

# Global Optimisation of Pessimistic Bi-Level Problems

Angelos Tsoukalas, Wolfram Wiesemann and Berç Rustem

Department of Computing  
Imperial College of Science, Technology and Medicine  
180 Queen's Gate, London SW7 2RH, United Kingdom

## Abstract

We study the solution of non-convex, pessimistic bi-level problems. After providing several motivating examples, we relate the problem to existing research in optimisation. We analyse key properties of the optimisation problem, such as closedness of the feasible region and computational complexity. We then present and investigate a semi-infinite solution approach that solves  $\epsilon$ -approximations of the problem. To the best of our knowledge, this represents the first solution technique proposed for this problem class. We close with numerical results and a discussion on fruitful directions for future research.

## 1 Introduction

This paper is concerned with problems of the following type:

$$\min_{x \in X} f(x) \tag{1a}$$

subject to

$$g(x, y) \leq 0 \quad \forall y \in \arg \max_{y' \in Y} h(x, y'), \tag{1b}$$

where  $X \subseteq \mathbb{R}^n$ ,  $Y \subseteq \mathbb{R}^m$ ,  $f : X \mapsto \mathbb{R}$  and  $g, h : X \times Y \mapsto \mathbb{R}$ . We assume that  $X$  and  $Y$  are non-empty and compact, while  $f$ ,  $g$  and  $h$  are continuous in their arguments. In case multiple constraints  $g_1, \dots, g_p : X \times Y \mapsto \mathbb{R}$  should be considered, they can be condensed to  $g(x, y) = \max\{g_i(x, y) : i = 1, \dots, p\}$ . Similarly, the case of multiple constraints  $h_i$  with individual second-stage objectives  $h_i$  can be handled by a straightforward extension of what follows. We assume that the feasible region of the subordinate optimisation problem does not depend on  $x$ , the variables of the master problem.

Model (1) is understood best in the context of non-cooperative game theory: two players  $A$  and  $B$  participate in a dynamic game where  $A$  (the leader) acts first and  $B$  (the follower) responds. In the game, information is assumed to be complete (i.e., the players' payoff functions  $f_A$  and  $f_B$ , as well as the constraint sets  $X_A$  and  $X_B$ , are common knowledge) and perfect (i.e., the follower knows the leader's decision). Having observed the leader's action  $x_A$ , the follower chooses a feasible decision  $x_B \in X_B$  that maximises his payoff function  $f_B(\cdot)$  (i.e., his 'best response'). Knowing that the follower does so, the leader chooses

an admissible decision  $x_A \in X_A$  which maximises his payoff function  $f_A(\cdot)$ , taking into account that the follower will choose his best response afterwards. The setting just described is known as ‘Stackelberg leader-follower game’ [1] and can be formalised as follows.

$$\max_{x_A \in X_A} \min_{x_B \in X_B} \left\{ f_A(x_A, x_B) : x_B \in \arg \max_{x'_B \in X_B} f_B(x_A, x'_B) \right\}. \quad (2)$$

Problem (2) is readily identified as a special case of model (1): the feasible region of the first-stage problem becomes  $x = (x_A, \tau) \in X = X_A \times \mathbb{R}$ , while the feasible region of the second-stage problem is described by  $y = x_B \in Y = X_B$ . The objective function is represented by  $f((x_A, \tau)) = -\tau$ . The constraint becomes  $g((x_A, \tau), x_B) = \tau - f_A(x_A, x_B)$ , which has to be less than or equal to zero for all maxima of  $h((x_A, \tau), x'_B) = f_B(x_A, x'_B)$ .

If the follower’s decision problem is convex but its objective function  $h$  lacks the property of strict convexity, the follower can possess several different globally optimal responses. If the follower’s decision problem is non-convex, the set of optimal follower responses might even be disconnected. In either situation, the follower is indifferent between several equally beneficial responses. Multiple optimal responses pose a problem to the leader since he cannot anticipate which of these optima will be chosen by the follower, and the corresponding optimal leader decisions can differ significantly. In the optimistic version of the Stackelberg leader-follower game, the leader anticipates that the follower chooses a best response that is most suitable for the leader. In essence, this means that the leader can decide which of the best responses is implemented by the follower. In (1) and (2), on the other hand, the leader does not have this possibility. Instead, he prepares for the worst case, that is, he chooses a decision  $x_A \in X_A$  which performs best in view of the ‘worst’ (globally optimal) follower response. Because of this, (1) and (2) are called pessimistic versions of the bi-level problem and Stackelberg game, respectively.

Stackelberg games and, more generally, bi-level problems naturally arise in economics, engineering, optimisation under uncertainty and several other application areas [8, 9, 33]. In the following, we consider two examples that illustrate some of the difficulties which have to be addressed by solution procedures.

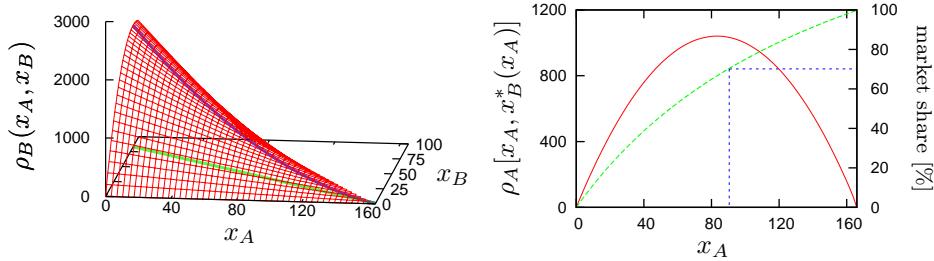
**Production Planning.** Two companies  $A$  and  $B$  produce the same product and need to decide on their production quantities. Company  $A$  is the market leader and has to decide first on a quantity  $x_A \in [0, 200]$ . Company  $B$ , on the other hand, is a follower and decides on a quantity  $x_B \in [0, 100]$  after observing  $A$ ’s decision. The market price for the product depends on the aggregated supply  $x_\Sigma = x_A + x_B$  and is determined by  $p(x_\Sigma) = 100 - 0.3x_\Sigma$ , while the (constant) unit production costs are 50 for both companies.  $A$  and  $B$  want to maximise their profits. Additionally, company  $A$  wishes to preserve its market dominance by providing at least 70% of the aggregated supply.

The profit of company  $A$  and  $B$  amounts to  $\rho_i(x_A, x_B) = (p(x_A + x_B) - 50) x_i$ ,  $i \in \{A, B\}$ . Company  $A$ ’s decision problem can be written as follows.

$$\max_{x_A \in [0, 200]} \min_{x_B \in x_B^*(x_A)} \left\{ (50 - 0.3(x_A + x_B)) x_A \right\}$$

subject to

$$\frac{x_A}{x_A + x_B} \geq 0.7 \quad \forall x_B \in x_B^*(x_A),$$



**Figure 1:** Production planning without investment option. The left graph visualises the follower's optimal response to the leader's production quantity, while the right graph shows the leader's profit (red, straight line) and market share (green, dashed line) as functions of the leader's production output.  $x_B^*(x_A)$  denotes the set of optimal responses of  $B$  to a production quantity  $x_A$  of  $A$ .

where

$$x_B^*(x_A) = \arg \max_{x_B \in [0,100]} \{(50 - 0.3(x_A + x_B)) x_B\}$$

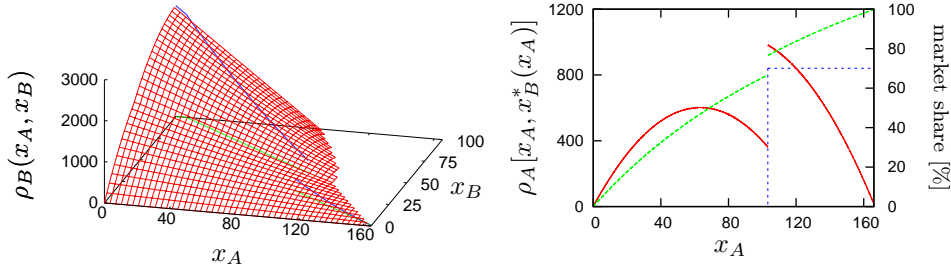
denotes  $B$ 's set of optimal responses to a given production quantity of  $A$ . The objective is to maximise company  $A$ 's worst-case profit. The constraint ensures that  $A$  preserves a market share of at least 70%, no matter which optimal response is chosen by  $B$ . It is easy to rewrite this model into the form of (1).

Figure 1 (left) illustrates company  $B$ 's optimal decision  $x_B^*$  as function of company  $A$ 's production quantity. As expected,  $B$ 's optimal output is monotonically decreasing in  $A$ 's production quantity. Figure 1 (right) shows  $A$ 's profit and market share as a function of  $A$ 's production quantity, anticipating  $B$ 's optimal response from Figure 1 (left). One can see that it would be optimal for  $A$  to produce  $\approx 83.33$  units (with a profit of  $\approx 1041.67$ ), which would cause  $B$  to produce  $\approx 41.66$  units (with a profit of  $\approx 520.87$ ). Since  $A$ 's market share would be  $\approx 66.66\%$  in this case, however,  $A$  instead has to produce the higher quantity of  $\approx 89.75$  units (with a lower profit of  $\approx 1035.49$ ), which causes  $B$  to produce  $\approx 38.45$  units (resulting in a profit of  $\approx 443.71$ ).

Let us now assume that  $B$  has the option to invest in new production facilities (at a cost of 500) that reduce  $B$ 's per-unit production costs to 38. In this case,  $\rho_B$  changes to

$$\rho_B(x_A, x_B) = \max \{(p(x_A + x_B) - 50) x_B, (p(x_A + x_B) - 38) x_B - 500\},$$

while the other problem parameters remain the same. Figure 2 (left) illustrates  $B$ 's new response function.  $B$ 's optimisation problem is clearly non-convex. Furthermore,  $B$ 's response function has a discontinuity at  $x_A \approx 103.33$ : if  $A$  produces more than this quantity, the investment is not beneficial for  $B$ . If  $A$  produces less than this quantity, on the other hand,  $B$  should invest in new facilities. In this case, however, the investment costs have to be amortised by a considerably higher production quantity. Note that at  $x_A \approx 103.33$ ,  $B$  might or might not invest since he is indifferent between both choices. Figure 2 (right) shows how  $A$ 's profit and market share depend on  $A$ 's production quantity in the

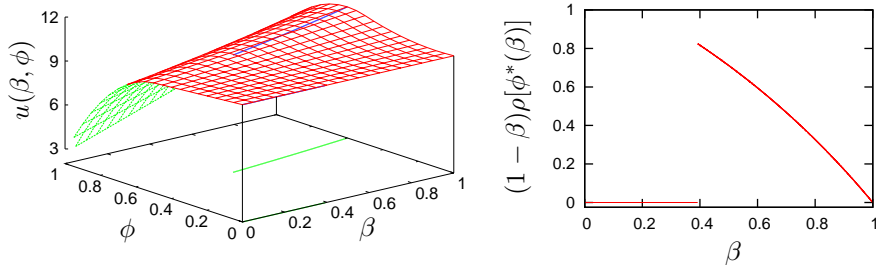


**Figure 2:** Production planning with investment option. The left graph visualises the follower's optimal response to the leader's production quantity, while the right graph shows the leader's profit (red, straight line) and market share (green, dashed line) as functions of the leader's production output. The follower invests if the leader produces less than  $\approx 103.33$  units. As in Figure 1,  $x_B^*(x_A)$  denotes the set of optimal responses of  $B$  to a production quantity  $x_A$  of  $A$ .

new situation. Note that due to the discontinuity in  $B$ 's response function, both  $A$ 's profit and his market share possess similar discontinuities at  $x_A \approx 103.33$ . With and without the market penetration constraint,  $A$ 's optimal production quantity is 'just above' this discontinuity. Since  $A$  optimises in view of the worst of  $B$ 's globally optimal responses, the point of discontinuity itself is *not* optimal: at this point,  $A$  has to consider the possibility that  $B$  might indeed invest and thus decrease the market price (by producing a large quantity in order to amortise his investment costs). As a result,  $A$ 's decision problem contains a discontinuous objective function and a non-closed feasible region (due to the market penetration constraint).  $A$ 's optimal production quantity is  $\approx 103.34$  units (resulting in a profit of  $\approx 981.69$  and a market share of  $\approx 76.54\%$ ), while  $B$ 's optimal response is  $\approx 31.67$  units (leading to a profit of 360.84).

It is interesting to note that  $B$ 's additional investment option results in a smaller profit for  $B$ , even though  $B$  does not exercise this option in the realised solution. This seems to contradict the principles of decision theory, where the presence of additional alternatives should not decrease the achievable profit. The reason for this surprising (but well-known) effect is that  $A$  is aware of  $B$ 's investment opportunity. Hence,  $A$  knows that the previously optimal production quantity of  $\approx 89.75$  would cause  $B$  to exercise his investment option and produce a higher amount. This, in turn, would result in a lower market price, lower profits and a violation of the market penetration constraint. Thus,  $A$  is forced to produce more, which results in lower profits for both companies but nevertheless constitutes the best achievable result for  $A$  in this new setting.

**Principal-Agent Problem.** The principal-agent problem illustrates the difficulty to design contracts under incomplete and asymmetric information [16]. Assume that agent  $A$  is supposed to perform a task for principal  $P$ . The generated profit  $\rho$  of that task is a function of the effort  $\phi \in [0, 1]$  that  $A$  puts into the task. Since  $P$  cannot observe  $\phi$ , he designs an incentive-based contract, that is, apart from a fixed salary, he pays  $A$  a fraction  $\beta$  of the observable task profit  $\rho(\phi)$ .  $P$  wants to choose  $\beta$  in such a way that his part of the profit is maximised. The setting can be cast as a Stackelberg game:  $P$ , the leader, moves first by deciding on the contract parameter  $\beta$ . After he has observed  $\beta$ ,  $A$  (the follower) chooses the effort  $\phi^*(\beta)$  that maximises his utility  $u(\beta, \phi)$ . Anticipating  $A$ 's



**Figure 3:** Principal-agent problem. The left and right graph visualise the agent's optimal response and the principal's profit, both as functions of the contract design chosen by the principal. The agent does not put any effort into the task if the profit share is below  $\approx 39.2\%$ .

decision,  $P$  wants to choose a contract parameter  $\beta$  that maximises his part of the profit, that is,  $(1 - \beta)\rho[\phi^*(\beta)]$ .

In the following, assume that  $A$ 's utility is determined by  $u(\beta, \phi) = s(\beta, \phi) - r(\phi)$ , where  $s(\beta, \phi) = 10 + \beta\rho(\phi)$  denotes his salary and  $r(\phi) = \phi + (5[\phi - 1/2]^+)^2$  represents  $A$ 's reluctance to work. The profit generated by the task outcome is determined by  $\rho(\phi) = 5\phi^2$ . With this notation,  $P$ 's optimisation problem can be described as follows.

$$\max_{\beta \in [0,1]} \min_{\phi \in \phi^*(\beta)} \{5(1 - \beta)\phi^2\},$$

where

$$\phi^*(\beta) = \arg \max_{\phi \in [0,1]} \{10 + 5\beta\phi^2 - \phi - (5[\phi - 1/2]^+)^2\}$$

denotes the set of optimal efforts for a given contract design  $\beta$ . The model maximises  $P$ 's worst-case profit, anticipating that  $A$  will choose any of his optimal responses. Again, it is straightforward to rewrite the model into the form of (1).

Figure 3 (left) illustrates  $A$ 's optimal effort levels for different contract designs  $\beta$ . One can see that  $A$ 's optimal strategy is to put no effort into the task if the profit share  $\beta$  is less than  $\approx 39.2\%$ . At  $\beta \approx 39.2\%$ ,  $A$  is indifferent between  $\phi = 0$  and  $\phi \approx 0.520$ . Above this margin,  $A$ 's optimal response increases with  $\beta$ . Figure 3 (right) shows  $P$ 's profit as a function of  $\beta$ , anticipating the optimal response of  $A$  from Figure 3 (left). As in the production planning example with investment opportunity, the leader's objective function is discontinuous. The optimal contract design is 'just above' this discontinuity at  $\beta \approx 39.2\%$ , causing  $A$  to put an effort of  $\phi \approx 0.521$  into the task. Again, the point of discontinuity itself is not optimal since  $P$  has to consider that  $A$  puts no effort into the task.

In the examples, we have encountered two cases where the follower's optimal response function is discontinuous. Such discontinuities lead to set-valued response functions since at the points of discontinuity, the follower is indifferent between two or more globally optimal responses. Furthermore, we have seen that quite naturally, the leader's optimal strategy can consist in 'pushing' the follower as close as possible to a particular side of such a discontinuity. In Stackelberg games of type (2), this results in (leader) optimisation problems whose projections on the  $x_A$ -space possess discontinuous objective functions. More

generally, in pessimistic bi-level problems, this results in non-closed feasible regions. Hence, one can expect (1) to be generically very difficult to solve.

The remainder of this text is organised as follows. In the next section, we relate formulation (1) to existing research in optimisation. In particular, we will highlight its connection with variants of the minimax problem, bi-level optimisation and mathematical programming with equilibrium constraints. In Section 3, we analyse properties of problem (1), such as closedness of the feasible region and the computational complexity under various assumptions. In Section 4, we present a solution procedure based on semi-infinite programming. Numerical results are provided in Section 5. We conclude in Section 6. Throughout this chapter, the discussion will take place on an ‘intuitive’ level. Formal justification for the claims made here is provided in the accompanying paper [31].

## 2 Relation to Existing Research

In this section, we relate model (1) to existing research. In particular, we consider various minimax formulations, bi-level models and equilibrium constraint problems. It will become apparent that our problem formulation fills a gap in the literature on optimisation under uncertainty and dynamic games.

### 2.1 Minimax Problems

Minimax problems [11, 25] are very similar to the Stackelberg game discussed in Section 1. Like Stackelberg games, minimax problems can be regarded as dynamic games with two players  $A$  (the leader) and  $B$  (the follower). Contrary to Stackelberg games, however, minimax problems constitute constant-sum games, i.e., both players possess the same objective function, which is to be minimised by player  $A$  and maximised by player  $B$ . Apart from game-theoretic applications similar to those outlined in Section 1, minimax problems arise naturally in optimisation under uncertainty. In such settings, player  $B$  plays the role of ‘nature’ which reacts to player  $A$ ’s decision in the most destructive way. Hence, player  $A$  (the decision maker) is interested in a strategy which performs best in view of the worst possible future contingency.

Based on the involved constraints, one distinguishes between uncoupled and coupled minimax problems. The uncoupled minimax problem is defined as:

$$\min_{x \in X} \max_{y \in Y} f(x, y). \quad (3)$$

The characteristic feature of uncoupled minimax problems is that the feasible regions  $X$  and  $Y$  are independent. Note, however, that the optimal strategy of player  $B$  nevertheless depends on the decision of player  $A$  since they are coupled by the objective function. Similar to Stackelberg games, a purely local optimum with respect to both  $x$  and  $y$  is typically not meaningful for practical applications. Although suboptimal solutions with respect to  $x$  are acceptable, a global maximum with respect to  $y$  is needed in order to correctly anticipate player  $B$ ’s response. As a result, solution procedures for (3) typically determine global optima with respect to  $y$ .

Depending on the optimality guarantees provided with respect to  $x$ , one distinguishes between local and global optimisation approaches. Local optimisation

algorithms for (3) have been proposed in [19, 22, 25, 26, 28]. A global optimisation procedure is described in [23]. Problem (3) can furthermore be regarded as a special case of semi-infinite programming, which implies that the respective solution procedures are applicable as well [5, 6, 15]. Due to their favourable convergence properties, however, specialised local optimisation procedures for (3) are preferable to semi-infinite programming techniques.

Contrary to (3), the coupled minimax problem contains coupled constraints which involve the decisions of both players. Coupled constraints can arise in game-theoretic settings (both players have to obey certain regulations, the leader wants to retain a certain market share regardless of the follower's response, etc.) and in optimisation under uncertainty (a constraint needs to be satisfied under any possible future contingency). In the following, we distinguish between two different types of coupled constraints: type *A*-constraints restrain player *A*, whereas type *B*-constraints restrain (primarily) player *B*.

A type *A*-constraint  $g_A : X \times Y \mapsto \mathbb{R}$  can be included in (3) as follows.

$$\min_{x \in X} \max_{y \in Y} f(x, y) \tag{4a}$$

subject to

$$g_A(x, y') \leq 0 \quad \forall y' \in Y. \tag{4b}$$

Hence, player *A* has to ensure that the coupled constraint  $g_A$  is satisfied for all of player *B*'s possible responses. Note that  $g_A$  does not couple player *A*'s decision to player *B*'s eventual response. Instead, any possible response of the follower is considered. Hence,  $g_A$  can be considered as part of the description of  $X$ , player *A*'s feasible region, and the term 'coupled constraint' is misleading. Type *A*-constraints can be regarded as a natural extension of the uncoupled minimax-problem (3) and have various applications in game theory (similar to our discussion in Section 1) and optimisation under uncertainty (e.g., a certain constraint has to be satisfied under any possible future contingency). Problem (4) is typically solved as a semi-infinite program [34].

Often, player *B* does not represent nature (whose most destructive responses should be anticipated), but rather a rational player whose objective is to maximise  $f$ . In this case, type *A*-constraints are unnecessarily restrictive. Indeed, in such settings player *A* merely has to ensure that all *optimal* responses of player *B* satisfy the coupled constraint. The resulting formulation is a special case of (2) and hence represents an instance of the pessimistic bi-level problem (1).

Contrary to type *A*-constraints, a type *B*-constraint  $g_B : X \times Y \mapsto \mathbb{R}$  restricts the admissible decisions of player *B*:

$$\min_{x \in X} \max_{y \in Y} f(x, y) \tag{5a}$$

subject to

$$g_B(x, y) \leq 0. \tag{5b}$$

More precisely, (5) is to be understood as

$$\min_{x \in X} \max_{y \in Y} \{f(x, y) : g_B(x, y) \leq 0\}.$$

Hence, player *A* can choose any strategy  $x \in X$ , while player *B*'s response  $y \in Y$  additionally has to satisfy the coupled constraint  $g_B(x, y) \leq 0$ . Contrary to type *A*-constraints, type *B*-constraints indeed couple the decisions of the players.

Note that (5) implicitly assumes the existence of a  $y \in Y$  that satisfies  $g_B(x^*, y) \leq 0$  for the  $x^*$  chosen by player  $A$  since otherwise, the objective value would be undefined. If non-emptiness of  $\{y \in Y : g_B(x, y) \leq 0\}$  is not vacuously satisfied for every  $x \in X$ , player  $A$  needs to ensure that player  $B$  has a feasible response that satisfies the coupled constraint:

$$\min_{x \in X} \left\{ \max_{y \in Y} \{f(x, y) : g_B(x, y) \leq 0\} : \exists y \in Y . g_B(x, y) \leq 0 \right\}.$$

This problem type seems less common in practice, however, since it enforces cooperation (in view of the satisfaction of the coupled constraint) among players that otherwise behave as perfect competitors (due to the objective function). Furthermore, this ‘cooperation’ is assumed to take place sequentially since player  $A$  decides before player  $B$  responds. A local optimisation procedure for minimax problems with type  $B$ -constraints can be found in [24].

## 2.2 Bi-Level Problems

Minimax problems are special instances of bi-level problems. Structurally, a bi-level problem is an optimisation problem that contains a constraint which requires a subset of the variables to optimise a subordinate optimisation problem. The optimistic (cooperative) bi-level problem can be defined as follows [9].

$$\min_{x, y} f(x, y) \tag{6a}$$

subject to

$$y \in \arg \min_{y' \in Y} \{h(x, y') : q(x, y') \leq 0\} \tag{6b}$$

$$g(x, y) \leq 0 \tag{6c}$$

$$x \in X. \tag{6d}$$

Again, the problem can be interpreted in a game-theoretic setting: for a fixed leader decision  $x^* \in X$ , player  $B$  (the follower) implements a decision

$$y^*(x^*) \in \arg \min_{y' \in Y} \{h(x^*, y') : q(x^*, y') \leq 0\}.$$

Player  $A$  (the leader), on the other hand, solves the problem

$$\min_{x \in X} \{f(x, y^*(x)) : g(x, y^*(x)) \leq 0\}$$

for *some* optimal follower response  $y^*(x)$ . Problem (6) is optimistic in the sense that the leader can choose which of the optimal responses  $y^*(x)$  is implemented by the follower. While this might reflect reality in some settings, it is difficult to justify in a game-theoretic context. One could argue that the leader is allowed to cooperate with the follower in order to decide which of his optimal responses is implemented. Note, however, that absence of cooperation is a fundamental assumption of the bi-level problem. Indeed, if cooperation was allowed, the two-level structure would be inappropriate and the players would rather agree on a Pareto-optimal solution obtained by means of a bargaining process.

Alternatively, one could keep up the assumption of perfect competition but postulate that binding contracts are feasible. In this case, the leader can pay an

arbitrarily small amount  $\epsilon$  to the follower if the latter implements an optimal response that is favourable for the leader. If transfer payments are allowed, however, the leader would have an incentive to exploit their full potential and rather solve the following optimisation problem:

$$\min_{x, \theta(\cdot), y} \{f(x, y) + \theta(y)\}$$

subject to

$$\begin{aligned} y \in \arg \min_{y' \in Y} \{h(x, y') - \theta(y') : q(x, y') \leq 0\} \\ g(x, y) \leq 0 \\ \theta(y) \geq 0 \quad \forall y \in Y \\ x \in X. \end{aligned}$$

Here,  $\theta(y)$  represents the amount of money that the leader pays to the follower in case the latter implements decision  $y \in Y$ .

If there is a unique globally optimal follower response for every leader decision  $x \in X$ , the optimistic bi-level problem (6) is equivalent to (1).<sup>1</sup> This is the case, for example, if  $Y$  is convex and  $h(x, \cdot)$  and  $q(x, \cdot)$  are strictly convex and quasi-convex in their second arguments, respectively, for every  $x \in X$ . If the follower's optimal response is not guaranteed to be unique, the interpretation of the obtained solution needs to be carefully considered. In any case, the optimistic bi-level problem provides a lower (optimistic) bound for (1).

Another difference between (6) and (1) deserves attention: in the former problem, the feasible region of the second-stage problem depends on the solution chosen in the first-stage problem through the constraint ' $q(x, y') \leq 0$ '. This is not the case in (1) since the feasible region of the second-stage problem is defined through ' $y \in Y$ ' only, i.e., it is independent of the first-stage decision  $x$ . In the following, we refer to the former problem type as 'dependent' bi-level problems, while the latter problem class is termed 'independent' bi-level problems. In this paper, we analyse and solve independent pessimistic bi-level problems, see (1).

If the second-stage problem is convex, optimistic bi-level problems can be expressed as mathematical programs with equilibrium constraints. Hence, the solution approaches described in the next section can be used for the local optimisation of (6). For problems with linear and quadratic constraints, global optimisation procedures have been suggested in [12, 13]. Global optimisation procedures for general instances of (6) have been proposed in [21, 30].

An alternative to (6) is the pessimistic bi-level problem presented in [21]:

$$\min_{x \in X} \max_{y \in \mathcal{Y}(x)} f(x, y), \tag{7a}$$

where for  $x \in X$ ,

$$\mathcal{Y}(x) = \left\{ y \in \arg \min_{y' \in Y} \{h(x, y') : q(x, y') \leq 0\} : g(x, y) \leq 0 \right\}. \tag{7b}$$

Contrary to (6), the leader cannot decide anymore which optimal response is implemented by the follower. A subtle but important problem arises when interpreting ' $g(x, y) \leq 0$ ', which implicitly represents a 'cooperative' constraint

---

<sup>1</sup>More precisely the *independent* version of (6) is equivalent to (1), see below.

in (7). Clearly, it restricts the set of admissible follower responses. Since the objective value is undefined for leader decisions  $x \in X$  for which there is no admissible follower response, however, the constraint implicitly restricts the leader as well. This entails two problems. Firstly, we already saw that ‘cooperative constraints’ are difficult to justify in a two-stage game-theoretic setting. Secondly, (7) could be regarded as inconsistent since its objective function anticipates the (non-cooperative) worst case, whereas its constraint assumes (some form of) cooperation between the players.

It seems more reasonable to assume that cautious leaders want to implement a decision  $x$  for which *every* optimal follower response  $y$  vacuously satisfies the constraint ‘ $g(x, y) \leq 0$ ’. This would result in the following formulation:

$$\min_{x \in X} \left\{ \max_{y \in \mathcal{Y}'(x)} \{f(x, y)\} : g(x, y) \leq 0 \quad \forall y \in \mathcal{Y}'(x) \right\},$$

where for  $x \in X$ ,

$$\mathcal{Y}'(x) = \arg \min_{y' \in Y} \{h(x, y') : q(x, y') \leq 0\}.$$

The independent version of this problem is readily identified as a special case of (1). Indeed, assume that there is a function  $\delta : Y \mapsto \mathbb{R}$  such that  $q(x, \cdot) = \delta(\cdot)$  for all  $x \in X$ . Then the problem can be reformulated as follows.

$$\min_{(x, \tau) \in X \times \mathbb{R}} \tau$$

subject to

$$\max \{f(x, y) - \tau, g(x, y)\} \leq 0 \quad \forall y \in \arg \min_{y' \in Y} \{h(x, y') : \delta(y') \leq 0\}.$$

### 2.3 Mathematical Programs with Equilibrium Constraints

Mathematical programming with equilibrium constraints (MPEC) is concerned with problems of the following type [20].

$$\min_{x, y} f(x, y) \tag{8a}$$

subject to

$$y \in \{z \in C(x) : F(x, z)^T(z' - z) \geq 0 \quad \forall z' \in C(x)\} \tag{8b}$$

$$(x, y) \in Z. \tag{8c}$$

Here,  $C$  is a set-valued function that maps vectors  $x$  to closed convex subsets of admissible vectors  $y$ . Constraint (8b) is a parametric variational inequality. Its value becomes apparent when recalling the first-order optimality conditions for optimisation problems: for a convex problem of type

$$\min_{y \in C(x)} \theta(x, y),$$

$y^* \in C(x)$  is an optimal solution if and only if

$$[\nabla_y \theta(x, y^*)]^T (y - y^*) \geq 0 \quad \forall y \in C(x). \tag{9}$$

For general non-convex problems, this variational inequality represents a necessary (but not sufficient) optimality condition.

Employing variational inequality (9), MPECs of type (8) can be used to solve the optimistic bi-level problem (6) if the follower's optimisation problem is convex. In case the follower's objective function is furthermore strictly convex, the optimal solution of (8) coincides with the solutions obtained from the pessimistic bi-level formulations (1) and (7). If the follower's objective function is convex (but not strictly convex), MPECs are valuable if the leader can choose which optimal response is implemented by the follower. As we discussed in the previous section, however, this seems difficult to justify in game-theoretic settings. If the follower's optimisation problem is non-convex, (8) allows the leader to choose any local optimum of the follower's optimisation problem. While this can provide useful lower (optimistic) bounds for the optimal value of (1) and (7), the obtained solution itself seems to be of little interest for both game-theoretic applications and optimisation under uncertainty. Indeed, any rational follower has an incentive to deviate from strategies prescribed by non-global optima of the follower's decision problem. Summing up, for the application areas envisioned in this paper, MPECs are primarily of interest if the follower's optimisation problem is convex and possesses a strictly convex objective function. These assumptions are rather restrictive; indeed, they are not even fulfilled by linear programs in general.

MPECs are generically non-convex and non-smooth optimisation problems. Without further assumptions, the feasible region of (8) is not even guaranteed to be closed. Hence, MPECs are very difficult to solve in general. In recent years, a variety of local optimisation procedures have been proposed that efficiently find feasible (but in general suboptimal) solutions for (8), see [20].

### 3 Problem Analysis

In this section, we analyse the theoretical properties of the independent pessimistic bi-level problem (1):

$$\min_{x \in X} f(x)$$

subject to

$$g(x, y) \leq 0 \quad \forall y \in \arg \max_{y' \in Y} h(x, y'),$$

where  $X \subseteq \mathbb{R}^n$ ,  $Y \subseteq \mathbb{R}^m$ ,  $f : X \mapsto \mathbb{R}$  and  $g, h : X \times Y \mapsto \mathbb{R}$ . As stated in Section 1, we assume that  $X$  and  $Y$  are non-empty and compact, while  $f$ ,  $g$  and  $h$  are continuous in their arguments. We also relate our findings to other variants of the bi-level problem (see Section 2.2).

In the first part of this section, we analyse the closedness of the feasible region of (1). Afterwards, we study the computational complexity of the problem. We close with a discussion of an  $\epsilon$ -approximation of the problem. This approximation will provide the basis for our solution procedure in the next section.

### 3.1 Closedness of the Feasible Region

We denote the point-to-set mapping that relates a first-stage decision  $x \in X$  to the set of optimal second-stage decisions  $y \in Y$  by

$$\mathcal{Y}(x) = \arg \max_{y \in Y} h(x, y).$$

Note that the maximum exists since  $Y$  is compact and  $h$  continuous. For the same reasons,  $\mathcal{Y}(x) = \{y' \in Y : h(x, y') \geq h^*(x)\}$  is compact for every  $x \in X$ . Thus, we can denote the maximum value of  $g(x, \cdot)$  over all optimal second-stage solutions  $y \in \mathcal{Y}(x)$  by  $g^*(x)$ :

$$g^*(x) = \max_{y \in \mathcal{Y}(x)} g(x, y).$$

With this notation, the independent pessimistic bi-level problem (1) equals

$$\min_{x \in X} \{f(x) : g^*(x) \leq 0\}.$$

This problem is closed if  $g^*$  is lower semicontinuous, see Figure 4. Closedness of the feasible region is a crucial property of optimisation problems. Indeed, the minimum or maximum of a continuous function over a non-closed domain might not be attained. The following example illustrates that the feasible region of (1) is generically non-closed:

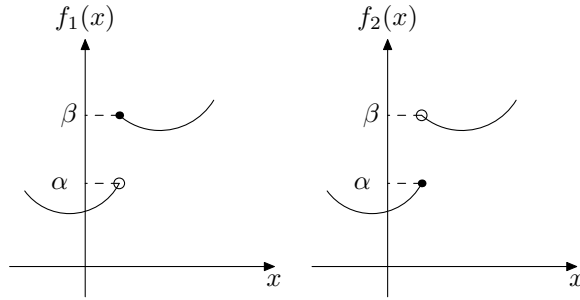
$$\max_{x \in [-1, 1]} x$$

subject to

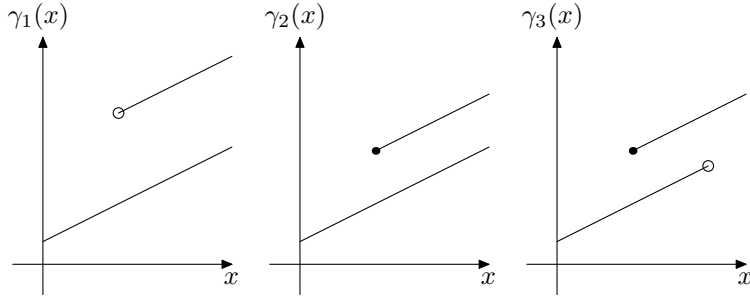
$$x \leq y \quad \forall y \in \arg \min_{y' \in [-1, 1]} \{xy'\}.$$

The model is readily identified as an instance of (1). Note that points  $x > 0$  are infeasible since the constraint becomes ' $x \leq -1$ ' in that case. Similarly,  $x = 0$  is infeasible since the constraint becomes ' $0 \leq y \forall y \in [-1, 1]$ ' in that case. For points  $x < 0$  the constraint becomes ' $x \leq 1$ ' and is as such vacuously satisfied. As a result, the admissible solutions are  $x \in [0, 1)$ , that is, the feasible region is not closed. Furthermore, since the objective is to maximise  $x$ , the optimum is undefined. Note that optimal solutions do exist for other objective functions, such as ' $\max -x$ '.

The non-closedness of the feasible region of (1) can be further analysed by considering Berge's Maximum theorem. For this we need the notions of hemicontinuity and closedness of point-to-set mappings. A point-to-set mapping  $\gamma : X \mapsto \mathcal{P}(Y)$  maps points  $x \in X$  to subsets of  $Y$ .  $\gamma$  is upper hemicontinuous at  $x \in X$  if  $\gamma(x) \subseteq S$  for some open set  $S \subseteq Y$  implies that  $\gamma(x') \subseteq S$  for all  $x'$  in a neighbourhood of  $x$ . Similarly,  $\gamma$  is lower hemicontinuous at  $x \in X$  if  $\gamma(x) \cap S \neq \emptyset$  for some open set  $S \subseteq Y$  implies that  $\gamma(x') \cap S \neq \emptyset$  for all  $x'$  in a neighbourhood of  $x$ .  $\gamma$  is continuous at  $x \in X$  if it is lower and upper hemicontinuous at  $x$ . The concepts of upper and lower hemicontinuity are illustrated in Figure 5.  $\gamma$  is closed at  $x \in X$  if  $x_n \rightarrow x$  and  $y_n \rightarrow y$  for  $y_n \in \gamma(x_n)$  imply that  $y \in \gamma(x)$ . Note that closedness of the point-to-set mapping  $\gamma$  at  $x \in X$  (as just described) implies closedness of the set  $\gamma(x)$ , but not vice versa. In order to avoid confusion, we will refer to the latter concept as *closed-valued* (or, in connection with boundedness, as *compact-valued*) in



**Figure 4:** Semicontinuity of constraints determines closedness of the feasible region. Function  $f_1$  in the left graph is upper semicontinuous, while function  $f_2$  in the right graph is lower semicontinuous.  $\{x : f_1(x) \leq \delta\}$  is only closed for  $\delta < \alpha$  and  $\delta \geq \beta$ , whereas  $\{x : f_1(x) \geq \delta\}$  is closed for all  $\delta$ . Similarly,  $\{x : f_2(x) \geq \delta\}$  is only closed for  $\delta \leq \alpha$  and  $\delta > \beta$ , whereas  $\{x : f_2(x) \leq \delta\}$  is closed for all  $\delta$ .



**Figure 5:** Upper and lower hemicontinuity. Point-to-set mapping  $\gamma_1$  is lower, but not upper hemicontinuous. Point-to-set mapping  $\gamma_2$ , on the other hand, is upper, but not lower hemicontinuous. Point-to-set mapping  $\gamma_3$ , finally, is neither upper nor lower hemicontinuous.

the following. Both concepts are illustrated in Figure 6. We say that  $\gamma$  is lower/upper hemicontinuous or closed if it is lower/upper hemicontinuous or closed at all  $x \in X$ .

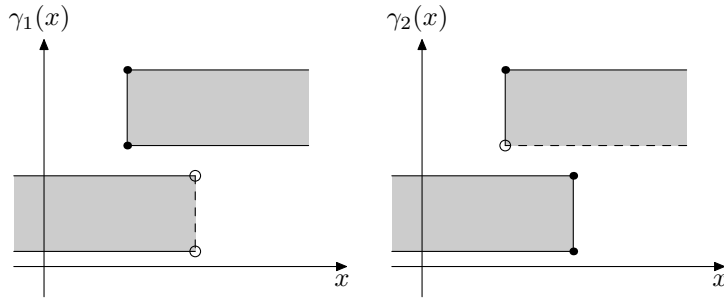
With this notation, Berge's Maximum theorem can be stated as follows.

**Theorem 3.1 (Berge's Maximum Theorem [3, 7])** For  $X \subseteq \mathbb{R}^n$  and  $Y \subseteq \mathbb{R}^m$ , let  $\gamma : X \mapsto \mathcal{P}(Y)$  be a compact-valued point-to-set mapping. Furthermore, for a continuous function  $\eta : X \times Y \mapsto \mathbb{R}$ , let

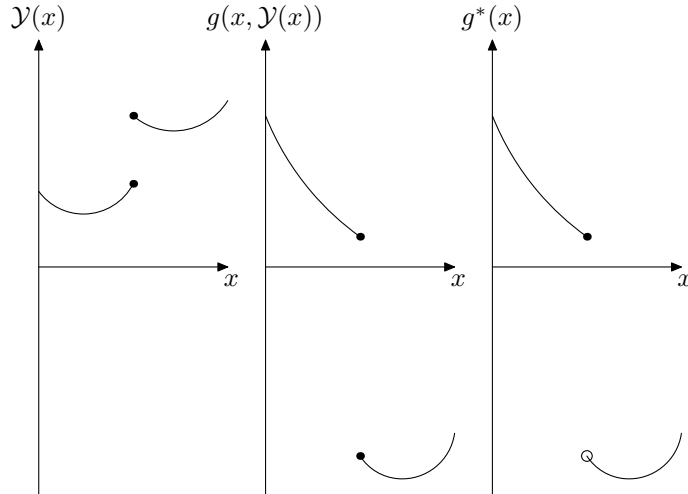
$$\Gamma(x) = \arg \max_{y \in \gamma(x)} \eta(x, y).$$

Then  $\Gamma$  is compact-valued for all  $x \in X$ . Furthermore, if  $\gamma$  is continuous at  $x \in X$ , then  $\Gamma$  is closed and upper hemicontinuous at  $x$ .

In our setting,  $X$  and  $Y$  coincide with the respective sets of (1),  $\gamma(x) = Y$  for all  $x \in X$ ,  $\eta(x, y) = h(x, y)$  and  $\Gamma(x) = \mathcal{Y}(x)$ . Since  $Y$  is compact,  $\gamma$  is compact-valued in our case. Continuity of  $\gamma$  is vacuously satisfied. Hence, Berge's Maximum theorem is applicable and  $\mathcal{Y}$  is closed and upper hemicontinuous. These properties do *not* imply lower semicontinuity of  $g^*$ , however. This is illustrated in Figure 7.



**Figure 6:** Closed and closed-valued point-to-set mappings. Point-to-set mapping  $\gamma_1$  is closed-valued, but not closed. Point-to-set mapping  $\gamma_2$  is neither closed-valued nor closed.



**Figure 7:** Closedness and upper hemicontinuity of  $\mathcal{Y}$  do not imply lower semicontinuity of  $g^*$ . For the point-to-set mapping  $\mathcal{Y}$  in the left graph, the middle and right graphs visualise  $x \mapsto \{g(x, y) : y \in \mathcal{Y}(x)\}$  and  $x \mapsto g^*(x)$ , respectively. Although  $\mathcal{Y}$  is closed and upper hemicontinuous,  $g^*$  fails to be lower semicontinuous.

Let us contrast this with the independent *optimistic* version of the bi-level problem. Here, the optimisation problem is

$$\min_{x \in X} \{f(x) : g'(x) \leq 0\} \quad \text{with} \quad g'(x) = \min_{y \in \mathcal{Y}(x)} g(x, y).$$

Again, we can employ Berge's Maximum theorem to show that  $\mathcal{Y}$  is closed and upper hemicontinuous. With the new choice of functions, however, this implies that  $g'$  is lower semicontinuous. Hence, the independent optimistic bi-level problem has a closed feasible region. Note that the same reasoning can be employed to show the *upper* semicontinuity of  $g^*$  in the case of the independent pessimistic bi-level problem.

Let us compare this result with the well-known fact that *dependent* optimistic bi-level problems have non-closed feasible regions in general. This can be

illustrated by the following example, which is borrowed from [29]:

$$\min_{\substack{x \in \mathbb{R}, \\ y \in [0, 2]}} \{y - x\}$$

subject to

$$\begin{aligned} x &\leq 1 \\ y &\in \arg \max_{y' \in [0, 2]} \{x + y' : xy' \geq 0\}. \end{aligned}$$

One can readily verify that the feasible region of this problem is described by the non-closed set  $(x, y) \in (-\infty, 0) \times \{0\} \cup [0, 1] \times \{2\}$ . In the terminology of Theorem 3.1, the point-to-set mapping  $\gamma$  is not continuous at zero and Berge's Maximum theorem is as such not applicable.<sup>2</sup> Hence, the added generality of the dependent optimistic bi-level problem over its independent counterpart comes at the price of generically non-closed feasible regions. Of course the dependent pessimistic bi-level problem has a generically non-closed feasible region as well since it encompasses its independent counterpart as a special case.

Summing up, we have seen that the pessimistic bi-level problem seems more involved than its optimistic counterpart. The feasible region of both problems can be expressed as  $\{x \in X : g(x) \leq 0\}$  for some function  $g : X \mapsto \mathbb{R}$ . For independent optimistic bi-level problems, we can employ Berge's Maximum theorem to show that this  $g$  is lower semicontinuous. This immediately implies that the feasible region is closed. Although Berge's Maximum theorem is applicable to independent pessimistic bi-level problems as well, this would only allow us to prove upper semicontinuity of  $g$ . Indeed, a simple example demonstrated that the feasible region of independent pessimistic bi-level problems is generically non-closed. Dependent bi-level problems, finally, have generically non-closed feasible regions, no matter whether the optimistic or pessimistic variant is considered.

It turns out that even in benign cases where the feasible region is closed, bi-level problems are difficult to solve in general. In the next section, we examine the computational complexity of several variants of the bi-level problem.

## 3.2 Computational Complexity

In this section, we consider the computational complexity of optimistic and pessimistic bi-level problems. Clearly, these problems are difficult to solve in general, i.e., without further assumptions about the feasible regions and involved functions. In the following, we will look at special cases and show that even in seemingly benign settings, bi-level problems remain difficult to solve. We assume some familiarity with the basic notions of complexity theory, in particular with  $\mathcal{NP}$ -hardness. For a textbook introduction to complexity theory, the reader is referred to [14]. Complexity results for dependent optimistic bi-level problems are compiled in [10, 33]; we will mention some of them further below.

We will base our discussion on the Knapsack problem: imagine  $n$  different goods with weights  $w_i$  and utilities  $u_i$ ,  $i = 1, \dots, n$ . The goal is to find a

---

<sup>2</sup>More precisely,  $\gamma$  is not lower hemicontinuous in this case. To see this, consider the open set  $S = (1, 2)$ :  $\gamma(x) \cap S \neq \emptyset$  for  $x \geq 0$ , but  $\gamma(x) \cap S = \emptyset$  for  $x < 0$ .

utility-maximising subset of goods whose collective weight does not exceed  $W$ :

$$\max_{x \in \{0,1\}^n} \left\{ \sum_{i=1}^n u_i x_i : \sum_{i=1}^n w_i x_i \leq W \right\}. \quad (10)$$

The Knapsack problem is known to be  $\mathcal{NP}$ -hard [14]. Intuitively speaking, this is caused by the binary requirement, which forces (current) solution procedures to consider many of the elements of  $\{0,1\}^n$  in the worst case. Since this set is of size  $2^n$ , this results in exponential solution procedures. For the remainder of this section, we will be interested in the following reformulation:

$$\max_{x \in [0,1]^n} \left\{ \sum_{i=1}^n u_i x_i - M \sum_{i=1}^n \min \{x_i, 1 - x_i\} : \sum_{i=1}^n w_i x_i \leq W \right\}. \quad (10')$$

Here, the binary requirement has been dropped in favour of the penalty function  $M \sum_{i=1}^n \min \{x_i, 1 - x_i\}$ . Note that this function vanishes at binary solutions, while other  $x$  are penalised with a multiplicity of the constant  $M$ . It can be shown that the penalty function is exact for finite  $M$ , i.e., for a suitably large (but polynomial)  $M$ , the optimal solution of the latter problem is guaranteed to be binary [2, 4]. For such  $M$ , the optimal solutions of (10) and (10') coincide. Due to the penalty term, (10') remains a non-convex optimisation problem.

Let us now consider linear bi-level problems:

$$\min_{x \in \mathbb{R}_+^n} c^T x \quad (11a)$$

subject to

$$Ax + By \leq b \quad \forall y \in \arg \max_{z \in \mathbb{R}_+^m} \{d^T x + f^T z : Cx + Dz \leq g\}. \quad (11b)$$

Model (11) represents a dependent pessimistic bi-level problem. If the ‘ $\forall$ ’ in the first constraint is replaced by an ‘ $\exists$ ’, we obtain the corresponding dependent optimistic bi-level problem. Let us first assume that  $C = 0$ . In this case, we consider independent bi-level problems, i.e., the set of admissible second-stage decisions does not depend on the first-stage variables. Moreover, since the contribution of the first-stage variables to the second-stage objective is constant, the set of optimal second-stage decisions does not depend on the first-stage variables either. In itself, however, this does not imply that (11) can be solved efficiently. Indeed, it could well be that the (vertex) set of optimal second-stage solutions has exponential cardinality, and the mere enumeration of its elements would be a difficult problem. It turns out, however, that one can employ duality arguments to show that both the optimistic and the pessimistic version of the linear independent bi-level problem can be solved in polynomial time [31].

Let us now assume that  $C \neq 0$ . In [10] it has been shown that the linear dependent optimistic bi-level problem can be solved in polynomial time if the number of second-stage variables is regarded as a constant. The authors show that the result can be extended to the linear dependent pessimistic bi-level problem. If we drop the assumption of a constant number of second-stage variables, however, the linear dependent optimistic bi-level problem becomes strongly  $\mathcal{NP}$ -hard [18]. In the following, we show that the linear dependent

pessimistic bi-level problem is  $\mathcal{NP}$ -hard as well, which is an immediate consequence of [2]. Consider the following formulation:

$$\max_{(x,\tau) \in \mathbb{R}_+^{n+1}} \left\{ \sum_{i=1}^n u_i x_i - M\tau \right\}$$

subject to

$$\begin{aligned} & \sum_{i=1}^n w_i x_i \leq W \\ \tau & \geq \sum_{i=1}^n y_i \quad \forall y \in \arg \max_{z \in [0,1]^n} \left\{ \sum_{i=1}^n z_i : z \leq x, z \leq e - x \right\}, \end{aligned}$$

where  $e$  denotes the vector of all ones. Note that for a given first-stage solution  $x$ , the optimal second-stage solution is  $y^*$  with  $y_i^* = \min\{x_i, 1 - x_i\}$ . Thus, this problem is equivalent to (10'). As a result, the linear dependent pessimistic bi-level problem is indeed  $\mathcal{NP}$ -hard.

Let us now consider nonlinear versions of the bi-level problem. Clearly, nonlinear dependent bi-level problems are  $\mathcal{NP}$ -hard since they encompass the linear dependent bi-level problems as special cases. In the following, we show that contrary its linear counterpart, the nonlinear independent pessimistic bi-level problem is  $\mathcal{NP}$ -hard as well. More precisely, we consider

$$\min_{x \in X} f(x) \tag{12a}$$

subject to

$$g(x, y) \geq 0 \quad \forall y \in \arg \max_{z \in Y} h(x, z) \tag{12b}$$

with  $f, g$  linear,  $X$  and  $Y$  polyhedral and  $h$  concave and smooth. Let us take a look at the following problem:

$$\max_{(x,\tau) \in \mathbb{R}_+^{n+1}} \left\{ \sum_{i=1}^n u_i x_i - M\tau \right\}$$

subject to

$$\begin{aligned} & \sum_{i=1}^n w_i x_i \leq W \\ \tau & \geq \sum_{i=1}^n y_i^2 \quad \forall (y^1, y^2) \in \arg \max_{(z^1, z^2) \in \mathbb{R}_+^{2n}} \left\{ -\|z^1 - x\|_2 : z^2 \leq z^1, z^2 \leq e - z^1 \right\}, \end{aligned}$$

where  $\|\cdot\|_2$  denotes the Euclidean norm. Clearly, the problem is an instance of (12). For a given first-stage solution  $x$ , the optimal second-stage solution is  $(y^1, y^2)$  with  $y_i^1 = x_i$  and  $y_i^2 \in [0, \min\{x_i, 1 - x_i\}]$ . Since  $\tau$  has to be greater than or equal to the sum of  $y^2$ 's components for *every* optimal  $y^2$ , this problem is equivalent to (10'), too. We note that the problem remains  $\mathcal{NP}$ -hard if we allow the second-stage problem to be solved only with (polynomially bounded) accuracy  $\epsilon$ . The corresponding proof can be found in [31].

Summing up, linear independent bi-level problems are efficiently solvable. The nonlinear independent pessimistic bi-level problem, however, is  $\mathcal{NP}$ -hard even in very elementary settings. Dependent bi-level problems are in general more difficult than their independent counterparts. Indeed, the only efficiently solvable case we are aware of are linear dependent bi-level problems under the additional assumption of a constant number of second-stage variables.

### 3.3 $\epsilon$ -Approximations

We have seen in Section 3.1 that the feasible region of the independent pessimistic bi-level problem (1),

$$\min_{x \in X} f(x)$$

subject to

$$g(x, y) \leq 0 \quad \forall y \in \arg \max_{y' \in Y} h(x, y'),$$

is generically non-closed. In this section, we construct an  $\epsilon$ -approximation of (1) that is closed and, under certain technical assumptions, ‘converges’ (with respect to the set of optimal solutions) to problem (1) as  $\epsilon \rightarrow 0$ . This approximation serves as the basis for the solution procedure which we will discuss in Section 4.

Using the notation from Section 3.1, we can reformulate (1) as follows.

$$\min_{x \in X} \{f(x) : g(x, y) \leq 0 \quad \forall y \in \mathcal{Y}(x)\},$$

where

$$\mathcal{Y}(x) = \{y \in Y : h(x, y) \geq h(x, y') \quad \forall y' \in Y\}.$$

In the remainder of this section, we refer to this problem as  $P$ . Let us denote by  $\theta^*$  the tightest bound on the objective value of  $P$  that we can obtain:

$$\theta^* = \inf_{x \in X} \{f(x) : g(x, y) \leq 0 \quad \forall y \in \mathcal{Y}(x)\}.$$

Since  $P$  is not guaranteed to be closed,  $\theta^*$  might or might not be attainable.

We consider the following approximation  $P(\epsilon)$  of  $P$ ,  $\epsilon > 0$ :

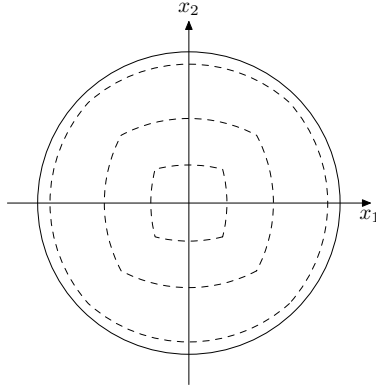
$$\min_{x \in X} \{f(x) : g(x, y) \leq 0 \quad \forall y \in \mathcal{Y}_\epsilon(x)\}, \quad (13a)$$

where

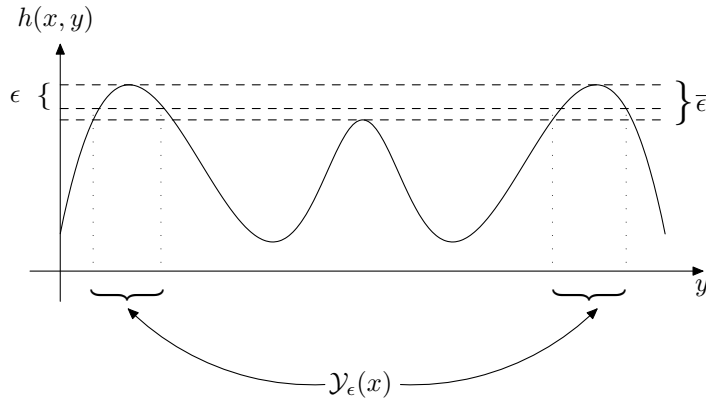
$$\mathcal{Y}_\epsilon(x) = \{y \in Y : h(x, y) > h(x, y') - \epsilon \quad \forall y' \in Y\}. \quad (13b)$$

We denote the optimal objective value of this problem by  $\theta(\epsilon)$  if it exists and set  $\theta(\epsilon) = \infty$  otherwise. Note that  $\mathcal{Y}_\epsilon(x)$  is a superset of  $\mathcal{Y}(x)$ ,  $x \in X$ . Hence,  $P(\epsilon)$  enforces more ‘ $g(x, y) \leq 0$ ’-constraints and as such constitutes a conservative approximation of  $P$ , see Figure 8. This, together with the fact that the objective functions of both problems coincide, implies that  $\theta(\epsilon) \geq \theta^*$  for all  $\epsilon > 0$ . Decreasing the value of  $\epsilon$  leads to better approximations of  $P$  in the following sense:

$$\forall \kappa > 0 \exists \lambda > 0 : \forall x \in X, \epsilon \in (0, \lambda], y \in \mathcal{Y}_\epsilon(x) \exists y' \in \mathcal{Y}(x) : \|y - y'\| \leq \kappa. \quad (14)$$



**Figure 8:** Feasible region of  $P$  and  $P(\epsilon)$  for different values of  $\epsilon$ . Shown is the feasible region (with respect to the  $X$ -space) induced by  $g(x, y) \leq 0$  for  $y \in \mathcal{Y}(x)$  (straight line) and  $y \in \mathcal{Y}_\epsilon(x)$  for different  $\epsilon$  (dashed lines). In this example,  $x, y \in \mathbb{R}^2$ ,  $g(x, y) = \|x\|_2^2 + \|y\|_2^2 - 50$  and  $h(x, y) = \|x - y\|_2$ .



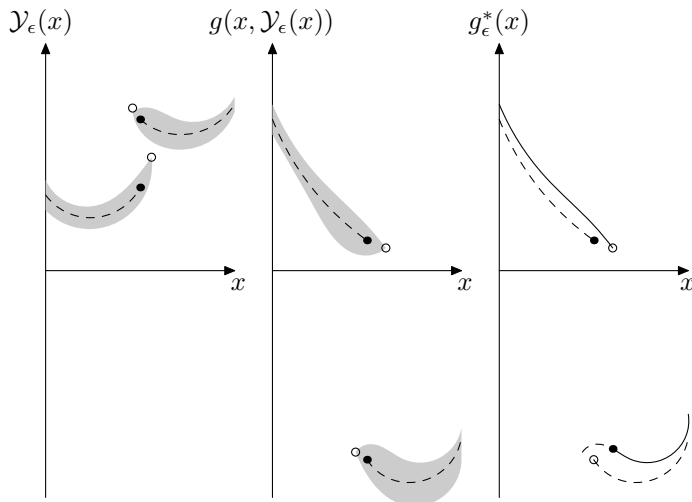
**Figure 9:** Convergence of  $\mathcal{Y}_\epsilon(x)$  to  $\mathcal{Y}(x)$  for given  $x$ . For  $\epsilon$  small enough (in this case, for  $\epsilon < \bar{\epsilon}$ ),  $\mathcal{Y}_\epsilon(x)$  contains only second-stage solutions ‘close’ to global maxima of  $h$ .

Hence, the distance between every  $y \in \mathcal{Y}_\epsilon(x)$  and its ‘closest’  $y' \in \mathcal{Y}(x)$  can be made arbitrarily small by decreasing  $\epsilon$ . The nature of  $\mathcal{Y}_\epsilon$  and the approximation thereby obtained is further illustrated in Figure 9.

Contrary to  $P$ ,  $P(\epsilon)$  has a closed feasible region for every  $\epsilon > 0$ . For this property, the strict inequality in the definition of  $\mathcal{Y}_\epsilon$  turns out to be crucial. To see this, note that (13) can be equivalently expressed as

$$\min_{x \in X} \{f(x) : g_\epsilon^*(x) \leq 0\} \quad \text{with} \quad g_\epsilon^*(x) = \sup_{y \in \mathcal{Y}_\epsilon(x)} g(x, y). \quad (13')$$

Figure 10 (left) shows that  $\mathcal{Y}_\epsilon$  is ‘open from the sides’ in the following sense: if  $y \in \mathcal{Y}_\epsilon(x)$ , then there is a neighbourhood  $N(x)$  of  $x$  such that  $y \in \mathcal{Y}_\epsilon(x')$  for all  $x' \in N(x)$ . This property is a direct consequence of the strict inequality in the definition of  $\mathcal{Y}_\epsilon$ . As Figure 10 (middle) illustrates, this property is passed on to the mapping  $x \mapsto \{g(x, y) : y \in \mathcal{Y}_\epsilon(x)\}$ . Figure 10 (right), finally, shows that this property ensures lower semicontinuity of  $g_\epsilon^*$ . Lower semicontinuity of



**Figure 10:** Employing  $\mathcal{Y}_\epsilon$  instead of  $\mathcal{Y}$  leads to closed, conservative approximations of  $P$ . The graphs illustrate the consequences of using  $\mathcal{Y}_\epsilon$  instead of  $\mathcal{Y}$  (dashed lines) for the example from Figure 7. Function  $g_\epsilon^*$  is lower semicontinuous and bounds  $g^*$  from above.

$g_\epsilon^*$ , however, guarantees that (13') (and thus, (13)) has a closed feasible region.

The difference between  $\theta(\epsilon)$  and  $\theta^*$  depends on the value of  $\epsilon$ . From the relation between  $\mathcal{Y}_\epsilon$  and  $\mathcal{Y}$ , one expects that smaller values of  $\epsilon$  lead to smaller gaps between  $\theta(\epsilon)$  and  $\theta^*$ . It can happen, however, that  $P(\epsilon)$  is infeasible for every  $\epsilon > 0$ , whereas  $P$  contains feasible solutions.<sup>3</sup> Similarly, optimal solutions of  $P(\epsilon)$  might 'substantially deviate' from those of  $P$  for every  $\epsilon > 0$ . In the following, we discuss a technical condition under which such problems cannot arise. More precisely, we give a condition under which

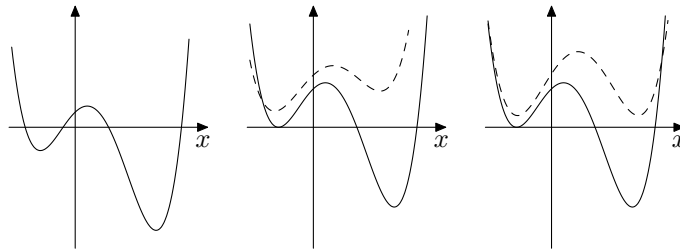
$$\lim_{\epsilon \rightarrow 0} \theta(\epsilon) = \theta^*.$$

By construction,  $\theta(\epsilon) \geq \theta^*$  for all  $\epsilon > 0$ . Moreover, feasible solutions of  $P$  have objective values of  $\theta^*$  or larger. Since  $P$  is not necessarily closed, however,  $\theta^*$  might not be achievable. Hence, we cannot hope for more than convergence of  $\theta(\epsilon)$  to  $\theta^*$  for  $\epsilon \rightarrow 0$ . In order to show such a convergence, we introduce another approximation  $P'(\delta)$  of  $P$ ,  $\delta > 0$ :

$$\min_{x \in X} \{f(x) : g(x, y) \leq -\delta \quad \forall y \in \mathcal{Y}(x)\}.$$

Let us denote the optimal objective value of  $P'(\delta)$  by  $\psi(\delta)$ ; in case  $P'(\delta)$  is infeasible, we set  $\psi(\delta) = \infty$ . For every  $\delta > 0$ , we can find an  $\epsilon(\delta) > 0$  such that  $\theta(\epsilon(\delta)) \leq \psi(\delta)$ . Indeed, let  $x \in X$  be feasible for  $P'(\delta)$ , that is,  $g(x, y) \leq -\delta$  for all  $y \in \mathcal{Y}(x)$ . We want to show that  $x$  is feasible for some  $P(\epsilon)$  as well, that is,  $g(x, y) \leq 0$  for all  $y \in \mathcal{Y}_\epsilon(x)$ . Due to the continuity of  $g$ , we can find a  $\phi > 0$  such that  $g(x, y') \leq 0$  for all  $y' \in Y$  with  $\|y - y'\| \leq \phi$ . Hence, we need to choose  $\epsilon$  such that all  $y' \in \mathcal{Y}_\epsilon(x)$  have a distance of at most  $\phi$  to some  $y \in \mathcal{Y}(x)$ . This is exactly what property (14) allows us to do. Hence,  $P'(\delta)$  is indeed a

<sup>3</sup>This can happen, for example, if the feasible region of  $P$  contains no interior points.



**Figure 11:** Technical condition for the convergence of  $P(\epsilon)$  to  $P$ ,  $\epsilon \rightarrow 0$ . The straight and dashed lines represent  $g^*$  and  $f$ , respectively. In the first graph,  $g^*$  does not possess local minima of value zero and hence, convergence can be guaranteed for any  $f$ . In the second graph, the unique global optimum of  $f$  is located at a solution where  $g^*$  attains a local minimum of value zero. As a result, convergence cannot be ensured in this case. In the third graph, finally, there are global optima of  $f$  where  $g^*$  does not attain a local minimum of value zero and hence, convergence is guaranteed.

conservative approximation of some  $P(\epsilon(\delta))$ , and we have shown that

$$x \text{ is feasible for } P'(\delta) \rightsquigarrow x \text{ is feasible for } P(\epsilon(\delta)) \rightsquigarrow x \text{ is feasible for } P.$$

Furthermore, the objective functions of all three problems coincide.

It remains to be shown that under a certain condition,

$$\lim_{\delta \rightarrow 0} \psi(\delta) = \theta^*.$$

Tuy [17, 32] calls such problems  $P$  ‘stable’. A sufficient condition for the stability of  $P$  is that the function  $g^*$  from Section 3.1,

$$g^*(x) = \max_{y \in \mathcal{Y}(x)} g(x, y),$$

has no local minimum of value zero, see [17, 32]. A tighter description can be found in [31], where it is shown that  $P$  is stable as long as not all of its global optima (if such global optima exist) are attained at local minima of  $g^*$  with value zero. If  $P$  does not possess any global optima (remember that  $P$  might have a non-closed feasible region),  $\psi(\delta)$  will always converge to  $\theta^*$  for  $\delta \rightarrow 0$ . The stability condition is illustrated in Figure 11. Since  $\theta(\epsilon(\delta))$  is ‘sandwiched’ between  $\psi(\delta)$  and  $\theta^*$ , stability of  $P$  implies convergence of  $\theta(\epsilon)$  to  $\theta^*$  for  $\epsilon \rightarrow 0$ .

Summing up, we have discussed a conservative approximation  $P(\epsilon)$  of the independent pessimistic bi-level problem  $P$ . Contrary to the latter problem, this approximation is closed and as such amenable to optimisation techniques. In the next section, we will present a semi-infinite solution procedure that allows us to solve  $P(\epsilon)$  for fixed  $\epsilon$ . The resulting solution represents a feasible but in general suboptimal solution to  $P$ . Better solutions can be found for smaller values of  $\epsilon$ , and if a technical assumption is satisfied, these solutions are guaranteed to converge to the global optima of  $P$  (if existent) for  $\epsilon \rightarrow 0$ .

Throughout this section, our key motivation for the  $\epsilon$ -approximation (13) was computational tractability. Indeed, contrary to the original problem (1), the approximation (13) has a closed feasible region. Apart from computational aspects, there are at least two further reasons why the  $\epsilon$ -approximation is a valuable tool in the setting of Stackelberg games. Firstly, it is very likely that

the leader does not know the follower's optimisation problem exactly, that is, he might not be sure about the precise shape of all involved functions. In this setting, which is termed 'imperfect information' in game theory, the leader might want to include a 'safety margin' by solving the  $\epsilon$ -approximation (13) instead of the original problem (1). Secondly and closely related, real-world decision makers suffer from 'bounded rationality'. In our setting, this means that the follower might not possess the resources to solve his optimisation problem to global optimality. Instead, he might settle for slightly suboptimal, e.g.,  $\epsilon$ -optimal decisions. By considering not only optimal, but also  $\epsilon$ -optimal follower decisions, the leader can hedge against followers with potentially bounded rationality.

## 4 Semi-Infinite Solution Procedure

In this section, we present a solution procedure for the  $\epsilon$ -approximation (13) of problem (1) with  $\epsilon > 0$  fixed:

$$\min_{x \in X} \{f(x) : g(x, y) \leq 0 \quad \forall y \in \mathcal{Y}_\epsilon(x)\},$$

where

$$\mathcal{Y}_\epsilon(x) = \{y \in Y : h(x, y) > h(x, y') - \epsilon \quad \forall y' \in Y\}.$$

We will employ semi-infinite programming principles [15] to solve this problem. Note that the constraint ' $g(x, y) \leq 0 \quad \forall y \in \mathcal{Y}_\epsilon(x)$ ' is equivalent to

$$\begin{aligned} & [y \in \mathcal{Y}_\epsilon(x) \Rightarrow g(x, y) \leq 0] && \forall y \in Y \\ \Leftrightarrow & [y \notin \mathcal{Y}_\epsilon(x)] \vee [g(x, y) \leq 0] && \forall y \in Y \\ \Leftrightarrow & [\exists y' \in Y : h(x, y) \leq h(x, y') - \epsilon] \vee [g(x, y) \leq 0] && \forall y \in Y, \end{aligned}$$

which bears structural similarity to the discussion in [21].<sup>4</sup> We are thus lead to the following reformulation of (13):

$$\min_{(x, y') \in X \times Y} f(x)$$

subject to

$$[g(x, y) \leq 0] \vee [h(x, y) \leq h(x, y') - \epsilon] \quad \forall y \in Y.$$

Note that this problem contains an infinite number of constraints in general, namely one for every  $y \in Y$ . In the spirit of semi-infinite programming, we solve this problem iteratedly for finite subsets  $Y_K$  of  $Y$ . More precisely, at the  $K$ th iteration, we consider

$$\min_{(x, y') \in X \times Y} f(x) \tag{15a}$$

subject to

$$[g(x, y_k) \leq 0] \vee [h(x, y_k) \leq h(x, y') - \epsilon] \quad \forall y_k \in Y_K \tag{15b}$$

---

<sup>4</sup>In the context of [21], however, disjunctive constraints are needed to enforce dependent second-stage constraints, whereas we employ disjunctive constraints in order to enforce  $g(x, y) \leq 0$  for all second-stage optimisers, i.e., to reflect the 'pessimism' of the leader.

for a finite subset  $Y_K = \{y_1, \dots, y_K\} \subseteq Y$ . Clearly, if some  $x \in X$  is feasible for (13), then it is also feasible for (15) and the objective values coincide. If  $x \in X$  is feasible for (15), however, we cannot conclude that  $x$  is also feasible for (13). Hence, we refer to (15) as  $Y_K$ -relaxation of (13). The optimal solution of (15),  $(x_K, y'_K) \in X \times Y$ , either satisfies

$$g(x_K, y) \leq 0 \quad \forall y \in \mathcal{Y}_\epsilon(x_K)$$

or there is an  $y_{K+1}$  such that

$$y_{K+1} \in \mathcal{Y}_\epsilon(x_K) \wedge g(x_K, y_{K+1}) > 0.$$

In the former case,  $x_K$  is feasible for the  $\epsilon$ -approximation (13) and at the same time optimal for its  $Y_K$ -relaxation (15). Since the objective functions of both problems coincide, we can conclude that  $x_K$  is optimal for (13) and terminate. In the latter case,  $y_{K+1}$  serves as a certificate for  $x_K$ 's infeasibility. In this case, we set  $Y_{K+1} = Y_K \cup \{y_{K+1}\}$  and iterate.

Before we consider the details of our solution procedure, we want to address two issues of practical importance. Firstly, we discuss how the disjunctive constraint (15b) can be expressed explicitly. Secondly, we consider a ‘tightened’ version of the  $Y_K$ -relaxation that can lead to faster convergence in practice.

For a given  $y_k \in Y_K$ , constraint (15b) is a disjunctive constraint of type

$$[\alpha(x) \leq 0] \vee [\beta(x) \leq 0]. \quad (16)$$

Disjunctions of this type are equivalent to the existence of a  $\lambda \in [0, 1]$  such that

$$\lambda\alpha(x) + (1 - \lambda)\beta(x) \leq 0.$$

Similarly, (16) is equivalent to the existence of a  $\lambda \in \{0, 1\}$  such that

$$\alpha(x) - M\lambda \leq 0 \wedge \beta(x) - M(1 - \lambda) \leq 0,$$

where  $M$  denotes a sufficiently large number. Which reformulation leads to ‘easier’ problems depends on the remaining constraints and the available optimisation software. In our numerical tests, we employ the first variant since it does not require any additional constants.

We now consider the issue of tightening relaxation (15). In this problem, both  $x$  and  $y'$  are decision variables, but only  $x$  enters the objective function. Indeed,  $y'$  plays the role of an auxiliary variable that decides which constraints ‘ $g(x, y_k) \leq 0$ ’,  $y_k \in Y_K$ , are to be enforced.<sup>5</sup> This implies that the choice  $y' \in \arg \max_{y \in Y} h(x, y)$  is always optimal since it leads to the least restrictions on  $x$ . In the extreme case,  $y' \in Y$  satisfies  $h(x, y') \geq h(x, y_k) + \epsilon$  for all  $y_k \in Y_K$ , which reduces (15) to the essentially unconstrained problem  $\min_{x \in X} f(x)$ . The effect just described can result in a slow convergence until large parts of the  $Y$ -space are included in  $Y_K$ . Hence, it is desirable to impose further constraints on  $y'$ . In the following, we consider problem (15) with the additional constraint

$$g(x, y') \leq 0. \quad (15c)$$

---

<sup>5</sup>Here, our setting deviates from [21], where the auxiliary variable naturally occurs in the objective function. As a result, the issue to be addressed in the following does not arise there.

We have already seen that feasibility of  $x \in X$  for (13) implies existence of  $y' \in Y$  such that  $(x, y')$  satisfies

$$[g(x, y) \leq 0] \vee [h(x, y) \leq h(x, y') - \epsilon] \quad \forall y \in Y.$$

Since  $Y_K \subseteq Y$ , this directly implies feasibility of  $(x, y')$  for (15) without (15c). Moreover, since  $y' \in Y$ , the condition is satisfied for  $y = y'$  as well, leading to

$$[g(x, y') \leq 0] \vee [h(x, y') \leq h(x, y') - \epsilon].$$

Since the second expression cannot be satisfied for  $\epsilon > 0$ , this implies that  $g(x, y') \leq 0$ . Thus, constraint (15c) is indeed implied by the  $\epsilon$ -approximation (13). Although (15c) is not needed to guarantee convergence of our semi-infinite solution procedure, it can lead to significant performance improvements in practice. We will come back to this point in Section 5.

We are now ready to present the solution procedure for the  $\epsilon$ -approximation (13) of the independent pessimistic bi-level problem (1).

**Algorithm.** Semi-infinite solution procedure for problem (13).

1. **Initialisation.** Set  $Y_0 = \emptyset$  and  $K = 0$  (iteration counter).
2. **Master Problem.** Solve problem (15):

$$\min_{(x, y') \in X \times Y} f(x)$$

subject to

$$\begin{aligned} [g(x, y_k) \leq 0] \vee [h(x, y_k) \leq h(x, y') - \epsilon] & \quad \forall y_k \in Y_K \\ g(x, y') \leq 0. & \end{aligned}$$

Let  $(x_K, y'_K)$  denote the/an optimal solution.

3. **Subproblem.** Calculate  $h_K = \max_{y \in Y} h(x_K, y)$ . Afterwards, solve the following optimisation problem.

$$\max_{y \in Y} \min \{g(x_K, y), h(x_K, y) - h_K + \epsilon\}.$$

Let  $y_{K+1}$  denote the/an optimal solution and  $\vartheta$  its objective value.

4. **Termination Criterion.** If  $\vartheta \leq 0$ , terminate:  $x_K$  is optimal for (13). Otherwise, set  $Y_{K+1} = Y_K \cup \{y_{K+1}\}$  and go to Step 2.

The algorithm proceeds by iterating between Steps 2 and 3. In Step 2, the current  $Y_K$ -relaxation is solved. Step 3 checks the existence of  $y_{K+1}$  with

$$y_{K+1} \in \mathcal{Y}_\epsilon(x_K) \wedge g(x_K, y_{K+1}) > 0.$$

If such a  $y_{K+1}$  exists, it is added to  $Y_K$  in Step 4 and the algorithm iterates. Otherwise,  $g(x_K, y) \leq 0$  is satisfied for all  $y \in \mathcal{Y}_\epsilon(x_K)$  and the algorithm terminates with  $x_K$  as optimal solution for (13).

It is commonly suggested to determine the ‘worst’ constraint violation when solving the subproblem in Step 3, that is, to solve

$$\max_{y \in Y} \{g(x_K, y) : h(x_K, y) > h_K - \epsilon\},$$

$K$	objective	$x_K$	$y_K$	$g(x_K, y_K)$
1	0.000	(0.000, 0.000)	(-1.000, -1.000)	1.414
2	0.000	(-1.000, -1.000)	(1.000, 1.000)	2.828
3	1.414	(0.000, 0.000)	(0.000, 0.000)	0.000

**Table 1:** Progress of our solution technique with and without tightened  $Y_K$ -relaxation (15c) when applied to Problem 1. The columns describe (in order) the iteration number, objective value, first-stage and second-stage solution and the degree of constraint violation.

instead. Note that the feasible region of this problem is not closed in general. For most practical applications, however, it suffices to solve the approximation

$$\max_{y \in Y} \{g(x_K, y) : h(x_K, y) \geq h_K - \epsilon\}, \quad (17)$$

which has a closed feasible region, instead. Although this approach typically leads to a superior performance in practice, convergence cannot be guaranteed. Indeed, if a maximiser  $y^*$  of (17) satisfies  $g(x_K, y^*) > 0$  and  $h(x_K, y^*) = h_K - \epsilon$ , the algorithm would enter an infinite loop. These issues notwithstanding, we will employ (17) in our numerical tests in Section 5.

Let us analyse the convergence properties of the algorithm. If the procedure terminates after  $K < \infty$  iterations, our previous discussion implies that the final solution  $x_K$  is optimal for (13). Assume that the algorithm does not terminate. Since  $X$  is compact, the resulting sequence  $\{x_K\}_K$  contains accumulation points. We can directly apply the argumentation developed by Blankenship and Falk in [6] to show that every accumulation point thus generated is optimal for (13).

## 5 Numerical Results

In this section, we employ the semi-infinite solution procedure from Section 4 to solve the  $\epsilon$ -approximation (13) of several independent pessimistic bi-level problems (1). The results will allow us both to verify the performance of the overall algorithm and to examine the use of tightening the  $Y_K$ -relaxation (15) by constraint (15c). All master and subproblems generated by our solution procedure were solved with BARON 6.0 [27].<sup>6</sup> We set the parameter  $\epsilon$  to  $10^{-3}$ .

**Problem 1.** *Convex first-stage and non-convex second-stage problem; the second-stage solution does not depend on the first-stage variables.*

$$\min_{-1 \leq x_1, x_2 \leq 1} \max_{-1 \leq y_1, y_2 \leq 1} \left\{ \|x - y\|_2 : y \in \arg \max_{-1 \leq y'_1, y'_2 \leq 1} \|y\|_2 \right\}.$$

The second-stage problem has four global optima  $(y_1^*, y_2^*) \in \{-1, 1\}^2$  that do not depend on the value of  $(x_1, x_2)$ , the first-stage variables. The first-stage problem aims to determine  $(x_1, x_2)$  whose maximal distance to any of these optima is minimised. Hence, the optimal solution is  $(x_1^*, x_2^*) = (0, 0)$ . The progress of our algorithm is documented in Table 1. For this example, the progress is the same for both the original and tightened  $Y_K$ -relaxation. In either case, the algorithm quickly determines the optimal solution.

<sup>6</sup>BARON Optimiser: <http://www.andrew.cmu.edu/user/ns1b/baron/baron.html>.

$K$	objective	$x_K$	$y_K$	$g(x_K, y_K)$
1	1.000	1.000	1.000	1.000
2	0.800	0.800	0.800	0.279
3	0.737	0.737	0.737	0.087
4	0.717	0.717	0.717	0.028
5	0.710	0.710	0.710	0.008
6	0.708	0.708	0.708	0.002
7	0.707	0.707	0.707	0.000

**Table 2:** Progress of our solution technique with tightened  $Y_K$ -relaxation (15c) when applied to Problem 2. See Table 1 for further details.

$K$	objective	$x_K$	$y_K$	$g(x_K, y_K)$
1	1.000	1.000	1.000	1.000
2	0.999	0.999	0.999	0.996
3	0.998	0.998	0.998	0.992
4	0.997	0.997	0.997	0.988
5	0.996	0.996	0.996	0.984
6	0.995	0.995	0.995	0.980
7	0.994	0.994	0.994	0.976
8	0.993	0.993	0.993	0.972
9	0.992	0.992	0.992	0.968
10	0.991	0.991	0.991	0.964
11	0.990	0.990	0.990	0.961
12	0.988	0.988	0.988	0.954
13	0.987	0.987	0.987	0.947
14	0.986	0.986	0.986	0.943
15	0.985	0.985	0.985	0.939

**Table 3:** Progress of our solution technique with original  $Y_K$ -relaxation, that is, without constraint (15c), when applied to Problem 2. See Table 1 for further details.

**Problem 2.** *Convex first-stage and second-stage problem.*

$$\max_{-1 \leq x \leq 1} \left\{ x : x^2 + y^2 \leq 1 \quad \forall y \in \arg \min_{-1 \leq y \leq 1} \|x - y\|_2 \right\}.$$

Contrary to Problem 1, the second-stage problem is convex. For a given first-stage decision  $x$ , the optimal second-stage solution is  $y^* = x$ . Hence, the optimal second-stage solution depends on the first-stage variables. The first-stage objective function maximises  $x$ , but the constraint limits  $x$  to  $x^* = 1/\sqrt{2} \approx 0.707$ , which is the optimal solution. The progress of our algorithm with tightened  $Y_K$ -relaxation is documented in Table 2. We see that the algorithm quickly determines the optimal solution. Table 3 shows part of the progress of our algorithm with the original  $Y_K$ -relaxation. In this case, large parts of the  $Y$ -space need to be examined, which leads to a very slow progress. Here, the value of the tightened relaxation becomes clearly visible.

**Problem 3.** *Production planning without investment option (see Section 1).*

$$\max_{x \in [0, 200]} \left\{ \min_{y \in \mathcal{Y}(x)} \{(50 - 0.3[x + y])x\} : \frac{x_A}{x + y} \geq 0.7 \quad \forall y \in \mathcal{Y}(x) \right\},$$

where

$$\mathcal{Y}(x) = \arg \max_{y \in [0, 100]} \{(50 - 0.3[x + y])y\}.$$

$K$	objective	$x_K$	$y_K$	$g(x_K, y_K)$
1	10000.00	0.000	82.755	10000.00
2	2083.04	84.315	41.753	1056.12
3	1544.56	125.712	21.057	794.16
4	1270.12	105.467	31.177	320.21
5	1170.25	62.457	52.682	204.78
6	1122.01	96.888	35.467	124.68
7	1066.08	81.091	43.366	39.21
8	1036.11	94.749	36.537	30.40
9	1026.87	81.091	43.367	6.03
10	1006.45	94.707	36.557	0.59
11	1005.86	94.707	36.559	0.00

**Table 4:** Progress of our solution technique with tightened  $Y_K$ -relaxation (15c) when applied to Problem 3. See Table 1 for further details.

The problem is equivalent to the first production planning problem from Section 1. Both the first-stage and the second-stage problem are convex. The progress of our algorithm with tightened  $Y_K$ -relaxation is reported in Table 4. The algorithm terminates at  $(x^*, y^*) = (94.707, 36.559)$  with a leader profit of 1005.86 and a market share of 72.15%. In Section 1 we have seen that the optimal solution is  $(x^*, y^*) = (89.75, 38.45)$  with a leader profit of 1035.49 and a market share of 70.01%. Here, we can see the conservative nature of the  $\epsilon$ -approximation: the determined solution constitutes a feasible but suboptimal solution for the original problem. More accurate results can be obtained for smaller values of  $\epsilon$ .

**Problem 4.** *Production planning with investment option (see Section 1).*

$$\max_{x \in [0, 200]} \left\{ \min_{y \in \mathcal{Y}(x)} \left\{ (50 - 0.3[x + y])x \right\} : \frac{x}{x + y} \geq 0.7 \quad \forall y \in \mathcal{Y}(x) \right\},$$

where

$$\mathcal{Y}(x) = \arg \max_{y \in [0, 100]} \max \left\{ (50 - 0.3[x + y] - 50)y, (50 - 0.3[x + y] - 38)y - 500 \right\}.$$

The problem is equivalent to the second production planning problem from Section 1. The second-stage problem is non-convex, which causes the feasible region of the first-stage problem to be non-closed. The progress of our algorithm with tightened  $Y_K$ -relaxation is reported in Table 5. The algorithm terminates at  $(x^*, y^*) = (104.864, 31.481)$  with a leader profit of 950.03 and a market share of 76.91%. In section 1 we have seen that the optimal solution is  $(x^*, y^*) = (103.34, 31.67)$  with a leader profit of 981.69 and a market share of 76.54%. Again, we can see the conservative nature of the  $\epsilon$ -approximation.

**Problem 5.** *Principal-agent problem (see Section 1).*

$$\max_{x \in [0, 1]} \min_{y \in \mathcal{Y}(x)} \left\{ 5(1 - x)y^2 \right\},$$

where

$$\mathcal{Y}(x) = \arg \max_{y \in [0, 1]} \left\{ 10 + 5xy^2 - y - (5[y - 1/2]^+)^2 \right\}.$$

The problem is equivalent to the principal-agent problem from Section 1. The second-stage problem is non-convex, which causes the objective function of the

$K$	objective	$x_K$	$y_K$	$g(x_K, y_K)$
1	10000.00	0.000	99.950	10000.00
2	2068.46	90.375	58.723	1592.13
3	1677.15	120.129	23.849	859.48
4	1199.00	104.614	31.604	243.38
5	996.54	102.726	52.548	645.43
6	971.34	144.215	11.806	510.74
7	950.03	104.864	31.481	0.00

**Table 5:** Progress of our solution technique with tightened  $Y_K$ -relaxation (15c) when applied to Problem 4. See Table 1 for further details.

$K$	objective	$x_K$	$y_K$	$g(x_K, y_K)$
1	5.000	0.000	0.000	5.000
2	0.949	0.473	0.529	0.210
3	0.849	0.394	0.521	0.029
4	0.822	0.393	0.521	0.000
5	0.822	0.393	0.521	0.000

**Table 6:** Progress of our solution technique with tightened  $Y_K$ -relaxation (15c) when applied to Problem 5. See Table 1 for further details.

first-stage problem to be discontinuous. The progress of our algorithm with tightened  $Y_K$ -relaxation is reported in Table 6. The algorithm terminates at  $(x^*, y^*) = (0.393, 0.521)$  with a principal profit of 0.822. We determined the same solution graphically in Section 1.

Summing up, all test instances were solved within moderate numbers of iterations. Moreover, the potential value of employing the tightened  $Y_K$ -relaxation has been demonstrated in one of the examples. Although we did not provide results for the application of the original  $Y_K$ -relaxation to Problems 3–5, we have encountered similar effects as in Problem 2 for these instances as well.

## 6 Conclusion

In this chapter, we studied the independent pessimistic bi-level problem. We analysed several properties of this problem class and related it to existing models from the literature. We discussed a convergent  $\epsilon$ -approximation of the problem and presented a semi-infinite solution approach. Although the  $\epsilon$ -approximation was primarily employed to guarantee tractability, it is also valuable in view of imperfect information and bounded rationality, two key issues in real-world decision making. To the best of our knowledge, this represents the first solution technique proposed for this problem class. We demonstrated the usability of our solution procedure with several examples.

We identified several promising directions for future research. First and foremost, the development of efficient local optimisation techniques for pessimistic bi-level problems would greatly enhance the practical relevance of this problem class. One possible approach is to employ the semi-infinite framework suggested in Section 4, but to solve the master problem (Step 2) merely to local optimality. In this case, significantly larger problems could be solved since the subproblems are of constant size and can be solved in parallel. The size of master problem, on the other hand, grows linearly with the number of iterations and becomes

increasingly non-convex due to the inclusion of disjunctive constraints. The key challenge is to analyse the convergence properties of the overall procedure.

Apart from the development of local optimisation procedures, several other directions seem interesting. Firstly, our investigation should be extended to dependent pessimistic bi-level and general multi-level problems. Secondly, a more thorough complexity analysis is needed for a better understanding of independent bi-level problems. Indeed, to the best of our knowledge, the complexity of the nonlinear independent optimistic bi-level problem is unknown. Similarly, there might be further relevant subclasses of the bi-level problem that can be solved in polynomial time. Finally, little has been reported about the constraint semantics of minimax and bi-level problems (see Section 2). A better understanding of the similarities and differences between these models could lead to novel problem formulations.

## 7 Acknowledgements

Support from EPSRC grant EP/C513584/1 is gratefully acknowledged.

## References

- [1] T. Basar and G. J. Olsder. *Dynamic Noncooperative Game Theory*. SIAM, 1999.
- [2] O. Ben-Ayed and C. E. Blair. Computational difficulties of bilevel linear programming. *Operations Research*, 38(3):556–560, 1990.
- [3] C. Berge and E. M. Patterson. *Topological Spaces*. Courier Dover Publications, 1997.
- [4] D. P. Bertsekas. Necessary and sufficient conditions for a penalty method to be exact. *Mathematical Programming*, 9(1):87–99, 1975.
- [5] B. Bhattacharjee, W. H. Green P. Lemonidis, and P. I. Barton. Global solution of semi-infinite programs. *Mathematical Programming*, 103(2):283–307, 2005.
- [6] J. W. Blankenship and J. E. Falk. Infinitely constrained optimization problems. *Journal of Optimization Theory and Applications*, 19(2):261–281, 1976.
- [7] K. C. Border. Correspondences. Technical report, California Institute of Technology, <http://www.hss.caltech.edu/~kcb/notes/correspondences.pdf>, February 2001.
- [8] B. Colson, P. Marcotte, and G. Savard. An overview of bilevel optimization. *Annals of Operations Research*, 153:235–256, 2007.
- [9] S. Dempe. *Foundations of Bilevel Programming*. Kluwer Academic Publishers, 2002.
- [10] X. Deng. Complexity issues in bilevel linear programming. In A. Migdalas, P. M. Pardalos, and P. Värbrand, editors, *Multilevel Optimization: Algorithms and Applications*, pages 149–164. Kluwer Academic Publishers, 1998.
- [11] D.-Z. Du and P. M. Pardalos, editors. *Minimax and Applications*. Kluwer Academic Publishers, 1995.
- [12] V. Dua, K. P. Papalexandri, and E. N. Pistikopoulos. Global optimization issues in multiparametric continuous and mixed-integer optimization problems. *Journal of Global Optimization*, 30(1):59–89, 2004.

- [13] N. P. Fáisca, V. Dua, B. Rustem, P. M. Saraiva, and E. N. Pistikopoulos. Parametric global optimisation for bilevel programming. *Journal of Global Optimization*, 38(4):609–623, 2007.
- [14] M. R. Garey and D. S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co, 1979.
- [15] R. Hettich and K. O. Kortanek. Semi-infinite programming: Theory, methods, and applications. *SIAM Review*, 35(3):380–429, 1993.
- [16] B. Holmstrom. Moral hazard and observability. *Bell Journal of Economics*, 10(1):74–91, 1979.
- [17] R. Horst and H. Tuy. *Global Optimization: Deterministic Approaches*. Springer, 1996.
- [18] R. G. Jeroslow. The polynomial hierarchy and a simple model for competitive analysis. *Mathematical Programming*, 32(2):146–164, 1985.
- [19] K. C. Kiwiel. A direct method of linearization for continuous minimax problems. *Journal of Optimization Theory and Applications*, 55(2):1573–2878, 1987.
- [20] Z.-Q. Luo, J.-S. Pang, and D. Ralph. *Mathematical Programs with Equilibrium Constraints*. Cambridge University Press, 1996.
- [21] A. Mitsos, P. Lemonidis, and P. I. Barton. Global solution of bilevel programs with a nonconvex inner program. *Journal of Global Optimization (accepted)*.
- [22] G. E. Monahan. Finding saddle points on polyhedra: Solving certain continuous minimax problems. *Naval Research Logistics*, 43(6):821–837, 1996.
- [23] P. Parpas and B. Rustem. An algorithm for the global optimization of a class of continuous minimax problems. *Journal of Optimization Theory and Applications (forthcoming)*.
- [24] J. O. Royset, E. Polak, and A. Der Kiureghian. Adaptive approximations and exact penalization for the solution of generalized semi-infinite min-max problems. *SIAM Journal on Optimization*, 14(1):1–34, 2003.
- [25] B. Rustem and M. Howe. *Algorithms for Worst-Case Design and Applications to Risk Management*. Princeton University Press, 2002.
- [26] B. Rustem, S. Zakovic, and P. Parpas. Convergence of an interior point algorithm for continuous minimax. *Journal of Optimization Theory and Applications*, 136(1):87–103, 2008.
- [27] N. V. Sahinidis. Baron: A general purpose global optimization software package. *Journal of Global Optimization*, 8(2):201–205, 1996.
- [28] H. Sasai. An interior penalty method for minimax problems with constraints. *SIAM Journal on Control and Optimization*, 12(4):643–649, 1974.
- [29] O. Stein and G. Still. On generalized semi-infinite optimization and bilevel optimization. *European Journal of Operational Research*, 142(3):444–462, 2002.
- [30] A. Tsoukalas, B. Rustem, and E. N. Pistikopoulos. A global optimization algorithm for generalized semi-infinite, continuous minimax with coupled constraints and bi-level problems. *Journal of Global Optimization*, 2008 (available via Online First).
- [31] A. Tsoukalas, W. Wiesemann, and B. Rustem. Global optimisation of pessimistic bi-level problems. Working Paper, Imperial College of Science, Technology and Medicine.
- [32] H. Tuy. Convex programs with an additional reverse convex constraint. *Journal of Optimization Theory and Applications*, 52(3):463–486, 1987.

- [33] L. N. Vicente and P. H. Calamai. Bilevel and multilevel programming: A bibliography review. *Journal of Global Optimization*, 5(3):291–306, 1994.
- [34] S. Žaković and B. Rustem. Semi-infinite programming and applications to min-max problems. *Annals of Operations Research*, 124:81–110, 2003.