Optimality criteria for bilevel programming problems using the radial subdifferential

D. Fanghänel*

October 28, 2005

Abstract

The discrete bilevel programming problems considered in this paper have discrete parametric lower level problems with linear constraints and a strongly convex objective function. Using both the optimistic and the pessimistic approach this problem is reduced to the minimization of auxiliary nondifferentiable and generally discontinuous functions. To develop necessary and sufficient optimality conditions for the bilevel problem the radial-directional derivative and the radial subdifferential of these auxiliary functions are used.

Key words: Bilevel programming, necessary and sufficient optimality conditions, discrete parametric optimization, minimization of discontinuous functions, radial-directional derivative.

1 Introduction

Bilevel programming problems are hierarchical optimization problems where the constraints of one problem (the so-called upper level problem) are defined in part by a second parametric optimization problem (the lower level problem) [1, 2]. These problems occur in a large variety of practical situations [3]. Many approaches are known to attack continuous bilevel programming problems. But, the number of references for bilevel programming problems with discrete variables is rather limited. Focus in the paper [15] is on existence of optimal solutions for problems which have discrete variables in the upper resp. the lower level problems. Solution algorithms have been developed in [5, 8, 9, 16]. The position of constraints in the upper resp. in the lower level problems is critical. The implications of and gains obtained from shifting a 0-1 variable from the lower to the upper level problems have been investigated in [4].

*Technical University Bergakademie Freiberg, Freiberg, Germany

Focus in this paper is on optimality conditions for bilevel programming problems with discrete variables in the lower level problem. Verification of optimality conditions for continuous linear problems is $\mathcal{NP}$–hard [14] even if the optimal solution of the lower level problem is unique for all upper level variable values.

If the lower level problems may have nonunique optimal solutions, useful concepts are the optimistic and the pessimistic approaches. Both concepts lead to the minimization of a discontinuous auxiliary function φ. In the case of a linear bilevel programming problem, this function is a generalized PC^1–function and the formulation of optimality conditions can be based on the radial-directional derivative [2, 6].

In this paper a similar approach is investigated for discrete bilevel programming problems.

The outline of the paper is as follows. In Section 2 the investigated bilevel program is formulated and some introductory examples are given. Structural properties of the solution set mapping of the lower level problem are investigated in Section 3. In Sections 4 and 5 focus is on properties of the auxiliary function φ. Optimality conditions using the radial-directional derivative of the function φ are developed in Section 6, and in Section 7 the same is done by the help of the radial subdifferential of the function φ.

Throughout this paper the gradient of a function is the row vector of the partial derivatives. Further we will use the abreviation $\{z^k\}$ for a sequence $\{z^k\}_{k=1}^{\infty}$ if this will not cause any confusion.

2 A bilevel problem with discrete lower level

In this paper we consider the following bilevel programming problem

$$\begin{cases} \min\{g(x,y) : y \in Y,\ x \in \Psi_D(y)\} \\ \\ \Psi_D(y) = \operatorname{Argmin}\{f(x,y) :\ x \in S_D\} \end{cases} \tag{1}$$

with the following requirements:

1. $Y \subseteq \mathbb{R}^n$ is convex, closed and $\operatorname{int} Y \neq \emptyset$.
2. $f(x,y) = F(x) - y^\top x$ with $F : \mathbb{R}^n \to \mathbb{R}$ being differentiable and strongly convex [10] with modulus $\theta > 0$, i.e. for all $x, x^0 \in \mathbb{R}^n$ it holds

$$F(x) \geq F(x^0) + \nabla F(x^0)(x - x^0) + \theta\|x - x^0\|^2.$$

3. $g(x,y)$ is continuously differentiable with respect to y.

4. The set $S_D \subseteq \mathbb{R}^n$ is required to be nonempty and discrete, i.e. there exists some $\omega > 0$ with $\|x - x'\| \geq \omega$ for all $x, x' \in S_D$, $x \neq x'$.

   S_D denotes the set of all feasible solutions of the lower level problem.

Thus, the problem under consideration is continuous in the upper level and discrete with some special structure in the lower level.

In general the solution of the lower level is not unique. This causes some uncertainty in the definition of the upper level objective function [2]. Thus, instead of $g(x, y)$, we will investigate the following functions

$$\phi_o(y) = \min_{x \in \Psi_D(y)} g(x, y), \tag{2}$$

$$\phi_p(y) = \max_{x \in \Psi_D(y)} g(x, y). \tag{3}$$

The function $\phi_o(y)$ is called optimististic solution function and $\phi_p(y)$ pessimistic solution function. While most of the papers on bilevel programming with possible nonunique lower level solutions investigate (implicitly) the optimistic approach (see e.g. [1] and the references therein), focus for instance in the paper [11] is on the pessimistic approach and both approaches have been compared in [12]. A local optimal solution of the optimististic/pessimistic solution function is a local optimistic/pessimistic solution of (1).

In this paper we investigate necessary and sufficient conditions under which some point $y^0 \in Y$ is a local optimistic/pessimistic solution of (1).

We will use the notation $\phi(y)$ if the statement holds for both $\phi_o(y)$ and $\phi_p(y)$.

For our considerations the so-called regions of stability are very important. They are defined as follows.

Definition 1. *Let $x^0 \in S_D$. Then the set*

$$\begin{aligned} R(x^0) &= \{y \in \mathbb{R}^n : f(x^0, y) \leq f(x, y) \text{ for all } x \in S_D\} \\ &= \{y \in \mathbb{R}^n : x^0 \in \Phi_D(y)\} \end{aligned}$$

is called region of stability for the point x^0.

Thus the set $R(x^0)$ denotes the set of all parameters for which the point x^0 is optimal.

To make the subject more clear consider the following example.

Example 1.

$$\min\{\sin(xy) : y \in [0, 5], x \in \Psi_D(y)\}$$

$$\Psi_D(y) = \operatorname*{Argmin}_x \left\{ \frac{1}{2}x^2 - xy : 0 \leq x \leq 5, x \in \mathbb{Z} \right\}$$

Since the upper level objective function is continuous on the regions of stability the latter ones can be seen in figure 1. Formally the regions of stability are

$$R(0) = (-\infty, 0.5],\ R(1) = [0.5, 1.5],\ R(2) = [1.5, 2.5],\ R(3) = [2.5, 3.5],$$
$$R(4) = [3.5, 4.5] \text{ and } R(5) = [4.5, \infty).$$

Using the definitions of the optimistic and pessimistic solution functions at the intersection points of the regions of stability, we get

$$\phi_o(y) = \begin{cases} 0 & y \le 0.5 \\ \sin(y) & 0.5 < y < 1.5 \\ \sin(2y) & 1.5 \le y \le 2.5 \\ \sin(3y) & 2.5 < y \le 3.5 \\ \sin(4y) & 3.5 < y \le 4.5 \\ \sin(5y) & y > 4.5 \end{cases} \qquad \phi_p(y) = \begin{cases} 0 & y < 0.5 \\ \sin(y) & 0.5 \le y \le 1.5 \\ \sin(2y) & 1.5 < y < 2.5 \\ \sin(3y) & 2.5 \le y < 3.5 \\ \sin(4y) & 3.5 \le y < 4.5 \\ \sin(5y) & y \ge 4.5 \end{cases} .$$

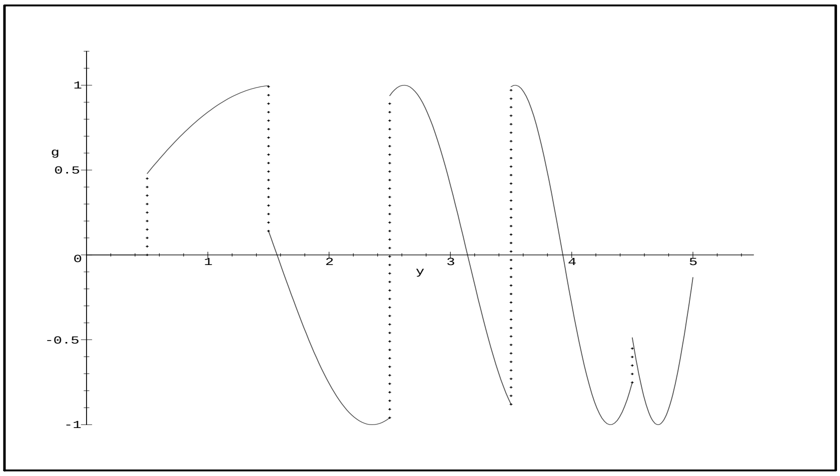

Figure 1: solution function ϕ for example 1

As it can be seen in figure 1 the local optimal solutions of ϕ_o are

$$y \in [0, 0.5],\ y = \frac{3\pi}{4},\ y = 3.5,\ y = \frac{11\pi}{8},\ y = \frac{3\pi}{2}$$

and

$$y \in [0, 0.5),\ y = \frac{3\pi}{4},\ y = \frac{11\pi}{8},\ y = \frac{3\pi}{2}$$

are the local optimal solutions of ϕ_p.

In Example 1 the optimistic and the pessimistic solution function are not continuous but rather selections of finitely many continuously differentiable functions.

3 Some remarks on the sets $\Psi_D(y)$ and $R(x)$

In this section we want to derive some properties of the sets $\Psi_D(y)$ and $R(x)$ which we will need later.

Lemma 1. *For each $x^0 \in S_D$ the set $R(x^0)$ is a closed convex set with $\nabla F(x^0)^\top$ in its interior.*

Proof. Let $x^0 \in S_D$. Then for all $y \in R(x^0)$ it holds $f(x^0, y) \leq f(x, y)$ for all $x \in S_D$ and therefore

$$(x - x^0)^\top y \leq F(x) - F(x^0) \quad \forall x \in S_D.$$

Thus, $R(x^0)$ corresponds to the intersection of (maybe infinitely many) half-spaces. This implies that $R(x^0)$ is convex and closed.

Now we want to show that $\nabla F(x^0)^\top \in \operatorname{int} R(x^0)$. Since $F : \mathbb{R}^n \to \mathbb{R}$ is strongly convex there exists some $\theta > 0$ with $F(x) \geq F(x^0) + \nabla F(x^0)(x - x^0) + \theta \|x - x^0\|^2$ for all $x \in \mathbb{R}^n$.

Consider $y = \nabla F(x^0)^\top + \alpha h$ with $h \in \mathbb{R}^n$, $\|h\| = 1$ and $\alpha \in [0, \theta\omega]$. Then, for all $x \in S_D$, $x \neq x^0$, the following sequence of inequalities is valid by $\|x - x^0\| \geq \omega$ for $x \neq x^0$:

$$\begin{aligned}
F(x) &\geq F(x^0) + \nabla F(x^0)(x - x^0) + \theta \|x - x^0\|^2 \\
&= F(x^0) + y^\top (x - x^0) - \alpha h^\top (x - x^0) + \theta \|x - x^0\|^2 \\
&\geq F(x^0) + y^\top (x - x^0) - \alpha \|x - x^0\| + \theta \|x - x^0\|^2 \\
&\geq F(x^0) + y^\top (x - x^0) + (\theta\omega - \alpha) \|x - x^0\| \\
&\geq F(x^0) + y^\top (x - x^0).
\end{aligned}$$

Thus we obtain $(\nabla F(x^0)^\top + \alpha h) \in R(x^0)$ for all $\alpha \in [0, \theta\omega]$, i.e. the assumption holds. □

Lemma 2. *1. For each $y \in \mathbb{R}^n$ the set $\Psi_D(y)$ has finite cardinality.*

2. If $y^0 \in \operatorname{int} R(x^0)$ for some $x^0 \in S_D$, then $\Psi_D(y^0) = \{x^0\}$.

3. Let some point $y^0 \in \mathbb{R}^n$ be given. Then there exists a positive real number $\epsilon > 0$ such that $\Psi_D(y) \subseteq \Psi_D(y^0)$ for all $y \in U_\epsilon(y^0) = \{y : \|y - y^0\| < \epsilon\}$.

Proof. 1. If $S_D = \emptyset$ the assumption holds obviously. Assume that $S_D \neq \emptyset$ and take a point $x^0 \in S_D$. Let an arbitrary $y \in \mathbb{R}^n$ be given. Then for all $x \in \Psi_D(y)$ it holds

$$F(x) - y^\top x \leq F(x^0) - y^\top x^0$$

implying

$$F(x^0) + \nabla F(x^0)(x - x^0) + \theta \|x - x^0\|^2 \leq F(x^0) + y^\top (x - x^0)$$

for some $\theta > 0$ since F is strongly convex. Thus,

$$\begin{aligned} \theta\|x - x^0\|^2 &\leq (y^\top - \nabla F(x^0))(x - x^0) \leq \|y - \nabla F(x^0)^\top\|\|x - x^0\| \\ \|x - x^0\| &\leq \frac{1}{\theta}\|y - \nabla F(x^0)^\top\|. \end{aligned}$$

Therefore $\Psi_D(y)$ has finite cardinality.

2. The inclusion $y^0 \in \operatorname{int} R(x^0)$ implies $\{x^0\} \subseteq \Psi_D(y^0)$ by definition. To prove the opposite direction assume that there exists a point $x \in \Psi_D(y^0)$, $x \neq x^0$. Then,

$$\begin{aligned} F(x) - {y^0}^\top x &= F(x^0) - {y^0}^\top x^0 \\ F(x) - F(x^0) &= {y^0}^\top (x - x^0) > \nabla F(x^0)(x - x^0) \end{aligned}$$

since F is strongly convex. Due to $y^0 \in \operatorname{int} R(x^0)$ there exists some $\epsilon > 0$ such that
$y := y^0 + \epsilon(y^0 - \nabla F(x^0)^\top) \in R(x^0)$. Now we obtain

$$\begin{aligned} f(x^0, y) &= F(x^0) - y^\top x^0 = F(x) - y^\top x^0 - {y^0}^\top (x - x^0) \\ &= f(x, y) + (y - y^0)^\top (x - x^0) \\ &= f(x, y) + \epsilon({y^0}^\top - \nabla F(x^0))(x - x^0) > f(x, y) \end{aligned}$$

which is a contradiction to $y \in R(x^0)$.

3. Assume that the assertion does not hold. Then there exist sequences $\{y^k\}_{k=1}^\infty$ with $y^k \to y^0, k \to \infty$, and
$\{x^k\}_{k=1}^\infty$ with $x^k \in \Psi_D(y^k)$ but $x^k \notin \Psi_D(y^0)$ for all k.
Thus, for fixed $x^0 \in S_D$, it holds

$$\begin{aligned} F(x^k) - {y^k}^\top x^k &\leq F(x^0) - {y^k}^\top x^0 \\ F(x^0) + \nabla F(x^0)(x^k - x^0) + \theta\|x^k - x^0\|^2 &\leq F(x^0) + {y^k}^\top (x^k - x^0) \\ \|x^k - x^0\| &\leq \frac{\|y^k - \nabla F(x^0)^\top\|}{\theta}. \end{aligned}$$

This yields

$$\|x^k - x^0\| \leq \underbrace{\frac{\|y^k - y^0\|}{\theta}}_{\to 0} + \frac{\|y^0 - \nabla F(x^0)^\top\|}{\theta},$$

i.e. $\{x^k\}$ is bounded and has finitely many elements. Therefore we can assume that all x^k are equal, i.e. $\exists x \in S_D$ with $x \in \Psi_D(y^k)$ $\forall k$ but $x \notin \Psi_D(y^0)$.
That means $y^k \in R(x)$ $\forall k$ but $y^0 \notin R(x)$. This is a contradiction to Lemma 1.

□

4 Basic properties of $\phi(y)$

In this section we want to show first that for each $y^0 \in \mathbb{R}^n$ there exists some $\epsilon > 0$ such that in the neighborhood $U_\epsilon(y^0)$ the optimistic/pessimistic solution function is a selection of finitely many continuously differentiable functions. Further, for this special $\epsilon > 0$ we will investigate the support set

$$Y_x(y^0) := \{y \in U_\epsilon(y^0) \cap R(x) : \ g(x,y) = \phi(y)\}$$

and its contingent cone

$$T_x(y^0) := \{r : \exists \{y^s\} \subseteq Y_x(y^0) \ \exists \{t_s\} \subseteq \mathbb{R}_+ : \ y^s \to y^0, t_s \downarrow 0, \lim_{s\to\infty} \frac{y^s - y^0}{t_s} = r\}.$$

That means, $Y_x(y^0)$ is the set of all $y \in U_\epsilon(y^0)$ for which both $x \in \Psi_D(y)$ and $g(x,y) = \phi(y)$ hold for a fixed point $x \in S_D$. Properties of these sets are essential for the investigation of generalized PC^1–functions (in short: GPC^1–functions) in the paper [6] leading to optimality conditions for linear bilevel programming problems in [2]. The following two theorems show that the objective functions in the two auxiliary problems (2) and (3) have many properties of GPC^1–functions, but they are not GPC^1–functions as it is shown by Example 2 below.

Theorem 3. *For the function ϕ and each $y^0 \in \mathbb{R}^n$ it holds:*

1. *There exists an open neighborhood $U_\epsilon(y^0)$ of y^0 and a finite number of points $x \in \Psi_D(y^0)$ with*

$$\phi(y) \in \{g(x,y)\}_{x\in\Psi_D(y^0)} \quad \forall y \in U_\epsilon(y^0).$$

2. $\operatorname{int} Y_x(y^0) = U_\epsilon(y^0) \cap \operatorname{int} R(x)$ *and* $Y_x(y^0) \subseteq \operatorname{cl}\ \operatorname{int} Y_x(y^0)$ *for* $x, y^0 \in \mathbb{R}^n$.
3. $T_x(y^0) \subseteq \operatorname{cl}\ \operatorname{int} T_x(y^0)$ *for* $y^0 \in R(x)$.

Proof. Let an arbitrary $y^0 \in \mathbb{R}^n$ be given.

1.) Because of Lemma 2, $\Psi_D(y^0)$ has finite cardinality and there exists some $\epsilon > 0$ with $\Psi_D(y^0) \supseteq \Psi_D(y)$ for all $y \in U_\epsilon(y^0)$. With $\phi(y) \in \{g(x,y)\}_{x\in\Psi_D(y)}$ it follows $\phi(y) \in \{g(x,y)\}_{x\in\Psi_D(y^0)} \quad \forall y \in U_\epsilon(y^0)$.

2.) Let $\bar{y} \in \operatorname{int} Y_x(y^0)$. Then there exists some $\delta > 0$ with $U_\delta(\bar{y}) \subseteq Y_x(y^0)$. Thus, $\bar{y} \in U_\epsilon(y^0)$ and $U_\delta(\bar{y}) \subseteq R(x)$, i.e. $\bar{y} \in U_\epsilon(y^0) \cap \operatorname{int} R(x)$.

Let $\bar{y} \in U_\epsilon(y^0) \cap \operatorname{int} R(x)$. Then there exists some $\delta > 0$ with $U_\delta(\bar{y}) \subseteq U_\epsilon(y^0)$ and $U_\delta(\bar{y}) \subseteq \operatorname{int} R(x)$. From Lemma 2 it follows $\Psi_D(y) = \{x\}$ $\forall y \in U_\delta(\bar{y})$. Thus, $\phi(y) = g(x,y)$ $\forall y \in U_\delta(\bar{y})$, i.e. $y \in Y_x(y^0)$ $\forall y \in U_\delta(\bar{y})$. Therefore, $\bar{y} \in \operatorname{int} Y_x(y^0)$. This implies the first equation of part 2.

Now let $\bar{y} \in Y_x(y^0)$. This means $\bar{y} \in R(x)$, $\bar{y} \in U_\epsilon(y^0)$ and $\phi(\bar{y}) = g(x,\bar{y})$. Since $R(x)$ is convex with nonempty interior (cf. Lemma 1) there exists some sequence

$\{y^k\} \in \operatorname{int} R(x)$ with $y^k \to \bar{y}$, $k \to \infty$. W.l.o.g. we can further assume that $y^k \in U_\epsilon(y^0)\ \forall k$. Consequently, $y^k \in \operatorname{int} Y_x(y^0)\ \forall k$ and thus $\bar{y} \in \operatorname{cl}\operatorname{int} Y_x(y^0)$.

3.) Let an arbitrary $r \in T_x(y^0)$ be given. Then there exist sequences $\{y^s\} \subseteq Y_x(y^0)$ and $\{t_s\} \subseteq \mathbb{R}_+$ with $y^s \to y^0$, $t_s \downarrow 0$ and $\lim_{s\to\infty} \frac{y^s - y^0}{t_s} = r$. We can assume w.l.o.g. that $t_s \in (0,1)\ \forall s$.

Take any $\tilde{y} \in \operatorname{int} Y_x(y^0)$ and let $\hat{y}^s := t_s\tilde{y} + (1-t_s)y^0 = y^0 + t_s(\tilde{y} - y^0)$. Then, $\lim_{s\to\infty} \hat{y}^s = y^0$ and $\frac{\hat{y}^s - y^0}{t_s} = \tilde{y} - y^0 =: \tilde{r}\ \forall s$. Since $R(x)$ is convex it follows easily that $\hat{y}^s \in \operatorname{int} Y_x(y^0)\ \forall s$ and $\tilde{r} \in \operatorname{int} T_x(y^0)$.

Now consider $z^s_\lambda := \lambda y^s + (1-\lambda)\hat{y}^s$ with $\lambda \in (0,1)$. Since $R(x)$ is convex and $\hat{y}^s \in \operatorname{int} Y_x(y^0)$ it follows $z^s_\lambda \in \operatorname{int} Y_x(y^0)\ \forall \lambda\ \forall s$. Then it holds $z^s_\lambda \to y^0$ for $s \to \infty$ and $\lim_{s\to\infty} \frac{z^s_\lambda - y^0}{t_s} = \lambda r + (1-\lambda)\tilde{r} =: r_\lambda \in T_x(y^0)$ for all $\lambda \in (0,1)$. Moreover, $r_\lambda \to r$ for $\lambda \to 1$.

Now, from $z^s_\lambda \in \operatorname{int} Y_x(y^0)$ it follows easily that $z^s_\lambda - y^0 \in \operatorname{int} T_x(y^0)$ and thus $\frac{z^s_\lambda - y^0}{t_s} \in \operatorname{int} T_x(y^0)\ \forall s\ \forall \lambda \in (0,1)$.
Hence, $r_\lambda \in \operatorname{cl}\operatorname{int} T_x(y^0)\ \forall \lambda \in (0,1)$. This together with $r_\lambda \to r$ for $\lambda \to 1$ implies $r \in \operatorname{cl}\operatorname{cl}\operatorname{int} T_x(y^0) = \operatorname{cl}\operatorname{int} T_x(y^0)$. □

Theorem 4. *$\operatorname{int} T_{x^1}(y^0) \cap \operatorname{int} T_{x^2}(y^0) = \emptyset$ for all $x^1, x^2 \in \Psi_D(y^0)$, $x^1 \neq x^2$.*

Proof. Let $r \in T_{x^1}(y^0) \cap T_{x^2}(y^0)$ be arbitrary. Due to $r \in T_{x^1}(y^0)$ there exists sequences $\{y^s\} \subseteq Y_{x^1}(y^0)$, $y^s \to y^0$ and $\{t_s\}$, $t_s \downarrow 0$ with $r^s := \frac{y^s - y^0}{t_s} \to r$.

From $y^s \in Y_{x^1}(y^0)\ \forall s$ it follows $y^s \in R(x^1)\ \forall s$, i.e. $F(x^1) - {y^s}^\top x^1 \leq F(x^2) - {y^s}^\top x^2$. Since $x^1, x^2 \in \Psi_D(y^0)$ it holds $F(x^1) - {y^0}^\top x^1 = F(x^2) - {y^0}^\top x^2$. Hence,

$$
\begin{aligned}
{y^s}^\top(x^1 - x^2) \geq F(x^1) - F(x^2) &= {y^0}^\top(x^1 - x^2)\\
(y^s - y^0)^\top(x^1 - x^2) &\geq 0 \quad \forall s\\
{r^s}^\top(x^1 - x^2) &\geq 0 \quad \forall s.\\
\text{With } r^s \to r \text{ this yields } r^\top(x^1 - x^2) &\geq 0.
\end{aligned}
$$

From $r \in T_{x^2}(y^0)$ it follows analogously $(x^1 - x^2)^\top r \leq 0$. Therefore it holds

$$(x^2 - x^1)^\top r = 0 \text{ for all } r \in T_{x^1}(y^0) \cap T_{x^2}(y^0).$$

Assume that there exists some $r \in \operatorname{int} T_{x^1}(y^0) \cap \operatorname{int} T_{x^2}(y^0)$. Then for all $t \in \mathbb{R}^n$, $||t|| = 1$ there exists a real number $\delta > 0$ with $r + \delta t \in T_{x^1}(y^0) \cap T_{x^2}(y^0)$, i.e.

$$
\begin{aligned}
(x^2 - x^1)^\top(r + \delta t) &= 0\\
\delta(x^2 - x^1)^\top t &= 0\\
(x^2 - x^1)^\top t &= 0 \quad \forall t
\end{aligned}
$$

and therefore $x^1 = x^2$. □

Next we show that the function ϕ is not a GPC^1-function (cf. [2],[6]). For GPC^1-functions one requires additionally to the results in the Theorems 3 and 4 that there exists a number $\delta > 0$ such that for all $r \in T_{x^1}(y^0) \cap T_{x^2}(y^0)$, $\|r\| = 1$, $x^1 \neq x^2$ some $t_0 = t(r) \geq \delta$ can be found with $y^0 + tr \in Y_{x^1}(y^0)$ or $y^0 + tr \in Y_{x^2}(y^0)$ $\forall t \in (0, t_0)$. We will show that the functions ϕ usually do not have this property.

Example 2. *Consider the lower level problem in (1) with the feasible set* $S_D = \{x^1 = (0,0,0)^\top, x^2 = (1,0,0)^\top, x^3 = (0,1,0)^\top\}$ *and* $f(x,y) = \frac{1}{2}x^\top x - x^\top y$. *Then we obtain the following regions of stability:*

$$\begin{aligned} R(x^1) &= \{y \in \mathbb{R}^3 : y_1 \leq 1/2, y_2 \leq 1/2\} \\ R(x^2) &= \{y \in \mathbb{R}^3 : y_1 \geq 1/2, y_2 \leq y_1\} \\ R(x^3) &= \{y \in \mathbb{R}^3 : y_2 \geq 1/2, y_2 \geq y_1\}. \end{aligned}$$

Let $g(x,y) = (1/2, -1, 0)^\top x$ *be the objective function of the upper level problem. Then,*

$$\phi_o(y) = \begin{cases} -1 & y \in R(x^3) \\ 0 & y \in R(x^1) \backslash R(x^3) \\ 1/2 & else \end{cases} .$$

Set $r = (0,0,1)^\top$ *and* $y^0 = (1/2, 1/2, 0)^\top$.
Further, $y^1(\epsilon) := (1/2 - \epsilon^2, 1/2 - \epsilon^2, \epsilon + \epsilon^2)^\top \in Y_{x^1}(y^0)$ $\forall \epsilon > 0$. *Then*

$$\begin{aligned} \lim_{\epsilon \to 0} y^1(\epsilon) &= (1/2, 1/2, 0)^\top = y^0, \\ \lim_{\epsilon \to 0} \frac{y^1(\epsilon) - y^0}{\epsilon} &= \lim_{\epsilon \to 0} (-\epsilon, -\epsilon, 1 + \epsilon)^\top = r, \ \textit{i.e.}\ r \in T_{x^1}(y^0). \end{aligned}$$

Analogously $y^2(\epsilon) := (1/2 + \epsilon^2, 1/2, \epsilon + \epsilon^2)^\top \in Y_{x^2}(y^0)$ $\forall \epsilon > 0$. *Then*

$$\begin{aligned} \lim_{\epsilon \to 0} y^2(\epsilon) &= (1/2, 1/2, 0)^\top = y^0, \\ \lim_{\epsilon \to 0} \frac{y^2(\epsilon) - y^0}{\epsilon} &= \lim_{\epsilon \to 0} (\epsilon, 0, 1 + \epsilon)^\top = r, \ \textit{i.e.}\ r \in T_{x^2}(y^0). \end{aligned}$$

Therefore, $r \in T_{x^1}(y^0) \cap T_{x^2}(y^0)$, $\|r\| = 1$, $x^1 \neq x^2$ *but* $\phi_o(y^0 + tr) = -1 < g(x^i, y^0 + tr)$, $i = 1, 2$, $\forall t > 0$, *i.e.* $y^0 + tr \notin Y_{x^1}(y^0)$ *and* $y^0 + tr \notin Y_{x^2}(y^0)$ *for all* $t > 0$.

Until now the description of the contingent cones has been more theoretical. Thus, for calculation we will need some better formula. In [10] many statements are given concerning contingent cones to closed convex sets. But, in general the sets $Y_x(y^0)$ are neither convex nor closed. Using $T_x(y^0) \subseteq \mathrm{cl}\{r \in \mathbb{R}^n : \exists t_0 > 0 \text{ with } y^0 + tr \in Y_x(y^0)\ \forall t \in [0, t_0]\}$ we obtain the following Lemma:

Lemma 5. *Let* $\bar{x} \in \Psi_D(y^0)$. *Then it holds*

$$T_{\bar{x}}(y^0) = \{r \in \mathbb{R}^n : \ 0 \leq (\bar{x} - x)^\top r \quad \forall x \in \Psi_D(y^0)\}.$$

Proof. Let $r \in T_{\bar{x}}(y^0)$. Then there exists some sequence $\{r^k\}$ with $\lim_{k\to\infty} r^k = r$ and $y^0 + tr^k \in R(\bar{x})$ for all k if $t > 0$ is sufficiently small. Hence,

$$\begin{aligned} F(\bar{x}) - (y^0 + tr^k)^\top \bar{x} &\leq F(x) - (y^0 + tr^k)^\top x \quad \forall x \in S_D \\ F(\bar{x}) - {y^0}^\top \bar{x} - t{r^k}^\top \bar{x} &\leq F(x) - {y^0}^\top x - t{r^k}^\top x \quad \forall x \in S_D . \end{aligned}$$

On the other hand it holds $F(\bar{x}) - {y^0}^\top \bar{x} = F(x) - {y^0}^\top x \; \forall x \in \Psi_D(y^0)$. Thus, ${r^k}^\top(\bar{x} - x) \geq 0 \; \forall k \, \forall x \in \Psi_D(y^0)$. Consequently it holds $r^\top(\bar{x} - x) \geq 0 \; \forall x \in \Psi_D(y^0)$.

Let $0 \leq (\bar{x} - x)^\top r \; \forall x \in \Psi_D(y^0)$. Then it holds

$$F(\bar{x}) - (y^0 + tr)^\top \bar{x} \leq F(x) - (y^0 + tr)^\top x \quad \forall x \in \Psi_D(y^0) \; \forall t \geq 0.$$

Further there exists some $\epsilon > 0$ with $\Psi_D(y) \subseteq \Psi_D(y^0) \; \forall y \in U_\epsilon(y^0)$. Thus, for all $t \in (0, \epsilon/\|r\|)$ it holds $F(\bar{x}) - (y^0 + tr)^\top \bar{x} \leq F(x) - (y^0 + tr)^\top x \quad \forall x \in \Psi_D(y^0 + tr)$, i.e. $y^0 + tr \in R(\bar{x}) \; \forall t \in (0, \epsilon/\|r\|)$. Now we will show that $r \in T_{\bar{x}}(y^0)$. Let $\tilde{y} = y^0 + t_0 r$ for some fixed $t_0 \in (0, \epsilon/\|r\|)$. Since $R(\bar{x})$ is convex with nonempty interior there exists some sequence $\{y^s\} \in \operatorname{int} R(\bar{x})$ with $y^s \to \tilde{y}$, $s \to \infty$ and $y^s \in U_\epsilon(y^0)$. Then it holds $(y^s - y^0)\lambda + y^0 \in \operatorname{int} Y_{\bar{x}}(y^0) \; \forall \lambda \in (0, 1) \; \forall s$. Consequently, $y^s - y^0 \in T_{\bar{x}}(y^0) \; \forall s$. Since $T_{\bar{x}}(y^0)$ is a closed cone and $\tilde{y} - y^0 = \lim_{s\to\infty} y^s - y^0$ it follows $\tilde{y} - y^0 = t_0 r \in T_{\bar{x}}(y^0)$, i.e. it holds $r \in T_{\bar{x}}(y^0)$. □

Consequently, the cones $T_x(y^0)$ are polyhedral cones with nonempty interior for all $x \in \Psi_D(y^0)$.

5 The radial-directional derivative

In the following we formulate criteria for local optimality. For this we want to use the radial-directional derivative which was introduced by Recht [13]. Such kind of considerations have even been done for GPC^1-functions [2, 6]. But as shown, although our functions ϕ have some properties in common with GPC^1-functions they are in general not GPC^1-functions.

Definition 2. *Let $U \subseteq \mathbb{R}^n$ be an open set, $y^0 \in U$ and $\phi : U \to \mathbb{R}$. We say that ϕ is radial-continuous at y^0 in direction $r \in \mathbb{R}^n$, $\|r\| = 1$, if there exists a real number $\phi(y^0; r)$ such that*

$$\lim_{t \downarrow 0} \phi(y^0 + tr) = \phi(y^0; r).$$

If the radial limit $\phi(y^0; r)$ exists for all $r \in \mathbb{R}^n$, $\|r\| = 1$, ϕ is called radial-continuous at y^0.

ϕ is radial-directionally differentiable at y^0, if there exists a positively homogeneous function $d\phi_{y^0} : \mathbb{R}^n \to \mathbb{R}$ such that for all $r \in \mathbb{R}^n$, $\|r\| = 1$ and all $t > 0$ it holds

$$\phi(y^0 + tr) - \phi(y^0; r) = t d\phi_{y^0}(r) + o(y^0, tr)$$

with $\lim_{t\downarrow 0} o(y^0, tr)/t = 0$. *Obviously,* $d\phi_{y^0}$ *is uniquely defined and is called the radial-directional derivative of* ϕ *at* y^0.

Theorem 6. *Both the optimistic solution function* ϕ_o *and the pessimistic solution function* ϕ_p *are radial-continuous and radial-directionally differentiable.*

Proof. Consider y^0 and some direction $r \in \mathbb{R}^n$, $\|r\| = 1$. Further let

$$\begin{aligned} I_r(y^0) &:= \{x \in \Psi_D(y^0) : \forall \epsilon > 0\ \exists t \in (0,\epsilon) \text{ with } y^0 + tr \in Y_x(y^0)\} \\ \text{and } G(y^0 + tr) &:= \min_{x \in I_r(y^0)} g(x, y^0 + tr). \end{aligned}$$

Since $\Psi_D(y^0)$ has finite cardinality and the sets $R(x)$ are convex it holds $\phi_o(y^0 + tr) = G(y^0 + tr)$ for all sufficiently small real numbers $t > 0$. Since the function $G(\cdot)$ is the minimum function of finitely many continuously differentiable functions it is continuous and quasidifferentiable (cf. [7]) and thus directionally differentiable in $t = 0$. Therefore the limits

$$\lim_{t\downarrow 0} G(y^0 + tr) = G(y^0) \text{ and } \lim_{t\downarrow 0} \frac{G(y^0 + tr) - G(y^0)}{t} = G'(y^0; r)$$

exist. Moreover, since for all $x \in I_r(y^0)$ it exists some sequence $\{t_k\} \downarrow 0$: $y^0 + t_k r \in Y_x(y^0)$ and

$$\lim_{t\downarrow 0} G(y^0 + tr) = \lim_{k\to\infty} G(y^0 + t_k r) = \lim_{k\to\infty} g(x, y^0 + t_k r) = g(x, y^0)$$

we derive

$$\phi_o(y^0; r) = \lim_{t\downarrow 0} G(y^0 + tr) = G(y^0) = g(x, y^0)\ \forall x \in I_r(y^0). \tag{4}$$

Concerning the radial-directional derivative we obtain

$$\begin{aligned} d\phi_{o y^0}(r) &= \lim_{t\downarrow 0} \frac{\phi_o(y^0 + tr) - \phi_o(y^0; r)}{t} = \lim_{t\downarrow 0} \frac{G(y^0 + tr) - G(y^0)}{t} \\ &= \nabla_y g(x, y^0) r\ \forall x \in I_r(y^0) \end{aligned} \tag{5}$$

since g is continuously differentiable with respect to y.

For $\phi_p(y)$ we can prove the assertions analogously. □

Example 3. *Let some feasible set* $S_D = \{x^1 = (0,0)^\top, x^2 = (0,1)^\top, x^3 = (-1,0)^\top\}$ *be given with functions* $f(x,y) = \frac{1}{2}x^\top x - x^\top y$ *and*

$$g(x,y) = x_1 + x_2 \cdot \begin{cases} y_1^3 \sin \frac{1}{y_1} & y_1 > 0 \\ 0 & y_1 \le 0\,. \end{cases}$$

Then the function $g(x,y)$ *is continuously differentiable with respect to* y. *The regions of stability are*

$$\begin{aligned} R(x^1) &= \{y \in \mathbb{R}^2 : y_1 \ge -0.5,\ y_2 \le 0.5\} \\ R(x^2) &= \{y \in \mathbb{R}^2 : y_1 + y_2 \ge 0,\ y_2 \ge 0.5\} \\ R(x^3) &= \{y \in \mathbb{R}^2 : y_1 \le -0.5,\ y_1 + y_2 \le 0\}. \end{aligned}$$

Let $y^0 = (0, \frac{1}{2})^\top$ *and* $r = (1,0)^\top$*. Then* $I_r(y^0) = \{x^1, x^2\}$ *for both the optimistic and the pessimistic solution function. Thus it holds*

$$\phi_o(y^0; r) = \phi_p(y^0; r) = g(x^1, y^0) = g(x^2, y^0) = 0$$

and

$$\phi_{o_{y^0}}(r) = \phi_{p_{y^0}}(r) = \nabla_y g(x^i, y^0) r, \ i = 1, 2.$$

Further it holds $\phi_o(y^0) = \phi_p(y^0) = 0$*. Remarkable in this example is the fact that for all* $\epsilon > 0$ *there exists some* $t \in (0, \epsilon)$ *with either* $\phi(y^0 + tr) \neq g(x^1, y^0 + tr)$ *or* $\phi(y^0 + tr) \neq g(x^2, y^0 + tr)$*.*

Now let $\bar{y} = (-\frac{1}{2}, \frac{1}{2})^\top$ *and* $r = (-1, 1)^\top$*. Then, for the optimistic solution function it holds*

$$I_r(\bar{y}) = \{x^3\} \text{ and } \phi_o(\bar{y}) = \phi_o(\bar{y}; r) = -1$$

and for the pessimistic solution function it holds

$$I_r(\bar{y}) = \{x^2\} \text{ and } \phi_p(\bar{y}) = \phi_p(\bar{y}; r) = 0.$$

Considering the direction $r = (0, 1)$ *we obtain* $I_r(\bar{y}) = \{x^2\}$ *and* $\phi(\bar{y}; r) = 0$ *for both the optimistic and the pessimistic case, but* $\phi_o(\bar{y}) = -1 \neq 0 = \phi_p(\bar{y})$*.*

Lemma 7. *For all* $y^0 \in \mathbb{R}^n$ *and for all* $r \in \mathbb{R}^n$ *it holds:*

1. $\phi_o(y^0) \leq \phi_o(y^0; r)$
2. $\phi_p(y^0) \geq \phi_p(y^0; r)$

Proof. Assume there exists some y^0 and some r with $\phi_o(y^0) > \phi_o(y^0; r)$. Then from $I_r(y^0) \subseteq \Psi_D(y^0)$ and the proof of Theorem 6 it follows that there exists some $x \in \Psi_D(y^0)$ with $\phi_o(y^0; r) = g(x, y^0)$. Hence, $\phi_o(y^0) > g(x, y^0)$ for some $x \in \Psi_D(y^0)$. This is a contradiction to the definition of ϕ_o.

The proof for ϕ_p is similar. □

6 Optimality criteria based on the radial-directional derivative

Let $\text{locmin}\{\phi(y) : \ y \in Y\}$ denote the set of all local minima of the function $\phi(\cdot)$ over the region $Y \subseteq \mathbb{R}^n$.

Theorem 8. *It holds*

$$\text{locmin}\{\phi_p(y) : \ y \in \mathbb{R}^n\} \subseteq \text{locmin}\{\phi_o(y) : \ y \in \mathbb{R}^n\}.$$

Proof. Arguing by contradiction we assume that there is some y^0 with $y^0 \in \text{locmin}\{\phi_p(y) : y \in \mathbb{R}^n\}$ but $y^0 \notin \text{locmin}\{\phi_o(y) : y \in \mathbb{R}^n\}$. Then there exists some sequence $\{y^k\} \subseteq \mathbb{R}^n$ with $y^k \to y^0, k \to \infty$ and $\phi_o(y^k) < \phi_o(y^0)$. Since $\Psi_D(y^0)$ has finite cardinality and $\Psi_D(y^0) \supseteq \Psi_D(y)$ for all y in a neighborhood of y^0 we can assume w.l.o.g. that there exists some $x \in \Psi_D(y^0)$ with $x \in \Psi_D(y^k)$ and $\phi_o(y^k) = g(x, y^k)\ \forall k$, i.e. $y^k \in Y_x(y^0)\ \forall k$. Since $g(x, \cdot)$ is differentiable with respect to y and $Y_x(y^0) \subseteq \text{cl int } Y_x(y^0)$ we can further assume that $y^k \in \text{int } Y_x(y^0)\ \forall k$. Thus it holds $\Psi_D(y^k) = \{x\}\ \forall k$, i.e. $\phi_o(y^k) = \phi_p(y^k) = g(x, y^k)$ $\forall k$. Consequently,

$$\phi_p(y^k) = \phi_o(y^k) < \phi_o(y^0) \leq \phi_p(y^0) \quad \forall k.$$

This is a contradiction to $y^0 \in \text{locmin}\{\phi_p(y) : y \in \mathbb{R}^n\}$. □

Thus, if $Y = \mathbb{R}^n$ and we know that some point y^0 is a local pessimistic solution then clearly y^0 is a local optimistic solution, too.

Further, for $y \in \text{int}\, Y$ and $y \in \text{locmin}\{\phi_p(y) : y \in Y\}$ it follows analoguously that $y \in \text{locmin}\{\phi_o(y) : y \in Y\}$. As the next example will show we indeed need the condition $y \in \text{int}\, Y$.

Example 4. *Let*

$$\begin{aligned} S_D &= \{(0,1)^\top, (0,-1)^\top\}, \\ Y &= \{y \in \mathbb{R}^2 : y_2 \geq 0\}, \\ f(x,y) &= \frac{1}{2} x^\top x - x^\top y \text{ and} \\ g(x,y) &= x_2 y_1^2 + y_2. \end{aligned}$$

Then it holds for $y \in Y$

$$\phi_p(y) = y_1^2 + y_2 \qquad \text{and} \qquad \phi_o(y) = \begin{cases} y_1^2 + y_2 & \text{if } y_2 > 0 \\ -y_1^2 + y_2 & \text{if } y_2 = 0\,. \end{cases}$$

Thus, $y^0 = (0,0)^\top$ is a local pessimistic but not a local optimistic solution. Moreover, y^0 is a global pessimistic solution. Some local optimistic solution does not exist.

For further considerations we will need the contingent cone of the set Y. For each given point $y^0 \in Y$ this cone is defined as follows:

$$\begin{aligned} T_Y(y^0) \quad := \quad & \{r \in \mathbb{R}^n : \exists \{y^s\} \subseteq Y\ \exists \{t_s\} \downarrow 0 : y^s \to y^0, s \to \infty, \\ & \text{with } \lim_{s \to \infty} \frac{y^s - y^0}{t_s} = r\}. \end{aligned}$$

The set $T_Y(y^0)$ is a convex, closed, nonempty cone [10]. Since Y is convex it holds

$$T_Y(y^0) = \text{cl } \mathbb{T}_Y(y^0)$$

with

$$\mathbb{T}_Y(y^0) = \{r \in \mathbb{R}^n : \exists t_0 > 0 \text{ with } y^0 + tr \in Y \ \forall t \in [0, t_0]\}.$$

Theorem 9. *Let $y^0 \in \mathbb{R}^n$ and let $\phi : \mathbb{R}^n \to \mathbb{R}$ denote the optimistic or the pessimistic solution function. Then $y^0 \notin \operatorname{locmin}\{\phi(y) : y \in Y\}$ if there exists some $r \in \mathbb{T}_Y(y^0)$, $\|r\| = 1$, such that one of the following conditions 1,2 is satisfied:*

1. $d\phi_{y^0}(r) < 0$ and $\phi(y^0; r) = \phi(y^0)$
2. $\phi(y^0; r) < \phi(y^0)$

Proof. Let the vector $r^0 \in \mathbb{T}_Y(y^0)$ with $\|r^0\| = 1$ satisfy condition 1. That means $d\phi_{y^0}(r) = \lim_{t\downarrow 0} t^{-1}(\phi(y^0 + tr^0) - \phi(y^0; r^0)) < 0$. Then there exists some $t' \in (0, t_0)$ such that $\phi(y^0 + tr^0) < \phi(y^0; r^0)$ and $y^0 + tr^0 \in Y \ \forall t \in (0, t')$. Because of $\phi(y^0; r^0) = \phi(y^0)$ we have $\phi(y^0 + tr^0) < \phi(y^0)$ and $y^0 + tr^0 \in Y$ for all $t \in (0, t')$. Thus, y^0 cannot be a local minimum of ϕ.

Now let the vector $r^0 \in \mathbb{T}_Y(y^0)$ with $\|r^0\| = 1$ satisfy condition 2. Then it holds

$$\phi(y^0) - \phi(y^0; r^0) = \phi(y^0) - \lim_{t\downarrow 0} \phi(y^0 + tr^0) > 0.$$

Hence there exists some $t' \in (0, t_0)$ such that $y^0 + tr^0 \in Y$ and $\phi(y^0) > \phi(y^0 + tr^0)$ for all $t \in (0, t')$. Thus, y^0 cannot be a local minimum of ϕ. □

Since $d\phi_{y^0}(\cdot)$ is not continuous it is indeed necessary to consider only the set $\mathbb{T}_Y(y^0)$. The consideration of $T_Y(y^0)$ would not lead to correct results as we will see in the next example.

Example 5. *Let*

$$\begin{aligned} S_D &= \{(-1, 0)^\top, (1, 0)^\top\}, \\ Y &= \{y \in \mathbb{R}^2 : (y_1 - 1)^2 + y_2^2 \leq 1\}, \\ f(x, y) &= \frac{1}{2} x^\top x - x^\top y \text{ and} \\ g(x, y) &= (x_1 + 1)(y_1^2 + y_2^2) + (x_1 - 1) y_2. \end{aligned}$$

Then $T_Y(y^0) = \{y \in \mathbb{R}^2 : y_1 \geq 0\}$ and for $r^0 = (0, 1)^\top$ it holds $\phi_o(y^0) = \phi_o(y^0; r^0) = 0$ and $d\phi_{o_{y^0}}(r^0) = -2 < 0$. Thus, condition 1 is satisfied for ϕ_o and r^0 but $y^0 = (0, 0)^\top$ is a global optimistic and pessimistic optimal solution.

Specifying the conditions of Theorem 9 by using Lemma 7 we obtain the following **necessary optimality conditions:**

Let $y^0 \in \operatorname{locmin}\{\phi_p(y) : y \in Y\}$. Then it holds

$$\phi_p(y^0) = \phi_p(y^0; r) \text{ and } d\phi_{p_{y^0}}(r) \geq 0 \quad \forall r \in \mathbb{T}_Y(y^0).$$

Let $y^0 \in \text{locmin}\{\phi_o(y) : y \in Y\}$. Then for all $r \in \mathbb{T}_Y(y^0)$ it holds

$$\phi_o(y^0) < \phi_o(y^0; r) \text{ or } d\phi_{o y^0}(r) \geq 0.$$

To prove the next theorem we will need the following lemma.

Lemma 10. *Assume it holds $\phi_o(y^0) = g(x^0, y^0)$ for $y^0 \in \mathbb{R}^n$ and $x^0 \in \Psi_D(y^0)$. Then $r \in T_{x^0}(y^0)$ implies*

$$\phi_o(y^0) = \phi_o(y^0; r).$$

Proof. Since ϕ_o is radial-continuous there exists some $\tilde{x} \in \Psi_D(y^0)$ and some sequence $\{t_k\} \downarrow 0$ with $\tilde{x} \in \Psi_D(y^0 + t_k r)$, $\phi_o(y^0 + t_k r) = g(\tilde{x}, y^0 + t_k r)$ and $\phi_o(y^0; r) = \lim_{k\to\infty} \phi_o(y^0 + t_k r) = \lim_{k\to\infty} g(\tilde{x}, y^0 + t_k r) = g(\tilde{x}, y^0)$. Clearly it holds $r \in T_{\tilde{x}}(y^0) \cap T_{x^0}(y^0)$. Then from the proof of Theorem 4 it follows that $r^\top(x^0 - \tilde{x}) = 0$. Further we know that $\tilde{x}, x^0 \in \Psi_D(y^0)$ and thus $F(x^0) - {x^0}^\top y^0 = F(\tilde{x}) - \tilde{x}^\top y^0$. Consequently, $F(x^0) - {x^0}^\top (y^0 + t_k r) = F(\tilde{x}) - \tilde{x}^\top (y^0 + t_k r)\ \forall k$, i.e. $x^0 \in \Psi_D(y^0 + t_k r)\ \forall k$. Thus we obtain $\phi_o(y^0 + t_k r) \leq g(x^0, y^0 + t_k r)\ \forall k$, i.e.

$$\phi_o(y^0; r) = \lim_{k\to\infty} \phi_o(y^0 + t_k r) \leq \lim_{k\to\infty} g(x^0, y^0 + t_k r) = g(x^0, y^0) = \phi_o(y^0).$$

Now from Lemma 7 it follows the equality. □

Theorem 11. *Assume that $y^0 \in Y$ is a point which satisfies one of the following two conditions:*

1. $\phi(y^0) < \phi(y^0; r)\ \forall r \in T_Y(y^0)$
2. $\phi(y^0) \leq \phi(y^0; r)\ \forall r \in T_Y(y^0)$ and $d\phi_{y^0}(r) > \gamma\ \forall r \in T_Y(y^0) : \phi(y^0) = \phi(y^0; r)$, $\|r\| = 1$ with $\gamma = 0$ in the optimistic case and $\gamma > 0$ in the pessimistic case.

Then, ϕ achieves a local minimum at y^0.

Proof. Suppose $y^0 \in Y$ satisfies one of the two conditions of the theorem. Arguing by contradiction we assume that there is a sequence $\{y^k\}_{k\geq 1}$ with $y^k \to y^0, k \to \infty$ and $\phi(y^k) < \phi(y^0)\ \forall k$. Since $\Psi_D(y^0) \supseteq \Psi_D(y)$ for all y in a neighborhood of y^0 and $\Psi_D(y^0)$ has finite cardinality there exists some $x^0 \in \Psi_D(y^0)$ such that $Y_{x^0}(y^0)$ contains infinitely many of the points y^k, i.e. $\phi(y^k) = g(x^0, y^k)$. In the following we consider the sequence $\{y^k\} \cap Y_{x^0}(y^0)$ and denote it by $\{y^k\}$ again. Because of the continuity of $g(x^0, \cdot)$ it follows

$$g(x^0, y^0) = \lim_{k\to\infty} g(x^0, y^k) = \lim_{k\to\infty} \phi(y^k) \leq \phi(y^0). \tag{6}$$

Let $r^k := \frac{y^k - y^0}{\|y^k - y^0\|}$, $k = 1, \ldots, \infty$. Then it holds $r^k \in \mathbb{T}_Y(y^0) \cap T_{x^0}(y^0)$. Further, let $\hat{r}$ an accumulation point of the sequence $\{r^k\}$. Clearly, $\hat{r} \in T_Y(y^0) \cap T_{x^0}(y^0)$.

i) Let ϕ denote the optimistic solution function. Then inequality (6) yields $g(x^0, y^0) = \phi_o(y^0)$ since $x^0 \in \Psi_D(y^0)$. Now from Lemma 10 it follows $\phi_o(y^0) = \phi_o(y^0; r^k)$ $\forall k$ and $\phi_o(y^0) = \phi_o(y^0; \hat{r})$. Thus, the first condition does not hold. Then the second condition must be satisfied. Since $\hat{r} \in T_{x^0}(y^0)$ it holds $y^0 + t\hat{r} \in R(x^0)$ for all $t \in [0, \epsilon)$. Therefore, $\phi_o(y^0 + t\hat{r}) \leq g(x^0, y^0 + t\hat{r})$ for all $t \in [0, \epsilon)$. Hence,

$$\begin{aligned} 0 < d\phi_{y^0}(\hat{r}) &= \lim_{t \downarrow 0} \frac{\phi_o(y^0 + t\hat{r}) - \phi_o(y^0; \hat{r})}{t} \\ &= \lim_{t \downarrow 0} \frac{\phi_o(y^0 + t\hat{r}) - \phi_o(y^0)}{t} \\ &\leq \lim_{t \downarrow 0} \frac{g(x^0, y^0 + t\hat{r}) - g(x^0, y^0)}{t} \\ &= \nabla_y g(x^0, y^0)\hat{r}. \end{aligned}$$

On the other hand it holds

$$\phi_o(y^0) > \phi_o(y^k) = g(x^0, y^0) + \|y^k - y^0\| \nabla_y g(x^0, y^0) r^k + o(\|y^k - y^0\|)$$

which together with $g(x^0, y^0) = \phi_o(y^0)$ and $\lim\limits_{k \to \infty} \frac{o(\|y^k - y^0\|)}{\|y^k - y^0\|} = 0$ leads to

$$\nabla_y g(x^0, y^0)\hat{r} \leq 0.$$

But this is a contradiction, i.e. if y^0 is no local optimistic solution none of the two conditions holds.

ii) Let ϕ denote the pessimistic solution function. Then from Lemma 7 it follows $\phi_p(y^0) \geq \phi_p(y^0; r)$ for all $r \in T_Y(y^0)$, i.e. the first condition is not satisfied. Then the second condition must be satisfied, i.e. it holds $\phi_p(y^0) = \phi_p(y^0; r)$ and $d\phi_{p_{y^0}}(r) > \gamma > 0$ for all $r \in T_Y(y^0)$.

Since ϕ_p is radial-continuous and radial differentiable for all k there exists some $x^k \in I_{r^k}(y^0)$. Because of $I_{r^k}(y^0) \subseteq \Psi_D(y^0)$ $\forall k$ and the finite cardinality of $\Psi_D(y^0)$ we can assume w.l.o.g. that there exists some $\bar{x} \in \Psi_D(y^0)$ with $\bar{x} \in I_{r^k}(y^0)$ $\forall k$. Thus, for all k it holds

$$\phi_p(y^0; r^k) = \phi_p(y^0) = g(\bar{x}, y^0) \quad \text{and} \quad 0 < \gamma < d\phi_{p_{y^0}}(r^k) = \nabla_y g(\bar{x}, y^0) r^k.$$

Since $\hat{r}$ is an accumulation point of $\{r^k\}$ we obtain

$$0 < \gamma \leq \nabla_y g(\bar{x}, y^0)\hat{r}.$$

Further we have $r^k \in T_{x^0}(y^0) \cap T_{\bar{x}}(y^0)$. Thus, $y^0 + tr^k \in R(x^0) \cap R(\bar{x})$ $\forall t \in [0, \epsilon)$ which yields $\phi_p(y^k) \geq g(\bar{x}, y^0)$ $\forall k$. Consequently, for all k it holds

$$\phi_p(y^0) > \phi_p(y^k) = g(\bar{x}, y^k) = g(\bar{x}, y^0) + \|y^k - y^0\| \nabla_y g(x^0, y^0) r^k + o(\|y^k - y^0\|)$$

which together with $g(\bar{x}, y^0) = \phi_p(y^0)$ and $\lim\limits_{k\to\infty} \frac{o(||y^k - y^0||)}{||y^k - y^0||} = 0$ leads to

$$\nabla_y g(\bar{x}, y^0)\hat{r} \leq 0.$$

But this is a contradiction, i.e. if y^0 is no local pessimistic solution none of the two conditions holds. □

Specifying the conditions of Theorem11 by using Lemma 7 we obtain the following **sufficient optimality conditions:**

Let

$$\phi_p(y^0) = \phi_p(y^0; r) \text{ and } d\phi_{p_{y^0}}(r) > \gamma > 0 \quad \forall r \in T_Y(y^0).$$

Then $y^0 \in \text{locmin}\{\phi_p(y) : y \in Y\}$.

Let

$$\phi_o(y^0) < \phi_o(y^0; r) \text{ or } d\phi_{o_{y^0}}(r) > 0 \quad \forall r \in T_Y(y^0).$$

Then $y^0 \in \text{locmin}\{\phi_o(y) : y \in Y\}$.

Example 6. *Consider the bilevel programming problem*

$$\begin{cases} \min\{g(x,y) : y \in \mathbb{R}^2, \ x \in \Psi_D(y)\} \\ \\ \Psi_D(y) = \text{Argmin}\{\frac{1}{2}||x||^2 - y^\top x : x_1 \leq 0, x_2 \geq 0, -x_1 + x_2 \leq 1, \ x \in \mathbb{Z}^2\} \end{cases}$$

with $g(x, y) = x_2(y_2 - (y_1 + 0.5)^2 - 0.5) + (1 - x_2)(y_1 - y_2 + 1) + x_1(3y_1 + 1.5)$. *We obtain*

$$\begin{aligned} S_D &= \{x^1 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, x^2 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, x^3 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}\} \text{ with} \\ R(x^1) &= \{y \in \mathbb{R}^2 : y_2 \geq 0.5, y_1 + y_2 \geq 0\}, \\ R(x^2) &= \{y \in \mathbb{R}^2 : y_2 \leq 0.5, y_1 \geq -0.5\} \text{ and} \\ R(x^3) &= \{y \in \mathbb{R}^2 : y_1 \leq -0.5, y_1 + y_2 \leq 0\}. \end{aligned}$$

Then we have

$$\phi_p(y) = \begin{cases} y_2 - (y_1 + 0.5)^2 - 0.5 & \text{if} \quad y_2 > 0.5, y_1 + y_2 > 0 \\ y_1 - y_2 + 1 & \text{if} \quad y_2 \leq 0.5, y_1 \geq -0.5 \\ -2y_1 - y_2 - 0.5 & \text{if} \quad y_1 + y_2 \leq 0, y_1 < -0.5\ . \end{cases}$$

Let $y^0 = (-1/2, 1/2)^\top$. *Then it holds* $\phi_p(y^0) = \phi_p(y^0; r) = 0 \ \forall r \in \mathbb{R}^2$ *and*

$$0 < d\phi_{p_{y^0}}(r) = \begin{cases} r_2 & \text{if} \quad r_2 > 0, r_1 + r_2 > 0 \\ r_1 - r_2 & \text{if} \quad r_2 \leq 0, r_1 \geq 0 \\ -2r_1 - r_2 & \text{if} \quad r_1 < 0, r_1 + r_2 \leq 0\ . \end{cases}$$

However y^0 *is no local minimum of* ϕ_p *since* $y(t) = (t - 0.5, 0.5(1 + t^2))^\top \to y^0$ *for* $t \downarrow 0$ *but* $\phi_p(y(t)) = -\frac{1}{2}t^2 < \phi_p(y^0) \ \forall t > 0$. *This is no contradiction to Theorem 11 since there does not exist any* $\gamma > 0$ *with* $\gamma < d\phi_{p_{y^0}}(r) \ \forall r$.

7 Optimality criteria using radial subdifferential

Definition 3. *Let $U \subseteq \mathbb{R}^n$, $y^0 \in U$ and $\phi : U \to \mathbb{R}$ be radial-directionally differentiable at y^0. We say that $d \in \mathbb{R}^n$ is a radial subgradient of ϕ at y^0 if*

$$\phi(y^0) + \langle r, d\rangle \leq \phi(y^0; r) + d\phi_{y^0}(r)$$

is satisfied for all $r : \phi(y^0) \geq \phi(y^0; r)$.

The set of all subgradients is called subdifferential and denoted by $\partial_{rad}\phi(y^0)$.

The following necessary criterion for the existence of a radial subgradient is valid:

Theorem 12 ([6]). *If there exists some $r \in \mathbb{R}^n$ with $\phi(y^0; r) < \phi(y^0)$ then it holds $\partial_{rad}\phi(y^0) = \emptyset$.*

With this theorem we get the following equivalent definition of the radial subgradient:

$$\partial_{rad}\phi(y^0) = \{d \in \mathbb{R}^n : \langle r, d\rangle \leq d\phi_{y^0}(r)\, \forall r \text{ satisfying } \phi(y^0) = \phi(y^0; r)\},$$

if there is no direction such that the radial limit in this direction is less than the function value.

Using Lemma 7 we obtain that for the pessimistic solution function either $\partial_{rad}\phi_p(y^0) = \emptyset$ if there exists some r with $\phi_p(y^0) > \phi_p(y^0; r)$ or $\partial_{rad}\phi_p(y^0) = \{d \in \mathbb{R}^n : \langle d, r\rangle \leq d\phi_{p_{y^0}}(r)\ \forall r\}$.
For the optimistic solution function the condition of Theorem 12 is never valid.

Thus,

$$\begin{aligned}\partial_{rad}\phi_o(y^0) = \{d \in \mathbb{R}^n : \langle r, d\rangle \leq d\phi_{o_{y^0}}(r)\, \forall r \\ \text{satisfying } \phi_o(y^0) = \phi_o(y^0; r)\}\end{aligned}$$

and

$$\partial_{rad}\phi_p(y^0) = \{d \in \mathbb{R}^n : \langle r, d\rangle \leq d\phi_{p_{y^0}}(r)\, \forall r\}$$

if there is no r such that $\phi_p(y^0; r) < \phi_p(y^0)$.

Next want to give further descriptions for the set $\partial_{rad}\phi(y^0)$ by using equation (5). To do this we will need the following notations:

$$\begin{aligned}\mathcal{T}(y^0) &:= \{r \in \mathbb{R}^n : \phi(y^0) = \phi(y^0; r)\} \\ I(y^0) &:= \bigcup_{r \in \mathcal{T}(y^0)} I_r(y^0)\end{aligned}$$

Then the following Lemma holds:

Lemma 13. *1.* $I(y^0) = \{x \in \Psi_D(y^0) : g(x, y^0) = \phi(y^0)\}$

2. $\text{cl}\ \mathcal{T}(y^0) = \bigcup_{x \in I(y^0)} T_x(y^0)$

3. $\partial_{rad}\phi(y^0) = \bigcap_{x \in I(y^0)} \{d \in \mathbb{R}^n : \langle d, r\rangle \leq \nabla_y g(x, y^0) r\ \forall r \in T_x(y^0)\}$

Proof. 1. Let $x \in I(y^0)$. Then there exists some $r \in \mathcal{T}(y^0)$ with $x \in I_r(y^0)$. Because of the definitions of $I_r(y^0)$ and $\mathcal{T}(y^0)$ it holds $\phi(y^0) = \phi(y^0; r) = g(x, y^0)$ and $x \in \Psi_D(y^0)$, i.e. $x \in \{x \in \Psi_D(y^0) : g(x, y^0) = \phi(y^0)\}$.

Let $x^0 \in \Psi_D(y^0)$ with $g(x^0, y^0) = \phi(y^0)$. Let $r = \nabla F(x^0)^\top - y^0$. Since $\nabla F(x^0)^\top \in \text{int}\ R(x^0)$ and $y^0 \in R(x^0)$ it holds $\lambda \nabla F(x^0)^\top + (1-\lambda) y^0 \in \text{int}\ R(x^0)\ \forall \lambda \in (0,1)$, i.e. $y^0 + \lambda r \in \text{int}\ R(x^0)\ \forall \lambda \in (0,1)$. Consequently, $\phi(y^0 + \lambda r) = g(x^0, y^0 + \lambda r)\ \forall \lambda \in (0,1)$. But this means $x^0 \in I_r(y^0)$ and $\phi(y^0; r) = g(x^0, y^0) = \phi(y^0)$, i.e. $r \in \mathcal{T}(y^0)$ and thus $x^0 \in I(y^0)$.

2. Let $\hat{r} \in \text{cl}\ \mathcal{T}(y^0)$. Then there exists some sequence $\{r^k\}_{k=1}^\infty \subseteq \mathcal{T}(y^0)$ with $\lim_{k\to\infty} r^k = \hat{r}$. Since $I(y^0) \subseteq \Psi_D(y^0)$ and $\text{card}\ \Psi_D(y^0) < \infty$ there exists w.l.o.g. some $x \in I(y^0)$ with $x \in I_{r^k}(y^0)\ \forall k$, i.e. $r^k \in T_x(y^0)\ \forall k$. Then from $T_x(y^0)$ being closed it follows that $\hat{r} \in T_x(y^0)$, i.e. $\hat{r} \in \bigcup_{x \in I(y^0)} T_x(y^0)$.

Let $\hat{r} \in \bigcup_{x \in I(y^0)} T_x(y^0)$. Then there exists some $x \in I(y^0)$ with $\hat{r} \in T_x(y^0)$. Since $T_x(y^0) \subseteq \text{cl int}\ T_x(y^0)$ there exists some sequence $\{r^k\}_{k=1}^\infty \subseteq \text{int}\ T_x(y^0)$ with $\lim_{k\to\infty} r^k = \hat{r}$. Thus, for all k it holds $y^0 + t r^k \in \text{int}\ R(x)$ for all $t > 0$ being sufficiently small. This means $x \in I_{r^k}(y^0)$ and $r^k \in \mathcal{T}(y^0)$ for all k. Consequently, $\hat{r} \in \text{cl}\ \mathcal{T}(y^0)$.

3. Let $d \in \bigcap_{x \in I(y^0)} \{d \in \mathbb{R}^n : \langle d, r\rangle \leq \nabla_y g(x, y^0) r\ \forall r \in T_x(y^0)\}$. Then for all $x \in I(y^0)$ it holds $\langle d, r\rangle \leq \nabla_y g(x, y^0) r = d\phi_{y^0}(r)\ \forall r \in T_x(y^0)$. Thus, $\langle d, r\rangle \leq d\phi_{y^0}(r)\ \forall r \in \bigcup_{x \in I(y^0)} T_x(y^0) \supseteq \mathcal{T}(y^0)$, i.e. $d \in \partial_{rad}\phi(y^0)$.

Let $d \in \partial_{rad}\phi(y^0)$. Now consider some arbitrary $x \in I(y^0)$ and some $r \in T_x(y^0)$. Then there exist some sequence $\{r^k\}_{k=1}^\infty \subseteq \text{int}\ T_x(y^0)$ and $\lim_{k\to\infty} r^k = r$. Since $\text{int}\ T_x(y^0) \subseteq \mathcal{T}(y^0)$ and $d \in \partial_{rad}\phi(y^0)$ it follows $\langle d, r^k\rangle \leq d\phi_{y^0}(r^k) = \nabla_y g(x, y^0) r^k\ \forall k$ and thus $\langle d, r\rangle \leq \nabla_y g(x, y^0) r$. Consequently, $d \in \bigcap_{x \in I(y^0)} \{d \in \mathbb{R}^n : \langle d, r\rangle \leq \nabla_y g(x, y^0) r\ \forall r \in T_x(y^0)\}$.

□

Lemma 14. *For all points $y^0 \in \mathbb{R}^n$ and $\bar{x} \in \Psi_D(y^0)$ the set*

$$N_{\bar{x}}(y^0) := \text{cone}\,\{(x - \bar{x}) : x \in \Psi_D(y^0)\}$$

is the normal cone of the contingent cone $T_{\bar{x}}(y^0)$. Further it holds

$$\partial_{rad}\phi(y^0) = \bigcap_{x \in I(y^0)} (N_x(y^0) + \nabla_y g(x, y^0)^\top).$$

Proof. We know from Lemma 5 that the contingent cone $T_{\bar{x}}(y^0)$ is equal to

$$T_{\bar{x}}(y^0) = \{r \in \mathbb{R}^n : \ (x-\bar{x})^\top r \le 0 \quad \forall x \in \Psi_D(y^0)\}.$$

Obviously it is the normal cone of the polyhedral cone $N_{\bar{x}}(y^0)$. Since $N_{\bar{x}}(y^0)$ is convex and closed the normal cone of $T_{\bar{x}}(y^0)$ is $N_{\bar{x}}(y^0)$ again.

Let $d \in \partial_{rad}\phi(y^0)$. Because of Lemma 13 for all $x \in I(y^0)$ it holds

$$\langle d, r\rangle \le d\phi_{y^0}(r) = \nabla_y g(x, y^0) r \qquad \forall r \in T_x(y^0).$$

Consequently, $d - \nabla_y g(x, y^0)^\top$ lies in the normal cone of $T_x(y^0)$ for all $x \in I(y^0)$, i.e. $d - \nabla_y g(x, y^0)^\top \in N_x(y^0)$ for all $x \in I(y^0)$. But this means

$$d \in N_x(y^0) + \nabla_y g(x, y^0)^\top \text{ for all } x \in I(y^0).$$

Thus, $\partial_{rad}\phi(y^0) \subseteq \bigcap_{x \in I(y^0)} (N_x(y^0) + \nabla_y g(x, y^0)^\top)$. The reverse inclusion follows analogously. □

Example 7. *Let* $y^0 = (0,0)^\top$ *and*

$$\begin{aligned} S_D &= \{x^1 = \begin{pmatrix}1\\1\end{pmatrix}, x^2 = \begin{pmatrix}1\\-1\end{pmatrix}, x^3 = \begin{pmatrix}-1\\1\end{pmatrix}, x^4 = \begin{pmatrix}-1\\-1\end{pmatrix}\} \\ f(x,y) &= \frac{1}{2}\|x\|^2 - x^\top y \\ g(x,y) &= (\frac{3+x_1}{2})y_1 - 2y_2 + (x_1 - x_2)^2. \end{aligned}$$

Then it holds

$$\begin{array}{ll} R(x^1) = \{y : \ y_1 \ge 0, y_2 \ge 0\}, & g(x^1, y) = 2y_1 - 2y_2 \\ R(x^2) = \{y : \ y_1 \ge 0, y_2 \le 0\}, & g(x^2, y) = 2y_1 - 2y_2 + 4 \\ R(x^3) = \{y : \ y_1 \le 0, y_2 \ge 0\}, & g(x^3, y) = y_1 - 2y_2 + 4 \\ R(x^4) = \{y : \ y_1 \le 0, y_2 \le 0\}, & g(x^4, y) = y_1 - 2y_2. \end{array}$$

Consequently, for the optimistic solution function it holds $\phi_o(y^0) = 0$ *and* $I(y^0) = \{x^1, x^4\}$. *Further, since* $N_{x^1}(y^0) = R(x^4)$, $N_{x^4}(y^0) = R(x^1)$ *and* $\nabla_y g(x^1, y^0) = (2, -2)$, $\nabla_y g(x^4, y^0) = (1, -2)$ *it holds*

$$\begin{aligned} \partial_{rad}\phi(y^0) &= (N_{x^1}(y^0) + \nabla_y g(x^1, y^0)^\top) \cap (N_{x^4}(y^0) + \nabla_y g(x^4, y^0)^\top) \\ &= \{d \in \mathbb{R}^2 : \ d_1 \le 2, \ d_2 \le -2\} \cap \{d \in \mathbb{R}^2 : \ d_1 \ge 1, \ d_2 \ge -2\} \\ &= [1,2] \times \{-2\}. \end{aligned}$$

Now we derive optimality criteria in connection with the radial subdifferential.

Assume some point $y^0 \in \operatorname{locmin}\{\phi(y) : \ y \in Y\}$ is given. Then we know from Theorem 9 that for all $r \in \mathbb{T}_Y(y^0)$, $\|r\| = 1$ it holds $\phi(y^0; r) \ge \phi(y^0)$ and $d\phi_{y^0}(r) \ge 0$ if $\phi(y^0; r) = \phi(y^0)$. Consequently,

$$\begin{array}{lll} & 0 \le \nabla_y g(x, y^0) r & \forall x \in I(y^0) \quad \forall r \in T_x(y^0) \cap \mathbb{T}_Y(y^0) \\ \text{and thus} & 0 \le \nabla_y g(x, y^0) r & \forall x \in I(y^0) \quad \forall r \in T_x(y^0) \cap T_Y(y^0). \end{array}$$

This means that $-\nabla_y g(x,y^0)^\top$ lies in the normal cone of $T_x(y^0)\cap T_Y(y^0)$ for all $x\in I(y^0)$. Let $\bar{I}(y^0) := \{x\in I(y^0):\ \text{ri}\,T_Y(y^0)\cap \text{ri}\,T_x(y^0)\neq\emptyset\}$. Since both cones are convex and closed the normal cone of $T_x(y^0)\cap T_Y(y^0)$ is equal to $N_Y(y^0)+N_x(y^0)$ for $x\in\bar{I}(y^0)$ where $N_Y(y^0)$ denotes the normal cone of $T_Y(y^0)$. Consequently,

$$\begin{aligned}
-\nabla_y g(x,y^0)^\top &\in N_Y(y^0)+N_x(y^0) \quad \forall x\in\bar{I}(y^0)\\
0 &\in N_Y(y^0)+(N_x(y^0)+\nabla_y g(x,y^0)^\top) \quad \forall x\in\bar{I}(y^0)\\
0 &\in \bigcap_{x\in\bar{I}(y^0)} \left[N_Y(y^0)+(N_x(y^0)+\nabla_y g(x,y^0)^\top)\right].
\end{aligned}$$

If it holds $y^0\in\text{int}\,Y$ we have $N_Y(y^0)=\{0\}$, $T_Y(y^0)=\mathbb{R}^n$ and $I(y^0)=\bar{I}(y^0)$. Thus, it holds the following theorem:

Theorem 15. *Let ϕ denote the optimistic or pessimistic solution function for the bilevel programming problem (1). If $y^0\in locmin\{\phi(y):\ y\in Y\}$ then*

$$0\in \bigcap_{x\in\bar{I}(y^0)} \left[N_Y(y^0)+(N_x(y^0)+\nabla_y g(x,y^0)^\top)\right].$$

If additionally $y^0\in\text{int}\,Y$ then $0\in\partial_{rad}\phi(y^0)$.

Example 8. *Let S_D denote the vertex set of a regular hexagon with radius 2, i.e. let S_D be equal to*

$$\{x^1=\begin{pmatrix}2\\0\end{pmatrix}, x^2=\begin{pmatrix}1\\\sqrt{3}\end{pmatrix}, x^3=\begin{pmatrix}-1\\\sqrt{3}\end{pmatrix}, x^4=\begin{pmatrix}-2\\0\end{pmatrix}, x^5=\begin{pmatrix}-1\\-\sqrt{3}\end{pmatrix}, x^6=\begin{pmatrix}1\\\sqrt{3}\end{pmatrix}\}.$$

Further let

$$g(x,y)=\begin{cases} y_1+4y_2 & \text{if } x=x^1\\ 2y_1-y_2 & \text{if } x=x^2\\ 1 & \text{else}\ . \end{cases}$$

Consider the optimistic solution function $\phi_o(y)$ and the set $Y=\{y\in\mathbb{R}^2:\ y_2\geq 0, y_2\leq y_1, y_1\leq 1\}$. Then $y^0=(0,0)^\top$ is an optimistic optimal solution. It holds $I(y^0)=\{x^1,x^2\}=\bar{I}(y^0)$. Further,

$$\begin{aligned}
N_{x^1}(y^0) &= \{y: y_2\leq -\sqrt{3}y_1, y_2\geq\sqrt{3}y_1\},\\
N_{x^2}(y^0) &= \{y: y_2\leq -\sqrt{3}y_1, y_2\leq 0\}\\
\textit{and}\quad N_Y(y^0) &= \{y: y_1+y_2\leq 0, y_1\leq 0\}.
\end{aligned}$$

Then it holds

$$\begin{aligned}
0 &\in N_Y(y^0)+N_{x^1}(y^0)+\nabla_y g(x^1,y^0)^\top\\
&= \{y\in\mathbb{R}^2:\ y_1\leq 1,\ y_2\leq 4-(y_1-1)\sqrt{3}\},\\
0 &\in N_Y(y^0)+N_{x^2}(y^0)+\nabla_y g(x^2,y^0)^\top\\
&= \{y\in\mathbb{R}^2:\ y_1+y_2\leq 1,\ y_2\leq -1-(y_1-2)\sqrt{3}\}.
\end{aligned}$$

Thus, the conditions of Theorem 15 are satisfied. Further,

$$\partial_{rad}\phi_o(y^0) = \{d \in \mathbb{R}^2 : \ d_2 \leq -1, \ d_2 \geq 4 + (d_1 - 1)\sqrt{3}\}.$$

In optimization one has very often necessary optimality criteria of the form $0 \in \partial\phi(y^0) + N_Y(y^0)$*. Such kind of necessary optimality criterium is usually not fulfilled for our problem. For instance in this example it holds* $0 \notin \partial_{rad}\phi_o(y^0) + N_Y(y^0)$*.*

Theorem 16. *Let* ϕ *denote the optimistic or pessimistic solution function for the bilevel programming problem (1). If* $0 \in int\,(\partial_{rad}\phi(y^0) + N_Y(y^0))$ *then* ϕ *achieves at* y^0 *a local minimum.*

Proof. Clearly $\partial_{rad}\phi(y^0) \neq \emptyset$. Thus it holds $\phi(y^0) \leq \phi(y^0; r)\ \forall r \in \mathbb{R}^n, \|r\| = 1$ because of Theorem 12.

Let $0 \in int\,(\partial_{rad}\phi(y^0) + N_Y(y^0))$. Then there exists some $\gamma > 0$ such that for all $r \in \mathbb{R}^n$, $\|r\| = 1$ it holds $\gamma r \in (\partial_{rad}\phi(y^0) + N_Y(y^0))$. Now fix some $\hat{r} \in \mathcal{T}(y^0) \cap T_Y(y^0)$, $\|\hat{r}\| = 1$. Then there exists some $s \in N_Y(y^0)$ with $(\gamma\hat{r} - s) \in \partial_{rad}\phi(y^0)$. Using the definition of $\partial_{rad}\phi(y^0)$ we obtain

$$\gamma\langle \hat{r}, r\rangle - \langle s, r\rangle = \langle \gamma\hat{r} - s, r\rangle \leq d\phi_{y^0}(r) \quad \forall r \in \mathcal{T}(y^0).$$

Because of $\hat{r} \in T_Y(y^0)$ and $s \in N_Y(y^0)$ it holds $\langle \hat{r}, s\rangle \leq 0$ and thus

$$0 < \gamma \leq \gamma\|\hat{r}\|^2 - \langle s, \hat{r}\rangle \leq d\phi_{y^0}(\hat{r}).$$

Thus, since $\hat{r}$ was arbitrary the sufficient optimality criterium is satisfied (Theorem 11), i.e.

$$0 < \gamma \leq d\phi_{y^0}(r) \qquad \forall r \in \mathcal{T}(y^0) \cap T_Y(y^0).$$

Hence, $y^0 \in \text{locmin}\{\phi(y) : \ y \in Y\}$. □

References

[1] J. F. Bard, Practical Bilevel Optimization: Algorithms and Applications, Kluwer Academic Publishers, Dordrecht, 1998

[2] S. Dempe, Foundations of Bilevel Programming, Kluver Academic Publishers, Dordrecht, 2002

[3] S. Dempe, Annotated Bibliography on Bilevel Programming and Mathematical Programs with Equilibrium Constraints, Optimization, 2003, Vol. 52, pp. 333-359

[4] S. Dempe, V. Kalashnikov and Roger Z. Ríos-Mercado, Discrete Bilevel Programming: Application to a Natural Gas Cash-Out Problem, to appear in European Journal of Operational Research

[5] S. Dempe and K. Richter, Bilevel programming with knapsack constraints, Central European Journal of Operations Research, 2000, Vol. 8, pp. 93-107

[6] S. Dempe, T. Unger, Generalized PC^1-Functions, Optimization, 1999, Vol. 46, pp. 311-326

[7] V. F. Dem'yanov, A. N. Rubinov, Quasidifferential Calculus, Optimization Software, Publications Division, New York, 1986

[8] T. Edmunds and J.F. Bard, An algorithm for the mixed-integer nonlinear bilevel programming problem, Annals of Operations Research, 1992, Vol. 34, pp. 149–162

[9] R.-H. Jan and M.-S. Chern, Nonlinear integer bilevel programming, European Journal of Operational Research, 1994, Vol. 72, pp. 574-587

[10] J.-P. Hiriart-Urruty and C. Lemarechal, Convex Analysis and Minimization Algorithms, Vol. 1, Springer-Verlag, Berlin et. al., 1993

[11] P. Loridan and J. Morgan, On strict ε-solutions for a two-level optimization problem, In: W. Buhler et at. (eds.), Proceedings of the International Conference on Operations Research 90, Springer Verlag, Berlin, 1992, pp. 165-172

[12] P. Loridan and J. Morgan, Weak via strong Stackelberg problem: New results, Journal of Global Optimization, 1996, Vol. 8, pp. 263-287

[13] P. Recht, Generalized Derivatives: An Approach to a New Gradient in Nonsmooth Optimization, volume 136 of Mathematical Systems in Economics, Anton Hain, Frankfurt am Main, 1993

[14] G. Savard and J. Gauvin, The steepest descent direction for the nonlinear bilevel programming problem, Operations Research Letters, 1994, Vol. 15, pp. 265-272

[15] L. N. Vicente and G. Savard and J. J. Judice, The discrete linear bilevel programming problem, Journal of Optimization Theory and Applications, 1996, Vol. 89, pp. 597-614

[16] U. Wen and Y. Yang, Algorithms for solving the mixed integer two-level linear programming problem, Computers and Operations Research, 1990, Vol. 17, pp. 133–142