to upper level
  - Theoretical convergence studies
- Evolutionary Multi-Level Optimization (EMLO)

Past Studies in Multi-Objective Evolutionary Bilevel Optimization (MEBO)

- Plenty of studies in single-objective bilevel optimization (BO), not much in multi-objective BO
- Optimality theory difficult
  - KKT conditions involve second derivative of lower level objectives and constraints
  - Dempe et al. (2006) developed KKT conditions
  - Impractical to implement (abstract terms)
- Fliege and Vincent (2006): BP → 4-obj MOP
  - Involves derivatives and unclear of extensions to multi-objective and higher level problems

---

Past MEBO Algorithms

- A common approach: Nested optimization
  - For every x_u, lower level problem is solved completely
  - Computationally expensive, extension to multi-objective BO impractical
- KKT conditions of lower level problem used as constraints (Herskovits et al., 2000)
  - Lagrange multiplers increase number of of decision variables
  - Constrained search space
  - Applicable to differentiable problems only
- Intelligent exhaustive upper level search
  - Subdivision approach on upper level and numerical optimization in lower level (Eichfelder, 2007)

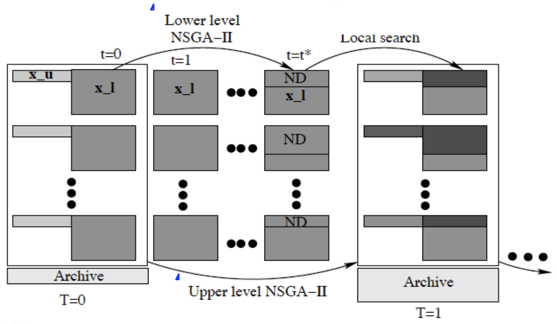
Single Versus Multi-Objective BO Problems

- Single-objective BO
  - Scalar F and scalar f
  - Usually one target solution x_u* and x_l*
- Multi-objective BO
  - Usually multiple solutions x_l* for each x_u*
  - Find and maintain many solutions for each x_u
  - Not an easy matter

BLEMO
(Deb and Sinha, 2009)

- Both levels use NSGA-II iteratively

- Population structure
- Lower and upper level NSGA-II
- Archiving



BLEMO Features: Different from a Pure Nested Procedure

- A population of x_u is initialized
  - Lower level NSGA-II attempts to move towards their PO fronts
- Based on intermediate information, x_u population is updated to move towards feasible and PO region of upper level problem
  - Again, lower level NSGA-II attempts to move their own PO fronts
  - And so on …
- Towards the end x_u does not change much, but x_l gets towards Pareto-optimality with iterations

---

Test Problem

(Eichfelder, 2007)

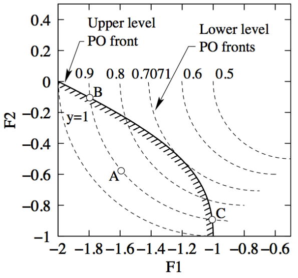
$$\min \mathbf{F}(\mathbf{x}) = \left\{ \begin{array}{c} x_1 - y \\ x_2 \end{array} \right\},$$

y upper, (x1,x2) lower

$$\text{st } (x_1, x_2) \in \text{argmin}_{(x_1,x_2)} \left\{ \begin{array}{c} \mathbf{f}(\mathbf{x}) = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \\ g_1(\mathbf{x}) = y^2 - x_1^2 - x_2^2 \geq 0 \end{array} \right\},$$

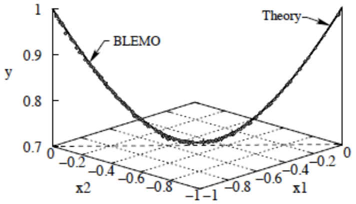
$$G_1(\mathbf{x}) = 1 + x_1 + x_2 \geq 0,$$
$$-1 \leq x_1, x_2 \leq 1, \quad 0 \leq y \leq 1.$$

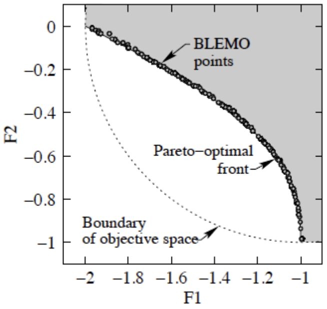
- Dashed lines are lower level PO fronts
  - Depends on y
- Upper level PO front on constraint G1
  - Maximum two solutions from each y
  - Not all y in upper PO front
- Solutions below PO front exist

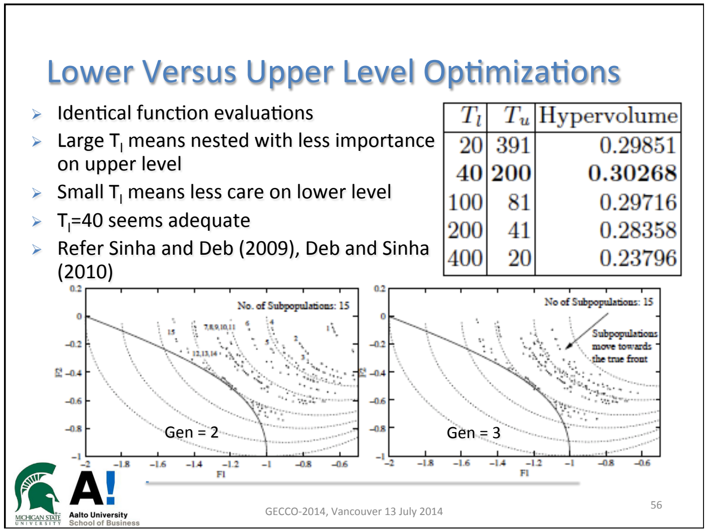


Results (Eichfelder, 2007)

- N_u=400, T_u=200
- N_l=40, T_l=40
- Near PO solutions found



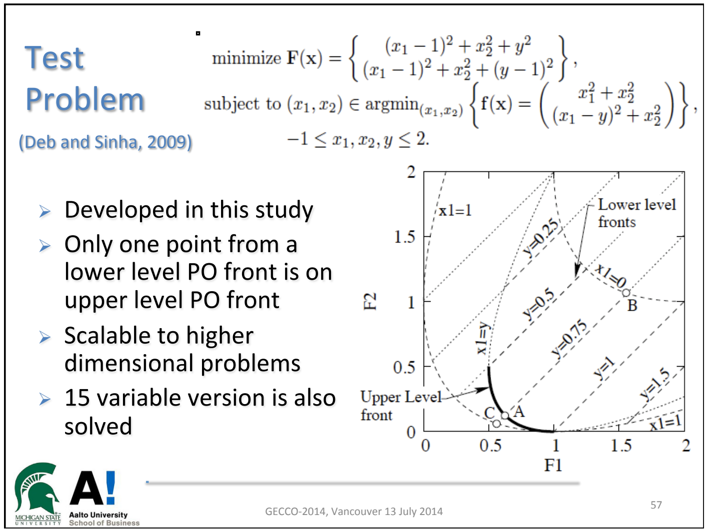




Lower Versus Upper Level Optimizations

- Identical function evaluations
- Large T_l means nested with less importance on upper level
- Small T_l means less care on lower level
- T_l=40 seems adequate
- Refer Sinha and Deb (2009), Deb and Sinha (2010)

T_l	T_u	Hypervolume
20	391	0.29851
40	**200**	**0.30268**
100	81	0.29716
200	41	0.28358
400	20	0.23796

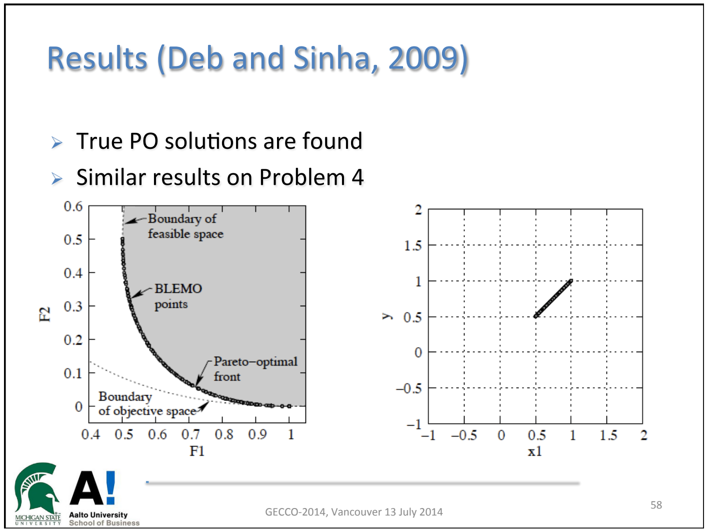


Test Problem

(Deb and Sinha, 2009)

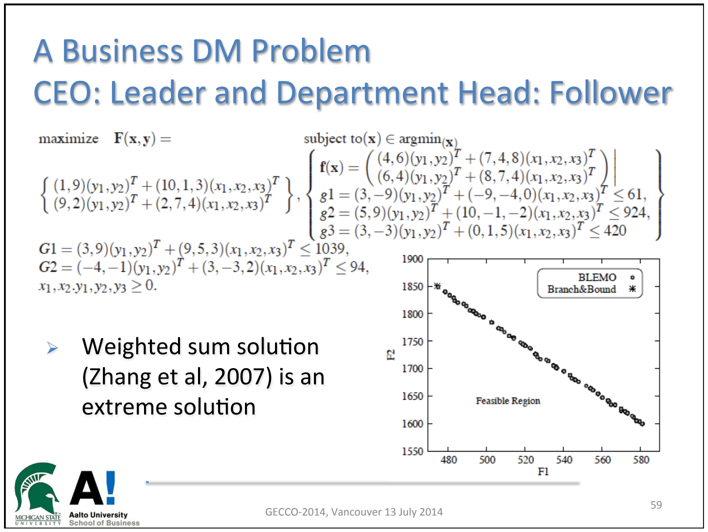
$$\text{minimize } \mathbf{F}(\mathbf{x}) = \left\{ \begin{array}{c} (x_1-1)^2 + x_2^2 + y^2 \\ (x_1-1)^2 + x_2^2 + (y-1)^2 \end{array} \right\},$$
$$\text{subject to } (x_1,x_2) \in \text{argmin}_{(x_1,x_2)} \left\{ \mathbf{f}(\mathbf{x}) = \left(\begin{array}{c} x_1^2 + x_2^2 \\ (x_1-y)^2 + x_2^2 \end{array} \right) \right\},$$
$$-1 \le x_1, x_2, y \le 2.$$

- Developed in this study
- Only one point from a lower level PO front is on upper level PO front
- Scalable to higher dimensional problems
- 15 variable version is also solved



Results (Deb and Sinha, 2009)

- True PO solutions are found
- Similar results on Problem 4



A Business DM Problem
CEO: Leader and Department Head: Follower

$$\text{maximize} \quad \mathbf{F}(\mathbf{x},\mathbf{y}) = \left\{ \begin{array}{l} (1,9)(y_1,y_2)^T + (10,1,3)(x_1,x_2,x_3)^T \\ (9,2)(y_1,y_2)^T + (2,7,4)(x_1,x_2,x_3)^T \end{array} \right\},$$
$$\text{subject to}(\mathbf{x}) \in \text{argmin}_{(\mathbf{x})} \left\{ \begin{array}{l} \mathbf{f}(\mathbf{x}) = \left(\begin{array}{l} (4,6)(y_1,y_2)^T + (7,4,8)(x_1,x_2,x_3)^T \\ (6,4)(y_1,y_2)^T + (8,7,4)(x_1,x_2,x_3)^T \end{array} \right) \Bigg| \\ g1 = (3,-9)(y_1,y_2)^T + (-9,-4,0)(x_1,x_2,x_3)^T \le 61, \\ g2 = (5,9)(y_1,y_2)^T + (10,-1,-2)(x_1,x_2,x_3)^T \le 924, \\ g3 = (3,-3)(y_1,y_2)^T + (0,1,5)(x_1,x_2,x_3)^T \le 420 \end{array} \right\}$$
$$G1 = (3,9)(y_1,y_2)^T + (9,5,3)(x_1,x_2,x_3)^T \le 1039,$$
$$G2 = (-4,-1)(y_1,y_2)^T + (3,-3,2)(x_1,x_2,x_3)^T \le 94,$$
$$x_1,x_2.y_1,y_2,y_3 \ge 0.$$

- Weighted sum solution (Zhang et al, 2007) is an extreme solution

Test Problem (Deb and Sinha, 2009)

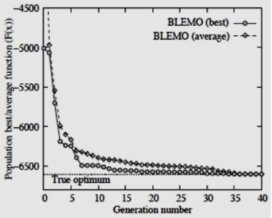
- $$\begin{aligned} &\text{minimize } F(\mathbf{x}) = -(200 - x_1 - x_2)(x_1 + x_3) - (160 - x_2 - x_4)(x_2 + x_4), \\ &\text{subject to} \\ &\quad \mathbf{x}_l \in \text{argmin}_{(\mathbf{x}_l)} \{\mathbf{f}(\mathbf{x}) = (x_1 - 4)^2 + (x_2 - 13)^2 + (x_3 - 35)^2 + (x_4 - 2)^2 | \\ &\qquad g_1(\mathbf{x}) = 0.4x_1 + 0.7x_2 \le y_1,\ g_2(\mathbf{x}) = 0.6x_1 + 0.3x_2 \le y_2, \\ &\qquad g_3(\mathbf{x}) = 0.4x_3 + 0.7x_4 \le y_3,\ g_4(\mathbf{x}) = 0.6x_3 + 0.3x_4 \le y_4\}, \\ &\quad G(\mathbf{x}) = y_1 + y_2 + y_3 + y_4 \le 40, \\ &\quad 0 \le y_1 \le 10,\ 0 \le y_2 \le 5,\ 0 \le y_3 \le 15, 0 \le y_4 \le 20, \\ &\quad 0 \le x_1 \le 20,\ 0 \le x_2 \le 20,\ 0 \le x_3 \le 40,\ 0 \le x_4 \le 40. \end{aligned}$$

(Colson, 2002)

- Known optimum:

$$\mathbf{x}_u^* = (7.36, 3.55, 11.64, 17.45)^T$$
$$\mathbf{x}_l^* = (0.91, 10, 29.09, 0)^T$$
$$F(\mathbf{x}^*) = -6600.0 \text{ and } f(\mathbf{x}^*) = 57.48.$$

- Identical solution obtained by BLEMO
- BLEMO scales to 1-obj problems as well

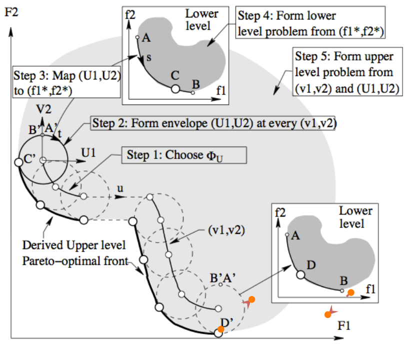


Extensions to This Study

- Difficult test problems suggested (Deb and Sinha, 2009, 2010)
  - Allows an adequate test to BP algorithms
- Self-adaptive and local search based BLEMO suggested (Deb and Sinha, 2010)
  - Much faster approach
  - No additional parameter
- Interaction with the decision maker at upper level leads to substantial savings in function evaluations (Sinha, 2011)

Test Problem Construction Principle

- Difficulties identified
- Bottom-up approach
- Five-step procedure
- *Conflict between lower and upper levels*


Step 4: Form lower level problem from (f1*,f2*)

Step 5: Form upper level problem from (v1,v2) and (U1,U2)

Step 3: Map (U1,U2) to (f1*,f2*)

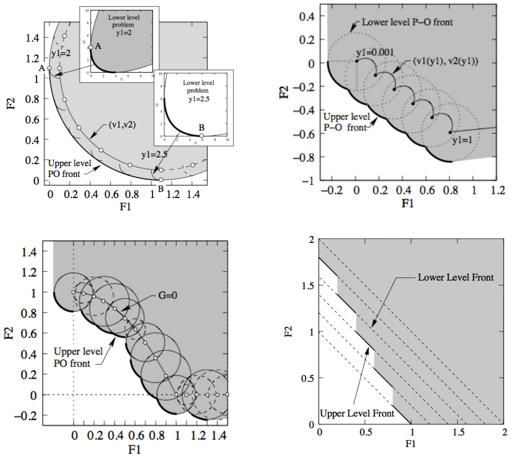
Step 2: Form envelope (U1,U2) at every (v1,v2)

Step 1: Choose Φ_U

Derived Upper level Pareto–optimal front

Some Problems

- Scalable in terms of variables and objectives
- Controllable difficulties

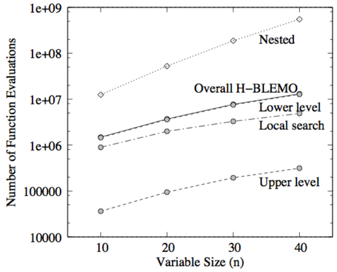


Hybrid, Self-Adaptive BLEMO

- Subpopulation size not fixed
  - Based on current population with respect to archive
- All parameters (N_l, T_l, T_u) are self-adapted
- Lower level best solutions are improved by a local search
  - Achievement scalarizing function method
- Parametric study on N_u suggests N_u=20n

Hallmark of Hybrid BLEMO

- Handle larger variables (<= 40), scalable performance
- One to two orders of magnitude less function evaluations
- Interactive lower upper level NSGA-II quicker
- Proposed method better than nested



Mine Taxation Strategy Problem
(Sinha, Malo, Frantsev and Deb, 2013)

- Kuusamo has natural beauty and a famous tourist resort
  - Contains large amounts of gold deposits
- Dragon Mining is interested in mining in the region
- Pros:
  - Generate a large number of jobs
  - Monetary gains
- Cons:
  - Run-off water from mining will pollute Kitkajoki river
  - Ore contains Uranium, mining may blemish reputation
  - Open pit mines located next to Ruka slopes will be a turn-off for skiing and hiking enthusiasts
  - Permanent damage to the nature

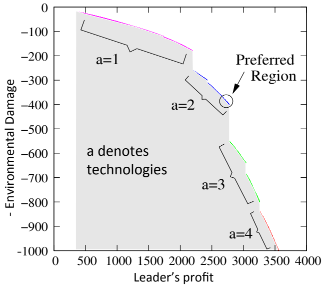
Leader-Follower Problem

- Government needs to make a decision about
  - Whether to allow mining in the region
  - If yes, then to what extent?
  - How to tax the mining company to meet its objectives?
- Mining company wants to maximize its gains given the taxes

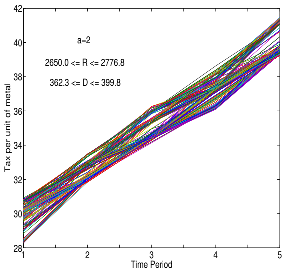
Leader: Government
Maximize revenue from taxes
Minimize Pollution

Follower: Mining Company
Maximize Profit

BLEMO Results



Pareto-frontier for the leader's problem

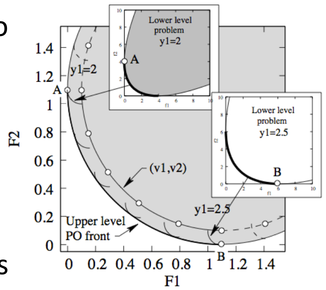


Taxation strategies in preferred region.

Preferred srategy
~75% profit to the government, ~25% to company

MEBO with Decision Making

- Preference in LL Pareto front may not lead to UL Pareto solutions
- Converse is not true
- Raises interesting hierarchy among UL and LL decision-makers



Raises very interesting scenarios, which we are currently pursuing!

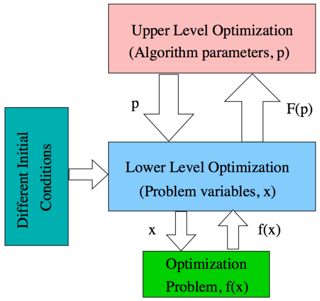
Robust EBO

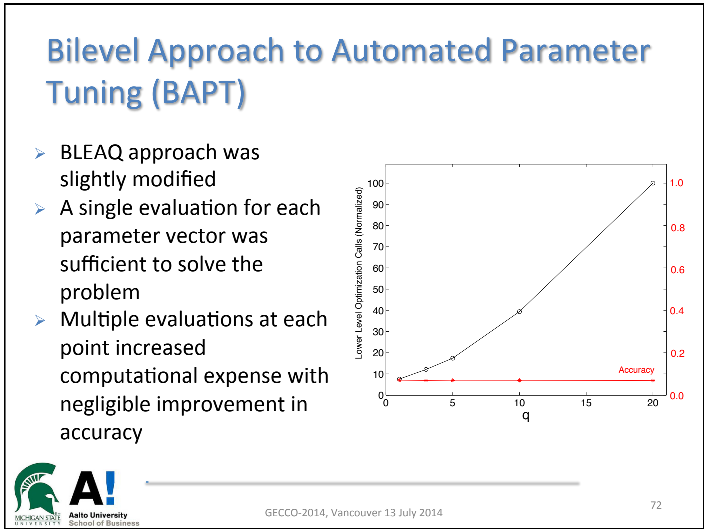
- Uncertainty in LL variables/parameters may make LL optimum sensitive
  - Find robust LL solution, but sacrifice in UL function value
- Uncertainty in UL variables/parameters may make UL optimum sensitive
  - Robust UL solution may not come from robust LL solution
- Uncertainty in both LL and UL variables
  - Create interesting scenarios, which we are currently pursuing

---

EBO Applications: Parameter Tuning of Algorithms

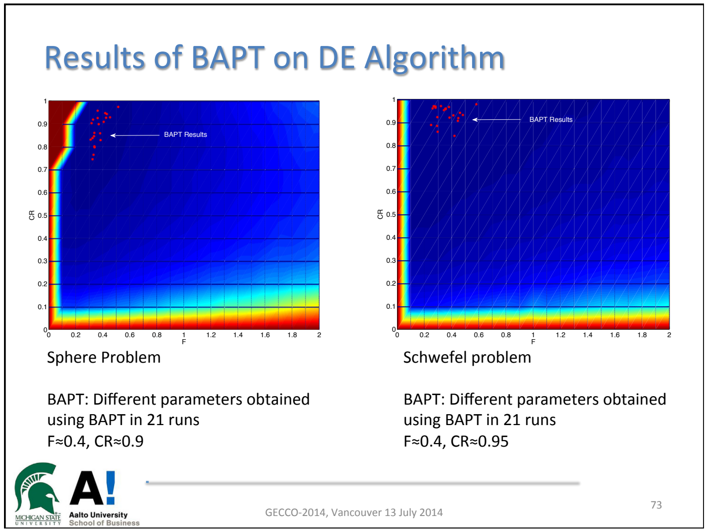
- GECCO-2014 paper (Sinha, Malo, Xu and Deb)
- Upper Level: Find optimal parameters that maximize algorithm performance over a number of initial cond.
- Lower Level: Run the optimization algorithm to find optimized solution





Bilevel Approach to Automated Parameter Tuning (BAPT)

- BLEAQ approach was slightly modified
- A single evaluation for each parameter vector was sufficient to solve the problem
- Multiple evaluations at each point increased computational expense with negligible improvement in accuracy



Results of BAPT on DE Algorithm

Sphere Problem

BAPT: Different parameters obtained using BAPT in 21 runs
F≈0.4, CR≈0.9

Schwefel problem

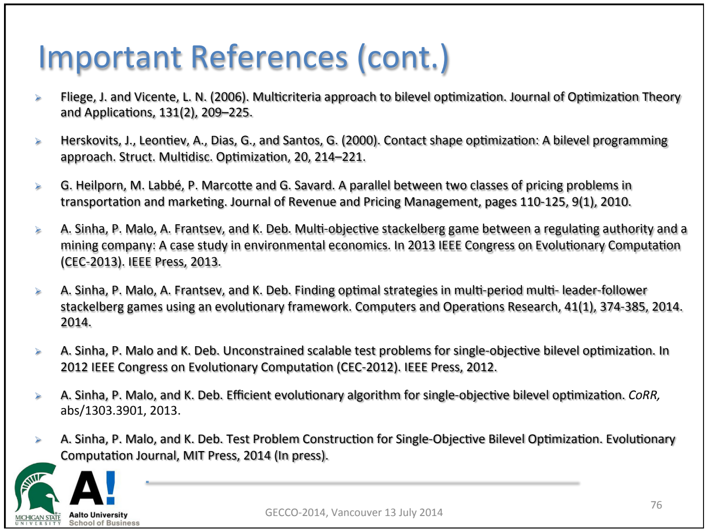
BAPT: Different parameters obtained using BAPT in 21 runs
F≈0.4, CR≈0.95

Conclusions

- Bilevel problems are plenty in practice, but are avoided due to lack of efficient methods
- Bilevel optimization received lukewarm interest by EA researchers so far
- Population approach of EA makes tremendous potential
- Nested nature of the problem makes the task computationally expensive
- Meta-modeling based EBO and its extensions show promise
- More collaborative efforts are needed

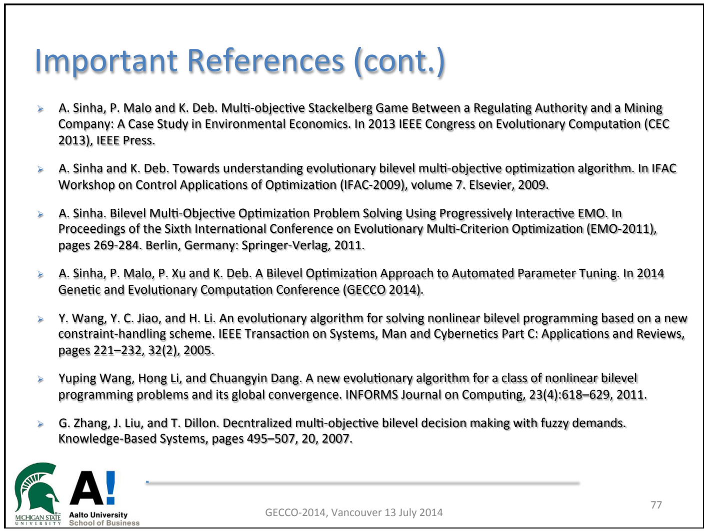
Important References

- J.F. Bard. Practical bilevel optimization: Algorithms and applications. The Netherlands: Kluwer, 1998.
- Brotcorne, M. Labbé, P. Marcotte, G. Savard and L. M. Houy. A Bilevel Model for Toll Optimization on a Multicommodity Transportation Network, Transportation Science, pages 345-358, 35(4), 2001.
- S. Dempe. Foundations of bilevel programming. Kluwer, Dordrecht, 2002.
- Dempe, S., Dutta, J., and Lohse, S. (2006). Optimality conditions for bilevel programming problems. Optimization, 55(5-6), 505–524.
- K. Deb and A. Sinha. Constructing test problems for bilevel evolutionary multi-objective optimization. In 2009 IEEE Congress on Evolutionary Computation (CEC-2009), pages 1153--1160. IEEE Press, 2009.
- K. Deb and A. Sinha. Solving bilevel multi-objective optimization problems using evolutionary algorithms. In Evolutionary Multi-Criterion Optimization (EMO-2009), pages 110--124. Berlin, Germany: Springer-Verlag, 2009.
- K. Deb and A. Sinha. An efficient and accurate solution methodology for bilevel multi-objective programming problems using a hybrid evolutionary-local-search algorithm. Evolutionary Computation Journal, pages 403-449, 14(5). MIT Press, 2010.
- G. Eichfelder. Solving Nonlinear Multiobjective Bilevel Optimization Problems with Coupled Upper Level Constraints, Preprint No. 320, Preprint-Series of the Institute of Applied Mathematics, Univ. Erlangen-Nürnberg, Germany, 2007.



Important References (cont.)

- Fliege, J. and Vicente, L. N. (2006). Multicriteria approach to bilevel optimization. Journal of Optimization Theory and Applications, 131(2), 209–225.
- Herskovits, J., Leontiev, A., Dias, G., and Santos, G. (2000). Contact shape optimization: A bilevel programming approach. Struct. Multidisc. Optimization, 20, 214–221.
- G. Heilporn, M. Labbé, P. Marcotte and G. Savard. A parallel between two classes of pricing problems in transportation and marketing. Journal of Revenue and Pricing Management, pages 110-125, 9(1), 2010.
- A. Sinha, P. Malo, A. Frantsev, and K. Deb. Multi-objective stackelberg game between a regulating authority and a mining company: A case study in environmental economics. In 2013 IEEE Congress on Evolutionary Computation (CEC-2013). IEEE Press, 2013.
- A. Sinha, P. Malo, A. Frantsev, and K. Deb. Finding optimal strategies in multi-period multi- leader-follower stackelberg games using an evolutionary framework. Computers and Operations Research, 41(1), 374-385, 2014. 2014.
- A. Sinha, P. Malo and K. Deb. Unconstrained scalable test problems for single-objective bilevel optimization. In 2012 IEEE Congress on Evolutionary Computation (CEC-2012). IEEE Press, 2012.
- A. Sinha, P. Malo, and K. Deb. Efficient evolutionary algorithm for single-objective bilevel optimization. *CoRR,* abs/1303.3901, 2013.
- A. Sinha, P. Malo, and K. Deb. Test Problem Construction for Single-Objective Bilevel Optimization. Evolutionary Computation Journal, MIT Press, 2014 (In press).



Important References (cont.)

- A. Sinha, P. Malo and K. Deb. Multi-objective Stackelberg Game Between a Regulating Authority and a Mining Company: A Case Study in Environmental Economics. In 2013 IEEE Congress on Evolutionary Computation (CEC 2013), IEEE Press.
- A. Sinha and K. Deb. Towards understanding evolutionary bilevel multi-objective optimization algorithm. In IFAC Workshop on Control Applications of Optimization (IFAC-2009), volume 7. Elsevier, 2009.
- A. Sinha. Bilevel Multi-Objective Optimization Problem Solving Using Progressively Interactive EMO. In Proceedings of the Sixth International Conference on Evolutionary Multi-Criterion Optimization (EMO-2011), pages 269-284. Berlin, Germany: Springer-Verlag, 2011.
- A. Sinha, P. Malo, P. Xu and K. Deb. A Bilevel Optimization Approach to Automated Parameter Tuning. In 2014 Genetic and Evolutionary Computation Conference (GECCO 2014).
- Y. Wang, Y. C. Jiao, and H. Li. An evolutionary algorithm for solving nonlinear bilevel programming based on a new constraint-handling scheme. IEEE Transaction on Systems, Man and Cybernetics Part C: Applications and Reviews, pages 221–232, 32(2), 2005.
- Yuping Wang, Hong Li, and Chuangyin Dang. A new evolutionary algorithm for a class of nonlinear bilevel programming problems and its global convergence. INFORMS Journal on Computing, 23(4):618–629, 2011.
- G. Zhang, J. Liu, and T. Dillon. Decntralized multi-objective bilevel decision making with fuzzy demands. Knowledge-Based Systems, pages 495–507, 20, 2007.

Join the Club!

Contact any of the authors for further information:

- Ankur Sinha: ankur.sinha@aalto.fi
- Kalyanmoy Deb: kdeb@egr.msu.edu

EBO Website: http://bilevel.org

- Matlab Codes of BLEAQ
- Technical papers
- Introductory materials on EBO
- Active research groups
- Register your name, if you are interested