Optimization
Vol. 58, No. 2, February 2009, 133–161

INVITED REVIEW

Semi-infinite programming, duality, discretization and optimality conditions†

Alexander Shapiro*

School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332-0205, USA

(*Received 3 July 2008; final version received 4 December 2008*)

The aim of this article is to give a survey of some basic theory of semi-infinite programming. In particular, we discuss various approaches to derivations of duality, discretization, and first- and second-order optimality conditions. Some of the surveyed results are well known while others seem to be less noticed in that area of research.

Keywords: semi-infinite programming; conjugate duality; convex analysis; discretization; Lagrange multipliers; first- and second-order optimality conditions

1. Introduction

The aim of this article is to give a survey of some basic theory of semi-infinite programming (SIP) problems of the form

$$\operatorname*{Min}_{x\in\mathbb{R}^n} f(x) \text{ subject to } g(x,\omega) \le 0, \quad \omega \in \Omega. \tag{1.1}$$

Here Ω is a (possibly infinite) index set, $\overline{\mathbb{R}} = \mathbb{R} \cup \{+\infty\} \cup \{-\infty\}$ denotes the extended real line, $f : \mathbb{R}^n \to \overline{\mathbb{R}}$ and $g : \mathbb{R}^n \times \Omega \to \mathbb{R}$. The above optimization problem is performed in the finite-dimensional space $\mathbb{R}^n$ and, if the index set Ω is infinite, is a subject to an infinite number of constraints, therefore it is referred to as a SIP problem.

There are numerous applications which lead to SIP problems. We can refer the interested reader to survey papers [11,12,21,27] where many such examples are described. There are also several books where SIP is discussed from theoretical and computational points of view (e.g. [4,9,10,22,23,31]). Compared with recent surveys [11,21], we use a somewhat different approach, although, of course, there is a certain overlap with these papers. For some of the presented results, for the sake of completeness, we outline proofs while more involved assertions will be referred to the literature.

It is convenient to view the objective function $f(x)$ as an extended real-valued function which is allowed to take $+\infty$ or $-\infty$ values. In fact, we always assume in the subsequent analysis that $f(x)$ is *proper*, i.e. its domain

$$\operatorname{dom} f = \left\{x \in \mathbb{R}^n : f(x) < +\infty\right\}$$

*Email: ashapiro@isye.gatech.edu
†Invited survey.

ISSN 0233–1934 print/ISSN 1029–4945 online

DOI: 10.1080/02331930902730070
http://www.informaworld.com

is nonempty and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$. Of course, it suffices to perform optimization in (1.1) over $x \in \mathrm{dom}\, f$. In that formulation an additional constraint of the form $x \in X$, where X is a subset $\mathbb{R}^n$, can be absorbed into the objective function by adding to it the indicator function $\mathbb{I}_X(\cdot)$ (recall that $\mathbb{I}_X(x) = 0$, if $x \in X$, and $\mathbb{I}_X(x) = +\infty$, if $x \neq X$). As we progress in the analysis we will need to impose more structure on the involved functions.

It is said that the SIP problem (1.1) is *linear* if the objective function and the constraints are linear in x, i.e. it can be written in the form

$$\operatorname*{Min}_{x \in \mathbb{R}^n} c^{\mathrm{T}} x \text{ subject to } a(\omega)^{\mathrm{T}} x + b(\omega) \leq 0, \quad \omega \in \Omega, \tag{1.2}$$

for some vector $c \in \mathbb{R}^n$ and functions $a : \Omega \to \mathbb{R}^n$, $b : \Omega \to \mathbb{R}$.

Definition 1.1 We say that the SIP problem (1.1) is convex if for every $\omega \in \Omega$ the function $g(\cdot, \omega) : \mathbb{R}^n \to \mathbb{R}$ is convex and the objective function $f(\cdot)$ is proper convex and lower semicontinuous.

Of course, the linear SIP problem (1.2) is convex.

This article is organized as follows. In the next section, we discuss duality of convex SIP problems from two points of view. Namely, first without assuming any particular structure of the index set Ω, and then by assuming certain topological properties of Ω and $g(x, \omega)$. There exists an extensive literature on duality of convex SIP problems (see, e.g. [12] and [10], and more recent surveys [11,21]). The approach that we use is based on conjugate duality (cf. [24,25]). In Section 3, we review some results on discretization of SIP problems with relation to their duality properties. Section 4 is devoted to first-order optimality conditions for convex and for smooth (differentiable) SIP problems. In Section 5, we review second-order necessary and/or sufficient optimality conditions. The material of that section is based, to some extent, on [4]. Finally, in Section 6 we discuss rates of convergence of optimal solutions of finite discretizations of SIP problems.

We use the following notation and terminology throughout this article. The notation ':=' stands for 'equal by definition'. We denote by $\mathcal{F}$ the feasible set of problem (1.1),

$$\mathcal{F} := \{x \in \mathrm{dom} f : g(x, \omega) \leq 0, \ \omega \in \Omega\}.$$

Of course, if $f : \mathbb{R}^n \to \mathbb{R}$ is real valued, then $\mathrm{dom}\, f = \mathbb{R}^n$. For a matrix (vector) A we denote by A^{T} its transpose. A vector $x \in \mathbb{R}^n$ is assumed to be a column vector, so that $x^{\mathrm{T}} y = \sum_{i=1}^n x_i y_i$ is the scalar product of two vectors $x, y \in \mathbb{R}^n$. For a function $v : \mathbb{R}^n \to \overline{\mathbb{R}}$ we denote by $v^*(x^*) = \sup_x \{(x^*)^{\mathrm{T}} x - v(x)\}$ its conjugate, and by $v^{**}(x)$ its biconjugate, i.e. conjugate of $v^*(x^*)$. The subdifferential $\partial v(x)$, at a point x where $v(x)$ is finite, is defined as the set of vectors γ such that

$$v(y) \geq v(x) + \gamma^{\mathrm{T}}(y - x), \quad \forall y \in \mathbb{R}^n.$$

It is said that $v(\cdot)$ is subdifferentiable at a point x if $v(x)$ is finite and $\partial v(x)$ is nonempty. Unless stated otherwise, the subdifferential $\partial g(x, \omega)$, gradient $\nabla g(x, \omega)$ and Hessian matrix $\nabla^2 g(x, \omega)$ of function $g(x, \omega)$ are taken with respect to x. By $Df(x)$ and $D^2 f(x)$ we denote the first- and second-order derivatives of the function (mapping) $f(x)$. Note that if $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then $Df(x)h = h^{\mathrm{T}} \nabla f(x)$ and $D^2 f(x)(h, h) = h^{\mathrm{T}} \nabla^2 f(x) h$.

For a set $S \subset \mathbb{R}^n$, we denote by int(S) its interior, by conv(S) its convex hull and by dist$(x, S) := \inf_{y \in S} \|x - y\|$. For two sets $A, B \subset \mathbb{R}^n$ we denote by

$$\mathbb{D}(A, B) := \sup_{x \in A} \text{dist}(x, B)$$

deviation of set A from set B. The support function of set S is

$$\sigma_S(x) := \sup_{h \in S} h^{\mathrm{T}} x,$$

also denoted $\sigma(x, S)$. Note that this support function remains the same if the set S is replaced by the topological closure of conv(S). The contingent cone $T_S(x)$, to S at point $x \in S$, is formed by vectors h such that there exist sequences $t_k \downarrow 0$ and $h_k \to h$ such that $x + t_k h_k \in S$ for all $k \geq 1$. It is said that a function $f : \mathbb{R}^n \to \mathbb{R}$ is directionally differentiable at a point $x \in \mathbb{R}^n$ if its directional derivative

$$f'(x, h) := \lim_{t \downarrow 0} \frac{f(x + th) - f(x)}{t}$$

exists for every $h \in \mathbb{R}^n$. Moreover, it is said that $f(\cdot)$ is Hadamard directionally differentiable at x if

$$f'(x, h) = \lim_{\substack{t \downarrow 0 \\ h' \to h}} \frac{f(x + th') - f(x)}{t}.$$

For locally Lipschitz functions directional differentiability implies Hadamard directional differentiability (e.g. [28]). By $\delta(\omega)$ we denote the (Dirac) measure of mass one at the point $\omega \in \Omega$.

2. Duality

In order to formulate a dual of the SIP problem (1.1), we need to embed the constraints into an appropriate functional space paired with a dual space. That is, let $\mathcal{Y}$ be a linear space of functions $\gamma : \Omega \to \mathbb{R}$. Consider the mapping $G : x \mapsto g(x, \cdot)$ from $\mathbb{R}^n$ into $\mathcal{Y}$, i.e. $G(x) = g(x, \cdot) \in \mathcal{Y}$ is a real-valued function defined on the set Ω. Depending on what we assume about the index set Ω and the constraint functions $g(x, \cdot)$, we can consider various constructions of the space $\mathcal{Y}$. In the following analysis we deal with the following two constructions.

In the general framework when we do not make any structural assumptions, we can take $\mathcal{Y} := \mathbb{R}^\Omega$ to be the space of all functions $\gamma : \Omega \to \mathbb{R}$ equipped with natural algebraic operations of addition and multiplication by a scalar. We associate with this space as the linear space $\mathcal{Y}^*$ of functions $\gamma^* : \Omega \to \mathbb{R}$ such that only a finite number of values $\gamma^*(\omega)$, $\omega \in \Omega$, are nonzero. For $\gamma^* \in \mathcal{Y}^*$ we denote by

$$\text{supp}(\gamma^*) := \{\omega \in \Omega : \gamma^*(\omega) \neq 0\}$$

its support set, and for $\gamma^* \in \mathcal{Y}^*$ and $\gamma \in \mathcal{Y}$ define the scalar product

$$\langle \gamma^*, \gamma \rangle := \sum \gamma^*(\omega) \gamma(\omega), \tag{2.1}$$

where the summation in (2.1) is performed over ω in the (finite) set supp(γ^*).

Another important case, which we discuss in details, is when the following assumption holds.

(A1) The set Ω is a compact metric space and the function $g : \mathbb{R}^n \times \Omega \to \mathbb{R}$ is continuous on $\mathbb{R}^n \times \Omega$.

In that case we can take $\mathcal{Y} := C(\Omega)$, where $C(\Omega)$ denotes the space of continuous functions $\gamma : \Omega \to \mathbb{R}$ equipped with the sup-norm $\|\gamma\| := \sup_{\omega \in \Omega} |\gamma(\omega)|$. The space $C(\Omega)$ is a Banach space and its dual $\mathcal{Y}^*$ is the space of finite signed measures on $(\Omega, \mathcal{B})$, where $\mathcal{B}$ is the Borel sigma algebra of Ω, with the scalar product of $\mu \in \mathcal{Y}^*$ and $\gamma \in \mathcal{Y}$ given by the integral

$$\langle \mu, \gamma \rangle := \int_\Omega \gamma(\omega) \mathrm{d}\mu(\omega). \tag{2.2}$$

The dual norm of $\mu \in C(\Omega)^*$ is $|\mu|(\Omega)$, where $|\mu|$ is the total variation of measure μ. For a measure $\mu \in C(\Omega)^*$ we denote by $\mathrm{supp}(\mu)$ its support, i.e. $\mathrm{supp}(\mu)$ is the smallest closed subset Υ of Ω such that $|\mu|(\Omega \setminus \Upsilon) = 0$. Of course, if $\mu = \sum_{i=1}^m \lambda_i \delta(\omega_i)$, then μ has finite support consisting of points ω_i such that $\lambda_i \neq 0$, $i = 1, \dots, m$. Note that assumption (A1) implies that the mapping $G(x)$, from $\mathbb{R}^n$ into the normed space $C(\Omega)$, is continuous.

The constraints $g(x, \omega) \le 0$, $\omega \in \Omega$, can be written in the form $G(x) \in K$, where

$$K := \{\gamma \in \mathcal{Y} : \gamma(\omega) \le 0, \ \omega \in \Omega\} \tag{2.3}$$

is the cone of nonpositive-valued functions in the corresponding functional space $\mathcal{Y}$. The polar (negative dual) of this cone is the cone

$$K^* := \left\{\gamma^* \in \mathcal{Y}^* : \langle \gamma^*, \gamma \rangle \le 0, \ \forall \gamma \in K\right\}. \tag{2.4}$$

We have that $K^* = \{\lambda \in \mathcal{Y}^* : \lambda \succeq 0\}$, where for $\mathcal{Y} = \mathbb{R}^\Omega$ we mean by $\lambda \succeq 0$ that $\lambda(\omega) \ge 0$ for all $\omega \in \Omega$, and for $\mathcal{Y} = C(\Omega)$ we mean by $\lambda \succeq 0$ that measure λ is nonnegative (i.e. $\lambda(A) \ge 0$ for any $A \in \mathcal{B}$). Note also that in the same way we can define the negative dual $K^{**} \subset \mathcal{Y}$ of the cone K^*. In both cases of considered paired spaces we have that $K^{**} = K$.

We associate with problem (1.1) its Lagrangian $L(x, \lambda) := f(x) + \langle \lambda, G(x) \rangle$, where $(x, \lambda) \in \mathbb{R}^n \times \mathcal{Y}^*$. That is,

$$L(x, \lambda) = f(x) + \sum_{\omega \in \mathrm{supp}(\lambda)} \lambda(\omega) g(x, \omega), \quad \text{for } \mathcal{Y} = \mathbb{R}^\Omega, \tag{2.5}$$

and

$$L(x, \lambda) = f(x) + \int_\Omega g(x, \omega) \mathrm{d}\lambda(\omega), \quad \text{for } \mathcal{Y} = C(\Omega). \tag{2.6}$$

In both cases we have that

$$\sup_{\lambda \succeq 0} L(x, \lambda) = \begin{cases} f(x), & \text{if } g(x, \omega) \le 0, \ \forall \omega \in \Omega, \\ +\infty, & \text{otherwise}, \end{cases} \tag{2.7}$$

and hence the primal problem (1.1) can be written as

$$\operatorname*{Min}_{x \in \mathbb{R}^n} \sup_{\lambda \succeq 0} L(x, \lambda). \tag{2.8}$$

Its Lagrangian dual is obtained by interchanging the 'Min' and 'Max' operators, that is

$$\operatorname*{Max}_{\lambda \succeq 0} \inf_{x \in \mathbb{R}^n} L(x, \lambda). \tag{2.9}$$

We denote by (P) and (D) the primal problem (1.1) and its dual (2.9), respectively, and by val(P) and val(D) and Sol(P) and Sol(D) their respective optimal values and sets of optimal solutions. It follows immediately from the minimax formulations (2.8) and (2.9) that val$(P) \geq$ val(D), i.e. the weak duality always holds here. It is said that the 'no duality gap' property holds if val$(P) =$ val(D), and the 'strong duality' property holds if val$(P) =$ val(D) and the dual problem has an optimal solution.

In order to proceed, let us embed the dual problem into the parametric family

$$\operatorname*{Max}_{\lambda \in \mathcal{Y}^*} \varphi(\lambda, y), \tag{2.10}$$

where

$$\varphi(\lambda, y) := \begin{cases} \inf_{x \in \mathbb{R}^n} \left\{ L(x, \lambda) - y^{\mathrm{T}} \to x \right\}, & \text{if } \lambda \in K^*, \\ -\infty, & \text{if } \lambda \notin K^*. \end{cases} \tag{2.11}$$

Note that the function $\varphi : \mathcal{Y}^* \times \mathbb{R}^n \to \overline{\mathbb{R}}$ is the infimum of affine functions, and hence is concave. It follows that the min-function

$$\vartheta(y) := \inf_{\lambda \in \mathcal{Y}^*} \{-\varphi(\lambda, y)\} = -\sup_{\lambda \in \mathcal{Y}^*} \varphi(\lambda, y) \tag{2.12}$$

is a an extended real-valued convex function. Clearly val$(D) = -\vartheta(0)$.

It is not difficult to calculate (cf. [30]) that the conjugate of the function $\vartheta(y)$ is

$$\vartheta^*(y^*) = \sup_{\lambda \in K^*} L^{**}(y^*, \lambda), \tag{2.13}$$

where $L^{**}(\cdot, \lambda)$ is the biconjugate of the function $L(\cdot, \lambda)$. If, moreover, the SIP problem (1.1) is convex, then for every $\lambda \in K^*$, the function $L(\cdot, \lambda)$ is proper convex and lower semicontinuous. By the Fenchel–Moreau theorem it follows that for all $\lambda \in K^*$, the function $L^{**}(\cdot, \lambda)$ coincides with $L(\cdot, \lambda)$, and hence

$$\vartheta^*(y^*) = \sup_{\lambda \in K^*} L(y^*, \lambda). \tag{2.14}$$

Since $\vartheta^{**}(0) = -\inf_{y^* \in \mathbb{R}^n} \vartheta^*(y^*)$, it follows by (2.8) that val$(P) = -\vartheta^{**}(0)$. Moreover, if $\vartheta^{**}(0)$ is finite, then

$$\partial\vartheta^{**}(0) = \arg\max_{y^* \in \mathbb{R}^n} \{-\vartheta^*(y^*)\} = -\arg\min_{y^* \in \mathbb{R}^n} \left\{ \sup_{\lambda \in K^*} L(y^*, \lambda) \right\},$$

and hence Sol$(P) = -\partial\vartheta^{**}(0)$. By the theory of conjugate duality [25], we have the following results (cf. [30]).

THEOREM 2.1 *Suppose that the SIP problem* (1.1) *is convex and* $\vartheta^{**}(0) < +\infty$. *Then the following holds*: (i) val$(D) =$ val(P) *iff the function* $\vartheta(y)$ *is lower semicontinuous at* $y = 0$, (ii) val$(D) =$ val(P) *and* Sol(P) *is nonempty iff the function* $\vartheta(y)$ *is subdifferentiable at* $y = 0$, *in which case* Sol$(P) = -\partial\vartheta(0)$.

The assertion (i) of the above theorem gives necessary and sufficient conditions for the 'no duality gap' property in terms of lower semicontinuity of the min-function $\vartheta(y)$.

However, it might not be easy to verify the lower semicontinuity of $\vartheta(y)$ in particular situations. Of course, if $\vartheta(y)$ is continuous at $y=0$, then it is lower semicontinuous at $y=0$. By convexity of $\vartheta(\cdot)$ we have that if $\vartheta(0)$ is finite, then $\vartheta(y)$ is continuous at $y=0$ iff $\vartheta(y) < +\infty$ for all y in a neighbourhood of 0. This leads to the following result (cf. [30]).

THEOREM 2.2 *Suppose that the SIP problem* (1.1) *is convex and* val(P) *is finite. Then the following statements are equivalent*: (i) *the min-function* $\vartheta(y)$ *is continuous at* $y=0$, (ii) Sol(P) *is nonempty and bounded*, (iii) val$(D)=$val(P) *and* Sol(P) *is nonempty and bounded*, (iv) *the following condition holds*: *there exists a neighbourhood* $\mathcal{N}$ *of* $0\in\mathbb{R}^n$ *such that for every* $y\in\mathcal{N}$ *there exists* $\lambda\in K^*$ *such that*

$$\inf_{x\in\mathbb{R}^n}\left\{L(x,\lambda)-y^{\mathrm{T}}x\right\} > -\infty. \tag{2.15}$$

In particular, we have that if the SIP problem (P) is convex and its optimal solutions set Sol(P) is nonempty and bounded, then the 'no duality gap' property follows.

The above results do not involve any structural assumptions about the index set Ω and constraint functions $g(x,\cdot)$ and do not say anything about the existence of an optimal solution of the dual problem (D). Now suppose that the assumption (A1) holds. As it was discussed earlier, in that case we can take $\mathcal{Y}=C(\Omega)$ and use its dual space $\mathcal{Y}^*=C(\Omega)^*$ of finite signed Borel measures on Ω. Consider the following problem

$$\operatorname*{Min}_{x\in\mathbb{R}^n} f(x) \text{ subject to } g(x,\omega)+z(\omega)\le 0, \quad \omega\in\Omega, \tag{2.16}$$

parameterized by $z\in\mathcal{Y}$. Let $v(z)$ be the optimal value of the above problem (2.16). Clearly for $z=0$, problem (2.16) coincides with the SIP problem (P) and hence $v(0)=$val(P). By the standard theory of conjugate duality we have here that val$(D)=v^{**}(0)$, where the dual problem (D) and conjugate operations are evaluated with respect to the paired spaces $\mathcal{Y}=C(\Omega)$ and $\mathcal{Y}^*=C(\Omega)^*$. Also if problem (P) is convex, then the (extended real valued) function $v: C(\Omega)\to\overline{\mathbb{R}}$ is convex. Therefore, by the Fenchel–Moreau theorem, it follows that if problem (P) is convex and val(D) is finite, then val$(P)=$val(D) iff the function $v(z)$ is lower semicontinuous at $z=0$ (in the norm topology of $C(\Omega)$).

Definition 2.1 It is said that Slater condition holds for the problem (P) if there exists $\bar{x}\in\operatorname{dom} f$ such that $g(\bar{x},\omega)<0$ for all $\omega\in\Omega$.

Since, under assumption (A1), Ω is compact and $g(\bar{x},\cdot)$ is continuous, the condition '$g(\bar{x},\omega)<0$ for all $\omega\in\Omega$' implies that there is $\varepsilon>0$ such that $g(\bar{x},\omega)<-\varepsilon$ for all $\omega\in\Omega$. That is, Slater condition means that $G(\bar{x})$ belongs to the interior of the set $K\subset C(\Omega)$ (recall that $G(\bar{x})$ is the function $g(\bar{x},\cdot)$ viewed as an element of the space $C(\Omega)$). This, in turn, implies that $v(z)\le f(\bar{x})<+\infty$ for all z in a neighbourhood of $0\in C(\Omega)$, i.e. that

$$0\in\operatorname{int}(\operatorname{dom} v). \tag{2.17}$$

It is possible to show that the converse is also true and the following results hold (e.g. [4, Section 2.5.1].

THEOREM 2.3 *Suppose that problem* (P) *is convex*, *assumption* (A1) *is fulfilled and* val(P) *is finite. Then the following statements are equivalent*: (i) *the optimal value function* $v(z)$ *is continuous at* $z=0$, (ii) *the regularity condition* (2.17) *holds*, (iii) *Slater condition holds*, (iv) *the set* Sol(D) *is nonempty and bounded*, (v) val$(P)=$val(D) *and the set* Sol(D) *is nonempty and bounded.*

By boundedness of the set Sol(D) we mean that it is bounded in the total variation norm of $C(\Omega)^*$, which is the dual of the sup-norm of $C(\Omega)$. Note that if $\mu \in C(\Omega)^*$ is a nonnegative measure, then its total variation norm is equal to $\mu(\Omega)$, and if $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$, then the total variation norm of μ is equal to $\sum_{i=1}^{m} |\lambda_i|$.

Now consider the linear SIP problem (1.2). In the framework of $\mathcal{Y} = \mathbb{R}^{\Omega}$ its Lagrangian dual is

$$\operatorname*{Max}_{\lambda \geq 0} \sum \lambda(\omega) b(\omega) \text{ subject to } c + \sum \lambda(\omega) a(\omega) = 0. \tag{2.18}$$

Recall that $\lambda \in \mathcal{Y}^*$ and the summation in (2.18) is taken over $\omega \in \operatorname{supp}(\lambda)$. The min-function $\vartheta(y)$, defined in (2.12), takes here the form

$$\vartheta(y) = \inf\left\{ -\sum \lambda(\omega) b(\omega) : c + \sum \lambda(\omega) a(\omega) = y,\ \lambda \in K^* \right\}. \tag{2.19}$$

Therefore for the linear SIP, conditions (i)–(iv) of Theorem 2.2 are equivalent to the condition:

$$0 \in \operatorname{int}\left\{ y \in \mathbb{R}^n : y = c + \sum \lambda(\omega) a(\omega),\ \lambda \in K^* \right\}. \tag{2.20}$$

Condition (2.20) is well known in the duality theory of linear SIP (cf. [9]). It is equivalent to the condition that $-c \in \operatorname{int}(M)$, where M is the convex cone generated by the vectors $a(\omega)_{\omega \in \Omega}$ (cf. [10]).

3. Discretization

It turns out that the 'no duality gap' and 'strong duality' properties are closely related to the discretization of the semi-infinite problem (1.1). That is, for a given (nonempty) finite set $\{\omega_1, \ldots, \omega_m\} \subset \Omega$ consider the following optimization problem

$$\operatorname*{Min}_{x \in \mathbb{R}^n} f(x) \text{ subject to } g(x, \omega_i) \leq 0, \quad i = 1, \ldots, m, \tag{3.1}$$

denoted by (P_m). Clearly the feasible set of problem (P) is included in the feasible set of (P_m), and hence $\operatorname{val}(P) \geq \operatorname{val}(P_m)$. Together with [10] we use the following terminology.

Definition 3.1 It is said that problem (P) is reducible if there exists a discretization (P_m) such that $\operatorname{val}(P) = \operatorname{val}(P_m)$, and it is said that problem (P) is discretizable if for any $\varepsilon > 0$ there exists a discretization (P_m) such that $\operatorname{val}(P_m) \geq \operatorname{val}(P) - \varepsilon$.

In other words, the problem (P) is discretizable if *there exists* a sequence (P_m) of finite discretizations such that $\operatorname{val}(P_m) \to \operatorname{val}(P)$. In [21] this property is called *weak discretizability*, in order to distinguish it from convergence $\operatorname{val}(P_m) \to \operatorname{val}(P)$ for *any* sequence of finite discretizations with the corresponding meshsize tending to zero. We will discuss this further in Section 6.

The Lagrangian dual of problem (3.1) is

$$\operatorname*{Max}_{\lambda \geq 0} \inf_{x \in \mathbb{R}^n} \mathcal{L}_m(x, \lambda), \tag{3.2}$$

where

$$\mathcal{L}_m(x, \lambda) := f(x) + \sum_{i=1}^{m} \lambda_i g(x, \omega_i), \quad (x, \lambda) \in \mathbb{R}^n \times \mathbb{R}^m, \tag{3.3}$$

is the Lagrangian of the discretized problem. Problem (3.2) will be denoted as (D_m). By the weak duality for the discretized problem we have that $\text{val}(P_m) \geq \text{val}(D_m)$. In both frameworks of $\mathcal{Y} = \mathbb{R}^\Omega$ and $\mathcal{Y} = C(\Omega)$, for $\lambda \in \mathcal{Y}^*$ with $\text{supp}(\lambda) = \{\omega_1, \ldots, \omega_m\}$, we have that $L(x, \lambda) = \mathcal{L}_m(x, \lambda)$ and hence $\text{val}(D) \geq \text{val}(D_m)$. Here with some abuse of notation we denote by the same λ, an m-dimensional vector formed from nonzero elements of $\lambda(\omega)$ in case of $\lambda \in (\mathbb{R}^\Omega)^*$, and $\lambda = \sum_{i=1}^m \lambda_i \delta(\omega_i)$ in case of $\lambda \in C(\Omega)^*$.

In the subsequent analysis of this section we use the following condition.

(A2) For any discretization such that $\text{val}(P_m)$ is finite it holds that $\text{val}(P_m) = \text{val}(D_m)$ and (D_m) has an optimal solution.

This condition holds in the following two important cases: when the SIP problem (1.1) is linear, and hence its discretization (P_m) is a linear programming problem, or when problem (1.1) is convex and the Slater condition is satisfied. Note that, of course, if the Slater condition holds for the problem (P), then it holds for any discretization (P_m). For linear SIP problems the following result is given in [10, Theorems 8.3 and 8.4].

Theorem 3.1 *In the setting of* $\mathcal{Y} = \mathbb{R}^\Omega$, *let* (D) *be the corresponding dual of* (P). *Then the following holds*: (i) *if* $\text{val}(P) = \text{val}(D)$, *then problem* (P) *is discretizable*, (ii) *if* $\text{val}(P) = \text{val}(D)$ *and the dual problem* (D) *has an optimal solution*, *then problem* (P) *is reducible*. *Moreover*, *if condition* (A2) *is fulfilled*, *then the converse of* (i) *and* (ii) *also holds*.

Proof Suppose that $\text{val}(P) = \text{val}(D)$. By the definition of the dual problem (D) we have that for any $\varepsilon > 0$ there exists $\bar{\lambda} \in \mathcal{Y}^*$ such that $\bar{\lambda} \succeq 0$ and that

$$\inf_{x \in X} L(x, \bar{\lambda}) \geq \text{val}(D) - \varepsilon. \tag{3.4}$$

In the considered case of $\mathcal{Y} = \mathbb{R}^\Omega$, we have that $L(x, \bar{\lambda})$ is the Lagrangian of the discretized problem (P_m) associated with the set $\{\omega_1, \ldots, \omega_m\} = \text{supp}(\bar{\lambda})$. Hence $\text{val}(P_m) \geq \inf_{x \in X} L(x, \bar{\lambda})$. It follows that

$$\text{val}(P_m) \geq \inf_{x \in X} L(x, \bar{\lambda}) \geq \text{val}(D) - \varepsilon = \text{val}(P) - \varepsilon, \tag{3.5}$$

and hence (P) is discretizable.

Now suppose that $\text{val}(P) = \text{val}(D)$ and problem (D) has an optimal solution $\bar{\lambda}$. Then (3.4) holds with $\varepsilon = 0$. Consequently (3.5) holds with $\varepsilon = 0$, which together with the inequality $\text{val}(P) \geq \text{val}(P_m)$ implies that $\text{val}(P) = \text{val}(P_m)$, and hence (P) is reducible.

Conversely, suppose that condition (A2) holds and problem (P) is discretizable. That is, for $\varepsilon > 0$ there exists a discretization (P_m) such that $\text{val}(P_m) \geq \text{val}(P) - \varepsilon$. Then

$$\text{val}(P_m) + \varepsilon \geq \text{val}(P) \geq \text{val}(D) \geq \text{val}(D_m) = \text{val}(P_m).$$

It follows that $|\text{val}(P) - \text{val}(D)| \leq \varepsilon$, and since $\varepsilon > 0$ is arbitrary, $\text{val}(P) = \text{val}(D)$.

In order to show the converse of (ii) observe that if $\text{val}(P) = \text{val}(P_m)$ and $\lambda \in \mathbb{R}^m$ is an optimal solution of the dual problem (D_m), then the corresponding $\lambda \in \mathcal{Y}^*$ is an optimal solution of problem (D). ■

The above theorem together with results of Section 2 give various sufficient/necessary conditions for discretizability and reducibility of problem (P). In particular, by Theorem 2.2(ii) we have the following Corollary.

Corollary 3.1 *Suppose that problem* (P) *is convex and the set* $\mathrm{Sol}(P)$ *is nonempty and bounded. Then problem* (P) *is discretizable.*

In order to verify the reducibility of problem (P), we need to ensure that the corresponding dual problem (D) has an optimal solution. For that we need to impose additional conditions of a topological type. Let us assume now that condition (A1) holds and use the space $\mathcal{Y}=C(\Omega)$. The dual $\mathcal{Y}^*$ of this space is the space of finite signed Borel measures on Ω. In particular we can consider measures of the form $\mu=\sum_{i=1}^m \lambda_i\delta(\omega_i)$, i.e. measures with finite support (which is a subset of $\{\omega_1,\dots,\omega_m\}$ consisting of points ω_i such that $\lambda_i\neq 0$). For such measure μ we have that

$$L(x,\mu)=f(x)+\int_\Omega g(x,\omega)\mathrm{d}\mu(w)=f(x)+\sum_{i=1}^m \lambda_i g(x,\omega_i)=\mathcal{L}_m(x,\lambda), \tag{3.6}$$

and hence it follows that $\mathrm{val}(D)\geq \mathrm{val}(D_m)$. By Theorem 2.3 we have that if problem (P) is convex and Slater condition holds, then the set $\mathrm{Sol}(D)$ is nonempty and bounded. Note that here the dual problem (D) is performed over Borel measures and *a priori* there is no guarantee that the set $\mathrm{Sol}(D)$ contains a measure with a finite support. In that respect we have the following, quite a nontrivial, result due to Levin [20] (see also [5]).

Theorem 3.2 *Suppose that problem* (P) *is convex, the set* $\mathrm{dom}\, f$ *is closed, assumption* (A1) *is fulfilled,* $\mathrm{val}(P)<+\infty$ *and the following condition holds*:

(A3) *For any points* $\omega'_1,\dots,\omega'_{n+1}\in\Omega$ *there exists a point* $\bar{x}\in\mathrm{dom} f$ *such that* $g(\bar{x},\omega'_i)<0$, $i=1,\dots,n+1$.

Then there exist points $\omega_1,\dots,\omega_m\in\Omega$, *with* $m\leq n$, *such that for the corresponding discretization* (P_m) *and its dual* (D_m) *the following holds*

$$\mathrm{val}(P)=\mathrm{val}(P_m)=\mathrm{val}(D_m)=\mathrm{val}(D). \tag{3.7}$$

Condition (A3) means that the Slater condition holds for any discretization (P_m) with $m\leq n+1$. This in turn implies that $\mathrm{Sol}(D_m)$ is nonempty and bounded, provided that $\mathrm{val}(P_m)$ is finite (Theorem 2.3). Of course, Slater condition for the problem (P) implies condition (A3). Under the assumptions of Theorem 2.3 we have that $\mathrm{val}(P)=\mathrm{val}(D)$. If, moreover, $\mathrm{val}(P)$ is finite, then there exists a discretization with $m\leq n$ points such that $\mathrm{Sol}(D_m)$ is nonempty and is a subset of $\mathrm{Sol}(D)$, and hence the dual problem (D) (in the sense of the space $\mathcal{Y}=\mathbb{R}^\Omega$ and its dual $\mathcal{Y}^*$) has an optimal solution.

By the definition, problem (P) is reducible if there exists discretization (P_m) such that $\mathrm{val}(P)=\mathrm{val}(P_m)$. If, moreover, (P) is convex, then there exists such discretization with the following bounds on m. Recall that Helly's theorem says that if A_i, $i\in I$, is a finite family of convex subsets of $\mathbb{R}^n$ such that the intersection of any $n+1$ sets of this family is nonempty, then $\cap_{i\in I}A_i$ is nonempty (use of Helly's theorem to derive such bounds for semi-infinite programs seemingly is going back to [20]).

Theorem 3.3 *Suppose that problem* (P) *is convex and reducible. Then there exists discretization* (P_m) *such that* $\mathrm{val}(P)=\mathrm{val}(P_m)$ *and*: (i) $m\leq n+1$ *if* $\mathrm{val}(P)=+\infty$, (ii) $m\leq n$ *if* $\mathrm{val}(P)<+\infty$.

Proof Let (P_k) be a discretization of (P) such that $\mathrm{val}(P_k)=\mathrm{val}(P)$ and $\{\omega_1,\dots,\omega_k\}$ be the corresponding discretization set. Consider sets $A_0:=\{x\in\mathbb{R}^n : f(x)<\mathrm{val}(P)\}$ and

$A_i := \{x \in \mathbb{R}^n : g(x, \omega_i) \le 0\}$, $i = 1, \ldots, k$. Since functions $f(\cdot)$ and $g(\cdot, \omega_i)$, $i = 1, \ldots, k$, are convex, these sets are convex.

Suppose that $\text{val}(P) = +\infty$. Note that in this case $A_0 = \text{dom} f$. Since $\text{val}(P_k) = +\infty$, we have that the set $\cap_{i=0}^k A_i$ is empty. By Helly's theorem it follows that there exists a subfamily of the family $A_0, \ldots, A_k$ with empty intersection and no more than $n+1$ members. Depending on whether this subfamily contains set A_0 or not, we have the required discretization (P_m) with $\text{val}(P_m) = +\infty$ and $m \le n$ or $m \le n+1$. This proves (i).

Now suppose that $\text{val}(P) < +\infty$. In order to prove (ii) we argue by a contradiction. Suppose that the assertion is false. Then the intersection of A_0 and any n sets of the family A_i, $1 \le i \le k$, is nonempty. Note that the intersection of all sets A_i, $1 \le i \le k$, is nonempty since otherwise the feasible set of problem (P_k) will be empty and consequently $\text{val}(P_k)$ will be $+\infty$. It follows that the intersection of any $n+1$ sets of the family A_i, $i \in \{0, 1, \ldots, k\}$ is nonempty. By Helly's theorem this implies that the intersection of all sets A_i, $i \in \{0, 1, \ldots, k\}$, is nonempty. Let $\bar{x}$ be a point of the set $\cap_{i=0}^k A_i$. Since $\bar{x} \in \cap_{i=1}^k A_i$, the point $\bar{x}$ is a feasible point of problem (P_k), and since $\bar{x} \in A_0$, we have that $f(\bar{x}) < \text{val}(P_k)$. This is a required contradiction. ■

4. First-order optimality conditions

It follows from the minimax representations (2.8) and (2.9) that if $\bar{x}$ is an optimal solution of problem (P) and $\bar{\lambda}$ is an optimal solution of its dual (D) and $\text{val}(P) = \text{val}(D)$, then $(\bar{x}, \bar{\lambda})$ is a saddle point of the Lagrangian $L(x, \lambda)$, i.e.

$$\bar{x} \in \arg\min_{x \in \mathbb{R}^n} L(x, \bar{\lambda}) \quad \text{and} \quad \bar{\lambda} \in \arg\max_{\lambda \in K^*} L(\bar{x}, \lambda). \tag{4.1}$$

Conversely, if $(\bar{x}, \bar{\lambda})$ is a saddle point of $L(x, \lambda)$, then $\bar{x}$ is an optimal solution of (P), $\bar{\lambda}$ is an optimal solution of (D) and $\text{val}(P) = \text{val}(D)$. The second condition in (4.1) means that $\bar{\lambda}$ is a maximizer of $\langle \lambda, G(\bar{x}) \rangle$ over $\lambda \in K^*$. If $G(\bar{x}) \notin K$, then $\sup_{\lambda \in K^*} \langle \lambda, G(\bar{x}) \rangle = +\infty$, and hence necessarily $G(\bar{x}) \in K$. Since $G(\bar{x}) \in K$, we have that $\langle \lambda, G(\bar{x}) \rangle \le 0$ and hence this maximum is attained at $\lambda = 0$ and is 0. That is, the second condition in (4.1) holds iff $G(\bar{x}) \in K$, $\lambda \in K^*$ and $\langle \bar{\lambda}, G(\bar{x}) \rangle = 0$. Note that for $G(\bar{x}) \in K$ and $\bar{\lambda} \in K^*$, condition $\langle \bar{\lambda}, G(\bar{x}) \rangle = 0$ is equivalent to the condition $\text{supp}(\bar{\lambda}) \subset \Delta(\bar{x})$, where

$$\Delta(\bar{x}) := \{\omega \in \Omega : g(\bar{x}, \omega) = 0\} \tag{4.2}$$

is the index set of active at x constraints. The above arguments hold for both frameworks of $\mathcal{Y} = \mathbb{R}^\Omega$ and $\mathcal{Y} = C(\Omega)$.

In the subsequent analysis we will use the following result due to Rogosinsky [26].

THEOREM 4.1 *Let Ω be a metric space equipped with its Borel sigma algebra $\mathcal{B}$, $q_i : \Omega \to \mathbb{R}$, $i = 1, \ldots, k$, be measurable functions, and μ be a (nonnegative) measure on $(\Omega, \mathcal{B})$ such that $q_1, \ldots, q_k$ are μ-integrable. Then there exists a (nonnegative) measure η on $(\Omega, \mathcal{B})$ with a finite support of at most k points such that $\int_\Omega q_i \, d\mu = \int_\Omega q_i \, d\eta$ for all $i = 1, \ldots, k$.*

4.1. *Convex case*

We assume in this subsection that problem (P) is convex. If $\bar{\lambda} \in K^*$, then $L(\cdot, \bar{\lambda})$ is convex, and hence the first condition in (4.1) holds iff 0 belongs to the subdifferential $\partial L(\bar{x}, \bar{\lambda})$

of $L(\cdot,\bar\lambda)$ at the point $\bar x$. Therefore, conditions (4.1) can be written in the following equivalent form[1]

$$0 \in \partial L(\bar x,\bar\lambda);\ \ g(\bar x,\omega) \le 0,\ \omega \in \Omega; \quad \bar\lambda \succeq 0;\ \ \mathrm{supp}(\bar\lambda) \subset \Delta(\bar x). \tag{4.3}$$

We can view conditions (4.3) as first-order optimality conditions for both frameworks of $\mathcal{Y} = \mathbb{R}^{\Omega}$ and $\mathcal{Y} = C(\Omega)$. We denote by $\Lambda(\bar x)$ the set of (Lagrange multipliers) $\bar\lambda \in \mathcal{Y}^*$ satisfying conditions (4.3).

For the linear SIP problem (1.2) and the framework of $\mathcal{Y} = C(\omega)$, we have that

$$\partial L(\bar x,\mu) = \{\nabla L(\bar x,\mu)\} = \{c + \textstyle\int_{\Omega} a(\omega)\mathrm{d}\mu(\omega)\}, \tag{4.4}$$

and hence the above conditions (4.3) take the form

$$a(\omega)^{\mathrm{T}}\bar x + b(\omega) \le 0, \quad \omega \in \Omega, \tag{4.5}$$

$$c + \int_{\Omega} a(\omega)\mathrm{d}\mu(\omega) = 0, \quad \mu \succeq 0, \tag{4.6}$$

$$\int_{\Omega} [a(\omega)^{\mathrm{T}}\bar x + b(\omega)]\mathrm{d}\mu(\omega) = 0. \tag{4.7}$$

In the framework of $\mathcal{Y} = \mathbb{R}^{\Omega}$ the integrals in (4.6) and (4.7) should be replaced by the respective sums. Conditions (4.5) and (4.6) represent feasibility conditions for the linear SIP (1.2) and its dual (2.18), respectively, and condition (4.7) is the complementarity condition.

As it was discussed above, conditions (4.3) mean that $(\bar x,\bar\lambda)$ is a saddle point of $L(x,\lambda)$. Therefore we have the following result.

THEOREM 4.2 *Suppose that problem* (P) *is convex and let* $\bar x \in \mathcal{Y}$ *and* $\bar\lambda \in \mathcal{Y}^*$ *be points satisfying conditions* (4.3) (*in either of frameworks of* $\mathcal{Y} = \mathbb{R}^{\Omega}$ *or* $\mathcal{Y} = C(\Omega)$). *Then* $\bar x$ *and* $\bar\lambda$ *are optimal solutions of problems* (P) *and* (D), *respectively, and* $\mathrm{val}(P) = \mathrm{val}(D)$.

As the above theorem states, sufficiency of conditions (4.3) does not require a constraint qualification. On the other hand, in order to ensure existence of Lagrange multipliers there is a need for additional conditions. By Theorem 2.3 we have the following result.

THEOREM 4.3 *Suppose that problem* (P) *is convex, assumption* (A1) *is fulfilled and let* $\bar x$ *be an optimal solution of* (P). *Then, in the framework of* $\mathcal{Y} = C(\Omega)$, *the set* $\Lambda(\bar x)$ *of Lagrange multipliers is nonempty and bounded iff Slater condition holds.*

The above theorem shows that the Slater condition ensures existence of Lagrange multipliers in a form of measures. Let us denote by $\Lambda_m(\bar x)$ the set of measures satisfying conditions (4.3) and with a finite support of at most m points. That is, $\mu \in \Lambda_m(\bar x)$ if $\mu = \sum_{i=1}^m \lambda_i \delta(\omega_i)$ and

$$0 \in \partial\left[f(\bar x) + \sum_{i=1}^m \lambda_i g(\bar x,\omega_i)\right];\ \ g(\bar x,\omega) \le 0,\ \omega \in \Omega; \quad \lambda \ge 0; \quad \{\omega_1,\dots,\omega_m\} \subset \Delta(\bar x). \tag{4.8}$$

Note that by the Moreau–Rockafellar theorem [24] we have that for $\lambda \ge 0$,

$$\partial\left[f(\bar x) + \sum_{i=1}^m \lambda_i g(\bar x,\omega_i)\right] = \partial f(\bar x) + \sum_{i=1}^m \lambda_i \partial g(\bar x,\omega_i). \tag{4.9}$$

The required regularity conditions for the above formula to hold are satisfied here since functions $g(\cdot, \omega_i)$ are continuous.

THEOREM 4.4 *Suppose that problem* (*P*) *is convex*, *assumption* (A1) *is fulfilled and let* $\bar{x}$ *be an optimal solution of* (*P*). *Then the set* $\Lambda_n(\bar{x})$ *is nonempty and bounded if Slater condition holds. Conversely, if* $\Lambda_{n+1}(\bar{x})$ *is nonempty and bounded, then Slater condition holds.*

Proof Assuming the Slater condition let us show that $\Lambda_n(\bar{x})$ is nonempty. Recall that under the Slater condition, the set $\Lambda(\bar{x})$ is nonempty and bounded. Since $\Lambda_n(\bar{x})$ is a subset of $\Lambda(\bar{x})$, it follows that $\Lambda_n(\bar{x})$ is bounded. Let $\mu \in C(\Omega)^*$ be a measure satisfying conditions (4.3), i.e. $\mu \in \Lambda(\bar{x})$. Consider the function $h(x) := \int_\Omega g(x, \omega)\mathrm{d}\mu(\omega)$. Since $g(\cdot, \omega)$, $\omega \in \Omega$, are convex real-valued functions and $\mu \succeq 0$, the function $h(\cdot)$ is convex, and since $g(x, \cdot)$ is continuous and Ω is compact, the function $g(x, \cdot)$ is bounded and hence $h(x)$ is real valued. Therefore, by the Moreau–Rockafellar theorem, $\partial L(\bar{x}, \mu) = \partial f(\bar{x}) + \partial h(\bar{x})$. By a theorem due to Strassen [33] we have that $\partial h(\bar{x}) = \int_\Omega \partial g(\bar{x}, \omega)\mathrm{d}\mu(\omega)$, i.e. $\partial h(\bar{x})$ consists of vectors of the form $\int_\Omega \gamma(\omega)\mathrm{d}\mu(\omega)$, for measurable selections $\gamma(\omega) \in \partial g(\bar{x}, \omega)$. Therefore, the first condition of (4.3) means that

$$q + \int_\Omega \gamma(\omega)\mathrm{d}\mu(\omega) = 0 \tag{4.10}$$

for some $q \in \partial f(\bar{x})$ and a certain measurable selection $\gamma(\omega) \in \partial g(\bar{x}, \omega)$.

Consider a measurable selection $\gamma(\omega) \in \partial g(\bar{x}, \omega)$ satisfying (4.10). By Theorem 4.1 we have that there exists a measure $\eta \succeq 0$ with a finite support $\{\omega_1, \ldots, \omega_m\} \subset \mathrm{supp}(\mu)$ such that $m \leq n$ and $\int_\Omega \gamma(\omega)\mathrm{d}\mu(\omega) = \int_\Omega \gamma(\omega)\mathrm{d}\eta(\omega)$. Since

$$\int_\Omega \gamma(\omega)\mathrm{d}\eta(\omega) \in \int_\Omega \partial g(\bar{x}, \omega)\mathrm{d}\eta(\omega) = \partial \int_\Omega g(\bar{x}, \omega)\mathrm{d}\eta(\omega),$$

it follows by (4.10) that $0 \in q + \partial \int_\Omega g(\bar{x}, \omega)\mathrm{d}\eta(\omega)$ and hence $\eta \in \Lambda_n(\bar{x})$. This shows that $\Lambda_n(\bar{x})$ is nonempty.

Conversely, suppose that $\Lambda_{n+1}(\bar{x})$ is nonempty and bounded. We need to show that then $\Lambda(\bar{x})$ is bounded. Indeed, let every element of $\Lambda_{n+1}(\bar{x})$ have a norm less than a constant $c > 0$. Arguing by a contradiction suppose that there is an element $\mu \in \Lambda(\bar{x})$ having the (total variation) norm $c' > c$. Since $\mu \succeq 0$, its total variation norm is equal to $\mu(\Omega)$. Consider the set $\{\mu' \in \Lambda(\bar{x}) : \mu'(\Omega) = c'\}$. This set is nonempty, since μ belongs to this set, and by Theorem 4.1 this set contains a measure η with a finite support of at most $n+1$ points (note that we added one more constraint $\int_\Omega \mathrm{d}\mu' = c'$ to this set). It follows that $\eta \in \Lambda_{n+1}(\bar{x})$ and has norm $c' > c$. This is a contradiction. ∎

4.2. *Smooth case*

In this section we discuss first-order optimality conditions for smooth (not necessarily convex) SIP problems. We make the following assumption in this subsection.

(A4) The set Ω is a compact metric space, the functions $g(\cdot, \omega)$, $\omega \in \Omega$, and $f(\cdot)$ are real valued continuously differentiable, and $\nabla g(\cdot, \cdot)$ is continuous on $\mathbb{R}^n \times \Omega$.

The above condition (A4) implies that the mapping $G : x \mapsto g(x, \cdot)$ is differentiable and its derivative $DG(x) : h \mapsto h^{\mathrm{T}}\nabla g(x, \cdot)$.

Let $\bar{x}$ be a locally optimal solution of the SIP problem (P). Linearization of optimality conditions (4.8) lead to the following conditions

$$\nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g(\bar{x}, \omega_i) = 0; \quad g(\bar{x}, \omega) \leq 0,\ \omega \in \Omega;\ \lambda \geq 0; \quad \{\omega_1, \ldots, \omega_m\} \subset \Delta(\bar{x}). \tag{4.11}$$

We denote by $\Lambda_m(\bar{x})$ the set of measures $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ satisfying conditions (4.11). If the problem (P) is convex, then since $L(\cdot, \mu)$ is differentiable we have that $\partial L(\bar{x}, \mu) = \{\nabla L(\bar{x}, \mu)\}$, and hence in that case conditions (4.11) coincide with conditions (4.8). There are several ways to show the existence of Lagrange multipliers satisfying conditions (4.11). We proceed as follows.

Consider functions

$$\mathfrak{g}(x) := \sup_{\omega \in \Omega} g(x, \omega) \quad \text{and} \quad \mathfrak{f}(x) := \max\{f(x) - f(\bar{x}), \mathfrak{g}(x)\}. \tag{4.12}$$

The SIP problem (1.1) can be written in the following equivalent form:

$$\operatorname*{Min}_{x \in \mathbb{R}^n} f(x) \text{ subject to } \mathfrak{g}(x) \leq 0. \tag{4.13}$$

By the assumption (A4), the set Ω is compact and the function $g(x, \cdot)$ is continuous, and hence the set

$$\Omega^*(x) := \arg\max_{\omega \in \Omega} g(x, \omega) \tag{4.14}$$

is nonempty and compact for any $x \in \mathbb{R}^n$. Since $\bar{x}$ is a feasible point of problem (P), it follows that $\mathfrak{g}(\bar{x}) \leq 0$, and $\mathfrak{g}(\bar{x}) = 0$ iff the index set $\Delta(\bar{x})$, defined in (4.2), of active at $\bar{x}$ constraints, is nonempty, in which case $\Delta(\bar{x}) = \Omega^*(x)$. By the Danskin theorem [8] the max-function $\mathfrak{g}(x)$ is directionally differentiable and its directional derivatives are given by

$$\mathfrak{g}'(x, h) = \sup_{\omega \in \Omega^*(x)} h^{\mathrm{T}} \nabla g(x, \omega). \tag{4.15}$$

Moreover, $\mathfrak{g}(x)$ is locally Lipschitz continuous and hence is directionally differentiable in the Hadamard sense.

By feasibility of $\bar{x}$ we have that $\mathfrak{g}(\bar{x}) \leq 0$ and hence $\mathfrak{f}(\bar{x}) = 0$. Moreover, it follows from local optimality of $\bar{x}$, that $\mathfrak{f}(x) \geq \mathfrak{f}(\bar{x})$ for all x in a neighbourhood of $\bar{x}$, i.e. $\bar{x}$ is a local minimizer of $\mathfrak{f}(x)$. Unless stated otherwise, we assume that the index set $\Delta(\bar{x})$, of active at $\bar{x}$ constraints, is *nonempty*, and hence $\mathfrak{g}(\bar{x}) = 0$.

Consider the set

$$\mathcal{A} := \{\nabla f(\bar{x})\} \cup \left\{\nabla g(\bar{x}, \omega),\ \omega \in \Delta(\bar{x})\right\}. \tag{4.16}$$

By (4.15) the function $\mathfrak{f}(x)$ is directionally differentiable at $x = \bar{x}$ and

$$\mathfrak{f}'(\bar{x}, \cdot) = \sigma_{\mathcal{A}}(\cdot) \tag{4.17}$$

(recall that $\sigma_{\mathcal{A}}(\cdot)$ denotes the support function of set $\mathcal{A}$). Since $\bar{x}$ is a local minimizer of $\mathfrak{f}(\cdot)$ it follows that $\mathfrak{f}'(\bar{x}, h) \geq 0$ for all $h \in \mathbb{R}^n$, which together with (4.16) imply that

$$0 \in \operatorname{conv}(\mathcal{A}). \tag{4.18}$$

Note that the set $\mathcal{A}$ is compact and therefore its convex hull is also compact and hence is closed.

Condition (4.18) means that there exist multipliers $\lambda_i \geq 0, i = 0, 1, \ldots, m$, not all of them zeros, and points $\omega_i \in \Delta(\bar{x})$, $i = 1, \ldots, m$, such that

$$\lambda_0 \nabla f(\bar{x}) + \sum_{i=1}^{m} \lambda_i \nabla g(\bar{x}, \omega_i) = 0. \tag{4.19}$$

The above condition (4.19) is Fritz John-type optimality condition. In order to ensure that the multiplier λ_0 in (4.19) is not zero we need a constraint qualification.

Definition 4.1 It is said that the extended Mangasarian–Fromovitz constraint qualification (MFCQ) holds at the point $\bar{x}$ if there exists $h \in \mathbb{R}^n$ such that

$$h^{\mathrm{T}} \nabla g(\bar{x}, \omega) < 0, \quad \forall \omega \in \Delta(\bar{x}). \tag{4.20}$$

This is a natural extension of the MFCQ used in non-linear programming when the index set Ω is finite. In the considered case (under condition (A4)), the extended MFCQ is equivalent to Robinson's constraint qualification (e.g. [17], [4, Example 2.102]), and in the convex case to Slater condition. We have the following result (cf. [18, Theorem 2.3], [4, Theorem 5.111]).

THEOREM 4.5 *Let $\bar{x}$ be a locally optimal solution of problem (P) such that the index set $\Delta(\bar{x})$ is nonempty. Suppose that condition (A4) is fulfilled and the MFCQ holds. Then the set $\Lambda_n(\bar{x})$ is nonempty and bounded. Conversely, if condition (A4) is fulfilled and the set $\Lambda_{n+1}(\bar{x})$ is nonempty and bounded, then the MFCQ holds.*

Proof Suppose that the MFCQ holds. Let $\lambda_i \geq 0$, $i = 0, 1, \ldots, m$, be multipliers and $\omega_i \in \Delta(\bar{x})$, $i = 1, \ldots, m$, be points satisfying conditions (4.17). By the above discussion, under condition (A4), such multipliers (not all of them zeros) always exist. We need to show that $\lambda_0 \neq 0$. Arguing by a contradiction suppose that $\lambda_0 = 0$. Let h be a vector satisfying condition (4.20). Then since $\lambda_0 = 0$, we have that $h^{\mathrm{T}}(\sum_{i=1}^{m} \lambda_i \nabla g(\bar{x}, \omega_i)) = 0$. On the other hand, because of (4.20) we have that $h^{\mathrm{T}}(\sum_{i=1}^{m} \lambda_i \nabla g(\bar{x}, \omega_i)) < 0$, which gives us a contradiction. This shows that, for some positive integer m, the set $\Lambda_m(\bar{x})$ is nonempty. To conclude that we can take $m \leq n$ observe that any extreme point of the set of vectors $\lambda \geq 0$ satisfying first equation in (4.11) (for fixed points ω_i) has at most n nonzero components.

Let us show that $\Lambda_m(\bar{x})$ is bounded (for any positive integer m). Since $\Delta(\bar{x})$ is compact, it follows by (4.20) that there exists $h \in \mathbb{R}^n$ and $\varepsilon > 0$ such that $h^{\mathrm{T}} \nabla g(\bar{x}, \omega) < -\varepsilon$ for all $\omega \in \Delta(\bar{x})$. Then by the first equation of (4.11) we have

$$h^{\mathrm{T}} \nabla f(\bar{x}) = -\sum_{i=1}^{m} \lambda_i h^{\mathrm{T}} \nabla g(\bar{x}, \omega_i) \geq \varepsilon \sum_{i=1}^{m} \lambda_i,$$

and hence $\sum_{i=1}^{m} \lambda_i$ is bounded by the constant $\varepsilon^{-1} h^{\mathrm{T}} \nabla f(\bar{x})$.

The converse assertion can be proved in a way similar to the proof of Theorem 4.5 (see [4, Theorem 5.111] for details). ■

Let us finally discuss first-order sufficient conditions. We will need the following useful result. It is difficult to give a correct reference for this result since it has been discovered and rediscovered by many authors. Recall that $\mathcal{F}$ denotes the feasible set of problem (*P*).

LEMMA 4.1 *Suppose that condition (A4) is fulfilled. Let $\bar{x}$ be a feasible point of problem (P) such that the index set $\Delta(\bar{x})$ is nonempty. Then $T_{\mathcal{F}}(\bar{x}) \subset \Gamma(\bar{x})$, where*

$$\Gamma(\bar{x}) := \left\{ h \in \mathbb{R}^n : h^{\mathrm{T}} \nabla g(\bar{x}, \omega) \leq 0, \ \omega \in \Delta(\bar{x}) \right\}. \tag{4.21}$$

If, moreover, the MFCQ holds at $\bar{x}$, then $T_{\mathcal{F}}(\bar{x}) = \Gamma(\bar{x})$.

Proof Let $h \in T_{\mathcal{F}}(\bar{x})$. Then there exist sequences $t_k \downarrow 0$ and $h_k \to h$ such that $\bar{x} + t_k h_k \in \mathcal{F}$, and hence $\mathfrak{g}(\bar{x} + t_k h_k) \leq 0$. Since $\mathfrak{g}(\cdot)$ is Hadamard directionally differentiable at $\bar{x}$ and $\mathfrak{g}(\bar{x}) = 0$, it follows that $\mathfrak{g}'(\bar{x}, h) \leq 0$. Together with (4.15) this implies that $h \in \Gamma(\bar{x})$. This shows that $T_{\mathcal{F}}(\bar{x}) \subset \Gamma(\bar{x})$.

In order to show the equality $T_{\mathcal{F}}(\bar{x}) = \Gamma(\bar{x})$ under MFCQ, we argue as follows. Note that $\mathcal{F} = G^{-1}(K)$. Also $T_K(\gamma) = \{\eta \in C(\Omega) : \eta(\omega) \leq 0,\ \omega \in \Delta(\bar{x})\}$, where $\gamma(\cdot) := g(\bar{x}, \cdot)$ (e.g. [4, Example 2.63]). In the considered framework the MFCQ is equivalent to Robinson's constraint qualification (e.g. [4, Example 2.102]), and hence $T_{\mathcal{F}}(\bar{x}) = DG(\bar{x})^{-1}[T_K(\gamma)]$ (e.g. [4, p. 66]). It remains to note that $DG(\bar{x}) : h \mapsto h^{\mathrm{T}} \nabla g(\bar{x}, \cdot)$. ■

Definition 4.2 For $p > 0$ it is said that the p-th order growth condition holds at a feasible point $\bar{x} \in \mathcal{F}$ if there exist constant $c > 0$ and a neighbourhood $\mathcal{V}$ of $\bar{x}$ such that

$$f(x) \geq f(\bar{x}) + c\|x - \bar{x}\|^p, \quad \forall x \in \mathcal{F} \cap \mathcal{V}. \tag{4.22}$$

In the literature, the first-order (i.e. for $p = 1$) growth condition at $\bar{x}$ is also referred to as $\bar{x}$ being a *strongly unique local solution* of (P) in [12], and *strict local minimizer of order* $p = 1$ in [21]. The second-order (i.e. for $p = 2$) growth condition is referred to as the *quadratic* growth condition.

THEOREM 4.6 *Suppose that condition* (A4) *is fulfilled. Let* $\bar{x}$ *be a feasible point of problem* (P) *such that the index set* $\Delta(\bar{x})$ *is nonempty. Then condition*

$$h^{\mathrm{T}} \nabla f(\bar{x}) > 0, \quad \forall h \in \Gamma(\bar{x}) \setminus \{0\} \tag{4.23}$$

is sufficient and, if the MFCQ holds at $\bar{x}$, *is necessary for the first-order growth condition to hold at* $\bar{x}$.

Proof Let us observe that the first-order growth condition holds at a point $\bar{x} \in \mathcal{F}$ iff

$$h^{\mathrm{T}} \nabla f(\bar{x}) > 0, \quad \forall h \in T_{\mathcal{F}}(\bar{x}) \setminus \{0\}. \tag{4.24}$$

Indeed, suppose that (4.22) holds (for $p = 1$) and let $h \in T_{\mathcal{F}}(\bar{x}) \setminus \{0\}$. Then there exist sequences $t_k \downarrow 0$ and $h_k \to h$ such that $\bar{x} + t_k h_k \in \mathcal{F}$. By (4.22) we have that $f(\bar{x} + t_k h_k) - f(\bar{x}) \geq c t_k \|h_k\|$ for all k large enough. Since $f(x)$ is continuously differentiable, we also have that

$$f(\bar{x} + t_k h_k) - f(\bar{x}) = t_k h^{\mathrm{T}} \nabla f(\bar{x}) + o(t_k). \tag{4.25}$$

It follows that $t_k h^{\mathrm{T}} \nabla f(\bar{x}) + o(t_k) \geq c t_k \|h_k\|$, which implies that $h^{\mathrm{T}} \nabla f(\bar{x}) \geq c\|h\|$.

To show the converse we argue by contradiction. Suppose that condition (4.22) does not hold. Then there exist sequences $c_k \downarrow 0$ and $\mathcal{F} \ni x_k \to \bar{x}$ such that

$$f(x_k) < f(\bar{x}) + c_k \|x_k - \bar{x}\|. \tag{4.26}$$

Consider $t_k := \|x_k - \bar{x}\|$ and $h_k := (x_k - \bar{x})/t_k$. Note that $\|h_k\| = 1$. By passing to a subsequence if necessary, we can assume that h_k converges to a vector h. It follows that $h \in T_{\mathcal{F}}(\bar{x})$ and $\|h\| = 1$, and hence $h \neq 0$. Moreover, by (4.25) we have $[f(\bar{x} + t_k h_k) - f(\bar{x})]/t_k \leq c_k$. Together with (4.25) this implies that $h^{\mathrm{T}} \nabla f(\bar{x}) \leq 0$, a contradiction with (4.24).

Now by Lemma 4.1 we have that $T_{\mathcal{F}}(\bar{x}) \subset \Gamma(\bar{x})$, and hence sufficiency of (4.23) follows. If, moreover, the MFCQ holds, then $T_{\mathcal{F}}(\bar{x}) = \Gamma(\bar{x})$, and hence the necessity of (4.22) follows. ■

By Farkas lemma, condition (4.23) is equivalent to

$$-\nabla f(\bar{x}) \in \text{int}[\text{conv}(\mathcal{A}')], \tag{4.27}$$

where $\mathcal{A}' := \{\nabla g(\bar{x}, \omega),\ \omega \in \Delta(\bar{x})\}$. Therefore, condition (4.27) is sufficient, and under the MFCQ is necessary, for the first-order growth to hold at $\bar{x}$. The sufficiency of conditions (4.23) and (4.27) is well known (see, e.g. [12, Lemma 3.4 and Theorem 3.6]).

Let us finally mention the following result about uniqueness of Lagrange multipliers, in the framework of $\mathcal{Y} = C(\Omega)$, [29] (see also [4, Theorem 5.114]). Note that if $\mu \in C(\Omega)^*$ is a unique Lagrange multipliers measure, then necessarily vectors $\nabla g(\bar{x}, \omega)$, $\omega \in \text{supp}(\mu)$, are linearly independent, and hence the support of μ has no more than n points.

Theorem 4.7 *Suppose that condition* (A4) *is fulfilled and let* $\mu = \sum_{i=1}^{m} \lambda_i \delta(\omega_i)$ *be a Lagrange multipliers measure satisfying the first-order necessary conditions* (4.11) *with* $\lambda_i > 0$, $i = 1, \dots, m$. *Then the set* $\Lambda(\bar{x}) = \{\mu\}$ *is a singleton* (*i.e.* μ *is unique*) *if and only if the following two conditions hold*: (i) *the gradient vectors* $\nabla g(\bar{x}, \omega_i)$, $i = 1, \dots, m$, *are linearly independent*, (ii) *for any neighbourhood* $\mathcal{W} \subset \Omega$ *of the set* $\{\omega_1, \dots, \omega_m\}$ *there exists* $h \in \mathbb{R}^n$ *such that*

$$\begin{aligned} &h^{\mathrm{T}} \nabla g(\bar{x}, \omega_i) = 0, \quad i = 1, \dots, m, \\ &h^{\mathrm{T}} \nabla g(\bar{x}, \omega) < 0, \quad \omega \in \Delta(\bar{x}) \setminus \mathcal{W}. \end{aligned} \tag{4.28}$$

If the set Ω is finite, then (4.28) is equivalent to

$$\begin{aligned} &h^{\mathrm{T}} \nabla g(\bar{x}, \omega_i) = 0, \quad i = 1, \dots, m, \\ &h^{\mathrm{T}} \nabla g(\bar{x}, \omega) < 0, \quad \omega \in \Delta(\bar{x}) \setminus \{\omega_1, \dots, \omega_m\}, \end{aligned} \tag{4.29}$$

and conditions (i)–(ii) of the above theorem become standard necessary and sufficient conditions for uniqueness of the Lagrange multipliers vector (cf. [19]). For SIP problems dependence of vector h on the neighbourhood $\mathcal{W}$ in (4.28) is essential (see [29, Example 3.1]).

5. Second-order optimality conditions

In this section, we discuss second-order necessary and/or sufficient optimality conditions for the SIP problem (P). We make the following assumption throughout this section and, unless stated otherwise, use the framework of the space $\mathcal{Y} = C(\Omega)$ and its dual space of measures.

(A5) The set Ω is a compact metric space, the functions $g(\cdot, \omega)$, $\omega \in \Omega$, and $f(\cdot)$ are real-valued twice continuously differentiable, and $\nabla^2 g(\cdot, \cdot)$ is continuous on $\mathbb{R}^n \times \Omega$.

The above condition (A5) implies that the mapping $G : x \mapsto (x, \cdot)$, from $\mathbb{R}^n$ into $C(\Omega)$, is twice continuously differentiable and its second-order derivative $D^2 G(x)(h, h) = h^{\mathrm{T}} \nabla^2 g(x, \cdot) h$. Also recall that $Df(x)h = h^{\mathrm{T}} \nabla f(x)$ and $D^{2f}(x)(h, h) = h^{\mathrm{T}} \nabla^2 f(x) h$.

We can write problem (P) as

$$\underset{x \in \mathcal{F}}{\text{Min}}\, f(x), \tag{5.1}$$

where $\mathcal{F} = G^{-1}(K)$ is the feasible set of problem (*P*). We will use the following concepts in this section. The set

$$T^2_{\mathcal{F}}(x,h) := \{z \in \mathbb{R}^n : \operatorname{dist}(x + th + \tfrac{1}{2}t^2 z, \mathcal{F}) = o(t^2),\ t \geq 0\} \tag{5.2}$$

is called the (inner) *second-order tangent set* to $\mathcal{F}$ at the point $x \in \mathcal{F}$ in the direction h. That is, the set $T^2_{\mathcal{F}}(x,h)$ is formed by vectors z such that $x + th + \frac{1}{2}t^2 z + r(t) \in \mathcal{F}$ for some $r(t) = o(t^2)$, $t \geq 0$. Note that this implies that $x + th + o(t) \in \mathcal{F}$, and hence $T^2_{\mathcal{F}}(x,h)$ can be nonempty only if $h \in T_{\mathcal{F}}(x)$. In a similar way are defined second-order tangent sets to the set $K \subset C(\Omega)$.

The upper and lower (parabolic) second-order directional derivatives of a (directionally differentiable) function $\phi : \mathbb{R}^n \to \mathbb{R}$ are defined as

$$\phi''_+(x; h, z) := \limsup_{t \downarrow 0} \frac{\phi(x + th + \frac{1}{2}t^2 z) - \phi(x) - t\phi'(x,h)}{\frac{1}{2}t^2}, \tag{5.3}$$

and

$$\phi''_-(x; h, z) := \liminf_{t \downarrow 0} \frac{\phi(x + th + \frac{1}{2}t^2 z) - \phi(x) - t\phi'(x,h)}{\frac{1}{2}t^2}, \tag{5.4}$$

respectively. Clearly $\phi''_+(x; h, z) \geq \phi''_-(x; h, z)$. If $\phi''_+(x; h, \cdot) = \phi''_-(x; h, \cdot)$, then it is said that ϕ is second-order directionally differentiable at x in direction h, and the corresponding second-order directional derivative is denoted $\phi''(x; h, z)$. If $\phi(\cdot)$ is twice continuously differentiable at x, then

$$\phi''(x; h, z) = z^{\mathrm{T}}\phi(x) + h^{\mathrm{T}}\nabla^2\phi(x)h. \tag{5.5}$$

5.1. *Second-order necessary conditions*

We assume in this subsection that $\bar{x} \in \mathcal{F}$ is a locally optimal solution of problem (*P*) and that the index set $\Delta(\bar{x})$, of active at $\bar{x}$ constraints, is nonempty. It follows from local optimality of $\bar{x}$ that

$$h^{\mathrm{T}}\nabla f(\bar{x}) \geq 0, \quad \forall h \in T_{\mathcal{F}}(\bar{x}). \tag{5.6}$$

Consider the set (cone)

$$C(\bar{x}) := \{h \in T_{\mathcal{F}}(\bar{x}) : h^{\mathrm{T}}\nabla f(\bar{x}) = 0\}. \tag{5.7}$$

The cone $C(\bar{x})$ represents those feasible directions along which the first-order approximation of $f(x)$ at $\bar{x}$ is zero, and is called the *critical cone*. Note that because of (5.6), we have that $C(\bar{x}) = \{0\}$ iff condition (4.24) holds, which in turn is a necessary and sufficient condition for first-order growth at $\bar{x}$ (see the proof of Theorem 4.7).

For some $h \in C(\bar{x})$ and $z \in T^2_{\mathcal{F}}(\bar{x}, h)$ consider the (parabolic) curve $x(t) := \bar{x} + th + \frac{1}{2}t^2 z$. By the definition of the second-order tangent set, we have that there exists $r(t) = o(t^2)$ such that $x(t) + r(t) \in \mathcal{F}$, $t \geq 0$. It follows by local optimality of $\bar{x}$ that $f(x(t) + r(t)) \geq f(\bar{x})$ for all $t \geq 0$ small enough. By the second-order Taylor expansion we have

$$f(x(t) + o(t^2)) = f(\bar{x}) + tDf(\bar{x})h + \tfrac{1}{2}t^2[Df(\bar{x})z + D^2 f(\bar{x})(h,h)] + o(t^2). \tag{5.8}$$

Since for $h \in C(\bar{x})$ the second term on the right-hand side of (5.8) vanishes, this implies the following second-order necessary condition:

$$Df(\bar{x})z + D^2 f(\bar{x})(h,h) \geq 0, \quad \forall h \in C(\bar{x}),\ \forall z \in T^2_{\mathcal{F}}(\bar{x}, h). \tag{5.9}$$

This condition can be written in the form:

$$\inf_{z \in T^2_{\mathcal{F}}(\bar{x},h)} \left\{ Df(\bar{x})z + D^2 f(\bar{x})(h,h) \right\} \geq 0, \quad \forall h \in C(\bar{x}). \tag{5.10}$$

The term

$$\inf_{z \in T^2_{\mathcal{F}}(\bar{x},h)} Df(\bar{x})z = -\sigma\big(-\nabla f(\bar{x}), T^2_{\mathcal{F}}(\bar{x}, h)\big) \tag{5.11}$$

corresponds to a curvature of the set $\mathcal{F}$ at $\bar{x}$. Of course, the second-order necessary condition (5.10) can be written in the following equivalent form

$$h^{\mathrm{T}} \nabla^2 f(\bar{x}) h - \sigma\big(-\nabla f(\bar{x}), T^2_{\mathcal{F}}(\bar{x}, h)\big) \geq 0, \quad \forall h \in C(\bar{x}). \tag{5.12}$$

We are going now to calculate this curvature term in a dual form.

Similar to (5.8), by the second-order Taylor expansion of $G(x)$ along the curve $x(t)$ we have

$$G\big(x(t) + o(t^2)\big) = G(\bar{x}) + tDG(\bar{x})h + \tfrac{1}{2}t^2\big[DG(\bar{x})z + D^2 G(\bar{x})(h,h)\big] + o(t^2). \tag{5.13}$$

If $z \in T^2_{\mathcal{F}}(\bar{x}, h)$, then $x(t) + o(t^2) \in \mathcal{F}$ and hence $G\ (x(t) + o(t^2)) \in K$ for $t \geq 0$ small enough, and thus

$$DG(\bar{x})z + D^2 G(\bar{x})(h,h) \in T^2_K(G(\bar{x}), DG(\bar{x})h). \tag{5.14}$$

It is possible to show that the converse implication follows under the MFCQ, and hence we have the following chain rule for the second-order tangent sets (e.g. [6])

$$T^2_{\mathcal{F}}(\bar{x}, h) = DG(\bar{x})^{-1}\big[T^2_K(G(\bar{x}), DG(\bar{x})h) - D^2 G(\bar{x})(h,h)\big]. \tag{5.15}$$

Consequently, assuming the MFCQ, we can write the minimization problem on the left-hand side of (5.10) in the form

$$\begin{aligned} \operatorname*{Min}_{z \in \mathbb{R}^n} \quad & Df(\bar{x})z + D^2 f(\bar{x})(h,h) \\ \text{s.t.} \quad & DG(\bar{x})z + D^2 G(\bar{x})(h,h) \in T^2_K(G(\bar{x}), DG(\bar{x})h). \end{aligned} \tag{5.16}$$

The second-order tangent set $T^2_K(G(\bar{x}), DG(\bar{x})h)$, to the cone $K \subset C(\Omega)$ at the point $\gamma = G(\bar{x}) \in K$ in the direction $\eta = DG(\bar{x})h$, is computed in [14,15] (see also [7] and [4, pp. 387–400] for a further discussion). That is,

$$T^2_K(G(\bar{x}), DG(\bar{x})h) = \big\{\alpha \in C(\Omega) : \alpha(\omega) \leq \tau_{\bar{x},h}(\omega),\ \omega \in \Omega\big\}, \tag{5.17}$$

where $\tau_{\bar{x},h} : \Omega \to \overline{\mathbb{R}}$ is a lower semicontinuous extended real-valued function, given by

$$\tau_{\bar{x},h}(\omega) := \begin{cases} 0, & \text{if } \omega \in \operatorname{int}(\Delta(\bar{x})) \text{ and } \eta(\omega) = 0, \\ \liminf\limits_{\substack{\omega' \to \omega \\ \gamma(\omega') < 0}} \dfrac{\big([\eta(\omega')]_+\big)^2}{2\gamma(\omega')}, & \text{if } \omega \in \operatorname{bdr}(\Delta(\bar{x})) \text{ and } \eta(\omega) = 0, \\ +\infty, & \text{otherwise,} \end{cases} \tag{5.18}$$

and $\mathrm{bdr}(\Delta(\bar{x})) = \Delta(\bar{x}) \setminus \mathrm{int}(\Delta(\bar{x}))$. Note that since $g(\bar{x}, \cdot)$ is continuous, the index set $\Delta(\bar{x})$, of active at $\bar{x}$ constraints, is closed and hence is compact. If $\Delta(\bar{x})$ is empty, then $\tau_{\bar{x},h}(\cdot) \equiv +\infty$ and $T_K^2(G(\bar{x}), DG(\bar{x})h) = C(\Omega)$. This is not surprising since in that case γ is an interior point of the set K.

It follows that problem (5.16) is a linear SIP problem:

$$\begin{aligned} \operatorname*{Min}_{z \in \mathbb{R}^n} \quad & z^{\mathrm{T}} \nabla f(\bar{x}) + h^{\mathrm{T}} \nabla^2 f(\bar{x}) h \\ \text{s.t.} \quad & z^{\mathrm{T}} g(\bar{x}, \omega) + h^{\mathrm{T}} \nabla^2 g(\bar{x}, \omega) h \leq \tau_{\bar{x},h}(\omega), \quad \omega \in \Omega. \end{aligned} \tag{5.19}$$

Its dual is the problem

$$\operatorname*{Max}_{\lambda \in \Lambda(\bar{x})} \left\{ h^{\mathrm{T}} \nabla_{xx}^2 L(\bar{x}, \lambda) h - \sigma(\lambda, \mathcal{T}^2(h)) \right\}, \tag{5.20}$$

where $\mathcal{T}^2(h) := T_K^2(G(\bar{x}), DG(\bar{x})h)$. In the present case, because of (5.17), for any $\lambda \in \Lambda(\bar{x})$ we have that

$$\sigma(\lambda, \mathcal{T}^2(h)) = \sup_{\alpha \in C(\Omega)} \left\{ \int_\Omega \alpha(\omega) \mathrm{d}\lambda(\omega) : \alpha(\omega) \leq \tau_{\bar{x},h}(\omega),\ \omega \in \Omega \right\}. \tag{5.21}$$

It can happen that the function $\tau_{\bar{x},h}(\cdot)$ is unbounded from below. Since $\tau_{\bar{x},h}(\cdot)$ is lower semicontinuous and Ω is compact, this happens iff there exists $\omega \in \Omega$ such that $\tau_{\bar{x},h}(\omega) = -\infty$. In that case there is no $\alpha \in C(\Omega)$ such that $\alpha(\cdot) \leq \tau_{\bar{x},h}(\cdot)$, and hence the second-order tangent set $\mathcal{T}^2(h)$ is empty and $\sigma(\cdot, \mathcal{T}^2(h)) \equiv -\infty$. If $\tau_{\bar{x},h}(\omega) > -\infty$ for all $\omega \in \Omega$, then $\tau_{\bar{x},h}(\cdot)$ is uniformly bounded from below and the set $\mathcal{T}^2(h)$ is nonempty. In that case, since $\tau_{\bar{x},h}(\cdot)$ is lower semicontinuous, it follows that

$$\sigma(\lambda, \mathcal{T}^2(h)) = \int_\Omega \tau_{\bar{x},h}(\omega) \mathrm{d}\lambda(\omega). \tag{5.22}$$

Recall that the support of $\lambda \in \Lambda(\bar{x})$ is a subset of $\Delta(\bar{x})$. Therefore, if $\Delta(\bar{x}) = \{\omega_1, \ldots, \omega_m\}$ is finite and $\tau_{\bar{x},h}(\omega_i) > -\infty$, $i = 1, \ldots, m$, then

$$\sigma(\lambda, \mathcal{T}^2(h)) = \sum_{i=1}^m \lambda_i \tau_{\bar{x},h}(\omega_i), \tag{5.23}$$

with $\lambda_i = \lambda(\omega_i)$.

Under the MFCQ, there is no duality gap between problems (5.19) and (5.20), and by Lemma 4.1 the critical cone can be written as

$$C(\bar{x}) = \left\{ h \in \mathbb{R}^n : h^{\mathrm{T}} \nabla g(\bar{x}, \omega) \leq 0,\ \omega \in \Delta(\bar{x}), \quad h^{\mathrm{T}} \nabla f(\bar{x}) = 0 \right\}, \tag{5.24}$$

or equivalently as

$$C(\bar{x}) = \left\{ h \in \mathbb{R}^n : \begin{array}{ll} h^{\mathrm{T}} \nabla g(\bar{x}, \omega) = 0, & \omega \in \mathrm{supp}(\lambda), \\ h^{\mathrm{T}} \nabla g(\bar{x}, \omega) \leq 0, & \omega \in \Delta(\bar{x}) \setminus \mathrm{supp}(\lambda) \end{array} \right\} \tag{5.25}$$

for any $\lambda \in \Lambda(\bar{x})$. This leads to the following second-order necessary conditions.

THEOREM 5.1 *Let $\bar{x}$ be a locally optimal solution of problem (P) such that the index set $\Delta(\bar{x})$ is nonempty. Suppose that condition* (A5) *and the MFCQ are fulfilled. Then the following second-order necessary conditions hold*

$$\sup_{\lambda\in\Lambda(\bar{x})}\left\{h^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},\lambda)h-\sigma(\lambda,\mathcal{T}^2(h))\right\}\geq 0,\quad \forall h\in C(\bar{x}). \tag{5.26}$$

The term $\sigma(\lambda,\mathcal{T}^2(h))$ is referred to as the *sigma* or *curvature* term. For any $\lambda\in\Lambda(\bar{x})$ and $h\in C(\bar{x})$ we have by (5.25) that $h^{\mathrm{T}}\nabla g(\bar{x},\omega)=0$ for all $\omega\in\mathrm{supp}(\lambda)$, and hence by (5.18) that $\int_\Omega \tau_{\bar{x},h}\mathrm{d}\lambda\leq 0$, which in turn implies that $\sigma(\lambda,\mathcal{T}^2(h))\leq 0$. That is, the sigma (curvature) term is always nonpositive. Therefore, conditions (5.26) are implied by the 'standard' second-order necessary conditions where this term is omitted. If the index set Ω is finite, and hence problem (*P*) becomes a nonlinear programming problem, the sigma term vanishes. In a sense the sigma term measures the curvature of K at the point $G(\bar{x})\in K$.

For nonlinear programming problems second-order necessary conditions in the form (5.26), without the sigma term, are well known (cf. [1,13]). Existence of an additional term in second-order optimality conditions for SIP problems was known for a long time. Usually it was derived by the so-called reduction method under quite restrictive assumptions (e.g. [12, Section 5]). Second-order parabolic directional derivatives were used in [1,2] and a prototype of the sigma term was given in [2, Theorem 2.1], although Hessian of the Lagrangian and the curvature (sigma) term are not clearly distinguished there. In an abstract form the sigma term was introduced and calculated by Kawasaki [14,16]. In the dual form considered here this term was derived, under Robinson constraint qualification, by Cominetti [6]. For a detailed development of that theory we may refer to [4, Section 3.2].

Unfortunately it may not be easy to use representation (5.22) in order to compute the sigma term. Moreover, we would like to have second-order sufficient conditions in the form (5.26) merely by replacing the inequality sign '≥ 0' by the strict inequality '>0'. In that case, we say that there is *no gap* between the corresponding second-order necessary and sufficient optimality conditions. We derived second-order necessary conditions by verifying (local) optimality along parabolic curves. There is no reason *a priori* that in this way we can ensure local optimality of the considered point $\bar{x}$ and hence to derive respective second-order sufficient conditions. In order to deal with this we proceed as follows.

By Lemma 4.1 and formulas (4.15) and (5.24) we have, under the MFCQ, that

$$C(\bar{x})=\left\{h:\mathfrak{g}'(\bar{x},h)\leq 0,\ h^{\mathrm{T}}\nabla f(\bar{x})=0\right\}. \tag{5.27}$$

Consider the max-function

$$\mathfrak{v}(\alpha):=\sup_{\omega\in\Omega}\alpha(\omega),\quad \alpha\in C(\Omega). \tag{5.28}$$

Note that since the set Ω is compact, any function $\alpha\in C(\Omega)$ attains its maximum over Ω and hence indeed the function $\mathfrak{v}:C(\Omega)\to\mathbb{R}$ is real valued. It is also straightforward to verify that the function $\mathfrak{v}(\cdot)$ is convex and Lipschitz continuous with Lipschitz constant one. Clearly, the cone $K\subset C(\Omega)$ can be written as $K=\{\alpha\in C(\Omega):\mathfrak{v}(\alpha)\leq 0\}$ and the max-function $\mathfrak{g}(x)$, defined in (4.12), can be written as $\mathfrak{g}(x)=\mathfrak{v}(G(x))$.

Consider functions $\gamma(\cdot):=g(\bar{x},\cdot)$, $\eta(\cdot):=h^{\mathrm{T}}\nabla g(\bar{x},\cdot)$ and $\zeta(\cdot):=z^{\mathrm{T}}\nabla^2 g(\bar{x},\cdot)z$ for some vectors h, $z\in\mathbb{R}^n$. Since the point $\bar{x}$ is feasible, we have that $\gamma\in K$, and since $\Delta(\bar{x})$

is nonempty, it follows that $\mathfrak{v}(\gamma) = 0$ and $\Delta(\bar{x}) = \arg\max_{\omega\in\Omega} \gamma(\omega)$. By the Danskin theorem we have that $\mathfrak{v}(\cdot)$ is directionally differentiable and

$$\mathfrak{v}'(\gamma, \eta) = \sup_{\omega\in\Delta(\bar{x})} \eta(\omega). \tag{5.29}$$

Since $\mathfrak{g}(x) = \mathfrak{v}(G(x))$ and by (5.13) we have

$$\mathfrak{g}\big(\bar{x} + th + \tfrac{1}{2}t^2 z\big) \quad = \mathfrak{v}\big(G(\bar{x}) + tDG(\bar{x})h + \tfrac{1}{2}t^2\big[DG(\bar{x})z + D^2G(\bar{x})(h,h)\big] + o(t^2)\big).$$

Since $\mathfrak{g}'(\bar{x}, h) = \mathfrak{v}'(\gamma, \eta)$ and because of Lipschitz continuity of $\mathfrak{v}(\cdot)$, we obtain the following chain rule for second-order directional derivatives (cf. [4, Proposition 2.53])

$$\mathfrak{g}''_+(\bar{x}; h, z) = \mathfrak{v}''_+\big(G(\bar{x}); DG(\bar{x})h, DG(\bar{x})z + D^2G(\bar{x})(h,h)\big). \tag{5.30}$$

Also, under the MFCQ, the chain rule (5.15) for the second-order tangent sets holds. Moreover, for γ, $\eta \in C(\Omega)$ such that $\mathfrak{v}(\gamma) = 0$ we have that (cf. [4, Proposition 3.30])

$$T^2_K(\gamma, \eta) = \big\{\zeta \in C(\Omega) : \mathfrak{v}''_+(\gamma; \eta, \zeta) \le 0\big\}, \quad \text{if } \mathfrak{v}'(\gamma, \eta) = 0, \tag{5.31}$$

and $T^2_K(\gamma, \eta) = C(\Omega)$, if $\mathfrak{v}'(\gamma, \eta) < 0$. Putting it all together we obtain the following.

Lemma 5.1 *Suppose that condition* (A5) *and the MFCQ are fulfilled, and consider points $\bar{x}, h \in \mathbb{R}^n$ such that $\mathfrak{g}(\bar{x}) = 0$. Then*

$$T^2_{\mathcal{F}}(\bar{x}, h) = \big\{z : \mathfrak{g}''_+(\bar{x}; h, z) \le 0\big\}, \quad \textit{if } \mathfrak{g}'(\bar{x}, h) = 0, \tag{5.32}$$

and $T^2_{\mathcal{F}}(\bar{x}, h) = \mathbb{R}^n$ if $\mathfrak{g}'(\bar{x}, h) < 0$.

Note that $f'(\bar{x}, h) = h^{\mathrm{T}}\nabla f(\bar{x})$ and $f''(\bar{x}; h, z) = z^{\mathrm{T}} f(\bar{x}) + h^{\mathrm{T}}\nabla^2 f(\bar{x})h$, and if $f'(\bar{x}, h) = 0$, then

$$\mathfrak{f}''_+(\bar{x}; h, z) = \begin{cases} \max\big\{f''(\bar{x}; h, z),\ \mathfrak{g}''_+(\bar{x}; h, z)\big\}, & \text{if } \mathfrak{g}'_+(\bar{x}, h) = 0, \\ f''(\bar{x}; h, z), & \text{if } \mathfrak{g}'_+(\bar{x}, h) < 0. \end{cases}$$

Because of that and by Lemma 5.1, conditions (5.10) imply that

$$\inf_{z\in\mathbb{R}^n} \mathfrak{f}''_+(\bar{x}; h, z) \ge 0, \quad \forall h \in C(\bar{x}). \tag{5.33}$$

Recall that, under the MFCQ, conditions (5.10) and (5.26) are equivalent. It is straightforward to derive necessity of conditions (5.33) from the fact that the local optimality of $\bar{x}$ implies that $\bar{x}$ is a local minimizer of $\mathfrak{f}(x)$ and that if $h \in C(\bar{x})$, then $\mathfrak{f}'(\bar{x}, h) = 0$. In such derivations there is no need to assume the MFCQ and it is possible to replace the upper second-order directional derivative of $\mathfrak{f}$ in (5.33) by the respective lower second-order directional derivative. On the other hand, conditions (5.33) are weaker than conditions (5.10) in the sense that they allow for situations when $\mathfrak{g}'_+(\bar{x}, h) = 0$ and $\mathfrak{g}''_+(\bar{x}; h, z) = 0$, while $f''(\bar{x}; h, z) < 0$.

In order to apply second-order necessary conditions (5.33) we need to calculate second-order directional derivatives of the max-function $\mathfrak{g}(\cdot)$. A relatively simple case is discussed in the following example.

Example 5.1 Assume that: (i) $\Omega \subset \mathbb{R}^\ell$, (ii) the set $\Delta(\bar{x}) = \{\omega_1, \ldots, \omega_m\}$ is nonempty and finite, (iii) each point ω_i, $i = 1, \ldots, m$, is an interior point of the set Ω, (iv) $g(x, \omega)$ is twice continuously differentiable on $\mathbb{R}^n \times \mathbb{R}^\ell$, (v) Hessian matrices $\nabla^2_{\omega\omega} g(\bar{x}, \omega_i)$, $i = 1, \ldots, m$, are nonsingular. Then locally, for x in a neighbourhood of $\bar{x}$, the max-function $\mathfrak{g}(x)$ can be represented as $\mathfrak{g}(x) = \max\{\psi_1(x), \dots, \psi_m(x)\}$, where $\psi_i(\cdot)$ are twice continuously differentiable in a neighbourhood of $\bar{x}$ functions with

$$\nabla\psi_i(\bar{x}) = \nabla g(\bar{x}, \omega_i), \tag{5.34}$$

$$\nabla^2\psi_i(\bar{x}) = \nabla^2_{xx} g(\bar{x}, \omega_i) - \left[\nabla^2_{x\omega} g(\bar{x}, \omega_i)\right]\left[\nabla^2_{\omega\omega} g(\bar{x}, \omega_i)\right]^{-1}\left[\nabla^2_{\omega x} g(\bar{x}, \omega_i)\right], \tag{5.35}$$

$i = 1, \dots, m$. The above is not difficult to show by employing the Implicit Function Theorem and basically is the reduction approach used in SIP (see, e.g. [12, Section 4]). In that case for such h that $\mathfrak{g}'(\bar{x}, h) = 0$ we have

$$\mathfrak{g}''(\bar{x}; h, z) = \max_{i \in I(\bar{x}, h)} \left\{z^{\mathrm{T}}\nabla g(\bar{x}, \omega_i) + h^{\mathrm{T}}\left[\nabla^2\psi_i(\bar{x})\right]h\right\} \tag{5.36}$$

where $I(\bar{x}, h) := \{i : h^{\mathrm{T}}\nabla g(\bar{x}, \omega_i) = 0,\ i = 1, \dots, m\}$, and assuming the MFCQ,

$$T^2_{\mathcal{F}}(\bar{x}, h) = \left\{z : z^{\mathrm{T}}\nabla g(\bar{x}, \omega_i) + h^{\mathrm{T}}\left[\nabla^2\psi_i(\bar{x})\right]h \le 0 : i \in I(\bar{x}, h)\right\}. \tag{5.37}$$

Moreover, assuming the MFCQ, the corresponding second-order necessary conditions (5.26) can be written as

$$\sup_{\lambda \in \Lambda(\bar{x})} \left\{h^{\mathrm{T}}\nabla^2 L(\bar{x}, \lambda)h - \sum_{i=1}^{m} \lambda_i h^{\mathrm{T}} H_i h\right\} \ge 0, \quad \forall h \in C(\bar{x}). \tag{5.38}$$

where

$$H_i := \left[\nabla^2_{x\omega} g(\bar{x}, \omega_i)\right]\left[\nabla^2_{\omega\omega} g(\bar{x}, \omega_i)\right]^{-1}\left[\nabla^2_{\omega x} g(\bar{x}, \omega_i)\right]. \tag{5.39}$$

The sigma term here is $\sigma(\lambda, \mathcal{T}^2(h)) = \sum_{i=1}^{m} \lambda_i h^{\mathrm{T}} H_i h$. Note that in the considered case the Hessian matrices $\nabla^2_{\omega\omega} g(\bar{x}, \omega_i)$ are negative definite and hence this sigma term is less than or equal to zero, as it should be.

It is possible to derive second-order directional derivatives of the max-function $\mathfrak{g}(\cdot)$ in more involved cases and hence to write the corresponding second-order necessary conditions. We will discuss this further in the next subsection.

5.2. *Second-order sufficient conditions*

In this section we assume that $\bar{x} \in \mathcal{F}$ is a feasible point of problem (P) satisfying the first-order necessary conditions (5.6). Consider the following condition

$$\inf_{z \in T^2_{\mathcal{F}}(\bar{x}, h)} \left\{Df(\bar{x})z + D^2 f(\bar{x})(h, h)\right\} > 0, \quad \forall h \in C(\bar{x}) \setminus \{0\}. \tag{5.40}$$

This condition is obtained from the second-order necessary condition (5.10) by replacing '≥ 0' sign with the strict inequality sign '> 0'. Necessity of (5.10) was obtained by verifying optimality of $\bar{x}$ along parabolic curves. There is no reason *a priori* that verification of (local) optimality along parabolic curves is sufficient to ensure local optimality of $\bar{x}$. Therefore, in order to ensure sufficiency of (5.40) we need an additional condition. The following concept of second-order regularity was introduced in [3] and developed further in [4].

Definition 5.1 It is said that the set $\mathcal{F}$ is second-order regular at $\bar{x} \in \mathcal{F}$ if for any sequence $x_k \in \mathcal{F}$ of the form $x_k = \bar{x} + t_k h + \frac{1}{2}t_k^2 r_k$, where $t_k \downarrow 0$ and $t_k r_k \to 0$, it follows that

$$\lim_{k\to\infty} \operatorname{dist}\big(r_k, T^2_{\mathcal{F}}(\bar{x}, h)\big) = 0. \tag{5.41}$$

Note that in the above definition the term $\frac{1}{2}t_k^2 r_k = o(t_k)$, and hence such a sequence $x_k \in \mathcal{F}$ can exist only if $h \in T_{\mathcal{F}}(\bar{x})$.

It turns out that second-order regularity can be verified in many interesting cases and ensures sufficiency of conditions (5.40) (cf. [3], [4, Section 3.3.3]). Proof of the following result is relatively easy, so we give it for the sake of completeness.

THEOREM 5.2 *Let $\bar{x} \in \mathcal{F}$ be a feasible point of problem* (*P*) *satisfying first-order necessary conditions* (5.6). *Suppose that $\mathcal{F}$ is second-order regular at $\bar{x}$. Then the second-order conditions* (5.40) *are necessary and sufficient for the quadratic growth at $\bar{x}$ to hold.*

Proof Suppose that conditions (5.40) hold. In order to verify the quadratic growth condition we argue by a contradiction, so suppose that it does not hold. Then there exists a sequence $x_k \in \mathcal{F} \setminus \{\bar{x}\}$ converging to $\bar{x}$ and a sequence $c_k \downarrow 0$ such that

$$f(x_k) - f(\bar{x}) \le c_k \|x_k - \bar{x}\|^2. \tag{5.42}$$

Denote $t_k := \|x_k - \bar{x}\|$ and $h_k := t_k^{-1}(x_k - \bar{x})$. By passing to a subsequence if necessary we can assume that h_k converges to a vector h. Clearly $h \neq 0$, and by the definition of $T_{\mathcal{F}}(\bar{x})$ it follows that $h \in T_{\mathcal{F}}(\bar{x})$. Moreover, by (5.42) we have

$$c_k t_k^2 \ge f(x_k) - f(\bar{x}) = t_k Df(\bar{x})h + o(t_k)$$

and hence $Df(\bar{x})h \le 0$. Because of the first-order necessary conditions it follows that $Df(\bar{x})h = 0$, and hence $h \in C(\bar{x})$.

Denote $r_k := 2t_k^{-1}(h_k - h)$. We have that $x_k = \bar{x} + t_k h + \frac{1}{2}t_k^2 r_k \in \mathcal{F}$ and $t_k r_k \to 0$. Consequently, it follows by the second-order regularity that there exists a sequence $z_k \in T^2_{\mathcal{F}}(\bar{x}, h)$ such that $r_k - z_k \to 0$. Since $Df(\bar{x})h = 0$, by the second-order Taylor expansion we have

$$f(x_k) = f(\bar{x} + t_k h + \tfrac{1}{2}t_k^2 r_k) = f(\bar{x}) + \tfrac{1}{2}t_k^2[Df(\bar{x})z_k + D^2 f(\bar{x})(h,h)] + o(t_k^2).$$

Moreover, since $z_k \in T^2_{\mathcal{F}}(\bar{x}, h)$ we have that

$$Df(\bar{x})z_k + D^2 f(\bar{x})(h,h) \ge c,$$

where c is equal to the left-hand side of (5.40), which by the assumption is positive. It follows that

$$f(x_k) \ge f(\bar{x}) + \tfrac{1}{2}c\|x_k - \bar{x}\|^2 + o(\|x_k - \bar{x}\|^2),$$

a contradiction with (5.42).

Conversely, suppose that the quadratic growth condition (4.22) (with $p = 2$) holds at $\bar{x}$. It follows that the function $\phi(x) := f(x) - \frac{1}{2}c\|x - \bar{x}\|^2$ also attains its local minimum over $\mathcal{F}$ at $\bar{x}$. Note that $\nabla\phi(\bar{x}) = \nabla f(\bar{x})$ and $h^{\mathrm{T}}\nabla^2\phi(\bar{x})h = h^{\mathrm{T}}\nabla^2 f(\bar{x})h - c\|h\|^2$. Therefore, by the second-order necessary conditions (5.10), applied to the function ϕ, it follows that the left-hand side of (5.40) is greater than or equal to $c\|h\|^2$. This completes the proof. ■

Now consider the following counterpart of second-order necessary conditions (5.26):

$$\sup_{\lambda\in\Lambda(\bar{x})}\left\{h^{\mathrm{T}}\nabla^2_{xx}L(\bar{x},\lambda)h-\sigma(\lambda,\mathcal{T}^2(h))\right\}>0,\quad \forall h\in C(\bar{x})\setminus\{0\}. \tag{5.43}$$

As it was argued in Section 5.1, the left-hand sides of the second-order necessary conditions (5.10) and (5.26) do coincide and the critical cone can be written in the form (5.24), provided that the MFCQ holds. Therefore, under the MFCQ, conditions (5.40) and (5.43) are equivalent. It is interesting to observe that even without the MFCQ, conditions (5.43) imply conditions (5.40) and hence are sufficient for local optimality of $\bar{x}$.

LEMMA 5.2 *Let $\bar{x}\in\mathcal{F}$ be a feasible point of problem* (P) *satisfying first-order necessary conditions* (5.6) *and such that the set* $\Delta(\bar{x})$ *is nonempty. Then conditions* (5.43) *imply conditions* (5.40). *If, moreover, the MFCQ holds, then conditions* (5.40) *and* (5.43) *are equivalent.*

Proof Recall that for the inclusion (5.14) to hold there is no need for the MFCQ. Therefore the feasible set of problem (5.16) includes the set $T_{\mathcal{F}}(\bar{x},h)$, and hence the optimal value of (5.16) is less than or equal to the optimal value of (5.10). Moreover, the optimal value of (5.16) is always greater than or equal to the optimal value of its dual problem (5.20). That is, the optimal value of the left-hand side of (5.26) is always greater than or equal to the optimal value of the left-hand side of (5.26). Also, the set in the right-hand side of (5.24) always includes the critical cone $C(\bar{x})$. This completes the arguments that (5.43) implies (5.40). Assuming the MFCQ, the equivalence of (5.40) and (5.43) was discussed in Section 5.1. ■

It follows that, under the assumption of second-order regularity, conditions (5.43) are sufficient for local optimality of $\bar{x}$. Without the MFCQ it can happen that the set $\Lambda(\bar{x})$ of Lagrange multipliers is empty. In that case, the left-hand side of (5.43) is $-\infty$ and hence conditions (5.43) cannot hold. Therefore, conditions (5.43) are applicable only if $\Lambda(\bar{x})$ is nonempty.

It is also possible to approach derivations of second-order sufficient conditions by employing the max-functions $\mathfrak{g}$ and $\mathfrak{f}$. Observe that if $\mathfrak{f}(\bar{x})=0$, i.e. $\bar{x}\in\mathcal{F}$, and there exist constant $c>0$ and a neighbourhood $\mathcal{V}$ of $\bar{x}$ such that

$$\mathfrak{f}(x)\geq c\|x-\bar{x}\|^2,\quad \forall x\in\mathcal{V}, \tag{5.44}$$

then the quadratic growth condition (for the problem (P)) holds at $\bar{x}$.

In the remainder of this section we assume that:

(A6) The function f is twice continuously differentiable, the set Ω is a compact subset of $\mathbb{R}^\ell$ and the function $g(x,\omega)$ is twice continuously differentiable on $\mathbb{R}^n\times\mathbb{R}^\ell$ (jointly in x and ω).

Consider a point $\bar{x}\in\mathcal{F}$ such that the set $\Delta(\bar{x})$ is nonempty. Recall that, in such case, $\Delta(\bar{x})$ coincides with the set of maximizers of $g(\bar{x},\cdot)$ over Ω and $\mathfrak{g}(\bar{x})=0$. For $h\in\mathbb{R}^n$ and $\bar{\omega}\in\Delta(\bar{x})$, let $\mathfrak{s}(h,\bar{\omega})$ be the optimal value of the problem

$$\operatorname*{Max}_{\eta\in\mathcal{C}(\bar{\omega})}\left\{2h^{\mathrm{T}}\nabla^2_{x\omega}g(\bar{x},\bar{\omega})\eta+\eta^{\mathrm{T}}\nabla^2_{\omega\omega}g(\bar{x},\bar{\omega})\eta+\sigma\big(\nabla_\omega g(\bar{x},\bar{\omega}),T^2_\Omega(\bar{\omega},\eta)\big)\right\}, \tag{5.45}$$

where

$$\mathcal{C}(\bar{\omega})=\left\{\eta\in\mathbb{R}^\ell:\eta\in T_\Omega(\bar{\omega}),\ \eta^{\mathrm{T}}\nabla_\omega g(\bar{x},\bar{\omega})=0\right\}$$

is the critical cone of the problem of maximization of $g(\bar{x}, \cdot)$ over Ω . Then for any $h \in \mathbb{R}^n$ and $\bar{\omega} \in \Delta(\bar{x})$ the following inequality holds (cf. [4, Proposition 4.129]):

$$\liminf_{\substack{t \downarrow 0 \\ \tilde{h} \to h}} \frac{\mathfrak{g}(\bar{x} + t\tilde{h}) - t\tilde{h}^{\mathrm{T}} \nabla_x g(\bar{x}, \bar{\omega})}{\frac{1}{2}t^2} \geq h^{\mathrm{T}} \nabla^2_{xx} g(\bar{x}, \bar{\omega}) h + \mathfrak{s}(h, \bar{\omega}). \tag{5.46}$$

Note that for $\eta = 0$, the quadratic and sigma terms inside (5.45) are zero, and hence $\mathfrak{s}(h, \bar{\omega})$ is always nonnegative. It can happen, however, that $\mathfrak{s}(h, \bar{\omega}) = +\infty$.

By employing (5.46) and (5.44) it is possible to derive the following second-order sufficient conditions (cf. [4, Theorem 5.116]).

THEOREM 5.3 *Let $\bar{x} \in \mathcal{F}$ be a feasible point of problem* (P) *satisfying first-order necessary conditions* (5.6) *and such that the index set $\Delta(\bar{x})$ and the Lagrange multipliers set $\Lambda(\bar{x})$ are nonempty. Then the following conditions are sufficient for the quadratic growth to hold at the point $\bar{x}$: for every $h \in C(\bar{x}) \setminus \{0\}$ there exists $\lambda = \sum_{i=1}^m \lambda_i \delta(\omega_i) \in \Lambda(\bar{x})$ such that*

$$h^{\mathrm{T}} \nabla^2 L(\bar{x}, \lambda) h + \sum_{i=1}^m \lambda_i \mathfrak{s}(h, \omega_i) > 0. \tag{5.47}$$

Compared with the sigma term $\sigma(\lambda, \mathcal{T}^2(h))$ of second-order conditions (5.43), the above conditions (5.47) have the additional term of the form $-\sum_{i=1}^m \lambda_i \mathfrak{s}(h, \omega_i)$. In some cases the optimal value $\mathfrak{s}(h, \bar{\omega})$, for $\bar{\omega} \in \Delta(\bar{x})$, can be calculated in a closed form. Note that the sigma term in (5.45) vanishes if there is a neighbourhood of $\bar{\omega}$ such that the sets Ω and $\bar{\omega} + T_\Omega(\bar{w})$ do coincide in that neighbourhood. In particular, this sigma term vanishes if the set Ω is polyhedral.

Suppose, for instance, that $\bar{\omega}$ is an interior point of Ω. Then $\nabla_\omega g(\bar{x}, \bar{\omega}) = 0$, and the sigma term in (5.45) vanishes and $\mathcal{C}(\bar{\omega}) = \mathbb{R}^\ell$. If, moreover, the matrix $\nabla^2_{\omega\omega} g(\bar{x}, \bar{\omega})$ is nonsingular, and hence is negative definite, then

$$\mathfrak{s}(h, \bar{\omega}) = -h^{\mathrm{T}} \left[\nabla^2_{x\omega} g(\bar{x}, \bar{\omega})\right] \left[\nabla^2_{\omega\omega} g(\bar{x}, \bar{\omega})\right]^{-1} \left[\nabla^2_{\omega x} g(\bar{x}, \bar{\omega})\right] h.$$

In the setting of Example 5.1 this gives the additional term of (5.47) exactly in the same form as the sigma term of (5.38), and hence in that case there is no gap between the second-order necessary and sufficient conditions.

If the matrix $\nabla^2_{\omega\omega} g(\bar{x}, \bar{\omega})$ is singular, then it can happen that $\mathfrak{s}(h, \bar{\omega}) = +\infty$. This happens if there exists vector η such that $\eta^{\mathrm{T}}[\nabla^2_{\omega\omega} g(\bar{x}, \bar{\omega})]\eta = 0$ while $h^{\mathrm{T}}[\nabla^2_{x\omega} g(\bar{x}, \bar{\omega})]\eta \neq 0$.

As to the question of 'no gap' second-order conditions, it is possible to show the following (cf. [4, Theorem 5.118]).

THEOREM 5.4 *The second-order sufficient conditions of Theorem* 5.3 *are 'no gap' second-order conditions if*: (i) *the MFCQ holds at $\bar{x}$*, (ii) *the set $\Delta(\bar{x})$ is finite*, (iii) *for every point $\bar{\omega} \in \Delta(\bar{x})$ the set Ω is second-order regular at $\bar{\omega}$, and* (iv) *the quadratic growth condition, for the problem of minimization of $\phi(\omega) = -g(\bar{x}, \omega)$ over $\omega \in \Omega$, holds at every point $\bar{\omega} \in \Delta(\bar{x})$.*

6. Rates of convergence of solutions of discretized problems

We assume in this section that the optimal value val(P) of the SIP problem (P) is finite and condition (A1) holds. Consider a sequence of discretizations $\Omega_m = \{\omega_1, \ldots, \omega_m\} \subset \Omega$ of problem (P). Let $\varepsilon_m \downarrow 0$ and $\hat{x}_m$ be an ε_m-optimal solution of the corresponding discretized problem (P_m). That is, $g(\hat{x}_m, \omega) \leq 0$ for all $\omega \in \Omega_m$, $f(\hat{x}_m)$ is finite and $f(\hat{x}_m) \leq \text{val}(P_m) + \varepsilon_m$.

What can be said about convergence of $\hat{x}_m$ to the set of optimal solutions of problem (P) as the meshsize

$$\varrho_m := \sup_{\omega \in \Omega} \operatorname{dist}(\omega, \Omega_m) \tag{6.1}$$

tends to zero. Here $\operatorname{dist}(\omega, \Omega_m)$ denotes the distance from point ω to the set Ω_m with respect to the metric ρ of the space Ω, i.e. $\operatorname{dist}(\omega, \Omega_m) = \min_{1 \le i \le m} \rho(\omega, \omega_i)$. Since Ω_m is a subset of Ω, the deviation of the set Ω from the set Ω_m, written in the right-hand side of (6.1), in fact is the Hausdorff distance between sets Ω_m and Ω.

The following result is well known, since its proof is easy we give it for the sake of completeness.

Lemma 6.1 *Suppose that assumption* (A1) *is fulfilled and the function* $f(\cdot)$ *is lower semicontinuous. If* $\varepsilon_m \downarrow 0$ *and* $\varrho_m \to 0$ *as* $m \to \infty$, *then any accumulation point of a sequence* $\hat{x}_m$, *of* ε_m*-optimal solutions of the discretized problems* (P_m), *is an optimal solution of the problem* (P).

Proof Let $\bar{x}$ be an accumulation point of the sequence $\{\hat{x}_m\}$. By passing to a subsequence if necessary, we can assume that $\hat{x}_m \to \bar{x}$. Let us observe that $g(\bar{x}, \omega) \le 0$ for any $\omega \in \Omega$. Indeed, consider $\omega \in \Omega$. Since $\varrho_m \to 0$, there exists $\omega_m \in \Omega_m$ such that $\omega_m \to \omega$. We have that $g(\hat{x}_m, \omega_m) \le 0$ and $g(\hat{x}_m, \omega_m) \to g(\bar{x}, \omega)$. It follows that $g(\bar{x}, \omega) \le 0$. Now let x be a point such that $g(x, \omega) \le 0$ for all $\omega \in \Omega$. Then $g(x, \omega) \le 0$ for all $\omega \in \Omega_m$, and hence $f(\hat{x}_m) \le f(x) + \varepsilon_m$ for all m. Since, because of the lower semicontinuity, $f(\bar{x}) \le \liminf_{m \to \infty} f(\hat{x}_m)$ and $\varepsilon_m \downarrow 0$, it follows that $f(\bar{x}) \le f(x)$. Since x was an arbitrary point of $\mathcal{F}$, it follows that $\bar{x}$ is an optimal solution of the problem (P). ■

Assume from now on that the function $f : \mathbb{R}^n \to \mathbb{R}$ is real valued and *continuous*. Denote by $\hat{\mathcal{F}}_m$ the feasible set of problem (P_m), i.e.

$$\hat{\mathcal{F}}_m := \{x \in \mathbb{R}^n : g(x, \omega_i) \le 0, \ i = 1, \dots, m\}.$$

Since $g(\cdot, \omega)$ is continuous, the set $\hat{\mathcal{F}}_m$ is closed. It is said that the sets $\hat{\mathcal{F}}_m$ are *uniformly bounded* if there is a bounded set $C \subset \mathbb{R}^n$ such that $\hat{\mathcal{F}}_m \subset C$ for all m. It is straightforward to show, by the above lemma and compactness arguments, that if $\varrho_m \to 0$ and the sets $\hat{\mathcal{F}}_m$ are uniformly bounded, then $\mathbb{D}(\hat{\mathcal{F}}_m, \mathcal{F}) \to 0$ and $\operatorname{dist}(\hat{x}_m, \operatorname{Sol}(P)) \to 0$. Moreover, it is possible to estimate the rate at which $\mathbb{D}(\hat{\mathcal{F}}_m, \mathcal{F})$ tends to zero.

Suppose that the MFCQ holds at a point $\bar{x} \in \mathcal{F}$. Then there exist a neighbourhood $\mathcal{V}$ of $\bar{x}$ and a constant α such that for all $x \in \mathcal{V}$ the following holds

$$\operatorname{dist}(x, \mathcal{F}) \le \alpha\big(\sup\nolimits_{\omega \in \Omega} [g(x, \omega)]_+\big) \tag{6.2}$$

(e.g. [4, Section 2.3, and Example 2.94]). Suppose, further, that the set $\operatorname{Sol}(P)$ is nonempty and bounded (and hence is compact) and the MFCQ is satisfied at every point of the set $\operatorname{Sol}(P)$. It follows then by compactness arguments that there exist α and a neighbourhood $\mathcal{W}$ of $\operatorname{Sol}(P)$ such that (6.2) holds for any $x \in \mathcal{W}$. Suppose, further, that $g(x, \cdot)$ is Lipschitz continuous on Ω uniformly in $x \in \mathcal{W}$, i.e. there exists a constant $\kappa > 0$ such that

$$|g(x, \omega) - g(x, \omega')| \le \kappa \rho(\omega, \omega'), \qquad \forall \omega, \omega' \in \Omega, \ \forall x \in \mathcal{W}. \tag{6.3}$$

Consider now a point $x \in \hat{\mathcal{F}}_m \cap \mathcal{W}$. We have that for any $\omega \in \Omega$ there exists a point $\omega' \in \Omega_m$ such that $\rho(\omega, \omega') \le \varrho_m$. Moreover, since $x \in \hat{\mathcal{F}}_m$, it follows that $g(x, \omega') \le 0$. Together with (6.3) this implies that $g(x, \omega) \le \kappa \varrho_m$. It follows by (6.2) that

$$\mathbb{D}(\hat{\mathcal{F}}_m \cap \mathcal{W}, \mathcal{F} \cap \mathcal{W}) = O(\varrho_m). \tag{6.4}$$

For a constant $p > 0$, we say that the p-th order growth condition (for the problem (P)) holds at a (nonempty) set $\mathcal{S} \subset \mathcal{F}$ if there exist constant $c > 0$ and neighbourhood $\mathcal{V}$ of $\mathcal{S}$ such that

$$f(x) \ge \text{val}(P) + c[\text{dist}(x, \mathcal{S})]^p, \quad \forall x \in \mathcal{F} \cap \mathcal{V}. \tag{6.5}$$

Clearly, it follows from (6.2) that $\mathcal{S}$ is a set of locally optimal solutions of problem (P). In particular, if $\mathcal{S} = \{\bar{x}\}$ is a singleton, then condition (6.5) coincides with condition (4.22) and hence the above definition is consistent with Definition 4.2.

For singleton $\mathcal{S}$ the following result is given in [32] (see also [21, Theorem 12]).

THEOREM 6.1 *Let $\mathcal{S}$ be a nonempty and bounded subset of $\mathcal{F}$. Suppose that*: (i) *p-th order growth condition holds at $\mathcal{S}$ with $p = 1$ or $p = 2$*, (ii) *condition* (A4) *is fulfilled and* $g(x, \cdot)$ *is Lipschitz continuous on Ω uniformly in $x \in \mathcal{V}$*, (iii) *the MFCQ holds at every $x \in \mathcal{S}$*, (iv) *for $\varepsilon_n = O(\varrho_m)$, problem (P_m) has an ε_m-optimal solution $\hat{x}_m$ such that* $\text{dist}(\hat{x}_m, \mathcal{S}) \to 0$. *Then for $p = 1$ and $p = 2$,*

$$\text{dist}(\hat{x}_m, \mathcal{S}) = O\big(\varrho_m^{1/p}\big). \tag{6.6}$$

Proof Since the function $f(x)$ is continuously differentiable, by reducing the neighbourhood $\mathcal{V}$ if necessary we can assume that the function $f(x)$ is Lipschitz continuous on $\mathcal{V}$, with Lipschitz constant denoted η. By the assumption (iv) we have that $\hat{x}_m \in \mathcal{V}$ for all m large enough. Then the following estimates hold (cf. [4, Proposition 4.37 and Remark 4.39]):

$$\text{dist}(\hat{x}_m, \mathcal{S}) \le (1 + c^{-1}\eta) d_m + c^{-1}\varepsilon_m, \quad \text{for } p = 1, \tag{6.7}$$

$$\text{dist}(\hat{x}_m, \mathcal{S}) \le 2d_m + c^{-1/2}\eta^{1/2} d_m^{1/2} + c^{-1/2}\varepsilon_m^{1/2}, \quad \text{for } p = 2, \tag{6.8}$$

where $d_m := \mathbb{D}(\hat{\mathbb{F}}_m \cap \mathcal{V}, \mathcal{F} \cap \mathcal{V})$.. By (6.4) we have (reducing the neighbourhood $\mathcal{V}$ if necessary) that $d_m = O(\varrho_m)$. Together with (6.7) and (6.8) this completes the proof. ■

Suppose, further, that $\Omega \subset \mathbb{R}^\ell$, the function $g(\cdot, \cdot)$ is twice continuously differentiable and for every $\bar{x} \in \mathcal{S}$ the set $\Omega^*(\bar{x})$, defined in (4.14), is contained in the interior of Ω. In that case, we have that $\nabla_\omega g(\bar{x}, \bar{\omega}) = 0$ for every $\bar{\omega} \in \Omega^*(\bar{x})$. It follows by (6.2) that in such case we can estimate d_m, in (6.7) and (6.8), as $d_m = O(\varrho_m^2)$. Consequently, for $p = 1$ and $p = 2$, the estimate (6.6) can be improved to

$$\text{dist}(\hat{x}_m, \mathcal{S}) = O\big(\varrho_m^{2/p}\big). \tag{6.9}$$

For singleton set $\mathcal{S}$ this result is due to Still [32]. It was also argued in [32] that if some of the elements of $\Omega^*(\bar{x})$ are boundary points of Ω, then one needs to add in Ω_m boundary points of Ω in order to achieve such rates of convergence.

Acknowledgements

The author would like to thank Frederic Bonnans for a discussion of second-order optimality conditions, and Diethard Klatte and an anonymous referee for many helpful suggestions. This work was supported in part by the National Science Foundation award DMI-0619977.

Note

1. All subdifferentials and gradients are taken here with respect to x.

References

[1] A. Ben-Tal and J. Zowe, *Second order and related extremality conditions in nonlinear programming*, J. Optim. Theory Appl. 31 (1980), pp. 143–165.
[2] A. Ben-Tal and J. Zowe, *A unified theory of first and second-order conditions for extremum problems in topological vector spaces*, Math. Prog. Study 19 (1982), pp. 39–76.
[3] J.F. Bonnans, R. Cominetti, and A. Shapiro, *Second order optimality conditions based on parabolic second-order tangent sets*, SIAM J. Optim. 9 (1999), pp. 466–492.
[4] J.F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer-Verlag, New York, 2000.
[5] J. Borwein, *Direct theorems in semi-infinite convex programming*, Math. Progr. 21 (1981), pp. 301–318.
[6] R. Cominetti, *Metric regularity, tangent sets and second-order optimality conditions*, J. Appl. Math. Optim. 21 (1990), pp. 265–287.
[7] R. Cominetti and J.P. Penot, *Tangent sets to unilateral convex sets*, Compt. Rend. Acad. Sci. Paris, Sér. I 321 (1995), pp. 1631–1636.
[8] J.M. Danskin, *The Theory of Max–Min and its Applications to Weapons Allocation Problems*, Springer-Verlag, New York, 1967.
[9] K. Glashoff and S.A. Gustafson, *Linear Optimization and Approximation*, Springer, New York, 1983.
[10] M.A. Goberna and M.A. López, *Linear Semi-Infinite Optimization*, Wiley, Chichester, 1998.
[11] M.A. Goberna and M.A. López, *Linear semi-infinite programming theory: An updated survey*, Eur. J. Oper. Res. 143 (2002), pp. 390–405.
[12] R. Hettich and K.O. Kortanek, *Semi-infinite programming: theory, methods and applications*, SIAM Rev. 35 (1993), pp. 380–429.
[13] A.D. Ioffe, *Second order conditions and augmented duality*, SIAM J. Contr. Optim. 17 (1979), pp. 266–288.
[14] H. Kawasaki, *An envelope-like effect of infinitely many inequality constraints on second order necessary conditions for minimization problems*, Math. Prog. 41 (1988), pp. 73–96.
[15] H. Kawasaki, *The upper and lower second order directional derivative of sup-type function*, Math. Prog. 41 (1988), pp. 327–339.
[16] H. Kawasaki, *Second order necessary optimality conditions for minimizing a sup-type function*, Math. Prog. 49 (1990), pp. 213–229.
[17] D. Klatte, *Stable local minimizers in semi-infinite optimization: regularity and second-order conditions*, J. Comp. Appl. Math. 56 (1994), pp. 137–157.
[18] D. Klatte and R. Henrion, *Regularity and stability in non-linear semi-infinite optimization*, in *Semi-infinite Programming*, R. Reemtsen and J.-J. Rueckmann, eds., Kluwer Academic Publishers, Boston, London, Dordrecht, 1998, pp. 69–102.
[19] J. Kyparisis, *On uniqueness of Kuhn-Tucker multipliers in non-linear programming*, Math. Prog. 32 (1985), pp. 242–246.

[20] V.L. Levin, *Application of a theorem of E. Helly in convex programming, problems of best approximation and related topics*, Mat. Sbornik 79 (1969), pp. 250–263.
[21] M. López and G. Still, *Semi-infinite programming*, Eur. J. Oper. Res. 180 (2007), pp. 491–518.
[22] E. Polak, *Optimization, Algorithms and Consistent Approximations*, Springer, Berlin, 1997.
[23] R. Reemtsen and J.-J. Rückmann, *Semi-Infinite Programming*, Kluwer, Boston, 1998.
[24] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.
[25] R.T. Rockafellar, *Conjugate Duality and Optimization*, in *Regional Conference Series in Applied Mathematics*, SIAM, Philadelphia, 1974.
[26] W.W. Rogosinsky, *Moments of non-negative mass*, Proc. Roy.Soc. Lond. Ser. A 245 (1958), pp. 1–27.
[27] G.F. Vázqueza, J.-J. Rückmann, O. Stein, and G. Still, *Generalized semi-infinite programming: A tutorial*, J. Comput. Appl. Math. 217 (2008), pp. 394–419.
[28] A. Shapiro, *On concepts of directional differentiability*, J. Optim. Theory Appl. 66 (1990), pp. 447–487.
[29] A. Shapiro, *On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints*, SIAM J. Optim. 7 (1997), pp. 508–518.
[30] A. Shapiro, *On duality theory of convex semi-infinite programming*, Optimization 54 (2005), pp. 535–543.
[31] O. Stein, *Bi-level Strategies in Semi-infinite Programming*, Kluwer, Boston, 2003.
[32] G. Still, *Discretization in semi-infinite programming: The rate of approximation*, Math. Progr. 91 (2001), pp. 53–69.
[33] V. Strassen, *The existence of probability measures with given marginals*, Ann. Math. Stat. 38 (1965), pp. 423–439.