Computational Optimization and Applications 6, 227–249 (1996)

On an Algorithm Solving Two-Level Programming Problems with Nonunique Lower Level Solutions

S. DEMPE AND H. SCHMIDT
Department of Mathematics, Technical University Chemnitz-Zwickau, PSF 964, 09009 Chemnitz, Germany

Received May 11, 1993; Revised February 1, 1995

Abstract. In the paper, an algorithm is presented for solving two-level programming problems. This algorithm combines a direction finding problem with a regularization of the lower level problem. The upper level objective function is included in the regularzation to yield uniqueness of the follower's solution set. This is possible if the problem functions are convex and the upper level objective function has a positive definite Hessian. The computation of a direction of descent and of the step size is discussed in more detail. Afterwards the convergence proof is given.

Last but not least some remarks and examples describing the difficulty of the inclusion of upper-level constraints also depending on the variables of the lower level are added.

Keywords:

1. Introduction

Two-level programming problems occur in a large variety of different situations in economics (cf., e.g., certain problems in the principal—agency theory [20]), in technical branches (as for instance various design problems [13, 17]) as well as in chemistry and physics [18]. The last problem is related to the computation of the equilibrium state of chemical reactions at higher temperatures [23]. Although the chemists technically are not able to observe in situ the single reactions, they described the final point of the system by a convex programming problem. Its optimal solution gives the amounts of each substance of the equilibrium state. Now, if we are forced to construct a chemical equilibrium having additional properties such as, e.g., a large amount of one single substance or a lack of another, we have to consider this system as being a parametric one, the parameters of which are determined by the input into the chemical reactor. Now, for realizing the goal just mentioned, we have to select values for these parameters such that the optimal solution of the resulting problem has the desired property. This leads to a two-level optimization problem described as follows:

Let, in a slightly more general setting, the parametric optimization problem for computing, e.g., the equilibrium state in chemical reactions be given as

$$\Psi(y) := \operatorname*{Argmin}_{x \in R^n} \{f(x, y) \mid g(x, y) \leq 0,\ h(x, y) = 0\}. \tag{1.1}$$

Then, the problem of evaluating a parameter vector y such that an element $x \in \Psi(y)$ has a certain desired property reads as

$$\begin{aligned} &F(x, y) \to \text{“}\min_{y \in R^m}\text{”} \\ &G(y) \le 0,\ x \in \Psi(y), \end{aligned} \tag{1.2}$$

where $F, f, g_i, h_j : R^n \times R^m \to R,\ i = 1, \dots, p,\ j = 1, \dots, q,\ G : R^m \to R^l, g(x, y) = (g_1(x, y), \dots, g_p(x, y))^{\mathrm{T}},\ h(x, y) = (h_1(x, y), \dots, h_q(x, y))^{\mathrm{T}},\ G(y) = (G_1(y), \dots, G_l(y))^{\mathrm{T}}$, and $F(\cdot, \cdot)$ gives a mathematical description of the global goal to be achieved. If $\Psi(y)$ does not reduce to a singleton for each y, problem (1.2) is not well-posed. Namely, the nonuniqueness of the lower-level solution $x(y) \in \Psi(y)$ for at least one parameter vector y causes that $x(y)$ cannot be substituted in the upper-level problem (1.2). Thus, for evaluating the objective function value $F(x(y), y)$ at some point y we have to give a finer rule for selecting $x \in \Psi(y)$. One such rule (i.e., one auxiliary problem for treating problem (1.1), (1.2)) consists in solving the hierarchical problem **(P)**:

$$\begin{aligned} &F(x(y), y) \to \min_{y} \\ &G(y) \le 0, \end{aligned} \tag{1.3}$$

where

$$x(y) \in \operatorname*{Argmin}_{x \in R^n} \{F(x, y) \mid x \in \Psi(y)\}. \tag{1.4}$$

Although there are other interesting attempts for attacking two-level programming problems (cf., e.g., [14, 15, 18, 19]), we prefer solving problem **(P)** which will be explained in Section 2 (cf. also [11]). The bibliography of Vicente/Calamai [24] summarizes a large part of the theory of bilevel programming problems especially concerning algorithms.

The content of this paper is as follows: In Section 2 we describe some auxiliary results as well as a certain regularization approach for solving problem **(P)** which will be outlined and discussed in Sections 3 and 4. Some remarks concerning the constraints of the type $G(x, y) \le 0$ instead of $G(y) \le 0$ will be given in Section 5.

2. The regularization approach

2.1. Preliminaries

The mathematical programming problem described by **(P)** is in fact a three-level problem. For computational reasons, one tries to reduce the problem to a two-level one by means of approximation techniques. Therefore we prefer an approximation achieving one important aim, namely that by regularizing problem (1.1) the solution of problem (1.4) will be approximated. Having this aim in mind, consider the following problem:

$$\Psi^{\alpha}(y) := \operatorname*{Argmin}_{x \in R^n} \{f(x, y) + \alpha F(x, y) \mid g(x, y) \le 0,\ h(x, y) = 0\}. \tag{2.1}$$

This problem is a regularization for (1.1) if, e.g., the following conditions are satisfied:

(A1) All problem functions are sufficiently smooth: $F,\ f,\ g_i,\ h_j \in C^2(R^n \times R^m, R)$, $i = 1, \dots, p,\ j = 1, \dots, q,\ G_i \in C^2(R^m, R),\ i = 1, \dots, l$.
(A2) Certain convexity assumptions are satisfied: $f(\cdot, y),\ g_i(\cdot, y),\ i = 1, \dots, p$, are convex on R^n, $h_j(\cdot, y),\ j = 1, \dots, q$, are affine, $\nabla^2_{xx} F(\cdot, y)$ is positive definite on R^n for each fixed y.

Denote by

$$M(y) := \{x \mid g(x, y) \le 0,\ h(x, y) = 0\} \tag{2.2}$$

the feasible set of problem (1.1) for each y. Then, for $\alpha > 0$, the objective function of problem (2.1) is strongly convex. Thus, for each y such that $M(y) \neq \emptyset$, problem (2.1) has a unique optimal solution. The following theorem shows that problem (2.1) realizes the aim stated above. Thereinafter, the following notion is used:

A point-to-set mapping $\Sigma : R^s \to 2^{R^t}$ is said to be *upper semicontinuous* at $w \in R^s$ if and only if for each open set $V \subseteq R^t$ satisfying $\Sigma(w) \subset V$ there exists an open neighbourhood $U \subset R^s$ of w such that $\Sigma(w') \subset V$ for each $w' \in U$.

Theorem 2.1. [1] *Consider the point-to-set mapping* $\Gamma : R^m \times R_+ \to 2^{R^n}$ *defined by* $\Gamma(y, \alpha) := \Psi^\alpha(y)$. *Let the assumptions* (A1), (A2),
(A3) $M := \{(x, y) \mid G(y) \le 0,\ x \in M(y)\}$ *is nonempty and bounded, and the following generalized Slater's condition at* y^0 *be satisfied:*
(A4) *There exists* $\hat{x}$ *satisfying* $g(\hat{x}, y^0) < 0,\ h(\hat{x}, y^0) = 0$ *and the gradients* $\{\nabla_x h_j(\hat{x}, y^0) : j = 1, \dots, q\}$ *are linearly independent.*

Then,
1. $\Gamma(\cdot\,, \cdot)$ *is upper semicontinuous at* $(y^0, \alpha^0),\ y^0 \in R^m,\ \alpha^0 \ge 0$.
2. *For each* $(y, \alpha) \in R^m \times R$ *with* $\alpha > 0$ *and each optimal solution* $x(y, \alpha)$ *of problem* (2.1) *we have the following two inequalities*

$$F(x(y, \alpha), y) \le \min_x\{F(x, y) \mid x \in \Psi(y)\}, \tag{2.3}$$

$$f(x(y, \alpha), y) \ge \varphi(y) := \min_x\{f(x, y) \mid x \in M(y)\}. \tag{2.4}$$

3. *Under the conditions of part* 2,

$$\lim_{\alpha \searrow 0} \Gamma(y, \alpha) \subseteq \operatorname*{Argmin}_{x \in R^n}\{F(x, y) \mid x \in \Psi(y)\}.$$

2.2. *Solvability*

In the next theorem, we show the existence of optimal solutions for problem **(P)**.

Theorem 2.2. *Consider problem* **(P)** *under the conditions* (A1)–(A4) (*for each* y). *Then, problem* **(P)** *has an optimal solution.*

Remark 2.3. In general, the solvability of a problem $\min_z\{\kappa(z) \mid \eta(z) \leq 0\}$, where $\kappa(z)$ is defined as an optimal value function of a certain parametric optimization problem $\kappa(z) := \min_r\{\xi(r, z) \mid r \in \mathcal{M}(z)\}$ does not imply the solvability of the problem

$$\min_{r,z}\{\xi(r, z) \mid r \in \mathcal{M}(z),\ \eta(z) \leq 0\}$$

(cf., e.g., [16]). In the case of problem **(P)**, we obtain the following:

1. The feasible set of the problem

$$\min_{x,y}\{F(x, y) \mid G(y) \leq 0, x \in \Psi(y)\} \tag{2.5}$$

   is bounded by (A3) and closed by [1, Theorems 3.1.1, 3.1.5, 4.3.3].
2. By (A1), $F(\cdot, \cdot)$ is continuous.

Thus, problem (2.5) has a (global) optimal solution which coincides with a global optimal solution of problem **(P)**.

Remark 2.4. Due to (A3), the regularization

$$\min_{x,y}\{F(x, y) \mid G(y) \leq 0,\ x \in \Psi^\alpha(y)\} \tag{2.6}$$

has an optimal solution $(x(\alpha), y(\alpha))$ for each $\alpha > 0$ which is contained in the set M. Thus, for $\alpha \searrow 0$, the sequence $\{(x(\alpha), y(\alpha))\}_{\alpha>0}$ has only bounded accumulation points and the function values $F(x(\alpha), y(\alpha))$ are also bounded.

3. The algorithm

3.1. *Motivation*

Motivated by Theorem 2.1, we will try to compute an optimal solution for problem **(P)** by searching for an accumulation point of the sequence of optimal solutions of the problems (2.6) for $\alpha \searrow 0$. By strong convexity of $f(\cdot, y) + \alpha F(\cdot, y)$ on R^n for each fixed y, problem (2.6) itself is equivalent to the following nondifferentiable optimization problem with an implicitly defined objective function:

$$\min_y\{\Phi_\alpha(y) \mid G(y) \leq 0\}, \tag{3.1}$$

where $\Phi_\alpha(y) := F(x_\alpha(y), y),\ x_\alpha(y) \in \Psi^\alpha(y)\ \forall\ y$. If we assume that the function $\Phi_\alpha(\cdot)$ possesses a directional derivative

$$\Phi'_\alpha(y; r) := \lim_{t\searrow 0} t^{-1}[\Phi_\alpha(y + tr) - \Phi_\alpha(y)]$$

for each direction r and each point y, then the problem (3.1) can be solved using the following prototype of an algorithm:

Descent algorithm:

Step 1. Select y^0 solving $G(y^0) \leq 0$, set $k := 0$, choose α^0 and $\varepsilon, \delta \in (0, 1)$.
Step 2. Compute a direction r^k, $\|r^k\| \leq 1$, satisfying

$$\Phi'_\alpha(y^k; r^k) \leq s^k,$$
$$\nabla_y G_i(y^k) r^k \leq -G_i(y^k) + s^k, \ i = 1, \ldots, l,$$

and $s^k < 0$.
Step 3. Choose a step-size t^k such that

$$\Phi_\alpha(y^k + t^k r^k) \leq \Phi_\alpha(y^k) + \varepsilon t^k s^k,$$
$$G(y^k + t^k r^k) \leq 0.$$

Step 4. Set $y^{k+1} := y^k + t^k r^k$, $k := k + 1$.
Step 5. If a stopping criterion is satisfied:

if α is sufficiently small, then stop
else set $\alpha := \delta\alpha$ and compute $x_\alpha(y^k)$.

Goto Step 2.

To inspire life into the algorithm a more detailed description of Steps 2 and 3 is necessary. We do not intend to give a more concrete rule for determining the values of ε and δ, since, on the one hand, the choice is dependent on the concrete nature of the functions f, F, G, g, h and, on the other hand, this presupposes a comprehensive computational experience which we do not have yet.

3.2. *A direction of descent*

3.2.1. The directional derivative of the optimal solution of the lower level. If the optimal solution of problem (2.1) is directionally differentiable then the directional derivative of the function $\Phi_\alpha(\cdot)$ is guaranteed to exist because of the smoothness of the function $F(\cdot, \cdot)$. The following theorem gives sufficient conditions therefore.

Let

$$L_\alpha(x, y, \lambda, \mu) := f(x, y) + \alpha F(x, y) + \lambda^\top g(x, y) + \mu^\top h(x, y)$$

be the Lagrangian of problem (2.1). If assumptions (A1)–(A4) are satisfied at $y = y^0$ then, a feasible point x of (2.1) is optimal if and only if the set of Lagrange multiplier vectors

$$\Lambda_\alpha(x, y^0) := \{(\lambda, \mu) \mid \nabla_x L_\alpha(x, y^0, \lambda, \mu) = 0, \ \lambda \geq 0, \ \lambda^\top g(x, y^0) = 0\}$$

is not empty. Moreover, in this case, $\Lambda_\alpha(x, y^0)$ is a bounded polyhedron [7], i.e., it is equal to the convex hull of its finite set $E\Lambda_\alpha(x, y^0)$ of vertices.

Lemma 3.1 [4, 22]. *Consider problem* (2.1) *and let the assumptions* (A1)–(A4) *be satisfied at* $y = y^0$, $x_\alpha(y^0) \in \Psi^\alpha(y^0)$. *Then, the vector function* $x_\alpha(\cdot)$, *defined by the unique optimal solutions of problems* (2.1) *for fixed* $\alpha > 0$, *is directionally differentiable at* y^0.

Actually, Lemma 3.1 is a corollary of the results in [22] and is independently due to [4]. As an immediate result of this lemma the directional derivative of the function $\Phi_\alpha(\cdot)$ exists. But, for computing a direction of descent which could successfully be used in Steps 2 and 3 of the algorithm, we need a rule for the calculation of the directional derivative of the function $x_\alpha(\cdot)$. One such rule is given in the following theorem using the additional assumption

(A5) For each nonempty set $K \subseteq I(x_\alpha(y^0), y^0) := \{j : g_j(x_\alpha(y^0), y^0) = 0\}$, all matrices

$$\begin{pmatrix} \nabla_x g_i(x_\alpha(y), y), & i \in K \\ \nabla_x h_j(x_\alpha(y), y), & j = 1, \dots, q \end{pmatrix},$$

composed by the gradients with respect to x are locally of constant rank around y^0.

Lemma 3.2 [4]. *If the assumptions* (A1)–(A5) *are satisfied for problem* (2.1) *at* $y = y^0$, $x_\alpha(y^0) \in \Psi^\alpha(y^0)$ *then, for each direction* r^0 *there exists a vertex* $(\lambda^0, \mu^0) \in E\Lambda_\alpha(x_\alpha(y^0), y^0)$ *such that:* $x'_\alpha(y^0; r^0)$ *is the unique optimal solution of the following convex quadratic programming problem* $Q_\alpha(y^0, r^0, \lambda^0, \mu^0)$:

$$\frac{1}{2} d^\top \nabla^2_{xx} L_\alpha(x_\alpha(y^0), y^0, \lambda^0, \mu^0) d + d^\top \nabla^2_{xy} L_\alpha(x_\alpha(y^0), y^0, \lambda^0, \mu^0) r^0 \to \min_d$$

$$\nabla_x g_i(x_\alpha(y^0), y^0) d + \nabla_y g_i(x_\alpha(y^0), y^0) r^0 \begin{cases} = 0, & \text{if } \lambda_i^0 > 0 \\ \le 0, & \text{if } i \in I(x_\alpha(y^0), y^0) \end{cases}$$

$$\nabla_x h_j(x_\alpha(y^0), y^0) d + \nabla_y h_j(x_\alpha(y^0), y^0) r^0 = 0, \; j = 1, \dots, q.$$

Remark 3.3. This lemma was given earlier by [21] and [12] but they imposed the linear independence constraint qualification w.r.t. x. Vice versa (A5) and (A6) below are weaker than this assumption, since they are satisfied for instance for linear inequality constraints of an arbitrary number with right hand side perturbations [5].

*3.2.2. **The direction finding problem.*** Now, for computing the direction of descent r^k used in the descent algorithm, we can use the following problem, in which the value of s^k is minimized. The additional constraints of this problem are given by the necessary and sufficient optimality conditions of first order for the quadratic convex programming problem $Q_\alpha(y^0, r, \lambda, \mu)$ used for the computation of $x'_\alpha(y^0; r)$, which are modified by the application of an active-set strategy for dropping the complementarity constraints.

Problem $D_\alpha(y^0, \lambda, \mu, K)$:

$$s \to \min_{d,r,\nu,\omega,s} \tag{3.2}$$

$$\Phi'_\alpha(y^0; r) := \nabla_x F(x_\alpha(y^0), y^0)d + \nabla_y F(x_\alpha(y^0), y^0)r \le s \tag{3.3}$$

$$\nabla_y G_i(y^0)r \le -G_i(y^0) + s,\ i = 1, \dots, l \tag{3.4}$$

$$\nabla^2_{xx} L_\alpha(x_\alpha(y^0), y^0, \lambda, \mu)d + \nabla^2_{xy} L_\alpha(x_\alpha(y^0), y^0, \lambda, \mu)r + \nabla_x^{\mathsf{T}} g(x_\alpha(y^0), y^0)\nu + \nabla_x^{\mathsf{T}} h(x_\alpha(y^0), y^0)\omega = 0 \tag{3.5}$$

$$\nabla_x g_i(x_\alpha(y^0), y^0)d + \nabla_y g_i(x_\alpha(y^0), y^0)r \begin{cases} = 0, & i \in K \\ \le 0, & i \in I(x_\alpha(y^0), y^0) \end{cases} \tag{3.6}$$

$$\nabla_x h_j(x_\alpha(y^0), y^0)d + \nabla_y h_j(x_\alpha(y^0), y^0)r = 0,\ j = 1, \dots, q \tag{3.7}$$

$$\nu_i \ge 0,\ i \in K \setminus \{j : \lambda_j > 0\}, \qquad \nu_i = 0,\ i \notin K,\ \|r\| \le 1. \tag{3.8}$$

Unfortunately, the optimality of $(d^0, r^0, \nu^0, \omega^0, s^0)$ for $D_\alpha(y^0, \lambda, \mu, K)$ is not sufficient for the fact that d^0 is equal to $x'_\alpha(y^0; r^0)$. This is due to the possibly discontinuous behaviour of $x'_\alpha(y^0; r)$ with respect to r if $p > 0$. To overcome this difficulty, we use the following theorem:

Theorem 3.4 [4]. *Let for problem* (2.1), *in addition to* (A1)–(A5), *also*
(A6) *The gradients* $\{\nabla g_i(x_\alpha(y^0), y^0), i \in I(x_\alpha(y^0), y^0)\} \cup \{\nabla h_j(x_\alpha(y^0), y^0), j = 1, \dots, q\}$ *with respect to both x and y are linearly independent*

be satisfied, then,

$$\{d \mid \exists \{r^k\}_{k=1}^{\infty} \text{ converging to } r^0 \text{ such that } \lim_{k\to\infty} x'_\alpha(y^0; r^k) = d\}$$
$$= \bigcup_{(\lambda,\mu)\in E\Lambda_\alpha(x_\alpha(y^0), y^0)} \{d \mid d \text{ is an optimal solution of problem } Q_\alpha(y^0, r^0, \lambda, \mu)\}.$$

Note that assumptions (A4)–(A6) are weaker than the linear independence constraint qualification with respect to x which can easily be seen in the very simple

Example.

$$\min_x\{(x-1)^2 : x + y \le 0,\ x - y \le 0\} \text{ at } y = 0.$$

Consequently, for computing a direction of descent needed in the above algorithm, in the worst case, we have to solve problems $D_\alpha(y^0, \lambda, \mu, K)$ for all vertices of $\Lambda_\alpha(x_\alpha(y^0), y^0)$ and all sets K satisfying $\{j : \lambda_j > 0\} \subseteq K \subseteq I(x_\alpha(y^0), y^0)$. Let $\theta_\alpha(y^0, \lambda, \mu, K)$ denote the optimal value of problem $D_\alpha(y^0, \lambda, \mu, K)$. Then, we also have the following necessary optimality condition for problem (2.1):

Theorem 3.5 [5]. *If* $(x_\alpha(y^0), y^0)$ *is an optimal solution of problem* (2.1) *and if assumptions* (A1)–(A6) *are satisfied at* $y = y^0$, *then*

$$\begin{aligned}&\min\{\theta_\alpha(y^0, \lambda, \mu, K) \mid (\lambda, \mu) \in E\Lambda_\alpha(x_\alpha(y^0), y^0),\\ &\qquad \{j : \lambda_j > 0\} \subseteq K \subseteq I(x_\alpha(y^0), y^0)\} \geq 0.\end{aligned}$$

*3.2.3. **A modified direction finding problem.*** Define the optimal value function of problem (2.1) as

$$\varphi_\alpha(y) := \min_x\{f(x, y) + \alpha F(x, y) \mid x \in M(y)\}.$$

Then, by [6, Theorem 2.3.2.], or [10], the directional derivative of this function exists provided that assumptions (A1)–(A4) are valid. This directional derivative can be computed by solving the following linear programming problem:

$$\varphi'_\alpha(y^0; r) = \max_{\lambda,\mu}\{\nabla_y L_\alpha(x_\alpha(y^0), y^0, \lambda, \mu)r \mid (\lambda, \mu) \in E\Lambda_\alpha(x_\alpha(y^0), y^0)\}. \tag{3.9}$$

By linear programming

$$\begin{aligned}&\max\{\nabla_y L(x(y^0), y^0, \lambda, \mu)r : (\lambda, \mu) \in \Lambda(x(y^0), y^0)\}\\ &\quad = \max\{\nabla_y L(x(y^0), y^0, \lambda, \mu)r : (\lambda, \mu) \in E\Lambda(x(y^0), y^0)\}.\end{aligned}$$

Denote the set of optimal solutions of problem (3.9) by

$$S_\alpha(y^0; r) := \operatorname*{Argmax}_{\lambda,\mu}\{\nabla_y L_\alpha(x_\alpha(y^0), y^0, \lambda, \mu)r \mid (\lambda, \mu) \in \Lambda_\alpha(x_\alpha(y^0), y^0)\}. \tag{3.10}$$

For our convergence proof in Section 4 we need modified direction finding problems which will be developed in the sequel.

The necessary (and sufficient) optimality conditions for (3.10) are given as: There exists a vector $d \in R^n$ satisfying

$$\begin{aligned}&\nabla_x g_i(x_\alpha(y^0), y^0)d + \nabla_y g_i(x_\alpha(y^0), y^0)r \begin{cases} = 0, & \text{if } \lambda_i > 0\\ \leq 0, & \text{if } i \in I(x_\alpha(y^0), y^0)\end{cases}\\ &\nabla_x h_j(x_\alpha(y^0), y^0)d + \nabla_y h_j(x_\alpha(y^0), y^0)r = 0,\ j = 1, \ldots, q.\end{aligned}$$

These are exactly the conditions guaranteeing nonemptiness of the feasible set of problem $Q_\alpha(y^0, r, \lambda, \mu)$. Thus,

$$\begin{aligned}&\left\{d \mid \exists\, \{r^k\}_{k=1}^\infty \text{ converging to } r \text{ such that } \lim_{k\to\infty} x'_\alpha(y^0; r^k) = d\right\}\\ &\quad = \bigcup_{(\lambda,\mu)\in ES_\alpha(y^0;r)} \{d \mid d \text{ is an optimal solution of problem } Q_\alpha(y^0, r, \lambda, \mu)\},\end{aligned}$$

where $ES_\alpha(y^0; r)$ is the vertex set of $S_\alpha(y^0; r)$. It should be mentioned that for proving Theorem 3.4 for each optimal solution d of problem $Q_\alpha(y^0, r, \lambda^0, \mu^0)$ the existence of a sequence $\{r^k\}_{k=1}^{\infty}$ converging to r has been shown such that $S_\alpha(y^0; r^k) = \{(\lambda^0, \mu^0)\}$ for each k. Then, $x'_\alpha(y^0; r^k)$ is equal to the unique optimal solution of problem $Q_\alpha(y^0, r^k, \lambda^0, \mu^0)$. Consider the inverse mapping to $S_\alpha(y^0; r)$ with respect to r:

$$T_\alpha(\lambda^0, \mu^0) := \{r : \{(\lambda^0, \mu^0)\} = S_\alpha(y^0; r)\}.$$

Let cl A denote the closure of a set A. Then, if problem $Q_\alpha(y^0, r, \lambda^0, \mu^0)$ has an optimal solution, then we have $r \in \text{cl } T_\alpha(\lambda^0, \mu^0)$ and for $r' \in T_\alpha(\lambda^0, \mu^0)$, the directional derivative $x'(y^0; r')$ is computed by $Q_\alpha(y^0, r', \lambda^0, \mu^0)$.

Lemma 3.6. *The set $T_\alpha(\lambda^0, \mu^0)$ is open, its closure is a convex cone with apex at zero.*

Let, for $r \in \text{cl } T_\alpha(\lambda^0, \mu^0)$, the index set of active constraints in the problem $Q_\alpha(y^0, r, \lambda^0, \mu^0)$ be denoted by $K(r)$ and set

$$T_\alpha^K(\lambda^0, \mu^0) := \{r \in T_\alpha(\lambda^0, \mu^0) : K(r) = K\}$$

for some set $K \in \mathcal{N} := \{K : \{i : \lambda_i^0 > 0\} \subseteq K \subseteq I(x_\alpha(y^0), y^0)\}$. If the index set of active constraints is fixed, then the necessary (and sufficient) conditions for optimality of the problem $Q_\alpha(y^0, r, \lambda^0, \mu^0)$ are linear in d, r, v, ω. Hence, $T_\alpha^K(\lambda^0, \mu^0)$ is also a convex set with cl $T_\alpha^K(\lambda^0, \mu^0)$ being a convex cone with apex at zero and we have

$$\text{cl } T_\alpha(\lambda^0, \mu^0) = \text{cl} \bigcup_{K \in \mathcal{N}} T_\alpha^K(\lambda^0, \mu^0).$$

Lemma 3.7. *Let* (A6) *be satisfied and let $T_\alpha^K(\lambda^0, \mu^0)$ have a nonempty interior. Then, the gradients $\{\nabla_x g_i(x_\alpha(y^0), y^0) : i \in K\} \cup \{\nabla_x h_j(x_\alpha(y^0), y^0) : j = 1, \dots, q\}$ are linearly independent.*

Let $\mathcal{M} = \{K \in \mathcal{N} : T_\alpha^K(\lambda^0, \mu^0)$ has a nonempty interior$\}$.

Lemma 3.8. *Let assumptions* (A1)–(A6) *be satisfied. Then,*

$$\text{cl } T_\alpha(\lambda^0, \mu^0) = cl \bigcup_{K \in \mathcal{M}} T_\alpha^K(\lambda^0, \mu^0).$$

In the sequel we will use the following assumption

(A7) Let $y^0 \in Y$, $(\lambda^0, \mu^0) \in E\Lambda_\alpha(x_\alpha(y), y)$ and let $K \in \mathcal{M}$. Then, strict complementarity slackness is satisfied for the optimal solution $x'_\alpha(y^0; r)$ of problem $Q_\alpha(y^0, r, \lambda^0, \mu^0)$ for each $r \in \text{int } T_\alpha^K(\lambda^0, \mu^0)$.

Now, if the assumptions (A1)–(A7) are satisfied, and if $(d^0, r^0, v^0, \omega^0, s^0)$ is an optimal solution of the problem $D_\alpha(y^0, \lambda^0, \mu^0, K)$ with $s^0 < 0$ then, $r^0 \in \text{cl } T_\alpha(\lambda^0, \mu^0)$ and, for each sufficiently small $\varepsilon > 0$ there exists r' such that $\|r^0 - r'\| \le \varepsilon$ and r' is an interior point of some set $T_\alpha^K(\lambda^0, \mu^0)$, $K \in \mathcal{M}$ by Lemma 3.8. Then, problem $Q_\alpha(y^0, r', \lambda^0, \mu^0)$ has a unique multiplier by Lemma 3.7 and by (A7) we have

$$\nabla_x g_i(x_\alpha(y^0), y^0)x'(y^0; r') + \nabla_y g_i(x_\alpha(y^0), y^0)r' < 0, \ \nu_i = 0$$

for each $i \in I(x_\alpha(y^0), y^0) \setminus K$ and

$$\nabla_x g_i(x_\alpha(y^0), y^0)x'(y^0; r') + \nabla_y g_i(x_\alpha(y^0), y^0)r' = 0, \ \nu_i > 0$$

for each $i \in K \setminus \{j : \lambda_j^0 > 0\}$. Thus, using the positive homogeneity of d with respect to r as well as a similar property for ν and ω, we obtain

Corollary 3.9. *Let the assumptions* (A1)–(A7) *be satisfied for problem* (2.1) *at* $y = y^0$, $x_\alpha(y^0) \in \Psi^\alpha(y^0)$, $\alpha > 0$. *Then, for computing a direction of descent we have to choose* $(\lambda^0, \mu^0) \in E\Lambda_\alpha(x_\alpha(y^0), y^0)$ *and a set* K *satisfying*

1. $\{j : \lambda_j^0 > 0\} \subseteq K \subseteq I(x_\alpha(y^0), y^0)$ *and*
2. *the gradients* $\{\nabla_x g_i(x_\alpha(y^0), y^0) : i \in K\} \cup \{\nabla_x h_j(x_\alpha(y^0), y^0) : j = 1, \ldots, q\}$ *are linearly independent*

such that the optimal value of the problem $D_\alpha^0(y^0, \lambda^0, \mu^0, K)$ *is less than zero, where problem* $D_\alpha^0(y^0, \lambda^0, \mu^0, K)$ *is given as follows*:

$$\begin{aligned}
& s \to \min_{d,r,\nu,\omega,s} \\
& \Phi'_\alpha(y^0; r) := \nabla_x F(x_\alpha(y^0), y^0)d + \nabla_y F(x_\alpha(y^0), y^0)r \le s \\
& \qquad \nabla_y G_i(y^0)r \le -G_i(y^0) + s, \ i = 1, \ldots, l \\
& \nabla^2_{xx} L_\alpha(x_\alpha(y^0), y^0, \lambda^0, \mu^0)d + \nabla^2_{xy} L_\alpha(x_\alpha(y^0), y^0, \lambda^0, \mu^0)r \\
& \qquad + \nabla_x^{\top} g(x_\alpha(y^0), y^0)\nu + \nabla_x^{\top} h(x_\alpha(y^0), y^0)\omega = 0 \\
& \nabla_x g_i(x_\alpha(y^0), y^0)d + \nabla_y g_i(x_\alpha(y^0), y^0)r \begin{cases} = 0, & i \in K \\ \le -g_i(x_\alpha(y^0), y^0) + s, & i \notin K \end{cases} \\
& \nabla_x h_j(x_\alpha(y^0), y^0)d + \nabla_y h_j(x_\alpha(y^0), y^0)r = 0, \ j = 1, \ldots, q \\
& \lambda_i + \nu_i + s \ge 0, \ i \in K, \qquad \nu_i = 0, \ i \notin K, \ \|r\| \le 1.
\end{aligned}$$

Consequently, the direction finding problem to be solved in Step 2 has some similarities with the minimization problem in Theorem 3.5: In Step 2 we are searching for a set K satisfying 1. and 2. of the above corollary and a vector (λ^0, μ^0) such that the optimal value of problem $D_\alpha^0(y^0, \lambda^0, \mu^0, K)$ is sufficiently small. It is indeed possible to detect such a set if and only if the optimality condition in Theorem 3.5 is not satisfied. In fact, it would also be acceptable to search for the minimal optimal function value of problem $D_\alpha^0(y^0, \lambda^0, \mu^0, K)$ subject to $(\lambda^0, \mu^0) \in E\Lambda_\alpha(x_\alpha(y^0), y^0)$ and all sets K with the above properties. But this is a combinatorial problem whose solution is time-consuming. The above strategy could be used to minimize this effort.

It should also be noticed that assumption (A7) does in general not imply that a strict complementarity slackness assumption is satisfied for the lower level problem. If this condition should be violated then there exist an index $k \in I(x_\alpha(y^0), y^0)$ and a set K with the above properties such that

$$\nabla_x g_k(x_\alpha(y^0), y^0)x'(y^0; r) + \nabla_y g_k(x_\alpha(y^0), y^0)r = 0$$

and $\nu_k = 0$ for each $r \in T_\alpha^K(\lambda^0, \mu^0)$, where the set $T_\alpha^K(\lambda^0, \mu^0)$ has a nonempty interior.

4. Convergence of the algorithm and computation of the step-size

4.1. Convergence for fixed α

In the prototype of the descent algorithm we used a kind of Armijo step-size rule, i.e., we proposed the selection of the largest number t^k in $\{\rho, \rho^2, \rho^3, \rho^4, \ldots\}$, where $\rho \in (0, 1)$, such that

$$\Phi_\alpha(y^k + t^k r^k) \le \Phi_\alpha(y^k) + \varepsilon t^k s^k, \ \varepsilon \in (0, 1), \ \text{ and } G(y^k + t^k r^k) \le 0. \tag{4.1}$$

We first show convergence of the algorithm for fixed $\alpha > 0$ thereby clarifying one possible choice of the stopping criterion in Step 5 of the algorithm. Clearly, the algorithm computes a sequence $\{(x_\alpha(y^k), y^k)\}_{k=1}^\infty$ having an accumulation point $(x_\alpha(y^0), y^0)$ by (A3) and Theorem 2.1.

Lemma 4.1. *Let $\alpha > 0$ and let $\{y^k, \lambda^k, \mu^k\}_{k=1}^\infty$ be a convergent sequence such that besides* (A1)–(A7) *the following assumptions are satisfied*:
1. *there exists a set K satisfying the conditions* 1. *and* 2. *of Corollary* 3.9
2. *the optimal values s^k of the problems $D_\alpha^0(y^k, \lambda^k, \mu^k, K)$ are less than zero and bounded away from zero, i.e., there exists $s < 0$ such that $s^k \le s$, $k = 1, 2, \ldots$.*

Then the following properties are valid:
1. *Let $\hat{x}_\alpha(y)$ be the optimal solution and $\hat{\lambda}(y)$, $\hat{\mu}(y)$ be the corresponding multipliers of the enlarged problem*

$$\min_x\{f(x, y) + \alpha F(x, y) : g_i(x, y) = 0, i \in K, \ h_j(x, y) = 0, \ j = 1, \ldots, q\}. \tag{4.2}$$

Then, for each k there exists $\tau^k > 0$ such that

$$\hat{x}_\alpha(y^k + tr^k) = x_\alpha(y^k + tr^k), \tag{4.3}$$
$$\hat{\lambda}_i(y^k + tr^k) = \lambda_i(y^k + tr^k), \ i \in K \tag{4.4}$$
$$\lambda_i(y^k + tr^k) = 0, \ i \notin K \tag{4.5}$$
$$\hat{\mu}(y^k + tr^k) = \mu(y^k + tr^k) \tag{4.6}$$

for each $0 \le t \le \tau^k$, $k = 1, 2, \ldots$, *and some multipliers*

$$(\lambda(y^k + tr^k), \mu(y^k + tr^k)) \in \Lambda_\alpha(x_\alpha(y^k + tr^k), y^k + tr^k).$$

2. *Zero cannot be an accumulation point of the sequence* $\{\tau^k\}_{k=1}^\infty$, *i.e., there exists* $\tau > 0$ *such that* $\tau^k > \tau$, $k = 1, 2, \ldots$.

Proof: The first part of the proof consists of the following: if r^k is determined by $D_\alpha^0(y^k, \lambda^k, \mu^k, K)$ then the conditions $g_i(\hat{x}_\alpha(y), y) \le 0$, $i \notin K$ and $\hat{\lambda}_i(y) \ge 0$, $i \in K$ are satisfied for sufficiently small $t > 0$. Afterwards in the second part we show that τ^k can be calculated as stated in the theorem.

By the properties of the set K, Lemma 3.7 and (A2), the optimal solution $\hat{x}_\alpha(\cdot)$ and the optimal multipliers $\hat{\lambda}(\cdot)$, $\hat{\mu}(\cdot)$ of problem (4.2) are continuously differentiable at $y = y^k$ [6].

If for the solution $\hat{x}_\alpha(y)$ and the multipliers $(\hat{\lambda}(y), \hat{\mu}(y))$ of problem (4.2) additionally the inequalities $g_i(\hat{x}_\alpha(y), y) \le 0$, $i \notin K$ and $\hat{\lambda}_i(y) \ge 0, i \in K$ are satisfied then the triple $(\hat{x}_\alpha(y), \hat{\lambda}(y), \hat{\mu}(y))$ with $\hat{\lambda}_i(y) = 0$, $i \notin K$, gives the optimal solution of problem (2.1) together with optimal multipliers since the necessary conditions for optimality for the convex programming problem (2.1) are satisfied in this case, too. This will be shown for $y = y^k + tr^k$, $t \in [0, \tau^k]$ $(\tau^k > 0)$ in what follows. It is obvious, that $\hat{x}_\alpha(y^k) = x_\alpha(y^k)$ and $(\hat{\lambda}(y^k), \hat{\mu}(y^k)) \in \Lambda_\alpha(x_\alpha(y^k), y^k)$. By the choice of r^k as part of an optimal solution $(d^k, r^k, v^k, \omega^k, s^k)$ of problem $D_\alpha^0(y^k, \lambda^k, \mu^k, K)$, we have $r^k \in T_\alpha^K(\lambda^k, \mu^k)$. Thus, $d^k = x'_\alpha(y^k; r^k)$. Moreover, we can drop the constraints with index $i \notin K$ in problem $Q_\alpha(y^k, r^k, \lambda^k, \mu^k)$ since they are inactive by

$$\nabla_x g_i(x_\alpha(y^k), y^k)x'(y^k; r^k) + \nabla_y g_i(x_\alpha(y^k), y^k)r^k \le s^k < 0. \tag{4.7}$$

But then, problem $Q_\alpha(y^k, r^k, \lambda^k, \mu^k)$ turns out to be equivalent to the convex quadratic optimization problem used for computing the directional derivative of $\hat{x}_\alpha(\cdot)$. This shows that

$$d^k = x'_\alpha(y^k; r^k) = \nabla_y \hat{x}(y^k)r^k,\ v_i^k = \nabla_y \hat{\lambda}_i(y^k)r^k,\ i \in K,\ \omega^k = \nabla_y \hat{\mu}(y^k)r^k. \tag{4.8}$$

Thus, by $\hat{\lambda}_i(y^k) \ge 0$, $i \in I(x_\alpha(y^k), y^k)$ and $v_i^k > 0, i \in K \setminus \{j : \hat{\lambda}_i(y^k) > 0\}$ we have

$$\hat{\lambda}_i(y^k + tr^k) = \hat{\lambda}_i(y^k) + t\nabla_y \hat{\lambda}_i(y^k)r^k + o(t) > 0,\ i \in K$$

and

$$\begin{aligned} g_i(\hat{x}_\alpha(y^k + tr^k), y^k + tr^k) &= g_i(\hat{x}_\alpha(y^k), y^k) \\ &+ t(\nabla_x g_i(\hat{x}_\alpha(y^k), y^k)\nabla_y \hat{x}(y^k)r^k + \nabla_y g_i(\hat{x}_\alpha(y^k), y^k)r^k) + o(t) < 0,\ i \notin K \end{aligned}$$

for sufficiently small $t > 0$ by (4.7). Hence, for each k there exists $\tau^k > 0$ such that $\hat{x}_\alpha(y^k + tr^k)$ and $\hat{\lambda}(y^k + tr^k)$ indeed satisfy $g_i(\hat{x}_\alpha(y^k + tr^k), y^k + tr^k) \le 0$, $i \notin K$ and $\hat{\lambda}_i(y^k + tr^k) \ge 0$, $i \in K$, $\forall t \in [0, \tau^k]$. As a result the optimal solutions of (2.1) and (4.2) coincide for $t \in [0, \tau^k]$.

By (A2), (A5) and the results in [8, Chapter 3] the bound τ^k is determined by

$$\tau^k = \min\{\min\{t > 0 : \exists i \notin K \text{ with } g_i(\hat{x}_\alpha(y^k + tr^k), y^k + tr^k) = 0\},\\ \min\{t > 0 : \exists i \in K \text{ with } \hat{\lambda}_i(y^k + tr^k) = 0\}\}. \quad (4.9)$$

This implies Part 1.

Let (y^0, λ^0, μ^0) be the limit point of the sequence $\{(y^k, \lambda^k, \mu^k)\}_{k=1}^{\infty}$. Let r^0 be an accumulation point of the bounded sequence $\{r^k\}$. By (A2), (A5) and well-known results in parametric optimization [6], the optimal solution $\hat{x}_\alpha(\cdot)$ as well as the optimal multipliers $\hat{\lambda}(\cdot)$ and $\hat{\mu}(\cdot)$ are continuously differentiable at $y = y^0$. Hence, by (4.8) we have $\lim \nabla_y \hat{x}_\alpha(y^k) = \nabla_y \hat{x}_\alpha(y^0)$, $\lim \nabla_y \hat{\lambda}(y^k) = \nabla_y \hat{\lambda}(y^0)$, $\lim \nabla_y \hat{\mu}(y^k) = \nabla_y \hat{\mu}(y^0)$. Thus, by (4.8) and by our assumptions, $(\nabla_y \hat{x}_\alpha(y^0) r^0, r^0, \nabla_y \hat{\lambda}(y^0) r^0, \nabla_y \hat{\mu}(y^0) r^0, s^0)$ is an optimal solution of problem $D_\alpha^0(y^0, \lambda^0, \mu^0, K)$ with $s^0 \le s$.

Now, arguing by contradiction, let $\lim_{k\to\infty} \tau^k = 0$. Let $i(k)$ be defined such that the minimum in (4.9) is attained for $i = i(k)$. Take an infinite subsequence $\{k \in \mathcal{K}\}$ such that $i^* = i(k)$ and $0 < \tau^k \le 1$ for all $k \in \mathcal{K}$. Then, $i^* \in K$ or $i^* \notin K$ for all $k \in \mathcal{K}$. Hence, either one of the following two conditions is satisfied for all $k \in \mathcal{K}$:

$$g_i(\hat{x}_\alpha(y^k + \tau^k r^k), y^k + \tau^k r^k) = 0 \quad \text{if } i = i^* \notin K \quad (4.10)$$

or

$$\hat{\lambda}_i(y^k + \tau^k r^k) = 0 \quad \text{if } i \in K. \quad (4.11)$$

In dependence of the occurring case we have

$$\begin{aligned} 0 &= g_i(\hat{x}_\alpha(y^k + \tau^k r^k), y^k + \tau^k r^k) = g_i(\hat{x}_\alpha(y^k), y^k) \\ &\quad + \tau^k \left(\nabla_x g_i(\hat{x}_\alpha(y^k), y^k) \nabla_y \hat{x}_\alpha(y^k) r^k + \nabla_y g_i(\hat{x}_\alpha(y^k), y^k) r^k + \frac{o(\tau^k)}{\tau^k} \right) \\ &\le g_i(\hat{x}_\alpha(y^k), y^k) + \nabla_x g_i(\hat{x}_\alpha(y^k), y^k) \nabla_y \hat{x}_\alpha(y^k) r^k + \nabla_y g_i(\hat{x}_\alpha(y^k), y^k) r^k \\ &\quad + \frac{o(\tau^k)}{\tau^k} \le s^k + \frac{o(\tau^k)}{\tau^k} \end{aligned} \quad (4.12)$$

by $g_i(\hat{x}_\alpha(y^k), y^k) \le 0$, $1 \ge \tau^k > 0$, i.e., $\nabla_x g_i(\hat{x}_\alpha(y^k), y^k) \nabla_y \hat{x}_\alpha(y^k) r^k + \nabla_y g_i(\hat{x}_\alpha(y^k), y^k) r^k + \frac{o(\tau^k)}{\tau^k} \ge 0$ if $i = i^* \notin K$ or

$$\begin{aligned} 0 &= \hat{\lambda}_i(y^k + \tau^k r^k) = \hat{\lambda}_i(y^k) + \tau^k \left(\nu_i^k + \frac{o(\tau^k)}{\tau^k} \right) \\ &\ge \lambda_i(y^k) + \nu_i^k + \frac{o(\tau^k)}{\tau^k} \ge -s^k + \frac{o(\tau^k)}{\tau^k} \end{aligned} \quad (4.13)$$

by $\lambda_i(y^k) \ge 0$, and $1 \ge \tau^k > 0$, i.e., $\nu_i^k + \frac{o(\tau^k)}{\tau^k} \le 0$ if $i = i^* \in K$. Both conditions contradict $\frac{o(\tau^k)}{\tau^k} \to 0$ and $s^k \to s^0 < 0$. This proves the Lemma. □

It should be noticed that the functions $\hat{x}(\cdot)$, $\hat{\lambda}(\cdot)$, $\hat{\mu}(\cdot)$ are continuously differentiable on $[0, \tau^k]$ provided that the assumptions of the above lemma are satisfied.

Lemma 4.2. *Let the assumptions* (A1)–(A7) *be satisfied on the set M. Then, for the sequence $\{(x_\alpha(y^k), y^k, r^k, d^k, \lambda^k, \mu^k, \nu^k, \omega^k, s^k, K^k)\}_{k=1}^{\infty}$ computed by the algorithm for fixed $\alpha > 0$, the sequence $\{s^k\}_{k=1}^{\infty}$ has zero as the only accumulation point.*

Proof: Arguing by contradiction, let $s < 0$ be a number such that

$$\Phi'_\alpha(y^k; r^k) \le s^k \le s \text{ and } \lim_{k\to\infty} s^k = s \tag{4.14}$$

for all sufficiently large k and at least one subsequence of the sequence of points computed by the algorithm (which, without loss of generality, can be identified with the sequence itself). Now we can consider two cases:
First let us assume that $\underline{\lim}_{k\to\infty} t^k > 0$. Then

$$\Phi_\alpha(y^k + t^k r^k) \le \Phi_\alpha(y^o) + \sum_{j=1}^{k} \varepsilon t^j s^j \le \Phi_\alpha(y^o) + \varepsilon s \sum_{j=1}^{k} t^j.$$

can be calculated and $\lim_{k\to\infty} \Phi_\alpha(y^k + t^k r^k) = -\infty$, which is an obvious contradiction.

Secondly, let $\underline{\lim}_{k\to\infty} t^k = 0$. Assume without loss of generality, that $\{t^k\}$ itself converges to zero. But then we have $t^k = \rho^{j_k}$, $k = 1, 2, \dots,$ and $\lim_{k\to\infty} j_k = \infty$ and at least one of the inequalities in Step 3 of the algorithm must be violated by ρ^{j_k-1}, i.e., at least one of the following inequalities must be true:

$$\frac{\Phi_\alpha(y^k + \rho^{j_k-1} r^k) - \Phi_\alpha(y^k)}{\rho^{j_k-1}} > \varepsilon s^k \tag{4.15}$$

$$G_i(y^k + \rho^{j_k-1} r^k) > 0,\ i = 1, \dots, l. \tag{4.16}$$

Since k tends to infinity and $l < \infty$, at least one of the inequalities holds infinitely often.

Let this be (4.15). Consider the sequence $\{(\lambda^k, \mu^k, K^k)\}_{k=1}^{\infty}$ used in the direction finding problems $D_\alpha^0(y^k, \lambda^k, \mu^k, K^k)$ in the steps of the algorithm. Choose an arbitrary infinite subsequence $\{k \in \mathcal{K}\}$ such that $K^k \equiv K$ for each $k \in \mathcal{K}$.

As a result of the last lemma, τ^k cannot converge to zero. Thus, there exists k^* such that for each $k \ge k^*$, $k \in \mathcal{K}$ we have $\rho^{j_k-1} < \tau^k$. But then,

$$\begin{aligned}
\varepsilon s^k &< \frac{\Phi_\alpha(y^k + \rho^{j_k-1} r^k) - \Phi_\alpha(y^k)}{\rho^{j_k-1}} \\
&= \frac{F(\hat{x}_\alpha(y^k + \rho^{j_k-1} r^k), y^k + \rho^{j_k-1} r^k) - F(\hat{x}_\alpha(y^k), y^k)}{\rho^{j_k-1}} \\
&= \nabla_x F(\hat{x}_\alpha(y^k), y^k)\nabla_y \hat{x}_\alpha(y^k) r^k + \nabla_y F(\hat{x}_\alpha(y^k), y^k) r^k \\
&\quad + \frac{o(\rho^{j_k-1})}{\rho^{j_k-1}} \le s^k + \frac{o(\rho^{j_k-1})}{\rho^{j_k-1}}
\end{aligned}$$

by (4.14), where $\hat{x}_\alpha(y)$ is the optimal solution of the problem (4.2). Passing to the limit on both sides of this inequality leads to $\varepsilon s < s$ which contradicts $\varepsilon \in (0, 1)$ and $s < 0$.

Thus for some $i \in \{1, \dots, l\}$,

$$\begin{aligned} 0 < G_i(y^k + \rho^{j_k-1} r^k) &= G_i(y^k) + \rho^{j_k-1} \nabla_y G_i(y^k) r^k + o(\rho^{j_k-1}) \\ &\leq G_i(y^k) + \rho^{j_k-1}(-G_i(y^k) + s^k) + o(\rho^{j_k-1}) \\ &\leq \rho^{j_k-1}(-G_i(y^k) + s^k) + o(\rho^{j_k-1}) \end{aligned}$$

which means

$$-G_i(y^k) + s^k > -\frac{o(\rho^{j_k-1})}{\rho^{j_k-1}}$$

or $\overline{\lim}_{k\to\infty} G_i(y^k) \leq \lim_{k\to\infty} s^k = s < 0$. Thus, $G_i(y^k + \rho^{j_k-1} r^k) \leq 0$ for each sufficiently large k since ρ^{j_k-1} tends to zero.

This contradiction to (4.16) completes the proof. □

It should be noticed that, if the linear independence constraint qualification with respect to x as well as strict complementarity hold for problem (2.1) at $y = y^0$, then the optimal multiplier and the set K^k in problem $D^0_\alpha(y^k, \lambda^k, \mu^k, K^k)$ are uniquely determined for each sufficiently large k. Hence, in this case, Lemma 4.2 gives the idea of a convergence proof for the descent algorithm.

It is a direct consequence of Lemma 4.2 that, as a stopping criterion in Step 5 of the descent algorithm, we can take a rule based on the distance of the optimal values of problems $D_\alpha(y^k, \lambda, \mu, K)$ from zero for all possible selections of λ, μ, K.

4.2. *Convergence for variable* α

Now, we can turn over to state the convergence result for the above algorithm. But, here again the possibly discontinuous behaviour of the directional derivative $x'_\alpha(\cdot\,;\cdot)$ causes an additional difficulty. Namely, if the optimal values of problems $D^0_{\alpha^k}(y^k, \lambda^k, \mu^k, K^k)$ tend to zero for $\{(\alpha^k, y^k, \lambda^k, \mu^k, K^k)\}_{k=1}^\infty$ converging to $(0, y^0, \lambda^0, \mu^0, K^0)$ there is no guarantee that the necessary optimality condition of Theorem 3.5 is satisfied even if $x'_0(y^0; r)$ exists for each r. Then we only have that $\theta_0(y^0, \lambda^0, \mu^0, K^0) \geq 0$. This difficulty is due to the vacancy of lower semicontinuity of the point-to-set mapping $\Lambda_\cdot(\cdot, \cdot)$. To overcome it we introduce the notion of γ-active constraints and approximate the optimal solutions of problems $D^0_\alpha(y, \lambda, \mu, K)$ by solving problems $D^0_\alpha(\hat{y}, \hat{\lambda}, \hat{\mu}, K)$ at perturbed points $(\hat{y}, \hat{\lambda}, \hat{\mu})$. Let $I_\gamma(x, y) := \{j : 0 \geq g_j(x, y) \geq -\gamma\}$ for $\gamma > 0$ sufficiently small. Now, the whole descent algorithm can be stated more concretely as:

Step 1. Select y^0 solving $G(y^0) \leq 0$, choose a starting value for α, a small ε', a sufficiently small γ, $\varepsilon \in (0, 1)$, a factor $\delta \in (0, 1)$, a $w < 0$, and set $k := 0$.

Step 2. Choose (K^k, λ^k, μ^k) with

$$(\lambda^k, \mu^k) \in E\Lambda_\alpha(x_\alpha(y^k), y^k) \text{ and } \left\{j : \lambda^k_j > 0\right\} \subseteq K^k \subseteq I(x_\alpha(y^k), y^k)$$

satisfying condition 2 of Corollary 3.9, such that $(d^k, r^k, v^k, \omega^k, s^k)$ is an optimal solution for problem $D^0_\alpha(y^k, \lambda^k, \mu^k, K^k)$.

If $s^k < w$ then goto Step 3′. If $s^k \geq w$ and not all possible (λ^k, μ^k, K^k) are tried then continue with Step 2. If all (λ^k, μ^k) and all K^k tried to be used then set $w := w/2$. If $|w| < \varepsilon'$ then goto Step 2′ else to Step 2.

Step 2′. Choose (K^k, λ^k, μ^k) satisfying

$$K^k \subseteq I_\gamma(x_\alpha(y^k), y^k)$$

and condition 2 of Corollary 3.9 as well as

$$(\lambda^k, \mu^k) \in \operatorname*{Argmin}_{\lambda,\mu}\{\|\nabla_x L_\alpha(x_\alpha(y^k), y^k, \lambda, \mu)\|^2 \mid \lambda_j = 0,\ j \notin K^k\},$$

such that $(d^k, r^k, v^k, \omega^k, s^k)$ is an optimal solution for the problem $D^0_\alpha(y^k, \lambda^k, \mu^k, K^k)$.

If $s^k < w$ then goto Step 3′. If $s^k \geq w$ and not all possible (λ^k, μ^k, K^k) are tried then continue with Step 2′. If all K^k tried to be used then set $w := w/2$. If $|w| < \varepsilon'$ then stop.

Step 3′. Select the largest number t^k in $\{\rho, \rho^2, \rho^3, \rho^4, \dots\}$, where $\rho \in (0, 1)$, such that $\Phi_\alpha(y^k + t^k r^k) \leq \Phi_\alpha(y^k) + \varepsilon t^k s^k$, and $G(y^k + t^k r^k) \leq 0$. If $t^k < \varepsilon'$ then drop the actual set K^k and continue searching for a new set K^k in Step 2′.

Step 4. Set $y^{k+1} := y^k + t^k r^k$, $k := k + 1$.

Step 5′. If $\alpha > \varepsilon'$ then $\alpha := \delta\alpha$, $x_\alpha(y^k) \in \Psi^\alpha(y^k)$, goto Step 2, else verify $|s^k| \leq \varepsilon'$. If true then goto Step 2′, else to Step 2.

A new variable w for the control of access to Step 2 or 2′ is additionally included. Its value is compared with s at the end of Steps 2 and 2′.

The value ε' must be so small that according to Lemma 4.1 the exit in Step 3′ can only be used if a set K is selected in the Step 2′ such that the problem $D^0_\alpha(y^k, \lambda^k, \mu^k, K)$ has a negative optimal value, but the corresponding direction r is a direction of ascent. This is obviously possible, if K is nowhere a set of active constraints locally around y^k.

The choice of γ seems to be a difficult task. But, on the first hand, there is a positive γ^0 such that $I_\gamma(x(y^0), y^0) = I(x(y^0), y^0)$ for each $0 < \gamma < \gamma^0$. Moreover, the set $I(x(y^0), y^0) \setminus I(x(y^k), y^k)$ should contain only a few elements (one or two) for sufficiently large k for most of the instances. On the one hand, searching for a direction of descent by use of the Step 2′ of the algorithm can result in a drastic increase of the numerical effort at least if γ is too large. Thus, we suggest to use the Step 2′ only in the case when the values of $\Phi'_\alpha(y^k; r^k)$ and α are sufficiently small and then only for small γ. On the other hand, if the Step 2′ successfully terminates with a useful direction r and with a set $K \not\subseteq I(x_\alpha(y^k), y^k)$, then the calculated descent in the objective function value can be expected to be much larger than during the last iterations.

For proving the convergence of the descent algorithm to a stationary point y^0 of problem **(P)** we need the additional assumption

(A8) For each (λ, μ, d, x) satisfying

$$x \in \Psi^0(y^0),\ (\lambda, \mu) \in \Lambda_0(x, y^0),\ d \neq 0,$$
$$\nabla_x g_i(x, y^0) d = 0, \quad \text{for } i \text{ satisfying } \lambda_i > 0,$$
$$\nabla_x h_i(x, y^0) d = 0, \quad \text{for } i = 1, \dots, q,$$

we have

$$d^{\mathrm{T}}\nabla^2_{xx}L_0(x, y^0, \lambda, \mu)d > 0.$$

This assumption guarantees that problems (1.1) and (2.1) have unique optimal solutions locally around y^0. Hence, convergence of y to y^0 and α to zero from above imply the convergence of $x_\alpha(y)$ to $x(y^0)$ and that the optimal solutions of the problems (1.1) and (1.4) coincide.

Theorem 4.3. *Let the assumptions* (A1)–(A7) *be satisfied for problem* (**P**) *on* M. *Take* $\gamma > 0$ *sufficiently small and fixed. Let the sequence* $\{(y^k, \lambda^k, \mu^k, d^k, r^k, s^k, t^k, v^k, \omega^k, K^k)\}_{k=1}^{\infty}$ *be computed by the descent algorithm where* $\alpha^k \searrow 0$. *Thus if* y^0 *is an accumulation point of* $\{y^k\}_{k=1}^{\infty}$ *satisfying the additional assumption* (A8), *then* $(x_0(y^0), y^0)$ *is a stationary point of problem* (**P**).

Proof: The proof will be divided into three parts: In the first one we show that, if $(x_0(y^0), y^0)$ is not stationary then a problem $D^0_{\alpha^k}(y^k, \lambda^k, \mu^k, K)$ with negative optimal value s^k is possible to construct for k sufficiently large. In the second part it will be derived that a sufficient decrease in the objective function value can be obtained by use of these selections (λ^k, μ^k, K). The third part proves that the fixed set K can infinitely often be taken.

Assume that $(x_0(y^0), y^0)$ is not a stationary point of problem (**P**). By the assumptions and according to Lemma 3.1, $\Psi^0(y^0)$ reduces to a singleton and the solution function $x_0(\cdot)$ is directionally differentiable at $y = y^0$. Thus, by Corollary 3.9, there exist $(\lambda^0, \mu^0) \in E\Lambda_0(x_0(y^0), y^0)$ and a set K having the appropriate properties such that the optimal function value of problem $D^0_0(y^0, \lambda^0, \mu^0, K)$ is less than zero. Let, without loss of generality, the set K with this property be unique (else, choose a suitable subsequence, which is a choice of an infinite number of elements out of finitely many possibilities, in what follows). Due to $\gamma > 0$ we have $K \subseteq I(x_0(y^0), y^0) \subseteq I_\gamma(x_\alpha(y^k), y^k)$ for sufficiently large k. Moreover, by [1, Theorem 5.3.2.], $\{(\lambda^k, \mu^k)\}$, computed by the algorithm in the Step 2′, converges to (λ^0, μ^0). Thus, the optimal function value of problem $D^0_{\alpha^k}(y^k, \lambda^k, \mu^k, K)$ also converges to the optimal function value of problem $D^0_0(y^0, \lambda^0, \mu^0, K)$, i.e., it will be less than zero for sufficiently large k.

Let $(d^k, r^k, v^k, \omega^k, s^k)$ be an optimal solution of $D^0_{\alpha^k}(y^k, \lambda^k, \mu^k, K)$. Let r^0 be an accumulation point of the bounded sequence $\{r^k\}_{k=1}^{\infty}$. Analogously to the proof of Lemma 4.1, $(\nabla_y\hat{x}_0(y^0)r^0, \nabla_y\hat{\lambda}_0(y^0)r^0, \nabla_y\hat{\mu}_0(y^0)r^0, s^0)$ is an accumulation point of $\{(d^k, v^k, \omega^k, s^k)\}_{k=1}^{\infty}$ with $(\nabla_y\hat{x}_0(y^0)r^0, r^0, \nabla_y\hat{\lambda}_0(y^0)r^0, \nabla_y\hat{\mu}_0(y^0)r^0, s^0)$ being an optimal solution of $D^0_0(y^0, \lambda^0, \mu^0, K)$, since assumption (A8) is satisfied at $y = y^0$. But then, due to the proof of Lemma 4.1, there exists $t^0 > 0$ such that

$$g_i(\hat{x}_0(y^0 + t^0r^0), y^0 + t^0r^0) < 0,\ i \notin K,\ \hat{\lambda}_i(y^0 + t^0r^0) > 0,\ i \in K. \tag{4.17}$$

Hence, the optimal solution $\hat{x}_\alpha(y)$ of the enlarged problem (4.2) is also optimal for (2.1) for each y in a sufficiently small neighbourhood of $y = y^0 + t^0r^0$ and sufficiently small

$\alpha > 0$. Moreover, by Lemma 4.2, $t^0 > 0$ can be chosen such that

$$\Phi_0(y^0 + t^0 r^0) = \Phi_0(y^0) + t^0 \Phi_0'(y^0 + \xi r^0; r^0) < \Phi_0(y^0) + t^0 \varepsilon s^0 \tag{4.18}$$

by $\Phi_0'(y^0 + \xi r^0; r^0) = \nabla_x F(\hat{x}_0(y^0 + \xi r^0), y^0 + \xi r^0) \nabla_y \hat{x}_0(y^0 + \xi r^0) r^0 + \nabla_y F(\hat{x}_0(y^0 + \xi r^0), y^0 + \xi r^0) r^0$ with $\xi \in (0, t^0)$ tending to s^0 for $t^0 \to 0$, and

$$G_i(y^0 + t^0 r^0) = G_i(y^0) + t^0 \nabla_y G_i(y^0 + \xi r^0) r^0 < 0 \tag{4.19}$$

for all i. Hence, by continuity of $x.(\cdot)$, there exists an open neighbourhood U of $(y^0, r^0, 0)$ such that, for each $(y, r, \alpha) \in U$, $\alpha > 0$, the inequalities (4.17)–(4.19) are satisfied as well.

Now, by $\lim y^k = y^0$, $\lim \alpha^k = 0$ and r^0 being an accumulation point of $\{r^k\}_{k=1}^\infty$ we have $(y^k, r^k, \alpha^k) \in U$ for sufficiently large k. Thus, the point $\hat{y}^{k+1} := y^k + t^0 r^k$ will be a suitable choice for the next iteration point for sufficiently large k, i.e., the selection of a step-size in Step 3′ of the modified algorithm is indeed possible. Hence we come up to the following conclusion: If the set K is taken in the direction finding problem, then the algorithm selects a step-size $t^k \geq t^0 > 0$ and a new point $y^{k+1} = y^k + t^k r^k$ is computed for which

$$\Phi_{\alpha^k}(y^k + t^k r^k) \leq \Phi_{\alpha^k}(y^k) + t^k \varepsilon s^k$$

holds. Moreover, by $\{\alpha^k\}_{k=1}^\infty$ tending to zero and $\lim_{\alpha \searrow 0} x_\alpha(y^k + t^k r^k) = x_0(y^k + t^k r^k)$ we have

$$\begin{aligned} &|\Phi_{\alpha^k}(y^k + t^k r^k) - \Phi_{\alpha^{k+1}}(y^k + t^k r^k)| \\ &\quad = |F(x_{\alpha^k}(y^k + t^k r^k), y^k + t^k r^k) - F(x_{\alpha^{k+1}}(y^k + t^k r^k), y^k + t^k r^k)| \leq -t^k \varepsilon s^k / 2 \end{aligned}$$

by $s^k < 0$ for sufficiently large k. Hence

$$\Phi_{\alpha^{k+1}}(y^k + t^k r^k) \leq \Phi_{\alpha^k}(y^k) + t^k \varepsilon s^k / 2 \tag{4.20}$$

for sufficiently large k.

Now, since the optimal value of the direction finding problems $D^0_{\alpha^k}(y^k, \lambda^k, \mu^k, K^k)$ with $K^k \subseteq I(x_{\alpha^k}(y^k), y^k)$ and $\{j : \lambda_j > 0\} \subseteq K^k$, $(\lambda, \mu) \in E\Lambda_{\alpha^k}(x_{\alpha^k}(y^k), y^k)$ converge to zero for $k \to \infty$, the modified direction finding problem $D^0_{\alpha^k}(y^k, \lambda^k, \mu^k, K)$ will be used for computing the direction of descent infinitely often in the algorithm. It can be shown in full analogy to the proof of Lemma 4.1 that the sequence $\{t^k\}$ of step-sizes taken in these iterations cannot converge to zero since the optimal values s^k are bounded away from zero. But then, $\Phi_{\alpha^{k+1}}(y^{k+1})$ converges to minus infinity by (4.20) which cannot be true by Remark 2.4. □

5. Remarks about the inclusion of $G(x, y)$ instead of $G(y)$ in the upper level

The more general problem (PG) consists in the following:

$$\varphi(y) \to \min_y$$

where

$$\varphi(y) = \min_x \{F(x, y) \mid G(x, y) \leq 0, x \in \Psi(y)\}$$

1. The next examples will be used to show the more complicated nature of this kind of problems:

   Example 1.

   $$x^2 + y \to \min_y$$
   $$-x - y \leq 0$$

   where x solves

   $$x \to \min_x$$
   $$x \geq 0$$

   Its optimal solution $x = y = 0$ differs from the optimal solution of the problem where the upper level constraints are shifted into the lower level. A shift of $x \geq 0$ leads to unsolvability. For this reason $G(x, y) \leq 0$ are really additional constraints which can not be shifted.
2. The problem (PG) can be embedded in a class of multilevel programming problems (similar to [2]). In this case a distinction to which minima the upper-level constraints belong is necessary. Primary this is a decision depending on the model, but the place of the constraints may also change the solvability and the number of applicable mathematical methods. If the constraints belong to $\min_y$ then the feasibility of the choice of y is checked after computing the "answer" $x(y)$ on y by solving the first minimization problem. Otherwise, the feasible set of the lower level may become smaller. — Although the practice is the most important influence to this decision the existing programming techniques are easier to apply to the constrained minimization over x.
3. Considering such more complicated problems the difficulty of the choice of a starting point arises. Therefore we propose to solve

   $$\|G(x, y) + z\|^2 \to \min_{y, z \geq 0,\ x \in \Psi(y)}$$

   which can be done by the suggested algorithm because every pair (y, z) with $\Psi(y) \neq \emptyset$ and $z = 0$ is a starting point of this new two-level problem.

   Another possibility is a penalty approach [11] which seems to be less adapted to the most applications because of the peculiar optimality definition which was necessary in [11] to prove the convergence of their algorithm. In comparison with the most other authors [15, 18, 19] their optimal solution is only a feasible one.

Now, problem (PG) is shown as a sequence of problems $(\mathrm{PG})_\alpha$:

$$\varphi_\alpha(y) \to \min_y$$

where

$$\varphi_\alpha(y) = F(x_\alpha(y), y)$$

and

$$x_\alpha(y) \in \mathrm{Argmin}\{f(x, y) + \alpha F(x, y) \mid G(x, y) \leq 0, x \in M(y)\}.$$

4. But an additional difficulty arises: Changing α to a smaller one it may happen that the new feasible set of the two-level problem becomes empty.

   The influence of this phenomenon can be partly avoided by a supplementary decrease of α or a small change of y. The full extent of this difficulty shows the following example:

   Example 2.

$$(x - 1)^2 + y^2 \to \min_y$$

where x solves

$$(x - 1)^2 + y^2 \to \min_x$$
$$x - y \leq 0$$
$$0 \leq y \leq 0.75$$

where x also solves

$$y \to \min_x$$
$$0 \leq x \leq 2$$

In the case $\alpha = 0$ the feasible set of the second minimization is equal to $[0, y]$ in contrast to its emptiness whenever $\alpha \neq 0$.

Appendix

Proof of Theorem 2.1:

1. Upper semicontinuity of $\Gamma(\cdot, \cdot)$ follows by straightforward application of Theorems 3.1.1., 3.1.5., and 4.3.3. of [1].
2. Inequality (2.4) is implied by $x(y, \alpha) \in M(y)$. Inequality (2.3) follows since otherwise at least one point in $\Psi(y)$ would give a better solution for problem (2.1) than $x(y, \alpha)$.
3. Part 3 is a simple consequence of Part 2 and upper semicontinuity of $\Gamma(y, \cdot)$. □

Proof of Theorem 2.2:

1. If the optimal value function $\Phi(\cdot)$ defined by $\Phi(y) = F(x(y), y)$ with $x(y)$ given by (1.4) is lower semicontinuous on $Y := \{y \mid \exists\, x$ such that $(x, y) \in M\}$ in the sense that $\forall\, y \in Y$
$$\underline{\lim}_{y' \to y} \Phi(y') \geq \Phi(y),$$
then **(P)** has an optimal solution, since Y is bounded by (A3) and $\Phi(y) > -\infty$ on Y with $\Phi(\hat{y}) < \infty$ for at least one $\hat{y} \in Y$ by (A1) and (A3).
2. Now, we show that $\Phi(\cdot)$ is lower semicontinuous on Y. Let $y^0 \in Y$ be arbitrarily fixed. Then, by (A1)–(A4), and Theorems 3.1.1., 3.1.5., and 4.3.3. of [1] $\Psi(\cdot)$ is upper semicontinuous at y^0. Thus, by (A1) and Theorem 4.2.2. in [1], $\Phi(\cdot)$ is indeed lower semicontinuous at y^0. □

Proof of Lemma 3.6: Let $r \in T_\alpha(\lambda^0, \mu^0)$. We have to show the existence of an open neighbourhood $U(r)$ of r with $U(r) \subseteq T_\alpha(\lambda^0, \mu^0)$. Arguing by contradiction, let there exist a sequence $\{r^k\}_{k=1}^\infty$ converging to r with $r^k \notin T_\alpha(\lambda^0, \mu^0)$ for each k. But then, since the linear programming problem (3.9) has an optimal solution for each k, there exist vertices $(\lambda^k, \mu^k) \in S_\alpha(y^0; r^k)$ with $(\lambda^0, \mu^0) \neq (\lambda^k, \mu^k)$ for each k. Fix an infinite subsequence $\{k \in \mathcal{K}\}$ to guarantee that $(\lambda^k, \mu^k) = (\lambda^1, \mu^1)$ for each $k \in \mathcal{K}$. But then, $(\lambda^1, \mu^1) \in S_\alpha(y^0; r)$ by upper semicontinuity of the solution set mapping of parametric linear programming problems with fixed feasible sets [1, Theorem 5.2.2.]. This contradiction proves the openness of $T_\alpha(\lambda^0, \mu^0)$.

Clearly, if $r \in T_\alpha(\lambda^0, \mu^0)$, then $tr \in T_\alpha(\lambda^0, \mu^0)$ for $t > 0$.

Let $r', r'' \in T_\alpha(\lambda^0, \mu^0)$. Then

$$\nabla_y L_\alpha(x_\alpha(y), y, \lambda^0, \mu^0) r > \nabla_y L_\alpha(x_\alpha(y), y, \lambda, \mu) r \tag{A.1}$$

for each $(\lambda, \mu) \in E\Lambda_\alpha(x_\alpha(y), y) \setminus \{(\lambda^0, \mu^0)\}$ for $r = r'$, $r = r''$ implies the validity of (A.1) also for $r = \delta r' + (1 - \delta) r''$ and $\delta \in [0, 1]$. □

Proof of Lemma 3.7: Let r' be an interior point in $T_\alpha^K(\lambda^0, \mu^0)$ and let there exist γ_i, $i \in K$, δ_j, $j = 1, \ldots, q$, $\sum_{i \in K} \gamma_i^2 + \sum_{j=1}^q \delta_j^2 > 0$ such that

$$\sum_{i \in K} \gamma_i \nabla_x g_i(x_\alpha(y^0), y^0) + \sum_{j=1}^q \delta_j \nabla_x h_j(x_\alpha(y^0), y^0) = 0. \tag{A.2}$$

Then, for each feasible point (d, r, v, ω, s) of $D_\alpha(y^0, \lambda^0, \mu^0, K)$ with $r \in T_\alpha^K(\lambda^0, \mu^0)$ we have by using (3.6) and (3.7),

$$\begin{aligned} &\sum_{i \in K} \gamma_i \nabla_x g_i(x_\alpha(y^0), y^0) d + \sum_{j=1}^q \delta_j \nabla_x h_j(x_\alpha(y^0), y^0) d \\ &\quad + \sum_{i \in K} \gamma_i \nabla_y g_i(x_\alpha(y^0), y^0) r + \sum_{j=1}^q \delta_j \nabla_y h_j(x_\alpha(y^0), y^0) r \\ &= \sum_{i \in K} \gamma_i \nabla_y g_i(x_\alpha(y^0), y^0) r + \sum_{j=1}^q \delta_j \nabla_y h_j(x_\alpha(y^0), y^0) r = 0. \end{aligned}$$

Since this equation is valid for each r in some neighbourhood of r', we have

$$\sum_{i\in K} \gamma_i \nabla_y g_i(x_\alpha(y^0), y^0) + \sum_{j=1}^{q} \delta_j \nabla_y h_j(x_\alpha(y^0), y^0) = 0. \tag{A.3}$$

Equations (A.2) and (A.3) together contradict (A6), which proves the lemma. □

Proof of Lemma 3.8: By definition,

$$T_\alpha(\lambda^0, \mu^0) \supseteq \bigcup_{K\in\mathcal{M}} T_\alpha^K(\lambda^0, \mu^0).$$

Due to Lemma 3.6 it is sufficient to show

$$T_\alpha(\lambda^0, \mu^0) \subseteq \bigcup_{K\in\mathcal{M}} \operatorname{cl} T_\alpha^K(\lambda^0, \mu^0).$$

Let $r \in T_\alpha(\lambda^0, \mu^0)$. We show the existence of a set $K \in \mathcal{M}$ such that $r \in \operatorname{cl} T_\alpha^K(\lambda^0, \mu^0)$. Using this inclusion, we obtain $r \in \operatorname{cl} T_\alpha^K(\lambda^0, \mu^0) \subseteq \operatorname{cl} \bigcup_{K\in\mathcal{M}} T_\alpha^K(\lambda^0, \mu^0)$ for each $r \in T_\alpha(\lambda^0, \mu^0)$, i.e., the lemma is valid. Arguing by contradiction, let there be no set $K \in \mathcal{M}$ with the desired property. Then,

$$r \in T_\alpha(\lambda^0, \mu^0) \setminus \bigcup_{K\in\mathcal{M}} \operatorname{cl} T_\alpha^K(\lambda^0, \mu^0)$$

and this set is open due to Lemma 3.6 and finiteness of $\mathcal{M}$. Hence, there exists an open neighbourhood $U(r)$ of r with

$$U(r) \subseteq T_\alpha(\lambda^0, \mu^0) \setminus \bigcup_{K\in\mathcal{M}} \operatorname{cl} T_\alpha^K(\lambda^0, \mu^0)$$
$$\subseteq \bigcup_{K\in\mathcal{N}\setminus\mathcal{M}} \operatorname{cl} T_\alpha^K(\lambda^0, \mu^0) \subseteq \bigcup_{K\in\mathcal{N}\setminus\mathcal{M}} \operatorname{aff} T_\alpha^K(\lambda^0, \mu^0),$$

where $\operatorname{aff} A$ denotes the affine hull of a set A, since $(A \cup B) \setminus B \subseteq A$ for arbitrary sets A and B. But, the last inclusion cannot be valid since an open set cannot be contained in the union of a finite number of lower-dimensional subspaces. □

Acknowledgments

The authors thank the referees for carefully reading of the paper and their valuable suggestions especially leading to a correction of the proof of Lemma 4.2.

Note

1. This and some other proofs can be found in the Appendix.

References

1. B. Bank, J. Guddat, D. Klatte, B. Kummer, and K. Tammer, Non-Linear Parametric Optimization, Akademie-Verlag: Berlin, 1982.
2. J.F. Bard, "Geometric and algorithmic developments for a hierarchical planning problem," European J. Oper. Res., vol. 19, pp. 372–383, 1985.
3. J.F. Bard, "Convex two-level optimization," Math. Programming, vol. 40, no. 1, pp. 15–27, 1988.
4. S. Dempe, "Directional differentiability of optimal solutions under Slater's condition," Math. Programming, vol. 59, no. 1, pp. 49–69, 1993.
5. S. Dempe, "A necessary and a sufficient optimality condition for bilevel programming problems," Optimization, vol. 25, pp. 341–354, 1992.
6. A.V. Fiacco, Introduction to Sensitivity and Stability Analysis in Nonlinear Programming, Academic Press: New York, 1983.
7. J. Gauvin, "A necessary and sufficient regularity condition to have bounded multipliers in nonconvex programming," Math. Programming, vol. 12, no. 1, pp. 136–139, 1977 .
8. J. Guddat, F.G. Vasquez, and H.Th. Jongen, Parametric Optimization: Singularities, Pathfollowing and Jumps, B.G. Teubner: Stuttgart, 1990.
9. P.T. Harker and J.-S. Pang, "Existence of optimal solutions to mathematical programs with equilibrium constraints," Oper. Res. Lett., vol. 7, no. 2, pp. 61–64, 1988.
10. W.W. Hogan, "Directional derivatives for extremal-value functions with applications to the completely convex case," Oper. Res., vol. 21, no. 1, pp. 188–209, 1973.
11. Y. Ishizuka and E. Aiyoshi, "Double penalty method for bilevel optimization problems," Ann. of Oper. Res., vol. 34, pp. 73–88, 1992.
12. K. Jittorntrum, "Solution point differentiability without strict complementarity in nonlinear programming," Math. Programming Stud., no. 21, pp. 127–138, 1984.
13. I. Kaneko, "On some recent engineering applications of complementarity problems," Math. Programming Stud., vol. 17, pp. 111–125, 1982.
14. P. Loridan and J. Morgan, "ε–regularized two-level optimization problems: Approximation and existence results," in Optimization—Fifth French-German Conference, Varetz, 1988, Lect. Notes in Mathematics no. 1405, pp. 99–113, 1989.
15. R. Lucchetti, F. Mignanego, and G. Pieri, "Existence theorem of equilibrium points in Stackelberg games with constraints," Optimization, vol. 18, no. 6, pp. 857–866, 1987.
16. B. Luderer, "Über die Äquivalenz nichtlinearer Optimierungsaufgaben," Working Paper, Techn. Univ. Karl-Marx-Stadt, 1983.
17. P. Marcotte, "Network design problem with congestion effects: A case of bilevel programming," Math. Programming, vol. 34, no. 2, pp. 142–162, 1986.
18. W. Oeder, Ein Verfahren zur Lösung von Zwei-Ebenen-Optimierungsaufgaben in Verbindung mit der Untersuchung von chemischen Gleichgewichten, Ph.D.-Thesis, Techn. Univ. Karl-Marx-Stadt, 1988.
19. J.S. Pang, S.P. Han, and N. Rangaraj, "Minimization of locally Lipschitzian functions," SIAM J. Optim. vol. 1, pp. 57–82, 1991.
20. T. Petersen, Optimale Anreizsysteme, Gabler Verlag: Wiesbaden, 1989.
21. S.M. Robinson, "Local structure of feasible sets in nonlinear programming, Part III: Stability and sensitivity," Math. Programming Stud., vol. 30, pp. 45–66, 1987.
22. A. Shapiro, "Sensitivity analysis of nonlinear programs and differentiability properties of metric projections," SIAM J. Control Optim., vol. 26, pp. 628–645, 1988.
23. W.R. Smith and R.W. Missen, Chemical Reaction Equilibrium Analysis: Theory and Algorithms, J. Wiley & Sons: New York, 1982.
24. L.N. Vicente and P.H. Calamai, "Bilevel and Multilevel Programming: A Bibliography Review," Report, Department of Systems Design Engineering, University of Waterloo, Waterloo, N2L 3G1 Ontario, Canada, 1993.