Chapter 2
Linear Bilevel Optimization Problem

2.1 The Model and First Properties

The linear bilevel optimization problem illustrated in Example 1.1 is the problem of the following structure

$$\min_{x,y}\{a^\top x + b^\top y : Ax + By \leq c, \ (x, y) \in \mathbf{gph}\, \Psi\}, \tag{2.1}$$

where $\Psi(\cdot)$ is the solution set mapping of the lower level problem

$$\Psi(x) := \underset{y}{\mathrm{Argmin}}\,\{d^\top y : Cy \leq x\}. \tag{2.2}$$

Here, A is a (p, n)-, B a (p, m)- and C a (n, m)-matrix and all variables and vectors used are of appropriate dimensions. Note that we have used here the so-called optimistic bilevel optimization problem, which is related to problem (1.4).

We find so-called connecting constraints $Ax + By \leq c$ in the upper level problem. Validity of such constraints is beyond the choice of the leader and can be verified only after the follower has selected his/her possibly not unique optimal solution. Especially in the case when $\Psi(x)$ does not reduce to a singleton this can be difficult. For investigating the bilevel programming problem in the case that $\Psi(x)$ does not reduce to a singleton, Ishizuka and Aiyoshi [153] introduced their double penalty method. In general, *connecting constraints* may imply that the feasible set of the bilevel programming problem is disconnected. This situation is illustrated by the following example:

Example 2.1 (Mersha and Dempe [227]). Consider the problem

$$\min_{x,y} -x - 2y$$

$$\text{subject to} \qquad \begin{aligned} 2x - 3y &\geq -12 \\ x + y &\leq 14 \end{aligned}$$

$$\text{and } y \in \underset{y}{\mathrm{Argmin}}\,\{-y : -3x + y \leq -3, \ 3x + y \leq 30\}.$$

Fig. 2.1 The problem with *upper level* connecting constraints. The feasible set is depicted with *bold lines*. The point C is global optimal solution, point A is a local optimal solution
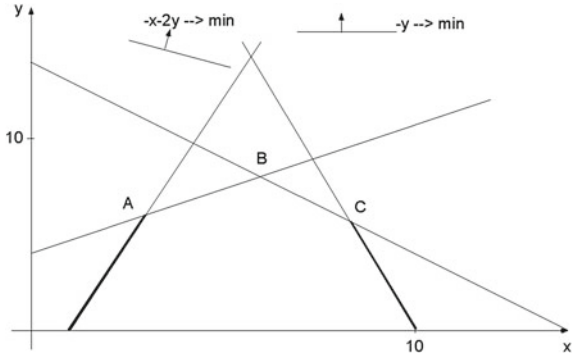
Fig. 2.2 The problem when the *upper level* connecting constraints are shifted into the *lower level* problem. The feasible set is depicted with *bold lines*. The global optimal solution is point B
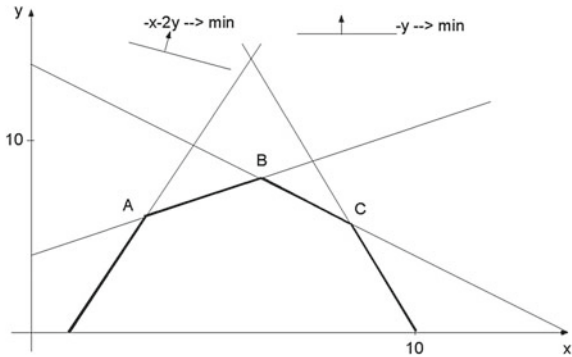
The optimal solution for this problem is point C at $(\overline{x}, \overline{y}) = (8, 6)$ (see Fig. 2.1). But if we shift the two upper level constraints to the lower level we get point B at $(\widetilde{x}, \widetilde{y}) = (6, 8)$ as an optimal solution (see Fig. 2.2). From this example it can easily be noticed that if we shift constraints from the upper level to the lower one, the optimal solution obtained prior to shifting is not optimal any more in general. Hence ideas based on shifting constraints from one level to another will lead to a solution which may not be a solution prior to shifting constraints.   □

In Example 2.1 the optimal solution of the lower level problem was unique for all x. If this is not the case, feasibility of a selection of the upper level decision maker possibly depends on the selection of the follower. In the optimistic case this means that the leader selects within the set of optimal solutions of the follower's problem one point which is at the same time feasible for the upper level connecting constraints and gives the best objective function value for the upper level objective function.

As we can see in Example 2.1 the existence of connecting upper level constraints will lead in general to disconnected feasible sets in the bilevel programming problem. Therefore, solution algorithms will live in one of the connected components of the feasible set (i.e. a sequence of feasible points which all belong to one of the connected parts is computed) or they need to jump from one of the connected parts of the feasible set to another one. This would use then ideas of discrete optimization.

In the following we will avoid this additional difficulty in assuming that the upper level constraints will depend on the upper level variables only. Hence, we consider the *linear bilevel optimization problem*

$$\min_{x,y}\{a^\top x + b^\top y : Ax \le c, \ (x, y) \in \mathbf{gph}\ \Psi\}, \qquad (2.3)$$

where $\Psi(\cdot)$ is the solution set mapping of the lower level problem

$$\Psi(x) := \underset{y}{\mathrm{Argmin}}\ \{d^\top y : Cy \le x\}. \qquad (2.4)$$

In this problem, parametric linear optimization (see e.g. Nožička et al. [257]) can be used to show that the graph of the mapping $\Psi(\cdot)$ equals the connected union of faces of the set $\{(x, y)^\top : Cy \le x\}$.

Here, a set M is connected if it is not contained in the union of two disjoint open sets $M \subset M_1 \cup M_2$, M_1, M_2 are open and not empty, $M_1 \cap M_2 = \emptyset$, having nonempty intersection with both of these sets: $M \cap M_i \ne \emptyset$, $i = 1, 2$.

Hence, the convex hull of this set is a convex polyhedron implying that problem (2.3) is a linear optimization problem. Thus, its optimal solution can be found at a vertex of the set

$$\{(x, y)^\top : Cy \le x, \ Ax \le c\}.$$

Theorem 2.1 *If problem (2.3) has an optimal solution, at least one global optimal solution occurs at a vertex of the set*

$$\{(x, y)^\top : Cy \le x, \ Ax \le c\}.$$

This theorem can be found in the article [40] by Candler and Townsley, it is the basis of many algorithms using (implicit or not complete) enumeration to compute a global optimum of problem (2.3) (see e.g. Bard [10]).

This property is lost if problem (2.1) with upper level connecting constraints is considered.

As it can be seen in Fig. 2.2, the bilevel optimization problem is a nonconvex optimization problem, it has a feasible set which is not given by explicit constraints. As a result, besides a global optimal solution bilevel optimization problems can have local extrema and stationary solutions which are not local optimal solutions.

In Sect. 1.2, the bilevel optimization problem has been interpreted as an hierarchical game of two players, the leader and the follower where the leader is the first to make a choice and the follower reacts optimally on the leader's selection. It has been shown in the article [11] by Bard and Falk that the solution strongly depends on the order of play: the leader may take advantage from having the first selection.

The following theorem shows that the (linear) bilevel optimization problem is $\mathcal{NP}$-hard in the strong sense which implies that it is probably not possible to find a polynomial algorithm for computing a global optimal solution of it. For more results on complexity theory the interested reader is referred to the monograph [126] by Garey and Johnson.

Theorem 2.2 (Deng [85]) *For any $\varepsilon > 1$ it is $\mathcal{NP}$-hard to find a solution of the linear bilevel optimization problem* (2.3) *with not more than ε times the global optimal function value of this problem.*

In the next example we will see that the bilevel programming problem depends on constraints being not active in the lower level problem. Hence, a global optimal solution of the bilevel problem can loose its optimality if an inequality is added which is not active at the global minimum. This behavior may be a bit surprising since it is not possible in problems on continuous (nonsmooth) optimization.

Example 2.2 (*Macal and Hurter* [210]) Consider the unconstrained bilevel optimization problem

$$
\begin{aligned}
&(x-1)^2 + (y-1)^2 \rightarrow \min_{x,y}, \\
&\text{where } y \text{ solves} \\
&0.5y^2 + 500y - 50xy \rightarrow \min_{y}.
\end{aligned}
\tag{2.5}
$$

Since the lower level problem is unconstrained and convex we can replace it by its necessary optimality conditions. Then, problem (2.5) becomes

$$
\min_{x,y}\{(x-1)^2 + (y-1)^2 : y - 50x + 500 = 0\}.
$$

The unique optimal solution of this problem is $(x^*, y^*) = (50102/5002, 4100/5002)$ with an optimal objective function value of $z^* = 81, 33$.

Now, add the constraint $y \geq 0$ to the lower level problem and consider the problem

$$
\begin{aligned}
&(x-1)^2 + (y-1)^2 \rightarrow \min_{x,y}, \\
&\text{where } y \text{ solves} \\
&y \in \underset{y}{\text{Argmin}}\,\{0.5y^2 + 500y - 50xy : y \geq 0\}.
\end{aligned}
\tag{2.6}
$$

The unique global optimal solution of problem (2.6) is $(\overline{x}, \overline{y}) = (1, 0)$. This point is not feasible for (2.5). Its objective function value in problem (2.6) is 1 showing that (x^*, y^*) is a local optimum but not the global optimal solution of problem (2.6).   □

In the next theorem we need the notion of an inner semicontinuous mapping.

Definition 2.1 (*Mordukhovich* [241]) A point-to-set mapping $\Gamma : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is said to be *inner semicontinuous* at $(\overline{z}, \overline{\alpha}) \in \mathbf{gph}\,\Gamma$ provided that, for each sequence $\{z^k\}_{k=1}^{\infty}$ converging to $\overline{z}$ there is a sequence $\{\alpha^k\}_{k=1}^{\infty}$, $\alpha^k \in \Gamma(z^k)$ converging to $\overline{\alpha}$.

Theorem 2.3 (Dempe and Lohse [68]) *Let $(\overline{x}, \overline{y})$ be a global optimal solution of the problem* (1.4). *Let Ψ be inner semicontinuous at $(\overline{x}, \overline{y})$. Then, $(\overline{x}, \overline{y})$ is a local optimal solution of the problem*

$$
\min_{x,y}\{F(x, y) : x \in X, \ (x, y) \in \mathbf{gph}\,\Psi^1\}
\tag{2.7}
$$

with

$$\Psi^1(x) := \operatorname*{Argmin}_{y} \{f(x, y) : g(x, y) \leq 0, \ h(x, y) \leq 0\}$$

with $h : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}$ provided that $h(\overline{x}, \overline{y}) < 0$ and that the function h is continuous.

Proof A point $y \in \Psi(x)$ is an optimal solution of the lower level problem of (2.7) if it is feasible for this problem, $\overline{y} \in \Psi^1(\overline{x})$. Hence, the point $(\overline{x}, \overline{y})$ is feasible for (2.7).

Assume that $(\overline{x}, \overline{y})$ is not a local optimum of problem (2.7). Then, there exists a sequence $\{(x^k, y^k)\}_{k=1}^{\infty}$ converging to $(\overline{x}, \overline{y})$ such that $x^k \in X$, $y^k \in \Psi^1(x^k)$ and $F(x^k, y^k) < F(\overline{x}, \overline{y})$. Note that (x^k, y^k) is feasible for problem (1.4) for large k.

Since Ψ is inner semicontinuous at $(\overline{x}, \overline{y})$ there exists a sequence $\widehat{y}^k \in \Psi(x^k)$ converging to $\overline{y}$. By continuity of the function h, $h(x^k, \widehat{y}^k) < 0$ and $\widehat{y}^k \in \Psi^1(x^k)$. Hence, $f(x^k, \widehat{y}^k) = f(x^k, y^k)$,

$$\Psi^1(x^k) = \{y : g(x, y) \leq 0, \ h(x, y) \leq 0, \ f(x^k, y) = f(x^k, \widehat{y}^k)\}$$
$$\subseteq \{y : g(x, y) \leq 0, \ f(x^k, y) = f(x^k, \widehat{y}^k)\} = \Psi(x^k)$$

and, hence,

$$\min_{y}\{F(x^k, y) : y \in \Psi(x^k)\} \leq \min_{y}\{F(x^k, y) : y \in \Psi^1(x^k)\} \leq F(x^k, y^k) < F(\overline{x}, \overline{y})$$

for sufficiently large k. This contradicts global optimality of $(\overline{x}, \overline{y})$. $\qquad\square$

In the article Dempe and Lohse [68] an example is given which shows that the restrictive assumption of inner semicontinuity of the solution set mapping of the lower level problem is essential.

A similar result to Example 2.2 can be shown if one variable is added in the lower level problem: a global optimal solution can loose global optimality.

Consider the bilevel programming problem

$$\min_{x,y}\{F(x, y) : x \in X, \ (x, y) \in \mathbf{gph}\,\Psi_L\}, \tag{2.8}$$

with a linear lower level problem parameterized in the objective function

$$\Psi_L(x) := \operatorname*{Argmin}_{y} \{x^\top y : Ay = b, \ y \geq 0\}, \tag{2.9}$$

where $X \subseteq \mathbb{R}^n$ is a closed set,

Let $(\overline{x}, \overline{y})$ be a global optimal solution of problem (2.8). Now, add one new variable y_{n+1} to the lower level problem with objective function coefficient x_{n+1} and a new column A_{n+1} in the coefficient matrix of the lower level problem, i.e. replace the lower level problem with

$$\Psi_{NL}(x) := \operatorname*{Argmin}_{y} \{x^\top y + x_{n+1} y_{n+1} : Ay + A_{n+1} y_{n+1} = b, \ y, y_{n+1} \geq 0\}$$

$$\tag{2.10}$$

and investigate the problem

$$\min_{x,y}\{\widetilde{F}(x, x_{n+1}, y, y_{n+1}) : (x, x_{n+1}) \in \widetilde{X}, \ (x, x_{n+1}, y, y_{n+1}) \in \mathbf{gph}\, \Psi_{NL}\}.$$

$$(2.11)$$

Here $\widetilde{X} \subseteq \mathbb{R}^{n+1}$ and $\widetilde{X} \cap \mathbb{R}^n \times \{0\} = X$.

Example 2.3 (*Dempe and Lohse* [68]) Consider the following bilevel programming problem with the lower level problem

$$\Psi_L(x) := \operatorname*{Argmin}_{y} \{x_1 y_1 + x_2 y_2 : y_1 + y_2 \le 2, \ -y_1 + y_2 \le 0, \ y \ge 0\} \quad (2.12)$$

and the upper level problem

$$\min\{(x_1 - 0.5)^2 + (x_2 - 0.5)^2 - 3y_1 - 3y_2 : (x, y) \in \mathbf{gph}\, \Psi_L\}. \qquad (2.13)$$

Then, the unique global optimum is $\overline{x} = (0.5; 0.5)$, $\overline{y} = (1; 1)$ with optimal objective function value -6. Now, adding one variable to the lower level problem

$$\Psi_{NL}(x) := \operatorname*{Argmin}_{y} \{x_1 y_1 + x_2 y_2 + x_3 y_3 : y_1 + y_2 + y_3 \le 2, \ -y_1 + y_2 \le 0, \ y \ge 0\}$$

$$(2.14)$$

and investigating the bilevel optimization problem

$$\min\{(x_1 - 0.5)^2 + (x_2 - 0.5)^2 + x_3^2 - 3y_1 - 3y_2 - 6y_3 : (x, y) \in \mathbf{gph}\, \Psi_{NL}\} \quad (2.15)$$

the point $x = (0.5; 0.5; 0.5)$, $y = (0; 0; 2)$ has objective function value -11.75. Hence, global optimality of $(\overline{x}, \overline{y})$ is destroyed. But, the point $((\overline{x}, 0), (\overline{y}, 0))$ remains feasible and it is a strict local minimum. $\square$

Theorem 2.4 (Dempe and Lohse [68]) *Let $(\overline{x}, \overline{y})$ be a global optimal solution for problem* (2.8) *and assume that the functions F, $\widetilde{F}$ are concave, X, $\widetilde{X}$ are polyhedra. Let*

$$\overline{x}_B^\top B^{-1} A_{n+1} < 0 \text{ for each basic matrix } B \text{ for } \overline{y} \text{ and } \overline{x} \qquad (2.16)$$

and $(\overline{x}, 0)$ be a local minimum of the problem

$$\min\{\widetilde{F}((x, x_{n+1}), (y, 0)) : (x, x_{n+1}) \in \widetilde{X}, \ y \in \Psi_L(\overline{x})\}.$$

Then, the point $((\overline{x}, 0), (\overline{y}, 0))$ is a local optimal solution of problem (2.11).

Proof Assume that $((\overline{x}, 0), (\overline{y}, 0))$ is not a local optimum. Then, there exists a sequence $((x^k, x_{n+1}^k), (y^k, y_{n+1}^k))$ converging to $((\overline{x}, 0), (\overline{y}, 0))$ with

$$F((x^k, x_{n+1}^k), (y^k, y_{n+1}^k)) < F((\overline{x}, 0), (\overline{y}, 0)) \text{ for all } k.$$

Since $((x^k, x^k_{n+1}), (y^k, y^k_{n+1}))$ is feasible for (2.11) and **gph** Ψ_{NL} equals the union of faces of the set (see e.g. Dempe [52])

$$\{(x, y) : x \in \widetilde{X}, \; Ay + A_{n+1} y_{n+1} = b, \; y, y_{n+1} \geq 0\},$$

then, since $((x^k, x^k_{n+1}), (y^k, y^k_{n+1}))$ converges to $((\overline{x}, 0), (\overline{y}, 0))$ there exists, without loss of generality, one facet M of this set with $((x^k, x^k_{n+1}), (y^k, y^k_{n+1})) \in M$ for all k. Moreover, by upper semicontinuity of $\Psi_{NL}(\cdot)$, $((\overline{x}, 0), (\overline{y}, 0)) \in M$. By Schrijver [285] there exists $c \in \mathbb{R}^{n+1}$ such that M equals the set of optimal solutions of the problem

$$\min\{c^\top (y, y_{n+1})^\top : Ay + A_{n+1} y_{n+1} = b, \; y, y_{n+1} \geq 0\}.$$

Since $(\overline{y}, 0) \in M$ there exists a basic matrix for $(\overline{y}, 0)$ and c. Then, the assumptions of the theorem imply that $(\overline{x}, 0) \neq c$ if x_{n+1} is a basic variable in (y^k, y^k_{n+1}) (since this implies that $c_B^\top B^{-1} A_{n+1} - c_{n+1} = 0$ by linear optimization). This implies that there is an open neighborhood V of $(\overline{x}, 0)$ such that $\Psi_{NL}(x, x_{n+1}) \subseteq \{(y, y_{n+1}) : y_{n+1} = 0\}$ for $(x, x_{n+1}) \in V$.

Hence, $y^k_{n+1} = 0$ for sufficiently large k.

By parametric linear optimization, $\Psi_L(x) \subseteq \Psi_L(\overline{x})$ for x sufficiently close to $\overline{x}$. Hence, the assertion follows. $\square$

Similar results are shown in the paper Dempe and Lohse [68] in the case when the lower level problem is a right-hand side perturbed linear optimization problem.

2.2 Optimality Conditions

Consider the bilevel optimization problem

$$\min_{y, b, c}\{F(y) : b \in \mathscr{B}, \; c \in \mathscr{C}, \; y \in \Psi(b, c)\}, \tag{2.17}$$

where

$$\mathscr{B} = \{b : Bb = \widetilde{b}\}, \; \mathscr{C} = \{Cc = \widetilde{c}\}$$

for some matrices B, C of appropriate dimension, $c \in \mathbb{R}^n$, $\widetilde{c} \in \mathbb{R}^q$ and $b \in \mathbb{R}^m$, $\widetilde{b} \in \mathbb{R}^p$. Here, the function $F : \mathbb{R}^n \to \mathbb{R}$ depends only on the optimal solution of the lower level problem. This makes the formulation of optimality conditions, which can be verified in polynomial time, possible.

The mapping $(b, c) \mapsto \Psi(b, c)$ is again the set of optimal solutions of a linear optimization problem:

$$\Psi(b, c) = \operatorname*{Argmin}_{y} \{c^\top y : Ay = b, \; y \geq 0\}.$$

We have $\widehat{y} \in \Psi(b, c)$ if and only if there is a vector $\widehat{z}$ such that $(\widehat{y}, \widehat{z})$ satisfies the following system of equations and inequalities:

$$Ay = b, \ y \geq 0,$$
$$A^\top z \leq c,$$
$$y^\top(A^\top z - c) = 0.$$

Thus, the graph **gph** Ψ of the mapping Ψ equals the projection of the union of faces of a certain polyhedron in $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m \times \mathbb{R}^n$ into the space $\mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^m$. Hence, the tangent (Bouligand) cone

$$C_{\mathcal{M}}(\widehat{u}) := \left\{ \begin{array}{l} d : \exists \{u^k\}_{k=1}^\infty \subset \mathcal{M}, \ \exists \{t_k\}_{k=1}^\infty \subset \mathbb{R}_+ \\[2mm] \text{with} \lim_{k \to \infty} t_k = 0, \ \lim_{k \to \infty} u^k = \widehat{u}, \ d = \lim_{k \to \infty} \frac{u^k - \widehat{u}}{t_k} \end{array} \right\}$$

at a point $(\overline{y}, \overline{b}, \overline{c})$ to the feasible set

$$\mathcal{M} := \{(y, b, c) : b \in \mathcal{B}, \ c \in \mathcal{C}, \ y \in \Psi(b, c)\}$$

equals the union of convex polyhedra, too. Thus, to check local optimality of some feasible point $(\overline{y}, \overline{b}, \overline{c}) \in \mathcal{M}$ for problem (2.17) it is necessary to verify that there is no direction of descent in any one of these convex polyhedra. Unfortunately, the number of these polyhedra cannot be bounded by a polynomial in the number of variables. This can be seen as a reason for $\mathcal{NP}$-hardness of proving local optimality in general bilevel optimization (see Hansen et al. [136] where an exact proof for this result is given).

The following result can be found in the paper [67] by Dempe and Lohse. Let for a moment $\mathcal{B} = \{\overline{b}\}$ reduce to a singleton. Take an arbitrary vertex $\overline{y}$ of the set $\{y : Ay = b, \ y \geq 0\}$. Then, by parametric linear optimization, there exists $\widehat{c}$ such that $\Psi(\overline{b}, c) = \{\overline{y}\}$ for all c sufficiently close to $\widehat{c}$, formally $\forall c \in U(\widehat{c})$ for some open neighborhood $U(\widehat{c})$ of $\widehat{c}$. Hence, if $U(\widehat{c}) \cap \mathcal{C} \neq \emptyset$, there exists $\overline{z}$ satisfying $A^\top \overline{z} \leq \overline{c}, \ \overline{y}^\top(A^\top \overline{z} - \overline{c}) = 0$ for some $\overline{c} \in U(\widehat{c}) \cap \mathcal{C}$ such that $(\overline{y}, \overline{z}, \overline{b}, \overline{c})$ is a local optimal solution of the problem

$$\begin{aligned} F(y) &\to \min_{y,z,b,c} \\ Ay &= b, \ y \geq 0, \\ A^\top z &\leq c, \\ y^\top(A^\top z - c) &= 0 \\ Bb &= \widetilde{b}, \\ Cc &= \widetilde{c}. \end{aligned} \tag{2.18}$$

Fig. 2.3 Definition of local optimality
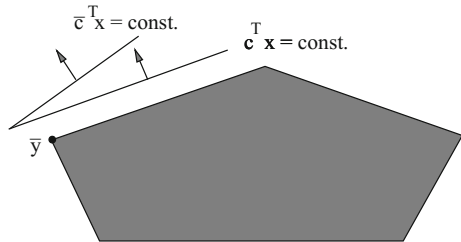
Theorem 2.5 (Dempe and Lohse [67]) *Let $\mathcal{B} = \{\bar{b}\}$, $\{\bar{y}\} = \Psi(\bar{b}, c)$ for all c in an open neighborhood $U(\hat{c})$ of $\hat{c}$ with $U(\hat{c}) \cap \mathcal{C} \neq \emptyset$. Then, $(\bar{y}, \bar{b}, \bar{c}, \bar{z})$ is a locally optimal solution of* (2.18) *for some dual variables $\bar{z}$ and a certain $\bar{c} \in U(\hat{c}) \cap \mathcal{C}$.*

Figure 2.3 can be used to illustrate this fact. The points $\bar{y}$ satisfying the assumptions of Theorem 2.5 are the vertices of the feasible set of the lower level problem given by the dashed area in this figure. Theorem 2.5 implies that each vertex of the set $\{y : Ay = b, \ y \geq 0\}$ is a local optimal solution of problem (2.17) which is not desired. To circumvent this difficulty the definition of a local optimal solution is restricted to variable y only:

Definition 2.2 (*Dempe and Lohse* [67]) A point $\bar{y}$ is a local optimal solution of problem (2.17) if there exists an open neighborhood $U(\bar{y})$ of $\bar{y}$ such that $F(y) \geq F(\bar{y})$ for all (y, b, c) with $b \in \mathcal{B}$, $c \in \mathcal{C}$ and $y \in U(\bar{y}) \cap \Psi(b, c)$.

To derive a necessary optimality condition for problem (2.17) according to this definition, a formula for a tangent cone to its feasible set depending only on y is needed. Let $(\bar{y}, \bar{z}, \bar{b}, \bar{c})$ be a feasible solution for problem (2.18) and define the index sets

$$
\begin{aligned}
I(\bar{y}) &= \{i : \bar{y}_i = 0\}, \\
I(z, c) &= \{i : (A^\top z - c)_i > 0\}, \\
\mathscr{I}(\bar{y}) &= \{I(z, c) : A^\top z \geq c, \ (A^\top z - c)_i = 0 \ \forall i \notin I(\bar{y}), \ c \in \mathcal{C}\} \\
I^0(\bar{y}) &= \bigcap_{I \in \mathscr{I}(\bar{y})} I.
\end{aligned}
$$

Remark 2.1 If an index set I belongs to the family $\mathscr{I}(\bar{y})$ then $I^0(\bar{y}) \subseteq I \subseteq I(\bar{y})$.

This remark and also the following one are obvious consequences of the definitions of the above sets.

Remark 2.2 We have $j \in I(\bar{y}) \setminus I^0(\bar{y})$, if and only if the system

$$
\begin{aligned}
(A^\top z - c)_i &= 0 \ \forall i \notin I(\bar{y}) \\
(A^\top z - c)_j &= 0 \\
(A^\top z - c)_i &\geq 0 \ \forall i \in I(\bar{y}) \setminus \{j\} \\
Cc &= \tilde{c}
\end{aligned}
$$

has a solution. Furthermore $I^0(\bar{y})$ is an element of $\mathscr{I}(\bar{y})$ if and only if the system

$$(A^\top z - c)_i = 0 \,\forall i \notin I^0(\bar{y})$$
$$(A^\top z - c)_i \geq 0 \,\forall i \in I^0(\bar{y})$$
$$Cc = \tilde{c}$$

has a solution.

This result makes an efficient computation of the set $I^0(\bar{y})$ possible.

Now, it turns out that the dual feasible solution $\bar{z}$ for the lower level problem as well as the objective function coefficients $\bar{c}$ are not necessary for solving problem (2.17), it is only necessary to consider possible index sets $I \in \mathscr{I}(\bar{y})$.

Theorem 2.6 (Dempe and Lohse [67]) $\bar{y}$ *is a local optimum for* (2.17) *if and only if $\bar{y}$ is a* (global) *optimal solution for all problems* (A_I):

$$F(y) \to \min_{y,b}$$
$$Ay = b$$
$$y \geq 0$$
$$y_i = 0 \quad \forall i \in I$$
$$Bb = \tilde{b}$$

with $I \in \mathscr{I}(\bar{y})$.

Proof Let $\bar{y}$ be a local optimal solution of (2.17) and assume that there is a set $I \in \mathscr{I}(\bar{y})$ with $\bar{y}$ being not optimal for (A_I). Then there exists a sequence $\{y^k\}_{k=1}^\infty$ of feasible solutions of (A_I) with $\lim_{k\to\infty} y^k = \bar{y}$ and $F(y^k) < F(\bar{y})$ for all k. Consequently $\bar{y}$ can not be local optimal for (2.17) since $I \in \mathscr{I}(\bar{y})$ implies that all y^k are also feasible for (2.18).

Conversely, let $\bar{y}$ be an optimal solution for all problems (A_I) and assume that there is a sequence $\{y^k\}_{k=1}^\infty$ of feasible points of (2.17) with $\lim_{k\to\infty} y^k = \bar{y}$ and $F(y^k) < F(\bar{y})$ for all k. For k sufficiently large the elements of this sequence satisfy the condition $y_i^k > 0$ for all $i \notin I(\bar{y})$ and due to the feasibility of y^k for (2.17) there are sets $I \in \mathscr{I}(\bar{y})$ such that y^k is feasible for problem (A_I). Because $\mathscr{I}(\bar{y})$ consists only of a finite number of sets, there is a subsequence $\{y^{k_j}\}_{j\in\mathbb{N}}$ where y^{k_j} are all feasible for a fixed problem (A_I). So we get a contradiction to the optimality of $\bar{y}$ for this problem (A_I). $\qquad\square$

Using the set I as a new variable in problem (A_I), the following problem is obtained which is equivalent to problem (2.18) by Theorem 2.6:

$$F(y) \to \min_{y,b,I}$$
$$Ay = b$$

$$y \geq 0 \tag{2.19}$$
$$y_i = 0 \quad \forall i \in I$$
$$Bb = \tilde{b}$$
$$I \in \mathscr{I}(\overline{y})$$

The following tangent cone can be used to express the feasible set of problem (A_I) near a feasible point $\overline{y}$ for a fixed set $I \in \mathscr{I}(\overline{y})$:

$$T_I(\overline{y}) = \{d| \, \exists r: \ Ad = r, \ Br = 0, \ d_i \geq 0, \ \forall i \in I(\overline{y}) \setminus I, \ d_i = 0, \ \forall i \in I\}$$

Using Theorem 2.6 a necessary optimality condition is derived:

Corollary 2.1 *If $\overline{y}$ is a local optimal solution of problem (2.17), and F is directionally differentiable then $F'(\overline{y}; d) \geq 0$ for all $d \in T(\overline{y}) := \bigcup_{I \in \mathscr{I}(\overline{y})} T_I(\overline{y})$.*

Since $d \in \operatorname{conv} T(\overline{y})$ is equal to a convex linear combination of elements in $T(\overline{y})$, $\nabla F(\overline{y})d < 0$ for some $d \in \operatorname{conv} T(\overline{y})$ only if $\nabla F(\overline{y})\overline{d} < 0$ for a certain $\overline{d} \in T(\overline{y})$. This leads to the necessary optimality condition

$$\nabla F(\overline{y})d \geq 0 \, \forall \, d \in \operatorname{conv} T(\overline{y})$$

provided that the objective function F is differentiable.

Consider the relaxed problem to (2.19):

$$F(y) \to \min_{y,b}$$
$$Ay = b$$
$$y_i \geq 0 \quad \forall i \notin I^0(\overline{y}) \tag{2.20}$$
$$y_i = 0 \quad \forall i \in I^0(\overline{y})$$
$$b \in \mathscr{B}$$

and the tangent cone

$$T_R(\overline{y}) = \{d: Ad = r, \ Br = 0, \ d_i \geq 0, \ i \in I(\overline{y}) \setminus I^0(\overline{y}), \ d_i = 0, \ i \in I^0(\overline{y})\}$$

to the feasible set of this problem at the point $\overline{y}$ again relative to y only.

Due to $I^0(\overline{y}) \subseteq I$ for all $I \in \mathscr{I}(y^0)$ we derive

$$\operatorname{conv} T(\overline{y}) = \operatorname{cone} T(\overline{y}) \subseteq T_R(\overline{y}), \tag{2.21}$$

where cone S denotes the conical hull of the set S, i.e. the set of all linear combinations of elements in S with nonnegative coefficients. Let span S denote the set of all linear combinations of elements in S.

Definition 2.3 The point $\overline{y}$ is said to satisfy the *full rank condition* (FRC), if

$$\text{span}(\{A_i \,:\, i \notin I(\overline{y})\}) = \mathbb{R}^m, \tag{2.22}$$

where A_i denotes the ith column of the matrix A.

Example 2.4 All non degenerated vertices of $Ay = b$, $y \ge 0$ satisfy the full rank condition.

This condition allows us now to establish equality between the cones above.

Theorem 2.7 (Dempe and Lohse [67]) *Let (FRC) be satisfied at the point $\overline{y}$. Then equality holds in* (2.21).

Proof Let $\overline{d}$ be an arbitrary element of $T_R(\overline{y})$, that means there is a $\overline{r}$ with $A\overline{d} = \overline{r}$, $B\overline{r} = 0$, $\overline{d}_i \ge 0$, $i \in I(\overline{y}) \setminus I^0(\overline{y})$, $\overline{d}_i = 0$, $i \in I^0(\overline{y})$. Without loss of generality assume $I(\overline{y}) = \{1, 2, \ldots, l\}$.

We consider the following linear systems (S_1)

$$Ad = \overline{r}$$
$$d_1 = \overline{d}_1$$
$$d_i = 0, \; i \in I(\overline{y}) \setminus \{1\}$$

and (S_j)

$$Ad = 0$$
$$d_j = \overline{d}_j$$
$$d_i = 0, \; i \in I(\overline{y}) \setminus \{j\}$$

for $j = 2, \ldots, l$. These systems have all feasible solutions since $\overline{y}$ satisfies the full rank condition.

Let $d^1, \ldots, d^l$ be (arbitrary) solutions of the systems (S_j) and define the direction $d = \sum_{j=1}^{l} d^j$. Then, $d_i = \overline{d}_i$ for $i \in I(\overline{y})$ as well as $Ad = A\overline{d} = \overline{r}$.

If $d = \overline{d}$ we are done since $d \in \text{cone}\, T(\overline{y}) = \text{conv}\, T(\overline{y})$. Assume that $d \neq \overline{d}$. (Fig. 2.4).

Define $\widehat{d}^1 := d^1 + \overline{d} - d$. Since d^1 is feasible for (S_1) and $d_i = \overline{d}_i$ for $i = 1, \ldots, k$ as well as $Ad = A\overline{d} = \overline{r}$ we obtain $\widehat{d}_i^1 = 0$ for all $i = 2, \ldots, k$ and

$$A\widehat{d}^1 = A(d^1 + \overline{d} - d) = \overline{r} + \overline{r} - \overline{r} = \overline{r}.$$

Hence $\widehat{d}^1$ is also a solution of (S_1).

Thus, $\widehat{d}^1 + \sum\limits_{j=2}^{l} d^j = \overline{d} - d + \sum\limits_{j=1}^{l} d^j = \overline{d}.$

Due to the definition of I and of the tangent cones $T(\overline{y})$ and $T_R(\overline{y})$ the conclusion $T_R(\overline{y}) \subseteq T(\overline{y})$ follows. $\square$

Fig. 2.4 Illustration (taken from Dempe and Lohse [67]) of the proof of Theorem 2.7
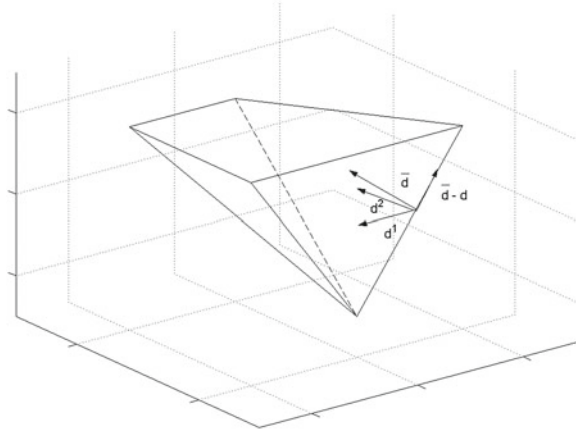
Due to Remark 2.2 at most n systems of linear (in-) equalities need to be investigated to compute the index set $I^0(\bar{y})$. Hence, by Theorem 2.7, verification of local optimality of a feasible point of problem (2.17) is possible in polynomial time.

2.3 Solution Algorithms

2.3.1 Computation of a Local Optimal Solution

We consider the linear bilevel optimization problem (2.3), (2.4). y^0 is an optimal solution of the lower level problem iff there exists u such that

$$C^\top u = d, \; u \le 0, \; u^\top (Cy - x) = 0.$$

Let the rank of the matrix C be equal to m: $r(C) = m$. An optimal solution of problem (2.4) can be found at a vertex of the feasible set, which means that there are m linearly independent rows C_i, $i = 1, \ldots, m$ (without loss of generality, these are the first m rows) of the matrix C such that

$$C_i y = x_i, \; i = 1, \ldots, m$$

and

$$C_i y \ge x_i, \; i = m + 1, \ldots, q.$$

Then, if the first m rows of C compose a matrix D, N is build up of the last $q - m$ rows, $x = (x_D \; x_N)$ is accordingly decomposed, we obtain $C = (D \; N)^\top$ and

$y = D^{\top-1}x_D$ is a solution of problem (2.4). A solution u^0 of the dual problem is given by $u = (u_D \ u_N)^\top$ with $u_D = D^{\top-1}d$, $u_N = 0$. Then,

$$u_D \leq 0, \ Dy = x_D, \ Ny \leq x_N, \ u_N = 0, \ Ny \leq x.$$

For D with $D^{\top-1}d \leq 0$ the set

$$\mathscr{R}_D = \{x : Dy = x_D, \ Ny \leq x_N \text{ for some } y \in \mathbb{R}^m\}$$

is the so-called *region of stability* for the matrix D. It consists of all parameter vectors x for which an optimal solution of the primal problem (2.4) can be computed using the matrix D.

For other values of x, the basic matrix D consists of other rows of C. This, together with basic matrices for the upper level constraints $Ax \leq c$ can be used to describe an algorithm enumerating all these basic matrices to find a global optimum of the bilevel optimization problem. For this, Theorem 2.1 is of course needed. Many algorithms for solving the linear bilevel optimization problem suggested in the last century used this approach (see e.g. Bard [10]; Candler and Townsley [40]; Bard and Falk [11]).

The idea of the following algorithm can be found in Dempe [49]:

Descent algorithm for the linear bilevel problem:
Input: Linear bilevel optimization problem (2.3).
Output: A local optimal solution.

 Algorithm: Start   Select an optimal basic solution (x^1, y) of
   the problem
$$\min\{a^\top x + b^\top y : Ax \leq c, \ Cy \leq x\}.$$

   Compute an optimal basic solution y^1 of the problem
   (2.4) for $x = x^1$. Set$k := 1$.
Step 1   Select a basic matrix D for y^k, compute the
   region of stability $\mathscr{R}_D$ and solve the problem

$$\min_x\{a^\top x + b^\top y : x = (x_D \ x_N)^\top, \ y = D^{\top-1}x_D, \ x \in \mathscr{R}_D\}.$$

   Let $(\widehat{x}, D^{\top-1}\widehat{x}_D)$ be an optimal solution.
Step 2   Set $x^{k+1} = \widehat{x}$ and compute an optimal basic solution
   y^{k+1} of the problem

$$\min_y\{b^\top y : y \in \Psi(x^{k+1})\}$$

   Stop if the optimal solution has not changed:

$$(x^{k+1}, y^{k+1}) = (x^k, y^k).$$

   Otherwise goto Step 1.

This algorithm computes a local optimal solution since either one of the problems in Steps 1 or 2 of the algorithm would lead to a better solution. For a rigorous proof, the interested reader is referred to the original paper Dempe [49].

2.3.2 A Global Algorithm

Consider the linear bilevel optimization problem

$$\min_{x,y}\{a^\top x + b^\top y : Ax \leq c, \ (x, y) \in \textbf{gph} \ \Psi^1\} \tag{2.23}$$

with

$$\Psi^1(x) = \underset{y}{\text{Argmin}} \ \{x^\top y : By \leq d\} \tag{2.24}$$

and the optimization problem

$$\min_{x,y}\{a^\top x + b^\top y : Ax \leq c, \ By \leq d, \ x^\top y \leq \varphi^1(x)\}, \tag{2.25}$$

where

$$\varphi^1(x) = \min_{y}\{x^\top y : By \leq d\}$$

is the optimal value function of problem (2.24). Both problems (2.23) and (2.25) are fully equivalent. It follows from parametric linear optimization (see e.g. Dempe and Schreier [77] and Beer [15]) that the function $\varphi(\cdot)$ is an in general nondifferentiable, concave, piecewise affine-linear and Lipschitz continuous function. It is equal to

$$\varphi^1(x) = \min\{x^\top y^1, x^\top y^2, \ldots, x^\top y^p\},$$

where $\{y^1, y^2, \ldots, y^p\}$ is the set of vertices of the convex polyhedron $\{y : By \leq d\}$. Strictly speaking, formula (2.25) is correct only on the set of all x for which $|\varphi^1(x)| \leq \infty$. If $\varphi^1(\widehat{x}) = \widehat{x}^\top y^k$, then $y^k \in \partial^{Cl}(\widehat{x})$ is an element of the generalized derivative in the sense of Clarke [see (3.10)]. Using the results from convex analysis (see Clarke [42] and Rockafellar [272]) we have

$$\varphi^1(x) \leq \varphi^1(\widehat{x}) + \widehat{y}^\top(x - \widehat{x}) \ \forall \ x, \ \forall \ \widehat{y} \in \partial^{cl}\varphi^1(\widehat{x}).$$

Hence,

$$\{(x, y) : \ Ax \leq c, \ By \leq d, \ x^\top y \leq \varphi^1(x)\}$$
$$\subseteq \{(x, y) : \ Ax \leq c, \ By \leq d, \ x^\top y \leq \widehat{y}^\top x\} \tag{2.26}$$

for $\widehat{y} \in \Psi^1(\widehat{x})$. This implies that the problem

$$\min_{x,y}\{a^\top x + b^\top y : Ax \leq c, \ By \leq d, \ x^\top y \leq \widehat{y}^\top x\} \qquad (2.27)$$

cannot have a worse objective function value than problem (2.25).

A solution algorithm for the linear bilevel optimization problem (2.23), (2.24)

Algorithm: Start   Select x^0 satisfying $Ax^0 \leq c$, compute $y^0 \in \Psi(x^0)$.
  Set $k := 1$ and $\mathscr{Y} := \{y^0\}$.
Step 1   Solve problem

$$\min_{x,y}\{a^\top x + b^\top y : Ax \leq c, \ By \leq d, \ x^\top y \leq \widehat{y}^\top x \ \forall \ \widehat{y} \in \mathscr{Y}\} \qquad (2.28)$$

  globally and let (x^k, y^k) be an optimal solution.
Step 2   If $y^k \in \Psi(x^k)$, stop: (x^k, y^k) is a global optimal solu-
  tion of problem (2.23), (2.24). Otherwise, set

$$\mathscr{Y} := \mathscr{Y} \cup \{\widehat{y}^k\}$$

  where $\widehat{y}^k \in \Psi(x^k)$ and go to Step 1.

Theorem 2.8 (Dempe and Franke [60]) *Let* $\{(x, y) : Ax \leq c, \ By \leq d\}$ *be bounded.*
The above algorithm computes a global optimal solution of the linear bilevel opti-
mization problem (2.23), (2.24).

Proof If the algorithm stops in Step 2, the last point (x^k, y^k) is feasible for the linear
bilevel optimization problem. Hence, due to (2.26) it is also globally optimal.

   Let $\{(x^k, y^k)\}_{k=1}^\infty$ be an infinite sequence computed by the algorithm. Since the
set $\mathscr{Y}$ is increasing, more and more constraints are added to problem (2.28) implying
that the sequence of its optimal objective function values is nondecreasing. On the
other hand, it is bounded from above by e.g. $a^\top x^1 + b^\top b^1$. Hence, this sequence
converges to, say, v^*. Let, without loss of generality (x^*, y^*) be a limit point of the
sequence $\{x^k, y^k\}_{k=1}^\infty$. Continuity of the function $\varphi(\cdot)$ leads to

$$\lim_{k\to\infty} \varphi(x^k) = \lim_{k\to\infty} x^{k\,\top}\widehat{y}^k = x^{*\,\top}\widehat{y}^*,$$

where $\widehat{y}^*$ is again without loss of generality a limit point of the sequence $\{\widehat{y}^k\}_{k=1}^\infty$.
Then, we have
$$x^{k\,\top}y^k \leq x^{k\,\top}\widehat{y}^{k-1}$$

by the formulae in the algorithm. Hence, by convergence of the sequences, we derive

$$x^{*\,\top}y^* \leq \varphi(x^*) = x^{*\,\top}\widehat{y}^*.$$

Consequently, the point (x^*, y^*) is feasible and, thus, also globally optimal. $\square$

It is clear that the algorithm can be implemented such that it stops after a finite number of iterations if the feasible set of the lower level problem is compact. The reason for this is that the feasible set has only a finite number of vertices which correspond to the vertices of the generalized derivative of the function $\varphi^1(\cdot)$.

One difficulty in realizing the above algorithm is that we need to solve the optimization problem (2.28) globally in each iteration. This is a nonconvex optimization problem and usually solution algorithms compute only stationary or local optimal solutions for such problems. Hence, it is perhaps more suitable to try to compute a local optimal solution of problem (2.23) respectively its equivalent problem (2.25). Often this is related to the use of a sufficient optimality condition. Since (2.25) is a nonsmooth optimization problem we can use a sufficient optimality condition of first order demanding that the directional derivative of the objective function is not negative on a suitable tangent cone to the feasible set.

Let M_R be the feasible set of problem (2.28) for some set $\mathscr{Y} \subseteq \{y : By \leq d\}$. Then, the Bouligand (or tangent) cone to M_R at some point (x^*, y^*) reads as

$$C_{M_R}(x^*, y^*) = \{d \in \mathbb{R}^{2n} : \exists \{(x^k, y^k)\}_{k=1}^{\infty} \subseteq M_R, \ \exists \{t_k\}_{k=1}^{\infty} \subseteq \mathbb{R}_+ \backslash \{0\}$$

$$\text{satisfying} \lim_{k \to \infty} (x^k, y^k) = (x^*, y^*), \ \lim_{k \to \infty} t_k = 0 \text{ and}$$

$$d = \lim_{k \to \infty} \frac{1}{t_k}((x^k, y^k) - (x^*, y^*))\}.$$

The local algorithm solving problem (2.23) is identical with the algorithm on the previous page with the only distinction that problem (2.28) in Step 1 of the algorithm is solved locally.

Theorem 2.9 (Dempe and Franke [60]) *Let the set $\{(x, y) : Ax \leq c, \ By \leq d\}$ be nonempty and compact. Assume that there is $\gamma > 0$ such that $(a^\top \ b^\top)d \geq \gamma$ for all $d \in C_{M_R}(x^k, y^k)$ and all sufficiently large k. Then, all accumulation points of the sequences computed using the local algorithm are locally optimal for problem (2.23).*

Proof The existence of accumulation points as well as their feasibility for problem (2.23) follow analogously to the proof of Theorem 2.8. Let

$$M_B := \{(x, y) : Ax \leq c, \ By \leq d, \ x^\top y \leq \varphi(x)\}$$

denote the feasible set of problem (2.23). Then, $M_B \subseteq M_R$ and C_{M_R} is a Bouligand cone to a convex set. Hence, for $(x, y) \in M_B$ sufficiently close to (x^k, y^k) we have $d^k := ((x, y) - (x^k, y^k))/\|(x, y) - (x^k, y^k)\| \in C_{M_R}(x^k, y^k)$ and $(a^\top \ b^\top)d^k \geq \gamma$ for sufficiently large k. The Bouligand cone to M_B is defined analogously to the Bouligand cone to M_R.

Let $(\overline{x}, \overline{y})$ be an arbitrary accumulation point of the sequence $\{(x^k, y^k)\}_{k=1}^{\infty}$ computed by the local algorithm. Assume that $(\overline{x}, \overline{y})$ is not a local optimal

solution. Then there exists a sequence $\{\bar{x}^k, \bar{y}^k\}_{k=1}^{\infty} \subset M_B$ converging to $(\bar{x}, \bar{y})$ with $a^\top \bar{x}^k + b^\top \bar{y}^k < a^\top \bar{x} + b^\top \bar{y}$ for all k. Then, by definition, without loss of generality

$$\lim_{k \to \infty} \frac{(\bar{x}^k, \bar{y}^k) - (\bar{x}, \bar{y})}{\|(\bar{x}^k, \bar{y}^k) - (\bar{x}, \bar{y})\|} = \bar{d} \in C_{M_B}(\bar{x}, \bar{y}) \subseteq C_{M_R}(\bar{x}, \bar{y})$$

and $(a^\top \ b^\top)\bar{d} \leq 0$. On the other hand, $\bar{d}$ is a limit point of a sequence $\{d^k\}_{k=1}^{\infty}$ with $d^k \in C_{M_R}(x^k, y^k)$ with $(a^\top \ b^\top)d^k \geq \gamma$ for all k by assumption.

This contradicts the assumption, thus proving the Theorem.   □

The following example is presented in Dempe and Franke [60] to illustrate the algorithm.

Example 2.5 Consider the bilevel optimization problem

$$\min_{x,y} 2x_1 + x_2 + 2y_1 - y_2$$
$$\text{s.t. } |x_1| \leq 1$$
$$-1 \leq x_2 \leq -0.75$$
$$y \in \Psi(x) := \operatorname*{Argmin}_{y} \{x^\top y : -2y_1 + y_2 \leq 0, \ y_1 \leq 2, \ 0 \leq y_2 \leq 2\}.$$

The concave optimal value function $\varphi(x)$ of the lower level problem reads

$$\varphi(x) = \begin{cases} 2x_1 + 2x_2 & \text{if } x_1 \in [-1, 0] \, , \, x_2 \in [-1, -0.75] \\ x_1 + 2x_2 & \text{if } x_1 \in (\, 0, 1 \,] \, , \, x_2 \in [-1, -0.75] \end{cases}$$

The values of the upper level objective function over the feasible set are

$$a^\top x + b^\top y = \begin{cases} -2 + 2x_1 + x_2 & \text{if } x_1 \in [-1, 0] \, , \, x_2 \in [-1, -0.75] \\ 2x_1 + x_2 & \text{if } x_1 \in (\, 0, 1 \,] \, , \, x_2 \in [-1, -0.75] \end{cases}$$

with the optimal solution at $x = (-1, -1)$ and the optimal function value 5. For $\mathcal{Y} \subseteq \{y : By \leq d\}$, the problem (2.28) is

$$\min_{x,y} 2x_1 + x_2 + 2y_1 - y_2$$
$$\text{s.t.} \quad |x_1| \leq 1$$
$$-1 \leq x_2 \leq -0.75$$
$$x^\top y \leq \min_{z \in \mathcal{Y}} x^\top z$$
$$-2y_1 + y_2 \leq 0$$
$$y_1 \leq 2$$
$$0 \leq y_2 \leq 2.$$

Now, the above algorithm works as follows.

Start $\mathscr{Y} := \{(0,0)\}$, $k := 1$.

Step 1 The optimal solution of (2.28) is $(x_1^1, x_2^1, y_1^1, y_2^1) = (-1, -1, 2, 0)$.

Step 2 The lower level with $(x_1^1, x_2^1) = (-1, -1)$ leads to $(z_1^1, z_2^1) = (2, 2)$ which is added to $\mathscr{Y}$. Go to Step 1.

Step 1 The optimal solution of (2.28) is $(x_1^2, x_2^2, y_1^2, y_2^2) = (-1, -1, 2, 2)$.

Step 2 The lower level with $(x_1^2, x_2^2) = (-1, -1)$ leads to $(z_1^2, z_2^2) = (2, 2)$ which coincides with the solution of Step 1, hence the algorithm terminates with the optimal solution $(x_1^2, x_2^2, y_1^2, y_2^2) = (-1, -1, 2, 2)$. $\qquad\square$