ON SOME TOPICS IN SEMI-INFINITE PROGRAMMING (A PRELIMINARY DISCUSSION)

November 23, 2005

GEORG STILL

ABSTRACT. A semi-infinite programming problem is an optimization problem in which finitely many variables appear in infinitely many constraints. This model naturally arises in an abundant number of applications in different fields of mathematics, economics and engineering. The present paper intends to give a short introduction into the field and to present some preliminary discussion on the complexity of linear SIP.

1. INTRODUCTION

1.1. **Problem formulation.** A *semi-infinite program* (SIP) is an optimization problem in finitely many variables $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ on a feasible set described by infinitely many constraints:

$$\text{(1)} \qquad \text{P:} \quad \min_x \; f(x) \qquad \text{s.t.} \quad g(x,y) \geq 0 \quad \forall y \in Y ,$$

where Y is an infinite *index set*. $\mathcal{F}$ will denote the feasible set, $v = \inf\{ f(x) \mid x \in \mathcal{F} \}$ the optimal value and $\mathcal{S} = \{\overline{x} \in \mathcal{F} \mid f(\overline{x}) = v\}$ the set of minimizers of the problem. We assume that $Y \subset \mathbb{R}^m$ is compact, $g(x,y)$ is continuous on $\mathbb{R}^n \times Y$ and for any fixed $y \in Y$ the gradient $\nabla_x g(x,y)$ (wrt. x) exists and is continuous on $\mathbb{R}^n \times Y$.

We also will consider generalizations of SIP, where the index set $Y = Y(x)$ is allowed to be dependent on x (GSIP for short),

$$\text{(2)} \qquad \min_x \; f(x) \qquad \text{s.t.} \quad g(x,y) \geq 0 \quad \forall y \in Y(x) .$$

During the last five decades the field of Semi-infinite Programming has known a tremendous development. More than 1000 articles have been published on the theory, numerical methods and applications of SIP. Originally the research on SIP was strongly related to (Chebyshev) approximation, see *e.g.*, [11], [7]. As excellent review articles on the field we refer to Polak [17] and Hettich/Kortanek [8]. Since a first contribution [12] the *generalized SIP problem* (2) became a topic of intensive research (see *e.g.*, [14] and [21]). For an extensive treatment of linear semi-infinite problems we refer to the book by Goberna/Lopez [9].

REMARK 1. For shortness and a clearer presentation we omit additional equality constraints $h_i(x) = 0$ in the formulation of SIP. It is not difficult to generalize (under appropriate assumptions on h_i) all results in this paper to the more general situation.

The paper is organized as follows. The next section presents some illustrative applications of SIP and GSIP. Section 3 shortly describes the structure of the feasible set of semi-infinite problems. Section 4 derives first order necessary and sufficient optimality conditions. Linear semi-infinite and linear semidefinite programming is treated in Section 5 also from the viewpoint of complexity. Section 6 presents an interior point method for LP and discusses the limits to extend such methods to (general) linear SIP. In Section 7 the discretization methods for solving SIP is surveyed.

2. APPLICATIONS

In [17], [8] and [9] many applications of SIP are mentioned in different fields such as (Chebyshev) approximation), robotics, boundary- and eigen value problems from Mathematical Physics, engineering design, optimal control, transportation problems, fuzzy sets, cooperative games, robust optimization and statistics. From this large list of applications we have chosen some illustrative examples.

Chebyshev Approximation : Let be given a function $f(y) \in C(\mathbb{R}^m, \mathbb{R})$ and a space of approximating functions $p(x, y)$, $p \in C(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R})$, parameterized by $x \in \mathbb{R}^n$. We want to approximate f by functions $p(x, \cdot)$ in the max-norm (Chebyshev-norm) $\| f\|_\infty = \max_{y\in Y} |f(y)|$ on a compact set $Y \subset \mathbb{R}^m$. To minimize the approximation error $\epsilon = \| f - p\|_\infty$, leads to the problem:

$$(3) \qquad \min_{x,\epsilon} \epsilon \ \text{ s. t. } \ g^{\pm}(x, y) := \pm\big(f(y) - p(x, y)\big) \le \epsilon \ \text{ for all } y \in Y \ .$$

This is a semi-infinite problem.

The so-called reverse Chebyshev problem consists of fixing the approximation error ϵ and making the region Y as large as possible (see [13] for such problems). Suppose, the set $Y = Y(d)$ is parameterized by $d \in \mathbb{R}^k$ and $v(d)$ denotes the volume of $Y(d)$ (e.g. $Y(d) = \Pi_{i=1}^k[-d_i, d_i]$). The reverse Chebyshev problem then leads to the GSIP (ϵ fixed).

$$(4) \qquad \max_{d,x} v(d) \text{ s. t. } \ g^{\pm}(x, y) := \pm\big(f(y) - p(x, y)\big) \le \epsilon \text{ for all } y \in Y(d) \ .$$

where the index set $Y(d)$ depends on the variable d. We refer the reader also to [13].

Mathematical Physics. The so-called defect minimization approach leads to semi-infinite programming models for solving problems from Mathematical Physics which is different from the common finite element and finite difference approaches. We give an example.

Shape optimization problem : Consider the following Boundary-value problem,

BVP: Given $G_0 \subset \mathbb{R}^m$ (G_0 a simply connected open region with smooth boundary ∂G_0, (closure $\overline{G}_0$)) and $k > 0$. Find a function $u \in C^2(\overline{G}_0, \mathbb{R})$ such that with the Laplacian $\Delta u = u_{y_1 y_1} + \ldots + u_{y_m y_m}$:

$$\begin{aligned} \Delta u(y) &= k, \qquad \forall y \in G_0 \\ u(y) &= 0, \qquad \forall y \in \partial G_0 \end{aligned}$$

By choosing a linear space of appropriate trial functions $u_j \in C^2(\mathbb{R}^m, \mathbb{R})$,

$$S = \left\{ u(x, y) = \sum_{i=1}^{n-1} x_i u_i(y) \right\}$$

this BVP can approximately be solved via the following SIP:

$$\min_{\varepsilon, x} \; \varepsilon \qquad \text{s.t.} \qquad \begin{aligned} \pm\big(\Delta u(x, y) - k\big) &\leq \varepsilon, \quad y \in G_0 \\ \pm u(x, y) &\leq \varepsilon, \quad y \in \partial G_0 \end{aligned}$$

In [5] the related but more complicated so-called *Shape Optimization Problem* has been considered theoretically.

SOP: Find a (simply connected) region $G \in \mathbb{R}^m$ with normalized volume $\mu(G) = 1$ and a function $u \in C^2(\overline{G}, \mathbb{R})$ which solves with a given objective function $F(G, u)$ the problem

$$\min_{\mu(G)=1, u} F(G, u) \qquad \text{s.t.} \qquad \begin{aligned} \Delta u(y) &= k, \qquad \forall y \in G \\ u(y) &= 0, \qquad \forall y \in \partial G \end{aligned}$$

This is a problem with variable region G and can be solved approximately via the following GSIP-problem:

Choose some appropriate set of regions $G(z)$ depending on a parameter $z \in \mathbb{R}^p$ and satisfying $\mu(G(z)) = 1$ for all z. Fix some small error bound $\varepsilon > 0$. Then we solve with trial functions $u(x, y) \in S$ the program

$$\min_{z, x} F(G(z), u(x, \cdot)) \qquad \text{s.t.} \qquad \begin{aligned} \pm\big(\Delta u(x, y) - k\big) &\leq \varepsilon, \quad \forall y \in G(z) \\ \pm u(x, y) &\leq \varepsilon, \quad \forall y \in \partial G(z) \end{aligned}$$

For further contributions to the theory and numerical results of this approach we refer *e.g.* to [20], [8].

Robotics. Control problems in robotics can often be modeled as a semi-infinite problem. We refer to the Ex.2.1 in [8] and the many references therein (cf. also [10]).

Geometry. Semi-infinite problems can be interpreted in a geometrical setting. Given a family of sets $S(x) \subset \mathbb{R}^p$ depending on $x \in \mathbb{R}^n$, we wish to find a 'body' Y of a certain class and a value $\overline{x}$ such that Y is contained in $S(\overline{x})$ and Y is as large as possible.

The mathematical formulation is as follows. Suppose $S(x)$ is defined by

$$S(x) = \{ y \in \mathbb{R}^p \mid g(x, y, t) \geq 0, \ \text{for all } t \in Q \}$$

where Q is a given compact set in $\mathbb{R}^s$ and $g \in C^2(\mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^s, \mathbb{R})$. Let the body $Y(d) \subset \mathbb{R}^p$ be parameterized by $d \in \mathbb{R}^q$ with $v(d)$, a measure for the size of $Y(d)$ (e.g. the volume). To maximize $v(d)$ for $Y(d) \subset S(x)$ then becomes:

$$\max_{d,x} \; v(d) \text{ s.t. } g(x, y, t) \geq 0 \text{ for all } y \in Y(d),\; t \in Q\,. \tag{5}$$

For the case that the set S is fixed this problem is known as *design centering* problem.

Optimization under uncertainty (Robust optimization). As a concrete situation we consider a linear program

$$\min_x \; c^T x \quad \text{s.t.} \quad a_j^T x - b_j \geq 0,\; \forall j \in J,$$

J a finite index set. Often in the model the data a_j and b_j are not known exactly. It is only known that the vectors (a_j, b_j) may vary in a set $Y_j \subset \mathbb{R}^{n+1}$. In a "pessimistic way" we now can restrict the problem to such x which are feasible for all possible data vectors leading to a SIP

$$\min_x \; c^T x \quad \text{s.t.} \quad a^T x - b \geq 0,\; \forall (a, b) \in Y := \cup_{j \in J} Y_j\,.$$

In the next example we discuss such a "robust optimization" model in economics. For more details we refer to [19], [23], [2], [16] and [1].

Economics. We consider the following *Portfolio problem*. We wish to invest K euros into n shares, say for a period of one year. We invest x_i euros in share i and expect at the end of the period a return of y_i euros per 1 euro investment in share i.
Our goal is to maximize the portfolio value $v = y^T x$ after a year, where $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. The problem is that the value y is not known in advance. However often models from economics suggest that the vector y will vary in some compact subset Y of $\mathbb{R}^n$. In this case we are led to solve the linear SIP:

$$\begin{aligned} \max_{v,x} \; v \quad \text{s.t.} \quad & y^T x - v \geq 0 \; \forall y \in Y \\ & \sum_i x_i = K,\; x \geq 0\,. \end{aligned}$$

Different other applications There are other interesting applications in game theory (see [18]) in data envelopment analysis (see [15]) and in probability theory (see [4]).

3. Structure of the feasible set

In this section we shortly discuss the structure of the feasible sets $\mathcal{F}$ of finite and semi-infinite optimization problems.

$\mathcal{F}$ in linear programming (LP): The feasible set is defined by finitely many linear constraints

$$\mathcal{F} = \{x \mid a_j^T x \geq b_j,\; j \in J\}$$

with a finite index set $J = \{1, \ldots, m\}$ and $a_j \in \mathbb{R}^n$, $b_j \in \mathbb{R}$.

$\mathcal{F}$ **in semi-infinite linear programming (LSIP):** The feasible set is defined by (possibly) infinitely many linear constraints

$$\mathcal{F} = \{x \mid a_y^T x \geq b_y,\ y \in Y\} \tag{6}$$

with an (possibly infinite) index set $Y \subset \mathbb{R}^m$ and functions $y \to a_y \in \mathbb{R}^n$, $y \to b_y \in \mathbb{R}$.

Any such set $\mathcal{F}$ is closed and convex but also the converse can be proven with an *separation theorem*.

LEMMA 1. *A subset $\mathcal{F} \subset \mathbb{R}^n$ is convex and closed iff it can be defined in the form (6).*

Proof. To prove that the closed convex set $\mathcal{F}$ can be written in the form (6) let us chose a point $y \notin \mathcal{F}$ arbitrarily. In view of the Separation Theorem (*e.g.*, [6, Th.10.1]) the point $y \notin \mathcal{F}$ can be separated from the closed convex set $\mathcal{F}$ by a separating halfspace $H_y^{\geq} = \{x \mid a_y^T x \geq b_y\}$, *i.e.*,

$$a_y^T x \geq b_y > a_y^T y, \quad \forall x \in \mathcal{F} \ .$$

So, the set $\mathcal{F}$ can be written as $\mathcal{F} = \bigcap_{y \notin \mathcal{F}} H_y^{\geq} = \{x \mid a_y^T x \geq b_y,\ \forall y \notin \mathcal{F}\}$.

□

$\mathcal{F}$ **in finite programming (FP):** The feasible set is defined by finitely many (nonlinear) constraints

$$\mathcal{F} = \{x \mid g_j(x) \geq 0,\ j \in J\} \tag{7}$$

where $J = \{1, \ldots, m\}$ is finite and $g_j : \mathbb{R}^n \to \mathbb{R}$ are continuous.

$\mathcal{F}$ **in semi-finite programming (SIP):** Feasible set defined by infinitely many (nonlinear) constraints

$$\mathcal{F} = \{x \mid g(x, y) \geq 0,\ y \in Y\} \tag{8}$$

with infinite index set $Y \subset \mathbb{R}^m$ and continuous $g : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$.

$\mathcal{F}$ **in generalized semi-finite programming (GSIP):**

$$\mathcal{F} = \{x \mid g(x, y) \geq 0,\ y \in Y(x)\}$$

with variable index set $Y(x)$ defined by a set valued mapping $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ and continuous $g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

The topological structure of the feasible sets $\mathcal{F}$ in FP and SIP are the same.

EX. 1. *The feasible sets $\mathcal{F}$ in FP and SIP are closed.*

The feasible set in GSIP need not be closed.
EX. 2. Consider the problem

$$\min_{x \in \mathbb{R}} x \quad \text{s.t.} \quad x \geq 1 - y, \qquad \forall y \in Y(x) = \{y \in \mathbb{R} \mid 0 \leq y,\ y \leq -x\}$$

Obviously, the feasible set consists of the open interval $(0, \infty)$ ($Y(x) = \emptyset$ means that x is feasible). A minimizer of the program does not exist. The problem here is that the index set $Y(x) = [0, -x]$ for $x \leq 0$, $Y(x) = \emptyset$ for $x > 0$, is not continuous at $x = 0$.

If $Y(x)$ is continuous then, non-closedness of $\mathcal{F}$ is excluded.

LEMMA 2. *Suppose the mapping $Y : K \rightrightarrows \mathbb{R}^m$ is continuous on a compact set $K \subset \mathbb{R}^n$. Then the feasible set $\mathcal{F} \cap K$ of GSIP is compact (in particular closed).*

4. FIRST ORDER OPTIMALITY CONDITIONS

In this section, first order optimality conditions are derived for the SIP problem P in (1). We assume $f, g \in C^1$.
A feasible point $\overline{x} \in \mathcal{F}$ is called a *local minimizer* of SIP if there is some $\varepsilon > 0$ such that

$$f(x) - f(\overline{x}) \geq 0 \qquad \text{for all } x \in \mathcal{F} \text{ with } \|x - \overline{x}\| < \varepsilon \,. \tag{9}$$

The minimizer $\overline{x}$ is said to be *global* if this relation holds for any $\varepsilon > 0$. We call $\overline{x} \in \mathcal{F}$ a *strict local minimizer of order* $p > 0$ if there exist some $q > 0$ and $\varepsilon > 0$ such that

$$f(x) - f(\overline{x}) \geq q \, ||x - \overline{x}||^p \qquad \text{for all } x \in \mathcal{F} \text{ with } \|x - \overline{x}\| < \varepsilon \,. \tag{10}$$

It is not difficult to see that near a point $\overline{x} \in \mathcal{F}$, where the active index set $Y_0(\overline{x}) = \{y \in Y \mid g(\overline{x}, y) = 0\}$ is empty, the SIP problem represents a common unconstrained minimization problem. So throughout the paper we assume that for any (candidate) minimizer of P the condition $Y_0(\overline{x}) \neq \emptyset$ is satisfied.

LEMMA 3. [Primal necessary optimality condition] *Let $\overline{x} \in \mathcal{F}$ be a local minimizer of P. Then there cannot exist a* strictly feasible descent direction d, i.e., *a vector $d \in \mathbb{R}^n$ satisfying the relations*

$$\nabla f(\overline{x})d < 0 \,, \;\; \nabla_x g(\overline{x}, y)d > 0, \;\; \forall y \in Y_0(\overline{x}) \,.$$

Proof. Let d be a strictly feasible descent direction. Then $f(\overline{x} + \tau d) < f(\overline{x})$ holds for small $\tau > 0$ (recall $f \in C^1$). We claim that there exists some $\tau_0 > 0$ with the property

$$g(\overline{x} + \tau d, y) \;\; > \;\; 0 \quad \text{for all } 0 < \tau \leq \tau_0 \text{ and } y \in Y$$

showing that $\overline{x}$ cannot be a local minimizer. Suppose that the claim is false. Then to each $k \geq 1$ there exists some $0 < \tau_k < 1/k$ and some $y_k \in Y$ such that $g(\overline{x} + \tau_k d, y_k) \leq 0$. Since Y is compact, there must exist some convergent subsequence (τ_{k_s}, y_{k_s}) such that $\tau_{k_s} \to 0$ and $y_{k_s} \to y^* \in Y$. The continuity of g then yields $g(\overline{x} + \tau_{k_s} d, y_{k_s}) \to g(\overline{x}, y^*)$, which implies $g(\overline{x}, y^*) = 0$ and $y^* \in Y_0(\overline{x})$. On the other hand, the Mean Value Theorem provides us with numbers $0 < \hat{\tau}_{k_s} < \tau_{k_s}$ such that

$$0 \geq g(\overline{x} + \tau_{k_s} d, y_{k_s}) - g(\overline{x}, y_{k_s}) = \tau_{k_s} \nabla_x g(\overline{x} + \hat{\tau}_{k_s} d, y_{k_s})d$$

and hence $\nabla_x g(\overline{x} + \hat{\tau}_{k_s} d, y_{k_s})d \leq 0$. So the continuity of $\nabla_x g$ entails

$$\nabla_x g(\overline{x}, y^*)d = \lim_{s \to \infty} \nabla_x g(\overline{x} + \hat{\tau}_{k_s} d, y_{k_s})d \;\leq\; 0 \,,$$

which contradicts the hypothesis on d.

□

THEOREM 1. [First order sufficient condition] *Let $\overline{x}$ be feasible for P. Suppose that there does not exist a vector $0 \neq d \in \mathbb{R}^n$ satisfying*

$$\nabla f(\overline{x})d \leq 0 \ , \ \nabla_x g(\overline{x}, y)d \geq 0, \ \forall y \in Y_0(\overline{x}) \ .$$

Then $\overline{x}$ is a strict local minimizer of SIP of order $p = 1$.

Proof. If $\overline{x}$ is an isolated feasible point than the result is trivially fulfilled. If not then, it is not difficult to see that the set $K(\overline{x}) := \{d \in \mathbb{R}^n \mid \|d\| = 1, \nabla_x g(\overline{x}, y)d \geq 0, \ \forall y \in Y_0(\overline{x}) \ \}$ is nonempty. (Consider with a sequence of $x_k \in \mathcal{F}$, $x_k \to \overline{x}$, an accumulation point d of $(x_k - \overline{x})/\|x_k - \overline{x}\|$). Since $K(\overline{x})$ is also compact, the value

$$c_1 = \min_{d \in K(\overline{x})} \ \nabla f(\overline{x})d$$

is attained by some $d' \in K(\overline{x})$. By assumption, $c_1 = \nabla f(\overline{x})d' > 0$ holds. Fix any $0 < c < c_1$. We claim that there is some $\varepsilon > 0$ with the property

$$f(x) - f(\overline{x}) \geq c \ \|x - \overline{x}\| \quad \text{for all } x \in \mathcal{F}, \quad \|x - \overline{x}\| < \varepsilon \ . \tag{11}$$

Suppose to the contrary that this claim is false and there is an infinite sequence of feasible points $x_k \to \overline{x}$ with the property $f(x_k) - f(\overline{x}) < c \ \|x_k - \overline{x}\|$. We write $x_k = \overline{x} + \tau_k d_k$ with $\tau_k > 0$ and $d_k \in S = \{d \in \mathbb{R}^n \mid \|d\| = 1\}$. Then, $x_k \to \overline{x}$ implies $\tau_k \to 0$. Moreover, the compactness of S ensures that some subsequence of (d_k) converges. Without loss of generality, we thus assume $d_k \to \overline{d}$, $\overline{d} \in S$.

The differentiability assumption together with the feasibility condition $g(x_k, y) \geq 0$ and the property $g(\overline{x}, y) = 0$ implies

$$\begin{array}{rcccccl} c\ |\tau_k| & > & f(x_k) - f(\overline{x}) & = & \tau_k \nabla f(\overline{x}) d_k & + & o(|\tau_k|) \ , \\ 0 & \leq & g(x_k, y) - g(\overline{x}, y) & = & \tau_k \nabla g(\overline{x}, y) d_k & + & o(|\tau_k|) \ , \quad y \in Y_0(\overline{x}) \ . \end{array}$$

Divide all these relations by τ_k. Letting $k \to \infty$ ($\tau_k \to 0$), we conclude that $\nabla g(\overline{x}, y)\overline{d} \geq 0$, *i.e.*, $\overline{d} \in K(\overline{x})$ and $\nabla f(\overline{x})\overline{d} \leq c$. This contradicts our choice of $c < c_1$. So (11) must be true.

□

REMARK. The assumptions of Theorem 1 are rather strong and can be expected to hold only in special cases (in Chebyshev approximation *cf. e.g.* [11]). It is not difficult to see that the assumptions imply that the set of gradients $\{\nabla_x g(\overline{x}, y) \mid y \in Y_0(\overline{x}) \ \}$ contain a basis of $\mathbb{R}^n$. So in particular $|Y_0(\overline{x})| \geq n$.

We introduce some *constraint qualifications*. We say that the *Linear Independence* constraint qualification holds at $\overline{x} \in \mathcal{F}$ if

$$\nabla_x g(\overline{x}, y), \ y \in Y_0(\overline{x}), \ \text{are linearly independent}$$

The (weaker) *Mangasarian Fromovitz CQ* (MFCQ) is said to hold at $\overline{x}$ if with some $d \in \mathbb{R}^n$

$$\nabla_x g(\overline{x}, y)d > 0, \quad \forall y \in Y_0(\overline{x}) \ . \tag{12}$$

More general sufficient optimality conditions need second order information.

We now derive the famous *Fritz John* (FJ) and the *Karush-Kuhn-Tucker* (KKT) optimality conditions.

THEOREM 2. [Dual Necessary Optimality Conditions] *Let $\overline{x}$ be a local minimizer of P. Then the following holds.*

(a) *There exist multipliers $\mu_0, \mu_1, \ldots, \mu_{k+1} \geq 0$ and indices $y_1, \ldots, y_{k+1} \in Y_0(\overline{x})$, $k \leq n$, such that $\sum_{j=0}^{k+1} \mu_j = 1$ and*

$$\mu_0 \nabla f(\overline{x}) - \sum_{j=1}^{k+1} \mu_j \nabla_x g(\overline{x}, y_j) = 0 \, . \qquad \text{(FJ- condition)} \tag{13}$$

(b) *If MFCQ holds at $\overline{x}$, then there exist multipliers $\mu_1, \ldots, \mu_k \geq 0$ and indices $y_1, \ldots, y_k \in Y_0(\overline{x})$, $k \leq n$, such that*

$$\nabla f(\overline{x}) - \sum_{j=1}^{k} \mu_j \nabla_x g(\overline{x}, y_j) = 0 \, . \qquad \text{(KKT-condition)} \tag{14}$$

Proof. Consider the set $S = \{\nabla f(\overline{x})\} \cup \{-\nabla_x g(\overline{x}, y) \mid y \in Y_0(\overline{x})\} \subseteq \mathbb{R}^n$. Since $\overline{x}$ is a local minimizer of P, there is no strictly feasible descent direction d at $\overline{x}$ (*cf.* Lemma 3). This means that

$$\text{there is no } d \in \mathbb{R}^n \text{ with } d^T s < 0 \text{ for all } s \in S \, .$$

Now $Y_0(\overline{x})$ is compact. Therefore (by continuity of $\nabla_x g(\overline{x}, \cdot)$) also S is compact. By Lemma 9 it follows $0 \in \text{conv } S$. In view of Caratheodory's Lemma 7, 0 is a convex combination of at most $n+1$ elements of S, *i.e.*,

$$\sum_{j=0}^{k} \mu_j s_j = 0 \qquad s_j \in S, \ \mu_j \geq 0, \ \sum_{j=0}^{k} \mu_j = 1 \quad \text{with } k \leq n \, , \tag{15}$$

which implies (a).

Now assume that $d \in \mathbb{R}^n$ is a strictly feasible direction at $\overline{x}$)*i.e.*, MFCQ holds). For statement (b) it suffices to show: $\mu_0 \neq 0$ in the representation (a) (as division by $\mu_0 > 0$ in (13) yields a representation of type (14)). Suppose to the contrary that $\mu_0 = 0$ is true, and multiply (13) with d. Since $\mu_j > 0$ holds for at least one $j \geq 1$, we obtain the contradiction

$$0 > -\sum_{j=1}^{k+1} \mu_j \nabla_x g(\overline{x}, y_j) d = 0^T d = 0 \, .$$

□

REMARK 2. Note that the results of this section in particular remain true for a finite program. In fact, given a finite program (see (7), we only have to chose a finite set $Y = \{y_1, \ldots, y_m\}$ and to identify $g(x, y_j) := g_j(x)$, $j = 1, \ldots, m$.

EX. 3. Assume that $\overline{x} \in \mathcal{F}$ is a KKT-point such that (14) holds with $k = n$ linear independent vectors $\nabla_x g(\overline{x}, y_j)$, $j = 1, \ldots, n$, (LICQ) and $\overline{\mu}_j > 0$. Show that then the sufficient conditions of Theorem 1 are satisfied.

5. CONVEX AND LINEAR SEMI-INFINITE PROGRAMS.

5.1. **Convex SIP.** The semi-infinite program is called *convex* if the objective function $f(x)$ is convex and, for every index $y \in Y$, the constraint function $g_y(x) = g(x, y)$ is concave (*i.e.*, $-g_y(x)$ is convex).

EX. 4. It is not difficult to show that for a convex SIP the feasible set is convex and that any local minimizer $\overline{x}$ is a global minimizer. Moreover, (if $g \in C^1$) the existence of a MFCQ vector d at a feasible point $\overline{x}$ is equivalent with the Slater condition: There exists a feasible point $\hat{x}$ such that $g(\hat{x}, y) > 0$ for all $y \in Y$.

Recall that a local minimizer of a convex program is actually a global one (see Ex. 4). As in finite programming, the KKT-conditions are sufficient for optimality.

THEOREM 3. *Assume that the SIP problem P is convex. If $\overline{x}$ is a feasible point that satisfies the Kuhn-Tucker condition (14), then $\overline{x}$ is a (global) minimizer of P.*

Proof. By the convexity assumption, we have for every feasible x and $\overline{y} \in Y_0(\overline{x})$ (*cf.* Lemma 8)

$$\begin{aligned} f(x) - f(\overline{x}) &\geq \nabla f(\overline{x})(x - \overline{x}) \\ 0 \leq g(x, \overline{y}) = g(x, \overline{y}) - g(\overline{x}, \overline{y}) &\leq \nabla_x g(\overline{x}, \overline{y})(x - \overline{x}) \ . \end{aligned}$$

Hence, if there are multipliers $\mu_j \geq 0$ and index points $\overline{y}_j \in Y_0(\overline{x})$ such that $\nabla f(\overline{x}) = \sum_{j=1}^{k} \mu_j \nabla_x g(\overline{x}, \overline{y}_j)$, we conclude

$$f(x) - f(\overline{x}) \geq \nabla f(\overline{x})(x - \overline{x}) = \sum_{j=1}^{k} \mu_j \nabla_x g(\overline{x}, \overline{y}_j)(x - \overline{x}) \geq 0 \ .$$

□

5.2. **Linear SIP.** An important special case of (convex) SIP is given by the *linear semi-infinite problem* (LSIP), where the objective function f and the function g are linear in x:

$$\text{(16)} \qquad \text{P:} \quad \min_x \ c^T x \qquad \text{s.t.} \quad a_y^T x \geq b_y \quad \forall y \in Y \ .$$

For an intensive treatment of LSIP we refer to the monograph [9]. Here we only derive strong duality results.

Recall that any convex optimization problem

$$\min_x \ c^T x \qquad \text{s.t.} \quad x \in \mathcal{F} \ , \qquad \mathcal{F} \subset \mathbb{R}^n \text{ a closed convex set,}$$

can be written as a LSIP. LSIP is said to be continuous, if Y is compact and the functions $y \to a_y = a(y)$, $y \to b_y = b(y)$ are continuous on Y.

EX. 5. *Let P be a continuous LSIP. Show that the function $\varphi(x) := \min_{y \in Y}\{a(y)^T x - b(y)\}$ is concave. So LSIP can be written as a convex program:* $\min\ c^T x$ *s.t.* $\varphi(x) \geq 0$.

Assume from now on that LSIP is continuous. We also write (16) formally in the familiar form

$$\text{(P)} \qquad \min_x \ c^T x \qquad \text{s.t.} \qquad Ax \geq b \ ,$$

where A is a matrix with infinitely many rows a_y^T, $y \in Y$ and b is a vector with infinitely many components b_y. We now define the (so-called Haar) *dual* of (P) as

$$\text{(D)} \qquad \max_u \ b^T u \qquad \text{s.t.} \qquad A^T u = c, \quad u \geq 0 \ ,$$

with the following understanding: $u = (u_y)$ is *dual feasible* if $u_y \geq 0$, $y \in Y$ and $u_y > 0$ for only *finitely many* $y \in Y$. So $b^T u = \sum_y b_y u_y$ is actually a finite sum and so is $A^T u = \sum_y a_y u_y$. Note that, by definition,

$$\text{(17)} \qquad \text{(D) is feasible} \quad \Longleftrightarrow \quad c \in \text{cone}\,\{a_y \mid y \in Y\}.$$

The optimal objective function values of (P) and (D) will be denoted by v_P and v_D.

REMARK. Recall from Caratheodory's Lemma 7 that the sum $c = \sum_y a_y u_y$ can be expressed as sums with at most n non-zero coefficients u_y.

As in LP, if $x \in \mathbb{R}^n$ and $u = (u_y)$ are primal resp. dual feasible, then

$$c^T x - u^T b = u^T(Ax - b) = \sum_y u_y(a_y^T x - b_y) \geq 0.$$

THEOREM 4. (Weak duality, complementary slackness) *If x and u are primal resp. dual feasible, then $c^T x \geq b^T u$. If $c^x = b^T u$ then x resp. u are optimal with $v_P = v_D$.*

We emphasize that in contrast to ordinary (finite) linear programs, linear semi-infinite problems do not necessarily have the strong duality property unless additional constraint qualifications hold (*cf.* Ex.6). A further notable difference to finite linear programming is the fact that the existence of primal and dual feasible solutions need not imply the existence of optimal solutions (*cf.* Ex. 7). To assure *e.g.* the existence of an optimal solution of (P) we may assume strict feasibility of the dual program.

The *Slater constraint qualification* for (P) assumes the existence of a strictly feasible primal solution (*Slater point*, see also Ex. 4):

(SC$_P$) There exists some $\hat{x} \in \mathbb{R}^n$ with $A\hat{x} > b$.

We also strengthen the dual feasibility condition (17) to the *dual Slater condition*:

(SC$_D$) $c \in \text{int cone}\,\{a_y \mid y \in Y\}$.

One can show that a feasible LSIP has an minimal solution if the condition SC$_D$ is fulfilled. This leads to the following strong duality result (see *e.g.*, [6] for a proof).

THEOREM 5 (Strong duality)**.** *Assume that the Slater conditions (SC$_P$) and (SC$_D$) are satisfied, i.e., (P) and (D) are strictly feasible. Then optimal primal and dual solutions $\overline{x}$ and $\overline{u}$ exist, and $v_P = v_D$ holds.*
Moreover in this case $\overline{x}$ is an optimal solution of (P) if and only if $\overline{x}$ is a KKT-point, i.e., *there exist a dually feasible $\overline{u}$ satisfying $\overline{u}_{y_j} \geq 0$ for $y_j \in Y_0(\overline{x})$, $j = 1, \dots, k$, $k \leq n$, and $\overline{u}_y = 0$ otherwise, such that*

$$c = \sum_{j=1}^{k} \overline{u}_{y_j} a_{y_j} \ .$$

In particular $\overline{u}^T(A\overline{x} - b) = 0$.

EX. 6. *Let $\kappa > 0$ be fixed. Show $v_P = 0$ and $v_D = \kappa$ and that neither the primal nor the dual Slater condition is satisfied for the linear (SIP)*

$$\min\ -x_2 \quad s.t. \quad -y^2 x_1 - y x_2 \geq 0\ , \quad y \in Y = [0, 1] \text{ and } x_2 \leq \kappa\ .$$

EX. 7. *Show that $\mathcal{F} = \{(x_1, x_2) \mid x_2 \geq e^{x_1} \text{ for } x_1 \leq 0 \text{ and } x_2 \geq x_1 + 1 \text{ for } x_1 \geq 0\}$ is the feasible set of the (LSIP)*

$$(P) \qquad \min\ x_2 \quad s.t. \quad -y x_1 + x_2 \geq y(1 - \ln y), \quad y \in Y = [0, 1]\ .$$

Show furthermore: (D) is feasible and (P) satisfies the primal Slater condition but does not have an optimal solution. (Consequently, (SC$_D$) cannot be satisfied).

5.3. **Linear semidefinite programs.** A *semidefinite program* (SDP) is of the type

$$(P): \quad \min_{x\in\mathbb{R}^n} c^T x \qquad \text{s.t.} \qquad \sum_{i=1}^{n} A_i x_i - B \succeq 0 \quad \text{(positive semidefinite)}, \tag{18}$$

where $B, A_1, \dots, A_n \in \mathbb{S}^{k\times k}$ are given $k \times k$ matrices. $\mathbb{S}^{k\times k}$ denotes the set of (real) symmetric $k \times k$ matrices. By the following trick a semidefinite program can be transformed into an LSIP. Recall that for $S = (s_{ij}) \in \mathbb{S}^{k\times k}$ the relation

$$S \succeq 0 \quad \Leftrightarrow \quad y^T S y = \sum_{i,j} s_{ij} y_i y_j \geq 0 \ \forall y \in Y := \{y \in \mathbb{R}^k \mid \|y\| = 1\} \tag{19}$$

holds. So we may state (18) equivalently as

$$\min_x c^T x \qquad \text{s.t.} \qquad a(y)^T x - b(y) \geq 0 \quad \forall y \in Y \tag{20}$$

with

$$a(y)^T = \left(y^T A_1 y, \dots, y^T A_n y\right) \quad \text{and} \quad b(y) = y^T B y \,.$$

Let us denote the inner product of $(k \times k)$-matrices $C = (c_{ij})$ and $D = (d_{ij})$ by $C \circ D = \sum_{i,j=1}^{k} c_{ij} d_{ij}$. Then it is not difficult to see that the LSIP dual of (20) can be written in the form (see [6])

$$(D): \quad \max_U \ B \circ U \qquad \text{s.t.} \quad A_i \circ U = c_i,\ i = 1, \dots, n, \qquad U \succeq 0. \tag{21}$$

The primal Slater condition takes the form:

$$(\mathrm{SC}_P) \qquad \text{There exists } \hat{x} \in \mathbb{R}^n \quad \text{s.t.} \quad \sum A_i \hat{x}_i - B \succ 0,$$

where $S \succ 0$ indicates that S is a positive definite matrix. Moreover it is not difficult to show (*cf.* [6]) that the dual slater condition can be written as

$$(\mathrm{SC}_D) \qquad \text{there exists } U \succ 0 \text{ such that } A_i \circ U = c_i,\ i = 1, \dots, n.$$

So in principle, optimality and duality results for SDP can directly be deduced from the corresponding results in LSIP. We obtain in this way (see [6] for a proof)

COROLLARY 1 (Strong duality)**.** *Assume that the Slater conditions (SC_P) and (SC_D) are satisfied, i.e., (P) and (D) are strictly feasible. Then optimal primal and dual solutions $\overline{x}$ and $\overline{U}$ exist, and $v_P = v_D$ holds.*
Moreover in this case $\overline{x}$ is an optimal solution of (P) if and only if there exist a dually feasible $\overline{U}$ satisfying $\overline{U} \cdot (\sum_i A_i \overline{x}_i - B) = 0$.

As a consequence of the previous considerations, the optimality conditions and duality results for these special cone constrained programs are easily obtained from the general theory for LSIP (*cf. e.g.* [6]).

Second order Cone programming. Similarly we discuss *second order conic programs*

$$(\mathrm{SOP}) \qquad \min\ c^\top x \quad \text{s.t.} \quad u := Ax - b \in L^m,$$

where L^m denotes the Lorentz cone

$$L^m := \{u \in \mathbb{R}^m \mid u_m \geq (u_1^2 + \ldots + u_{m-1}^2)^{1/2}\}.$$

Define $\tilde{u} = (u_1, \ldots, u_{m-1})$ and observe the identity

$$\|\tilde{u}\| = \max_{\tilde{y} \in \mathbb{R}^{m-1},\ \|\tilde{u}\|=1} \tilde{y}^\top \tilde{u}.$$

So the condition $u \in L^m$, or $u_m \geq \|\tilde{u}\|$, can be written as $u_m - \tilde{y}^\top \tilde{u} \geq 0\ \forall \|\tilde{y}\| = 1$, and the feasibility condition in SOP reads:

$$y^\top (Ax - b) \geq 0 \quad \text{or} \quad a(y)^T x \geq b(y) \quad \forall y \in Y = \{y = (\tilde{y}, 1) \mid \|\tilde{y}\| = 1\},$$

with $a_i(y) = y^\top A_i$, A_i the i-th column of A and $b(y) = y^\top b$.

5.4. **Complexity issues.** As a first general observation we emphasize that, from the numerical viewpoint, SIP is much more difficult than FP. The main reason is the difficulty associated with the feasibility test for $\overline{x}$. In a finite program,

$$FP: \qquad \min_x \ f(x) \qquad \text{s.t.} \quad g_j(x) \geq 0 \quad \forall j \in J = \{1, 2, \ldots, m\},$$

we only have to compute m function values $g_j(\overline{x})$ and to check whether all these values are nonnegative. In SIP, checking the feasibility of $\overline{x}$ is obviously equivalent to solve the global minimization problem $Q(\overline{x})$ in the y variable:

$$Q(\overline{x}): \qquad \min_y \ g(\overline{x}, y) \quad \text{s.t.} \quad y \in Y, \tag{22}$$

and to check whether for a global solution $\overline{y}$ the condition $g(\overline{x}, \overline{y}) \geq 0$ holds.
Note that, even for the LSIP, the problem $Q(\overline{x})$ is not in general a convex problem. As a consequence of this fact, the LSIP problem cannot be expected to be solvable in polynomial time. However, there are special subclasses of linear or convex semi-infinite programs which can be solved polynomially. Interesting examples are semidefinite and second order cone programming, as well as certain classes of robust optimization problems [2]. We refer to [25] for an extensive treatment of this issue.
Here for shortness we will try to motivate the fact that SDP can be solved efficiently from the viewpoint of the ellipsoid method. Firstly, we emphasize that it is not difficult to see that a convex program can be solved efficiently if the corresponding feasibility problem can be solved efficiently. To find a feasible point $\overline{x} \in \mathcal{F}$, the ellipsoid method proceeds as follows.

Ellipsoid Method

INIT: Start with a (sufficiently large) ellipsoid E_0 so that $\mathcal{F} \cap E_0 \neq \emptyset$.

ITER: Given the ellipsoid $E_j = \{x \mid (x - x_j)^T A_j (x - x_j) \leq 1\}$, $A_j \succ 0$, such that
$\mathcal{F} \cap E_j = \mathcal{F} \cap E_0$

(1) if $x_j \in \mathcal{F}$ STOP

(2) OTHERWISE: compute a separating halfspace $H_j^{\geq}$ such that
$\mathcal{F} \subseteq H_j^{\geq}$, $x_j \notin H_j^{\geq}$.
and find (the smallest) ellipsoid E_{j+1} with vol $E_{j+1} \leq q$ vol E_j $(0 < q < 1)$
containing $E_j \cap H_j^{\geq}$ and thus $\mathcal{F} \cap E_0$.

So the Ellipsoid Method generates ellipsoids $E_0, E_1, \ldots$ with the property

$$\text{vol } (\mathcal{F} \cap E_0) \le \text{vol } E_j << q^j \text{vol } E_0 \to 0 \,.$$

Hence, if $0 < \text{vol } (\mathcal{F} \cap E_0)$, it is clear that the algorithm will find a point $x_j \in \mathcal{F}$ after a finite number of iterations.
The overall ellipsoid method can be shown to be polynomial if the feasibility test $\overline{x} \in \mathcal{F}$ in step (1) and the construction of a separating halfspace in step (2) can be done in polynomial time. To consider the special case of SDP let $G(x) := \sum_{i=1}^n A_i x_i - B \in \mathbb{S}^{k \times k}$.

LEMMA 4. *The feasibility test $G(\overline{x}) \succeq 0$ and the construction of a separating halfspace can be done in time $O(k^3)$.*

Proof. It is well-known that by applying a modification of the Gauss algorithm to $\overline{G} = G(\overline{x})$ we can compute, in time $O(k^3)$, a decomposition

$$Q\overline{G}Q^T = D \qquad \text{with } Q \text{ regular and } D = \text{diag } (d_1, \ldots, d_k),$$

such that $\overline{G} \succeq 0$ if and only if $d_i \ge 0, \ \forall i$. So if $d_i \ge 0, \ \forall i$, the point $\overline{x}$ is feasible. If not say $d_1 < 0$, than by defining $a_1 = Q^T e_1$ (e_1 the unit vector) it follows for all feasible x, *i.e.*, $G(x) \succeq 0$:

$$a_1^T G(x) a_1 \ge 0 > d_1 = e_1^T D e_1 = e_1^T Q\overline{G}Q^T e_1 = a_1^T \overline{G} a_1 \,.$$

So the linear inequality $a_1^T G(x) a_1 = \sum_{i=1}^n (a_1^T A_i a_1) x_i - a_1^T B a_1 \ge 0$ defines a separating half-space.

□

Ex. Show for SOC: The feasibility test and the construction of a separating hyperplane for some $u \notin L^m$ is easily done.

Special case g linear in y. We consider the special case of a SIP with linear (or convex) objective and constraint function of the form

$$g(x, y) = a(x)^T y - b(x) \ge 0 \quad \forall y \in Y$$

So x is feasible if:

$$\varphi(x) := \min_{y \in Y} \ a(x)^T y - b(x) \ge 0$$

Suppose Y is given by a ball say $Y = \{y \mid \|y\| \le 1\}$. By observing the relation

$$\min_{\|y\|=1} a^T y - b \quad = \quad a^T \left[-\frac{a}{\|a\|} \right] - b = -\|a\| - b \ge 0$$

we find by using a result of Schur:

$$-\|a\| \ge b \Leftrightarrow \begin{pmatrix} -b \cdot I & a \\ a^T & -b \end{pmatrix} \text{psd.}$$

Consequently if $a(x) = Ax, \ b(x) = b^T x$ are linear in x the feasibility test $\varphi(x) \ge 0$ becomes the constraint of a SDP, and the SIP problem is equivalent to a (linear) semidefinite program.

Ex. Suppose that Y is given by (the ellipsoid) $Y = \{y \in \mathbb{R}^m \mid (y - y_0)^T Q^{-1}(y - y_0) \leq 1\}$ with a positive definite matrix $Q = CC^T$. Then the constraint

$$\varphi(x) := \min_{y \in Y} a(x)^T y - b(x) \geq 0$$

is equivalent to the SDP constraint

$$\begin{pmatrix} -\beta(x) \cdot I & Ca(x) \\ a(x)^T C & -\beta(x) \end{pmatrix} \quad \text{is positive semidefinite}$$

where $\beta(x) = (b(x) - a(x)^T y_0)$

In the case of SOC programming the feasibility test (and the construction of a separating hyperplane) is trivially polynomial.

6. A Primal-Dual Interior Point Method

It is well-known that interior point methods lead to polynomial solution methods for solving LP problems.
On has tried to extend this method to LSIP. In this section we will discuss such an extension. Essentially it is the approach discussed in Tuncel/Todd [24]. We however present this approach in a different and shorter form. We describe a version of the interior point method based on the pair of primal and dual programs.
We will see however that the extension to LSIP (at the current state) seem only of theoretical value because it relies (see also [24]) on an assumption which cannot be verified a priori.
Consider the linear program

$$(P) \qquad \max\ \mathbf{c}^T\mathbf{x} \quad s.t. \quad \mathbf{A}\mathbf{x} + \mathbf{s} = \mathbf{b} \tag{23}$$

and its dual

$$(D) \qquad \min\ \mathbf{b}^T\mathbf{z} \quad s.t. \quad \mathbf{A}^T\mathbf{z} = \mathbf{c} \quad and \quad \mathbf{z} \geq \mathbf{0} \tag{24}$$

with $\mathbf{A} \in \mathbb{R}^{m \times n}$, $\mathbf{c} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^m$ (m constraints).

The strong duality property of linear programs says that feasible solutions $\mathbf{x}$ for (23) and $\mathbf{z}$ for (24) are optimal if and only if they satisfy the complementary slackness condition $z_j s_j = 0$ for $j = 1, \ldots, m$. So we seek to find $\mathbf{z} \geq \mathbf{0}$, $\mathbf{s} \geq \mathbf{0}$ and $\mathbf{x}$ solving the system of equations

$$\begin{array}{rcl} \mathbf{A}\mathbf{x} + \mathbf{s} & = & \mathbf{b} \\ \mathbf{A}^T\mathbf{z} & = & \mathbf{c} \\ z_j s_j & = & 0\,, \quad j = 1, \ldots, m. \end{array} \tag{25}$$

Suppose we can find feasible solutions $\mathbf{x}$ and $\mathbf{y}$ that satisfy complementary slackness ”approximately” in the sense that

$$z_j s_j = \mu \qquad \text{for } j = 1, \ldots, m \tag{26}$$

holds for some $\mu > 0$. Then the solutions might still be acceptably good if μ is sufficiently small ($\mu < \varepsilon/m$, say). Indeed, we observe for the associated duality gap:

$$0 \leq \mathbf{b}^T\mathbf{z} - \mathbf{c}^T\mathbf{x} = \mathbf{z}^T\mathbf{s} = \sum_{j=1}^{m} z_j s_j = m\mu < \varepsilon\ . \tag{27}$$

Note that $\mu > 0$ in (26) implies $z_j > 0$ and $s_j > 0$ for all j, *i.e.*, $\mathbf{z} > \mathbf{0}$ and $\mathbf{s} > \mathbf{0}$. So we are dealing with primal and dual solutions that lie in the (relative) *interior* of the respective feasibility regions.

We therefore want to determine a solution $\mathbf{z} > \mathbf{0}$, $\mathbf{s} > \mathbf{0}$, and $\mathbf{x}$ of the *perturbed system*

$$\begin{array}{rcl} \mathbf{A}\mathbf{x} + \mathbf{s} & = & \mathbf{b} \\ \mathbf{A}^T\mathbf{z} & = & \mathbf{c} \\ z_j s_j & = & \mu\ , \quad j = 1, \ldots, m. \end{array} \tag{28}$$

for the parameter $\mu > 0$.

In the following, we assume that there exist *strongly feasible* solutions $\mathbf{z}$, $\mathbf{z}$ of (23) and (24), *i.e.*, $\mathbf{z} > \mathbf{0}$ and $\mathbf{s} = \mathbf{c} - \mathbf{A}^T\mathbf{x} > \mathbf{0}$. Furthermore, we always assume that the matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$ has full rank n (after having removed redundant equations if necessary). So $\mathbf{x}$ is uniquely determined by $\mathbf{s}$.

Remark. Assuming that strictly feasible primal and dual solutions exist it can be shown that for any $\mu > 0$ there is a unique solution $(\mathbf{x}(\mu), \mathbf{z}(\mu), \mathbf{s}(\mu))$ (*central path*) of (28)

The idea of the method is to construct a sequence of values $\mu_k \downarrow 0$ together with (approximate) solutions $(\mathbf{x}_k, \mathbf{y}_k, \mathbf{s}_k)$ of (28) for $\mu = \mu_k$.

We now extend the previous analysis to linear semi-infinite programs and introduce the short hand notation

$$\begin{array}{ll} Ax \equiv a(j)^T x & b \equiv b(j) = b_j \quad j \in Y \\ b^T z \equiv \int_Y b_j z_j \, dj & P := \int_Y 1 \cdot dj \end{array}$$

Here for convenience, we have replaced the index variable y by j. Moreover (as in [24]) we assume $a_i, s, z \in C(Y)$ and that the functions $a_1(j), \ldots, a_n(j)$ are linear independent on Y. Note that at least for the limit $\mu = 0$ the assumption $z \in C(Y)$ is unrealistic because we expect the solution z of D to be a point functional.
With this definitions, P and D also represent a pair of primal and dual linear semi-infinite programs. Moreover the equations (25) and (28) make sense for LSIP if the third equation (*i.e.*, (26) is replaced by $z_j s_j = 0$ (or $= \mu$) for all $j \in Y$. The weak duality relation (27) then becomes

$$0 \le b^T z - c^T x = z^T s = \int_Y z(j)s(j)dj = P\mu$$

In what follows we give the whole proof for the LP case and only note the modifications needed for LSIP.

6.1. **Newton Steps.** We wish to solve the system (28) for fixed μ. The problem is that, due to the quadratic equations $z_j s_j = \mu$, the system is nonlinear. So we may apply the Newton method.

At this point, it is useful to introduce the following shorthand notation for any vectors $\mathbf{u}, \mathbf{w} \in \mathbb{R}^n$:

$$\begin{array}{rcll} \mathbf{uw} & = & (u_1 w_1, \ldots, u_m w_m)^T & \mid\ = u_j w_j \\ \mathbf{w}^{-1} & = & (1/w_1, \ldots, 1/w_m)^T & \mid\ = 1/w_j \\ \sqrt{\mathbf{u}} & = & (\sqrt{u_1}, \ldots, \sqrt{u_m})^T & \mid\ = \sqrt{u_j} \\ \boldsymbol{\mu} & = & (\mu, \ldots, \mu) & \mid\ = \mu \end{array} \quad j \in Y \text{ in case of LSIP}$$

With this notation, for example, we can write $\mathbf{u}^T\mathbf{w} = \mathbf{1}^T(\mathbf{uw})$.

The Newton method for solving the system (28) consists of computing solutions $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$ of the linear equation

$$(29)\qquad \begin{aligned} \mathbf{A}\Delta\mathbf{x}+\Delta\mathbf{s} &= \mathbf{0}\\ \mathbf{A}^T\Delta\mathbf{z} &= \mathbf{0}\\ \mathbf{s}\Delta\mathbf{z}+\mathbf{z}\Delta\mathbf{s} &= \boldsymbol{\mu}-\mathbf{zs}. \end{aligned}$$

leading to the next iterate $(\mathbf{x}^+, \mathbf{z}^+, \mathbf{s}^+) = (\mathbf{x}, \mathbf{z}, \mathbf{s}) + (\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$.

PROPOSITION 1. *For every $\mathbf{z} > \mathbf{0}$ and $\mathbf{s} > \mathbf{0}$, the system (29) of linear equations admits a unique solution $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$.*

Proof. The first two equations in (29) state $\Delta\mathbf{z} \in \ker \mathbf{A}^T$ and $\Delta\mathbf{s} \in \text{row } \mathbf{A}^T$. The third equation is equivalent with

$$\sqrt{\mathbf{sz}^{-1}}\Delta\mathbf{z} + \sqrt{\mathbf{zs}^{-1}}\Delta\mathbf{s} = \frac{\boldsymbol{\mu}-\mathbf{zs}}{\sqrt{\mathbf{zs}}}\ .$$

Pre-multiplying the vectors in each of the subspaces $\ker \mathbf{A}^T$ and row $\mathbf{A}^T$ with the fixed positive vectors $\sqrt{\mathbf{sz}^{-1}}$ resp. $\sqrt{\mathbf{zs}^{-1}}$, we obtain the linear subspaces

$$(30)\qquad \begin{aligned} U &= \{\sqrt{\mathbf{sz}^{-1}}\Delta\mathbf{z} \mid \Delta\mathbf{z} \in \ker \mathbf{A}^T\} \quad \textit{with} \quad \dim\ U = m - \text{rank}\ \mathbf{A}\\ V &= \{\sqrt{\mathbf{zs}^{-1}}\Delta\mathbf{s} \mid \Delta\mathbf{s} \in \text{row}\ \mathbf{A}^T\} \quad \textit{with} \quad \dim\ V = \text{rank}\ \mathbf{A}. \end{aligned}$$

Since $\ker \mathbf{A}^T$ and row $\mathbf{A}^T$ are complementary orthogonal subspaces, so are U and V. Indeed, for $\mathbf{u} = \sqrt{\mathbf{sz}^{-1}}\Delta\mathbf{z}$ and $\mathbf{v} = \sqrt{\mathbf{zs}^{-1}}\Delta\mathbf{s}$ we have $u_j v_j = \Delta z_j \Delta s_j$ and hence

$$(31)\qquad \mathbf{uv} = \Delta\mathbf{z}\Delta\mathbf{s} \qquad \text{and} \qquad \mathbf{u}^T\mathbf{v} = \Delta\mathbf{z}^T\Delta\mathbf{s} = 0.$$

So every vector in $\mathbb{R}^m$ can be (uniquely) written as $\mathbf{u}+\mathbf{v}$ with $\mathbf{u} \in U$ (its projection onto U) and $\mathbf{v} \in V$ (its projection onto $V = U^\perp$). In particular, there exist $\mathbf{u} \in U$ and $\mathbf{v} \in V$ with

$$(32)\qquad \mathbf{u}+\mathbf{v} = \frac{\boldsymbol{\mu}-\mathbf{zs}}{\sqrt{\mathbf{zs}}}.$$

Hence also $\Delta\mathbf{z}$ and $\Delta\mathbf{s}$ are uniquely determined, and $\Delta\mathbf{z}$ follows (uniquely) from $\mathbf{A}^T\Delta\mathbf{z} = -\Delta\mathbf{s}$. (Recall that $\mathbf{A}$ has full row rank n by assumption).
The proof in the LSIP case follows in a similar way by using a result in [3, Th.4.1].

The triple $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$ is said to be a *Newton step* if it solves (29). The Newton step is *feasible* if

$$\mathbf{z}^+ = \mathbf{z}+\Delta\mathbf{z} > \mathbf{0} \qquad \text{and} \qquad \mathbf{s}^+ = \mathbf{s}+\Delta\mathbf{s} > \mathbf{0}\ .$$

Convergence of Newton Steps. Ideally, we would like to have the Newton step $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$ to yield $\mathbf{z}^+\mathbf{s}^+ = \boldsymbol{\mu}$. However, we only obtain

$$(33)\qquad \mathbf{z}^+\mathbf{s}^+ = \boldsymbol{\mu} + \Delta\mathbf{z}\Delta\mathbf{s} \quad (= \boldsymbol{\mu} + \mathbf{uv} \text{ by } (31))$$

As a consequence of (33), we observe

$$\begin{aligned} (34)\quad (\mathbf{z}^+)^T\mathbf{s}^+ = \mathbf{1}^T(\mathbf{z}^+\mathbf{s}^+) = \mathbf{1}^T\boldsymbol{\mu} + \mathbf{1}^T(\Delta\mathbf{z}\Delta\mathbf{s}) = m\mu + \Delta\mathbf{z}^T\Delta\mathbf{s} &= m\mu\\ (35)\quad &= P\mu \text{ for LSIP.} \end{aligned}$$

Hence, provided the Newton step is feasible, $(\mathbf{z}^+, \mathbf{s}^+)$ yields a strictly feasible primal-dual pair of solutions with duality gap $m\mu$ or $P\mu$ (even if $\mathbf{z}^+\mathbf{s}^+ \neq \boldsymbol{\mu}$).

The next inequality between the max-norm and the Euclidian norm is essential for obtaining the polynomiality of the interior point method. For $\mathbf{u}, \mathbf{v} \in \mathbb{R}^m$ it trivially follows

$$|u_j v_j| \le (\sum (u_j v_j)^2)^{1/2} = \|\mathbf{uv}\| \ .$$

Unfortunately such a bound is not generally valid in (the infinite dimensional space) $C(Y)$. For the LSIP case we however have to make the following assumption (see also [24]) which cannot be verified a priori.

AS. All vectors $\mathbf{u} = \sqrt{\mathbf{sz}^{-1}}\Delta\mathbf{z}$ and $\mathbf{v} = \sqrt{\mathbf{zs}^{-1}}\Delta\mathbf{s}$ (satisfying $u_j v_j = \Delta z_j \Delta s_j$) computed during the interior point method satisfy with some constant C, $1 \le C < \infty$:

$$|u_j v_j| \le C\|\mathbf{uv}\| \quad \text{for all } j \in Y \ . \tag{36}$$

We now measure the quality of an approximation $\mathbf{zs} \approx \boldsymbol{\mu}$ by taking the relative (squared) Euclidean distance from the central path

$$\delta \ = \ \delta(\mathbf{z}, \mathbf{s}, \mu) \ = \ \frac{C^2}{\mu} \left\| \frac{\mathbf{zs} - \boldsymbol{\mu}}{\sqrt{\mathbf{zs}}} \right\|^2 \ = \ \frac{C^2}{\mu} \|\mathbf{u} + \mathbf{v}\|^2 \qquad \text{with} \qquad \begin{array}{l} C = 1 \text{ for LP} \\ C \text{ as in (36) for LSIP} \end{array}$$

where $\mathbf{u} \in U$ and $\mathbf{v} \in V$ are as in (32). We can now make our previous remark precise: If $\mathbf{zs} \approx \boldsymbol{\mu}$, then $\mathbf{z}^+\mathbf{s}^+$ is an even better approximation to $\boldsymbol{\mu}$.

THEOREM 6. *Let AS hold for LSIP. Then for $\delta = \delta(\mathbf{z}, \mathbf{s}, \mu) \le 1$ in the Newton iteration it follows:*

(a) $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$ *is a feasible Newton step.*
(b) $\delta^+ \ = \ \delta(\mathbf{z}^+, \mathbf{s}^+, \mu) \ \le \ \frac{1}{2}\delta^2$.

Proof. By the auxiliary Lemma 5 below, we have for each component j (in both cases LP ($C = 1$) and LSIP (see (36))

$$|\Delta z_j \Delta s_j| = |u_j v_j| \le C\|\mathbf{uv}\| \le \frac{C}{2}\|\mathbf{u} + \mathbf{v}\|^2 = \frac{1}{2C}\mu\delta \le \mu/(2C).$$

Hence using $C \ge 1$, (33) yields $\mathbf{z}^+\mathbf{s}^+ \ge \boldsymbol{\mu}/2$, which implies (again by Lemma 5 below):

$$\delta^+ = \frac{C^2}{\mu} \left\| \frac{\mathbf{z}^+\mathbf{s}^+ - \boldsymbol{\mu}}{\sqrt{\mathbf{z}^+\mathbf{s}^+}} \right\|^2 \le \frac{C^2}{\mu} \left\| \frac{\mathbf{uv}}{\sqrt{\mu/2}} \right\|^2 \le \frac{2C^2}{\mu^2} \cdot \frac{1}{4}\|\mathbf{u} + \mathbf{v}\|^4 = \frac{1}{2}\delta^2 \ .$$

To show feasibility of the Newton step, suppose to the contrary that $z_j^+ < 0$ holds. In view of $\mathbf{z}^+\mathbf{s}^+ \ge \boldsymbol{\mu}/2 > 0$, then also $s_j^+ < 0$ must hold. But $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$ is a Newton step. (29) therefore implies

$$\mu = z_j s_j + z_j \Delta s_j + s_j \Delta z_j = z_j s_j^+ + s_j z_j^+ - s_j z_j < 0 \ ,$$

a contradiction.

◇

LEMMA 5. *Let $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$ (or $\in C(Y)$) be orthogonal. Then*

$$\|\mathbf{uv}\| \ \le \ \frac{1}{2}\|\mathbf{u} + \mathbf{v}\|^2 \ .$$

Proof. The general identity $4\alpha\beta = (\alpha+\beta)^2 - (\alpha-\beta)^2$ yields

$$4|u_j v_j| = |(u_j + v_j)^2 - (u_j - v_j)^2| \le (u_j + v_j)^2 + (u_j - v_j)^2 \,.$$

Summation over $j = 1, \dots, n$ (or integration over $j \in Y$) gives

$$4\|\mathbf{uv}\| \le 4 \sum_j |u_j v_j| \le \|\mathbf{u}+\mathbf{v}\|^2 + \|\mathbf{u}-\mathbf{v}\|^2 \,.$$

The claim follows by observing that $\|\mathbf{u}+\mathbf{v}\| = \|\mathbf{u}-\mathbf{v}\|$ for orthogonal **u** and **v**.

$\diamond$

6.2. **The Algorithm IPM.** Theorem 6 has important algorithmic implications. Statement (b) indicates that $\delta(\mathbf{z}, \mathbf{s}, \mu)$ converges quickly to 0 (and hence $(\mathbf{x}, \mathbf{z}, \mathbf{s})$ to a solution of (28)). Since we really would like to solve (28) with $\mu > 0$ as small as possible, the question arises whether we could decrease the parameter μ after each Newton iteration.

Consider an arbitrary $0 < \theta < 1$ and assume $\delta(\mathbf{z}, \mathbf{s}, \mu) \le 1$. By (34) we have $(\mathbf{z}^+)^T \mathbf{s}^+ = \|\sqrt{\mathbf{z}^+\mathbf{s}^+}\|^2 = m\mu$ (or $= P\mu$), so we compute in the LP case :

$$\begin{aligned}
\delta(\mathbf{z}^+, \mathbf{s}^+, \theta\mu) &= \frac{1}{\theta\mu}\left\| \frac{\mathbf{z}^+\mathbf{s}^+ - \theta\boldsymbol{\mu}}{\sqrt{\mathbf{z}^+\mathbf{s}^+}} \right\|^2 \\
&= \frac{1}{\theta\mu}\left\| (1-\theta)\sqrt{\mathbf{z}^+\mathbf{s}^+} + \theta \frac{\mathbf{z}^+\mathbf{s}^+ - \boldsymbol{\mu}}{\sqrt{\mathbf{z}^+\mathbf{s}^+}} \right\|^2 \\
&= \frac{1}{\theta\mu}\|(1-\theta)\sqrt{\mathbf{z}^+\mathbf{s}^+}\|^2 + \frac{1}{\theta\mu}\left\| \theta \frac{\mathbf{z}^+\mathbf{s}^+ - \boldsymbol{\mu}}{\sqrt{\mathbf{z}^+\mathbf{s}^+}} \right\|^2 \\
&= \frac{(1-\theta)^2}{\theta} m + \theta\delta(\mathbf{z}^+, \mathbf{s}^+, \mu) \\
&\le \frac{(1-\theta)^2}{\theta} m + \frac{\theta}{2} \qquad \text{(by Theorem 6).}
\end{aligned}$$

(The third equation above uses $\|\mathbf{a}+\mathbf{b}\| = \|\mathbf{a}\| + \|\mathbf{b}\|$ for $\mathbf{a} \perp \mathbf{b}$.) For LSIP we obtain the same in equalities with m replaced by $m = C^2 P$.

For the choice $\theta^* = 1 - \dfrac{1}{m}$ we find the bound

$$\delta(\mathbf{z}^+, \mathbf{s}^+, \theta\mu) \le \frac{1}{m-1} + \frac{\theta^*}{2} \le 1 \qquad (\text{if } m > 2)\,. \tag{37}$$

So Theorem 6 guarantees that the Newton steps will remain feasible even when we reduce the current value of $\mu > 0$ by the factor θ^* in each iteration!

This discussion suggests the following algorithm, which will compute an approximately optimal solution of the linear program (23). To start the algorithm, we assume to have at our disposal a primal feasible vector $\mathbf{x}_0$, a vector $\mathbf{z}_0$ and a parameter $\mu_0 > 0$ with the properties

(i) $\mathbf{z}_0 > \mathbf{0}$.
(ii) $\mathbf{s}_0 = \mathbf{b} - \mathbf{A}\mathbf{x}_0 > \mathbf{0}$.
(iii) $\delta(\mathbf{z}_0, \mathbf{s}_0, \mu_0) \le 1$.

In every iteration, the algorithm computes a feasible Newton step $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$ with respect to the current parameters $(\mathbf{z}, \mathbf{s}, \mu)$ and then reduces μ by the factor θ^*. As a stopping criterion, we use a precision parameter $\varepsilon > 0$ (which can be chosen freely).

Algorithm IPM

INIT: $\mathbf{z} \leftarrow \mathbf{z}_0$, $\mathbf{s} \leftarrow \mathbf{s}_0$, $\mu \leftarrow \mu_0$;
ITER: Compute a solution $(\Delta\mathbf{x}, \Delta\mathbf{z}, \Delta\mathbf{s})$ of (29) and update
$\mathbf{z} \leftarrow \mathbf{z} + \Delta\mathbf{x}$;
$\mathbf{s} \leftarrow \mathbf{s} + \Delta\mathbf{s}$;
IF $m\mu \leq \varepsilon$ (or $P\mu \leq \varepsilon$) STOP, ELSE $\mu \leftarrow \theta^* \mu$.

By (37) and Theorem 6 every Newton step is feasible. When the algorithm stops, the current solution satisfies

$$\mathbf{b}^T\mathbf{z} - \mathbf{c}^T\mathbf{x} \;=\; \mathbf{z}^T\mathbf{s} \;=\; m\mu \;\leq\; \varepsilon \quad (\text{or} = \; P\mu \;\leq\; \varepsilon \text{ for LSIP}) \; .$$

So we have found a solution $\mathbf{x}$ for the linear program (23) whose objective function value $\mathbf{c}^T\mathbf{x}$ differs by at most ε from the optimum. We call such a solution ε-*approximate*.

The total number K of iterations of algorithm IPM can be estimated. From the general inequality $1 + x \leq e^x$ we find $(\theta^*)^m = (1 - 1/m)^m \leq e^{-1}$. So μ shrinks by at least a factor e^{-1} in any m subsequent iterations. Consequently, after

$$\text{(38)} \qquad \boxed{K \;\leq\; m \cdot \ln \frac{\mu_0 m}{\varepsilon}} \qquad (\text{ with } m = C^2 P \text{ for LSIP})$$

iterations, the current $\mu = (\theta^*)^K \mu_0$ will satisfy the stopping criterion $\mu \leq \varepsilon/m$ (or $\mu \leq \varepsilon/P$ in the LSIP case). So we have obtained a polynomial (in the number of Newton steps) approximation algorithm, which for LSIP however assumes that AS holds.

7. DISCRETIZATION METHODS

In a discretization method we choose finite subsets Y' of Y, and instead of $P \equiv P(Y)$ we solve the finite programs

$$\text{P}(Y'): \quad \min \; f(x) \qquad \text{s.t. } \; g(x, y) \geq 0, \; \forall y \in Y' \; .$$

Let $v(Y')$, $\mathcal{F}(Y')$ and $S(Y')$ denote the the minimal value, the feasible set and the set of (global) minimizers of $P(Y')$ respectively. We introduce the Hausdorff distance (meshsize) $\rho(Y')$ between Y' and Y by

$$\rho(Y') := \max_{y \in Y} \text{dist}\,(y, Y') \quad \text{where} \;\; \text{dist}\,(y, Y') = \min_{y' \in Y'} \|y - y'\| \; .$$

The following relation is trivial but important:

$$\text{(39)} \qquad Y_2 \subset Y_1 \quad \Rightarrow \quad \mathcal{F}(Y_1) \subset \mathcal{F}(Y_2) \text{ and } v(Y_2) \leq v(Y_1) \; .$$

We consider the discretization concept: $\text{P}(Y)$ is said to be *discretizable* if for each sequence of finite grids $Y_k \subset Y$ satisfying $\rho(Y_k) \to 0$ (for k lage enough) there exist solutions $\overline{x}_k$ of $\text{P}(Y_k)$ and for each sequence of solutions the relation

$$dist(\overline{x}_k, S(Y)) \to 0 \quad \text{ and } \quad v(Y_k) \to v(Y)$$

holds. To treat non-convex problems we also introduce a local concept. Given a local minimizer $\overline{x}$ of $P(Y)$ the SIP is called locally discretizable at $\overline{x}$ if the discretizability relation holds locally, i.e. if the problem $P^l(Y)$,

$$P^l(Y): \quad \min_{x \in U_{\overline{x}}} f(x) \quad \text{s.t.} \quad g(x,y) \geq 0 \; \forall y \in Y \; ,$$

obtained as the restriction of $P(Y)$ to an open neighborhood $U_{\overline{x}}$ of $\overline{x}$, is discretizable.
We give an illustrative example.

EXAMPLE 1. Consider the linear SIP (with some fixed $\varepsilon > 0$):

$$\min \; x_1 \text{ s.t. } x_1 \cos y + x_2 \sin y \geq 1, \; y \in Y := \begin{cases} [\pi, 3\pi/2] & \text{case A} \\ [\pi - \varepsilon, 3\pi/2] & \text{case B} \end{cases}$$

The minimizer of P$(Y)$is $\overline{x} = (-1, 0)$.

Case A: The problem is not discretizable (only weakly discretizable). For a grid Y' containing $y = \pi$ we have $v(Y') = v(Y)$. On the other hand for any Y' not containing π the value is unbounded, $v(Y') = -\infty$.
Case B: The problem is discretizable as is easily shown. Note that in case B the condition $c \in \text{int cone}\,\{a(y) \mid y \in Y\}$ is satisfied but not in case A (*cf.* also Theorem 8).

The following algorithm is based on the concept of discretizability.

Algorithm 1 (*Conceptual discretization method*)

Step k: Given a discretization $Y_k \subset Y$

i. Compute a solution x_k of P(Y_k).
ii. Stop, if x_k is feasible within a fixed accuracy $\alpha > 0$, i.e. $g(x_k, y) \geq -\alpha$, $y \in Y$. Otherwise, select a finer discretization $Y_{k+1} \subset Y$.

Under a compactness assumption on the feasible sets we obtain a general convergence result for this method. We begin with linear and convex problems.

THEOREM 7. *A convex SIP is discretizable if the feasible set $F(Y)$ is compact.*

The following result on discretizability of LSIP is contained in [11, p. 70-75].

THEOREM 8. *Let the LSIP problem $P(Y)$ be feasible and assume that the condition $c \in$ int cone $\{a(y) \mid y \in Y\}$ holds. Then $\mathcal{S}(Y)$ is non-empty and bounded, so a solution of $P(Y)$ exists. Moreover $P(Y)$ is discretizable.*

The next result is valid for general SIP.

THEOREM 9. *Let the sequence of discretizations Y_k satisfy*

$$Y_0 \subset Y_k \; \forall k \geq 1 \textit{ and } \rho(Y_k) \to 0 \quad \textit{for } k \to \infty \; .$$

Suppose $\mathcal{F}(Y_0)$ is compact. Then $P(Y)$ is discretizable, i.e., *the problems $P(Y_k)$ have solutions x_k and each such sequence of solutions satisfies* dist $(x_k, \mathcal{S}(Y)) \to 0$.

Proof. By assumption and using $Y_0 \subset Y_k \subset Y$ the feasible sets $\mathcal{F}(Y)$, $\mathcal{F}(Y_k)$, of $P(Y)$, $P(Y_k)$ respectively, are compact and satisfy $\mathcal{F}(Y) \subset \mathcal{F}(Y_k) \subset \mathcal{F}(Y_0)$, $\quad k \in \mathbb{N}$. Consequently, solutions x_k of $P(Y_k)$ exist. Suppose now that a sequence of such solutions does not satisfy dist $(x_k, \mathcal{S}(Y)) \to 0$. Then there exist $\varepsilon > 0$ and a subsequence x_{k_ν} such that

$$\text{dist } (x_{k_\nu}, \mathcal{S}(Y)) \geq \epsilon > 0 \quad \forall \nu \; .$$

Since $x_{k_\nu} \in \mathcal{F}(Y_0)$ we can select a convergent subsequence. Without restriction we can assume $x_{k_\nu} \to \overline{x},\ \nu \to \infty$. In view of $\mathcal{F}(Y) \subset \mathcal{F}(Y_k)$ the relation $f(x_{k_\nu}) \leq v(Y)$ holds and thus by continuity of f we find

$$f(\overline{x}) \leq v(Y) \ .$$

We no show that $\overline{x} \in \mathcal{S}(Y)$ in contradiction to our assumption. To do so it suffices to prove that $\overline{x} \in \mathcal{F}(Y)$. Let $\overline{y} \in Y$ be given arbitrarily. Since $\rho(Y_{k_\nu}) \to 0$ for $\nu \to \infty$ we can choose $y_{k_\nu} \in Y_{k_\nu}$, such that $y_{k_\nu} \to \overline{y}$. In view of $g(x_{k_\nu}, y_{k_\nu}) \geq 0$, by taking the limit $\nu \to \infty$, it follows $g(\overline{x}, \overline{y}) \geq 0$, i.e. $\overline{x} \in \mathcal{F}(Y)$.

□

We now consider discretizability for general (also nonlinear) semi-infinite problems. Here we have to make use of the local concept. The following can be easily proven (e.g. in [22])

LEMMA 6. *Let be given a sequence of grids $Y_k \subset Y$ with $\rho_k := \rho(Y_k) \to 0$.*

(a) *Let x_k be points in $\mathcal{F}(Y_k) \cap K$, where K is a compact subset of $\mathbb{R}^n$. Then there exists $c > 0$ such that for all $\rho_k > 0$ small enough*

$$g(x_k, y) \geq -c\ \rho_k \quad \forall y \in Y \ .$$

(b) *Let MFCQ be satisfied at $\overline{x}$ with the vector d (cf. (12)). Then there exist numbers $\tau > 0,\ \varepsilon_1 > 0$ such that for small ρ_k the points $x + \tau\rho_k d$ are feasible for $P(Y)$ for all $x_k \in \mathcal{F}(Y_k)$ with $\|x_k - \overline{x}\| < \varepsilon_1$.*

THEOREM 10. *Let $\overline{x}$ be a local minimizer of $P(Y)$ of order $p \geq 1$. Suppose MFCQ holds at $\overline{x}$. Then P is locally discretizable at $\overline{x}$. More precisely, there is some $\sigma > 0$ such that for any sequence of grids $Y_k \subset Y$ with $\rho(Y_k) \to 0$ and any sequence of solutions x_k of the locally restricted problem $P^l(Y_k)$ (see the definition of discretizability) the following relation holds:*

$$0 \leq f(\overline{x}) - f(x_k) \leq O(\rho(Y_k)) \quad \textit{and} \quad ||x_k - \overline{x}|| \leq \sigma\ \rho(Y_k)^{1/p} \ . \tag{40}$$

Proof. Consider the SIP restricted to the closed ball cl $B_\kappa(\overline{x})$ with small κ chosen such that $\kappa < \varepsilon, \varepsilon_1$ (with ε in (10) and ε_1 in Lemma 6) :

$$\mathrm{P}^l(Y_k): \quad \min f(x) \qquad \text{s.t.} \quad x \in \mathcal{F}(Y_k) \cap \mathrm{cl}\ B_\kappa(\overline{x}) \ .$$

Obviously, since $\overline{x} \in \mathcal{F}(Y_k)$ and $\mathcal{F}(Y_k) \cap$ cl $B_\kappa(\overline{x})$ is compact (and nonempty), a solution x_k^l exists. Note that $\overline{x}$ is the unique (global) minimizer of $\mathrm{P}^l(Y)$. Put $\rho_k := \rho(Y_k)$ and consider any sequence of solutions x_k^l of $P^l(Y_k)$. In view of $\mathcal{F}(Y) \subset \mathcal{F}(Y_k)$ and $x_k^l + \tau\rho_k d \in F(Y) \cap$ cl $B_\varepsilon(\overline{x})$ (for large k, see Lemma 6(b)) we find

$$f(x_k^l) \leq f(\overline{x}) \leq f(x_k^l + \tau\rho_k d) \ .$$

Since $\overline{x}$ is a minimizer of order p (see 10) it follows

$$||x_k^l + \tau\rho_k d - \overline{x}||^p \leq \frac{1}{q}\left(f(x_k^l + \tau\rho_k d) - f(\overline{x})\right) \leq \frac{1}{q}\left(f(x_k^l + \tau\rho_k d) - f(x_k^l)\right) = O(\rho_k) \ .$$

Finally, the triangle inequality yields

$$||x_k^l - \overline{x}|| \leq ||x_k^l + \tau\rho_k d - \overline{x}|| + ||\tau\rho_k d|| = O(\rho_k^{1/p}) \ . \tag{41}$$

In particular $||x_k^l - \overline{x}|| < \kappa$ (for large k) such that x_k^l are (global) minimizers of the problem $\mathrm{P}^l(Y_k)$ restricted to the open neighborhood $B_\kappa(\overline{x})$.

□

REMARK 3. Under additional assumptions on the quality of the discretizations Y_k one can prove a faster convergence than in (40). It has been shown in [22] that a convergence rate $||\overline{x}_k - \overline{x}|| = O(\rho_k^{2/p})$ occurs if the grids y_k of meshsizes ρ_k are chosen in a special way.

Complexity of the Discretization Method. Let us shortly discuss the complexity of an approximation algorithm which solve the discretized LSIP problem by the interior point method. Consider an LSIP and the corresponding discretized LP problems depending on the meshsize ρ of the grid Y_ρ,

$$P(Y_\rho) \min_x c^T x \text{ s.t. } a(y)^T x \geq b(y) \quad \forall y \in Y_\rho$$

Le x_ρ be its solution. For simplicity we assume $Y = [0,1]^s$. The number m of gridpoints of a grid Y_ρ of meshsize ρ is (roughly) given by

$$m \approx \left(\frac{1}{\rho}\right)^s$$

By the interior point method we can compute an approximation $\overline{x}_\rho$ of x_ρ such that

$$0 \leq f(\overline{x}_\rho) - f(x_\rho) \leq \rho$$

by a number of (basic) computation steps of order $O(m^3 \ln \frac{m}{\rho})$ (see Section 6).
The discretization error between x_ρ and a the solution $\overline{x}$ of LSIP was bounded by

$$0 \leq f(\overline{x}) - f(x_\rho) \leq C \cdot \rho \, .$$

It is not difficult to show that such a constant C does not depend on ρ. So the overall complexity for computing $\overline{x}_\rho$ satisfying (assume $C \geq 1$)

$$|f(\overline{x}_\rho) - f(\overline{x})| \leq C \cdot d$$

is given by

$$O\left(s \cdot \left(\frac{1}{\rho}\right)^{3\cdot s} \cdot \ln \frac{1}{\rho}\right)$$

which is polynomial for fixed s. Note however that the computation work grows fast with s.

7.1. **Exchange method.** We also outline the *exchange method* which is often more efficient than a pure discretization method. This method can be seen as a compromise between the discretization method in Section 7.3 and the so-called reduction approach.

Algorithm 3 (*Conceptual exchange method*)

Step k: Given a discretization $Y_k \subset Y$ and a fixed, small value $\alpha > 0$.

i. Compute a solution x_k of SIP(Y_k).
ii. Compute local solutions y_k^i, $i = 1, \ldots, i_k$ ($i_k \geq 1$) of $Q(x_k)$ (cf. (22)) such that one of them, say y_k^1, is a global solution, i.e., $g(x_k, y_k^1) = \max_{y \in Y} g(x_k, y)$
iii. Stop, if $g(x_k, y_k^1) \geq -\alpha$, with a solution $\overline{x} \approx x_k$. Otherwise, update

$$Y_{k+1} = Y_k \cup \{y_k^i, \ i = 1, \ldots, i_k\} \, . \tag{42}$$

Theorem 11. *Suppose that the (starting) feasible set $\mathcal{F}(Y_0)$ is compact. Then, the exchange method (with $\alpha = 0$) either stops with a solution $\overline{x} = x_{k_0}$ of $P(Y)$ or the sequence $\{x_k\}$ of solutions of $P(Y_k)$ satisfies* dist $(x_k, \mathcal{S}(Y)) \to 0$.

Proof. We consider the case that the algorithm does not stop with a minimizer of P(Y). As in the proof of Theorem 9, by our assumptions, a solution x_k of $P(Y_k)$ exists, $x_k \in \mathcal{F}(Y_0)$ and with the subsequence $x_{k_\nu} \to \overline{x}$ we find

$$f(\overline{x}) \leq v(Y)\ .$$

Again we have to show $\overline{x} \in \mathcal{F}$ or equivalently $\varphi(\overline{x}) \geq 0$ for the value function $\varphi(x)$ of $Q(x)$. In view of $\varphi(x_k) = g(x_k, y_k^1)$ (see Algorithm 2, step ii) we can write

$$\varphi(\overline{x}) = \varphi(x_k) + \varphi(\overline{x}) - \varphi(x_k) = g(x_k, y_k^1) + \varphi(\overline{x}) - \varphi(x_k)\ .$$

Since $y_k^1 \in Y_{k+1}$ we have $g(x_{k+1}, y_k^1) \leq 0$ and by continuity of g and φ we find

$$\varphi(\overline{x}) \geq \big(g(x_k, y_k^1) - g(x_{k+1}, y_k^1)\big) + \big(\varphi(\overline{x}) - \varphi(x_k)\big) \to 0 \text{ for } k \to \infty\ .$$

□

We refer to the review paper [8] for more details on this approach.

8. Appendix

This section contains some definitions and auxiliary results. A set C is called *convex* if C contains with any $x, y \in C$ also the whole *line segment* $[x, y] = \{(1-\lambda)x + \lambda y \mid 0 \leq \lambda \leq 1\}$. For an arbitrary set $A \subset \mathbb{R}^n$ we define its *convex cone* by

$$\text{cone } A = \{a = \sum_{j=1}^{k} \lambda_j a_j \mid k \geq 1,\ a_j \in A,\ \lambda_j \geq 0\}\ .$$

and its *convex hull* by

$$\text{conv } A = \{a = \sum_{j=1}^{k} \lambda_j a_j \mid k \geq 1,\ a_j \in A,\ \lambda_j \geq 0,\ \sum_{j=1}^{k} \lambda_j = 1\}\ .$$

Lemma 7. (Caratheodory) *For $A \subset \mathbb{R}^n$, each $a \in$* conv *A can be represented as a convex combination of (at most) $n+1$ vectors: $a = \sum_{j=1}^{n+1} \lambda_j a_j$, and each element $a \in$* cone *A as a conic combination of (at most) n vectors: $a = \sum_{j=1}^{n} \lambda_j a_j$*

A function $f : C \to \mathbb{R}$ defined on a convex set $C \subset \mathbb{R}^n$ is called *convex* if for all $x, y \in C$ and $0 \leq \lambda \leq 1$,

$$f((1-\lambda)x + \lambda y) \leq (1-\lambda) f(x) + \lambda f(y)\ .$$

Lemma 8. *A differentiable function $f : C \to \mathbb{R}$ ($C \subset \mathbb{R}^n$, open and convex) is convex on C if and only if for any $x, y \in C$: $f(x) \geq f(x) + \nabla f(y)(x - y)$.*

LEMMA 9. [Generalized Gordan Lemma] *Let $A \subset \mathbb{R}^n$ be a compact set. Then exactly one of the following alternatives is true.*

(i) $0 \in \operatorname{conv} A$.
(ii) *There exists some* $d \in \mathbb{R}^n$ *such that* $a^T d < 0$ *for all* $a \in A$.

REFERENCES

[1] Amaya J., Gómez J.A., *Strong duality for inexact linear programming problems.* Optimization 49, no. 3, 243-269, (2001).

[2] Ben-Tal A., Nemirovski A., *Robust solutions of uncertain linear programs*, Operations Research Letters 25, 1-13, (1999).

[3] Borwein J.M., Lewis A.S., *Convergence of best entropy estimates.* SIAM J. Optim. 1, 191-205 (1991).

[4] Dall'Aglio M., *On Some Applications of LSIP to Probability and Statistics*, Semi-infinite programming, Eds. Goberna/López, (Alicante, 1999), 237–254, Nonconvex Optim. Appl., 57, Kluwer Acad. Publ., Dordrecht, 2001.

[5] Dambrine M., Pierre M., *About stability of equilibrium shapes*, Mathem. Modelling and Num. Analysis Vol.34, N0 4, 811-834, (2000).

[6] Faigle U., Kern W., Still G., Algorithmic Principles of Mathematical Programming, Kluwer, Dordrecht, (2002).

[7] Glashoff K., Gustafson S.-A., Linear Optimization and Approximation, Springer Verlag, Berlin, (1983).

[8] Hettich R., Kortanek K., Semi-infinite programming: Theory, methods and applications, SIAM Review, vol 35, N0.3, 380-429, (1993).

[9] Goberna M.A., L opez M.A., *Linear Semi-Infinite Optimization*, John Wiley & Sons, Chichester, (1998).

[10] Haaren-Retagne E., *A semi-infinite programming algorithm for robot trajectory planning*, Dissertation, University Trier (1992).

[11] Hettich R. and Zencke P., *Numerische Methoden der Approximation und der semi-infiniten Optimierung*, Teubner, Stuttgart (1982).

[12] Hettich R., Still G., *On Generalized semi-infinite programming problems*, Working paper, University of Trier, (1986).

[13] Hoffmann A. and Reinhardt R., *On reverse Chebyshev approximation problems*, Technical University of Illmenau, Preprint No. M08/94, (1994).

[14] H. Th Jongen, J.-J. Rückmann, O. Stein, Generalized semi-infinite optimization: A first order optimality condition and examples, Mathematical Programming 83 (1998) 145-158.

[15] Jess A., Jongen H.Th., Neralic L., Stein O., *A semi-infinite programming model in data envelopment analysis*, Optimization 49, 369-385, (2001).

[16] Len T., Vercher E., *Optimization under uncertainty and linear semi-infinite programming: a survey.* in Semi-infinite programming, Eds. Goberna/López, (Alicante, 1999), 327–348, Nonconvex Optim. Appl., 57, Kluwer Acad. Publ., Dordrecht, 2001.

[17] Polak E., *On the mathematical foundation of nondifferentiable optimization in engineering design*, SIAM Rev., 29, 21-89, (1987).

[18] Sánchez-Soriano J., Llorca N., Tijs S., Timmer J., *Semi-infinite assignment and transportation games.* Semi-infinite programming, Eds. Goberna/López, (Alicante, 1999), 349–363, Nonconvex Optim. Appl., 57, Kluwer Acad. Publ., Dordrecht, (2001).

[19] Soyster A.L., *Convex programming with set-inclusive constraints and applications to inexact linear programming*, Operations research 21, 1154-1157, (1973).

[20] Still G., Haaren-Retagne E., Hettich R., *A numerical comparison of two approaches to compute membrane eigenvalues by defect-minimization*, in “Intern. Series of Numer. Math.” vol 96, Birkhäuser Verlag, 209-224, (1991).

[21] Still G., *Generalized semi-infinite programming: Theory and methods*, European Journal of Operational Research 119, 301-313, (1999).

[22] Still G., *Discretization in semi-infinite programming: the rate of convergence*, Math. Program. 91 (2001), no. 1, Ser. A, 53–69.

[23] Tichatschke R., Hettich R., Still G., *Connections between generalized inexact and semi-infinite linear programming*, ZOR-Methods and Models of OR 33, 367-382, (1989).

[24] Tuncel L. Todd M.J., *Asymptotic behavior of interior point methods: a view from semi-infinite programming*, Mathematics of Operations Research 21, No. 2, 354-380, (1996).

[25] Wolkowicz H., Saigal L., Vandenberghe (Eds.), Handbook of semidefinite programming. Theory, algorithms, and applications. International Series in Operations Research and Management Science 27, Kluwer Academic Publishers, Boston, (2000).