SEMI-INFINITE PROGRAMMING

MARCO LÓPEZ, 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 paper, which intends to make a compromise between an introduction and a survey, treats the theoretical basis, numerical methods, applications and historical background of the field.

1. INTRODUCTION

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

$$P: \quad \min_x \ f(x) \quad \text{s.t.} \quad g(x,t) \geq 0 \quad \forall t \in T, \tag{1}$$

where T is an infinite *index set*. For the sake of shortness, we omit additional equality constraints $h_i(x) = 0,\ i = 1, 2, \ldots, m$.
By $\mathcal{F}$ we denote the *feasible set* of P, whereas $v := \inf\{f(x) \mid x \in \mathcal{F}\}$ is the *optimal value*, and $\mathcal{S} := \{\overline{x} \in \mathcal{F} \mid f(\overline{x}) = v\}$ is the *optimal set* or set of *minimizers* of the problem. We say that P is *feasible* or *consistent* if $\mathcal{F} \neq \emptyset$, and set $v = +\infty$ when $\mathcal{F} = \emptyset$. With the only exception of Section 4 and Subsection 6.3, we assume that f is continuously differentiable on $\mathbb{R}^n$, that T is a compact set in $\mathbb{R}^m$, and that the functions $g(.,.)$ and $\nabla_x g(.,.)$ are continuous on $\mathbb{R}^n \times T$.

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

$$P: \quad \min_x \ c^\top x \quad \text{s.t.} \quad a(t)^\top x \geq b(t) \quad \forall t \in T. \tag{2}$$

We also study here the *generalized SIP* (GSIP), for which the index set $T = T(x)$ is allowed to be dependent on x,

$$P: \quad \min_x \ f(x) \quad \text{s.t.} \quad g(x,t) \geq 0 \quad \forall t \in T(x). \tag{3}$$

During the last five decades the field of Semi-infinite Programming has known a tremendous development. More than 1000 articles and 10 books have been published on the theory, numerical methods and applications of SIP.

Date: October 17, 2005.
[0]Marco López, University of Alicante,marco.antonio@ua.es
[0]Georg Still, University of Twente, g.still@math.utwente.nl

1.2. **Historical background.** Although the origins of SIP are related to Chebyshev approximation, to the classical work of Haar on linear semi-infinite systems [41], and to the Fritz John optimality condition [49], the term SIP was coined in 1962 by Charnes, Cooper and Kortanek in some papers devoted to LSIP [17, 18, 19]. The last author, who contributed significantly to the development of the first applications of SIP in economics, game theory, mechanics, statistical inference, etc., has recently pointed out [56] the historical role of a paper published by Tschernikow in 1963 [83]. Gustafson and Kortanek proposed, during the early 1970s, the first numerical methods for SIP models arising in applications (see, for instance, [40]). The publication around 1980 of the following six books converted SIP in a mature and independent subfield in optimization. Two volumes of *Lecture Notes on Mathematics*, edited by Hettich [42] and by Fiacco and Kortanek [25], and four monographs by Tichatschke [81], Glashoff and Gustafson [27], Hettich and Zencke [43] (providing numerical methods and applications to approximation problems), and Brosowski [12] (devoted to stability in SIP). More recently, Goberna and López presented in [32] an extensive approach to LSIP, including both theory and numerical aspects. Reputed optimization books devoted some chapters to SIP *e.g.*, Krabs [57], Anderson and Nash [3], Guddat *et al.* [39], Bonnans and Shapiro [8], and Polak [66]. We also mention the review article of Polak [65] (on mathematical foundations of feasible directions methods), and [45] where Hettich and Kortanek surveyed, in a superb manner, theoretical results, methods and applications of SIP. Recently Goberna [28] and Goberna and López ([33, 34]) reviewed the LSIP model. Following the tracks of [4] and [12], during the last years, the stability analysis in SIP became an important research issue (see *e.g.*, [13, 14, 15, 16, 26, 30, 35, 36, 52, 55], etc., as a sample of recent papers on this topic). Since a first contribution [44] the GSIP model (3) became a topic of intensive research (see *e.g.*, [53] and [54]).

1.3. **Summary.** The paper is organized as follows. After fixing the notation in §1.4, §2 gives an account of a representative collection of motivating SIP and GSIP models in many different application fields. §3 presents the first order (primal and dual) optimality conditions. §4 is focused on different families of LSIP problems (continuous, FM and LFM), and the Haar duality theory is discussed in detail. Also in §4, some cone constrained optimization problems (in particular, the semidefinite and the second order conic programming problems) are presented as special cases of LSIP. §5 surveys the second order optimality conditions, although proofs are not included due to the technical difficulties of the subject. §6 introduces the principles of the main algorithmic approaches. Special attention is paid to the discretization methods, including some results about the discretizability of the general LSIP (with arbitrary index set), and to the exchange methods which are, in general, more efficient than the pure discretization algorithms. The last section, §7, is devoted to explore the relationship between the GSIP model and other important classes of optimization problems, like bi-level problems and mathematical programs with equilibrium constraints. The paper finishes with the list of cited references, but the reader will find an exhaustive bibliography on SIP and GSIP in [87].

1.4. **Notation and preliminaries.** 0_p will denote the null-vector in the Euclidean space $\mathbb{R}^p$ and $\|.\|$ represents the Euclidean norm.
If $X \subset \mathbb{R}^p$, by aff X, conv X, cone X, $D(X,x)$, and X^0 we shall denote the affine hull of X, the convex hull of X, the conical convex hull of X (always including the null-vector), the cone of feasible directions of X at x, and the positive polar cone of X ($X^0 := \{d \in \mathbb{R}^p \mid d^\top x \geq 0 \ \forall x \in X\}$), respectively. From the topological side, int X, cl X and bd X represent the interior, the closure and the boundary of X, respectively, whereas rint X and rbd X

are the relative interior and the relative boundary of X, respectively (the interior and the boundary in the topology relative to aff X).
Given a function $f : \mathbb{R}^p \to [-\infty, +\infty[$, the set $\left\{\binom{x}{\alpha} \in \mathbb{R}^{p+1} \mid \alpha \leq f(x)\right\}$ is called hypograph of f and is denoted by hypo f. The function f is a proper concave function if hypo f is a convex set in $\mathbb{R}^{p+1}$ and dom $f := \{x \in \mathbb{R}^p \mid f(x) > -\infty\} \neq \emptyset$. The closure cl f of a proper concave function f is another proper concave function, defined as the upper-semicontinuous hull of f, i.e., $(\mathrm{cl}\ f)(x) = \limsup_{y\to x} f(y)$. The following facts are well-known: hypo (cl f) = cl (hypo f), dom $f \subset$ dom (cl f) $\subset$ cl (dom f), and both functions f and cl f coincide except perhaps at points of rbd (dom f).
The vector u is a subgradient of the proper concave function f at the point $x \in$ dom f if, for every $y \in \mathbb{R}^p$, $f(y) \leq f(x) - u^\top(y - x)$. The set of all the subgradients of f at x is called subdifferential of f at x, and is denoted by $\partial f(x)$. The subdifferential is a closed convex set, and the differentiability of f at x is equivalent to $\partial f(x) = \{\nabla f(x)\}$. Moreover, $\partial f(x) \neq \emptyset$ if $x \in$ rint (dom f), and $\partial f(x)$ is a nonempty compact set if and only if $x \in$ int (dom f).

For later purposes we give two theorems of the alternative (see [32] for a proof).

LEMMA 1. [Generalized Gordan Lemma] *Let $A \subset \mathbb{R}^p$ be a set such that* conv *A is closed (*e.g., *A compact). Then exactly one of the following alternatives is true:*

(i) $0 \in$ conv A.
(ii) *There exists some $d \in \mathbb{R}^p$ such that $a^\top d < 0\ \forall a \in A$.*

LEMMA 2. [Generalized Farkas Lemma] Let $S \subset \mathbb{R}^{p+1}$ be an arbitrary nonempty set of vectors (a, b), $a \in \mathbb{R}^p$, $b \in \mathbb{R}$ such that the system $\sigma = \{a^\top z \geq b\ ,\ (a, b) \in S\}$ is feasible. Then the following statements are equivalent:

(i) The inequality $c^\top z \geq \gamma$ is a consequent relation of the system σ.
(ii) $\binom{c}{\gamma} \in$ cl K where $K = \text{cone}\left\{\binom{a}{b} \in S\ ,\ \binom{0_p}{-1}\right\}$.

2. EXAMPLES AND APPLICATIONS

In the review papers [45, 65], as well as in [32], the reader will find many applications of SIP in different fields such as Chebyshev approximation, robotics, mathematical physics, engineering design, optimal control, transportation problems, fuzzy sets, cooperative games, robust optimization, etc. There are also significant applications in statistics ([20, 32]), *e.g.*, the generalized Neyman-Pearson (present in the origin of linear programming), optimal experimental design in regression, constrained multinomial maximum-likelihood estimation, robustness in Bayesian statistics, actuarial risk theory, etc. From this large list of applications we have chosen some for illustrative purposes.

Chebyshev approximation. Let be given a function $f \in C(\mathbb{R}^m, \mathbb{R})$ and a set of approximating functions $p(x, \cdot) \in C(\mathbb{R}^n \times \mathbb{R}^m, \mathbb{R})$, parameterized by $x \in \mathbb{R}^n$. We want to approximate f by functions $p(x, \cdot)$ using the max-norm (Chebyshev-norm) $\|f\|_\infty = \max_{t\in T} |f(t)|$ on a compact set $T \subset \mathbb{R}^m$. Minimizing the approximation error $\epsilon := \|f - p\|_\infty$ is a problem that can be equivalently expressed as the SIP problem:

$$\min_{x,\epsilon}\ \epsilon \quad \text{s.t.} \quad g^\pm(x, t) := \pm\big(f(t) - p(x, t)\big) \leq \epsilon \qquad \forall t \in T. \tag{4}$$

In the so-called *reverse Chebyshev problem* we fix the approximation error ϵ and make the region T as large as possible [48]. Suppose that $T = T(d)$ is parameterized by $d \in \mathbb{R}^k$ and that $v(d)$ denotes the volume of $T(d)$ (*e.g.* $T(d) = \Pi_{i=1}^k[-d_i, d_i]$). The reverse Chebyshev problem then leads to the GSIP (with a small $\epsilon > 0$, fixed):

$$\max_{d,x} \; v(d) \;\text{ s.t. }\; \pm\big(f(t) - p(x,t)\big) \le \epsilon \qquad \forall t \in T(d), \tag{5}$$

where the index set $T(d)$ depends on the variable d.
For more details on the relations between SIP and Chebyshev approximation we refer to [27] and [43]. In [79] also the connection between approximation and GSIP has been discussed.

The minimal norm problem in the space of polynomials. A relevant class of functions in optimization are the so called *d.c. functions* (difference of convex functions). From the infinite possible decompositions of such functions, a suitable d.c. representation of any d.c. polynomial can be obtained by solving a *minimal norm problem* (MNP) [59]. This minimal d.c. representation improves the computational efficiency of the global optimization algorithms conceived to solve d.c. programs, by reducing the number of iterations needed to find a global optimum [24]. A peculiarity of the MNP problem is that it can be transformed into an equivalent quadratic SIP with linear constraints.
Let $\mathbb{R}_m[x]$, $x = (x_1, ..., x_n)$ be the vector space of polynomials of degree less than or equal to m. Let $B := \{f_i(x), i \in I\}$, be the usual basis of monomials in $\mathbb{R}_m[x]$. Hence, each polynomial $z(x) \in \mathbb{R}_m[x]$ can be written as $z(x) = \sum_{i\in I} z_i f_i(x)$, $z_i \in \mathbb{R}$, $i \in I$. In $\mathbb{R}_m[x]$ we consider the norm $\|\cdot\|$ defined by $\|z\| = (\sum_{i\in I} z_i^2)^{1/2}$. Let $C \subset \mathbb{R}^n$ be a closed convex set and let $K_m(C)$ be the nonempty closed convex cone of the polynomials in $\mathbb{R}_m[x]$ which are convex on C. Let $(y_1(x), y_2(x)), (w_1(x), w_2(x)) \in K_m(C) \times K_m(C)$ be two pairs of d.c. representations of $z(x)$ on C, *i.e.*, $z(x) = y_1(x) - y_2(x) = w_1(x) - w_2(x)$ on C. It can be argued that the pair $(y_1(x), y_2(x))$ is *better* than the pair $(w_1(x), w_2(x))$ if

$$\|y_1 + y_2\| \le \|w_1 + w_2\|.$$

A d.c. representation of $z(x)$ on C is *minimal* if it is better that any other d.c. representation of $z(x)$ on C. Thus, the practical way of obtaining the minimal d.c. representation of $z(x)$ on C is to solve the following problem.
Consider $z(x) \in \mathbb{R}_m[x]$ and let $(y_1(x), y_2(x)) \in K_m(C) \times K_m(C)$ be such that $z(x) = y_1(x) - y_2(x)$, and define $v(x) := y_1(x) + y_2(x)$, which is a convex polynomial on C. Hence, we can write

$$y_1(x) = (v(x) + z(x))/2 \quad \text{and} \quad y_2(x) = (v(x) - z(x))/2,$$

so $v(x) = -z(x) + 2y_1(x) = z(x) + 2y_2(x)$. Thus, the MNP problem is expressed as follows:

$$\min\big\{\|v\| : v \in \{-z + 2K_m(C)\} \cap \{z + 2K_m(C)\}\big\}. \tag{6}$$

Next we describe the feasible set of the problem (6) by requiring the convexity of the polynomials $(v \pm z)$, which means that the Hessian matrices $\nabla^2(v \pm z)(x) = \sum_{i\in I}(v_i \pm z_i)\nabla^2 f_i(x)$ must be positive semidefinite. We can then write

$$u^\top\left(\sum_{i\in I}(v_i \pm z_i)\nabla^2 f_i(x)\right)u \ge 0, \text{ for all } u \in S^n \text{ and } x \in C, \tag{7}$$

where $S^n = \{z \in \mathbb{R}^n : \|z\| = 1\}$. In this way the problem (6) has been transformed into the equivalent quadratic SIP problem, with $T = S^n \times C$,

$$P: \quad \min\{\|v\|^2 \equiv \sum_{i\in I} v_i^2\} \quad \text{s.t.} \quad u^\top\left(\sum_{i\in I}(v_i \pm z_i)\nabla^2 f_i(x)\right)u \ge 0 \quad \forall (u,x) \in T.$$

Mathematical physics. The so-called defect minimization approach for solving mathematical physics problems leads to SIP models, and 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 is a simply connected open region, with smooth boundary bd G_0, and closure cl G_0), and a positive constant $k > 0$, find a function $u \in C^2(\text{cl } G_0, \mathbb{R})$ such that its Laplacian $\Delta u = \frac{\partial^2 u}{\partial t_1^2} + \ldots + \frac{\partial^2 u}{\partial t_m^2}$ satisfies

$$\begin{array}{rcll} \Delta u(t) &=& k, & \forall t \in G_0, \\ u(t) &=& 0, & \forall t \in \text{bd } G_0. \end{array}$$

By choosing a linear space $S = \left\{u(x,t) = \sum_{i=1}^n x_i u_i(t)\right\}$, generated by appropriate trial functions $u_i \in C^2(\mathbb{R}^m, \mathbb{R})$, $i = 1, 2, \ldots, n$, this BVP can approximately be solved via the following SIP:

$$\min_{\varepsilon, x} \varepsilon \quad \text{s.t.} \quad \begin{array}{rcll} \pm\big(\Delta u(x,t) - k\big) &\le& \varepsilon, & \forall t \in G_0, \\ \pm u(x,t) &\le& \varepsilon, & \forall t \in \text{bd } G_0. \end{array}$$

In [21] the following related, but more complicated, model has been considered theoretically. This is the so-called *shape optimization problem* (SOP).

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

$$\min_{G,u} F(G,u) \quad \text{s.t.} \quad \begin{array}{rcll} \Delta u(t) &=& k, & \forall t \in G, \\ u(t) &=& 0, & \forall t \in \text{bd } G, \\ \mu(G) &=& 1. & \end{array}$$

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

Choose some appropriate family 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 $u(x,t)$ from the set S above, the program

$$\min_{z,x} F(G(z), u(x,\cdot)) \quad \text{s.t.} \quad \begin{array}{rcll} \pm\big(\Delta u(x,t) - k\big) &\le& \varepsilon, & \forall t \in G(z), \\ \pm u(x,t) &\le& \varepsilon, & \forall t \in \text{bd } G(z). \end{array}$$

Similar models such as the *membrane packing problem with rigid obstacle* are to be found in [63]. For further contributions to the theory and numerical results of this defect minimization approach we refer *e.g.* to [78, 80].

Robotics. Many control problems in robotics lead to semi-infinite problems (*cf.* [46]). As an example we discuss the *maneuverability problem*, which in [44] originally has led to the concept of GSIP.

Let $\Theta = \Theta(\tau) \in \mathbb{R}^m$ denote the position of the so-called tool center point of the robot (in robot coordinates) at time τ. Let $\dot\Theta, \ddot\Theta$ be the corresponding velocities, accelerations (derivatives w.r.t. τ). The dynamical equations have often the form

$$g(\Theta, \dot\Theta, \ddot\Theta) := A(\Theta)\ddot\Theta + F(\Theta, \dot\Theta) = K,$$

with (external) forces $K \in \mathbb{R}^m$. Here $A(\Theta)$ is the inertia matrix, and F describes the friction, gravity, centrifugal forces, etc. The forces K are bounded: $K^- \le K \le K^+$.

For fixed $\Theta, \dot{\Theta}$, the set of feasible (possible) accelerations is given by

$$Z(\Theta, \dot{\Theta}) = \{\ddot{\Theta} \mid K^- \leq g(\Theta, \dot{\Theta}, \ddot{\Theta}) \leq K^+\}.$$

Note that, since g is linear in $\ddot{\Theta}$, for fixed $(\Theta, \dot{\Theta})$, the set $Z(\Theta, \dot{\Theta})$ is convex (intersection of half-spaces). Let now be given an *operating region* Q, e.g.

$$Q = \{(\Theta, \dot{\Theta}) \in \mathbb{R}^{2m} \mid (\Theta^-, \dot{\Theta}^-) \leq (\Theta, \dot{\Theta}) \leq (\Theta^+, \dot{\Theta}^+)\}.$$

Then, the set of feasible accelerations $\ddot{\Theta}$, *i.e.*, the set of accelerations which can be realized in every point $(\Theta, \dot{\Theta}) \in Q$, becomes

$$Z_0 = \bigcap_{(\Theta,\dot{\Theta}) \in Q} Z(\Theta, \dot{\Theta}) = \{\ddot{\Theta} \mid K^- \leq g(\Theta, \dot{\Theta}, \ddot{\Theta}) \leq K^+, \ \forall (\Theta, \dot{\Theta}) \in Q\}.$$

The set Z_0 is convex (as an intersection of the convex sets $Z(\Theta, \dot{\Theta})$). For controlling the robot one has to check whether a desired acceleration $\ddot{\Theta}$ is possible, *i.e.* whether $\ddot{\Theta} \in Z_0$. Often this test takes too much time due to the complicated description of Z_0. A faster test can be done as follows. First we have to find a simple body T (e.g. a ball or an ellipsoid) as large as possible, which is contained in Z_0. Then, instead of the test $\ddot{\Theta} \in Z_0$ one performs the faster check $\ddot{\Theta} \in T$.
The construction of an appropriate body $T \subset Z_0$ leads to a GSIP as follows. Suppose that the body $T(d)$ depends on the parameter $d \in \mathbb{R}^q$ and that $v(d)$ is the volume of $T(d)$. Then, we wish to maximize the volume subject to the condition $T(d) \subset Z_0$. This is the so-called *maneuverability problem* and leads to the GSIP:

$$\text{(8)} \qquad \max_d \ v(d) \quad \text{s.t.} \quad K^- \leq g(\Theta, \dot{\Theta}, \ddot{\Theta}) \leq K^+, \quad \forall (\Theta, \dot{\Theta}) \in Q, \ \ddot{\Theta} \in T(d).$$

Geometry. Semi-infinite problems naturally arise in a geometrical setting. More precisely, the outer approximation (covering) of a set $T \subset \mathbb{R}^m$ by a set $S(x)$, depending on a parameter $x \in \mathbb{R}^n$, leads to a SIP. To cover T from inside will yield a GSIP.
Suppose $S(x)$ is described by $S(x) = \{t \in \mathbb{R}^m \mid g(x, t) \geq 0\}$ and $v(x)$ denotes its volume. In order to find the set $S(x)$ of smallest volume covering T we have to solve the SIP:

$$\min_x v(x) \quad \text{s.t.} \quad g(x, t) \geq 0 \quad \forall t \in T.$$

In the inner approximation problem we maximize the volume such that the set $S(x)$ is contained in the set $T = \{t \in \mathbb{R}^m \mid g(t) \geq 0\}$, and it is modeled by the GSIP:

$$\max_x v(x) \quad \text{s.t.} \quad g(t) \geq 0 \quad \forall t \in S(x).$$

This problem is also known as *design centering problem* (see *e.g.* [75], [81]).

Optimization under uncertainty. We consider a linear program

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

where 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 $T_j \subset \mathbb{R}^{n+1}$. In a pessimistic model we now can restrict the problem to such x which are feasible for all possible data vectors. This leads to a SIP

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

For more details we refer to [1, 6, 58, 71, 82]. In the next example we discuss such a *robust optimization* model in economics.

Economics. In a *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 t_i euros per 1 euro investment in share i.
Our goal is to maximize the portfolio value $v = t^\top x$ after a year where $x = (x_1, \dots, x_n)$ and $t = (t_1, \dots, t_n)$. The problem is that the values t_i are not known in advance (otherwise we could invest all the money into the share with maximal value t_i). However, knowledge from the past and models from economics allow us to predict that the gain coefficients t_i vary between certain bounds. So we can assume that the vector t will be contained in a specific compact subset $T \subset \mathbb{R}^n$.
In this *robust optimization* model we now wish to maximize the gain v for the *worst case* vector $t \in T$, and we are led to solve the linear SIP:

$$\max_{v,x} \ v \quad \text{s.t.} \qquad t^\top x - v \geq 0 \quad \forall t \in T, \ \text{and} \ \sum_i x_i = K, \ x \geq 0.$$

We refer to [75] for details and numerical experiments.

3. First order optimality conditions

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

$$(9) \qquad f(x) - f(\overline{x}) \geq 0 \qquad \forall x \in \mathcal{F} \text{ such that } \|x - \overline{x}\| < \varepsilon.$$

The minimizer $\overline{x}$ is said to be *global* if this relation holds for every $\varepsilon > 0$. 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

$$(10) \qquad f(x) - f(\overline{x}) \geq q||x - \overline{x}||^p \qquad \forall x \in \mathcal{F} \text{ such that } \|x - \overline{x}\| < \varepsilon.$$

For $\overline{x} \in \mathcal{F}$ we consider the *active index set*:

$$T_a(\overline{x}) := \{t \in T \mid g(\overline{x}, t) = 0\}.$$

Since g is continuous and T is compact, the subset $T_a(\overline{x}) \subset T$ is also compact. The condition $T_a(\overline{x}) = \emptyset$ implies $\overline{x} \in \text{int}\, \mathcal{F}$ and, near $\overline{x}$, the problem P can be seen as an unconstrained minimization problem. So, throughout the paper, we assume that, at a candidate minimizer $\overline{x}$, the set $T_a(\overline{x})$ is nonempty.
Now we introduce the so-called constraint qualifications. The *linear independence constraint qualification* (LICQ) is said to be satisfied at $\overline{x} \in \mathcal{F}$ if the active gradients

$$(\text{LICQ}) \qquad \nabla_x g(\overline{x}, t), \quad t \in T_a(\overline{x}), \ \text{are linearly independent.}$$

The *Mangasarian-Fromovitz constraint qualification* (MFCQ) holds at $\overline{x} \in \mathcal{F}$ if there exists a direction $d \in \mathbb{R}^n$ such that

$$(\text{MFCQ}) \qquad \nabla_x g(\overline{x}, t) d > 0 \quad \forall t \in T_a(\overline{x}).$$

A direction $d \in \mathbb{R}^n$ satisfying the condition MFCQ is called a *strictly feasible direction.*

Remark The condition LICQ implies MFCQ. To see this note that by LICQ the set $T_a(\overline{x})$ cannot contain more than n points, $T_a(\overline{x}) = \{t_1, \dots, t_r\}$, $r \leq n$, and the system $\nabla_x g(\overline{x}, t_j) d = 1$, $j = 1, \dots, r$ has a solution d. Moreover, in the proof of Lemma 3 we show that if d is a strictly feasible direction at $\overline{x} \in \mathcal{F}$ and $\tau > 0$ is small enough, then $\overline{x} + \tau d \in \text{int}\, \mathcal{F}$, which, obviously, will be non-empty.

A vector $d \in \mathbb{R}^n$ is a *strictly feasible descent direction* if it satisfies simultaneously

$$\nabla f(\overline{x}) d < 0, \ \nabla_x g(\overline{x}, t) d > 0 \quad \forall t \in T_a(\overline{x}).$$

Lemma 3. [Primal necessary optimality condition] *Let $\overline{x} \in \mathcal{F}$ be a local minimizer of P. Then, there will not exist a strictly feasible descent direction.*

Proof. Assume, reasoning by contradiction, that d is a strictly feasible descent direction. Then, $f(\overline{x}+\tau d) < f(\overline{x})$ holds for small $\tau > 0$. On the other hand, we show next the existence of some $\tau_0 > 0$ with the property

$$g(\overline{x}+\tau d, t) > 0 \quad \forall \tau \in]0, \tau_0] \text{ and } \forall t \in T, \tag{11}$$

entailing that $\overline{x}$ can not be a local minimizer (contradiction). Suppose now that (11) does not hold. Then, for each natural number k, there exists some $t_k \in T$ such that $0 < \tau_k < 1/k$ and $g(\overline{x}+\tau_k d, t_k) \leq 0$. Since T is compact, there must exist some subsequence (τ_{k_s}, t_{k_s}) such that $\tau_{k_s} \to 0$ and $t_{k_s} \to t^* \in T$. The continuity of $g(.,.)$ then yields $g(\overline{x}+\tau_{k_s} d, t_{k_s}) \to g(\overline{x}, t^*)$, which itself implies $g(\overline{x}, t^*) = 0$ and, so, $t^* \in T_a(\overline{x})$. On the other hand, the Mean Value Theorem provides us with numbers $0 < \hat{\tau}_{k_s} < \tau_{k_s}$ such that

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

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

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

which contradicts the hypothesis on d. □

THEOREM 1. [First order sufficient condition] *Let $\overline{x}$ be feasible for P. Suppose that there is no $d \in \mathbb{R}^n \setminus \{0_n\}$ satisfying*

$$\nabla f(\overline{x}) d \leq 0 \text{ and } \nabla_x g(\overline{x}, t) d \geq 0 \ \ \forall t \in T_a(\overline{x}).$$

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

For a proof of this result we refer to [45] (and also to [73]).

Remark The assumptions of Theorem 1 are rather strong and can be expected to hold only in special cases (see, *e.g.,* [45]). It is not difficult to see that the assumptions imply that the set of gradients $\{\nabla_x g(\overline{x}, t) \mid t \in T_a(\overline{x})\}$ contains a basis of $\mathbb{R}^n$. So, in particular, $|T_a(\overline{x})| \geq n$.
More general sufficient optimality conditions need second order information (*cf.* Section 5 below).

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

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

(a) *There exist multipliers $\mu_0, \mu_1, \ldots, \mu_k \geq 0$ and indices $t_1, \ldots, t_k \in T_a(\overline{x})$, $k \leq n+1$, such that $\sum_{j=0}^k \mu_j = 1$ and*

$$\mu_0 \nabla f(\overline{x}) - \sum_{j=1}^{k} \mu_j \nabla_x g(\overline{x}, t_j) = 0^\top. \qquad \text{(FJ-condition)} \tag{12}$$

(b) *If MFCQ holds at $\overline{x}$, then there exist multipliers $\mu_1, \ldots, \mu_k \geq 0$ and indices $t_1, \ldots, t_k \in T_a(\overline{x})$, $k \leq n$, such that*

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

Proof. (a) Consider the set $S = \{\nabla f(\overline{x})\} \cup \{-\nabla_x g(\overline{x}, t) \mid t \in T_a(\overline{x})\} \subseteq \mathbb{R}^n$. 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

$$\text{there is no } d \in \mathbb{R}^n \text{ such that } s^\top d < 0 \ \forall s \in S.$$

$T_a(\bar{x})$ is compact and, by continuity of $\nabla_x g(\bar{x}, \cdot)$, S is also compact. Hence conv S is a compact convex set and $0_n \in$ conv S (Lemma 1). In view of Caratheodory's Theorem, 0_n is a convex combination of at most $n+1$ elements of S, *i.e.*,

$$\text{(14)} \qquad \sum_{j=1}^{k} \mu_j s_j = 0 \qquad s_j \in S,\ \mu_j \geq 0,\ \sum_{j=1}^{k} \mu_j = 1 \text{ with } k \leq n+1,$$

which implies (a).
(b) Now assume $\mu_0 = 0$ in (12), *i.e.*, for every j in (14) such that $\mu_j > 0$ we have $s_j \neq \nabla f(\bar{x})$ and, accordingly, there is an associated $t_j \in T_a(\bar{x})$ such that $s_j = -\nabla_x g(\bar{x}, t_j)$. Then, if d is the direction involved in MFCQ the following contradiction arises:

$$0 > -\sum_{j=1}^{k} \mu_j \nabla_x g(\bar{x}, t_j) d = 0^\top d = 0.$$

□

Convex semi-infinite programs. The semi-infinite program P is called *convex* if the objective function $f(x)$ is convex and, for every index $t \in T$, the constraint function $g_t(.) = g(., t)$ is concave (*i.e.*, $-g_t(.)$ is convex). A local minimizer of a convex program is actually a global one.
In this convex setting the following constraint qualification is usual. We say that P satisfies the *Slater condition* if

$$(\text{SCQ}) \qquad \text{there exists } \widehat{x} \text{ such that } g(\widehat{x}, t) > 0 \ \ \forall t \in T.$$

LEMMA 4. *Let $\mathcal{F}$ be non-empty. Then P satisfies SCQ if and only if MFCQ holds at every $\bar{x} \in \mathcal{F}$.*

Proof. Assume, first, that P satisfies SCQ. If $T_a(\bar{x}) = \emptyset$, then MFCQ is trivially satisfied. Otherwise, if $t \in T_a(\bar{x})$ we have

$$\nabla_x g(\bar{x}, t)(\widehat{x} - \bar{x}) \geq g(\widehat{x}, t) - g(\bar{x}, t) = g(\widehat{x}, t) > 0,$$

and $d := \widehat{x} - \bar{x}$ satisfies MFCQ. Now we choose a point $\bar{x} \in \mathcal{F}$. By assumption MFCQ holds at $\bar{x}$ with a vector d. In the proof of Lemma 3 (*cf.* (11)) it is shown that there exists some $\tau_0 > 0$ such that the point $\widehat{x} := \bar{x} + \tau_0 d$ satisfies SCQ. □

As in convex finite programming, the KKT-conditions are sufficient for optimality.

THEOREM 3. *Let P be a convex SIP. If $\bar{x}$ is a feasible point that satisfies the Kuhn-Tucker condition (13), then $\bar{x}$ is a (global) minimizer of P.*

Proof. By the convexity assumption, we have for every feasible x and $\bar{t} \in T_a(\bar{x})$:

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

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

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

□

More details on constraint qualifications for the SIP problem (like the semi-infinite versions of the Kuhn-Tucker and Abadie constraint qualifications) can be found in *e.g.* [38].

4. LINEAR SIP

4.1. Different models in LSIP. This section deals with the general LSIP

$$P: \qquad \min c^\top x \quad \text{s.t.} \quad a_t^\top x \ge b_t \qquad \forall t \in T.$$

Here T is an arbitrary (infinite) set, and the vectors $c, a_t \in \mathbb{R}^n$, as well as the scalars $b_t \in \mathbb{R}$, are also arbitrary. The functions $t \mapsto a_t \equiv a(t)$ and $t \mapsto b_t \equiv b(t)$ need not to have any special property. As an intersection of closed halfspaces, the feasible set of P is a closed convex set.

We introduce different families of LSIP problems through some properties of their constraint systems $\sigma = \{a_t^\top x \ge b_t\ ,\ t \in T\}$, which have a great influence on optimality, stability and on the efficiency of algorithms:

(a) P is called *continuous* ([17, 18]), when T is a compact Hausdorff topological space, and $a(t)$ and $b(t)$ are continuous functions on T.

(b) A feasible problem P is said to be *locally Farkas-Minkowski* (LFM) ([64], extensions to the convex SIP in [23] and [60]) when every linear inequality $a^\top x \ge b$ which is a *consequent relation* of σ, and such that $a^\top x = b$ is a supporting hyperplane of $\mathcal{F}$, is also a consequence of a finite subsystem of σ.

(c) P, assumed again to be feasible, is *Farkas-Minkowski* (FM) ([83, 86]) if every linear consequent relation of σ is a consequence of some finite subsystem.

An LFM problem exhibits a satisfactory behavior with respect to the duality theory, and every FM problem is LFM (the converse holds provided that $\mathcal{F}$ is bounded). On the other hand, many approximation problems are modeled as continuous problems, but these problems behave badly w.r.t. stability duality and numerical methods unless a Slater point x^0 exists ($a_t^\top x^0 > b_t\ \forall t \in T$), in which case they also belong to the FM family.

The *first-moment cone*, M, and the *characteristic cone*, K, play an important role in LSIP:

$$M := \text{cone}\,\{a_t, t \in T\}\ ,\ K := \text{cone}\left\{\begin{pmatrix} a_t \\ b_t \end{pmatrix}, t \in T; \begin{pmatrix} 0_n \\ -1 \end{pmatrix}\right\}.$$

According to Lemma 2, the program P is FM if and only if the cone K is closed. Another consequence of Lemma 2 is the characterization of the LFM problems as those feasible LSIP problems such that

$$A(x) = D(\mathcal{F}, x)^0,\ \forall x \in \mathcal{F}\ .$$

Here $A(x) = \text{cone}\,\{a_t\ ,\ t \in T_a(x)\}$ is called *active cone* at x and $D(\mathcal{F}, x)$ is the cone of feasible directions of $\mathcal{F}$ at x.

4.2. Optimality and duality. If P is a continuous LSIP problem, with $T \subset \mathbb{R}^m$, the first order optimality theory presented in Section 3 applies (*cf.* also [32, Chapter 7]).

In fact, the KKT condition now turns out to be $c \in A(\overline{x})$. If the LSIP problem is LFM we observe that $A(\overline{x})$ is a closed cone (because it coincides with $D(\mathcal{F}, \overline{x})^0$) and, since $\overline{x}$ is optimal if and only if $c \in D(\mathcal{F}, \overline{x})^0$, the LFM property itself becomes a constraint qualification.

Associated with P, different *dual* problems can be defined. For instance, if P is continuous, a natural dual problem is

$$D_0: \quad \max \int_T b(t)d(\lambda(t)) \quad \text{s.t.} \quad \int_T a(t)d(\lambda(t)) = c, \quad \lambda \in \mathcal{M}^+(T)\ ,$$

where $\mathcal{M}^+(T)$ denotes the cone of the nonnegative regular Borel measures on the compact Hausdorff space T. Nevertheless, our most general approach does not assume any particular property of the index set and, consequently, we introduce a dual problem that is always well-defined. This can be accomplished by restricting, in D_0, the feasible solutions to atomic measures concentrating their mass on a finite set and yields the so-called *Haar dual*

$$D: \quad \max \sum_{t\in T} \lambda_t b_t \quad \text{s.t.} \quad \sum_{t\varepsilon T} \lambda_t a_t = c,\ \lambda_t \geq 0,$$

where we allow only for a finite number of the *dual variables*, $\lambda_t,\ t \in T$, to take positive values. If $\lambda := (\lambda_t, t \in T)$ is a feasible solution of D (also called dual-feasible), we shall consider the *support* of λ :

$$\text{supp}\ \lambda := \{t \in T \mid \lambda_t > 0\}.$$

By v_{D_0} and v_D we denote the optimal values of D_0 and D, respectively. If P is a continuous problem with $T \subset \mathbb{R}^m$, a theorem of Rogosinsky [69] established $v_{D_0} = v_D$. Moreover, if v_{D_0} is attainable (i.e., if D_0 is solvable), then v_D is also attainable. For a general compact Hausdorff space T, the equivalence between D_0 and D (from the optimality point of view) is established in [11]. Because of these equivalences, the Haar dual is much more convenient.
Next we consider the objective vector c in P as a *parameter*. We analyze here the properties of the *optimal value function* $v : \mathbb{R}^n \to [-\infty, +\infty[$ and the *optimal set mapping* $\mathcal{S} : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$, assigning to each $c \in \mathbb{R}^n$, the optimal value $v(c)$ and optimal set (possibly empty) $\mathcal{S}(c)$ of the problem $P(c) : \min\{c^\top x \mid x \in \mathcal{F}\}$, respectively.
Obviously, $c \in \text{dom}\ \mathcal{S}$ if and only if $P(c)$ is solvable (i.e., if $P(c)$ has optimal solutions). We shall also assume that $\mathbb{R}^n \neq \mathcal{F} \neq \emptyset$.
A crucial element in duality theory is the so-called *duality gap* $\delta(c) := v(c) - v_D(c)$, where $v_D(c)$ denotes the optimal value of the associated dual problem

$$D(c): \quad \max \sum_{t\in T} \lambda_t b_t \quad \text{s.t.} \quad \sum_{t\varepsilon T} \lambda_t a_t = c,\ \lambda_t \geq 0.$$

For this parametric dual problem only the right-hand side terms in the equality constraints change, and we assume $v_D(c) = -\infty$ when $c \notin M$, in other words when de dual problem is not feasible.
Note that if $x \in \mathcal{F}$ and $\lambda = \{\lambda_t,\ t \in T\}$ is feasible for the dual problem $D(c)$, one has $c^\top x = \sum_{t\in T} \lambda_t a_t^\top x \geq \sum_{t\in T} \lambda_t b_t$, which gives rise to the *weak duality* inequality $v(c) \geq v_D(c)$ (i.e., $\delta(c) \geq 0$). When $\delta(c) = 0$, we say that there is *no-duality gap*.
Observe that, applying Lemma 2, (recall $\max \emptyset = -\infty$),

$$v(c) = \sup\{\alpha \in \mathbb{R} \mid c^\top x \geq \alpha \text{ is a consequence of } \sigma\} = \sup \left\{ \alpha \in \mathbb{R} \;\middle|\; \begin{pmatrix} c \\ \alpha \end{pmatrix} \in \text{cl}\ K \right\}.$$

Moreover, the dual optimal value can be rewritten as follows

$$v_D(c) = \sup \left\{ \alpha \in \mathbb{R} \;\middle|\; \begin{pmatrix} c \\ \alpha \end{pmatrix} \in K \right\}, \tag{15}$$

and basic results from convex analysis (*cf.* [68]) yield straightforwardly the following statements:

THEOREM 4. ([29, Theorem 8.1],[70]) *In relation to the parametric dual pair $P(c)$, $D(c)$, with $\mathbb{R}^n \neq \mathcal{F} \neq \emptyset$, the following properties hold:*

(i) *v and v_D are proper concave functions such that $v = \text{cl}\ v_D$ and, so, their values coincide at every point of* rint (dom v_D),
(ii) hypo $v = \text{cl}\ K$,
(iii) $\mathcal{S}(c) = \partial v(c)$,
(iv) rint $M \subset \text{dom}\ \mathcal{S} \subset \text{dom}\ v \subset \text{cl}\ M$, *and* rint (dom v_D) = rint M,

(v) *For every* $c \in \operatorname{rint} M$, *we have* $\delta(c) = 0$ *and* $\mathcal{S}(c) \neq \emptyset$,
(vi) $\mathcal{S}(c)$ *is a (non-empty) compact set if and only if* $c \in \operatorname{int} M$.

In contrast to (finite) linear programming, for LSIP, a strong duality result as in Theorem 4(v) needs not to hold unless some constraint qualification is satisfied. This can be seen from the following

EXAMPLE 1. $P: \quad \min\ x_2 \quad \text{s.t.} \quad t^2 x_1 + t x_2 \geq 0,\ \forall t \in T = [0,1] \text{ and } x_2 \geq -1.$

Here $v = 0$ and $v_D = -1$. Obviously, the condition $c \in \operatorname{rint} M$ does not hold.

Cone constrained programs as special cases of LSIP. A cone constrained linear problem is a program of the form

$$\min\ c^\top x \quad \text{s.t.} \quad y := Ax - b \in C, \tag{16}$$

where C is a closed convex cone in a normed space. If $C \subset \mathbb{R}^m$ and $\operatorname{int} C \neq \emptyset$ such a problem can be transformed into an LSIP satisfying the FM property (see [37]). As important special cases we discuss *semidefinite programs*

$$(\text{SDP}) \qquad \min\ c^\top x \quad \text{s.t.} \quad Y := \sum_{i=1}^{n} x_i A_i - B \in S_+^m,$$

with symmetric matrices $A_i, B \in \mathbb{R}^{m\times m}$ and S_+^m being the cone of positive semidefinite $m \times m$-matrices, as well as the *second order conic programs*

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

where L^m denotes the Lorentz cone $L^m := \{y \in \mathbb{R}^m \mid y_m \geq (y_1^2 + \ldots + y_{m-1}^2)^{1/2}\}$.
By definition, $Y \in S_+^m$ is equivalent to $t^\top Y t \geq 0\ \forall t \in C_m = \{t \in \mathbb{R}^m \mid \|t\| = 1\}$, where $\|\cdot\|$ denotes the Euclidean norm. So the feasibility condition for SDP becomes $\sum_i x_i t^\top A_i t \geq t^\top B t$ or

$$a^\top(t) x \geq b(t) \quad \forall t \in C_m \quad \text{with } a_i(t) = t^\top A_i t,\ b(t) = t^\top B t,$$

which turns SDP into a LSIP problem.
For the program SOP we define $\tilde{y} = (y_1, \ldots, y_{m-1})$ and observe the identity

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

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

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

with $a_i(t) = t^\top A_i$, A_i the i-th column of A and $b(t) = t^\top b$.
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.* [22]). Applications of the abstract duality theory to the problem of moments, LSIP, and continuous linear programming problems are discussed in [70].

5. SECOND ORDER OPTIMALITY CONDITIONS

A natural way to obtain optimality conditions for SIP is the so-called *reduction approach*. The advantage of this method is that we can deal in the same way with SIP and GSIP problems. Since the approach is described in detail in [45], here we only sketch the results. Consider the GSIP (3) and assume that the index set is given by

$$T(x) := \{t \in \mathbb{R}^m \mid u(x, t) \geq 0_q\}, \tag{17}$$

where $u : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^q$, and that f, g, u are C^2-functions. The approach is based on the fact that, for $\overline{x} \in \mathcal{F}$, each active index

$$\bar{t}_j \in T_a(\overline{x}) = \{t \in T(\overline{x}) \mid g(\overline{x}, t) = 0\}$$

is a global minimizer of the so-called *lower level problem*,

$$(18) \qquad Q(\overline{x}) : \qquad \min_t \; g(\overline{x}, t) \text{ s.t. } u_l(\overline{x}, t) \geq 0 \;, \; l \in L := \{1, \ldots, q\},$$

which represents a finite program depending on the parameter $\overline{x}$. So, under some constraint qualification each $\bar{t}_j \in T_a(\overline{x})$ must satisfy the *Karush-Kuhn-Tucker* (KKT) condition, *i.e.*, with the Lagrange function $\mathcal{L}_j^{(\overline{x})}$ of $Q(\overline{x})$, corresponding to $\bar{t}_j \in T_a(\overline{x})$, the relation

$$(19) \qquad \nabla_t \mathcal{L}_j^{(\overline{x})}(\overline{x}, \bar{t}_j, \overline{\gamma}_j) = \nabla_t g(\overline{x}, \bar{t}_j) - \sum_{l \in L_0(\overline{x}, \bar{t}_j)} \overline{\gamma}_{jl} \nabla_t u_l(\overline{x}, \bar{t}_j) = 0_m^\top$$

holds with associated multipliers $\overline{\gamma}_{jl} \geq 0$, and the active index set $L_0(\overline{x}, \bar{t}_j) := \{l \in L \mid u_l(\overline{x}, \bar{t}_j) = 0\}$. The approach depends on the *reduction assumptions*:

(RA): All $\bar{t}_j \in T_a(\overline{x})$ are nondegenerate minimizers of $Q(\overline{x})$ such that LICQ, SC (strict complementary slackness), and SOC (strong second order conditions) hold at them (see [45] and [47] for details).

Under the condition RA the following can be shown: The set $T_a(\overline{x})$ is finite, i.e. $T_a(\overline{x}) = \{\bar{t}_1, \ldots, \bar{t}_r\}$, and there are (locally defined) C^1-functions $t_j(x)$ and $\gamma_j(x)$ with $t_j(\overline{x}) = \bar{t}_j$, $\gamma_j(\overline{x}) = \overline{\gamma}_j$, such that, for every x near $\overline{x}$, $t_j(x)$ is a local minimizer of $Q(x)$ with corresponding unique multiplier $\gamma_j(x)$. With these functions the following *reduction* holds in a neighborhood $U_{\overline{x}}$ of $\overline{x}$:
The point $x \in U_{\overline{x}}$ is a local solution of the GSIP problem P if and only if $x \in U_{\overline{x}}$ is a local solution of the *locally reduced (finite) program*

$$(20) \qquad P_{red}(\overline{x}) : \quad \min f(x) \quad \text{s.t.} \quad G_j(x) := g(x, t_j(x)) \geq 0, \; j = 1, \ldots, r.$$

Moreover, for $j = 1, \ldots, r$, the following identity holds:

$$\nabla G_j(\overline{x}) = \nabla_x \mathcal{L}_j^{(\overline{x})}(\overline{x}, \bar{t}_j, \overline{\gamma}_j).$$

Based on this result, the standard optimality conditions for the finite problem $P_{red}(\overline{x})$ directly lead to optimality conditions for GSIP. To do so, we define the cone of critical directions

$$C_{\overline{x}} := \{d \in \mathbb{R}^n \mid \nabla f(\overline{x})d \leq 0, \; \nabla G_j(\overline{x})d \geq 0, \; j = 1, \cdots, r\},$$

and recall that LICQ is said to hold at $\overline{x}$ if the vectors $\nabla G_j(\overline{x})$, $j = 1, \cdots, r$, are linearly independent. We give the following necessary and sufficient optimality conditions of FJ-type for P (see also [76]).

THEOREM 5. *Let for $\overline{x} \in \mathcal{F}$ the assumptions RA be satisfied such that GSIP can locally be reduced to $P_{red}(\overline{x})$. Then the following conditions hold:*
(a) Suppose that $\overline{x}$ is a local minimizer of P and that LICQ is satisfied. Then there exist multipliers $\overline{\mu} \geq 0_r$ such that (with the expressions for $\nabla_x L$ and $\nabla_x^2 L$ below)

$$\nabla_x L(\overline{x}, \overline{\mu}) = 0_n^\top \quad \text{and} \quad d^\top \nabla_x^2 L(\overline{x}, \overline{\mu}) d \geq 0 \quad \forall d \in C_{\overline{x}}.$$

(b) Suppose that there exist multipliers $\overline{\mu} \geq 0_r$ such that

$$\nabla_x L(\overline{x}, \overline{\mu}) = 0_n^\top \quad \text{and} \quad d^\top \nabla_x^2 L(\overline{x}, \overline{\mu}) d > 0 \quad \forall d \in C_{\overline{x}} \setminus \{0\}.$$

Then $\overline{x}$ is a strict local minimizer of P of order 2.

The expressions for $\nabla_x L(\overline{x}, \overline{\mu})$ *and* $\nabla_x^2 L(\overline{x}, \overline{\mu})$ *read:*

$$\begin{aligned}
\nabla_x L(\overline{x}, \overline{\mu}) &= \nabla f(\overline{x}) - \sum_{j=1}^{r} \overline{\mu}_j \nabla_x \mathcal{L}_j^{(\overline{x})}(\overline{x}, \overline{t}_j, \overline{\gamma}_j), \\
\nabla_x^2 L(\overline{x}, \overline{\mu}) &= \nabla^2 f(\overline{x}) - \sum_{j=1}^{r} \overline{\mu}_j \left(\nabla_x^2 \mathcal{L}_j^{(\overline{x})}(\overline{x}, \overline{t}_j, \overline{\gamma}_j) - \nabla t_j^\top(\overline{x}) \nabla_t^2 \mathcal{L}_j^{(\overline{x})}(\overline{x}, \overline{t}_j, \overline{\gamma}_j) \nabla t_j(\overline{x}) \right) \\
&\quad + \sum_{j=1}^{r} \overline{\mu}_j \sum_{l \in L_0(\overline{x}, \overline{t}_j)} \left(\nabla^\top \gamma_{jl}(\overline{x}) \nabla_x u_l(\overline{x}, \overline{t}_j) + \nabla_x^\top u_l(\overline{x}, \overline{t}_j) \nabla \gamma_{jl}(\overline{x}) \right) .
\end{aligned}$$

Recall that under the assumption RA at $\overline{x} \in \mathcal{F}$, for each active index point $\overline{t}_j \in T_a(\overline{x})$ the KKT-equations (19) must be satisfied. Together with Theorem 5 we obtain the following *complete system of optimality conditions* for SIP and GSIP:
If $\overline{x} \in \mathcal{F}$ is a local minimizer with active index set $T_a(\overline{x}) = \{\overline{t}_1, \ldots, \overline{t}_r\}$, such that the assumptions of Theorem 5(a) are fulfilled, then $\overline{x}$ must satisfy the equations

$$(21) \qquad \begin{aligned}
\nabla f(x) - \textstyle\sum_{j=1}^{r} \mu_j \nabla_x \left(g(x, t_j) - \sum_{l \in L_0(x,t_j)} \gamma_{jl} u_l(x, t_j) \right) &= 0_n^\top, \\
\forall j: \qquad g(x, t_j) &= 0, \\
\forall j: \qquad \nabla_t g(x, t_j) - \textstyle\sum_{l \in L_0(x,t_j)} \gamma_{jl} \nabla_t u_l(x, t_j) &= 0_m^\top, \\
\forall j \text{ and } \forall l \in L_0(x, t_j): \qquad u_l(x, t_j) &= 0,
\end{aligned}$$

with appropriate Lagrange multipliers μ_j and γ_{jl}. This system has $K := n + r(m+1) + \sum_{j=1}^{r} |L_0(\overline{x}, \overline{t}_j)|$ equations and K unknowns x, μ_j, t_j, γ_j.

Remark Note that for common SIP-problems the system (21) simplifies. Since in this case the functions u_l do not depend on x, in the first equations the sum over $\gamma_{jl} \nabla_x u_l(x, t_j)$ vanishes. Observe that the first part of the system (21) has the structure of the KKT-equations in finite programming. However, since in SIP the index variable t is not discrete and it may vary in a whole continuum T, this system does not determine the active indices t_j. To fix also the t_j's, for all active index points t_j we have to add the second system of KKT-equations in the t variable.

6. NUMERICAL METHODS

Nowadays the numerical approach to SIP has become an active research area. An excellent review on SIP algorithms is [67]. Recently, the NEOS Server has included the program NSIPS, coded in AMPL [84, 85].
As in finite programming we can distinguish between primal and dual solution methods. In the so-called discretization methods the SIP problem is directly replaced by a finite problem (FP).
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 t variable:

$$Q(\overline{x}): \qquad \min_t \ g(\overline{x}, t) \quad \text{s.t.} \quad t \in T,$$

and to check whether for a global solution $\overline{t}$ the condition $g(\overline{x}, \overline{t}) \geq 0$ 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 [62], as well as certain classes of robust optimization problems [6].

6.1. **Primal methods.** We discuss the so-called method of *feasible directions* for the SIP problem. The idea is to move from a current (non-optimal) feasible point x_k to the next point $x_{k+1} = x_k + \tau_k d_k$ in such a way that x_{k+1} remains feasible and has a smaller objective value. The simplest choice would be to move along a strictly feasible descent direction which was defined as a vector d satisfying

$$\nabla f(\overline{x})d < 0, \qquad \nabla_x g(\overline{x}, t)d > 0 \quad \forall t \in T_a(\overline{x}). \tag{22}$$

Feasible Direction Method (Zoutendijk, in case of FP) Choose a starting point $x_0 \in \mathcal{F}$.

Step k: Stop if x_k is a FJ point.

(1) Choose a strictly feasible descent direction d_k.
(2) Determine a solution τ_k for the problem:
(*) $\min \{ f(x_k + \tau d_k) \mid t > 0,\ x_k + \tau d_k \in \mathcal{F} \}$.
Set $x_{k+1} = x_k + \tau_k d_k$.

Let us consider the following *stable* way to obtain a strictly feasible descent direction d that takes *all* constraints into account. For finite programs this is the method suggested by Topkis and Veinott (see *e.g.* [22]). We solve the LSIP:

$$\begin{array}{llrcll} \min_{d,z} z & \text{s.t.} & \nabla f(x_k)d - z & \leq & 0, & \\ & & -\nabla_x g(x_k, t)d - z & \leq & g(x_k, t), & \forall t \in T, \\ & & \pm d_i & \leq & 1, & i = 1, \dots, n. \end{array} \tag{23}$$

LEMMA 5. *If (23) has optimal value equal to zero, x_k satisfies the FJ necessary optimality conditions (12).*

Proof. Note that $(d, z) = (0_n, 0)$ is always feasible for (23). Suppose that the statement is false and that x_k is not a FJ point. Then by Lemma 1, there exists a strictly feasible descent direction d. With the same arguments as in the proof of Lemma 3, and by replacing $\overline{x}$ by x_k, we find that there is some $\tau_0 > 0$ such that

$$g(x_k, t) + \tau \nabla_x g(x_k, t)d > 0 \quad \forall \tau \in]0, \tau_0[,\ t \in T.$$

So, by choosing $\tau > 0$ such that $\tau |d_i| \leq 1,\ \forall i$, we have shown that there exist $z_0 < 0$ and a vector $d (= \tau d)$ with $|d_i| \leq 1,\ \forall i$, satisfying

$$\nabla f(x_k)d < z_0 \quad \text{and} \quad g(x_k, t) + \nabla_x g(x_k, t)d > -z_0 \quad \forall t \in T. \tag{24}$$

Consequently (23) has optimal value $z \leq z_0$.

□

THEOREM 6. *Assume that the Feasible Direction Method generates points x_k and computes directions d_k as solutions of (23). Suppose further that a subsequence $(x_s)_{s \in S}$ converges to $\overline{x}$, $S \subseteq \mathbb{N}$ being an infinite subset of indices s. Then, $\overline{x}$ is a FJ point.*

Proof. Since $\mathcal{F}$ is closed, the limit $\overline{x}$ of the points $x_s \in \mathcal{F}$ also belongs to $\mathcal{F}$. Suppose that the theorem is false and that $\overline{x}$ is not a FJ point. Then arguing as in the proof of Lemma 5, with x_k replaced by $\overline{x}$ and z_0 replaced by $2z$, it follows that there exist $z < 0$ and a vector d with $|d_i| \leq 1,\ \forall i$, satisfying

$$\nabla f(\overline{x})d < 2z \quad \text{and} \quad g(\overline{x},t) + \nabla_x g(\overline{x},t)d > -2z \quad \forall t \in T.$$

The continuity of $g(x,t),\ \nabla f(x)$ and $\nabla_x g(x,t)$, together with the compactness of T, also imply that

$$\nabla f(x_s)d < z \quad \text{and} \quad g(x_s,t) + \nabla_x g(x_s,t)d > -z \quad \forall t \in T,$$

must hold if s is sufficiently large. So (d,z) is feasible for (23), which in turn implies $z \geq z_s$ for every optimal solution (d_s, z_s) of (23) at x_s. In particular, we note

$$\nabla f(x_s)d_s < z \quad \text{and} \quad g(x_s,t) + \nabla_x g(x_s,t)d_s > -z \quad \forall t \in T.$$

Let $\delta > 0$ be such that $\|\nabla f(x) - \nabla f(\overline{x})\| < |z|/3\sqrt{n}$ holds whenever $\|x - \overline{x}\| < \delta$. Then $\|d_s\| \leq \sqrt{n}$, together with the inequality of Cauchy-Schwarz, yield

$$|(\nabla f(x) - \nabla f(\overline{x}))d_s| \leq \|\nabla f(x) - \nabla f(\overline{x})\| \cdot \|d_s\| < |z|/3.$$

So, if $\|x - \overline{x}\| < \delta$ and s is large enough, we have:

$$\nabla f(x)d_s \leq \nabla f(\overline{x})d_s + |z|/3 \leq \nabla f(x_s)d_s + 2|z|/3 < z/3.$$

Applying the same reasoning to $\nabla_x g(x,t)$, we therefore conclude that, for δ small enough and $\|x - \overline{x}\| < \delta$,

$$g(x_s,t) + \nabla_x g(x,t)d_s \geq -z/3 \quad \forall t \in T.$$

We are interested in points x of the form $x = x_s + \tau d_s,\ 0 < \tau < \min\left\{1, \delta/(2\sqrt{n})\right\}$. If s is so large that $\|x_s - \overline{x}\| < \delta/2$, then $\|x - \overline{x}\| < \delta$. Moreover, the Mean Value Theorem guarantees the existence of some $0 < \alpha < \tau$ such that

$$f(x_s + \tau d_s) = f(x_s) + \tau \nabla f(x_s + \alpha d_s)d_s < f(x_s) + \tau z/3.$$

Similarly, we find for any $t \in T$

$$\begin{aligned} g(x_s + \tau d_s, t) &= g(x_s,t) + \tau \nabla_x g(x_s + \alpha_t d_s, t)d_s \\ &= (1-\tau)g(x_s,t) + \tau[g(x_s,t) + \nabla_x g(x_s + \alpha_t d_s, t)d_s] \\ &\geq (1-\tau)g(x_s,t) - \tau z/3 > 0 \end{aligned}$$

(use the feasibility $g(x_s,t) \geq 0$), which tells us $x_s + \tau d_s \in \mathcal{F}$. So, the minimization step $(^*)$ in the algorithm produces some τ_s ensuring a decreasing of at least

$$f(x_{s+1}) - f(x_s) \ = \ f(x_s + \tau_s d_s) - f(x_s) \ \leq \ -\frac{\delta}{2\sqrt{n}}|z|/3$$

for all sufficiently large s. Hence, $f(x_s) \to -\infty$ as $s \to \infty$ contradicting our assumption that $x_s \to \overline{x}$ and, accordingly, $f(x_s) \to f(\overline{x})$. □

Some fundamentals of the descent methods based on nonsmooth analysis can be found in the review article [65].

6.2. **Dual methods.** The so-called dual or KKT-approaches try to compute a solution of the system of KKT optimality conditions. This is mostly achieved by applying some Newton-type iterations. The theoretical basis of this approach is closely related to so-called generic properties of P. So we shortly discuss two variants of this KKT-approach and the related genericity results for SIP or GSIP.

SQP-Method based on reduction. To solve SIP or GSIP we can apply any algorithm from finite programming to the locally reduced problem $P_{red}(x)$ described in Section 5. An efficient way to do so is the SQP-version. We give a conceptual description (see [45, Section 7.3] for more details).

Method based on the reduction:

Step k: Start from a given x_k (not necessarily feasible).

(1) Determine the local minima $t_1, \cdots, t_{r_k}$ of $Q(x_k)$ in (18).
(2) Apply N_k steps of a SQP-solver (for finite programs) to the problem (20)

$$P_{red}(x_k): \qquad \min_x f(x) \;\text{ s.t. }\; G_j(x) := g(x, t_j(x)) \geq 0, \;\; j = 1, \cdots, r_k,$$

leading to iterates $x_{k,i}, \;\; i = 1, \cdots, N_k$.
(3) Set $x_{k+1} = x_{k,N_k}$ and $k = k + 1$.

Note that in this procedure we have to trace out the minimizer functions $t_j(x_k)$ numerically by parametric optimization techniques. For a discussion of such a method combining global convergence and local superlinear convergence we refer to [45].

Methods based on the system of KKT-equations. To solve the GSIP (3) we also can try to compute a solution of the complete system of optimality conditions (21) by some Newton-type iteration. The problem here is that we have to find a rough approximation of a solution which can serve as a starting point for the locally (quadratically) convergent Newton iteration. A possible procedure, described in [45], performs as follows:

(1) Compute an approximate solution of the semi-infinite program by some other method (discretization or exchange method).
(2) Use this approximation as a starting point in the Newton iteration for solving (21).

Genericity results for SIP and GSIP. In the reduction approach presented in Section 5 the following has been assumed:

RC: For $x \in \mathcal{F}$ and a solution $x, \mu_j, y_j, \gamma_{jl}$ of (21) the following is true:

(1) LICQ holds: The vectors in the following set are linearly independent

$$\left\{ \nabla_x \mathcal{L}_j^{(x)}(x, t_j, \gamma_j) = \nabla_x g(x, t_j) - \sum_{l \in L_0(x,t_j)} \gamma_{jl} \nabla_x u_l(x, t_j), \;\; t_j \in T_a(x) \right\}$$

(2) The condition RA holds (see Section 5).

Note that in the Newton method for (21) the natural regularity conditions are that the Jacobian of the system is regular at the solution point. It is not difficult to see, that both assumptions are essentially equivalent and are related to the second order sufficient conditions in Theorem 5(b).

The genericity question now is to examine whether RC are natural assumptions in the sense that they are generically fulfilled. To answer this question we have to define the problem set properly. For fixed dimensions n, m, q the SIP problem P can be seen as an element of the space

$$\mathcal{P} := \{P = (f, g, u)\} \equiv C^{\infty}(\mathbb{R}^n, \mathbb{R}) \times C^{\infty}(\mathbb{R}^{n+m}, \mathbb{R}) \times C^{\infty}(\mathbb{R}^m, \mathbb{R}^q).$$

This function space is assumed to be endowed with the so-called strong Withney topology [50]. By a generic subset of the set $\mathcal{P}$ we mean a subset which is dense and open in $\mathcal{P}$. The following result states that for the class of SIP problems the reduction approach and the Newton method are generically applicable.

THEOREM 7. [51] *The set $\mathcal{P}$ of all C^{∞}-SIP problems contains an open and dense subset $\mathcal{P}_0 \subset \mathcal{P}$ such that for all programs $P \in \mathcal{P}_0$ the regularity condition RC is satisfied for all $x \in \mathcal{F}$.*

Unfortunately the situation is much more complicated for GSIP where a genericity result as in Theorem 7 is not true (see *e.g.*, [72, Section 3]).

6.3. **Discretization methods.** In a discretization method we choose *finite* subsets T' of T, and instead of $P \equiv P(T)$ we solve the finite programs

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

Let $v(T')$, $\mathcal{F}(T')$ and $\mathcal{S}(T')$ denote the minimal value, the feasible set, and the set of global minimizers of $P(T')$. We call these finite subsets T' *grids* or *discretizations*. The following relation is trivial:

$$T_2 \subset T_1 \quad \Rightarrow \quad \mathcal{F}(T_1) \subset \mathcal{F}(T_2) \text{ and } v(T_2) \leq v(T_1). \tag{25}$$

We consider different concepts here: $P \equiv P(T)$ is called *finitely reducible* if

there is a grid $T' \subset T$ such that $v(T') = v(T)$.

P is said to be *weakly discretizable* if there exists a sequence of grids T_k such that

$$v(T_k) \to v(T).$$

Obviously, if P is finitely reducible, then P is weakly discretizable. Note that the previous concepts apply for the general SIP, but the following notion requires T to be a subset of a space with a metric $d(\cdot,\cdot)$. We define a *meshsize* $\rho(T')$ of a grid T' by the Hausdorff distance between T' and T, $\rho(T') := \sup_{t \in T} \min_{t' \in T'} d(t,t')$,.

P is called *discretizable* if for each sequence of grids T_k satisfying $\rho(T_k) \to 0$ we have:

(i) $P(T_k)$ is solvable for k large enough,
(ii) for each sequence of solutions $\overline{x}_k \in \mathcal{S}(T_k)$ we have

$$d(\overline{x}_k, \mathcal{S}(T)) \to 0, \quad \text{and} \quad v(T_k) \to v(T) \quad \text{if } k \to \infty,$$

where $d(x, \mathcal{S}) := \min_{s \in \mathcal{S}} \|x - s\|$. Note that from a numerical viewpoint, only the concept of discretizability is useful.
We also introduce a local concept: Given a local minimizer $\overline{x}$ of $P(T)$, the SIP is called *locally discretizable* at $\overline{x}$ if the relations above hold locally, i.e. if the problem $P^{local}(T)$, obtained as the restriction of $P(T)$ to an open neighborhood $U_{\overline{x}}$ of $\overline{x}$, is discretizable.

We start the discussion with negative examples about the possibilities for discretization.

EXAMPLE 2. Let us consider the nonconvex SIP problem

$$P : \quad \min \quad x_2 \quad \text{s.t. } (x_1 - t)^2 + x_2 \geq 0, \ \forall t \in T = [0,1], \text{ and } 0 \leq x_1 \leq 1.$$

Obviously $\mathcal{F} = \{x \in \mathbb{R}^2 \mid 0 \leq x_1 \leq 1 \text{ and } x_2 \geq 0\}$, $v = 0$, and $\mathcal{S} = \{\overline{x} \in \mathbb{R}^2 \mid 0 \leq \overline{x}_1 \leq 1 \text{ and } \overline{x}_2 = 0\}$. On the other hand, whichever grid T' we take, it is evident that $v(T') < 0$

and $\mathcal{S}(T')$ is a (nonempty) finite set. So P is not finitely reducible but it is discretizable (as it is not difficult to see).

The preceeding problem is not a convex SIP because the concavity of the functions $g(\cdot, t)$ fails. Reconsidering the problem in Example 1 it is not difficult to see that here $v(T') = v_D = -1$, for every finite grid $T' \subset T$, whereas $v = v(T) = 0$ (*cf.* Theorem 9).
Concerning the finite reducibility, the following result ([10], [45, Theorem 4.2]) comes from Helly-type arguments: Assume that P is a convex SIP, with bounded optimal value, such that T is a compact Hausdorff topological space, and that a specific Slater-like condition holds (for every set of $n+1$ points $t_0, t_1, ..., t_n \in T$, a point $\widetilde{x}$ exists such that $g(\widetilde{x}, t_i) > 0$, $i = 0, 1, ..., n$). Then the problem P is finitely reducible (with some existing $T' \subset T$, $|T'| = n$).
Note that such a convex problem P is reducible if there exists a minimizer satisfying the KKT condition (13) with a subset $T' = \{t_1, \ldots, t_k\}$ of $T_a(\overline{x})$, $k \le n$. Indeed then, by Theorem 3 the point $\overline{x}$ is also a solution of $P(T')$.

For a general LSIP problem (with arbitrary T), we have the following result:

THEOREM 8. [32, Theorems 8.3 and 8.4]. *Let us consider the problem*

$$P: \quad \min c^\top x \quad s.t. \quad a_t^\top x \ge b_t \;\; \forall t \in T,$$

where T is an arbitrary set.
(a) Assume that the optimal value v is finite. Then the following statements are equivalent:

(i) *P is finitely reducible,*
(ii) *We have $\binom{c}{v} \in K$,*
(iii) *D is solvable and $v_D = v$ (no duality gap),*
(iv) *D is solvable and P is weakly discretizable.*

(b) Let us again consider $P = P(c)$ as a problem depending on c as a parameter. Then $P(c)$ is finitely reducible for every $c \in \mathbb{R}^n$ such that $v(c)$ is finite if and only if the system $\sigma = \{a_t^\top x \ge b_t,\ t \in T\}$ is FM.

Proof. (a) "$(i) \Rightarrow (ii)$" Let T' be a grid such that $v(T') = v(T) \equiv v$. According to the duality theory in ordinary LP, the associated dual $D(T')$ is solvable and there is no duality gap; i.e. $v(T') = v_D(T')$. In other words, there exist optimal values of the dual variables $\overline{\lambda}_t,\ t \in T'$, such that

$$\sum_{t \in T'} \overline{\lambda}_t \begin{pmatrix} a_t \\ b_t \end{pmatrix} = \begin{pmatrix} c \\ v_D(T') \end{pmatrix} = \begin{pmatrix} c \\ v \end{pmatrix},$$

and $\binom{c}{v} \in K$, holds.
"$(ii) \Rightarrow (iii)$" From $\binom{c}{v} \in K$ and (15) we get $v_D \ge v$. The weak duality inequality then yields $v_D = v$, and $\binom{c}{v} \in K$ entails the solvability of D.
"$(iii) \Rightarrow (iv)$" Next we proceed by proving that $v_D = v$ (no duality gap) implies that P is weakly discretizable. Let us consider sequences $\{x_k\}$ and $\{\lambda_k\}$ of feasible solutions of P and D, respectively, such that

$$\lim_{k\to\infty} \sum_{t \in T} \lambda_{kt} b_t = v_D = v = \lim_{k\to\infty} c^\top x_k. \tag{26}$$

Taking grids $T_k :=$ supp λ_k we write

$$\sum_{t \in T} \lambda_{kt} b_t \le v_D(T_k) = v(T_k) \le v_D(T) = v(T) \le c^\top x_k,$$

and (26) implies $v(T_k) \to v(T)$.

"$(iv) \Rightarrow (i)$" Assume that T_k, $k = 1, 2, \ldots$, is a sequence of grids such that $-\infty < v(T_k) \to v(T)$, and that $\overline{\lambda}$ is an optimal solution of D. Then the restriction of $\overline{\lambda}$ to $T_0 :=$ supp $\overline{\lambda}$ is also optimal for $D(T_0)$, and we have

$$v(T_k) = v_D(T_k) \leq v_D(T) = v_D(T_0) = v(T_0) \leq v(T).$$

In view of $v(T_k) \to v(T)$ we find $v(T_0) = v(T)$.
(b) Reasoning by contradiction, let us assume that $P(c)$ is finitely reducible, for every $c \in \mathbb{R}^n$ such that $v(c)$ is finite, and that $\sigma = \{a_t^\top x \geq b_t,\ t \in T\}$ is not FM. Then, since K is not closed (*cf.* Section 4.1), there will exist $\binom{c_0}{\gamma} \in (\text{cl } K) \setminus K$. According to Lemma 2, $c_0^\top x \geq \gamma$ is a linear consequence of σ and thus, $v(c_0) \geq \gamma$. Therefore, $v(c_0)$ is finite (because $P(c_0)$ is feasible) and, so, $P(c_0)$ is finitely reducible by assumption. Part (a) and (15) conjointly entail $-\infty < v_D(c_0) \leq \gamma$. Part (a) also yields $\binom{c_0}{v(c_0)} \in K$ and $v(c_0) = v_D(c_0)$. Hence $v(c_0) = v_D(c_0) = \gamma$, and we get a contradiction with $\binom{c_0}{\gamma} \notin K$.
Conversely, let us assume that σ is FM. If the optimal value $v(c)$ of $P(c)$ is finite, $c^\top x \geq v(c)$ is a consequence of σ and Farkas' Lemma, together with the FM assumption, yields

$$\begin{pmatrix} c \\ v(c) \end{pmatrix} = \text{cl } K = K.$$

Now, (a) applies. □

THEOREM 9. [32, Corollary 8.2.1] *Assume that the optimal value v of a general LSIP problem P is finite. Then P is weakly discretizable if and only if $v = v_D$ (no duality gap).*

Proof. " $\Leftarrow$ " Already established in the proof of "$(iii) \Rightarrow (iv)$" in part (a) of Theorem 8.
" $\Rightarrow$ " Assume that T_k, $k = 1, 2, \ldots$, is a sequence of grids such that $-\infty < v(T_k) \to v(T) \equiv v$. Then we can write

$$v(T_k) = v_D(T_k) \leq v_D(T) \leq v(T),$$

and taking limits for $k \to \infty$ we conclude $v_D(T) = v(T)$. □

We next give a sufficient condition for LSIP to be discretizable or weakly discretizable, partially based on Theorem 4.

THEOREM 10. [32, Theorem 8.6 and Corollary 8.6.1]. *Consider the feasible LSIP problem*

$$P: \quad \min c^\top x \quad s.t. \quad a_t^\top x \geq b_t \quad \forall t \in T,$$

T being an arbitrary index set. If $c \in$ rint M, then P is weakly discretizable.
Moreover, if M is full dimensional and $c \in$ int M, then $\mathcal{S}$ is bounded and P is discretizable, provided that P is continuous.

COROLLARY 1. *Let P be a continuous LSIP with bounded feasible set $\mathcal{F}$. Then P is discretizable.*

Proof. If $\mathcal{F}$ is bounded, one has $M = \mathbb{R}^n$ (cl M is the polar of the recession cone of $\mathcal{F}$), and $c \in$ int M holds trivially. □

The following example illustrates the difference between weak discretizability and discretizability.

EXAMPLE 3. Consider the LSIP (with some fixed $\varepsilon > 0$):

$$\min\ x_1 \quad \text{s.t.} \quad x_1 + t x_2 \geq 0 \ \ \forall t \in \text{T:=} \begin{cases} [0, 1]] & \textit{Case A} \\ [-\varepsilon, 1] & \textit{Case B} \end{cases}$$

$v = 0$, and a minimizer of $P(T)$ is, in both cases, given by $\overline{x} = (0, 0)$.

Case A: The problem is weakly discretizable but not discretizable. For a grid T' containing $t = 0$ we have $v(T') = v$. On the other hand for any T' not containing 0 the value is unbounded, $v(T') = -\infty$.
Case B: The problem is discretizable as it can easily be shown. Note that in Case B the condition $c \in \text{int}\, M$ is satisfied, but not in Case A.

The following algorithm is based on the discretizability concept and from now on we assume $T \subset \mathbb{R}^m$.

Conceptual discretization method

Step k: Given a grid $T_k \subset T$ and a small value $\alpha > 0$.

(1) Compute a solution x_k of $P(T_k)$.
(2) Stop if x_k is feasible within the fixed accuracy α, i.e. $g(x_k, t) \geq -\alpha \;\; \forall t \in T$.
   Otherwise, select a finer discretization T_{k+1}, *i.e.*, $\rho(T_{k+1}) < \rho(T_k)$.

We begin with some general convergence results for the discretization method.

LEMMA 6. *If $P \equiv P(T)$ is continuous, T_k, $k = 1, 2, ..,$ is a sequence in 2^T such that $\rho(T_k) \to 0$ as $k \to \infty$, and $x_k \in \mathcal{F}(T_k)$, $k = 1, 2, \ldots,$ converges to $\overline{x}$, then $\overline{x} \in \mathcal{F} \equiv \mathcal{F}(T)$.*

Proof. For a fixed $t \in T$, $\rho(T_k) \to 0$ entails the existence of a sequence of indices $t_k \in T_k$, $k = 1, 2, \ldots,$ such that $t_k \to t$ as $k \to \infty$. Thus, $g(x_k, t_k) \geq 0$, $k = 1, 2, \ldots,$ and taking limits, for $k \to \infty$, the continuity of P allows us to write $g(\overline{x}, t) \geq 0$. Since we took an arbitrary $t \in T$, we conclude $\overline{x} \in \mathcal{F}$. □

THEOREM 11. *Let $\mathcal{F}(T_1)$ be compact and let the sequence of discretizations T_k satisfy*

$$T_1 \subset T_k \;\forall k \geq 2 \;\text{and}\; \rho(T_k) \to 0 \quad \text{for}\; k \to \infty\,.$$

Then $P(T)$ is discretizable.
Proof. By assumption and using $T_1 \subset T_k \subset T$ the feasible sets $\mathcal{F}(T)$, $\mathcal{F}(T_k)$, of $P(T)$, $P(T_k)$ respectively, are compact and satisfy $\mathcal{F}(T) \subset \mathcal{F}(T_k) \subset \mathcal{F}(T_1)$, $k \in \mathbb{N}$. Consequently, solutions x_k of $P(T_k)$ exist. Suppose now that a sequence of such solutions does not satisfy $d(x_k, \mathcal{S}(T)) \to 0$. Then there exist $\varepsilon > 0$ and a subsequence x_{k_ν} such that

$$d(x_{k_\nu}, \mathcal{S}(T)) \geq \epsilon > 0 \quad \forall \nu\,.$$

Since $x_{k_\nu} \in \mathcal{F}(T_1)$ we can select a convergent subsequence. Without restriction we can assume $x_{k_\nu} \to \overline{x}$, $\nu \to \infty$. In view of $\mathcal{F}(T) \subset \mathcal{F}(T_k)$ the relation $f(x_{k_\nu}) \leq v(T)$ holds and thus by continuity of f we find

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

We now show that $\overline{x} \in \mathcal{S}(T)$ in contradiction to our assumption. To do so it suffices to prove that $\overline{x} \in \mathcal{F}(T)$. Let $\overline{t} \in T$ be given arbitrarily. Since $\rho(T_{k_\nu}) \to 0$ for $\nu \to \infty$ we can choose $t_{k_\nu} \in T_{k_\nu}$, such that $t_{k_\nu} \to \overline{t}$. In view of $g(x_{k_\nu}, t_{k_\nu}) \geq 0$, by taking the limit $\nu \to \infty$, it follows $g(\overline{x}, \overline{t}) \geq 0$, i.e. $\overline{x} \in \mathcal{F}(T)$. □

Next we consider local discretizability for general nonlinear SIP and we assume for the rest of this subsection that the function g is continuously differentiable on $\mathbb{R}^n \times T$. Recall that the *Mangasarian Fromovitz constraint qualification* (MFCQ) holds at $\overline{x} \in \mathcal{F}$ if there exists a vector $d \in \mathbb{R}^n$ such that $\nabla_x g(\overline{x}, \overline{t})d > 0$, $\forall \overline{t} \in T_a(\overline{x})$. Because $T_a(\overline{x})$ is compact and $\nabla_x g(\overline{x}, .)$ is continuous, there must exist κ such that

$$\nabla_x g(\overline{x}, \overline{t})d \geq \kappa > 0 \quad \forall \overline{t} \in T_a(\overline{x}). \tag{27}$$

LEMMA 7. *Let be given a sequence of grids $T_k \subset T$ with $\rho_k := \rho(T_k) \to 0$.*

(a) *Let K be a compact subset of $\mathbb{R}^n$. Then there exists $c > 0$ such that for all k large enough:*

$$g(x_k, t) \geq -c\rho_k \quad \forall t \in T \text{ and } \forall x_k \in \mathcal{F}(T_k) \cap K\ .$$

(b) *Let MFCQ be satisfied at $\overline{x}$ with the vector d. Then, there exist positive numbers τ and ε_1 such that for all k large enough:*

$$x_k + \tau\rho_k d \in \mathcal{F}(T) \quad \forall x_k \in \mathcal{F}(T_k) \text{ with } \|x_k - \overline{x}\| < \varepsilon_1\ .$$

Proof. (a) Let t_k be a solution of $\min_{t\in T}\ g(x_k, t)$ and let $\hat{t}_k$ be a point in T_k such that $||\hat{t}_k - t_k|| \leq \rho_k$. By Lipschitz continuity of g ($\nabla_t g(\cdot,\cdot)$ is continuous) and using $g(x_k, \hat{t}_k) \geq 0$ we find

$$g(x_k, t) \geq g(x_k, t_k) \geq g(x_k, t_k) - g(x_k, \hat{t}_k) \geq -c||\hat{t}_k - t_k|| \geq -c\rho_k \quad \forall t \in T,$$

for some Lipschitz constant $c > 0$.

(b) To prove the statement we proceed in two steps. Firstly, for $\varepsilon > 0$ we consider the relative open set

$$T_a^\varepsilon(\overline{x}) := \{t \in T \mid ||t - \overline{t}|| < \varepsilon \text{ for some } \overline{t} \in T_a(\overline{x})\}.$$

By MFCQ (see (27)) using the continuity of $\nabla_x g$ there is some $\varepsilon > 0$ such that

$$\nabla_x g(x, t)d \geq \frac{\kappa}{2} \quad \forall t \in T_a^\varepsilon(\overline{x}) \text{ and } \forall x \text{ such that } ||x - \overline{x}|| < \varepsilon.$$

Thus if $||x_k - \overline{x}|| < \varepsilon$, and for all $t \in T_a^\varepsilon(\overline{x})$ and small ρ_k, we find using (a)

$$\begin{aligned} g(x_k + \tau\rho_k d, t) &= g(x_k, t) + \tau\rho_k \nabla_x g(x_k, t)d + o(\tau\rho_k) \\ &\geq -c\rho_k + \tau\frac{\kappa}{2}\rho_k + o(\tau\rho_k) \\ &= \rho_k(\tau\frac{\kappa}{2} - c) + o(\tau\rho_k) \geq 0, \end{aligned} \tag{28}$$

provided that we choose τ such that $\tau\frac{\kappa}{2} > c$. Secondly, we consider the compact set $T \setminus T_a^\varepsilon(\overline{x})$. By continuity of g, for given $\varepsilon > 0$, there exists $\varepsilon_0 > 0$ such that

$$g(x, t) > 0 \quad \forall t \in T \setminus T_a^\varepsilon(\overline{x}) \text{ and } \forall x \text{ such that } ||x - \overline{x}|| < \varepsilon_0. \tag{29}$$

Now we chose $x_k \in \mathcal{F}(T_k)$ and ρ_k such that with $\varepsilon_1 := \min\{\varepsilon/2, \varepsilon_0/2\}$, the relations $\|x_k - \overline{x}\| < \varepsilon_1$ and $\|\tau\rho_k d\| < \varepsilon_1$ hold. Then, using

$$\|x_k + \tau\rho_k d - \overline{x}\| \leq \|x_k - \overline{x}\| + \tau\rho_k\|d\| < \min\{\varepsilon, \varepsilon_0\},$$

(28) and (29) yield $g(x_k + \tau\rho_k d, t) \geq 0\ \forall t \in T$. □

THEOREM 12. *Let $\overline{x}$ be a local minimizer of $P(T)$ of order p ($p \geq 1$). Suppose that 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 $T_k \subset T$ with $\rho(T_k) \to 0$ and any sequence of solutions $\overline{x}_k$ of the locally restricted problem $P^{local}(T_k)$ (see the definition of discretizability):*

$$||\overline{x}_k - \overline{x}|| \leq \sigma\rho(T_k)^{1/p}. \tag{30}$$

Proof. Consider the SIP restricted to the closed ball cl $B_\kappa(\overline{x})$ ($B_\kappa(\overline{x}) := \{x \mid \|x - \overline{x}\| < \kappa\}$) with small κ chosen such that $\kappa < \varepsilon, \varepsilon_1$ (with ε in (10) and ε_1 in Lemma 7) :

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

Obviously, since $\overline{x} \in \mathcal{F}(T_k)$ and $\mathcal{F}(T_k) \cap \text{cl } B_\kappa(\overline{x})$ is compact (and nonempty), a solution x_k^l of $P^{local}(T_k)$ exists. Note that $\overline{x}$ is the unique (global) minimizer of $P^{local}(T)$. Put $\rho_k :=$

$\rho(T_k)$ and consider any sequence of solutions x_k^l of $P^{local}(T_k)$. In view of $\mathcal{F}(T) \subset \mathcal{F}(T_k)$ and $x_k^l + \tau\rho_k d \in \mathcal{F}(T) \cap \text{cl } B_\kappa(\overline{x})$ (for large k, according to Lemma 7(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 (19)) it follows

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

Finally, the triangle inequality yields using $p \geq 1$,

$$||x_k^l - \overline{x}|| \leq ||x_k^l + \tau\rho_k d - \overline{x}|| + ||\tau\rho_k d|| = O(\rho_k^{1/p}). \tag{31}$$

In particular $||x_k^l - \overline{x}|| < \kappa$ for large k, and $\overline{x}_k := x_k^l$ are (global) minimizers of the problem $P^{local}(T_k)$, restricted to the open neighborhood $B_\kappa(\overline{x})$, *i.e.*, $\overline{x}_k$ are local minimizers of $P(T_k)$. □

Remark The result of Theorem 12 remains true for the global minimization problem P if the feasible set $\mathcal{F}$ of P is restricted to a compact subset in $\mathbb{R}^n$.

It has been shown in [77] that a convergence rate $||\overline{x}_k - \overline{x}|| = O(\rho_k^{2/p})$ occurs if the grids T_k of meshsizes ρ_k are chosen in a special way.

6.4. **Exchange methods.** 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 and the continuous reduction approach in Section 6.2.

Conceptual exchange method

Step k: Given a grid $T_k \subset T$ and a fixed small value $\alpha > 0$.

(1) Compute a solution x_k of $P(T_k)$.
(2) Compute local solutions t_j^k, $j = 1, \ldots, j_k$ ($j_k \geq 1$) of $Q(x_k)$ (cf. (18)) such that one of them, say t_1^k, is a global solution, i.e., $g(x_k, t_1^k) = \min_{t\in T} g(x_k, t)$.
(3) Stop, if $g(x_k, t_1^k) \geq -\alpha$, with an approximate solution $\overline{x} := x_k$. Otherwise, update

$$T_{k+1} = T_k \cup \{t_j^k, \; j = 1, \ldots, j_k\}. \tag{32}$$

THEOREM 13. *Suppose that the (starting) feasible set $\mathcal{F}(T_1)$ is compact. Then, the exchange method (with $\alpha = 0$) either stops with a solution $\overline{x} = x_{k_0}$ of $P(T)$ or the sequence $\{x_k\}$ of solutions of $P(T_k)$ satisfies $d(x_k, \mathcal{S}(T)) \to 0$.*

Proof. We consider the case that the algorithm does not stop with a minimizer of $P(T)$. As in the proof of Theorem 11, and by virtue of our assumptions, a solution x_k of $P(T_k)$ exists, $x_k \in \mathcal{F}(T_1)$, and with the subsequence $x_{k_\nu} \to \overline{x}$ we find

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

Without loss of generality, we write $x_k \to \overline{x}$. 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 $\varphi(x_k) = g(x_k, t_1^k)$ (see Algorithm, step ii) we can write

$$\varphi(\overline{x}) = \varphi(x_k) + \varphi(\overline{x}) - \varphi(x_k) = g(x_k, t_1^k) + \varphi(\overline{x}) - \varphi(x_k).$$

Since $t_1^k \in T_{k+1}$ we have $g(x_{k+1}, t_1^k) \geq 0$ and, by continuity of g and φ, we find

$$\varphi(\overline{x}) \quad \geq \quad \big(g(x_k, t_1^k) - g(x_{k+1}, t_1^k)\big) + \big(\varphi(\overline{x}) - \varphi(x_k)\big) \to 0 \ \text{ for } k \to \infty.$$

$\square$

We refer to the review paper [45] for more details on this approach.

7. GSIP AND RELATED PROBLEMS

In this section we shortly survey the recent developments in GSIP. Optimality conditions have been established in Section 5 and for applications we refer to Section 2.
We wish to discuss relations of GSIP with other important classes of optimization problems namely *bilevel problems* (BL) (*e.g.*, [5]) and *mathematical programs with equilibrium constraints* (MPEC) (*e.g.*, [61]). For shortness we only deal with the first class. Bilevel problems are of the form

$$\text{(33)} \qquad \begin{array}{lll} \text{BL:} & \min_{x,t}\ f(x,t) \quad \text{s.t.} & g(x,t) \geq 0, \\ & \text{and } t \text{ is a solution of} & Q(x):\ \min_t F(x,t)\ \text{ s.t. }\ t \in T(x). \end{array}$$

The problem GSIP can be transformed into a BL program as follows. Let us assume $T(x) \neq \emptyset,\ \forall x$, and recall the (lower level) problem (see (18))

$$\text{(34)} \qquad Q(x): \qquad \min_t\ g(x,t) \quad \text{s.t.} \quad t \in T(x),$$

depending on the parameter x. Then (assuming that $Q(x)$ is solvable) we can write

$$g(x,t) \geq 0\ \forall t \in T(x) \quad \Leftrightarrow \quad g(x,t) \geq 0 \text{ and } t \text{ solves } Q(x).$$

So GSIP takes the BL form:

$$\text{(35)} \qquad \begin{array}{lll} \text{GSIP}_{BL}\text{:} & \min_{x,t}\ f(x) \quad \text{s.t.} & g(x,t) \geq 0, \\ & \text{and } t \text{ is a solution of} & Q(x):\ \min_t\ g(x,t) \quad \text{s.t.} \quad t \in T(x). \end{array}$$

This problem is a BL program with the special property that the objective function f does not depend on t and that the constraint function g in the first level coincides with the lower level objective function.

Remark Note however that there is a subtle difference between the interpretation of the original constraints, $g(x,t) \geq 0\ \forall t \in T(x)$, and the feasibility condition in the bilevel form (35) for the case that $T(x)$ is empty. In the form (35), in this case, because of the additional condition $t \in T(x)$, no feasible point (x,t) exists (for this x). For the GSIP problem however an empty index set $T(x)$ means that there are no constraints and such points x are feasible.

For a comparison between BL and GSIP from a structural and generic viewpoint we refer to [74] (and between BL and MPEC to [7]).

To solve programs with bilevel structure numerically it is convenient to reformulate the problems as nonlinear programs. We will restrict ourselves to the GSIP problem (35) and we assume again that the sets $T(x)$ are defined explicitly via (17) with a C^1-function $u : \mathbb{R}^{n+m} \to \mathbb{R}^q$. Let also g be from C^1. If t is a solution of $Q(x)$ satisfying some constraint qualification, then t must necessarily satisfy the Kuhn-Tucker conditions:

$$\text{(36)} \qquad \nabla_t g(x,t) - \gamma^\top \nabla_t u(x,t) \quad = \quad 0_m^\top \ \text{ and } \gamma^\top u(x,t) = 0,$$

with some multiplier $0_q \leq \gamma \in \mathbb{R}^q$. So we can consider the program

$$
\begin{aligned}
\min_{x,t,\gamma} \ f(x) \qquad \text{s.t.} \quad & g(x,t) \geq 0, \\
& \nabla_t g(x,t) - \gamma^\top \nabla_t u(x,t) = 0_m^\top, \\
& \gamma^\top u(x,t) = 0, \\
& \gamma \geq 0_q, \ u(x,t) \geq 0_q.
\end{aligned} \tag{37}
$$

This program is a relaxation of GSIP_{BL} in the sense that, under a constraint qualification for $Q(x)$, the feasible set of (35) is contained in the feasible set of (37). In particular, any solution (x,t,γ) of (37) with the property that t is a minimizer of $Q(x)$, must also be a solution of the original program. If in addition to the constraint qualification, the problem $Q(x)$ is convex, then (37) is equivalent with the original GSIP program.
In the form (37), GSIP is transformed into a nonlinear program with complementarity constraints and the problems can be solved numerically, for instance by some interior point approach. For GSIP problems this has been done successfully in [75].

REFERENCES

[1] J. Amaya, J.A. Gómez, Strong duality for inexact linear programming problems, Optimization 49 (2001) 243-269.
[2] E.J. Anderson, A.B. Philpott, A.B. (Eds.), Infinite Programming, Springer-Verlag, Berlin, 1985.
[3] E.J. Anderson, P. Nash, Linear Programming in Infinite Dimensional Spaces, Wiley, New York, 1987.
[4] B. Bank, J. Guddat, D. Klatte, B. Kummer, K. Tammer, Non-Linear Parametric Optimization, Birkhäuser Verlag, Basel, 1983.
[5] J.F. Bard, Practical bilevel optimization. Algorithms and applications, Nonconvex Optimization and its Applications 30, Kluwer, Dordrecht, 1998.
[6] A. Ben-Tal, A. Nemirovski, Robust solutions of uncertain linear programs, Operations Research Letters 25 (1999) 1-13.
[7] S.I Birbil, G. Bouza, H. Frenk, G. Still, Equilibrium constrained optimization problems, European Journal of Operational Research 169 3 (2006) 1108-1127.
[8] J.F. Bonnans and A. Shapiro, Perturbation Analysis of Optimization Problems, Springer-Verlag, New York-Berlin-Heidelberg, 2000.
[9] J. M. Borwein, Direct theorems in semi-infinite convex programming, Mathematical Programming 21 (1981) 301-318.
[10] J.M. Borwein, The limiting Lagrangian as a consequence of the Helly's theorem, Journal of Optimization. Theory and Applications 33 (1981) 497-513.
[11] J.M. Borwein, Semi-Infinite Programming Duality: How Special Is It?, in: A.V. Fiacco, K.O. Kortanek (Eds.), Semi-Infinite Programming and Applications, Lecture Notes in Economics and Mathematical Systems 215, 1983, pp. 10-36.
[12] B. Brosowski, Parametric semi-infinite optimization, Verlag Peter Lang, Frankfurt, 1982.
[13] M.J. Cánovas, M.A. López, J. Parra, Upper semicontinuity of the feasible set mapping for linear inequality systems, Set-Valued Analysis 10 (2002), 361–378.
[14] M.J. Cánovas, M.A. López, J. Parra, Stability in the discretization of a parametric semi-infinite convex inequality system, Mathematics of Operations Research 27 (2002) 755–774.
[15] M.J. Cánovas, M.A. López, J, Parra, M.I. Todorov, Stability and Well-Posedness in Linear Semi-Infinite Programming, SIAM Journal on Optimization 10 (1999) 82-98.
[16] M.J. Cánovas, M.A. López, J. Parra, M.I. Todorov, Solving strategies and well-posedness in linear semi-infinite programming, Annals of Operations Research 101 (2001) 171–190.
[17] A. Charnes, W.W. Cooper, K.O.Kortanek, Duality, Haar programs and finite sequence spaces, Proceedings of the National Academy of Science 48 (1962) 783-786.
[18] A. Charnes, W.W. Cooper, K.O.Kortanek, Duality in semi-infinite programs and some works of Haar and Carathéodory, Management Sciences 9 (1963) 209-228.

[19] A. Charnes, W.W. Cooper, K.O. Kortanek, On the Theory of Semi-Infinite Programming and a Generalization of the Kuhn-Tucker Saddle Point Theorem for Arbitrary Convex Functions, Naval Research Logistics Quarterly 16 (1969) 41-51.
[20] M. Dall'Aglio, On Some Applications of LSIP to Probability and Statistics, in M.A. Goberna, M.A. López (Eds.), Semi-Infinite Programming. Recent Advances, Nonconvex Optimization and Its Applications 57, Kluwer 2001, pp. 237-254.
[21] M. Dambrine, M. Pierre, About stability of equilibrium shapes, Mathem. Modelling and Num. Analysis 34 (2000) 811-834.
[22] U. Faigle, W. Kern, G. Still, Algorithmic Principles of Mathematical Programming, Kluwer, Dordrecht, 2002.
[23] M.D. Fajardo, M.A. López, Locally Farkas-Minkowski systems in convex semi-infinite programming, Journal of Optimization Theory and Applications 103 (1999) 313-335.
[24] A. Ferrer, Applying global optimization to a problem in short-term hydrothermal scheduling. In Generalized convexity, generalized monotonicity and applications, Nonconvex Optimization and Its Applications 77, Springer, New York, 2005, pp. 263–285.
[25] A.V. Fiacco, K.O. Kortanek (Eds.), Semi-Infinite Programming and Applications, Lecture Notes in Economics and Mathematical Systems 215, 1983.
[26] V. Gayá, M.A. López, V.N. Vera de Serio, Stability in convex semi-infinite programming and rates of convergence of optimal solutions of discretized finite subproblems, Optimization 52 (2003) 693–713.
[27] K. Glashoff, S.-A. Gustafson, Linear Optimization and Approximation, Springer Verlag, Berlin, 1983.
[28] M.A. Goberna, Linear Semi-infinite Optimization: Recent Advances, in: V. Jeyakumar, A.M. Rubinov (Eds.), Continuous Optimization, Current Trends and Modern Applications Series: Applied Optimization, 99, 2005.
[29] M.A. Goberna, M.A. López, Optimal value function in semi-infinite programming, Journal of Optimization. Theory and Applications 59 (1988) 261-280.
[30] M.A. Goberna, M.A. López, Topological Stability of Linear Semi-Infinite Inequality Systems, Journal of Optimization Theory and Applications 89 (1996) 227-236.
[31] M.A. Goberna, M.A. López, Reduction and discrete optimization in linear semi-infinite programming, Optimization 18 (1987) 643-658.
[32] M.A. Goberna, M.A. López, Linear Semi-Infinite Optimization, Wiley, Chichester, 1998.
[33] M.A. Goberna, M.A. López, A comprehensive survey of linear semi-infinite optimization theory, in: R.M. Reemtsen and J. Rückmann (Eds.), Semi-infinite programming, Nonconvex Optimization and Its Applications 25, Kluwer, Boston, 1998, 3–27.
[34] M.A. Goberna, M.A. López, Linear semi-infinite programming theory: an updated survey, European Journal of Operations Research 143 (2002) 390–405.
[35] M.A. Goberna and M.A. López and M.I. Todorov, *Stability Theory for Linear Inequality Systems*, SIAM Journal on Matrix Analysis and Applications, 17, 730-743, (1996).
[36] M.A. Goberna and M.A. López and M.I. Todorov, *Stability Theory for Linear Inequality Systems II: Upper Semicontinuity of the Solution Set Mapping*, SIAM Journal on Optimization, 7, 1138-1151, (1997).
[37] W. Gómez, J.-A. Gómez, Cutting plane algorithms for robust conic convex optimization problems, Optimization Methods and Software, to appear (2005).
[38] F. Guerra, J.A. Orozco, J.-J. Rückmann, On constraint qualifications in semi-infinite optimization, Parametric optimization and related topics VII, Aportaciones Mat. Investig. 18, Soc. Mat. Mexicana, México, 2004, pp. 133-141.
[39] J. Guddat, F. Guerra-Vázquez, H.Th. Jongen, Parametric optimization: singularities, pathfollowing and jumps, B. G. Teubner, Stuttgart; Wiley, Chichester, 1990.
[40] S.-A. Gustafson, K.O. Kortanek, Numerical treatment of a class of semi-infinite programming problems, Naval Research Logistics Quarterly 20 (1973) 477-504.
[41] A. Haar, Über lineare Ungleichungen, Acta Math. Szeged 2 (1924) 1-14.
[42] R. Hettich (Ed.), Semi-infinite programming, Lecture Notes in Control and Information Sciences, Springer, Berlin, 1979.
[43] R. Hettich, P. Zencke, Numerische Methoden der Approximation und der semi-infiniten Optimierung, Teubner, Stuttgart, 1982.
[44] R. Hettich, G. Still, On Generalized semi-infinite programming problems, Working paper, University of Trier, 1986.

[45] R. Hettich, K.O. Kortanek, Semi-infinite programming: Theory, methods and applications, SIAM Review 35 (1993), 380-429.
[46] R. Hettich, G. Still, Semi-infinite programming models in Robotics, in: G. Guddat et al. (Eds.), Parametric Optimization and Related Topics II, Academie Verlag, Berlin, 1991.
[47] R. Hettich, G. Still, Second order optimality conditions for generalized semi-infinite programming problems, Optimization 34 (1995) 195-211.
[48] A. Hoffmann, R.Reinhardt, On reverse Chebyshev approximation problems, Technical University of Illmenau, Preprint No. M08/94, 1994.
[49] F. John, Extremum problems with inequalities as subsidiary conditions, Studies and Essays: Courant Anniversary Volume, Interscience Publisher (Wiley), New York, 1948, 187-204.
[50] H. Th. Jongen, P. Jonker, F. Twilt, Nonlinear Optimization in finite Dimensions, Kluwer, Dordrecht, 2000.
[51] H. Th Jongen, G. Zwier, On the local structure of the feasible set in semi-infinite optimization, in: Brosowski, Deutsch (Eds.), International Series of Numerical Mathematics 72, Birkhäuser Verlag, Basel, 1984, pp. 185-202.
[52] H. Th. Jongen, J.-J. Rückmann, On stability and deformation in semi-infinite optimization, in: R.M. Reemtsen, J.-J. Rückmann (Eds.), Semi-infinite programming, Nonconvex Optimization and Its Applications 25, Kluwer, Boston, 1998, pp. 29-67.
[53] 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.
[54] A. Kaplan, R. Tichatschke, On the numerical treatment of a class of semi-infinite terminal problems, Optimization 41 (1997) 1-36.
[55] D. Klatte, R. Henrion, Regularity and stability in nonlinear semi-infinite optimization, in: R.M. Reemtsen, J.-J. Rückmann (Eds.), Semi-infinite programming, Nonconvex Optimization and Its Applications 25, Kluwer, Boston, 1998, pp. 69-102.
[56] K.O. Kortanek, On the 1962-1972 decade of semi-infinite programming: a subjective view, in: M.A. Goberna, M.A. López (Eds.), Semi-infinite programming. Recent Advances, Nonconvex Optimization and Its Applications 57, Kluwer, Dordrecht, 2001, pp. 3-41.
[57] W. Krabs, Optimization and Approximation, Wiley, New York, 1979.
[58] T. León, E. Vercher, Optimization under uncertainty and linear semi-infinite programming: a survey, in: M.A. Goberna, M.A. López (Eds.), Semi-infinite programming. Recent Advances, Nonconvex Optimization and Its Applications 57, Kluwer, Dordrecht, 2001, pp. 327-348.
[59] D.T. Luc, J.E. Martínez-Legaz, A. Seeger, Least deviation decomposition with respect to a pair of convex set, Journal of Convex Analysis 6 (1999) 115–140.
[60] W. Li, C. Nahak, I. Singer, Constraint Qualifications for Semi-Infinite Systems of Convex Inequalities, SIAM Journal on Optimization 11 (2000) 31-52.
[61] Z.-Q. Luo, J.-S. Pang, D. Ralph, Mathematical Programs with Equilibrium Constraints, Cambridge University Press, Cambridge, 1997.
[62] Y. Nesterov, A. Nemirovski, Interior point polynomial algorithms in convex programming, SIAM Studies in Applied Mathematics 13, Philadelphia, 1994.
[63] J. Outrata, M. Kovcvara, J. Zowe, Nonsmooth approach to optimization problems with equilibrium constraints. Theory, applications and numerical results, Nonconvex Optimization and Its Applications 28, Kluwer, Dordrecht, 1998.
[64] R. Puente, V.N. Vera de Serio, Locally Farkas-Minkowski linear semi-infinite systems, TOP 7 (1999) 103-121.
[65] E. Polak, On the mathematical foundation of nondifferentiable optimization in engineering design, SIAM Review 29 (1987) 21-89.
[66] E. Polak, Optimization. Algorithms and consistent approximations, Applied Mathematical Sciences 124, Springer-Verlag, New York, 1997.
[67] R. Reemtsen, S. Görner, Numerical methods for semi-infinite programming: a survey, in: R. Reemtsen, J.-J. Rückmann (Eds.), Semi-infinite programming, Nonconvex Optimization and Its Applications 25, Kluwer, Boston, 1998, pp. 195-275.
[68] R.T. Rockafellar, Convex Analysis, Princeton University Press, Princeton, New Jersey, 1970.
[69] W.W. Rogosinski, Moments of non-negative mass, Proc. Roy. Soc. London (Ser. A) 245 (1958) 1-27.

[70] A. Shapiro, On duality theory of conic linear problems, in: M.A. Goberna, M.A. López (Eds.), Semi-infinite programming. Recent Advances, Nonconvex Optimization and Its Applications 57, Kluwer, Dordrecht, 2001, pp. 135-165.
[71] A.L. Soyster, Convex programming with set-inclusive constraints and applications to inexact linear programming, Operations Research 21 (1973) 1154-1157.
[72] O. Stein , Bilevel strategies in Semi-infinite Programming, Kluwer, Boston (2003).
[73] O. Stein, G. Still, On optimality conditions for generalized semi-infinite programming problems, Journal of Optimization Theory and Applications 104 (2000) 443-458.
[74] O. Stein, G. Still, On generalized semi-infinite optimization and bilevel optimization, European Journal of Operational Research 142(3) (2002) 444-462.
[75] O. Stein, G. Still, Solving semi-infinite optimization problems with interior point techniques, SIAM Journal on Control and Optimization 42(3) (2003) 769–788.
[76] Still G., Generalized semi-infinite programming: Theory and methods, European Journal of Operational Research 119, (1999), 301-313.
[77] Still G., Discretization in semi-infinite programming: the rate of convergence, Math. Program. 91 (2001), no. 1, Ser. A, 53–69.
[78] G. Still, Approximation theory methods for solving elliptic eigenvalue problems, Z. Angew. Math. Mech. 83 (7) (2003) 468-478.
[79] G. Still, Approximation and Optimization: Classical results and new developments, Parametric optimization and related topics VII, Aportaciones Mat. Investig. 18, Soc. Mat. Mexicana, México, 2004, pp. 207-233.
[80] G. Still, E. Haaren-Retagne, R. Hettich, A numerical comparison of two approaches to compute membrane eigenvalues by defect-minimization, International Series of Numerical Mathematics 96, Birkhä̈user Verlag, 1991, pp. 209-224.
[81] R. Tichatschke, Lineare Semi-Infinite Optimierungsaufgaben und ihre Anwendungen in der Approximationstheorie, Karl-Marx-Stadt: Wissenschaftliche Schriftenreihe der Technischen Hochschule, 1981.
[82] R. Tichatschke, R. Hettich, G. Still, Connections between generalized inexact and semi-infinite linear programming, ZOR-Methods and Models of OR 33 (1989) 367-382.
[83] S.N. Tschernikow, O teoreme chaara dlja beskonteschnych sistem linejnych neravenctv, Uspekhi Matematicheskikh Nauk 113 (1963) 199-200.
[84] A. Vaz, E. Fernandes, M. Gomes, SIPAMPL: Semi-Infinite Programming with AMPL, ACM Trans. Math. Software 30(1) (2004) 47–61.
[85] A. Vaz, E. Fernandes, M. Gomes, A sequential quadratic programming with a dual parametrization approach to nonlinear semi-infinite programming, Top 11(1) (2003) 109–130.
[86] Z. Yong-Jin, Generalizations of some fundamental theorems on linear inequalities, Acta Mathematica Sinica 16 (1966) 25-40.
[87] **Extended SIP bibliography:** http://wwwhome.math.utwente.nl/˜stillgj/sip.html.