Nonlinear Bilevel Programming

Alain B. Zemkoho

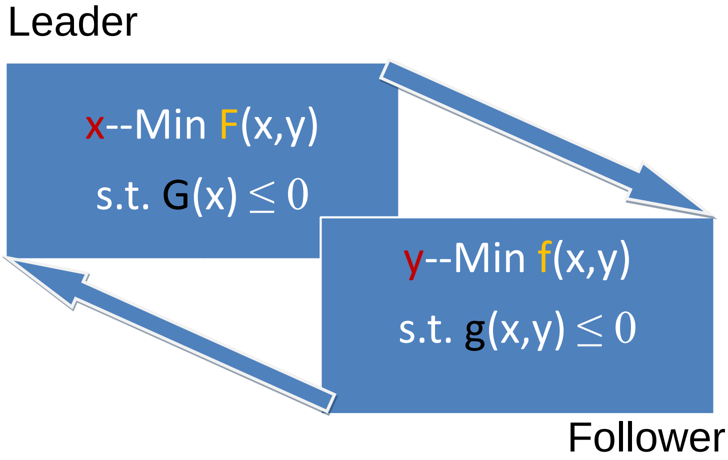
School of Mathematics, University of Birmingham (UK)

International Conference on Complementarity Problems

August 4–8, 2014, Berlin, Germany

Table of contents

Conceptual definition of a bilevel program



Mathematical model of the bilevel program

$$
\begin{aligned}
&\text{“}\min_{x}\text{”} \;\; F(x,y) \\
&\text{s.t.} \;\; x \in X := \{x \in \mathbb{R}^n : \;\; G(x) \leq 0\}, \\
&\qquad\; y \in S(x) := \arg\min_{y}\{f(x,y) : \;\; g(x,y) \leq 0\}.
\end{aligned}
$$

- x (resp. y) is the upper (resp. lower)–level variable.
- F (resp. f) is the upper (resp. lower)–level objective function.
- $G(x) \leq 0$ (resp. $g(x,y) \leq 0$) is the upper (resp. lower)–level constraint.

Mathematical model of the bilevel program

$$
\begin{aligned}
&\text{“}\min_x\text{”} \;\; F(x,y)\\
&\text{s.t.} \;\; x \in X := \{x \in \mathbb{R}^n : \;\; G(x) \leq 0\},\\
&\qquad\; y \in S(x) := \arg\min_y \{f(x,y) : \;\; g(x,y) \leq 0\}.
\end{aligned}
$$

- x (resp. y) is the upper (resp. lower)–level variable.
- F (resp. f) is the upper (resp. lower)–level objective function.
- $G(x) \leq 0$ (resp. $g(x,y) \leq 0$) is the upper (resp. lower)–level constraint.

Uniquely defined lower-level solution:

$$S(x) = \{y(x)\}, \quad \forall x \in X.$$

The problem is well-defined:

$$\min_x \mathcal{F}(x) := F(x, y(x)) \quad \text{s.t.} \;\; G(x) \leq 0.$$

Mathematical model of the bilevel program (contd ...)

Non-uniqueness in the lower-level problem:

$$S : X \rightrightarrows \mathbb{R}^m$$

- Original optimistic formulation

$$(\mathrm{P}_o) \qquad \min_{x \in X} \varphi_o(x) := \min_{y} \{F(x,y) |\ y \in S(x)\}$$

- Pessimistic formulation

$$(\mathrm{P}_p) \qquad \min_{x \in X} \varphi_p(x) := \max_{y} \{F(x,y) |\ y \in S(x)\}$$

Existence and approximation results for these models are available in the literature, see, e.g., Loridan and Morgan (1989, 1996).

Standard optimistic bilevel program

$$
(\mathrm{P}) \qquad \begin{array}{ll} \min\limits_{x,y} & F(x,y) \\ \text{s.t.} & x \in X := \{x \in \mathbb{R}^n : \ G(x) \leq 0\}, \\ & y \in S(x) := \arg\min\limits_{y}\{f(x,y) : \ g(x,y) \leq 0\}. \end{array}
$$

- Leader in control of both x and y.
- Model considered in most works in the literature.

More details on the reformulations can be found in Bard (1998), Dempe (2002, 2003), Marcotte et al. (2007).

Standard optimistic bilevel program

$$
\text{(P)} \qquad \begin{array}{ll} \min\limits_{x,y} & F(x,y) \\ \text{s.t.} & x \in X := \{x \in \mathbb{R}^n : \ G(x) \leq 0\}, \\ & y \in S(x) := \arg\min\limits_{y}\{f(x,y) : \ \ g(x,y) \leq 0\}. \end{array}
$$

- Leader in control of both x and y.
- Model considered in most works in the literature.

More details on the reformulations can be found in Bard (1998), Dempe (2002, 2003), Marcotte et al. (2007).

Some theoretical issues in bilevel programming

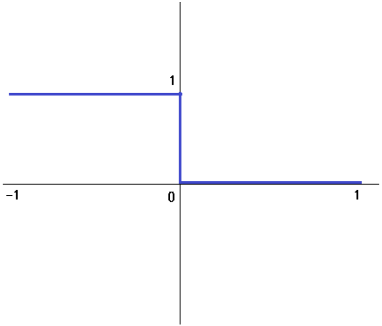
1. How are the above models related to each other?
2. How does standard CQs (e.g., MFCQ) behave?
3. How can the standard theory known in NLP be translated into bilevel programming? (e.g., optimality conditions, stability analysis)

On the link between the models

- (P$_o$) and (P$_p$) different from each other when $|S(x)| > 1$.
- Link between (P) and (P$_o$):

$$F(x,y) := x; \quad X := [-1,1]$$

$$S(x) := \arg\min_y \{xy \mid y \in [0,1]\}$$
$$= \begin{cases} [0,1] & \text{if } x = 0 \\ \{0\} & \text{if } x > 0 \\ \{1\} & \text{if } x < 0 \end{cases}$$



S_o is i.s.c. at $(0,0)$: $\forall x^k \to 0, \ \exists y^k \in S_o(x^k)$ s.t. $y^k \to 0$.

See Z. (2012) for more details.

One-level reformulations for (P)

- KKT reformulation:

$$\begin{aligned}\min_{x,y,u}\ & F(x,y)\\ \text{s.t.}\ \ & x\in X,\ \nabla_y f(x,y)+\textstyle\sum_{i=1}^{p} u_i\nabla_y g_i(x,y)=0,\\ & g(x,y)\le 0,\quad u\ge 0,\quad u^\top g(x,y)=0.\end{aligned}$$

- LLVF reformulation:

$$\begin{aligned}\min_{x,y}\ & F(x,y)\\ \text{s.t.}\ \ & x\in X,\ g(x,y)\le 0,\quad f(x,y)\le\varphi(x).\end{aligned}$$

One-level reformulations for (P)

- KKT reformulation:

$$\begin{aligned}&\min_{x,y,u} \ F(x,y)\\ &\text{s.t.} \quad x \in X,\ \nabla_y f(x,y) + \sum_{i=1}^{p} u_i \nabla_y g_i(x,y) = 0,\\ &\qquad\ \ g(x,y) \le 0,\quad u \ge 0,\quad u^\top g(x,y) = 0.\end{aligned}$$

- LLVF reformulation:

$$\begin{aligned}&\min_{x,y} \ F(x,y)\\ &\text{s.t.} \quad x \in X,\ g(x,y) \le 0,\quad f(x,y) \le \varphi(x).\end{aligned}$$

Which reformulation is the best for a solution process?

One-level reformulations for (P)

- KKT reformulation:

$$\begin{aligned} &\min_{x,y,u} \; F(x,y) \\ &\text{s.t.} \;\; x \in X, \; \nabla_y f(x,y) + \textstyle\sum_{i=1}^{p} u_i \nabla_y g_i(x,y) = 0, \\ &\qquad g(x,y) \leq 0, \quad u \geq 0, \quad u^\top g(x,y) = 0. \end{aligned}$$

- LLVF reformulation:

$$\begin{aligned} &\min_{x,y} \; F(x,y) \\ &\text{s.t.} \;\; x \in X, \; g(x,y) \leq 0, \quad f(x,y) \leq \varphi(x). \end{aligned}$$

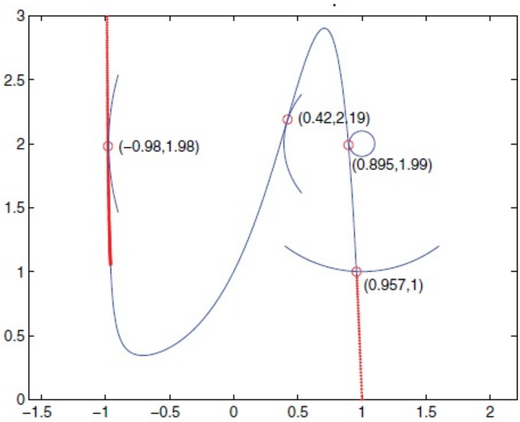
Which reformulation is the best for a solution process?

- KKT reformulation: (1) Convexity; (2) CQ in lower-level; (3) Ambiguity in the link with (P); (4) Demanding in terms of derivatives.
- LLVF reformulation: (1) Nonsmoothness; (2) Implicity nature of φ.

Why is convexity needed?

Mirrlees problem (1999):

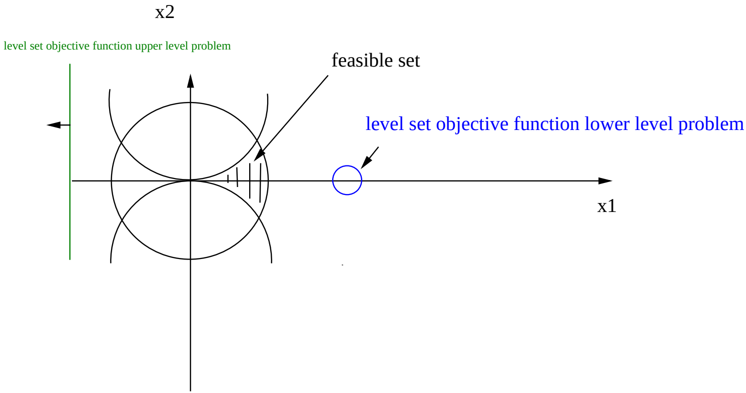
$$\min_{x,y} \; (x-2)^2 + (y-1)^2 \;\text{ s.t. }\; \min_{y} \; -xe^{-(y+1)^2} - e^{-(y-1)^2}$$



Remark: Observe that there is no lower-level constraint in this case.

Why is a constraint qualification required?

Allende and Still (2013): replace KKT conditions with FJ conditions.



$$\min\ x_1\ \text{ s.t. }\ \min\ \{\|x-(2,\ 0)^\top\| :\ \|x\| \le 1,\ x_2 \le x_1^2,\ x_2 \ge -x_1^2\}$$

See Dempe & Z. (2014) for details.

Link between KKT reformulation and (P)

When we have convexity and a CQ in the lower-level, then:

- (KKT) and (P) are globally equivalent in some sense;
- For $(\bar{x}, \bar{y})$ to be a local optimal solution of (P), $(\bar{x}, \bar{y}, u)$ has to be a local optimal solution of (KKT) for all $u \in \Lambda(\bar{x}, \bar{y})$;
- Hence, things would be great if LICQ is satisfied in the lower-level. But this condition is not generic in parametric optimization (see Dempe & Dutta 2012).

Link between KKT reformulation and (P)

When we have convexity and a CQ in the lower-level, then:

- (KKT) and (P) are globally equivalent in some sense;
- For $(\bar{x}, \bar{y})$ to be a local optimal solution of (P), $(\bar{x}, \bar{y}, u)$ has to be a local optimal solution of (KKT) for all $u \in \Lambda(\bar{x}, \bar{y})$;
- Hence, things would be great if LICQ is satisfied in the lower-level. But this condition is not generic in parametric optimization (see Dempe & Dutta 2012).

An alternative to the KKT reformulation would be the GE reformulation:

$$\min_{x,y} F(x,y) \text{ s.t. } x \in X, \quad 0 \in \nabla_y f(x,y) + N_{K(x)}(y)$$

- Burdens that remain: convexity & higher order derivatives

See, e.g., Henrion & Surowiec (2010), Mordukhovich & Outrata (2007) and Dempe & Zemkoho (2012)

Optimality conditions via the LLVF reformulation

Stage 1. Partial calmness: The exists λ such that

$$(\mathrm{P}) \quad \Longleftrightarrow \quad \begin{array}{l} \min_{x,y} \; F(x,y) + \lambda(f(x,y) - \varphi(x)) \\ \text{s.t.} \;\; x \in X, \; g(x,y) \le 0. \end{array}$$

Optimality conditions via the LLVF reformulation

Stage 1. Partial calmness: The exists λ such that

$$(\mathrm{P}) \quad \Longleftrightarrow \quad \begin{array}{ll} \min\limits_{x,y} & F(x,y) + \lambda(f(x,y) - \varphi(x)) \\ \text{s.t.} & x \in X,\ g(x,y) \le 0. \end{array}$$

Partial calmness automatically holds when lower-level linear w.r.t. y, see Dempe & Z. (2013).

Optimality conditions via the LLVF reformulation

Stage 1. Partial calmness: The exists λ such that

$$(\mathrm{P}) \quad \Longleftrightarrow \quad \begin{array}{ll} \min\limits_{x,y} & F(x,y) + \lambda(f(x,y) - \varphi(x)) \\ \text{s.t.} & x \in X,\ g(x,y) \leq 0. \end{array}$$

Partial calmness automatically holds when lower-level linear w.r.t. y, see Dempe & Z. (2013).

Stage 2. Use of standard CQ on remaining constraints.

Optimality conditions

The KKT conditions depend on the estimate of $\partial\varphi$; see, e.g., Ye & Zhu (1995), Dempe, Dutta & Mordukhovich (2007) and Dempe & Z. (2013)

- φ convex;
- φ non-convex.

Extending the results to the nonsmooth case is relatively easy; see, e.g. Dempe, Dutta & Mordukhovich (2007) and Dempe & Z. (2013).

Optimality conditions via (KKT): smooth case

$$\nabla F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j (\nabla G_j(\bar{x}), 0) + \sum_{i=1}^{p} \beta_i \nabla g_i(\bar{x},\bar{y})$$

$$+\sum_{l=1}^{m} \gamma_l \left[\nabla_{(x,y)}(\nabla_{y_l} f)(\bar{x},\bar{y}) + \sum_{i=1}^{p} u_i \nabla_{(x,y)}(\nabla_{y_l} g_i)(\bar{x},\bar{y})\right] = 0,$$

$$\nabla_y g_\nu(\bar{x},\bar{y})\gamma = 0,\ \beta_\eta = 0.$$

C-, M- and S-stationarity respectively determined by

$\forall i \in \theta :\ \beta_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) \geq 0,$

$\forall i \in \theta : \left(\beta_i > 0 \wedge \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) > 0\right) \vee \beta_i \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) = 0,$

$\forall i \in \theta :\ \beta_i \geq 0 \wedge \sum_{l=1}^{m} \gamma_l \nabla_{y_l} g_i(\bar{x},\bar{y}) \geq 0.$

Optimality conditions via (KKT): nonsmooth case

$$\min_{x,y,u} \{F(x,y) | \quad 0 \in \mathcal{L}(x,y,u),\ G_j(x) \leq 0,\ j = 1, \ldots, k,$$
$$u_i \geq 0,\ g_i(x,y) \leq 0,\ u_i g_i(x,y) = 0,\ i = 1, \ldots, p\}.$$

Optimality conditions via (KKT): nonsmooth case

$$\min_{x,y,u} \{F(x,y) | \quad 0 \in \mathcal{L}(x,y,u),\ G_j(x) \le 0,\ j = 1,\dots,k,$$
$$u_i \ge 0,\ g_i(x,y) \le 0,\ u_i g_i(x,y) = 0,\ i = 1,\dots,p\}.$$

$$0 \in \partial F(\bar{x},\bar{y}) + \sum_{j=1}^{k} \alpha_j \big(\partial G_j(\bar{x}), 0_m\big) + \partial\langle \beta, g\rangle(\bar{x},\bar{y})$$

$$+ D^*(\partial_y f)((\bar{x},\bar{y})|t^0)(\gamma) + \sum_{i=1}^{p} D^*(\partial_y g_i)((\bar{x},\bar{y})|t^i)(\bar{u}_i\gamma),$$

$$\forall i \in \nu : \sum_{l=1}^{m} t_l^i \gamma_l = 0,\ \beta_\eta = 0.$$

M-stationarity and S-stationarity:

$\forall i \in \theta$: $\big(\beta_i > 0 \wedge \sum_{l=1}^m t_l^i \gamma_l > 0\big) \vee \beta_i \big(\sum_{l=1}^m t_l^i \gamma_l\big) = 0,$

$\forall i \in \theta$: $\beta_i \ge 0 \wedge \sum_{l=1}^m t_l^i \ge 0.$

Here, $t^0 \in \partial_y f(\bar{x},\bar{y})$, $t^i \in \partial_y g_i(\bar{x},\bar{y})$, $i = 1,\dots,p$, $t^0 + \sum_{i=1}^p \bar{u}_i t^i = 0.$

On the S-stationarity

$$\beta_i \bar{\partial} g_i(\bar{x}, \bar{y}) - \xi \bar{u}_i \bar{\partial} g_i(\bar{x}, \bar{y}) \supset (\beta_i - \xi \bar{u}_i) \bar{\partial} g_i(\bar{x}, \bar{y})$$

On the S-stationarity

$$\beta_i \bar{\partial} g_i(\bar{x},\bar{y}) - \xi \bar{u}_i \bar{\partial} g_i(\bar{x},\bar{y}) \supset (\beta_i - \xi \bar{u}_i)\bar{\partial} g_i(\bar{x},\bar{y})$$

For the example

$$F(x,y) := |x-y|,\ G(x) := -x,\ f(x,y) := \max\{x,y\},\ g(x,y) := |y| - x,$$

$(0,0)$ is an optimal solution and $\partial g(0,0) = \text{co}\,\{(-1,1)^\top, (-1,-1)^\top\}$.

$$D^*(\partial_y f)((\bar{x},\bar{y})|\bar{z})(z^*) = \begin{cases} \{(x,-x):\ x \in \mathbb{R}\} & \text{if} \quad \bar{x}=\bar{y},\ 0<\bar{z}<1,\ z^*=0, \\ \emptyset & \text{if} \quad \bar{x}=\bar{y},\ 0<\bar{z}<1,\ z^*\neq 0, \\ \{(x,-x):\ x \in \mathbb{R}\} & \text{if} \quad \bar{x}=\bar{y},\ \bar{z}=0,\ z^*=0, \\ \{(x,-x):\ x < 0\} & \text{if} \quad \bar{x}=\bar{y},\ \bar{z}=0,\ z^*>0, \\ \{(0,0)\} & \text{if} \quad \bar{x}=\bar{y},\ \bar{z}=0,\ z^*<0, \\ \{(x,-x):\ x \in \mathbb{R}\} & \text{if} \quad \bar{x}=\bar{y},\ \bar{z}=1,\ z^*=0, \\ \{(x,-x):\ x > 0\} & \text{if} \quad \bar{x}=\bar{y},\ \bar{z}=1,\ z^*<0, \\ \{(0,0)\} & \text{if} \quad \bar{x}=\bar{y},\ \bar{z}=1,\ z^*>0. \end{cases}$$

S-stationarity holds at $(0,0)$ with $\bar{u}=1$, $\kappa=1$, $\xi=\beta$, $\alpha=1$ and $\beta=2$. But

$$\begin{bmatrix} 0 \\ 1 \end{bmatrix} = 2\begin{bmatrix} -1 \\ 1 \end{bmatrix} - 2\begin{bmatrix} -1 \\ \frac{1}{2} \end{bmatrix}.$$

See Dempe & Z. (2014) for more details.

Original optimistic and pessimistic cases

The notion of two-level value function

$$\varphi_o(x) := \min_y \{F(x,y) |\ y \in S(x)\}$$

is introduced and studied in Dempe, Mordukhovich and Z. (2012). It leads to results on

- Optimality conditions for (P_o)
- Optimality conditions for (P_p)
- Stability analysis of the value functions of (P), (P_o) and (P_p)

See Dempe, Mordukhovich and Z. (2014) for optimality conditions in the pessimistic case.

Solution methods: implicit function approach

$$\min \mathcal{F}(x) := F(x, y(x)) \text{ s.t. } x \in X$$

A key reference is the book by Outrata, Kocvara & Zowe (1998):

$$\partial\mathcal{F}(\bar{x}) = \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y})\partial y(\bar{x})$$

- An estimate of $\partial y(\bar{x})$ is considered
- and used to build a bundle function.

Solution methods: implicit function approach

$$\min \mathcal{F}(x) := F(x, y(x)) \text{ s.t. } x \in X$$

A key reference is the book by Outrata, Kocvara & Zowe (1998):

$$\partial\mathcal{F}(\bar{x}) = \nabla_x F(\bar{x}, \bar{y}) + \nabla_y F(\bar{x}, \bar{y}) \partial y(\bar{x})$$

- An estimate of $\partial y(\bar{x})$ is considered
- and used to build a bundle function.

Other references include:

- Bundle-type method: Dempe (2002) and Falk & Liu (1995)
- Steepest descent-type methods: Kolstad & Ladson (1990), Savard & Gauvin (1994), Vicente et al. (1994) and Mersha & Dempe (2011)

Forcing uniqueness in the lower-level problem

In the case where $|S(x)| > 1$ for some values of $x \in X$:

$$\min_{y} f(x,y) + \alpha\pi(y) \quad \text{s.t.} \quad g(x,y) \leq 0 \qquad (\alpha > 0)$$

Example: $\pi(y) := \|y\|^2$ (Tikhonov regularization)

- See Dempe & Schmidt (1996), Dempe & Bard (2001), Morgan & Patrone (2006), Bergounioux & Haddou (2008), Molodtsov (1976), etc.

Forcing uniqueness in the lower-level problem

In the case where $|S(x)| > 1$ for some values of $x \in X$:

$$\min_y f(x,y) + \alpha\pi(y) \quad \text{s.t.} \quad g(x,y) \le 0 \qquad (\alpha > 0)$$

Example: $\pi(y) := \|y\|^2$ (Tikhonov regularization)

- See Dempe & Schmidt (1996), Dempe & Bard (2001), Morgan & Patrone (2006), Bergounioux & Haddou (2008), Molodtsov (1976), etc.

Regularizing can lead very far from the solution

$$X := [-2,2], \quad Y := [-1,1]$$

$$F(x,y) := -x - y; \text{ and}$$

$$f(x,y) := \begin{cases} (x+\frac{7}{4})y & \text{if } x \in [-2,-\frac{7}{4}] \\ 0 & \text{if } x \in [\frac{7}{4}, -\frac{7}{4}] \\ (x-\frac{7}{4})y & \text{if } x \in [\frac{7}{4}, 2] \end{cases}$$

- Solution from Tikhonov regularization: $(\frac{7}{4}, 0)$
- Optimistic solution: $(2,-1)$
- Pessimistic solution: $(\frac{7}{4}, 1)$

See Morgan & Patrone (2006) for more details.

Solution methods for (P): LLVF reformulation

Mitsos et al. (2008) and Kleniati & Adjiman (2013*a*, 2013*b*) have proposed methods to compute $(\epsilon_F, \epsilon_f)-$optimal points:

$$\begin{aligned} &F(\bar{x}, \bar{y}) - F^* < \epsilon_F, \\ &f(\bar{x}, \bar{y}) - \varphi(\bar{x}) \leq \epsilon_f, \\ &\bar{x} \in X, \;\; g(\bar{x}, \bar{y}) \leq 0. \end{aligned}$$

Branch-and-bound-type techniques are applied on relaxations of the LLVF reformulation in the form:

$$\begin{aligned} &\min_{x,y} \; F(x, y) \\ &\text{s.t.} \;\; x \in X, \; g(x, y) \leq 0, \\ &\quad f(x, y) \leq f(x, z), \;\; \forall z : \; g(x, z) \leq 0. \end{aligned}$$

These methods generate global optima.

Solution methods for (P): LLVF reformulation

Mitsos et al. (2008) and Kleniati & Adjiman (2013*a*, 2013*b*) have proposed methods to compute $(\epsilon_F, \epsilon_f)-$optimal points:

$$
\begin{aligned}
&F(\bar{x},\bar{y}) - F^* < \epsilon_F,\\
&f(\bar{x},\bar{y}) - \varphi(\bar{x}) \le \epsilon_f,\\
&\bar{x} \in X, \;\; g(\bar{x},\bar{y}) \le 0.
\end{aligned}
$$

Branch-and-bound-type techniques are applied on relaxations of the LLVF reformulation in the form:

$$
\begin{aligned}
&\min_{x,y}\; F(x,y)\\
&\text{s.t.}\;\; x \in X,\; g(x,y) \le 0,\\
&\quad f(x,y) \le f(x,z),\;\; \forall z:\; g(x,z) \le 0.
\end{aligned}
$$

These methods generate global optima.

Stationary points/Local optimal solution:

- Nonlinear problems (with $K(x) := Y$): Xu, Ye & Zhang (2014, 2014) and Xu & Ye (2014)
- Linear and quadratic problems: Strekalovsky, Orlov & Malyshev (2010) and Dempe & Franke (2014)

Solution methods for the pessimistic problem

Červinka, Matonoha & Outrata (2013) propose a method to compute relaxed approximate solutions:

- KKT reformulation
- UFO solver is used to evaluate

$$\min_y \{F(x,y): \quad y \in S(x)\} \qquad (:= \varphi_p(x))$$

  for fixed values of x
- BOBYQA (derivative–free optimization) solver is used for the outer problem (P_p)

Solution methods for the pessimistic problem

Wiesemann et al. (2013) propose the following method:

Stage 1. Difficulty moved to the constraints:

$$\begin{aligned} \min_{x,v} \ & v \\ \text{s.t.} \ & F(x,y) \le v, \ \ \forall y \in S(x) := \arg\min_y \{f(x,y) : \ y \in Y\} \\ & x \in X. \end{aligned}$$

Stage 2. Approximation of the LLVF reformulation:

$$\begin{aligned} \min_{x,v} \ & v \\ \text{s.t.} \ & F(x,y) \le v, \ \forall y \in \{z \in Y : f(x,z) - f(x,z') < \epsilon, \forall z' \in Y\} \\ & x \in X. \end{aligned}$$

Stage 3. Infinite optimization reformulation:

$$\begin{aligned} \min_{x,z,v} \ & v \\ \text{s.t.} \ & \lambda(y).[f(x,z) - f(x,y) + \epsilon] + (1 - \lambda(x)).g(x,y) \le 0, \ \forall y \in Y \\ & x \in X, \ z \in Y, \ \lambda : Y \mapsto [0,1]. \end{aligned}$$

Stage 4. A semi-infinite optimization technique is then used.

Conclusions

- The mathematical models of bilevel programs are now better understood.
- There has been a considerable development in the derivation of optimality conditions.
- Stability analysis for optimal value functions has started.
- Noting is known yet for the stability analysis of solution mappings.
- An important amount of solution schemes is available for MPECs/MPCCs, but for most of them, it is not yet clear whether they effectively generate optimal solutions for bilevel programs.
- The LLVF reformulation seems quite promising and research is at initial stage.
- Research has also started on the development of optimality conditions and solution methods for original optimistic and pessimistic models.