Evolutionary Bilevel Optimization:
Applications and Methods

Ankur Sinha
Associate Professor
Production and Quantitative Methods
Indian Institute of Management Ahmedabad
Gujarat 380015, India
asinha@iima.ac.in

Kalyanmoy Deb
Professor and Koenig Endowed Chair
Electrical and Computer Engineering
Michigan State University
East Lansing, MI 48824, USA
kdeb@egr.msu.edu

Outline

- Bilevel Optimization: An Introduction
- Genesis
- Solution Methodologies
- Test Problem Construction
- Results
- Multi-objective Bilevel Optimization
- Applications

What is Bilevel Optimization?

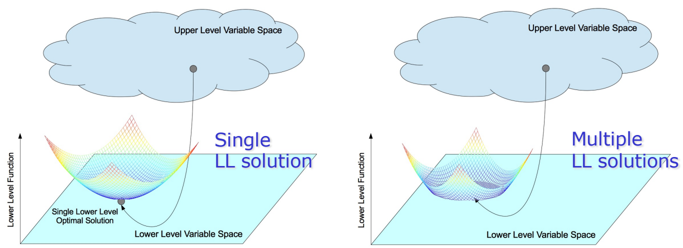
- Two levels of optimization tasks
  - Upper level: (x_u,x_l)
  - Lower level: (x_l), x_u is fixed
- An upper level feasible solution must be an optimal lower level solution: $(x_u, x_l^*(x_u))$

Min is default, can be max in any of the levels

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

An Illustration

- Lower level solution x_l can be a singleton or multi-valued
- Bilevel optimal solution corresponds to the best combination of lower level optimum and upper level values

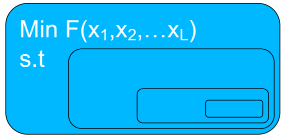


Properties of Bilevel Problems

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

Multi-level Optimization

- Multi-level (L levels) optimization
  - Two or more levels of optimization
  - Nested structure



Origin of Bilevel Programming

An Extension of Mathematical Programming

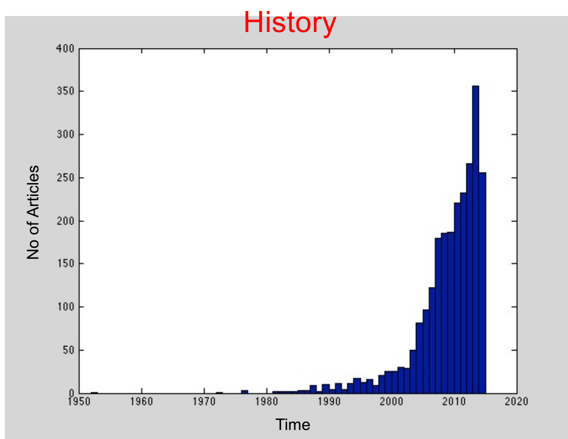
- All optimization problems are special cases of bilevel programming

–Bracken and McGill (1973)

Stackelberg Games

- Bilevel programs commonly appear in game theory when there is a leader and follower

–Stackelberg (1952)



Topic Models

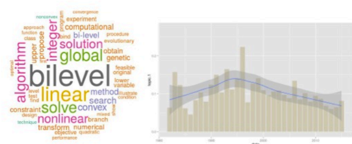


Fig. 1: Topic: Methods.

Fig. 3: Topic: Classical game theory.

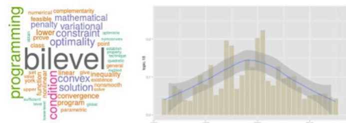


Fig. 2: Topic: Optimality conditions.

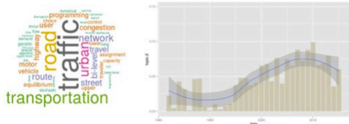


Fig. 4: Topic: Network design.

More than 10% share

Topic Models

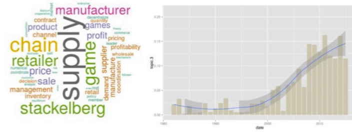


Fig. 5: Topic: Supply chain applications.

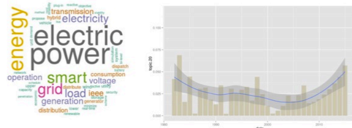


Fig. 7: Topic: Electricity transmission applications.

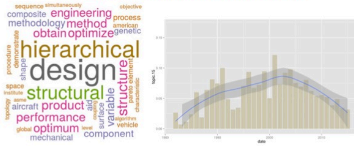


Fig. 6: Topic: Optimal design applications.

Fig. 8: Topic: Telecommunication applications.

Around 5-10% share

Topic Models

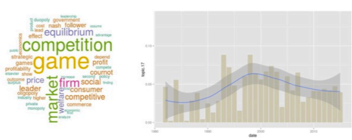


Fig. 9: Topic: Business applications.

Fig. 11: Topic: Hierarchical decision making applications.

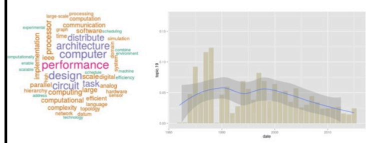


Fig. 10: Topic: Computer architecture and circuit design.

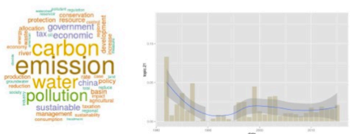


Fig. 12: Topic: Environment applications.

Around 5% share

Topic Models

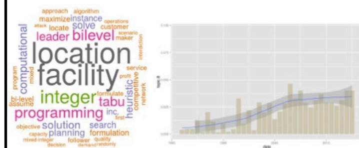


Fig. 13: Topic: Facility location applications.

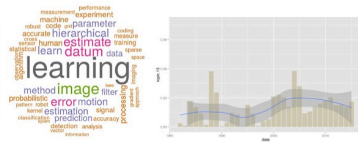


Fig. 15: Topic: Machine learning applications.

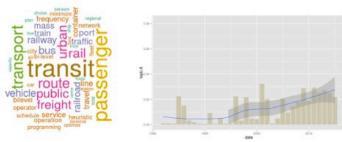


Fig. 14: Topic: Vehicle Routing applications.

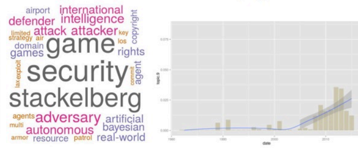


Fig. 16: Topic: Defense applications.

Less than 5% share

An Example

Bilevel optimization problem with lower level constraints!

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

Such that

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

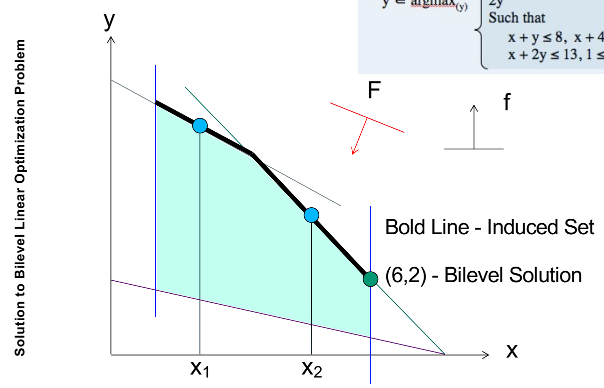
A Bilevel Linear Optimization Problem

Example (Cont.)

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

Such that

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



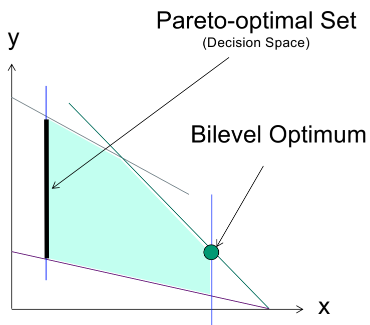
Bilevel vs. Multi-objective

Bilevel linear optimization problem is now modified as follows:

$$\text{Min}_{(x,y)}\ 3y + x$$
$$\text{Min}_{(x,y)}\ -2y$$

Such that

$x + y \le 8$
$x + 4y \ge 8$
$x + 2y \le 13$
$1 \le x \le 6$



Note that two-objective and bilevel problems are different in principle

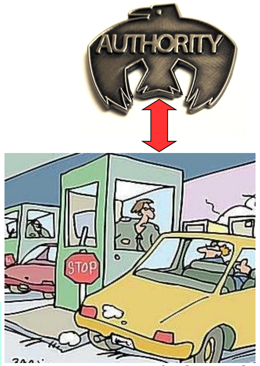
Some Applications

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

Toll Setting Problem

- Authority's (Upper level) 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 possible tolls
- Highway users' (Lower level) 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



Does it make sense to choose or not to choose a toll high-way before knowing the toll amount?

Brotcorne et al. (2001)

Stackelberg Competition

Competition between a leader and a follower firm (Duopoly)

Leader solves the following optimization problem to maximize its profit

$$\begin{aligned} \max_{q_l, q_f} \quad & \Pi_l = P(q_l, q_f)q_l - C(q_l) \\ \text{s.t.} \quad & q_f \in \arg\max_{q_f}\{\Pi_f = P(Q)q_f - C(q_f)\}, \\ & q_l + q_f \geq Q, \\ & q_l, q_f, Q \geq 0, \end{aligned}$$

where Q is the quantity demanded, $P(q_l, q_f)$ is the price of the goods sold, and $C(\cdot)$ is the cost of production of the respective firm. The variables in this model are the production levels of each firm q_l, q_f and demand Q.

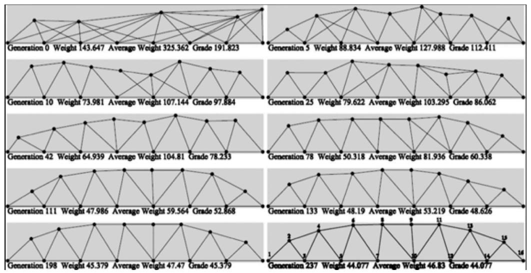
If the leader and follower have similar functions, leader always makes a higher profit.

- First mover's advantage

Can be extended to multiple leaders and multiple followers

Structural Optimization

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



Taxation Strategy

- Recently, there was a controversy 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 (UL), the mining company (LL) optimizes its own profits

Leader: Government Maximize revenue from taxes, Minimize Pollution

Can the government know the outcome for any tax policy chosen by them?

Follower: Mining Company Maximize Profit

Sinha, et al. (2013)

21

Defense Applications

Hub-and-Spoke Networks

Interdiction Problem
Attacker-Defender

Two levels

Attacker
Maximize operating costs post attack

Defender
Minimize operating cost

Protection Problem
Defender-Attacker

Two levels

Defender
Minimize the maximum damage by fortification

Attacker
Maximize damage

Robust Design
Defender-Attacker-Defender

Three levels

Defender
Take interdiction problem into account during design phase

Attacker
Maximize operating costs post attack

Defender
Minimize operating cost

Nine key substations out of 55,000 substations can result in a coast to coast blackout in US
WSJ Link

22

Agri-business Management

Upper Level
Regulator

Objective 1: Minimize Pollution (Fertilizer)
Objective 2: Maximize Revenues

Lower Level
Multiple Farms

Objective: Maximize Individual Profit

Decision Variables: Taxation, inputs, outputs

Contains multiple lower level optimization problems

Whittaker et al. (2016)

23

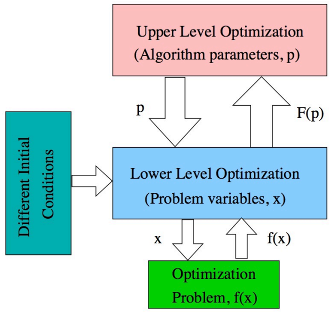
Parameter Tuning

Upper Level: Find optimal parameters that maximize algorithm performance over a number of initial conditions

Lower Level: Run the optimization algorithm to find optimized solution

Does it make sense to conclude an algorithm's performance without spelling out algorithm parameters?

Researchers commonly rely on grid search or random search



Bennett et al. (2008), Sinha et al. (2014)

24

Inverse Optimal Control

- While performing actions humans optimize certain unknown cost function
- It might be interesting to have an idea of the cost function that might help in designing efficient humanoids
- Given the data corresponding to the motion identifying the reward or cost function becomes an inverse problem

Mombaur et al. (2010), Suryan et al. (2016)

25

Solution Methodologies

- Single-level reduction using KKT
  - Bialas and Karwan (1984), Bard and Falk (1982), Bard and Moore (1990)
- Descent methods
  - Savard and Gauvin (1994), Vicente et al. (1994)
- Penalty function methods
  - Aiyoshi and Shimizu (1981, 1984), Ishijuka and Aiyoshi (1992), White and Anandalingam (1993)
- Trust region methods
  - Colson et al. (2005)
- Using lower level optimal value function
  - Mitsos (2010)

26

Special Cases

- Linear bilevel problems
  - Reducible to a mixed integer linear program
- Bilevel problems with combinatorial variables at upper level and linear program at lower level
  - Reducible to a mixed integer linear program
- Bilevel problems with combinatorial variables at both levels
  - Very hard to solve
- Bilevel problems with similar objectives at both levels
  - Reduces to minmax or minmin (min) problems
  - Ideas of duality can be utilized

27

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 should help solve BO problems better

28

Niche of Evolutionary Methods (cont.)

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

EAs for Bilevel Optimization

- Most of the EAs for bilevel optimization have been nested in nature
  - Using one algorithm for upper level and solving the lower level optimization problem for every upper level point
  - Not very interesting!
  - Expensive even for small instances!
  - Non-scalable!

Bilevel Optimization using EAs

EA at upper level and exact method at lower level
- Mathieu et al. (1994): LP for lower level and GA for upper level
- Yin (2000): Frank-Wolfe Algorithm for lower level and EA for upper level

EA at both upper and lower level
- Li et al. (2006): Particle Swarm Optimization at both levels
- Angelo et al. (2013): Differential Evolution at both levels
- Sinha et al. (2014): Genetic Algorithm at both levels

EA used after single-level reduction
- EA researchers have also tried replacing the lower level problems using KKT (Hejazi et al. (2002), Wang et al. (2008), Li et al. (2007))

Bilevel Optimization using EAs

Approximating lower level level rational response
- Sinha, Malo, Deb. (2013, 2014, 2017): Iteratively approximates lower level optimal response with upper level decision vector (Discussed later)

Approximating lower level optimal value function
- Sinha, Malo, Deb. (2016): Iteratively approximates lower level optimal function value with upper level decision vector (Discussed later)

Trust region method and Approximate KKT
- Sinha, Soun and Deb (2017)

Kriging based methods
- Sinha et al (2018), Islam et al. (2018)

Can EAs be really useful for bilevel optimization?

33

Can EAs be really useful for bilevel optimization?

- Nested approaches are certainly not the way forward

34

Can EAs be really useful for bilevel optimization?

- It is noteworthy that at each iteration an EA has a population of points
  - Can these population of points be put to use to approximate certain mappings in bilevel?
  - Exploiting the structure and properties of the problem is essential!

35

Approach 1 (Lower Level Reaction Set Mapping)

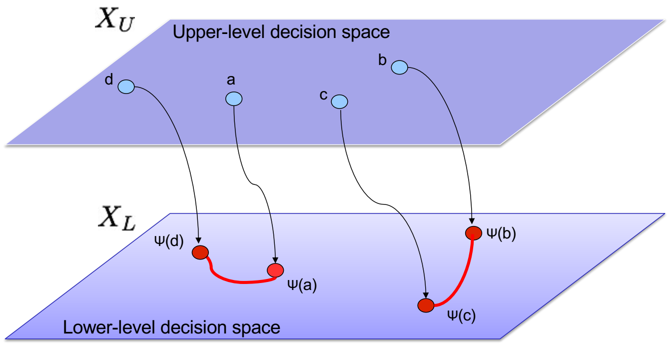
$$\Psi(x_u) = \underset{x_l}{\text{argmin}}\{f(x_u, x_l) : g_j(x_u, x_l) \leq 0, j = 1, \ldots, J\}.$$

$$\begin{aligned}
\min_{x_u, x_l} \quad & F(x_u, x_l) \\
\text{s.t.} \quad & \\
& x_l \in \Psi(x_u) \\
& G_k(x_u, x_l) \leq 0, k = 1, \ldots, K
\end{aligned}$$

Step 0: Solve the lower level problem completely for the initial population
Step 1: Use the population members to approximate the Ψ-mapping locally
Step 2: Solve the reduced single level problem for a few iterations
Step 3: Update the local Ψ-mappings and continue
Step 4: If termination criteria not met, go to Step 2

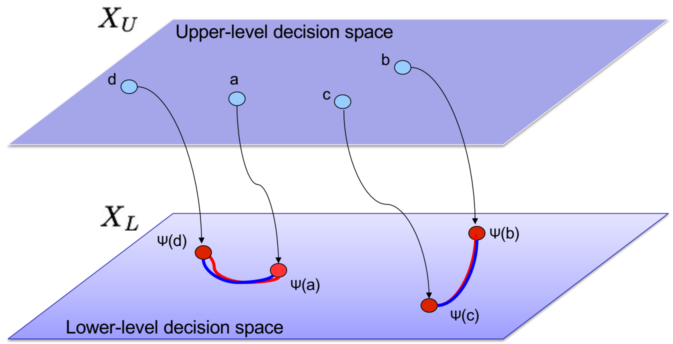
36

Actual Ψ-mapping



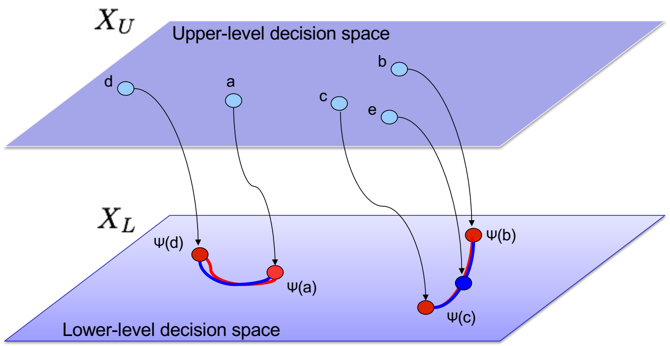
37

Approximate Ψ-mapping



38

Using approximate Ψ-mapping



For new upper level point 'e', we need not solve the lower level problem, as an approximate estimate is available

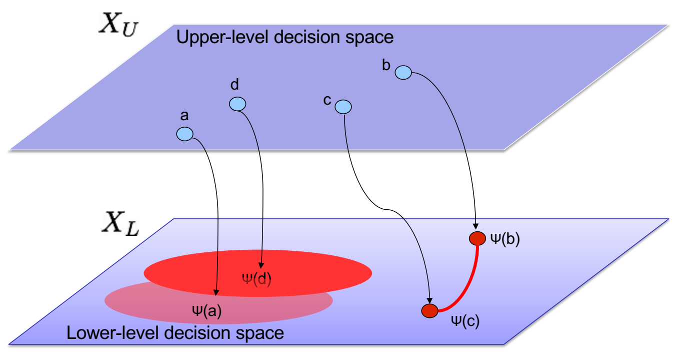
39

Approximation Choice

- Tried different strategies for localized approximation, like,
  - Linear Approximation
  - Piecewise linear approximation
  - Quadratic approximation
- Results were favorable and similar with piecewise-linear as well as quadratic approximation
- Decided to use quadratic approximation because of its simplicity
- More complex techniques like neural networks are an obvious extension but require large number of points

40

Set-valued Ψ becomes problematic



Dual challenge: 1. Finding the set and 2. Approximating the set

41

Approach 2 (Optimal Value Function Mapping)

$$\varphi(x_u) = \min_{x_l}\{f(x_u, x_l) : x_l \in \Omega(x_u)\}$$

$$\begin{aligned}
&\min_{x_u, x_l} \quad F(x_u, x_l)\\
&\text{s.t.}\\
&\quad f(x_u, x_l) \le \varphi(x_u)\\
&\quad g_j(x_u, x_l) \le 0, j = 1, \dots, J\\
&\quad G_k(x_u, x_l) \le 0, k = 1, \dots, K
\end{aligned}$$

Step 0: Solve the lower level problem completely for the initial population
Step 1: Use the population members to approximate the φ-mapping locally
Step 2: Solve the reduced single level problem for a few iterations
Step 3: Update the local φ-mappings and continue
Step 4: If termination criteria not met, go to Step 2

42

Issues

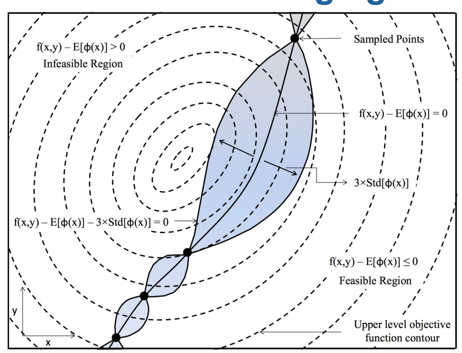
$$\varphi(x_u) = \min_{x_l}\{f(x_u, x_l) : x_l \in \Omega(x_u)\}$$

$$\begin{aligned}
&\min_{x_u, x_l} \quad F(x_u, x_l)\\
&\text{s.t.}\\
&\quad f(x_u, x_l) \le \varphi(x_u)\\
&\quad g_j(x_u, x_l) \le 0, j = 1, \dots, J\\
&\quad G_k(x_u, x_l) \le 0, k = 1, \dots, K
\end{aligned}$$

- The approximate φ-mapping makes the region highly constrained
- With errors in estimation of φ-mapping the reduced problem might become infeasible

43

Approximation of Φ-mapping through Kriging



Kriging provides both mean and standard deviation

$f(x_u, x_l) \le \varphi(x_u)$ + 3×Std[ϕ(x)]

Addition of the standard deviation term ensures feasibility of the auxiliary problem

Sinha et al. (2018)
Best paper award at WCCI 2018

44

Using Approximate KKT Conditions

- KKT conditions are hard to satisfy because of strict equality conditions
- It is possible to relax the KKT conditions using approximate KKT conditions (Dutta 2013)
- Bilvel problems can be replaced with approximate KKT conditions

$$\min_{y} \quad f(y)$$
subject to
$$g_j(y) \leq 0, j = 1, \ldots, J.$$

$\Rightarrow$

$$\min_{y} \quad \epsilon$$
subject to
$$g_j(y) \leq 0, j = 1, \ldots, J,$$
$$||\nabla_y L(y, \lambda)||^2 \leq \epsilon,$$
$$\sum_{j=1}^{J} \lambda_j g_j(y) \geq -\epsilon$$
$$\lambda_j \geq 0, j = 1, \ldots, J,$$
where $L(y, \lambda) = f(y) + \sum_{j=1}^{J} \lambda_j g_j(y)$

45

Next Part by K. Deb
https://www.coin-lab.org

46

Outline

- Bilevel Optimization: An Introduction
- Genesis
- Solution Methodologies (Ψ and Φ mappings)
- Test Problem Construction
- Results
- Multi-objective Bilevel Optimization
- Applications

47

Test Problems

- Given that a convergence proof is difficult, we can only use test problems to justify whether an an algorithm works or not!
- First, we begin with some simple test problems

48

8-Problem Test Suite (TP1-TP8)

Problem	Formulation	Best Known Sol.
TP1 2-var UL 2-var LL $n=2,\ m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=(x_1-30)^2+(x_2-20)^2-20y_1+20y_2,$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=(x_1-y_1)^2+(x_2-y_2)^2 \\ 0\le y_i\le 10,\quad i=1,2 \end{array}\right\},$ $x_1+2x_2\ge 30, x_1+x_2\le 25, x_2\le 15$	$F=225.0$ $f=100.0$
TP2 2-var UL 2-var LL $n=2,\ m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=2x_1+2x_2-3y_1-3y_2-60,$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=(y_1-x_1+20)^2+(y_2-x_2+20)^2 \\ x_1-2y_1\ge 10, x_2-2y_2\ge 10 \\ -10\ge y_i\ge 20,\quad i=1,2 \end{array}\right\},$ $x_1+x_2+y_1-2y_2\le 40,$ $0\le x_i\le 50,\quad i=1,2.$	$F=0.0$ $f=100.0$

Test Problems

Problem	Formulation	Best Known Sol.
TP3 2-var UL 2-var LL $n=2,\ m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=-(x_1)^2-3(x_2)^2-4y_1+(y_2)^2,$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=2(x_1)^2+(y_1)^2-5y_2 \\ (x_1)^2-2x_1+(x_2)^2-2y_1+y_2\ge -3 \\ x_2+3y_1-4y_2\ge 4 \\ 0\le y_i,\quad i=1,2 \end{array}\right\},$ $(x_1)^2+2x_2\le 4,$ $0\le x_i,\quad i=1,2$	$F=-18.6787$ $f=-1.0156$
TP4 2-var UL 3-var LL $n=2,\ m=3$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=-8x_1-4x_2+4y_1-40y_2-4y_3,$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=x_1+2x_2+y_1+y_2+2y_3 \\ y_2+y_3-y_1\le 1 \\ 2x_1-y_1+2y_2-0.5y_3\le 1 \\ 2x_2+2y_1-y_2-0.5y_3\le 1 \\ 0\le y_i,\quad i=1,2,3 \end{array}\right\},$ $0\le x_i,\quad i=1,2$	$F=-29.2$ $f=3.2$

Test Problems

Problem	Formulation	Best Known Sol.
TP5 2-var UL 2-var LL $n=2,\ m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=rt(x)x-3y_1-4y_2+0.5t(y)y,$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=0.5t(y)hy-t(b(x))y \\ -0.333y_1+y_2-2\le 0 \\ y_1-0.333y_2-2\le 0 \\ 0\le y_i,\quad i=1,2 \end{array}\right\},$ where $h=\begin{pmatrix} 1 & 3 \\ 3 & 10 \end{pmatrix}, b(x)=\begin{pmatrix} -1 & 2 \\ 3 & -3 \end{pmatrix}x, r=0.1$ $t(\cdot)$ denotes transpose of a vector	$F=-3.6$ $f=-2.0$
TP6 1-var UL 2-var LL $n=1,\ m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=(x_1-1)^2+2y_1-2x_1,$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=(2y_1-4)^2+ \\ (2y_2-1)^2+x_1y_1 \\ 4x_1+5y_1+4y_2\le 12 \\ 4y_2-4x_1-5y_1\le -4 \\ 4x_1-4y_1+5y_2\le 4 \\ 4y_1-4x_1+5y_2\le 4 \\ 0\le y_i,\quad i=1,2 \end{array}\right\},$ $0\le x_1$	$F=-1.2091$ $f=7.6145$

Test Problems

Problem	Formulation	Best Known Sol.
TP7 2-var UL 2-var LL $n=2,\ m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=-\frac{(x_1+y_1)(x_2+y_2)}{1+x_1y_1+x_2y_2},$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=\frac{(x_1+y_1)(x_2+y_2)}{1+x_1y_1+x_2y_2} \\ 0\le y_i\le x_i,\quad i=1,2 \end{array}\right\},$ $(x_1)^2+(x_2)^2\le 100$ $x_1-x_2\le 0$ $0\le x_i,\quad i=1,2$	$F=-1.96$ $f=1.96$
TP8 2-var UL 2-var LL $n=2,\ m=2$	$\underset{(x,y)}{\text{Minimize}}\ F(x,y)=\lvert 2x_1+2x_2-3y_1-3y_2-60\rvert,$ s.t. $y\in\underset{(y)}{\text{argmin}}\left\{\begin{array}{l} f(x,y)=(y_1-x_1+20)^2+ \\ (y_2-x_2+20)^2 \\ 2y_1-x_1+10\le 0 \\ 2y_2-x_2+10\le 0 \\ -10\le y_i\le 20,\quad i=1,2 \end{array}\right\},$ $x_1+x_2+y_1-2y_2\le 40$ $0\le x_i\le 50,\quad i=1,2$	$F=0.0$ $f=100.0$

Results on TPs

	UL Func. Evals.			LL Func. Evals.		
	φ-Appx Med	Ψ-Appx Med	No-Appx Med	φ-Appx Med	Ψ-Appx Med	No-Appx Med
TP1	134	150	-	1438	2061	-
TP2	148	193	436	1498	2852	5686
TP3	187	137	633	2478	1422	6867
TP4	299	426	1755	3288	6256	19764
TP5	175	270	576	2591	2880	6558
TP6	110	94	144	1489	1155	1984
TP7	166	133	193	2171	1481	2870
TP8	212	343	403	2366	5035	7996

Sinha et al. (2016)

Results on TPs (Cont.)

Approach 1: Ψ- Mapping (Approach 1)
Approach 2: φ –Mapping (Approach 2)

	UL Func. Evals.			LL Func. Evals.		
	φ-Appx Med	Ψ-Appx Med	No-Appx Med	φ-Appx Med	Ψ-Appx Med	No-Appx Med
TP1	134	150	-	1438	2061	-
TP2	148	193	436	1498	2852	5686
TP3	187	137	633	2478	1422	6867
TP4	299	426	1755	3288	6256	19764
TP5	175	270	576	2591	2880	6558
TP6	110	94	144	1489	1155	1984
TP7	166	133	193	2171	1481	2870
TP8	212	343	403	2366	5035	7996

In general, φ-Mapping approach is better

Sinha et al. (2016)

Comparison with other approaches

Approach 1: Ψ- Mapping (Approach 1)
Approach 2: φ –Mapping (Approach 2)

21 runs

	Mean Func. Evals. (UL+LL)				
	φ-appx.	Ψ-appx.	No-appx.	WJL	WLD
TP1	1595	2381	35896	85499	86067
TP2	1716	3284	5832	256227	171346
TP3	2902	1489	7469	92526	95851
TP4	3773	6806	21745	291817	211937
TP5	2941	3451	7559	77302	69471
TP6	1689	1162	1485	163701	65942
TP7	2126	1597	2389	1074742	944105
TP8	2699	4892	5215	213522	182121

In general, φ-Mapping approach is better

WJL – Wang et al. (2005), WLD – Wang et al. (2011)

Modified Test Problems (m-TP1 to m-TP8)

General Structure (x and y are vectors):

$$F^{new}(x,y) = F(x,y) + y_p^2 + y_q^2$$
$$f^{new}(x,y) = f(x,y) + (y_p - y_q)^2$$
$$y_p, y_q \in [-1, 1]$$

y_p and y_q are LL variables, in addition to y

- Modification leads to multiple lower level optimal solutions for each upper level decision vector
- May cause Ψ-Mapping to be difficult
    - Multiple y_p and y_q variables to be mapped to

Results (Modified Test Problems)

Optimal f is modeled   Optimal y is modeled   21 runs

	Upper Level Function Evaluations			Lower Level Function Evaluations			Both Methods Fail	
	φ-Appx.			φ-Appx.			Ψ-Appx.	No-Appx.
	Min	Med	Max	Min	Med	Max	Min/Med/Max	Min/Med/Max
m-TP1	130	172	338	2096	2680	8629	-	-
m-TP2	116	217	-	2574	4360	-	-	-
m-TP3	129	233	787	1394	3280	13031	-	-
m-TP4	198	564	2831	1978	5792	28687	-	-
m-TP5	160	218	953	3206	4360	17407	-	-
m-TP6	167	174	529	2617	3520	8698	-	-
m-TP7	114	214	473	1514	5590	11811	-	-
m-TP8	150	466	2459	2521	6240	35993	-	-

Only φ-Mapping works!
A single y^* is enough

Bilevel Test Problem Construction: A Systematic Approach

- Test problems with controllable difficulties are often required to evaluate evolutionary algorithms
- Controllable and segregated difficulties help to identify what aspects of the problem, the algorithm is unable to handle

Requirements

- Controllable difficulty in convergence at upper and lower levels
- Controllable 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 bilevel test problem can be easily obtained

Test Problem Framework

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

$$\mathbf{x}_u = (\mathbf{x}_{u1}, \mathbf{x}_{u2}) \quad \text{and} \quad \mathbf{x}_l = (\mathbf{x}_{l1}, \mathbf{x}_{l2})$$ - vectors

(Sinha, Malo and Deb, 2014)

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 at each level
- 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

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

Constant in LL

Multi-modal

Controlling Difficulty in Interactions

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

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

C

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

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

C

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

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 is also possible

Controlled Multimodality

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

$$f_1(\mathbf{x}_{u1}, \mathbf{x}_{u2}) = (x_{u1}^1)^2 + (x_{u1}^1)^2 + (x_{u2}^1)^2 + (x_{u2}^2)^2$$ Scales LL values

C $$f_2(\mathbf{x}_{l1}) = (x_{l1}^1 - x_{l1}^2)^2$$ Induces multiple solutions: $x^1_{l1} = x^2_{l1}$

I $$f_3(\mathbf{x}_{u2}, \mathbf{x}_{l2}) = (x_{u2}^1 - x_{l2}^1)^2 + (x_{u2}^2 - x_{l2}^2)^2$$

C $$F_1(\mathbf{x}_{u1}) = (x_{u1}^1)^2 + (x_{u1}^1)^2$$

$$F_2(\mathbf{x}_{l1}) = (x_{l1}^1)^2 + (x_{l1}^2)^2$$ Gives best UL solution: $x^1_{l1} = x^2_{l1} = 0$

I $$F_3(\mathbf{x}_{u2}, \mathbf{x}_{l2}) = (x_{u2}^1 - x_{l2}^1)^2 + (x_{u2}^2 - x_{l2}^2)^2$$

Difficulty due to Constraints

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

- Constraints exist, but are inactive at the optimum
- A subset of constraints 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
- Any other complexities

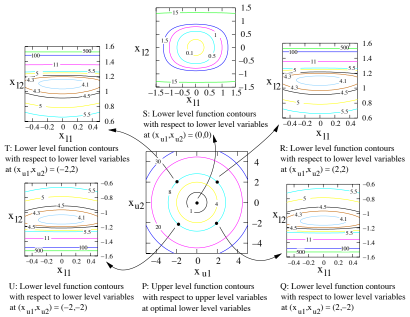
Problem 1 (SMD 1)

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

Upper and Lower Function Contours

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

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



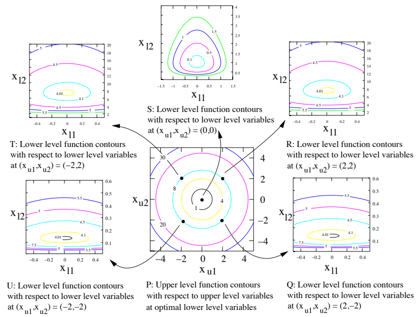
Problem 2 (SMD 2)

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

Upper and Lower Function Contours

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

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



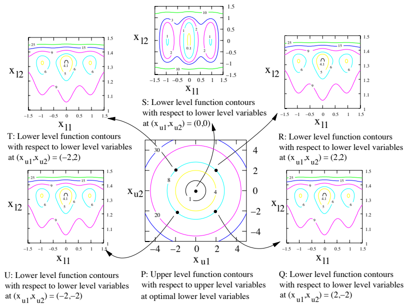
Problem 3 (SMD 3)

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

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

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

Upper and Lower Function Contours



69

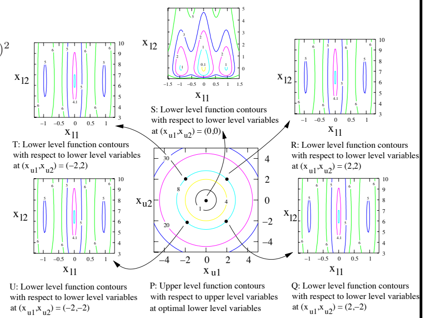
Problem 4 (SMD 4)

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

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

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

Upper and Lower Function Contours



70

Results Using BLEAQ

- Following are the results for 10 variable instances of the test problems (Sinha et al., 2014) using BLEAQ (Ψ-Mapping)
- Comparison performed against nested evolutionary approach

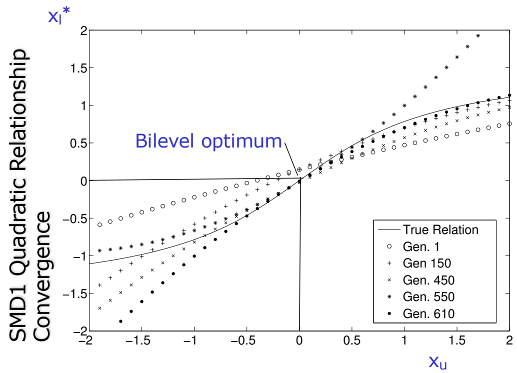
Number of Runs: 21
Savings: Ratio of FE required by nested approach against BLEAQ

Pr. No.	Best Func. Evals.		Median Func. Evals.		Worst Func. Evals.	
	LL	UL	LL (Savings)	UL (Savings)	LL	UL
SMD1	99315	610	110716 (14.71)	740 (3.34)	170808	1490
SMD2	70032	376	91023 (16.49)	614 (3.65)	125851	1182
SMD3	110701	620	125546 (11.25)	900 (2.48)	137128	1094
SMD4	61326	410	81434 (13.59)	720 (2.27)	101438	1050
SMD5	102868	330	126371 (15.41)	632 (4.55)	168401	1050
SMD6	95687	734	118456 (14.12)	952 (3.25)	150124	1410

For other problems as well, the improvement is more than an order of magnitude

71

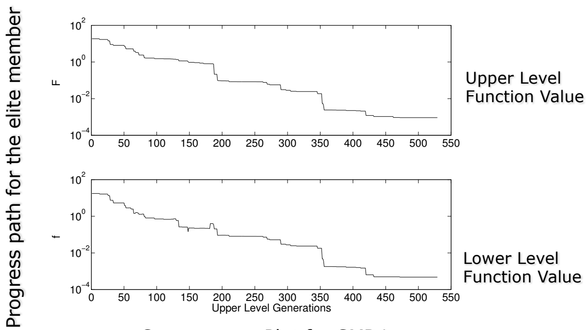
Results on SMD1



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

72

Convergence Plots on SMD1



Convergence Plot for SMD1

73

BLEAQ vs BLEAQ2

- BLEAQ (Ψ-mapping) works well on problems with single optimal solution at the lower level, but fails in the presence of multiple solutions.
- BLEAQ relies only on the approximation of the Ψ-mapping
- BLEAQ2 (combined (Ψ-φ Mappings) relies on the approximation of both Ψ and φ-mappings and is able to handle multiple lower level optimal solutions as well.

74

Advanced Topics of EBO

- Multi-objective EBO
  - At least one level has multiple objectives
- MEBO with decision-making
- Robust EBO
  - Uncertainty in at least one level
- EBO applications
  - Parameter tuning of algorithms
  - Practical applications

75

Advanced EBO Ideas (cont.)

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

76

Multi-objective EBO

- Bilevel problems may involve optimization of multiple objectives at one or both levels
- Dempe et al. (2006) developed KKT conditions
- Little work has been done in the direction of multi-objective bilevel algorithms (Eichfelder (2007), Deb and Sinha (2010))
- A general multi-objective bilevel problem may be formulated as follows:

$$\begin{aligned}
&\underset{x_u, x_l}{\text{Min}}\ F(x_u, x_l) = (F_1(x_u, x_l), \ldots, F_p(x_u, x_l)) \\
&\qquad \text{subject to} \\
&\qquad x_l \in \underset{x_l}{\text{argmin}}\{f(x_u, x_l) = (f_1(x_u, x_l), \ldots, f_q(x_u, x_l)) \\
&\qquad\qquad g_i(x_u, x_l) \geq 0, i \in I\} \\
&\qquad G_j(x_u, x_l) \geq 0, j \in J.
\end{aligned}$$

77

Optimistic Pareto Front

- Two levels of decision making
- Multiple objectives involved at both the levels

Leader

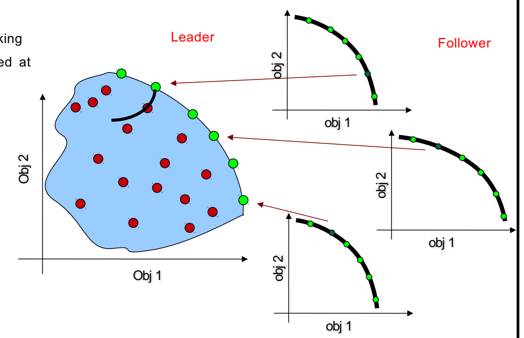
Objectives: Max Objective 1

Max Objective 2

Followers

Objectives: Max objective 1

Max objective 2



78

Pessimistic Pareto Front

- Two levels of decision making
- Multiple objectives involved at both the levels

Leader

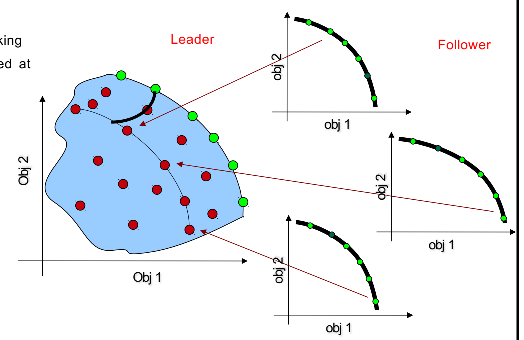
Objectives: Max Objective 1

Max Objective 2

Followers

Objectives: Max objective 1

Max objective 2



79

Preference Structure Known

- Two levels of decision making
- Multiple objectives involved at both the levels

Leader

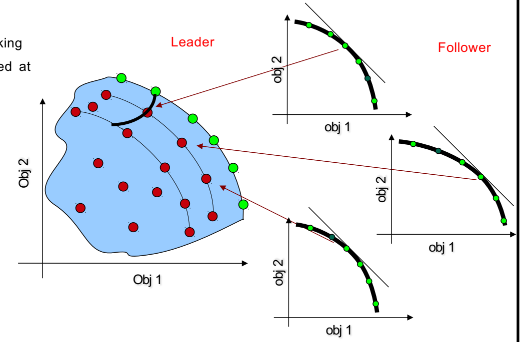
Objectives: Max Objective 1

Max Objective 2

Followers

Objectives: Max objective 1

Max objective 2



Lower level problem becomes single objective

80

Uncertainty from Unknown Preference Structure

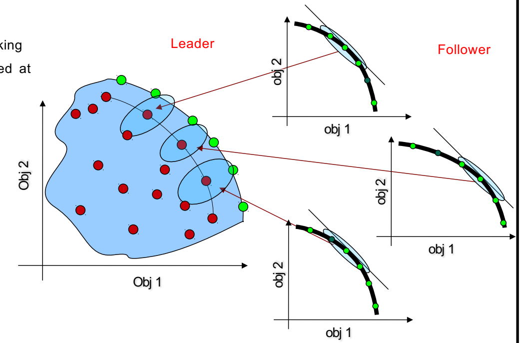
- Two levels of decision making
- Multiple objectives involved at both the levels

Leader

Objectives: Max Objective 1
Max Objective 2

Followers

Objectives: Max objective 1
Max objective 2



There is uncertainty around the frontier

(Sinha et al., 2015)

Challenges

- Such problems can be very difficult to handle
- Optimistic formulation makes little sense in these problems
- Considering a known preference structure (and accounting for uncertainties) might be a realistic and viable approach

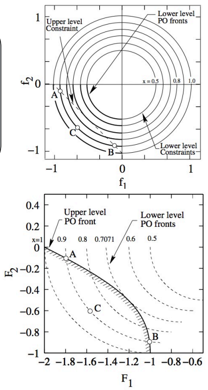
Test problem 1 and Results

$\mathbf{x} = (x_u, x_l)$ $x_u = (x)$ $x_l = (y)$

$$\text{Minimize} \quad \mathbf{F}(\mathbf{x}) = \begin{pmatrix} y_1 - x \\ y_2 \end{pmatrix},$$

$$\text{Subject to} \quad (y_1, y_2) \in \operatorname{argmin}_{(y_1,y_2)} \left\{ \mathbf{f}(\mathbf{x}) = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \middle| g_1(\mathbf{x}) = x^2 - y_1^2 - y_2^2 \geq 0 \right\},$$
$$G_1(\mathbf{x}) = 1 + y_1 + y_2 \geq 0,$$
$$-1 \leq y_1, y_2 \leq 1, \quad 0 \leq x \leq 1.$$

- Lower level Pareto front depends on x
- Upper level Pareto-optimal front lies on constraint G_1
  - Maximum two solutions from each x
  - Not all x in upper Pareto-optimal front
- Solutions possible even below the upper level Pareto-optimal front, but they are infeasible



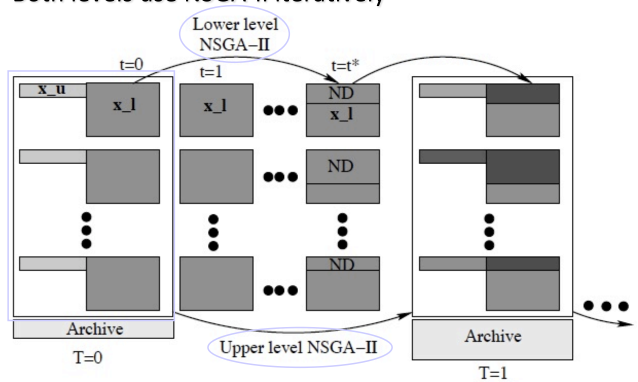
(Eichfelder, 2007)

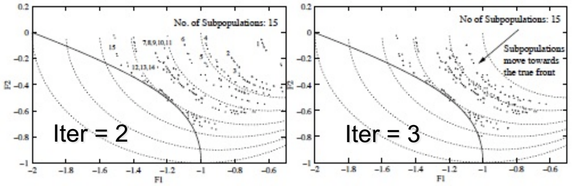
BLEMO

(Deb and Sinha, ECJ 2010)

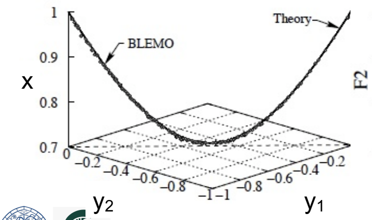
- Population structure
- Lower and upper level NSGA-II
- Archiving

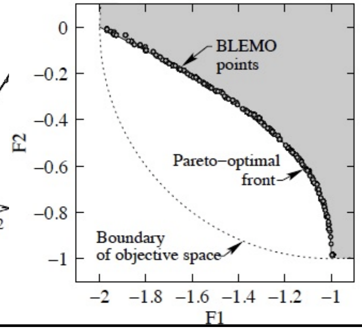
- Both levels use NSGA-II iteratively





(Deb and Sinha, 2010)





A Business DM Problem

CEO: Leader and Dept Head: follower

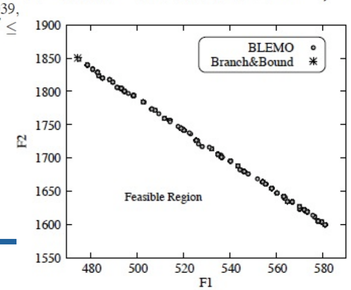
(Zhang et al, 2007)

$$\text{maximize} \quad \mathbf{F}(\mathbf{x},\mathbf{y}) = \left\{ \begin{array}{l} (1,9)(y_1,y_2)^T + (10,1,3)(x_1,x_2,x_3)^T \\ (9,2)(y_1,y_2)^T + (2,7,4)(x_1,x_2,x_3)^T \end{array} \right\},$$

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

$$G1 = (3,9)(y_1,y_2)^T + (9,5,3)(x_1,x_2,x_3)^T \le 1039,$$
$$G2 = (-4,-1)(y_1,y_2)^T + (3,-3,2)(x_1,x_2,x_3)^T \le$$
$$x_1,x_2,y_1,y_2,y_3 \ge 0.$$

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



Mine Taxation Strategy Problem from Finland

Kuusamo has natural beauty and a famous tourist resort

- Contains large amounts of gold deposits

Dragon Mining is interested in mining in the region

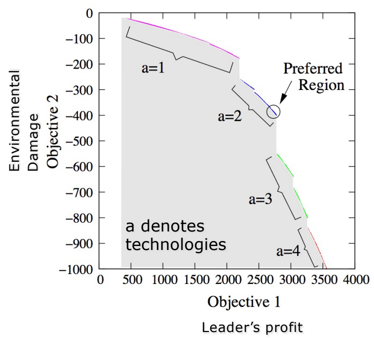
Pros:

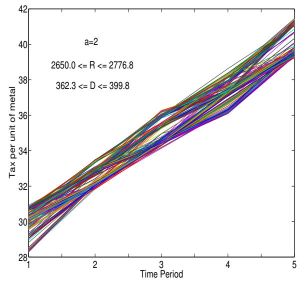
- Generate a large number of jobs
- Monetary gains

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

BLEAQ Results



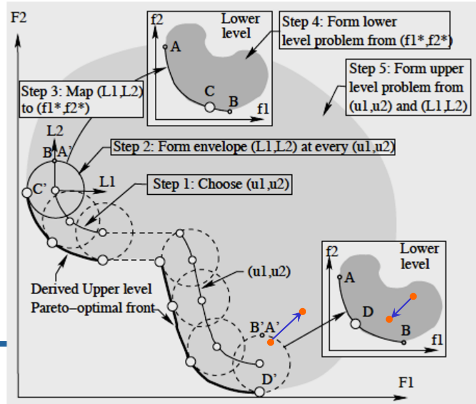


Taxation strategies in preferred region.

Preferred strategy: ~75% profit to the government, ~25% to company: Upper level has more decision-making power

EMBO Test Problem Construction Principle

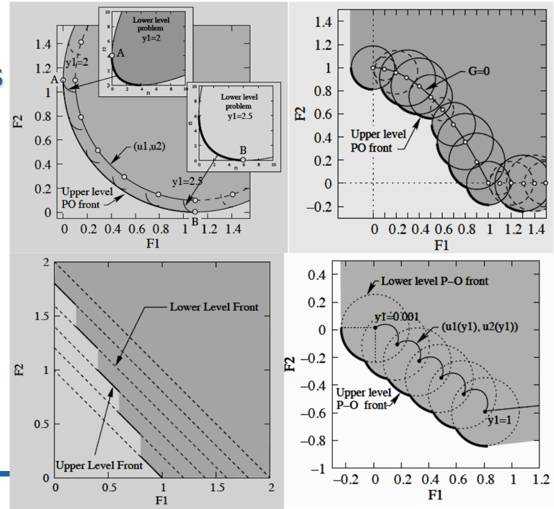
- Difficulties identified
- Bottom-up approach
- Five-step procedure
- *Conflict between lower and upper levels*



Some Problems

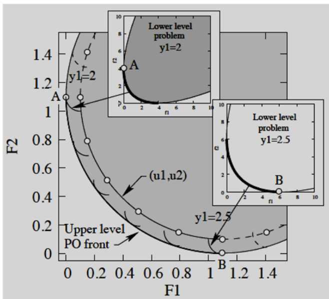
Scalable in terms of variables and objectives

Controllable difficulties



EMBO with Decision-Making

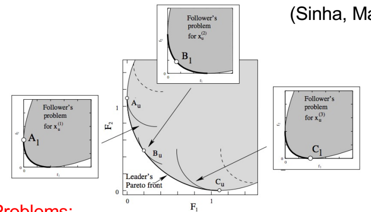
- Preference in LL Pareto front may not lead to UL Pareto solutions
- Converse is not true
  - Unequal importance among UL and LL DMs
- Raises interesting hierarchy among UL and LL decision-makers



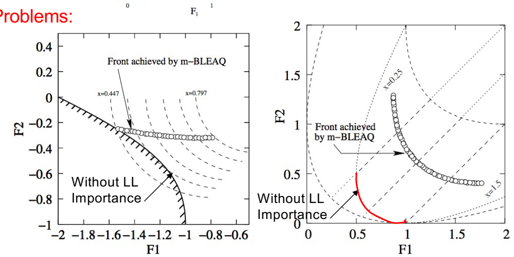
Importance of LL Decision-Maker

- LL Decision-maker can make a decision on her/his own
- M-BLEAQ

(Sinha, Malo, Deb, ECJ 2014)



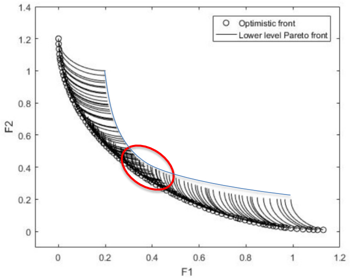
Two Difference Problems:

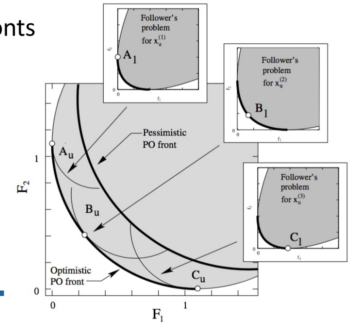


Variation in Expectation

(Sinha, Malo and Deb, 2016)

- Optimistic PO front: No power on LL DM
- Pessimistic PO front: Complete power on LL DM
  - Leader optimizes worst outcome from LL
- The difference between optimistic and pessimistic fronts provide DM ideas at UL





Bilevel Optimization with Uncertainties

- Uncertainty is, in most cases, inevitable in practical applications.
- Sources of uncertainties:
  - Imperfect implementation, changing environment, etc.
- In the context of bilevel optimization problems
  - Uncertainty in design variables and parameters
  - Uncertainty in objective and constraint function computations (Noise)
  - Uncertainty in decision making information
  - Uncertainty in control of decision-making preferences between two levels
- Uncertainties in the context of bilevel optimization have *NOT* been formally studied
- Clear mathematical definitions and formulations of robust/reliable bilevel solutions do *NOT* exist

Robust Bilevel Optimization

(Lu, Deb and Sinha, 2018)

- Both upper and lower-level variables are uncertain within their neighborhoods: Type-I Robustness:

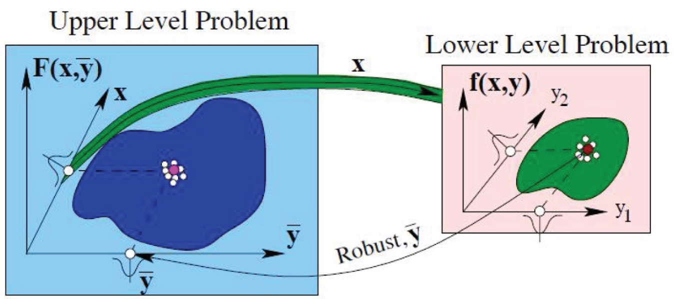
$$\begin{array}{rl} \text{Min.}_{(\mathbf{x},\mathbf{y})} & F^{\text{eff}}(\mathbf{x},\mathbf{y}), \\ \text{s.t.} & \mathbf{y} \in \text{argmin}_{(\mathbf{y})} \{ f^{\text{eff}}(\mathbf{x},\mathbf{y}) | g_j(\mathbf{x}+\Delta\mathbf{x},\mathbf{y}+\Delta\mathbf{y}) \le 0, \\ & \quad \forall\, \Delta\mathbf{x} \in \mathcal{B}_{\delta\mathbf{x}}, \Delta\mathbf{y} \in \mathcal{B}_{\delta\mathbf{y}},\ j = 1,\ldots,J_L \}, \\ & G_j(\mathbf{x}+\Delta\mathbf{x},\mathbf{y}+\Delta\mathbf{y}) \le 0,\ \forall\, \Delta\mathbf{x} \in \mathcal{B}_{\delta\mathbf{x}}, \Delta\mathbf{y} \in \mathcal{B}_{\delta\mathbf{y}}, \\ & \quad j = 1,\ldots,J_U. \end{array}$$

$$f^{\text{eff}}(\mathbf{x},\mathbf{y}) = \frac{1}{|(\mathcal{B}_{\delta\mathbf{x}},\mathcal{B}_{\delta\mathbf{y}})|} \int_{\mathbf{z}\in(\mathbf{x},\mathbf{y})+(\mathcal{B}_{\delta\mathbf{x}},\mathcal{B}_{\delta\mathbf{y}})} f(\mathbf{z})d\mathbf{z},$$

$$F^{\text{eff}}(\mathbf{x},\mathbf{y}) = \frac{1}{|(\mathcal{B}_{\delta\mathbf{x}},\mathcal{B}_{\delta\mathbf{y}})|} \int_{\mathbf{z}\in(\mathbf{x},\mathbf{y})+(\mathcal{B}_{\delta\mathbf{x}},\mathcal{B}_{\delta\mathbf{y}})} F(\mathbf{z})d\mathbf{z}.$$

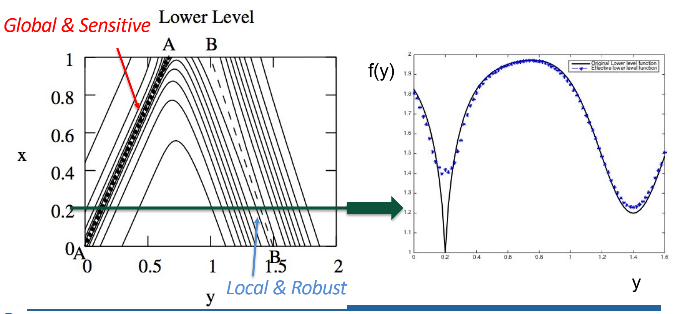
Note that even if $\Delta\mathbf{y} = 0$, LL is uncertain due to $\Delta\mathbf{x}$ perturbation, stays as parameter uncertainties at LL

Bilevel Optimization with Uncertainties: Big Picture

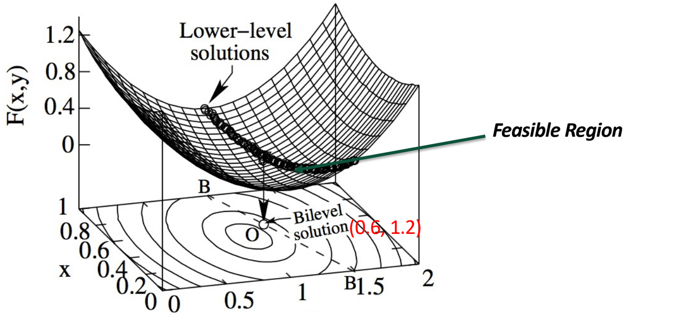


Robust Bi-Level Optimization

Lower-level variables are uncertain



Robustness-based (Cont.)

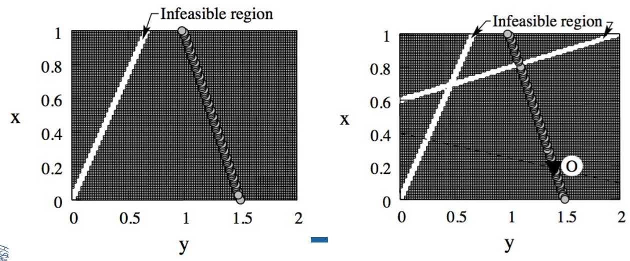


(b) Robust bilevel solutions for UnCaseB.

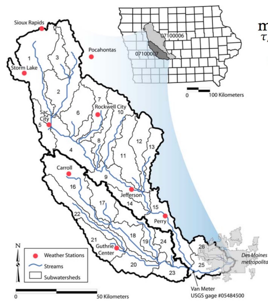
2-Variable		ULFV	LLFV	ULVS	LLVS
Grid		0.0474	1.2169	0.60	1.18
BLEAQ	Best	0.0473	1.1906	0.5992	1.1819
	Median	0.0459	1.1848	0.5923	1.1908
	Worst	0.0439	1.1826	0.5442	1.2116

Type-II Robustness

$$
\begin{aligned}
\text{Min.}_{(\mathbf{x},\mathbf{y})} \quad & F(\mathbf{x},\mathbf{y}), \\
\text{s.t.} \quad & \mathbf{y} = \operatorname{argmin}_{(\mathbf{y})} \left\{ \begin{array}{l} f(\mathbf{x},\mathbf{y}) \Big| \frac{|f^{\text{eff}}(\mathbf{x},\mathbf{y}) - f(\mathbf{x},\mathbf{y})|}{|f(\mathbf{x},\mathbf{y})|} \le \eta_L, \\ g_j(\mathbf{x}+\Delta\mathbf{x}, \mathbf{y}+\Delta\mathbf{y}) \le 0, \\ \forall\, \Delta\mathbf{x} \in \mathcal{B}_{\delta\mathbf{x}}, \Delta\mathbf{y} \in \mathcal{B}_{\delta\mathbf{y}},\ j = 1, \ldots, J_L \end{array} \right\} \\
& \frac{|F^{\text{eff}}(\mathbf{x},\mathbf{y}) - F(\mathbf{x},\mathbf{y})|}{|F(\mathbf{x},\mathbf{y})|} \le \eta_U, \\
& G_j(\mathbf{x}+\Delta\mathbf{x}, \mathbf{y}+\Delta\mathbf{y}) \le 0,\ \forall\, \Delta\mathbf{x} \in \mathcal{B}_{\delta\mathbf{x}}, \Delta\mathbf{y} \in \mathcal{B}_{\delta\mathbf{y}}, \\
& j = 1, 2, \ldots, J_U.
\end{aligned}
$$



IOWA Raccoon Watershed

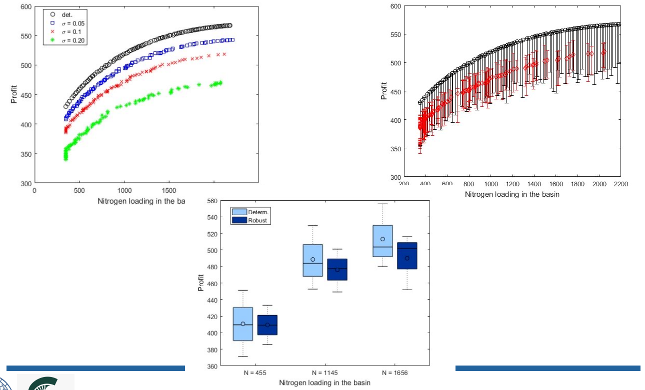


$$
\begin{aligned}
\max_{\tau,x,y} \quad & F(\tau, x) = \big(\Pi(\tau, x_N),\, B(x_N)\big) \\
\text{s.t.} \quad & (x^k, y^k) \in \operatorname*{argmax}_{x^k,\, y^k} \{\pi^k(\tau^k, x^k, y^k) : (\tau^k, x^k, y^k) \in \Omega^k\} \\
& \qquad \forall\, k \in \{1, \ldots, K\}, \\
& x_n^k \ge 0, \forall\, k \in \{1, \ldots, K\}, n \in \{1, \ldots, N\}, \\
& y_m^k \ge 0, \forall\, k \in \{1, \ldots, K\}, m \in \{1, \ldots, M\}, \\
& \tau^k \ge 1, \forall\, k \in \{1, \ldots, K\},
\end{aligned}
$$

$$
\begin{aligned}
\max_{x^k, y^k} \pi^k(p^k, w^k, \tau^k) = \; & p^k y^k - \sum_{n=1}^{N-1} w_n x_n^k - \tau_N^k w_N x_N^k \\
\text{s.t.} \quad & y^k \le P^k(x^k),
\end{aligned}
$$

IOWA Raccoon Watershed Results

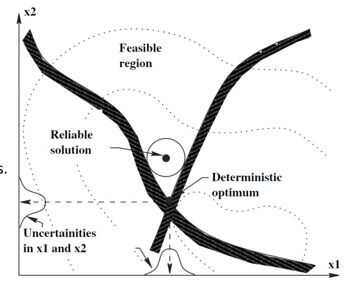
(GECCO-2017)



Constrained Bilevel Optimization

(Lu, Deb and Sinha, 2018)

- Constraints exist in almost every practical engineering design problem, and play a critical role in deciding the optimal solution.
- The deterministic optimum usually lies on a constraint surface or at the intersection of constraint surfaces.
  - Fail to remain feasible in many occasions.
- Many studies aim to handle this issue in single level optimization, none yet in the domain of bilevel optimization.



Reliability in Bilevel Problem

- Reliable bilevel solution definition:

$$\begin{aligned} \text{Minimize}_{(\mathbf{x},\mathbf{y})} \quad & F(\mathbf{x},\mathbf{y}), \\ \text{subject to} \quad & \mathbf{y} \in \text{argmin}_{(\mathbf{y})} \left\{ f(\mathbf{x},\mathbf{y}) | (P(\textstyle\bigwedge_{j=1}^{J_L} g_j(\mathbf{x},\mathbf{y}) \le 0)) \ge r \right\}, \\ & (P(\textstyle\bigwedge_{j=1}^{J_U} G_j(\mathbf{x},\mathbf{y}) \le 0)) \ge R. \end{aligned}$$

- *P ()* signify the joint probability of the solution (x, y) being feasible for all constraints.
- The effect of uncertainties in lower, upper or both levels are different because of the unequal importance of each level.
- Test problems proposed for the purpose of concept demonstration, *NOT* for algorithm performance assessment.

A Scenario

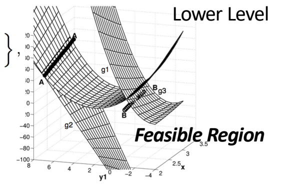
- Constraint functions at both levels

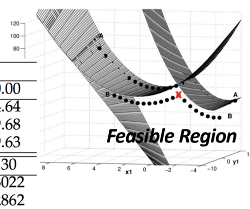
$$\begin{aligned} \text{Maximize}_{(\mathbf{x},\mathbf{y})} \quad & F(\mathbf{x},\mathbf{y}) = x_2, \\ \text{subject to} \quad & \mathbf{y} \in \text{argmax}_{(\mathbf{y})} \left\{ \begin{array}{l} f(\mathbf{x},\mathbf{y}) = y_2, \\ g_j(\mathbf{x},\mathbf{y}) \ge 0, \quad j = 1,2,3 \end{array} \right\}, \\ & G_j(\mathbf{x},\mathbf{y}) \ge 0, \quad j = 1,2,3, \\ & -4 \le x_1 \le 10, -100 \le x_2 \le 200, \\ & -4 \le y_1 \le 10, -100 \le y_2 \le 200. \end{aligned}$$

$$\begin{aligned} G_1(\mathbf{x},\mathbf{y}) &= (\tfrac{y_1}{14} + \tfrac{16}{7})(x_1 - 2)^2 - x_2, \\ G_2(\mathbf{x},\mathbf{y}) &= x_2 - 12.5(\tfrac{y_1}{14} + \tfrac{16}{7})(x_1 - 5), \\ G_3(\mathbf{x},\mathbf{y}) &= 5(x_1 + 4 - (\tfrac{y_1}{14} + \tfrac{16}{7}))(x_1 + 8 - (\tfrac{y_1}{14} + \tfrac{16}{7})) - x_2, \end{aligned}$$

$$\begin{aligned} g_1(\mathbf{x},\mathbf{y}) &= (\tfrac{x_1}{14} + \tfrac{16}{7})(y_1 - 2)^2 - y_2, \\ g_2(\mathbf{x},\mathbf{y}) &= y_2 - 12.5(\tfrac{x_1}{14} + \tfrac{16}{7})(y_1 - 5), \\ g_3(\mathbf{x},\mathbf{y}) &= 5(y_1 + 4 - (\tfrac{x_1}{14} + \tfrac{16}{7}))(y_1 + 8 - (\tfrac{x_1}{14} + \tfrac{16}{7})) - y_2. \end{aligned}$$

4-Variable			ULFV	LLFV	ULVS		LLVS	
Determ.	Grid		70.00	70.00	7.00	70.00	7.00	70.00
	BLEAQ	Best	69.64	64.64	7.00	69.64	7.00	64.64
		Median	69.61	69.68	6.9988	69.61	7.0011	69.68
		Worst	69.50	69.63	6.9952	69.50	7.0001	69.63
Reliable	Grid		0.32	0.32	−0.80	0.30	−0.80	0.30
	BLEAQ	Best	0.3022	0.3022	−0.7917	0.3022	−0.7917	0.3022
		Median	0.2745	0.2862	−0.7859	0.2745	−0.7903	0.2862
		Worst	0.2354	0.3009	−0.7859	0.2354	−0.7914	0.3009

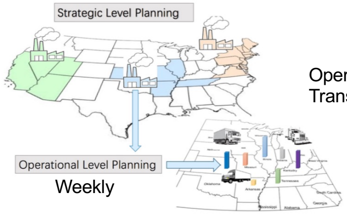




Tri-Level Optimization

- Three levels of optimization problems interlinked by two consecutive levels
- Min **$\mathcal{F}$(x,y,z)**
  - Min **F(y,z)**, given **x**
    - Min **f(z)**, given **x** and **y**
- Constraints are expected at every level
- To make an application realistic, we need to replace lowest level heuristic/rule based
- Not much work available, but all issues discussed before are applicable here too
  - BLEAQ can be extended
  - Currently pursuing

A Case Study: Supply Chain Management



Strategic Level Transport Cost

$$\text{Min.}\quad F(\mathbf{x},\mathbf{y}) = \sum_m \sum_n f_{m,n}(\mathbf{x},\mathbf{y}),$$

$$\text{s.t.}\quad \mathbf{y} \in \operatorname{argmin}\Big\{ f_{m,n}(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{c} f'_j(\mathbf{x},\mathbf{y}_j) \mid,$$

Operational Level Transport Cost

$$\sum_{i=1}^{d} w_i \cdot \mathbf{y}_{ij} \le C_j,\ i = 1,\ldots,d(estinations),$$

$$\sum_{i=j}^{c} \mathbf{y}_{ij} = D_i,\ j = 1,\ldots,c(arriers),$$

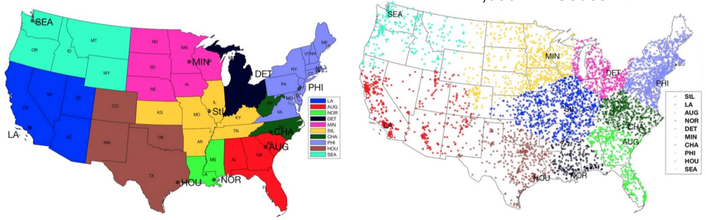
$$f'_j(\mathbf{x},\mathbf{y}_j) = \max\Big(\sum_{i=1}^{d} DD_i(\mathbf{x}) \cdot \mathbf{y}_{ij} \cdot CC_j,\ \tau_j\Big)\Big\},$$

$$m = 1,2,\ldots,p(lants),\quad n = 1,2,\ldots,52\ weeks.$$

	Trucks	Cost	Service	Min. Charge
Carrier1	7	\$1.93	0.90	\$4050
Carrier2	8	\$1.82	0.82	\$4370
Carrier3	6	\$1.88	0.86	\$3380
Carrier4	9	\$1.97	0.93	\$5320
Carrier5	5	\$1.94	0.91	\$2910
Carrier6*	N.A.	\$2.39	0.88	N.A.

- UL: Plant to destination layput
- LL: At each plant, carrier companies are pre-determined
- Allocation of goods to carrier is LL task

State-Level and Zip-Level Results

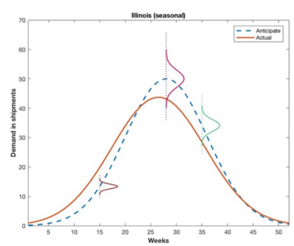


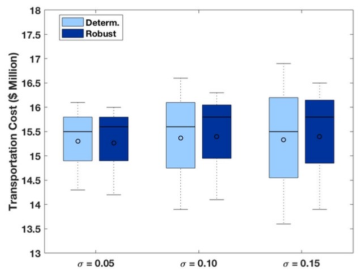
Advantage of Using Bilevel Optimization Over Single-Level Optimization:

Model	Single-level	Bilevel
Transportation Cost	\$ 16.2M	**\$ 15.5M** (4.3%)

Uncertainty in Demand

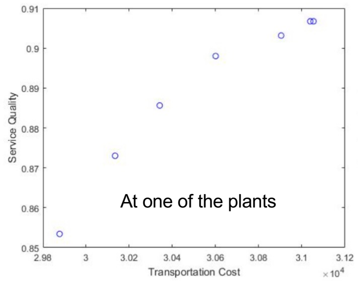
- Seasonal demand of goods assumed with uncertainty
- Robust bi-level optimization performed
- Robust solution is able to handle uncertainties better than the deterministic solution





Bi-Objective at Operational Level

- Operational level at each plant considers two objectives:
  - Transport cost
  - Service quality obtained carrier companies: $f^{(2)}_{m,n}(\mathbf{x},\mathbf{y}) = \sum_{j=1}^{c}\sum_{i=1}^{d} \mathbf{y}_{ij}\cdot CS_j$.
- Produces a PO front at each plant
- Strategic level chooses the best overall Transport Cost



Trade-off:

Model	Single-objective	Bi-objective
Transportation Cost	**$ 15.5M**	$ 15.9M
Service Quality*	0.86	**0.89**

Average service quality over all plants

109

Tri-Level Consideration

$$\text{Min.}\quad F(\mathbf{x},\mathbf{y},\mathbf{z}) = \sum_k f_k(\mathbf{x},\mathbf{y},\mathbf{z}), \qquad \text{(Strategic level)}$$

Yearly, Plant to destination layout

where $\mathbf{y},\mathbf{z}$ solve:

$$\text{Min.}\quad f_k(\mathbf{x},\mathbf{y},\mathbf{z}) = \sum_m\sum_n f_{m,n}(\mathbf{x},\mathbf{y},\mathbf{z}), \qquad \text{(Tactical level)}$$

Quarterly, Choice of Carriers

where $\mathbf{z}$ solves:

$$\text{Min.}\quad f_{m,n}(\mathbf{x},\mathbf{y},\mathbf{z}) = \sum_{j=1}^{c} f_j'(\mathbf{x},\mathbf{y},\mathbf{z}_j), \qquad \text{(Operational level)}$$

Daily, Allocation of goods to trucks

$$\text{s.t.}\quad \Big\{ \sum_{i=1}^{d} w_i\cdot \mathbf{z}_{ij} \le C_j(\mathbf{y}),\ i=1,\ldots,d(estinations),$$

$$\sum_{i=j}^{c} \mathbf{z}_{ij} = D_i,\ j=1,\ldots,c(arriers),$$

$$f_j'(\mathbf{x},\mathbf{y},\mathbf{z}_j) = \max\big(\sum_{i=1}^{d} DD_i(\mathbf{x})\cdot \mathbf{z}_{ij}\cdot CC_j(\mathbf{y}),\ \tau_j\big)\Big\}.$$

$$k=1,2,\ldots,t(erms),\quad m=1,2,\ldots,p(lants),\quad n=1,2,\ldots,\frac{52\ weeks}{terms}.$$

Advantage of Using Tri-level Optimization:

Finer optimization

Model	Single-level	Bilevel	Tri-level
Transportation Cost	$ 16.2M	$ 15.5M (4.3%)	**$ 15.3M** (5.5%)

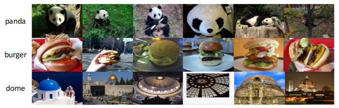
Most benefit

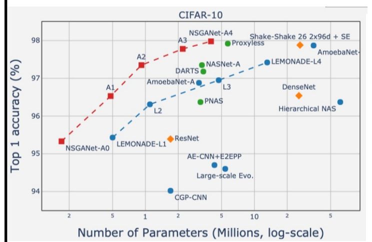
110

Bilevel Application in AI/ML

Neural Architecture Search (NAS):

$$\begin{aligned} \text{minimize}\quad & \mathbf{F}(\boldsymbol{\alpha}) = \big(f_1(\boldsymbol{\alpha};\boldsymbol{w}^*(\boldsymbol{\alpha})),\ldots,f_k(\boldsymbol{\alpha};\boldsymbol{w}^*(\boldsymbol{\alpha})),f_{k+1}(\boldsymbol{\alpha}),\ldots,f_m(\boldsymbol{\alpha})\big)^T,\\ \text{subject to}\quad & \boldsymbol{w}^*(\boldsymbol{\alpha}) \in \operatorname{argmin} \mathcal{L}(\boldsymbol{w};\boldsymbol{\alpha}),\\ & \boldsymbol{\alpha}\in\boldsymbol{\Omega}_\alpha,\quad \boldsymbol{w}\in\boldsymbol{\Omega}_w, \end{aligned}$$





(b) NSGANet-A1 (c) NSGANet-A2 (d) NSGANet-A4

111

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
- Extension to Tri-Level optimization is needed
- Application to industry would be beneficial

112

Some References

- J.F. Bard. Practical bilevel optimization: Algorithms and applications. The Netherlands: Kluwer, 1998.
- L. Brotcorne, M. Labbé, P. Marcotte, and G. Savard. 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.
- S. Dempe, J. Dutta, and B.S. Mordukhovich (2007). New necessary optimality conditions in optimistic bilevel programming. Optimization, 56(5-6), 577–604.
- J. Fliege, and L.N. Vicente (2006). Multicriteria approach to bilevel optimization. Journal of Optimization Theory and Applications, 131(2), 209–225.
- 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. An improved bilevel evolutionary algorithm based on quadratic approximations. In 2014 IEEE Congress on Evolutionary Computation (CEC-2014). IEEE Press, 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.