TRANSFORMATION OF BILEVEL OPTIMIZATION PROBLEMS INTO SINGLE-LEVEL ONES

STEPHAN DEMPE

ABSTRACT. Bilevel optimization problems are hierarchical problems with a constraint set which is a subset of the graph of the solution set mapping of a second optimization problem. To investigate their properties and derive solution algorithms, their transformation into single-level ones is necessary. For this, various approaches have been developed. The first and most often used approach is to replace the lower level problem using its Karush-Kuhn-Tucker conditions. It has been shown that this results in a nonconvex optimization problem which is equivalent to the bilevel optimization problem if a global optimal solution is searched for. In case of local optimal solutions this is no longer the case: a local optimal solution of the single-level problem does not need to be related to a local optimal solution of the bilevel optimization problem. In this article transformation approaches using different dual problems for the lower level optimization problem are investigated. The resulting nonconvex single-level optimization problems are again not equivalent to the bilevel optimization problem provided their local optimal solutions are considered.

1. Introduction

Bilevel optimization (or bilevel programming) problems are hierarchical optimization problems where the feasible set of the upper level (or leader's) problem is constrained by the graph of the solution set mapping of the lower level (or follower's) problem. This problem has been first formulated by H.v. Stackelberg [26] in 1934 in the context of an economic situation. The now so-called Stackelberg problem is a special case of the bilevel optimization problem where the sets of possible selections of both players do not depend on their opponents. About 40 years later, this problem has been introduced to the optimization community, see [4, 13]. A recent bibliography [8] counts more than 1500 references, especially also over 65 Ph.D. thesis at universities from all over the world.

If the solution set mapping of the lower level problem does not reduce to a (continuous) function, the problem is not well defined. Since, in the Stackelberg game, the leader cannot influence the follower's selection, the objective function value of the leader's problem is in general not uniquely defined, it can take different values for different selections of the follower. Different ways out are discussed in such a situation, best known are the optimistic (or strong) and the pessimistic (or weak) approaches. The one used here is the optimistic version of the bilevel optimization problem.

Key words and phrases. bilevel optimization; mathematical programs with complementarity constraints; necessary optimality conditions; Lagrange dual problem; Wolfe duality; Mond-Weir dual problem; optimal value function transformation.

Due to the implicit nature of its constraints, the bilevel optimization problem needs to be replaced by a single-level one. This can be done using different approaches: historically the first approach was to replace the lower level problem using its Karush-Kuhn-Tucker (KKT) conditions. Mirrlees [18] found that this is only possible if the lower level problem is convex. Then, the KKT conditions are sufficient and necessary optimality conditions provided that a regularity condition is satisfied for that problem for all possible selections of the leader. It is easy to see that this results in an equivalent single-level problem provided that a global optimal solution is searched for. Bilevel optimization problems are nonconvex ones which makes the investigation of local optimal solutions important. Dempe and Dutta investigated in [6] the relations between local optimal solutions of the bilevel optimization problem and its KKT transformation.

The resulting single-level optimization problem is a so-called mathematical program with complementarity (or equilibrium) constraints (MPCC). In [24], Scheel and Scholtes have shown that the Mangasarian-Fromovitz constraint qualification (MFCQ) is violated at all feasible solutions. This makes both the derivation of necessary optimality conditions and the convergence proof of solution algorithms difficult.

In relation with the derivation of solution algorithms for bilevel optimization problems the idea of replacing the lower level problem applying Lagrange duality has been developed, see e.g. White and Anandalingam [27]. Section 3.2 will be devoted to this transformation in more details. We will especially also find a relation of this transformation with the one using the optimal value function of the lower level problem. The latter one has been shown to be fully equivalent with the bilevel optimization problem by Outrata in [21]. We will see that the MFCQ is again violated at all feasible points.

Wolfe [28] has suggested another dual problem for convex optimization problems which has been applied by Li et al. [14] to bilevel optimization problems. Here it is shown by an example that the MFCQ can be satisfied sometimes. Nevertheless we will see in Section 3.3 that local optimal solutions of the transformed nonconvex problem are again in general not related to a local optimal solution of the bilevel optimization problem.

One further dual problem going back to Mond and Weir [19] has been used to transform the bilevel optimization problem to a single-level one by Li et al. [**?**]. Section 3.18 will be devoted to this approach. Again, focus will be on local optimal solutions for this nonconvex optimization problem.

2. The optimistic bilevel optimization problem

For sufficiently smooth functions $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^p$, we consider the lower level problem

$$\min_y \{f(x,y) : g(x,y) \le 0\}, \tag{2.1}$$

its optimal value function

$$\varphi(x) := \min_y \{f(x,y) : g(x,y) \le 0\} : \mathbb{R}^n \to \mathbb{R} \tag{2.2}$$

and its solution set mapping

$$\Psi(x) := \{y : f(x,y) \le \varphi(x),\ g(x,y) \le 0\} : \mathbb{R}^n \rightrightarrows \mathbb{R}^m. \tag{2.3}$$

Then, again for sufficiently smooth functions $F : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $G : \mathbb{R}^n \to \mathbb{R}^q$, the *(optimistic) bilevel optimization problem*

$$\min_{x,y}\{F(x,y) : G(x) \le 0,\ (x,y) \in \text{gph } \Psi\} \tag{2.4}$$

can be defined, where

$$\text{gph } \Psi := \{(x,y) : y \in \Psi(x)\}$$

denotes the graph of the solution set mapping of (2.1). Problem (2.4) is also called the upper level optimization problem.

The Mangasarian-Fromovitz constraint qualification (MFCQ) is satisfied for (2.1) at a point $(\overline{x}, \overline{y})$ if there exists $\overline{d}$ satisfying

$$\forall\ j \in I(\overline{x}, \overline{y}) :\ \nabla_y g_j(\overline{x}, \overline{y})\overline{d} < 0.$$

Here, $I(\overline{x}, \overline{y}) := \{j : g_j(\overline{x}, \overline{y}) = 0\}$ is the index set of active constraints.

Definition 2.1. A point-to-set (or multivalued) mapping $Z : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is called *upper semicontinuous* at $\overline{x}$, if for each open set V with $Z(\overline{x}) \subseteq V$ there is an open set $U \ni \overline{x}$ such that $Z(x) \subseteq V$ for all $x \in U$.

Theorem 2.2 (Bank et al. [2])**.** *Let $\overline{x}$ with $G(\overline{x}) \le 0$ be fixed. If for problem (2.1) the assumption (MFCQ) is satisfied at each $\widehat{y} \in Y(\overline{x})$ with $Y(\overline{x}) := \{y : g(\overline{x}, y) \le 0\}$ and $Y(\overline{x}) \ne \emptyset$ is compact then, the mapping $x \mapsto \Psi(x)$ is upper semicontinuous at $\overline{x}$ and the function $x \mapsto \varphi(x)$ is continuous at $\overline{x}$.*

If the assumptions of Theorem 2.2 are satisfied at every point $\overline{x}$ with $G(\overline{x}) \le 0$ then, the graph of the mapping Ψ is compact. As a consequence of this fact, using the famous Weierstraß' Theorem, we derive

Corollary 2.3. *Let $X := \{x : G(x) \le 0\}$ be nonempty and compact. If the assumptions in Theorem 2.2 are satisfied at all points $\overline{x} \in X$ then, problem* (2.4) *has an optimal solution.*

Problem (2.4) is an hierarchical, nonconvex optimization problem. For its investigation, the lower level problem needs to be replaced. This will be the topic in the rest of the article.

3. Transformations

3.1. **Use of Karush-Kuhn-Tucker conditions of the lower level problem.**

If the lower level problem (2.1) is replaced using the Karush-Kuhn-Tucker conditions, the following problem arises:

$$\begin{aligned} &F(x,y) \longrightarrow \min_{x,y,u} \\ &G(x) \le 0 \\ &\nabla_y f(x,y) + u^\top \nabla_y g(x,y) = 0 \\ &0 \ge g(x,y) \perp u \ge 0, \end{aligned} \tag{3.1}$$

where the last line means that $g_i(x,y) \le 0,\ u_i \ge 0,\ u_i g_i(x,y) = 0$ is satisfied for all $i = 1, 2, \ldots, p$. Clearly, this is possible, it the Mangasarian-Fromovitz constraint qualification is satisfied for the lower level problem at all points $(x,y) \in \text{gph}\Psi$. Problem (3.1) is a so-called mathematical program with complementarity constraints (MPCC), see e.g. the monographs [15, 20].

Using the unconstrained bilevel optimization problem

$$\begin{aligned} &F(x,y) := (x-2)^2 + (y-1)^2 \longrightarrow \min \\ &\text{s.t.: } y \text{ minimizes } f(x,y) := -xe^{-(y+1)^2} - e^{-(y-1)^2}, \end{aligned} \tag{3.2}$$

Mirrlees [18] has found that the global optimal solution of this problem cannot be computed using the KKT transformation (3.1). The reason for this is that, for equivalence between (2.4) and (3.1), the KKT conditions need to be necessary and sufficient optimality conditions for global optima. Nonconvex optimization problems can have local optima and even stationary points thus making the feasible set of (3.1) larger than that of (2.4) in general. Using recent results by Martínez-Legaz [17], the convexity assumptions can be weakened.

Theorem 3.1 (Dempe, Dutta [6])**.** *Let the lower level problem (2.1) be convex, MFCQ be satisfied at all points $(\overline{x}, y) \in$ gph Ψ. Then, for each global optimal solution $(\overline{x}, \overline{y})$ of* (2.4) *there exists $\overline{u} \in \Lambda(\overline{x}, \overline{y}) := \{u : \nabla_y L(\overline{x}, \overline{y}, u) = 0,\ 0 \le u \perp g(\overline{x}, \overline{y}) \le 0\}$ such that $(\overline{x}, \overline{y}, \overline{u})$ is a global optimal solution of* (3.1)*. Vice versa, for each global optimal solution $(\overline{x}, \overline{y}, \overline{u})$ of* (3.1)*, the point $(\overline{x}, \overline{y})$ is a global optimal solution of* (2.4)*.*

The following example shows that the relations between both problems are more difficult in case of local optimal solutions.

Example 3.2 (Dempe, J. Dutta, [6])**.** *Consider the linear lower level problem*

$$\min_y \{-y : x + y \le 1,\ -x + y \le 1\} \tag{3.3}$$

and the upper level problem

$$\min\{(x-1)^2 + (y-1)^2 : (x,y) \in \text{gph } \Psi\} \tag{3.4}$$

This problem has the unique optimal solution $(\overline{x}, \overline{y}) = (0.5, 0.5)$ and no local optimal solutions.

Consider the point $(x^0, y^0) = (0, 1)$. Then,

$$\Lambda(x^0, y^0) = \begin{cases} \{(1,0)\} & \text{if } x > 0 \\ \{(0,1)\} & \text{if } x < 0 \\ \text{conv}\{(1,0),(0,1)\} & \text{if } x = 0 \end{cases}$$

where conv$\,A$ denotes the convex hull of the set A.

Take $(x^0, y^0, u_1^0, u_2^0) = (0, 1, 0, 1)$

$u_2 > 0$ *implies* $y = x + 1 \Rightarrow x \le 0 \Rightarrow (x-1)^2 + (y-1)^2 = (x-1)^2 + x^2 \ge 1.$

Hence, this point is a local optimal solution of the MPCC but does <u>not</u> correspond to a local optimal solution of the bilevel optimization problem.

Based on results by Gauvin [10] and Robinson [23] it can be shown that the point-to-set mapping $(x, y) \mapsto \Lambda(x, y)$ is upper semicontinuous at $(\overline{x}, \overline{y}) \in$ gph Ψ provided that (MFCQ) is satisfied at all points $y \in \Psi(\overline{x})$. Let $(\overline{x}, \overline{y}, \overline{u})$ be a local minimum of (3.1) and $\{(x^k, y^k)\}_{k=1}^\infty \subset$ gph Ψ converging to $(\overline{x}, \overline{y}) \in$ gph Ψ. Then, there exists a sequence $\{u^k\}_{k=1}^\infty$ with $u^k \in \Lambda(x^k, y^k)$ for all k converging to $\widehat{u} \in \Lambda(\overline{x}, \overline{y})$. Unfortunately, if $\Lambda(\overline{x}, \overline{y})$ does not reduce to a singleton, $\overline{u} \ne \widehat{u}$ implying that (x^k, y^k, u^k) does not converge to $(\overline{x}, \overline{y}, \overline{u})$ implying that $(\overline{x}, \overline{y})$ is in general not a local minimum of (2.4). This implies

Theorem 3.3 (Dempe, Dutta, [6]). *Let the lower level problem (2.1) be convex, MFCQ be satisfied at all points* $(\overline{x}, y) \in$ *gph* Ψ*. Then, the point* $(\overline{x}, \overline{y})$ *is a local optimal solution of problem* (2.4) *if and only if* $(\overline{x}, \overline{y}, \overline{u})$ *is a local optimal solution for problem* (3.1) *for all* $\overline{u} \in \Lambda(\overline{x}, \overline{y})$.

The situation in Example 3.2 that a local optimal solution of problem (3.1) is not related to a local optimal solution of the bilevel problem (2.4) is possible in the case when the set of Lagrange multipliers in the lower level problem (2.1) does not reduce to a singleton which is not possible when the linear independence constraint qualification is satisfied.

Let $(\overline{x}, \overline{y}) \in$ grp Ψ. The linear independence constraint qualification (LICQ) is satisfied at $(\overline{x}, \overline{y})$ if the gradients $\nabla_y g_i(\overline{x}, \overline{y}) : i \in I(\overline{x}, \overline{y})$ are linearly independent. The LICQ is in nonlinear differentiable optimization generically satisfied at local optimal solutions of differentiable optimization problems, see e.g. [25]. The same result with other words is given in

Theorem 3.4 ([11]). *Consider the (sufficiently smooth) optimization problem*

$$\min_x \{f(x) + b^\top x : h(x) + d^1 = 0,\ g(x) + d^2 \leq 0\} \tag{3.5}$$

depending on $(b, d^1, d^2) \in \mathbb{R}^n \times \mathbb{R}^q \times \mathbb{R}^p$*. Then, the set of perturbations* (b, d^1, d^2) *such that LICQ fails at some feasible point of* (3.5) *has Lebesgue measure equal to zero.*

Using the parameterized Sard lemma [11] and linear perturbations it can be shown that the LICQ is satisfied for almost all feasible points of the lower level problem (2.1) provided that the functions g_i do not depend on x, see Aussel et al. [1]. Example 3.2 explains that this is not the case in general if the constraints depend also on x even if $g(x, y) = Ax + By - b$.

3.2. **Lagrange duality.** We can also use the related Lagrange duality for the lower level problem. Let, for $u \geq 0$,

$$\varphi_L(x, u) := \min_y \{L(x, y, u)\} \tag{3.6}$$

with $L(x, y, u) = f(x, y) + u^\top g(x, y)$. The dual optimization problem is

$$\max_u \{\varphi_L(x, u) : u \geq 0\} \tag{3.7}$$

and we have weak duality:

Theorem 3.5. *For* $x = \widehat{x}$ *let* $\widehat{y}$ *be feasible for* (2.1) *and* $\widehat{u} \geq 0$*. Moreover, let* $y \mapsto f(\widehat{x}, y)$ *and* $y \mapsto g_i(\widehat{x}, y)$ *be convex for all* i*. Then,*

$$f(\widehat{x}, \widehat{y}) \geq \varphi_L(\widehat{x}, \widehat{u}).$$

If the MFCQ is satisfied for (2.1) *and* $x = \widehat{x}$ *then,*

$$\varphi(\widehat{x}) = \max_u \{\varphi_L(\widehat{x}, u) : u \geq 0\}$$

provided that $\varphi(\widehat{x})$ *is finite.*

This makes it possible to replace (2.4) by

$$
\begin{aligned}
&F(x,y) \to \min_{x,y,u}\\
&G(x) \le 0\\
&g(x,y) \le 0 \qquad\qquad (3.8)\\
&u \ge 0\\
&f(x,y) \le \varphi_L(x,u),
\end{aligned}
$$

provided that the last inequality can be satisfied, i.e., there is no duality gap.

Theorem 3.6. *Let the problem* (2.1) *be convex for fixed x and let the MFCQ be satisfied for all feasible y . Then we have:*

(1) *If $(\overline{x},\overline{y})$ is a global optimal solution of* (2.4) *then there is $\overline{u}$ such that $(\overline{x},\overline{y},\overline{u})$ is a global optimal solution of* (3.8) *and the optimal objective function values coincide.*
(2) *Vice versa, if $(\overline{x},\overline{y},\overline{u})$ is a global optimal solution of* (3.8) *then, $(\overline{x},\overline{y})$ is a global optimal solution of* (2.4).
(3) *If $(\overline{x},\overline{y})$ is a local optimal solution of* (2.4) *and $\overline{u}\in\Lambda(\overline{x},\overline{y})$ then, $(\overline{x},\overline{y},\overline{u})$ is a local optimal solution of* (3.8).

Proof. (1) If $(\overline{x},\overline{y})$ is a global optimal solution of (2.4) then, due to strong duality, there is $\overline{u}$ such that $(\overline{x},\overline{y},\overline{u})$ is feasible for (3.8). If $(\overline{x},\overline{y},\overline{u})$ would not be globally optimal then, there exists a feasible point $(\widehat{x},\widehat{y},\widehat{u})$ for (3.8) with $F(\widehat{x},\widehat{y}) < F(\overline{x},\overline{y})$ and $\widehat{y}\in\Psi(\widehat{x})$ by weak duality. This violates global optimality of $(\overline{x},\overline{y})$ for (2.4).
(2) If $(\overline{x},\overline{y},\overline{u})$ is a global optimal solution of (3.8) then, by weak duality $\overline{y}\in\Psi(\overline{x})$ and $(\overline{x},\overline{y})$ is feasible for (2.4). If this point would not be globally optimal then, there exists a feasible point $(\widehat{x},\widehat{y})$ for (2.4) and, due to our assumptions, $\widehat{u}\ge 0$ such that $(\widehat{x},\widehat{y},\widehat{u})$ is feasible for (3.8) with $F(\widehat{x},\widehat{y}) < F(\overline{x},\overline{y})$. Since this violates our assumptions, the assertion is correct.
(3) If $\overline{u}\in\Lambda(\overline{x},\overline{y}) = \{u : \nabla_y L(\overline{x},\overline{y},u) = 0,\ 0\le u\perp g(\overline{x},\overline{y})\le 0\}$ then

$$f(\overline{x},\overline{y}) = L(\overline{x},\overline{y},\overline{u}) = \varphi_L(\overline{x},\overline{u})$$

by convexity. Consequently, $(\overline{x},\overline{y},\overline{u})$ is feasible for (3.8). If $(\overline{x},\overline{y},\overline{u})$ is not locally optimal then, there exists a sequence (x^k,y^k,u^k) of feasible points with $F(x^k,y^k) < F(\overline{x},\overline{y})$ for all k converging to $(\overline{x},\overline{y},\overline{u})$. Then, by weak duality, (x^k,y^k) is a sequence of feasible points for (2.4). This violates local optimality of $(\overline{x},\overline{y})$.

□

Example 3.7. *Consider Example* (3.2). *Then,*

$$L(x,y,u) = -y+u_1(x+y-1)+u_2(-x+y-1) = y(-1+u_1+u_2)+x(u_1-u_2)-u_1-u_2$$

which has a finite minimum only for $u_1+u_2 = 1$. Then,

$$\min_y L(x,y,u) = x(u_1-u_2)-u_1-u_2.$$

We consider points near $(x,y,u) = (0,1,0,1)$. Then, the last inequality in (3.8) *reads*

$$-y \le x(u_1-u_2)-u_1-u_2 \text{ or } x(u_1-u_2)+y = x(1-2u_2)+y \ge 1$$

by $u_1 + u_2 = 1$. Since the constraints $x + y \leq 1,\ -x + y \leq 1$ of the lower level problem are also part of (3.8) *we derive $1 \leq x + y - 2u_2x \leq 1 - 2u_2x$. Near $u_2 = 1$ we have $u_2 > 0$. This implies $x \leq 0$ or that $(0, 1, 0, 1)$ is a local minimum of* (3.8)*.*

By convexity, problem (3.8) can equivalently be written as follows:

$$
\begin{aligned}
&F(x,y) \to \min_{x,y,u}\\
&G(x) \leq 0\\
&g(x,y) \leq 0\\
&u \geq 0\\
&f(x,y) \leq L(x,z,u),\\
&\nabla_z L(x,z,u) = 0
\end{aligned}
\tag{3.9}
$$

The last inequality in (3.8) implies

$$f(x,y) \leq f(x,\widehat{y}) + u^\top g(x,\widehat{y})\ \forall\ \widehat{y}.$$

For $g(x,\widehat{y}) \leq 0$ we especially obtain

$$f(x,y) \leq \widehat{\varphi}(x) := \min_y \{f(x,y) : g(x,y) \leq 0\}.$$

Note that this is correct for all Lagrange multipliers at the same time. This has the consequence formulated above in Theorem 3.3.

The transformation using the optimal value function of the lower level problem follows:

$$
\begin{aligned}
&F(x,y) \longrightarrow \min_{x,y}\\
&G(x) \leq 0\\
&g(x,y) \leq 0\\
&f(x,y) \leq \widehat{\varphi}(x).
\end{aligned}
\tag{3.10}
$$

Problem (3.10) is fully equivalent to (2.4) even without convexity assumption for the lower level problem [21]. Note that we avoided the explicit use of the Lagrange multiplier here.

Theorem 3.8 ([22, 29])**.** *The (nonsmooth) Mangasarian-Fromovitz-constraint qualification is violated at all feasible points of problem* (3.10)*.*

The proof of this theorem is not very difficult. It is well known that the Mangasarian-Fromowitz CQ is satisfied iff there is no irregular Lagrange multiplier. The existence of an irregular Lagrange multiplier for (3.10) follows since each feasible point for problem (3.10) is a global optimal solution of the problem

$$\min_{x,y} \{f(x,y) - \varphi(x) : g(x,y) \leq 0\}$$

with optimal objective function value zero. The resulting necessary Fritz-John optimality conditions determine an irregular Lagrange multiplier for (3.10).

3.3. **Wolfe duality.** Wolfe [28] has formulated a dual problem for the differentiable convex optimization problem (2.1):

$$
\begin{aligned}
&f(x,y) + u^\top g(x,y) \to \max_{y,u}\\
&\nabla_y L(x,y,u) = 0,\ u \geq 0.
\end{aligned}
\tag{3.11}
$$

Again, $L(x,y,u) = f(x,y) + u^\top g(x,y)$. Weak duality for the pair of problems (2.1) and (3.11) can be shown as follows for feasible points $\widehat{y}$ of (2.1) and (y,u) of (3.11) using convexity, see Wolfe [28]:

$$\begin{aligned} f(x,\widehat{y}) - f(x,y) &\geq \nabla_y f(x,y)(\widehat{y}-y) \\ = -\sum_{i=1}^p u_i \nabla_y g_i(x,y)(\widehat{y}-y) &\geq -\sum_{i=1}^p u_i (g_i(x,\widehat{y}) - g_i(x,y)) \\ \geq \sum_{i=1}^p u_i g_i(x,y). \end{aligned}$$

Without convexity this is not correct since the local optimal solution of (2.1) with the largest objective function value is an optimal solution of (3.11).

Strong duality can be verified provided that a constraint qualification as Slater's condition is satisfied, cf. Wolfe [28]. Then, $\widehat{y}$ is an optimal solution of (2.1) if and only if the saddle point inequalities are satisfied:

$$L(x,\widehat{y},u) \leq L(x,\widehat{y},\widehat{u}) \leq L(x,y,\widehat{u}) \ \forall\ u \geq 0,\ \forall y.$$

Since $y \mapsto L(x,y,u)$ is convex we have

$$L(x,\widehat{y},u) - L(x,y,u) \geq \nabla_y L(x,y,u)(\widehat{y}-y)$$

implying $L(x,\widehat{y},u) = L(x,y,u)$ for all $\widehat{y}, y$ satisfying

$$\nabla_y L(x,y,u) = 0.$$

Consequently, for all optimal solutions $\widehat{y}$ of (2.1) there exists $\widehat{u} \geq 0$ with

$$\begin{aligned} L(x,\widehat{y},\widehat{u}) &= \max_u \{L(x,\widehat{y},u) : u \geq 0\} \\ &\geq \max_u \{L(x,\widehat{y},u) : \nabla_y L(x,\widehat{y},u) = 0,\ u \geq 0\} \\ &= \max_{y,u} \{L(x,y,u) : \nabla_y L(x,y,u) = 0,\ u \geq 0\} \\ &\geq L(x,\widehat{y},\widehat{u}). \end{aligned}$$

The reverse result is not correct in general.

Theorem 3.9 ([28, 5])**.** *Let $\widehat{y}$ be an optimal solution of the convex problem* (2.1) *and let Slater's constraint qualification be satisfied. Then, there exists $\widehat{u} \geq 0$ such that $(\widehat{y},\widehat{u})$ is an optimal solution of problem* (3.11)*. Conversely, let $(\widehat{y},\widehat{u})$ be an optimal solution of* (3.11) *and assume $\nabla_y^2 f(x,\widehat{y}) + \widehat{u}\nabla^2 g(x,\widehat{y})$ be regular. Then, $\widehat{y}$ is an optimal solution of* (2.1).

The first assertion has been shown above. For the second one: By Fritz-John optimality conditions, there exist $\alpha \geq 0, v, w \leq 0$ such that:

$$\alpha(\nabla_y f(x,\widehat{y}) + \widehat{u}^\top \nabla_y g(x,\widehat{y})) + v^\top(\nabla_y^2 f(x,\widehat{y}) + \widehat{u}^\top \nabla_y^2 g(x,\widehat{y})) = 0 \tag{3.12}$$

$$\alpha g(x,\widehat{y}) + v^\top \nabla_y g(x,\widehat{y}) - w = 0 \tag{3.13}$$

$$w^\top \widehat{u} = 0 \tag{3.14}$$

$$(\alpha, v, w) \neq 0 \tag{3.15}$$

Using the equality in (3.11), equation (3.12) gives $v = 0$ due to the regularity assumption. If $\alpha = 0$, then $w = 0$ contradicting (3.15). Equation (3.13) implies then $g(x,\widehat{y}) = \frac{w}{\alpha} \leq 0$ which means that $\widehat{y}$ is feasible for (2.1) and, inserting this into (3.14), that complementarity slackness $\widehat{u}^\top g(x,\widehat{y}) = 0$ is satisfied. Now, weak duality or (3.12) show that the Karush-Kuhn-Tucker conditions are satisfied which are sufficient for optimality of $\widehat{y}$.

Clearly, the regularity condition in the last theorem can be replaced by strong convexity of the objective function with respect to y in (2.1), see Kanniappan [12].

This can be a bit generalized, see [14]:

Definition 3.10. A differentiable function $f : \mathbb{R}^n \to \mathbb{R}$ is called pseudoconvex if $\nabla f(x^2)^\top (x^1 - x^2) \geq 0$ implies $f(x^1) \geq f(x^2)$.

Theorem 3.11 (Li et al. [14])**.** *Assume that the function $y \mapsto L(x, y, u)$ is pseudoconvex for any $u \geq 0$. Then, weak duality holds for the pair* (2.1) *and* (3.11)*.*

If (x, y, u) are feasible for (3.11) then $\nabla_y L(x, y, u) = 0$ which implies $L(x, z, u) \geq L(x, y, u)$ for all feasible solutions z of (2.1). Thus, by $u \geq 0$,

$$f(x, z) \geq f(x, z) + u^\top g(x, z) = L(x, z, u) \geq L(x, y, u)$$

which implies the assertion.

We also have strong duality provided some constraint qualification is satisfied.

Theorem 3.12 (Strong Wolfe duality, Li et al. [14])**.** *Assume that $y \mapsto L(x, y, u)$ is pseudoconvex for any $u \geq 0$ and that Slater's condition is satisfied for the primal problem* (2.1)*, and $\overline{y}$ is an optimal solution of* (2.1)*. Then, there exists an optimal solution $(\overline{z}, \overline{u})$ of* (3.11) *such that*

$$\begin{aligned} &\min_y \{f(x, y) : g(x, y) \leq 0\} = f(x, \overline{y}) = \\ &L(x, \overline{z}, \overline{u}) = \max_{y,u} \{L(x, y, u) : \nabla_y L(x, y, u) = 0, u \geq 0\}. \end{aligned}$$

Since Slater's condition is satisfied, for an optimal solution $\overline{y}$ of (2.1) there exists $\widehat{u} \geq 0$ such that

$$\nabla_y L(x, \overline{y}, \widehat{u}) = 0, \widehat{u} \geq 0, \widehat{u}^\top g(x, \overline{y}) = 0.$$

Hence, $(x, \overline{y}, \widehat{u})$ is feasible for (3.11). Assume that this point is not a global optimal solution of (3.11). The, there exists a feasible point (x, y, u) such that

$$L(x, y, u) > L(x, \overline{y}, \widehat{u}) = f(x, \overline{y})$$

which contradicts weak duality. Hence, the theorem is true.

Pseudoconvexity can be replaced by invexity [3] which essentially means that validity of the Karush-Kuhn-Tucker conditions at some point $(\overline{y}, \overline{u})$ implies that $\overline{y}$ is a global minimum. Also, the Slater condition can be replaced by the Guignard constraint qualification.

Note that the optimal solution of (3.11) does not need to be an optimal solution of (2.1) unless we have more assumptions, cf. the remarks above.

Now, replace the bilevel optimization problem by problem (3.9):

$$\begin{aligned} &F(x, y) \to \min_{x,y,z,u} \\ &G(x) \leq 0 \\ &g(x, y) \leq 0, \; u \geq 0 \\ &f(x, y) \leq L(x, z, u) \\ &\nabla_z f(x, z) + u^\top \nabla_z g(x, z) = 0 \end{aligned} \tag{3.16}$$

Theorem 3.13 (Li et al. [14])**.** *Let strong Wolfe duality be satisfied for the problems* (2.1) *and* (3.11)*. Then, a global optimal solution* (x^0, y^0) *of* (2.4) *is related to a global optimal solution* (x^0, y^0, z^0, u^0) *of* (3.9) *and vice versa.*

This is a reformulation of the first two assertions in Theorem 3.6.

Strong Wolfe duality cannot be satisfied if the lower level problem has local optimal solutions which are not globally optimal since weak duality is violated in that case. Moreover, if $g(x, y) \not\equiv 0$, problem (3.9) is a nonconvex optimization problem.

Example 3.14. *Consider the problem in Example 3.2. Then, problem* (3.9) *reads as*

(3.17)

$$\begin{aligned} &\min\{(x-1)^2 + (y-1)^2 : -y \le -z + u_1(x+z-1) + u_2(-x+z-1), \\ &\quad -1 + u_1 + u_2 = 0,\ u \ge 0,\ x+y \le 1,\ -x+y \le 1\} \\ =&\min\{(x-1)^2 + (y-1)^2 : -y \le (u_1 - u_2)x - 1,\ -1 + u_1 + u_2 = 0,\ u \ge 0, \\ &\quad x+y \le 1,\ -x+y \le 1\} \end{aligned}.$$

Consider the point $(0, 1, 0, 1)^\top$*. In a neighborhood of this point,* $u_1 - u_2 < 0$*. Hence, if* $x > 0$*, we have* $y > 1$ *from the first constraint. But this violates* $x + y \le 1$*. This means that* $x \le 0$ *in a neighborhood of this point. Adding the constraint* $x \le 0$ *to the last problem, the point* $(0, 1, 0, 1)^\top$ *is optimal. This implies that the point* $(0, 1, 0, 1)^\top$ *is a local optimum of problem* (3.17)*. But this point is not a local optimal solution in Example 3.2.*

Theorem 3.15. *Assume that problem* (2.1) *is convex and that Slater's condition is satisfied for the lower level problem. Let* $(\overline{x}, \overline{y})$ *be a local optimal solution for* (2.4)*. Then,* $(\overline{x}, \overline{y}, \overline{z}, \overline{u})$ *is a local optimal solution for each strong Wolfe dual solution* $(\overline{z}, \overline{u})$ *to* $\overline{y}$*.*

Proof. That is a repetition of the third assertion in Theorem 3.6. □

For the opposite implication we need to assume that $(\overline{x}, \overline{y}, \overline{z}, \overline{u})$ is locally optimal for all $(\overline{z}, \overline{u})$ which satisfy strong Wolfe duality for (2.1) at $(\overline{x}, \overline{y})$.

Lemma 3.16. *Let* $g, h : \mathbb{R}^m \mapsto \mathbb{R}^k$ *be differentiable functions and*

$$\{x : g(x) \le 0\} \subseteq \{x : h(x) \le 0\}.$$

If, for x^0 *with* $g(x^0) \le 0$ *the Mangasarian-Fromowitz CQ is violated for* $\{x : h(x) \le 0\}$ *then, it is also violated for* $\{x : g(x) \le 0\}$*.*

If, arguing by contradiction,

$$\nabla g_i(x^0) d^0 < 0 \text{ for all } i \text{ with } g_i(x^0) = 0,$$

then $g(x^0 + \alpha d^0) < 0$ for sufficiently small $\alpha > 0$. Existence of that direction and the assumed inclusion would imply that $h(x^0 + \alpha d^0) < 0$, too.

This implies the following result:

Theorem 3.17. *Assume that the lower level problem* (2.1) *is convex, Slater's condition be satisfied. Let* (x^0, y^0) *be a feasible solution of problem* (2.4)*. Let* (x^0, y^0, z^0, u^0) *be feasible for* (3.9)*. If* $y^0 = z^0$ *then, the Mangasarian-Fromovitz CQ is violated.*

Proof. Since $z^0 = y^0$ we derive

$$f(x^0, y^0) \le L(x^0, z^0, u^0) = L(x^0, y^0, u^0) = f(x^0, y^0) + u^{0\top} g(x^0, y^0)$$

and the feasible set of (3.9) is a subset of the set

$$\{(x, y, z, u) : g(x, y) \le 0,\ u \ge 0, u^\top g(x, y) \ge 0\}.$$

Since the MFCQ is violated for this set, it is also violated for (3.9). □

This is no longer correct in the case that $y^0 \ne z^0$. In that case, the MFCQ is not necessarily violated at all feasible solutions, see [14] for an example.

3.4. **Mond-Weir duality.** In [19] another duality concept has been introduced for the lower level problem (2.1).

Definition 3.18. A differentiable function f is called quasiconvex if

$$f(y) - f(x) \le 0 \Rightarrow \nabla f(x)^\top (y - x) \le 0.$$

Mangasarian [16] points out that whereas some results (such as sufficiency and converse duality) hold if, in (2.1) for fixed x, f is only pseudoconvex and g quasiconvex in y, Wolfe duality does not hold for such functions.

Mond and Weir [19] formulated the following problem using the data of (2.1):

$$\begin{aligned} f(x, y) &\to \max_{y,u} \\ \nabla_y f(x, y) + u^\top \nabla_y g(x, y) &= 0 \\ u^\top g(x, y) &\ge 0 \\ u &\ge 0. \end{aligned} \tag{3.18}$$

Theorem 3.19. *[Mond and Weir [19]] If $\overline{y}$ is feasible for (2.1), (y^0, u^0) is feasible for (3.18), f is pseudoconvex and all g_i are quasiconvex. Then, $f(x, \overline{y}) \ge f(x, y^0)$.*

Proof. For $u \ge 0$, for a feasible points (y^0, u^0) for (3.18) and $\overline{y}$ for (2.1), we have

$$u^{0\top} g(x^0, \overline{y}) - u^{0\top} g(x, y^0) \le 0.$$

Quasiconvexity then implies

$$u^{0\top} \nabla_y g(x, y^0)^\top (\overline{y} - y^0) \le 0.$$

Feasibility for (3.18) then leads to

$$\nabla_y f(x, y^0)^\top (\overline{y} - y^0) \ge 0.$$

Using pseudoconvexity we get $f(x, \overline{y}) \ge f(x, y^0)$. □

Weak duality in Theorem 3.19 implies the following: If, under the assumptions of Theorem 3.19 we have $f(x, \overline{y}) \le f(x, y^0)$ then, $\overline{y} \in \Psi(x)$ and (x, y^0, u^0) is a global optimal solution of (3.18).

Theorem 3.20. *[Egudo and Mond [9]] Let $\overline{y}$ be an optimal solution of (2.1) for fixed x and let Slater's condition be satisfied for the optimization problem (2.1) satisfying the assumptions of Theorem 3.19. Then, there exists $\overline{u}$ such that $(\overline{y}, \overline{u})$ is an optimal solution of (3.18) and both objective function values coincide.*

Proof. Since Slater's condition is satisfied and $\overline{y}$ is an optimal solution, the Karush-Kuhn-Tucker conditions are satisfied guaranteeing existence of $\overline{u}$. Then, $(\overline{y}, \overline{u})$ is feasible for (3.18). Weak duality proves the theorem. □

This makes it possible to replace the bilevel optimization problem by

$$
\begin{aligned}
F(x,y) &\to \min_{x,y,z,u}\\
G(x) &\le 0\\
g(x,y) &\le 0\\
u^\top g(x,z) &\ge 0 \qquad (3.19)\\
u &\ge 0\\
\nabla_y f(x,z) + u^\top \nabla_y g(x,z) &= 0\\
f(x,y) &\le f(x,z).
\end{aligned}
$$

Theorem 3.21. *Let $(\overline{x},\overline{y})$ be a global optimal solution of (2.4) with a lower level problem satisfying Slater's condition and the assumptions of Theorem 3.19. Then, there exists $(\overline{z},\overline{u})$ such that $(\overline{x},\overline{y},\overline{z},\overline{u})$ is a global optimal solution of (3.19) and both optimal objective function values are the same.*

Vice versa, if $(\overline{x},\overline{y},\overline{z},\overline{u})$ is a global optimal solution of (3.19) then, $(\overline{x},\overline{y})$ is a global optimal solution of (2.4) with the same objective function value.

Proof.
(1) Let $(\overline{x},\overline{y})$ be a global optimal solution of (2.4). By Theorem 3.20 there exists $(\overline{z},\overline{u})$ such that $(\overline{x},\overline{y},\overline{z},\overline{u})$ is feasible for (3.19).

If $(\overline{x},\overline{y},\overline{z},\overline{u})$ would not be optimal for (3.19) then, there exists a feasible point (x^0,y^0,z^0,u^0) with $F(x^0,y^0) < F(\overline{x},\overline{y})$. By Theorem 3.20, $f(x^0,z^0) = \varphi(x^0)$ and $y^0 \in \Psi(x^0)$ by feasibility, i.e. the point (x^0,y^0) is feasible for (2.4) violating the assumption.

(2) In the opposite direction, let $(\overline{x},\overline{y},\overline{z},\overline{u})$ be a global optimal solution of (3.19). Then, by Theorem 3.20, the optimal objective function values of (3.18) and (2.1) coincide. Hence, since $f(\overline{x},\overline{y}) = f(\overline{x},\overline{z})$ and $g(\overline{x},\overline{y}) \le 0$ the point $(\overline{x},\overline{y})$ is feasible for (2.4).

If $(\overline{x},\overline{y})$ is assumed to be not globally optimal for (2.4) then there exists a feasible point (x^0,y^0) with $F(x^0,y^0) < F(\overline{x},\overline{y})$. By Theorem (3.20) there is (z^1,u^1) such that (x^0,y^0,z^1,u^1) is feasible for (3.19) with the same objective function value. This is a contradiction to the assumed global optimality of $(\overline{x},\overline{y},\overline{z},\overline{u})$.

□

Theorem 3.22. *Let $(\overline{x},\overline{y})$ be a local optimal solution of (2.4) and Slater's condition as well as the assumptions of Theorem 3.19 be satisfied for the lover level problem (2.1). Then, all feasible points $(\overline{x},\overline{y},\overline{z},\overline{u})$ of (3.19) are locally optimal solutions of (3.19).*

Proof. Arguing by contradiction, let $(\overline{x},\overline{y},\overline{z},\overline{u})$ be not locally optimal, i.e. let there exist a sequence (x^k,y^k,z^k,u^k) of feasible points for (3.19) converging to $(\overline{x},\overline{y},\overline{z},\overline{u})$ with

$$F(x^k,y^k) < F(\overline{x},\overline{y}) \ \forall\ k.$$

Then, due to weak Mond-Weir duality in Theorem 3.19, $y^k \in \Psi(x^k)$ for all k violating local optimality of $(\overline{x},\overline{y})$ for (2.4). □

The opposite implication is not correct in general.

Example 3.23. *Consider again Example 3.2. Then, problem* (3.19) *reads as follows:*

$$
\begin{aligned}
(x-1)^2 + (y-1)^2 &\to \min \\
x + y &\le 1 \\
-x + y &\le 1 \\
u_1,\ u_2 &\ge 0 \\
u_1(x+z-1) + u_2(-x+z-1) &\ge 0 \\
u_1 + u_2 &= 1 \\
-y &\le -z.
\end{aligned} \tag{3.20}
$$

Consider the point $(x, y, z, u) = (0, 1, 1, 0, 1)$. *This point is feasible for* (3.20). *The last inequality implies*

$$z \le y \le 1 \text{ by the constraints of the lower level problem.}$$

Hence,

$$u_1(x+y-1)+u_2(-x+y-1) = x(u_1-u_2)+(u_1+u_2)y-u_1-u_2 \ge x(u_1-u_2)+z-1 \ge 0$$

or

$$u_1(x+y-1) = 0,\ u_2(-x+y-1) = 0.$$

If $u_1 > 0,\ u_2 > 0$ *then,* $y = 1,\ x = 0$. *Otherwise, since* $z - 1 \le 0$ *and* $u_1 - u_2 \le 0$ *near* $(x, y, z, u) = (0, 1, 1, 0, 1)$ *we have* $x \le 0$.

Hence, the point $(x, y, z, u) = (0, 1, 1, 0, 1)$ *is a local minimum.*

4. Conclusion

Bilevel optimization problems have a constraint set which is a subset of the graph of the solution set mapping of a second, parameter dependent optimization problem. To investigate them, this solution set mapping needs to be substituted. For doing that, various approaches can be found in literature as the use of the (necessary and sufficient) Karush-Kuhn-Tucker conditions for the lower level problem or certain duality relations. In the article, the Lagrange, Wolfe and Mond-Weir dual problems are suggested for replacing the lower level problem. Assumptions have been proposed guaranteeing that the bilevel optimization problem proves to be equivalent to the transformed ones. The latter problems are all nonconvex ones implying the need to investigate local optimal solutions. Local optimal solutions of the bilevel optimization are related to local optimal solutions of its replacements. The opposite relation is in general not correct which can be shown using the same example as for the respective investigation with respect to the KKT transformation. The transformed problem using the Lagrange dual problem for the lower level problem is closely related with the optimal objective function value transformation. Moreover, it reduces to the one using the Wolfe dual problem. For the KKT transformed problem an algorithm for computing a local optimal solution of the bilevel optimization problem can be found in Dempe and Franke [7]. The respective investigations remain open for the other approaches.

References

1. D. Aussel, G. Bouza, S. Dempe, and S. Lepaul, *Genericity analysis of multi-leader-disjoint-followers game*, SIAM Journal on Optimization **31** (2021), no. 3, 2055–2079.
2. B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, *Non-linear parametric optimization.*, Basel - Boston - Stuttgart: Birkhäuser Verlag, 1983.
3. A Ben-Israel and B Mond, *What is invexity?*, J. Austral. Math. Soc. Ser. B **28** (1986), 1–9.
4. J. Bracken and J. McGill, *Mathematical programs with optimization problems in the constraints*, Operations Research **21** (1973), 37–44.
5. B.D. Craven and B. Mond, *On converse duality in nonlinear programming*, Operations research **19** (1971), no. 4, 1075–1078.
6. S. Dempe and J. Dutta, *Is bilevel programming a special case of a mathematical program with complementarity constraints?*, Mathematical Programming **131** (2012), 37–48.
7. S. Dempe and S. Franke, *Solution of bilevel optimization problems using the KKT approach*, Optimization **68** (2019), 1471 – 1489.
8. Stephan Dempe, *Bilevel optimization: theory, algorithms, applications and a bibliography*, Bilevel optimization (S. Dempe and A. Zemkoho, eds.), Springer, 2020, pp. 581–672.
9. RR Egudo and B Mond, *Duality with generalized convexity*, The ANZIAM Journal **28** (1986), no. 1, 10–21.
10. J. Gauvin, *A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming*, Mathematical Programming **12** (1977), 136–139.
11. Jürgen Guddat, F Guerra Vazquez, and Hubertus Th Jongen, *Parametric optimization: singularities, pathfollowing and jumps*, Springer, 1990.
12. P. Kanniappan, *Duality theorems for convex programming without constraint qualification*, Journal of the Australian Mathematical Society **36** (1984), no. 2, 253–266.
13. J. Kornaj and T. Liptak, *Two-level planning*, Econometrica **33** (1965), 141–169.
14. Y. Li, G.-H. Lin, J. Zhang, and X. Zhu, *A novel approach for bilevel programs based on wolfe duality*, (2021).
15. Z.-Q. Luo, J.-S. Pang, and D. Ralph, *Mathematical programs with equilibrium constraints*, Cambridge University Press, Cambridge, 1996.
16. Olvi L Mangasarian, *Nonlinear programming*, (1969).
17. J.E. Martínez-Legaz, *Optimality conditions for pseudoconvex minimization over convex sets defined by tangentially convex constraints*, Optimization Letters **9** (2015), no. 5, 1017–1023.
18. J.A. Mirrlees, *The theory of moral hazard and unobservable bevaviour: part I*, Review of Economic Studies **66** (1999), 3–21.
19. B. Mond and T. Weir, *Generalized concavity and duality*, Generalized Concavity in Optimization and Economics (S. Schaible and W.T. Ziemba, eds.), Academic Press, New York, 1981, p. 263–279.
20. J. Outrata, M. Kočvara, and J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints*, Kluwer Academic Publishers, Dordrecht, 1998.
21. J.V. Outrata, *On the numerical solution of a class of Stackelberg problems*, ZOR - Mathematical Methods of Operations Research **34** (1990), 255–277.
22. M. Pilecka, *Combined reformulation of bilevel programming problems*, Master's thesis, TU Bergakademie Freiberg, Fakultät für Mathematrik und Informatik, 2011.
23. S.M. Robinson, *Generalized equations and their solutions, part II: Applications to nonlinear programming*, Mathematical Programming Study **19** (1982), 200–221.
24. H. Scheel and S. Scholtes, *Mathematical programs with equilibrium constraints: stationarity, optimality, and sensitivity*, Mathematics of Operations Research **25** (2000), 1–22.
25. V. Shikhman, *Topological aspects of nonsmooth optimization*, vol. 64, Springer Science & Business Media, 2011.
26. H.v. Stackelberg, *Marktform und Gleichgewicht*, Springer-Verlag, Berlin, 1934, engl. transl.: The Theory of the Market Economy, Oxford University Press, 1952.
27. D.J. White and G. Anandalingam, *A penalty function approach for solving bi-level linear programs*, Journal of Global Optimization **3** (1993), 397–419.
28. P. Wolfe, *A duality theorem for non-linear programming*, Quarterly of applied mathematics **19** (1961), no. 3, 239–244.
29. J.J. Ye and D.L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization **33** (1995), 9–27.

TU Bergakademie Freiberg, Germany
Email address: dempe@tu-freiberg.de