The Hong Kong Polytechnic University
Department of Applied Mathematics

Optimality Conditions of Semi-Infinite Programming and Generalized Semi-Infinite Programming

Zhangyou Chen

A thesis submitted in partial fulfilment of
the requirements for the degree of Doctor of Philosophy

May 2013

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgment has been made in the text.

ZHANGYOU CHEN

Abstract

Semi-infinite programming has been long an important model of optimization problems, arising from areas such as approximation, control, probability. Generalized semi-infinite programming has also been an active research area with relatively short history. Nonetheless it has been known that the study of a generalized semi-infinite programming problem is much more difficult than a semi-infinite programming problem. The purpose of this thesis is to develop necessary optimality conditions for semi-infinite and generalized semi-infinite programming problems with penalty functions techniques as well as other approaches.

We introduce two types of p-th order penalty functions ($0 < p \leq 1$), for semi-infinite programming problems, and explore various relations between them and their relations with corresponding calmness conditions. Under the exactness of certain type penalty functions and some other appropriate conditions especially second order conditions of the constraint functions, we develop optimality conditions for semi-infinite programming problems. This process is also applied to generalized semi-infinite programming problems after being equivalently transformed into standard semi-infinite programming problems.

Via the transformation of penalty functions of the lower level problems, we study some properties of the feasible set of the generalized semi-infinite programming problem which is known to possess unusual properties such as non-closedness, re-entrant corners, disjunctive structures, and further establish a sequence of approximate optimization problems and approximate properties for generalized semi-infinite programming problems.

We also investigate nonsmooth generalized semi-infinite programming problems via

generalization differentiation and derive corresponding optimality conditions via variational analysis tools. Finally, we characterize the strong duality theory of generalized semi-infinite programming problems with convex lower level problems via generalized augmented Lagrangians.

Acknowledgments

I would like to express my deep gratitude to my supervisor, Professor Xiaoqi Yang, for offering me the opportunity to study at PolyU, and for his ceaseless guidance and support throughout the course of this research. Professor Xiaoqi Yang has always been generous with his time and his knowledge, discussing with me various topics and details of the works. He has encouraged me and pushed me patiently whenever I needed it. Without him this thesis could not have been written.

I would also like to thank Professor Nanjing Huang at Sichuan University for introducing me to nonlinear analysis and optimization and for his encouragement in all these years.

My thanks also go to the fellow research students at Hong Kong for an enjoyable stay.

Contents

1 Introduction and Literature Review **1**

1.1 Semi-Infinite Programming 1

1.2 Generalized Semi-Infinite Programming 5

1.3 Penalty Functions 10

1.4 Purpose of the Thesis 12

1.5 Notation 14

2 Semi-Infinite Programming **18**

2.1 Optimality and Qualifications 18

2.2 Optimality without Penalty 22

2.3 Optimality with Penalty 25

2.4 Lower Order Penalization 27

2.4.1 Penalty Functions and Calmness 28

2.4.2 Optimality Conditions 37

2.5 Numerical Scheme 45

3 Generalized Semi-Infinite Programming **48**

3.1 Introduction 48

3.2 Calmness 51

3.3 Lower Level Penalty Transform 54

3.4 Nonsmooth Generalized Semi-Infinite Programming 60

3.5 Lower Order Penalization 73

3.6 GSIP with Convex Lower Level Problem 82

4 Conclusions **96**

Chapter 1

Introduction and Literature Review

1.1 Semi-Infinite Programming

Semi-infinite programming problems and generalized semi-infinite programming problems consist of an important branch of mathematical programming, arising in various fields such as (reverse) Chebyshev approximation, engineering design, physical science, probability, statistics, optimal control, robust optimization.

Throughout the thesis, we always treat semi-infinite programming problems and generalized semi-infinite programming problems separately. When the index set does not depend on the decision variables, generalized semi-infinite programming problems reduce to standard semi-infinite programming problems. However, generalized semi-infinite programming problems are much harder to solve than standard semi-infinite programming problems and all major methods for standard semi-infinite programming problems cannot be directly applied to the study of generalized semi-infinite programming problems.

The semi-infinite programming problem (SIP) is of the form:

$$\min f(x) \quad \text{s.t.} \quad g(x,t) \leq 0, t \in T, \tag{1.1}$$

where f is a function defined on $\mathbb{R}^n$ and T an arbitrary set. If $f(x) = c^T x$ and $g(x,t) = a(t)^T x + b(t)$ for some functions $a(t)$ and $b(t)$, then problem (1.1) is referred to as the linear semi-infinite programming problem (LSIP), otherwise, as the nonlinear semi-infinite programming problem. If f and $g(\cdot,t), t \in T$ are convex functions, then problem (1.1) is referred to as the convex semi-infinite programming problem. If T depends on x, it is referred to as the generalized semi-infinite programming problem (GSIP).

It has been over fifty years since the term, "semi-infinite" programs, was first introduced by Charnes, Cooper, and Kortanek in 1962 [17]. Since then, semi-infinite programming has been under intensive study and has been a mature area of optimization. There have been a large number of works on semi-infinite programming problems. Some general literature on semi-infinite programming problems include the books [51, 56, 11, 30, 35, 36, 113, 37], optimization textbooks with semi-infinite programming chapters [83, 2, 41, 108, 8] and review articles [107, 53, 125, 119, 89, 45, 126].

In general, we only review the works on nonlinear semi-infinite programming problems with the index set T having a topological structure. However, one point worthy of mentioning is the introduction of the generalized finite sequence space in the 1960s, attached with the formal Lagrange multipliers and dual variables, in linear semi-infinite programming problems irrespective of the structure of the index set. When the index set is a compact Hausdorff space, due to a result of Rogosinski [117], the generalized finite sequence space is related to the finite signed Borel measure space on the index set, which serves as the dual space for general semi-infinite programming problems.

Reduction methods The reduction methods intend to replace the semi-infinite system by a finite system. For a convex semi-infinite system, based on the Helly-type theorem of open convex sets [80], equivalence between SIP and the finite subproblem was established in [4]. The similar approach was extended in [9] to quasi-convex SIP.

Without convexity, the so-called reduction ansatz (linear independence, strict complementarity, second order sufficient conditions) was introduced by Wetterling [140] to describe the feasible set locally by a finite number of smooth functions, and was fully used to derive the second order optimality conditions by Hettich and Zencke [56], see also [124, 71, 78].

Abstract formulation Let $G(x) = g(x,\cdot)$ and $K = C(T)_-$. SIP can be put into the general form of abstract optimization problems in Banach spaces

$$\min f(x) \quad \text{s.t.} \quad G(x) \in K, \tag{1.2}$$

where $f : X \to \mathbb{R}$, $G : X \to Z$, X and Z are Banach spaces and $K \subset Z$ is a closed convex cone with non-empty interior points. Kawasaki [75] derived second order necessary conditions for problem (1.2) via the second order analysis with an emphasis of the envelope-like effect due to the infiniteness of constraints. A new term, also referred to as sigma or curvature term by other authors, occurs along the second order derivative of the Lagrangian function in the optimality conditions. The obtained results were applied to the following min-max problem

$$\min S(x) := \sup\{g(x,t) : t \in T\}, \tag{1.3}$$

involving the sup-tpye functions, see [76, 77] for more details, which can be equivalently reformulated as in form (1.2)

$$\min v \quad \text{s.t.} \quad G(x) - ve \in K,$$

where $e \equiv 1$, an element in $C(T)$. In [76] it was shown that the upper second order derivative of S is closed related to the curvature term.

A more general sup-type function

$$S(x) := \sup\{g(x,t) : t \in T(x)\}, \tag{1.4}$$

was consider by Shapiro [124]. Differentiability properties of S was considered under strong second order sufficient conditions and then was used to derive second order optimality conditions for semi-finite programs. Similar second order conditions are also given by Hettich and Jongen [52] under reduction ansatz.

Nonsmooth formulation Let $h(x) = \sup_{t\in T} g(x,t)$. SIP problem (1.1) can be reformulated as

$$\min f(x) \quad \text{s.t.} \quad h(x) \leq 0. \tag{1.5}$$

In general h is a nonsmooth function even if g is smooth. It is known that h is directional differentiable under the differentiability of g. Pschenichnyi [111] used this model to derive the FJ-type optimality for SIP. Using certain generalized second order directional derivative, such as the one following a parabolic curve, second order analysis can be processed for (1.5). Ben-Tal et al. [5] derived the second order conditions for SIP with C^2 data via a general second order conditions for problem (1.5).

The function h can be viewed as the composition of a max-function (a proper convex function) and a smooth function if g is smooth. Based on the ideas from [64, 65, 66], Ioffe [67] obtained the first and second order optimality conditions for SIP with $C^{1,1}$ data via the Clarke generalized directional derivative.

Converting the semi-infinite constraints into a $C^{1,1}$ constraint

$$\int_a^b \max\{g(x,t),0\}\mathrm{d}t = 0,$$

when g is C^2 and $T = [a,b]$, Yang [141] obtained the second order optimality conditions in terms of generalized second order directional derivatives, see Yang [142, 143], by directly applying the corresponding second order conditions for $C^{1,1}$ constrained optimization problems.

Penalty function methods As an early development of theory of penalty functions, Pietrzykowski [105] devised a sequence of penalty functions (referred to as potential functions) to approximate the constraint optimization problem in locally compact metric spaces. The author also proposed the integral-type penalty function for SIP with penalty parameter $c > 0$

$$f(x) + c\int_T g_+(x,t)\,\mathrm{d}t,$$

to approximate SIP (1.1). Extending the work [105] but requiring some geometric structure of variable spaces and convexity of defining functions, Pietrzykowski [106] obtained the convergence result of penalty functions. The author also suggested the p-order penalty function for some $p > 0$:

$$f(x) + c\int_T (g_+(x,t))^p\,\mathrm{d}t.$$

Conn and Gould [24] introduced a strengthened integral-type penalty function and derived sufficiency conditions for the exact penalization. Polak et al. [109] presented an interior penalty function algorithm for solving semi-infinite min-max problems. The penalty term for the max-type function $\max_{t\in[0,1]} g(x,t)$ takes the form of $\int_{[0,1]} 1/[c - g(x,t)]\,\mathrm{d}t$ with $c > g(x,t)$. A general review on exact penalty functions for SIP can also be found in Coope and Price [25]. Rückmann and Shapiro [122] proposed an augmented Lagrangian approach for SIP and studied necessary and sufficient conditions for existence of the augmented Lagrange multipliers—elements of the space of generalized finite sequence. Recently, Huy and Kim [62] derived sufficient conditions with respect to the limiting subdifferential and limiting normal cone along with other assumptions for the exactness of the sharp Lagrangian of the augmented Lagrangian approach as well as the stability of the perturbed infinite systems.

With the advancement of modern variational analysis, set-valued analysis and nonsmooth analysis, see for example [116, 93, 94, 97, 102], it is possible to use tools of generalized differentials/derivatives to analyze and study the semi-infinite systems and the semi-infinite problems especially those with nonsmooth data. For example, use the limiting/basic/Mordukhovich subdifferentials and/or coderivatives to derive the necessary optimality conditions for nonsmooth SIP problems, to characterize or estimate the Lipschitz-like properties of the set defined by infinite systems, the differentiability properties of the solution set and the optimal value of SIP problems, see [14, 16, 148, 15, 20, 28, 63, 61, 98, 100, 101, 99].

1.2 Generalized Semi-Infinite Programming

The generalized semi-infinite programming problem (GSIP) is of the form

$$\min f(x) \quad \text{s.t.} \quad g(x,y) \leq 0, y \in Y(x), \tag{1.6}$$

with the index set $Y(x)$ usually defined by inequality constraints

$$Y(x) := \{y \in \mathbb{R}^m \mid v(x,y) \leq 0\}. \tag{1.7}$$

This problem is called the generalized (or general) semi-infinite programming problem since the variable space is of finite dimension and the index set is variable dependent and in general is infinite. When Y is not x-dependent, it reduces to the standard (or ordinary) semi-infinite programming problem. The lower level problem associated with GSIP is

$$\max g(x,y) \quad \text{s.t.} \quad y \in Y(x). \tag{Q(x)}$$

Study of GSIP has a much shorter history than the study of SIP problems and GSIP turns out to be a much harder problem than SIP. Early studies on GSIP starting around the middle of 1980s came from applications, for example, maneuverability of robotics by Graettinger [39], Graettinger and Krogh [40], Hettich and Still [54], and the time minimal control problems by Krabs [84]. Other modeling sources also include the reverse Chebyshev approximation problems [58], terminal problems [73, 74], and parametric data envelopment analysis [103].

Some of the early systematic studies of GSIP are Klatte [78] and Hettich and Still [55]. Both articles used the reduction ansatz to transform GSIP into finite nonlinear programming problem ($C^{1,1}$ problem), and thus derived the stability properties of stationary points and the first and second order necessary optimality conditions, respectively. Jongen et al. [68] first derived the FJ-type optimality conditions for GSIP with a basic argument without any regularity assumptions and showed the possible difficulty and complexity of the structure of the feasible set with illustrating examples.

Speciality of GSIP. GSIP has some properties not shared by SIP and nonlinear programming problems. The feasible set of GSIP is not always closed, see examples from [68]. Theoretically, the feasible set of GSIP can be represented as the level set of the value function of the lower level problems. The value function is usually only upper semi-continuous, which is true if the data is continuous and Y is locally bounded, but not lower semi-continuous. Thus the loss of lower semi-continuity will lead to the nonclosedness of the feasible set.

The feasible set of GSIP can also be expressed in the following formula, see [128]:

$$[\mathrm{pr}_x(\mathrm{gph}Y \cap \mathcal{G}^C)]^C, \tag{1.8}$$

where pr_x denotes the projection onto the x-space, $\mathrm{gph}Y = \{(x,y) \mid v(x,y) \leq 0\}$ is the graph of the mapping Y, $\mathcal{G} := \{(x,y) \mid g(x,y) \leq 0\}$, and A^C denotes the set complement. This formula discloses some topological features of the feasible set of GSIP which are not known from standard SIP or finite nonlinear programming: an inherent *disjunctive structure* and *non-closedness*. Another feature of the feasible set of GSIP is the *re-entrant corners*. It has been shown that the properties of non-closedness and re-entrant corners are related to the Mangasarian-Fromovitz constraint qualification (MFCQ) and the linear independence constraint qualification (LICQ) of the lower level problem, respectively, see [135]. These structural observations have lead to closer looks at optimality conditions and solution methods for GSIP.

One of the recent advancements of the study of GSIP is the exploration via the symmetric representation of the closure of the feasible set M. A symmetric representation for the closure of M is as follows, see [46, 47]: For a generic data $(g, v_1, \cdots, v_l)$, and under some additional assumption, that the family $Y(x), x \in \mathbb{R}^n$, is locally bounded, the closure $\overline{M}$ of the feasible set can be described as follows

$$\overline{M} = \{x \in \mathbb{R}^n \mid g(x, y) \leq 0, y \in Y^{<}(x)\} \tag{1.9}$$

$$= \{x \in \mathbb{R}^n \mid \sigma(x, y) \leq 0, y \in \mathbb{R}^n\}, \tag{1.10}$$

where

$$Y^{<}(x) = \{y \in \mathbb{R}^m \mid v_i(x, y) < 0, i = 1, \cdots, l\},$$

and

$$\sigma(x, y) = \min\{g(x, y); -v_1(x, y), \cdots, -v_l(x, y)\}.$$

Moreover, for each $\bar{x} \in \overline{M}$, there exists some neighborhood U of $\bar{x}$ and a nonempty compact set V such that

$$\overline{M} \cap U = \{x \in U \mid \sigma(x, y) \leq 0, y \in V\}. \tag{1.11}$$

An earlier similar result from Stein [127] is that

$$\overline{M} \cap U \subset \{x \in U \mid \sigma(x, y) \leq 0, y \in V\}. \tag{1.12}$$

Motivated by the simple and symmetric representation of the closure of the feasible set of GSIP, some recent works focused on the characterizations of the closure of the feasible set of GSIP problem and the corresponding problem restricted on its closure. Günzel et al. [47] proposed the symmetric reduction ansatz and represented the $\overline{M}$ locally by a set defined by a finite number of inequalities. Guerra-Vázquez et al. [43] proposed the symmetric MFCQ assumption for points in the right hand of equation (1.10). Under symmetric MFCQ, relation (1.10) holds and the interior and boundary of M have easy representations related to σ. Jongen and Shikhman [69] introduced the nonsmooth symmetric reduction ansatz to generalize the ansatz in [47] and thus described M as a set defined by finitely many inequalities of max-type functions.

Approaches for GSIP One core idea of the study of GSIP is to transform GSIP problem into known problems, such as standard semi-infinite programming problem or nonlinear programming problem.

Reduction ansatz Reduction ansatz is a method introduced by Wetterling [140] as well as Hettich and Jongen [52] to transform the standard SIP into finite nonlinear programming problems. As mentioned earlier, Klatte [78] and Hettich and Still [55] applied the reduction method to GSIP case.

The basic assumptions of reduction ansatz at given point $\bar{x}$ are the assumptions of nondegeneracy (i.e. LICQ, strict complementarity slackness, and second order sufficiency condition) or the strong stability in the sense of Kojima [81] at all $\bar{y} \in Y_0(\bar{x})$ along with other mild conditions. Under reduction ansatz, $Y_0(\bar{x})$ is finite and the feasible set M can be finitely defined around $\bar{x}$ by

$$M \cap U = \{x \in U \mid g(x, y_i(x)) \leq 0, i = 1, \cdots, s\},$$

where U is a neighborhood of $\bar{x}$ and $y_i, i = 1, \cdots, s$, are continuously differentiable functions defined on U and thus their derivatives can be calculated explicitly.

It is possible that GSIP can be transformed into standard SIP problem if some stability property is present on the index set Y. In fact, $Y(x)$ is homeomorphic to $Y(\bar{x})$ for all x near $\bar{x}$ if Mangasarian-Fromovitz constraint qualification (MFCQ) holds on $Y(\bar{x})$, see, Guddat et al [42]. Assuming the compactness of the feasible set of GSIP and LICQ for lower level problem, Weber [139] transformed it into a standard SIP. So did Still [135] by replacing LICQ with MFCQ.

Under these schemes, these authors established the optimality conditions for GSIP via the reduced problems and also prepared ways to possible numerical aspects for GSIP. However, the reduction ansatz, by eliminating the variable dependence of the index set, has its drawbacks. It may destroy some properties of the defining function such as linearity, convexity; it may be too expensive to compute the transformation and the new index set explicitly or approximately; in contrast to SIP, the assumptions to guarantee the reduction cannot be interpreted as 'weak' assumptions.

Nonlinear programming formulation. Letting $\phi(x)$ be the value function of the lower level problem $Q(x)$, GSIP can be equivalently formulated as the standard (probably nonsmooth) mathematical programming problem

$$\min f(x) \quad \text{s.t.} \quad \phi(x) \leq 0. \tag{1.13}$$

Under this reformulation and with some estimates (the first and second order directional derivatives) of the value function of $Q(x)$ and other appropriate assumptions (some CQ's on lower level problems), Ruckmann and Shapiro [120, 121] derived the first and the second order optimality conditions for GSIP. They gave a shorter proof for the optimality conditions in [68]. Extending the results by Ruckmann and Shapiro [120], Stein [127] considered GSIP without CQ's on the lower level problems (referred to as degenerate).

GSIP with convex lower level problems. Another class of tractable GSIP is the one with convex lower level problems. The convexity means that for all $x \in \mathbb{R}^n$, all functions $-g(x, \cdot)$ and $v_i(x, \cdot), i = 1, \cdots, l$, are convex. However, the feasible set of GSIP is not necessarily convex even though the functions g and v are assumed to be linear with respect to (x, y).

A number of works explored this convexity property to replace the lower level problems via either the strong duality theory or KKT optimality system of the lower level problems. Thus GSIP problems are reduced into known optimization problems and various optimization techniques are introduced to deal with the reduced GSIP problems, such as the branch-and-bound approach by Levitin and Tichatschke [86], the smoothing technique via regularization by Levitin and Tichatschke [86], interior point technique by Stein and Still [131], and the semi-smooth approach by Stein and Tezel [132, 133]. More works in this aspect can be referred to Levitin [87] and Stein and Winterfeld [134].

Constraint qualifications A major topic in optimization theory is to explore properties, referred to as constraint qualifications or regularity conditions, among the feasible set and its defining functions in order to obtain more informative optimality conditions. The development of constraint qualifications is parallel to its study in nonlinear programming. The extended Mangasarian-Fromovitz constraint qualification (EMFCQ) was first introduced by Jongen et at. [68] to obtain the KKT-type optimality conditions for GSIP. Guerra-Vázquez and Rückmann [44] developed two extensions of Kuhn-Tucker constraint qualification that lead to KKT-type optimality conditions and discussed various relationships among them and the Abadie constraint qualification [127], EMFCQ. Recently, extending existing results, Ye and Wu [145] considered various constraint qualifications leading to a weaker form of KKT-type optimality conditions for GSIP which is natural under the setting of the infinite index.

1.3 Penalty Functions

One of the techniques we used in this thesis is the theory of penalty functions such as exact lower order penalty functions for standard and generalized semi-infinite programming problems, penalty functions as approximation and augmented Lagrangians for generalized semi-infinite programming problems.

By associating the constrained optimization problem $\min\{f(x) : x \in X\}$, with a sequence of unconstrained optimization problems $\min\{f(x) + cP(x) : x \in \mathbb{R}^n\}$ where P is a nonnegative function that characterizes the feasible set X by $P = 0$ and $c > 0$ is a penalty parameter, the penalization theory maintains that under mild assumptions there exists a solution path $x(c)$ of the unconstrained problems converging to $x(\infty)$ which belongs to the solution set of original constrained optimization problem, see for example Eremin [29], Zangwill [147], Pietrzykowski [105], and Fiacco and McCormick [31].

A more attractive observation is the possibility of existence of exact penalty functions where the constrained problem can be replaced by one unconstrained problem or the convergence of $x(c)$ is finite, see, e.g., [29, 147, 50, 104].

The common penalty functions include $P(x) = \text{dist}(x, X)$, the l_∞ penalty $P(x) = \max\{(g_i)_+(x), h_j(x) : i \in I, j \in J\}$, the l_1 penalty $P(x) = \sum_{i\in I}(g_i)_+(x) + \sum_{i\in J}|h_j(x)|$, and the quadratic penalty $P(x) = \sum_{i\in I}(g_i)_+(x)^2 + \sum_{i\in J}|h_j(x)|^2$, when $X = \{x|g_i(x) \leq 0, h_j(x) = 0, i \in I, j \in J\}$. The most radical one, always exact, is the indicator function $P(x) = \delta_X(x)$.

Another known type of penalty functions is the augmented Lagrangian, combining Lagrangians with penalty methods, or called method of multipliers, see, e.g., Hestenes [50] and Rockafellar [114] where the constrained optimization problems with feasible sets defined by equality constraints and equality/inequality constraints are considered, respectively. More details for method of multipliers are referred to the book by Bertsekas [6].

The importance of penalty methods not only lies in that it provides an effective approach to solve optimization problems but also its close relationship with basic theory of optimization. Ioffe [64] established the first order necessary optimality conditions of constrained optimization problems via variational analysis through the exact penalty function

$$\max\{f(x), \max_{i\in I} g_i(x)\} + c(\|h(x)\| + \text{dist}(S, x)),$$

where $X := \{x \in S \mid g(x) \leq 0, h(x) = 0\}$. Based on the same idea of the above exact penalty function, optimality conditions were further developed in [65, 66].

Han and Mangasarian [48] derived sufficient conditions for the exactness of a general class of penalty functions, and obtained the KKT-type optimality conditions via exact penalty functions working as a constraint qualification.

Fletcher [33] also provided a simple derivation of the first and second order optimality conditions of constrained optimization problems by virtue of the l_1 penalty function.

While providing a comprehensive review of the theory of exact penalty functions, Burke [13] obtained the fundamental theory of constrained optimization problems by use of l_1 exact penalty functions. Since the exactness of l_1 penalty functions of requires constraints regularity conditions, different penalty functions are desirable under weaker conditions or in case of degenerate problems. Warga [138] established the equivalence between constrained optimization problems with some generalized analytic functions and the exact penalty functions of lower order. Luo et al. [91] extended the work of Warga to mathematical programs with equilibrium constraints.

Rubinov and Yang [118] showed that the exactness of lower order penalty functions (or referred to as non-Lipschitz penalty functions) is equivalent to the generalized calmness-type condition. Clarke [22] and Burke [12] established the equivalence between the exactness of penalty functions of order one and the calmness condition. Huang and Yang [60] proposed a unified augmented Lagrangian scheme including the lower order penalty functions to derive the theory of duality and exact penalization extending the work of Rockafellar and Wets with convex augmenting functions. Recently, Yang and Meng [144] derived optimality conditions in terms of Dini directional derivatives via lower order exact penalty functions for mathematical programming problems and obtained the KKT-type optimality conditions by further introducing the second order qualifications via generalized second order directional derivatives. Meng and Yang [92] further developed KKT-type optimality conditions via the contingent derivative of lower order penalty functions for mathematical programming problems and mathematical programs with complementarity constraints.

The most extremal and simple penalty function involving the indicator function of X is

$$f(x) + \delta_X(x),$$

which is obviously equivalent to the original constrained optimization problem and serves as a basic variant model for constrained optimization. The development of optimality conditions via this penalty function is closely related to generalized differentiation and calculus rules from modern variational analysis, exemplified by Mordukhovich [94].

Another element permeating the process of analysis in this thesis is the direct and indirect use of generalized differentiation, a centerpiece of variational analysis and its applications. The presence of nonsmoothness is no matter that the optimization problem itself is smooth or nonsmooth. Considering nonsmooth semi-infinite programming, we present the fundamental theory of semi-infinite programming in the framework of generalized differentiation. The advancement of variational analysis invigorates optimization theory both in depth and width, see more details from the monographs of Rocakfellar and Wets [116] and Mordukhovich [93, 94].

1.4 Purpose of the Thesis

Optimality conditions at the core of optimization theory not only provide criteria to identify the optimal solutions but also prepare passages of numerical analysis. This thesis concentrates on the development of various necessary optimality conditions for both semi-infinite programming problems and generalized semi-infinite programming problems. We propose and analyze different approaches to the study of semi-infinite programming problems and generalized semi-infinite programming problems.

In the first part of the thesis, we propose two types of lower order penalty functions, max-type and integral-type, for SIP problems and analyze the relationships between each other and with the corresponding concepts of calmness. The max-type penalty function generalizes the l_∞ penalty function in nonlinear programming. Under the exactness of the 1-order max-type penalty functions, the KKT-type optimality conditions hold. The integral-type penalty functions, which are stronger than the max-type penalty functions, are more attuned to SIP. Under the exactness of the low order integral-type penalty functions along with some assumptions of second order derivatives of the defining functions, we obtain necessary optimality conditions which are slightly weaker than the KKT-type optimality conditions. Different from the penalty function theory of nonlinear programming, we show by examples that the proposed lower order penalty functions neither imply nor be implied by the Abadie constraint qualification. Thus, lower order penalty functions at least theoretically can deal with some of SIP that cannot be dealt with existing approaches.

In the second part of the thesis, we focus on the study of generalized semi-infinite programming problems. Firstly, applying the lower order penalty function technique used in the first part, we derive the necessary optimality conditions for GSIP after an appropriate transformation. The transformation is done through the augmented Lagrangians of the lower level problems of GSIP. The success of the lower order penalty approach depends on the calculus of the generalized second order directional derivatives of the augmented Lagrangians.

Secondly, we consider approximation schemes for GSIP. The feasible set of GSIP is in general of unusual properties not shared with standard semi-infinite programming problems and finite problems. Also, GSIP is in itself of bilevel structure and its every feasible point is a global solution of the lower level problems. Techniques to deal with the complex structure are required. We propose penalty functions of simple constructions to approximate of the irregular feasible set and thus give approximate problems and derive convergence results.

Thirdly, due to the intrinsic nonsmoothness, we may consider GSIP with Lipschitz continuous functions and derive corresponding optimality conditions by tools of modern variational analysis. We propose two different approaches. One is to reformulate the GSIP as a min-max problem via the optimal value functions of the lower level problems. The other one is to formulate GSIP as a bilevel problem. Both approaches depend on the estimates of the generalized subdifferentials of the optimal value functions of the lower level problems.

Finally, we consider the weak and strong duality theory for GSIP via the augmented Lagrangian approach. A generalized augmented Lagrangian will be constructed for a mathematical program with complementarity constraints which is derived from GSIP with convex lower level problems.

The outline of the thesis is as follows. Section 2.1 presents a basic review of the optimality conditions and qualifications for semi-infinite problems. Section 2.2 and 2.3 present no penalty-related and penalty-related approaches to derive necessary optimality conditions. Section 2.4 considers a lower order exact penalty functions approach for semi-infinite programming problems. In Chapter 3, we focus on generalized semi-infinite problems. Section 3.2 considers the concept of calmness for generalized semi-infinite problems. Section 3.3 considers approximations of generalized semi-infinite problems via the lower order penalty transformations. Section 3.4 derives optimality conditions for nonsmooth generalized semi-infinite problems. Section 3.5 continues the lower order penalty approach for generalized semi-infinite problems. Section 3.6 considers generalized semi-infinite problems with convex lower problems.

1.5 Notation

Lastly, we summarize some notations and definitions which will be used throughout the thesis.

Let B denote the closed unit ball of the n-dimensional vector space $\mathbb{R}^n$ and $B(x, r)$ denote the ball center at x with radius r. For $C \subset \mathbb{R}^n$, co$C$, cone$C$, int$C$, and cl$C$ denote the convex hull, convex cone hull, interior, and closure of C, respectively.

A set C is locally closed at $\bar{x}$ if there is a neighborhood V of $\bar{x}$ such that $C \cap V$ is closed. The indicator function δ_C of C is defined by

$$\delta_C(x) = \begin{cases} 0 & \text{if } x \in C, \\ \infty & \text{if } x \notin C. \end{cases}$$

The horizon cone C^∞ of C is defined by

$$C^\infty = \begin{cases} \{x \mid \exists x^k \in C, \lambda^k \downarrow 0, \text{with } \lambda^k x^k \to x\} & \text{if } C \neq \emptyset, \\ \{0\} & \text{if } C = \emptyset. \end{cases}$$

The polar cone C° of C is defined by $C^\circ = \{x^* \mid \langle x, x^* \rangle \leq 0 \text{ for all } x \in C\}$.

A cone $K \subset \mathbb{R}^n$ is pointed if the only solution of $x_1 + \cdots + x_p = 0$ with $x_i \in K$ is $x_i = 0$ for all $i = 1, \cdots, p$.

Let $\{C^k\}_{k=1}^{\infty}$ be a sequence of subsets in $\mathbb{R}^n$. The outer and inner limit sets of $\{C^k\}$ are respectively defined by

$$\limsup_{k\to\infty} C^k = \{x \in \mathbb{R}^n \mid \exists \text{ subsequence } \{C^i\}_{i=1}^{\infty} \text{ of } \{C^k\}, \exists x^i \in C^i \text{ with } x^i \to x\},$$
$$\liminf_{k\to\infty} C^k = \{x \in \mathbb{R}^n \mid \forall k \text{ large } \exists x^k \in C^k \text{ with } x^k \to x\}.$$

The limit of the sequence exists if the outer and inner limit sets are equal:

$$\lim_{k\to\infty} C^k := \limsup_{k\to\infty} C^k = \liminf_{k\to\infty} C^k.$$

The extended reals $\overline{\mathbb{R}} := \mathbb{R} \cup \{\pm\infty\}$. For a function $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$, the effective domain of f is denoted by

$$\mathrm{dom} f := \{x \in \mathbb{R}^n \mid f(x) < \infty\}.$$

We call f a proper function if $f(x) < \infty$ for at least one $x \in \mathbb{R}^n$ and $f(x) > -\infty$ for all $x \in \mathbb{R}^n$; otherwise it is improper. The epigraph, hypograph and graph of f are respectively defined by

$$\begin{aligned} \mathrm{epi} f &= \{(x, r) \in \mathbb{R}^{n+1} \mid f(x) \le r\}, \\ \mathrm{hypo} f &= \{(x, r) \in \mathbb{R}^{n+1} \mid r \le f(x)\}, \\ \text{and} \quad \mathrm{gph} f &= \{(x, f(x)) \mid x \in \mathbb{R}^n\}. \end{aligned}$$

A function f is lower semi-continuous at $\bar{x}$ (lsc) if $\liminf_{x\to\bar{x}} f(x) \ge f(\bar{x})$; it is upper semi-continuous at $\bar{x}$ (usc) if $-f$ is lsc at $\bar{x}$. We say f is continuous at $\bar{x}$ if it is both lsc and usc at $\bar{x}$. The level set of f is denoted by

$$\mathrm{lev}_{\le\alpha} f := \{x \in \mathbb{R}^n \mid f(x) \le \alpha\}.$$

A function f is (lower) level bounded if the level set $\mathrm{lev}_{\le\alpha} f$ is bounded for each $\alpha \in \mathbb{R}$.

A function f is Lipschitz continuous on $D \subset \mathbb{R}^n$ if there is a constant $\kappa \in \mathbb{R}_+ := [0, \infty)$ with

$$|f(x) - f(y)| \le \kappa \|x - y\|, \text{ for all } x, y \in D.$$

The constant κ is called a Lipschitz constant for f on D. We say f is Lipschitz continuous at $\bar{x}$ or locally Lipschitz at $\bar{x}$ if there is a neighborhood V of $\bar{x}$ and constant $\kappa \in \mathbb{R}_+$

with

$$|f(x) - f(y)| \leq \kappa \|x - y\|, \text{ for all } x, y \in V.$$

For set-valued mapping $S\colon \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, the graph of S is denoted by $\mathrm{gph} S := \{(x, y) \mid y \in S(x)\}$. The outer and inner limits of S at $\bar{x}$ are respectively defined by:

$$\limsup_{x \to \bar{x}} S(x) = \bigcup_{x^k \to \bar{x}} \limsup_{k \to \infty} S(x^k), \quad \text{and} \quad \liminf_{x \to \bar{x}} S(x) = \bigcap_{x^k \to \bar{x}} \liminf_{k \to \infty} S(x^k).$$

We say S is outer semi-continuous at $\bar{x}$ (osc) if

$$\limsup_{x \to \bar{x}} S(x) \subset S(\bar{x}),$$

or equivalently $\limsup_{x \to \bar{x}} S(x) = S(\bar{x})$, and inner semi-continuous at $\bar{x}$ (isc) if

$$\liminf_{x \to \bar{x}} S(x) \supset S(\bar{x}),$$

or equivalently when S is closed-valued, $\liminf_{x \to \bar{x}} S(x) = S(\bar{x})$. It is called continuous at $\bar{x}$ if both conditions hold, i.e., if $S(x) \to S(\bar{x})$ as $x \to \bar{x}$.

A mapping S has the Aubin property (or Lipschitz-like property) relative to X at $\bar{x}$ for $\bar{y}$, where $(\bar{x}, \bar{y}) \in \mathrm{gph} S$, if $\mathrm{gph} S$ is locally closed at $(\bar{x}, \bar{y})$ and there are neighborhoods V of $\bar{x}$ and W of $\bar{y}$ and a constant $\kappa \in \mathbb{R}_+$ such that

$$S(x') \cap W \subset S(x) + \kappa \|x - x'\| \mathbb{B} \quad \text{for all } x, x' \in X \cap V.$$

The tangent cone T_C and regular tangent cone $\hat{T}_C$ of C at $\bar{x}$ are respectively defined by

$$T_C(\bar{x}) = \limsup_{\lambda \downarrow 0} \frac{C - \bar{x}}{\lambda} \quad \text{and} \quad \hat{T}_C(\bar{x}) = \liminf_{x \xrightarrow{C} \bar{x}, \lambda \downarrow 0} \frac{C - x}{\lambda}.$$

The regular normal cone $\hat{N}_C$ of C at $\bar{x}$ is defined by

$$\hat{N}_C(\bar{x}) = \{v \in \mathbb{R}^n \mid \limsup_{x \xrightarrow{C} \bar{x}} \langle v, x - \bar{x} \rangle / \|x - \bar{x}\| \leq 0\},$$

The (general) normal cone N_C of C at $\bar{x}$ is defined by

$$N_C(\bar{x}) = \limsup_{x \to \bar{x}} \hat{N}_C(x).$$

A set C is regular at $\bar{x} \in C$ in the sense of Clarke if it is locally closed at $\bar{x}$ and every normal vector of C at $\bar{x}$ is a regular normal vector, i.e., $N_C(\bar{x}) = \hat{N}_C(\bar{x})$. A function

$f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ is called subdifferentially regular at $\bar{x}$ if $f(\bar{x})$ is finite and epi$f$ is Clarke regular at $(\bar{x}, f(\bar{x}))$ as a subset of $\mathbb{R}^n \times \mathbb{R}$.

The regular subdifferential $\hat{\partial} f$ of f at $\bar{x}$ is defined by

$$\hat{\partial} f(\bar{x}) = \{u \in \mathbb{R}^n \mid \liminf_{x \to \bar{x}, x \neq \bar{x}} (f(x) - f(\bar{x}) - \langle u, x - \bar{x}\rangle)/\|x - \bar{x}\| \geq 0\}.$$

The (general) subdifferential ∂f and singular subdifferential $\partial^\infty f$ of f at $\bar{x}$ are respectively defined by

$$\partial f(\bar{x}) = \limsup_{x \xrightarrow{f} \bar{x}} \hat{\partial} f(x) \quad \text{and} \quad \partial^\infty f(\bar{x}) = \limsup_{\substack{x \xrightarrow{f} \bar{x} \\ \lambda \downarrow 0}} \lambda \hat{\partial} f(x).$$

The upper regular subdifferential $\hat{\partial}^+ f$ and upper subdifferential $\partial^+ f$ of f at $\bar{x}$ are respectively defined by

$$\hat{\partial}^+ f(\bar{x}) = -\hat{\partial}(-f)(\bar{x}) \quad \text{and} \quad \partial^+ f(\bar{x}) = \limsup_{x \xrightarrow{f} \bar{x}} \hat{\partial}^+ f(x).$$

It turns out that [116]

$$\begin{aligned}
\hat{\partial} f(\bar{x}) &= \{v \mid (v, -1) \in \hat{N}_{\mathrm{epi} f}(\bar{x}, f(\bar{x}))\}, \\
\partial f(\bar{x}) &= \{v \mid (v, -1) \in N_{\mathrm{epi} f}(\bar{x}, f(\bar{x}))\}, \\
\partial^+ f(\bar{x}) &= \{v \mid (-v, 1) \in N_{\mathrm{hypo} f}(\bar{x}, f(\bar{x}))\},
\end{aligned}$$

which may as well serve as the definitions of these subdifferentials of f at $\bar{x}$. While

$$\partial^\infty f(\bar{x}) \subset \{v \mid (v, 0) \in N_{\mathrm{epi} f}(\bar{x}, f(\bar{x}))\}$$

which holds with equality when f is locally lsc at $\bar{x}$.

Chapter 2

Semi-Infinite Programming

In this chapter we consider the following semi-infinite programming problems

$$\min f(x) \quad \text{s.t.} \quad g(x,t) \leq 0, t \in T, \tag{2.1}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g\colon \mathbb{R}^n \times T \to \mathbb{R}$ are continuouly differentiable and T is a Hausdorff compact space.

Although the linear and convex SIP and semi-infinite systems are precursors of the study of SIP and are still active research aspects of SIP, see, e.g., [36, 38, 9, 90], we only concentrate on the nonlinear semi-infinite programming problems and its optimality conditions and related qualifications.

2.1 Optimality and Qualifications

In this section we provide a review of basic optimality conditions for semi-infinite programming problems and its various constraint qualifications. One purpose of this section is to emphasize the impact of the intrinsic nature of the 'infiniteness' in SIP, which leads to different types of optimality conditions for SIP compared with nonlinear programming problems.

First, we begin with some basic notations and definitions. Denote its feasible set by

$$M = \{x \in \mathbb{R}^n \mid g(x,t) \leq 0, t \in T\},$$

and the active index set at x

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

Recall the tangent cone of M at x is defined by

$$T_M(x) := \limsup_{\lambda \downarrow 0} \lambda^{-1}(M - x),$$

and define the linearized cone of M at x by

$$L_M(x) := \{d : \langle \nabla g(x,t), d\rangle \le 0, t \in T(x)\},$$

Denote the derivable cone of M at x by

$$\Gamma_M(x) := \{d : \text{there is a curve } x(s) \in M \text{ for } s \in [0, s_0] \text{ such that } x(0) = x, x'_+(0) = d\}$$

and the strictly linearized cone

$$l_M(x) := \{d : \langle \nabla g(x,t), d\rangle < 0, t \in T(x)\}.$$

Then the following inclusions hold:

$$l_M(x) \subset \Gamma_M(x) \subset T_M(x) \subset L_M(x).$$

Denote by $N'(x)$ the convex cone generated by $\{\nabla_x g(x,t), t \in T(x)\}$:

$$N'(x) := \operatorname{cone}(\cup_{t \in T(x)} \nabla_x g(x,t)).$$

It is easy to see that $L_M(x) = N'(x)^\circ = (\operatorname{cl} N'(x))^\circ$.

Definition 2.1.1. *We say that the FJ-type optimality condition of SIP problem* (2.1) *holds at a local optimal solution $\bar{x} \in \mathbb{R}^n$ if there exist $\lambda_i \ge 0, i = 0, \cdots, k$, and $t_i \in T(\bar{x}), i = 1, \cdots, k$ such that $k \le n$, $\sum_{i=0}^{k} \lambda_i = 1$, and*

$$\lambda_0 \nabla f(\bar{x}) + \sum_{i=1}^{k} \lambda_i \nabla g(\bar{x}, t_i) = 0. \tag{2.2}$$

We say that the KKT-type optimality condition of SIP problem (2.1) *holds at a local optimal solution $\bar{x} \in \mathbb{R}^n$ if there exist $\lambda_i \ge 0$, and $t_i \in T(\bar{x}), i = 1, \cdots, k$, such that $k \le n$ and*

$$\nabla f(\bar{x}) + \sum_{i=1}^{k} \lambda_i \nabla g(\bar{x}, t_i) = 0. \tag{2.3}$$

We also say that the feasible point $\bar{x}$ is an FJ/KKT point of SIP if the FJ/KKT-type optimality condition of SIP holds at $\bar{x}$.

It is easy to see that (2.2) and (2.3) are respectively equivalent to the conditions

$$0 \in \mathrm{co}\,\{\nabla f(\bar{x}), \nabla g(\bar{x}, t), t \in T(\bar{x})\}, \text{ and}$$
$$\nabla f(\bar{x}) \in N'(\bar{x}).$$

Below is a basic theorem in optimization theory.

Lemma 2.1.1 (Farkas lemma)**.** *Let $A \subset \mathbb{R}^n$ be arbitrary. Then for every $v \in \mathbb{R}^n$, exactly one of the following holds:*

(i) $v \in \mathrm{cl}(\mathrm{cone}A)$*;*

(ii) *there is a solution η of*

$$\langle \eta, v \rangle > 0, \quad \langle \eta, a \rangle \leq 0, a \in A.$$

The Farkas lemma, as well as its various variants, is a centerpiece in theory of optimization and the inequality/equality systems. One important feature here is that the 'cl' in the first relation is caused by the 'infiniteness' of the set A.

Proposition 2.1.1. *The function $\phi(x) := \max_{t \in T} g(x, t)$ is directionally differentiable with*

$$\phi'(x; d) = \max_{t \in T(x)} \langle \nabla_x g(x, t), d \rangle, \quad \textit{for all } d \in \mathbb{R}^n.$$

If $\bar{x}$ is a local solution of SIP (2.1), then $\bar{x}$ is also a local solution of the unconstrained optimization problem

$$\min_x \max\{f(x) - f(\bar{x}), \max_{t \in T} g(x, t)\}.$$

Applying Proposition 2.1.1 to the max-function $\max\{f(x) - f(\bar{x}), \max_{t \in T} g(x, t)\}$, we obtain the necessary optimality conditions

$$\max\{\langle \nabla f(\bar{x}), d \rangle, \max_{t \in T(\bar{x})} \langle \nabla g(\bar{x}, t), d \rangle\} \geq 0, \text{ for all } d \in \mathbb{R}^n,$$

which is equivalent to

$$0 \in \mathrm{co}\{\nabla f(\bar{x}); \nabla g(\bar{x}, t), t \in T(\bar{x})\}.$$

An alternative approach is given by Abadie [1] as follows.

Proposition 2.1.2. *Let $\bar{x}$ be a local optimal solution of SIP* (2.1)*. Then the following necessary optimality condition holds:*

$$\langle \nabla f(\bar{x}), d \rangle \geq 0 \quad \textit{for all } d \in T_M(\bar{x}).$$

To obtain more informative optimality conditions, certain constraint qualifications (CQ) are needed. We list the constraint qualifications as follows.

(i) $l_M(x) \neq \emptyset$,

(ii) $T_M(x) = L_M(x)$,

(iii) $\mathrm{cl}\,(\mathrm{co}T_M(x)) = L_M(x)$,

(iv) $T_M(x)^\circ = N'(x)$,

(v) $T_M(x)^\circ = \mathrm{cl}\, N'(x)$,

(vi) $T_M(x) = N'(x)^\circ$.

Compared with CQ's of nonlinear optimization, the first three are usually referred to as the extended Mangasarian-Fromovitz constraint qualification (EMFCQ), Abadie's constraint qualification (EACQ), and Guignard constraint qualification (GCQ), respectively. EMFCQ was first appeared in Jongen et al. [70] and was earlier introduced by Krabs [82] as regularity. For ACQ it may be referred to [88]. Unlike in nonlinear programming, ACQ is not sufficient to derive the KKT-tpye optimality conditions. The following proposition establishes the relationships between the preceding CQ's and various optimality conditions.

Proposition 2.1.3. (1) *The relations between the above various conditions are*

$$\text{(i)} \Rightarrow \text{(ii)} \Rightarrow \text{(iii)}; \text{(iv)} \Rightarrow \text{(iii)}; \text{(ii)} = \text{(vi)}; \text{(iii)} = \text{(v)}.$$

(2) *Under either* (i) *or* (iv)*, the KKT-tpye optimality condition holds; under either* (ii) *or* (iii)*, the primal type optimality holds:*

$$\langle \nabla f(x), d \rangle \geq 0 \quad \textit{for all } d \in L_M(x),$$

which by Farkas Lemma is equivalent to

$$-\nabla f(x) \in \mathrm{cl}\, N'(x).$$

Based on above discussions, the optimality conditions can be written in various forms, primal or dual forms, under various constraint qualifications, as follows.

(i) $\langle \nabla f(x), d\rangle \geq 0, \quad$ for all $d \in T_M(x)$,

(ii) $\langle \nabla f(x), d\rangle \geq 0, \quad$ for all $d \in L_M(x)$,

(iii) $\langle \nabla f(x), d\rangle \geq 0, \quad$ for all $d \in l_M(x)$,

(iv) $\max\{\langle \nabla f(x), d\rangle, \langle \nabla g(x,t), d\rangle, t \in T(x)\} \geq 0, \quad$ for all $d \in \mathbb{R}^n$,

(v) $0 \in \operatorname{co}\{\nabla f(x), \nabla g(x,t), t \in T(x)\}$,

(vi) $\nabla f(x) \in \operatorname{cl} N'(x)$,

(vii) $\nabla f(x) \in N'(x)$.

In all, we may have three different types of optimality conditions – (v), (vi), (vii). (v) is equivalent to (iv) and is called the FJ-type optimality. (vii) is the KKT-type optimality which holds under EMFCQ. (vi) is equivalent to (ii) which may hold under ACQ. (i), (iii) , and (v) hold without any qualifications. The closeness of $N'(x)$ is a key property which differs from nonlinear programming and is always closed. In semi-infinite settings, the system $\{\nabla g(x,t), t \in T(x)\}$ is also referred to as Farkas-Minkowski system if $N'(x)$ is closed.

2.2 Optimality without Penalty

Note that in this section the normal cones and the corresponding subdifferentials can be understood as in convex analysis as related sets are convex.

Let $\bar{x}$ be a local optimal solution of SIP (2.1). Let $K = C(T)_-$ be the nonempty closed convex subset of $C(T)$, which is the space of continuous functions on T. The polar cone of K is the set of nonnegative measures $\mu \in C(T)^*$, i.e., $\mu(A) \geq 0$ for any Borel set $A \subset T$.

Let $y \in K$. Denote by $I(y)$ the contact points of y:

$$I(y) = \{t \in T \mid y(t) = 0\}. \tag{2.4}$$

Then the tangent cone

$$T_K(y) = \{z \in C(T) \mid z(t) \leq 0, t \in I(y)\}, \tag{2.5}$$

and the normal cone

$$N_K(y) = \{\mu \in C(T)^* \mid \mathrm{supp}(\mu) \subset I(y), \mu \succeq 0\}. \tag{2.6}$$

Proposition 2.2.1. *For a sequence of convex sets A^k in a norm linear space X, converging to A, let $x^k \xrightarrow{A^k} x, y^k \in N_{A^k}(x^k), y^k \xrightarrow{w^*} y$ and $\{\|y^k\|\}$ be bounded. Then $y \in N_A(x)$.*

Proof. Assume the contrary, $y \notin N_A(x)$. Then there is $\epsilon > 0$ and $u \in A$ such that

$$\langle y, u - x\rangle = \epsilon.$$

Since $A^k \to A$, there is a sequence $u^k \in A^k$ converging to u. Then it is easy to see

$$\langle y^k, u^k - x^k\rangle = \langle y^k - y, u - x\rangle + \langle y^k, u^k - u + x - x^k\rangle + \langle y, u - x\rangle \geq \frac{\epsilon}{2}$$

for all k sufficiently large. This can't hold since $y^k \in N_{A^k}(x^k)$. □

Theorem 2.2.1 (subgradients of distant functions at out-of-set points [93, Theorem 1.99])**.** *For any $T \neq \emptyset$, a subset of a Banach space, and any $\bar{x} \notin T$,*

$$\hat{\partial}\mathrm{dist}(x, T) = \hat{N}_{T(\rho)}(\bar{x}) \cap \{x^* \in X^* \mid \|x^*\| = 1\}, \tag{2.7}$$

where $T(\rho) = \{x \in X \mid \mathrm{dist}(x, T) \leq \rho\}$ with $\rho = \mathrm{dist}(\bar{x}, T)$.

Theorem 2.2.2. *Let $\bar{x}$ be a local optimal solution of the SIP problem* (2.1)*. Suppose that $f(\cdot)$ and $g(\cdot, t), t \in T$ are continuously differentiable and that $\nabla g(x, t)$ is continuous in (x, t). Then there exist $\lambda_0 \geq 0$ and $\mu \in N_K(g(\bar{x}))$ with* $\max\{\lambda_0, \|\mu\|\} = 1$ *such that*

$$0 = \lambda_0 \nabla f(\bar{x}) + \langle \nabla g(\bar{x}, \cdot), \mu\rangle. \tag{2.8}$$

Proof. Define

$$\theta_\delta(x) = \mathrm{dist}((f(x), g(x)), C_{\delta, \bar{x}}), \tag{2.9}$$

where $g(x) := g(x,\cdot) \in C(T)$, $C_{\delta,\bar{x}} := (f(\bar{x}) - \delta + \mathbb{R}_-) \times K$ and $\delta > 0$. It is assumed that the norm chosen for $\mathbb{R} \times C(T)$ is such that $\|(r,0)\| = |r|$ (we take the max-norm). Then

$$\theta_\delta(\bar{x}) \leq \delta + \inf\ \theta_\delta(x). \tag{2.10}$$

Using Ekeland's variational principle, for each $\delta > 0$, there is an x_δ such that

$$\|\bar{x} - x_\delta\| \leq \sqrt{\delta}, \tag{2.11}$$

$$\theta_\delta(x) + \sqrt{\delta}\|x - x_\delta\| > \theta_\delta(x_\delta), \text{ for any } x \neq x_\delta. \tag{2.12}$$

The last inequality means that x_δ minimizes that function

$$\tilde{\theta}_\delta(x) := \theta_\delta(x) + \sqrt{\delta}\|x - x_\delta\|.$$

Note that $\theta_\delta(x_\delta) > 0$ otherwise it contradicts the optimum of $\bar{x}$ for problem (2.1) and so $(f(x_\delta), g(x_\delta)) \notin C_{\delta,\bar{x}}$. From the generalized Fermat rule for optimality: $0 \in \partial\tilde{\theta}_\delta(x_\delta)$, and in addition to some appropriate rules of subdifferential calculus (see, e.g. [23, Th3.2 of Chapter 2]), we have that there exists

$$(\lambda_\delta, \mu_\delta) \in N_{C_{\delta,\bar{x}} + \mathrm{dist}((f(x_\delta), g(x_\delta)), C_{\delta,\bar{x}})B}(x_\delta)$$

such that $\|(\lambda_\delta, \mu_\delta)\| = 1$ and

$$0 \in \lambda_\delta \nabla f(x_\delta) + \partial g(x_\delta)^* \mu_\delta + \sqrt{\delta} B, \tag{2.13}$$

where $\partial g(x_\delta) = \nabla_x g(x,\cdot) \in C(T)^n$. Applying the Alaoglu theorem, we may let (λ_0, μ) be a w^*-limit point of the sequence $\{(\lambda_{\delta^k}, \mu_{\delta^k})\}$ for some $\delta^k \to 0$. We also have that $\mu \in N_K(g(\bar{x}))$ by Proposition 2.2.1 and the fact that $C_{\delta,\bar{x}} + \mathrm{dist}((f(x_\delta), g(x_\delta)), C_{\delta,\bar{x}}) \to (f(\bar{x}) + \mathbb{R}_-) \times K$. Since $\nabla f(x)$ and $\nabla g(x,t)$ are continuous, then $\nabla f(x_{\delta^k}) \to \nabla f(x_{\bar{x}})$ and $\nabla g(x_{\delta^k}, t) \to \nabla g(\bar{x}, t)$ for each $t \in T$. Thus we complete the proof by letting $\delta \to 0$ in (2.13). □

A theorem from Rogosinski [117] is as follows.

Theorem 2.2.3 ([126, 117]). *Let T be a metric space equipped with its Borel sigma algebra $\mathcal{B}$, $q_i \colon T \to \mathbb{R}, i = 1, \cdots, k$, be measurable functions, and μ be a (nonnegative) measure on $(T, \mathcal{B})$ such that $q_1, \cdots, q_k$ are μ-integrable. Then there exists a (nonnegative) measure η on $(T, \mathcal{B})$ with a finite support of at most k points such that*

$$\int_T q_i \,\mathrm{d}\mu = \int_T q_i \,\mathrm{d}\eta, i = 1, \cdots, k.$$

Thus by applying Rogosinski's Theorem we obtain the following theorem.

Theorem 2.2.4. *Let the assumptions of Theorem 2.2.2 hold. Then there exist* $\lambda_0 \geq 0$ *and* $t_i \in I(g(\bar{x})), \lambda_i \geq 0, i = 1, \cdots, L(\leq n)$ *with* $\sum_{i=0}^{L} \lambda_i = 1$ *such that*

$$0 = \lambda_0 \nabla f(\bar{x}) + \sum_{i=1}^{L} \lambda_i \nabla g(\bar{x}, t_i). \tag{2.14}$$

2.3 Optimality with Penalty

In this section, we adapt an approach from Bertsekas and Ozdaglar [7], where the generalized Fritz-John optimality conditions are derived for nonlinear programming and subsequently new constraint qualifications are introduced and relations with other known qualifications are investigated. This approach is also adapted by Kanzow and Schwartz [72] to study mathematical programs with equilibrium constraints.

Let $\bar{x}$ be a local optimal solution of SIP problem (2.1) and associate it with a sequence of problems:

$$(P_k) \qquad \min_{x \in B(\bar{x}, \epsilon)} F_k(x) := f(x) + \frac{k}{2} \int_T g_+(x,t)^2 \, \mathrm{d}\mu(t) + \frac{1}{2}\|x - \bar{x}\|^2 \tag{2.15}$$

where μ is a nonnegative regular Borel measure on T. Let x^k sovle (P_k) (since the feasible set of problem (P_k) is compact, the solution exists). Then we have

$$F_k(x^k) \leq F_k(\bar{x}), \tag{2.16}$$

that is

$$f(x^k) + \frac{k}{2} \int_T g_+(x^k, t)^2 \, \mathrm{d}\mu(t) + \frac{1}{2}\|x^k - \bar{x}\|^2 \leq f(\bar{x}), \quad \text{for all } k. \tag{2.17}$$

As $\{x^k\}$ is bounded, we may also assume that $x^k \to x^*$, if needed by taking a subsequence, and thus

$$\int_T g_+(x^k, t)^2 \, \mathrm{d}\mu(t) \to 0.$$

Then by continuity, $\int_T g_+(x^*, t)^2 \, \mathrm{d}\mu(t) = 0$, and by the penality (under appropriate assumption): x^* is feasible. Then

$$f(x^*) + \frac{1}{2}\|x^* - \bar{x}\|^2 \leq f(\bar{x}), \tag{2.18}$$

$$f(\bar{x}) \leq f(x^*), \tag{2.19}$$

which imply that $x^* = \bar{x}$. By optimality for P_k at x^k and the calculus rule for the integral term [112], we have

$$0 = \nabla f(x^k) + k \int_T g_+(x^k, t) \nabla g(x^k, t) \, \mathrm{d}\mu(t) + (x^k - \bar{x}) \quad \text{for } k \text{ large enough.} \tag{2.20}$$

By Theorem 2.2.3, there exists nonnegative measure δ^k with a finite support of no larger than n points such that

$$\int_T g_+(x^k, t) \nabla g(x^k, t) \, \mathrm{d}\mu(t) = \int_T g_+(x^k, t) \nabla g(x^k, t) \, \mathrm{d}\delta^k(t). \tag{2.21}$$

As δ^k is of finite support, we assume that $\delta^k(t) = \sum_i \lambda_{ki} \chi_T^k(t_{ki})$ with $\mathrm{supp}\delta^k = \{t_{ki} : i\}$, $\lambda_{ki} > 0$, and $\chi_T^k(t) = 1$, for $t \in \mathrm{supp}\delta^k$, otherwise, 0, then

$$\int_T g_+(x^k, t) \nabla g(x^k, t) \, \mathrm{d}\mu(t) = \sum_i \lambda_{ki} g_+(x^k, t_{ki}) \nabla g(x^k, t_{ki}). \tag{2.22}$$

Let $\gamma^k = \sqrt{1 + k^2 \sum_i (\lambda_{ki} g_+(x^k, t_{ki}))^2}$, $\mu_{k0} = \frac{1}{\gamma^k}$, $\mu_{ki} = \frac{\lambda_{ki} g_+(x^k, t_{ki})}{\gamma^k}$. Then $\|\mu_k\| = 1$ for $\mu_k := (\mu_{ki})_{i=0}^{i_k}$, and

$$0 = \mu_{k0} \nabla f(x^k) + \mu_{ki} \nabla g(x^k, t_{ki}) + \frac{1}{\gamma^k}(x^k - \bar{x}). \tag{2.23}$$

Without loss of generality we may assume that $|\mathrm{supp}\delta^k| = L$, a constant no larger than n. By compactness, assume that

$$\mu_{k0} \to \lambda_0, \mu_{ki} \to \lambda_i, i = 1, \cdots, L, \tag{2.24}$$

$$t_{ki} \to \bar{t}_i, i = 1, \cdots, L. \tag{2.25}$$

Taking the limit in (2.23), we obtain

$$0 = \lambda_0 \nabla f(\bar{x}) + \sum_{i=1}^{L} \lambda_i \nabla g(\bar{x}, \bar{t}_i). \tag{2.26}$$

If $\lambda_i > 0$, then from the definition of μ_{ki} it follows that $\lambda_i g(x^k, t_{ki}) > 0$ for all k large enough. And we also have $f(x^k) < f(\bar{x})$.

Theorem 2.3.1 (Optimality)**.** *Let $\bar{x}$ be a local optimal solution of SIP problem* (2.1)*. Assume that the support of the measure μ is T, which implies $\int_T g_+(x, t) \, \mathrm{d}\mu(t)$ is a penalty term, see e.g.,* (2.34)*, then we have*

(i) *there are nonnegative real numbers $\lambda_0, \cdots, \lambda_L, L \le n$ not all zero, $\bar{t}_i \in T$ such that*

$$0 = \lambda_0 \nabla f(\bar{x}) + \sum_{i=1}^{L} \lambda_i \nabla g(\bar{x}, \bar{t}_i). \tag{2.27}$$

(ii) *if* $J := \{i : \lambda_i > 0, i \geq 1\} \neq \emptyset$*, there is* $\{x^k\}$*, converging to* $\bar{x}$*,* $\{t_{ki}\}$*, converging to* $\bar{t}_i$ *such that*

$$f(x^k) < f(\bar{x}), \lambda_i g(x^k, t_{ki}) > 0, i \in J. \tag{2.28}$$

Remark 2.3.1. *We can use the objective function*

$$F_k(x) := f(x) + \frac{k}{2} \max_{t \in T} [g_+(x,t)]^2 + \frac{1}{2} \|x - \bar{x}\|^2$$

along with the subdifferential calculus rule for max*-type functions to obtain the same results.*

Assumption 2.3.1. *Given* $\bar{x}$ *and any* $\bar{t}_i \in I(\bar{x})$*,* $i \leq L, L \leq n$*, there is no nonzero* $\lambda \in \mathbb{R}^L$*, no* $x^k \to \bar{x}, t_{ki} \to \bar{t}_i$*, such that*

(i) $\sum_{i=1}^{L} \lambda_i \nabla g(\bar{x}, \bar{t}_i) = 0, \lambda \geq 0,$

(ii) $\sum_{i=1}^{L} \lambda_i g(x^k, t_{ki}) > 0,$ *for all* k*.*

Theorem 2.3.2. *Let the assumptions of Theorem 2.3.1 and Assumption 2.3.1 hold. Then there are nonnegative real numbers* $\lambda_1, \cdots, \lambda_n$ *not all zero, and* $\bar{t}_i \in T$ *such that*

$$0 = \nabla f(\bar{x}) + \sum_{i=1}^{n} \lambda_i \nabla g(\bar{x}, \bar{t}_i). \tag{2.29}$$

2.4 Lower Order Penalization

In this section we study optimality conditions of SIP via exact penalty functions. We will investigate relations between the lower order exact penalty function and the calmness condition and consequently apply the lower order exact penalty functions to derive optimality conditions. More specifically, we will propose two types of pth-order ($0 < p \leq 1$) penalty functions, the max-type and integral-type, generalizations of the l_∞ and l_1 penalty functions in nonlinear programming problems, respectively. We will then study the relationship of exactness of the penalty functions and the equivalence between the exactness of penalty functions and the corresponding concept of calmnesses. Under the assumption of exactness of the penalty functions, the first-order optimality conditions for SIP are derived. For cases $p = 1$ and $p < 1$, we adopt different penalty functions. To a degree, we prefer the stronger one, integral-type penalty functions, to derive optimality conditions. It is also shown that, playing the roles of constraint qualifications, exact penalty function approaches are different from the traditional ones.

The penalty function methods for nonlinear programming problems have been widely investigated, see [29, 147, 104, 114, 6, 13, 118, 144]. An integral-type l_1 penalty function for SIP was proposed by Pietrzykowski [105], but not always exact. Improved by Conn and Gould [24], an exact penalty function, extending the l_1 penalty function of the nonlinear programming problem, was devised. Algorithms for SIP based on exact penalty functions was developed by Coope and Price [25]. Auslender et al. [3] considered the convex SIP, developed Remez-type and integral-tpye algorithms counpled with penalty and smoothing methods and proved the convergence of primal and dual sequences.

2.4.1 Penalty Functions and Calmness

Consider the following semi-infinite programming problems:

$$\min \ f(x) \quad \text{s.t.} \quad g(x,t) \le 0, \quad t \in T, \tag{2.30}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \times T \to \mathbb{R}$ are continuously differentiable real-valued functions, and T is a compact Hausdorff space.

Let $M = \{x \in R^n \mid g(x,t) \le 0, t \in T\}$ be the feasible region and

$$T(x) = \arg\max_{t\in T} g(x,t) = \{t \in T \mid g(x,t) = \max_{t\in T} g(x,t)\}$$

be the index set of active constraints of SIP. A penalty function $F\colon \mathbb{R}^n \to \mathbb{R}$ is of the following form

$$F(x) = f(x) + \rho P(x),$$

where $\rho > 0$ is a penalty parameter and $P\colon \mathbb{R}^n \to \mathbb{R}_+$ is a penalty term satisfying

$$P(x) = 0 \Leftrightarrow x \in M.$$

Let $\bar{x}$ be a local optimal solution of SIP (2.30). F is said to be locally exact at $\bar{x}$ if there exists $\bar{\rho} > 0$ such that $\bar{x}$ is a local optimal solution of F for any $\rho \ge \bar{\rho}$.

Let $0 < p \le 1$ and $[a]_+ = \max\{a, 0\}$. We consider two types of pth-order penalty functions, i.e., max-type penalty and integral-type penalty functions. A pth-order max-type penalty function for SIP is defined as,

$$F^p_{max}(x) = f(x) + \rho \max_{t\in T} g^p_+(x,t), \tag{2.31}$$

where $g_+^p(x,t) = \big(\max\{g(x,t),0\}\big)^p$.

Let μ be a nonnegative regular Borel measure defined on T with the support of μ being equal to T, that is $\text{supp}(\mu) = T$, where the support of μ is defined as the set of the points $t \in T$ such that any open neighborhood V of t has a positive measure:

$$\text{supp}\,(\mu) := \{t \in T \mid \mu(V) > 0, \text{ for any open neighborhood } V \text{ of } t\}.$$

Two pth-order integral-type penalty functions for SIP are defined respectively by

$$F_{int}^p(x) = f(x) + \rho \int_T g_+^p(x,t)\,\mathrm{d}\mu(t), \tag{2.32}$$

$$\bar{F}_{int}^p(x) = f(x) + \rho(\int_T g_+(x,t)\,\mathrm{d}\mu(t))^p. \tag{2.33}$$

With the assumption $\text{supp}(\mu) = T$, $\int_T g_+^p(x,t)\,\mathrm{d}\mu(t)$ and $(\int_T g_+(x,t)\,\mathrm{d}t)^p$ are penalty terms, i.e.,

$$\int_T g_+^p(x,t)\,\mathrm{d}\mu(t) = 0 \Longleftrightarrow g(x,t) \le 0, \quad \forall t \in T, \tag{2.34}$$

$$\int_T g_+(x,t)\,\mathrm{d}\mu(t) = 0 \Longleftrightarrow g(x,t) \le 0, \quad \forall t \in T.$$

It is obvious that the second equivalence follows from (2.34). Thus we show only the necessity of (2.34) since the sufficiency is trivial. Suppose that there exists $t_0 \in T$ such that $g(x,t_0) > 0$. The continuity of g ensures the existence of a neighborhood V of t_0 and $\alpha > 0$ such that $g(x,t) > \alpha$ for all $t \in V$. Thus,

$$\int_T g_+^p(x,t)\,\mathrm{d}\mu(t) \ge \int_V g_+^p(x,t)\,\mathrm{d}\mu(t) \ge \alpha^p \mu(V) > 0,$$

where the last step comes from the fact that $T = \text{supp}(\mu)$. This leads to a contradiction.

It is clear from the definition that if F_{max}^p (resp. F_{int}^p and $\bar{F}_{int}^p$) is locally exact at x^*, then so is $F_{max}^{\tilde{p}}$ (resp. $F_{int}^{\tilde{p}}$ and $\bar{F}_{int}^{\tilde{p}}$) for all $\tilde{p} \in (0,p)$.

If the set T is finite, say $T = \{t_1, \cdots, t_s\}$, then the feasible set M is defined by a finite number of inequalities and hence SIP becomes a standard nonlinear programming problem. Correspondingly, taking a measure $\mu = \sum_{i=1}^s \delta_{t_i}$, where δ_{t_i} is the Dirac measure defined on T, the pth-order penalty functions (2.31), (2.32) and (2.33) take the form of

$$F^p_{max}(x) = f(x) + \rho \max_{1\le i\le s} g^p_+(x, t_i), \tag{2.35}$$

$$F^p_{int}(x) = f(x) + \rho \sum_{i=1}^{s} g^p_+(x, t_i), \tag{2.36}$$

$$\bar{F}^p_{int}(x) = f(x) + \rho(\sum_{i=1}^{s} g_+(x, t_i))^p, \tag{2.37}$$

respectively, where (2.36) and (2.37) are the pth-order penalty functions for a nonlinear programming problem considered in Meng and Yang [92] and Yang and Meng [144]. It is easy to see that the exactness of all three penalty functions (2.35), (2.36) and (2.37) are equivalent.

Let T be a compact subset of $\mathbb{R}^m$ and μ be the Lebesgue measure. If $T = \mathrm{cl}(\mathrm{int}\, T)$, then $\mathrm{supp}(\mu) = T$. Here the 'support' is understood as the support of the Lebesgue measure with respect to T:

$$\mathrm{supp}(\mu) = \{t \in T \mid \int_{B(t,\delta)\cap T} \mathrm{d}\mu(t) > 0 \text{ for all } \delta > 0\}.$$

Especially, if T is a convex set with nonempty interior points, then $\mathrm{supp}(\mu) = T$, see, e.g., [116, Theorem 2.33].

Throughout the chapter we will make the assumptions that $T = \mathrm{supp}(\mu)$ whenever the integeral-type penalty function is dealt with and that in all examples T is assumed to be a compact subset of $\mathbb{R}^m$ and μ, for simplicity, is assumed to be the Lebesgue measure.

Note that for $p \in (0, 1)$, the following inequality

$$\int_T g_+(x,t)^p \,\mathrm{d}\mu(t) \le \mu(T)^{1-p}(\int_T g_+(x,t)\,\mathrm{d}\mu(t))^p$$

holds by Hölder's inequality, see [85, Theorem 5.1]. Thus the exactness of F^p_{int} implies that of $\bar{F}^p_{int}$. We give the following example to show that the exactness of penalty function $\bar{F}^p_{int}$ is indeed strictly weaker than that of F^p_{int}.

Example 2.4.1. *Consider SIP problem*

$$\min f(x) \quad \text{s.t.} \quad g(x,t) \le 0, t \in T,$$

where $f(x) = \begin{cases} x^2, & \text{if } x \geq 0, \\ -x^{\frac{5}{3}}, & \text{otherwise,} \end{cases}$ $g(x,t) = 2tx - t^2$ *and* $T = [-1, 1]$. *Then for* $x \leq 0$ *sufficiently small,* $\int_T g_+(x,t)^{1/2}\,\mathrm{d}t = \int_{2x}^{0}(2tx - t^2)^{1/2}\,\mathrm{d}t = \int_{-1}^{1} x^2(1-s^2)^{1/2}\,\mathrm{d}s = \frac{\pi}{2}x^2$, *where* $t = (1-s)x$, *and* $(\int_T g_+(x,t)\,\mathrm{d}t)^{1/2} = (-\frac{4}{3}x^3)^{1/2}$. *Thus it follows that* $\bar{F}_{int}^{1/2}$ *is exact at* $x^* = 0$ *and* $F_{int}^{1/2}$ *is not.*

Next, we explore the relationship between the two exact penalty functions F_{max}^p and F_{int}^p. First, let us introduce, for $x \in \mathbb{R}^n$,

$$A(x) = \arg\max_{t \in T} g_+(x,t) = \{t \in T \mid g_+(x,t) = \max_{t \in T} g_+(x,t)\}.$$

Theorem 2.4.1. *Let* $0 < p \leq 1$ *and* $supp(\mu) = T$. *We have*

(a) *if* F_{int}^p *is a local exact penalty function at* $\bar{x}$, *so is* F_{max}^p;

(b) *if* F_{max}^p *is a local exact penalty function at* $\bar{x}$ *and*

$$\liminf_{x \to \bar{x}} \mu(A(x)) > 0, \tag{2.38}$$

so is F_{int}^p.

Proof. (a) It is easy to see that $g_+^p(x,t) \leq \left(\max_{t \in T} g_+(x,t)\right)^p = \max_{t \in T} g_+^p(x,t), \forall x \in \mathbb{R}^n$. So,

$$\int_T g_+^p(x,t)\,\mathrm{d}\mu(t) \leq \int_T \max_{t \in T} g_+^p(x,t)\,\mathrm{d}\mu(t) = \mu(T)\max_{t \in T} g_+^p(x,t),$$

where $\mu(T) > 0$, since $\mathrm{supp}(\mu) = T$. By assumption that F_{int}^p is a local exact penalty function at $\bar{x}$, there exists $\bar{\rho} > 0$ such that for every $\rho \geq \bar{\rho}$ there exists $\delta_\rho > 0$ such that

$$F_{int}^p(x) = f(x) + \rho \int_T g_+^p(x,t)\,\mathrm{d}\mu(t) \geq f(\bar{x}), \quad \text{for all } x \in B(\bar{x}, \delta_\rho). \tag{2.39}$$

Let $\hat{\rho} = \mu(T)\bar{\rho}$. For each $\rho \geq \hat{\rho}$, we have

$$F_{max}^p(x) = f(x) + \rho \max_{t \in T} g_+^p(x,t) \geq f(x) + \frac{\rho}{\mu(T)} \int_T g_+^p(x,t)\,\mathrm{d}\mu(t). \tag{2.40}$$

Note that $\frac{\rho}{\mu(T)} \geq \bar{\rho}$. It follows from (2.39) and (2.40) that

$$F_{max}^p(x) \geq f(x) + \frac{\rho}{\mu(T)} \int_T g_+^p(x,t)\,\mathrm{d}\mu(t) \geq f(\bar{x}), \quad \text{for all } x \in B(\bar{x}, \delta_{\rho/\mu(T)}).$$

This means that F_{max}^p is a local exact penalty function at $\bar{x}$.

(b) The condition (2.38) is equivalent to that there exist $\varepsilon > 0$ and $\delta_1 > 0$ such that

$$\mu(A(x)) \geq \varepsilon, \qquad \text{for all } x \in B(\bar{x}, \delta_1).$$

Thus, for any $x \in B(\bar{x}, \delta_1)$, we have

$$\int_T g_+^p(x,t)\,\mathrm{d}\mu(t) \geq \int_{A(x)} g_+^p(x,t)\,\mathrm{d}\mu(t) = \mu(A(x)) \max_{t\in T} g_+^p(x,t) \geq \varepsilon \max_{t\in T} g_+^p(x,t). \quad (2.41)$$

Since F_{max}^p is a local exact penalty function, there exists $\bar{\rho} > 0$ such that $\bar{x}$ is also a local minimum of F_{max}^p for all $\rho \geq \bar{\rho}$; that is, for every $\rho \geq \bar{\rho}$, there exist $\delta_\rho > 0$ (with $\delta_\rho < \delta_1$) such that

$$F_{max}^p(x) = f(x) + \rho \max_{t\in T} g_+^p(x,t) \geq f(\bar{x}), \qquad \text{for all } x \in B(\bar{x}, \delta_\rho). \qquad (2.42)$$

Let $\hat{\rho} = \frac{\bar{\rho}}{\varepsilon}$. For any $\rho \geq \hat{\rho}$ (implying $\rho\varepsilon \geq \bar{\rho}$), it follows from (2.41) and (2.42) that for all $x \in B(\bar{x}, \delta_{\rho\varepsilon})$,

$$F_{int}^p(x) = f(x) + \rho \int_T g_+^p(x,t)\,\mathrm{d}\mu(t) \geq f(x) + \rho\varepsilon \max_{t\in T} g_+^p(x,t) \geq f(\bar{x}).$$

That is, F_{int}^p is a local exact penalty function at $\bar{x}$. □

A well-known result for linear programming is that the l_1 penalty function is always exact (so is for the lower-order penalty function). A natural question arises: whether this result remains true for the linear SIP. The answer is negative. The following example illustrates that for any $0 < p \leq 1$, there always exists a linear SIP such that the penalty function $F_{\max}^p$ is not exact.

Example 2.4.2. *Let $0 < p \leq 1$ and $q = \frac{1}{p} \geq 1$. Consider a linear SIP of the form*

$$\min x \quad \text{s.t.} \quad (2q)t^{2q-1}x - (2q-1)t^{2q} \leq 0, \quad t \in [-1,1],$$

with the local optimal solution $\bar{x} = 0$. The p-order max-type penalty function can be rewritten as

$$F_{max}^p(x) = f(x) + \rho[\big(\max_{t\in T} g(x,t)\big)_+]^p.$$

For any $x \in [-1,1]$, we have

$$\max_{t\in T} g(x,t) = \max_{t\in[-1,1]} (2q)t^{2q-1}x - (2q-1)t^{2q} = x^{2q} \geq 0, \qquad (2.43)$$

where the second equality follows from the fact that

$$\nabla_t g(x,t) = 0 \Rightarrow 2q(2q-1)t^{2q-2}x - (2q-1)2qt^{2q-1} \begin{cases} < 0, & t > x, \\ = 0, & t = x, \\ > 0, & t < x, \end{cases}$$

so the maximum in (2.43) is attained at $t = x$.

For any fixed $\rho > 0$*, we have*

$$F^p_{max}(x) = f(x) + \rho[\big(\max_{t\in T} g(x,t)\big)_+]^p = x + \rho x^{2pq} = (1+\rho x)x < 0, \tag{2.44}$$

when $x < 0$ *is sufficiently close to* 0*. This implies that* F^p_{max} *is not a local exact penalty function at* $\bar{x} = 0$*. Neither is* F^p_{int}*, by Theorem 2.4.1 (a).*

Clearly,

$$A(x) = \begin{cases} T(x), & \text{if } x \notin M, \\ T, & \text{otherwise.} \end{cases}$$

Since $\mu(T) > 0$ by the assumption $\text{supp}(\mu) = T$, the condition (2.38) can be rewritten as

$$\liminf_{x\to\bar{x}} \mu(A(x)) = \liminf_{\substack{x\to\bar{x}\\ x\notin M}} \mu(T(x)) > 0.$$

The following example shows that the condition (2.38) is not necessary.

Example 2.4.3. *Consider the following linear SIP problem*

$$\min x_1 \quad \text{s.t.} \quad tx_1 + t^3 x_2 \le 0, \quad t \in [-1,1],$$

with $M = \{(0,0)\}$*. The point* $\bar{x} = (0,0)$ *is the optimal solution. It is obvious that* $\liminf_{x\to\bar{x}} \mu(A(x)) = 0$*. Next, we check the exactness of integral-type penalty function*

$$F^1_{int}(x_1,x_2) := x_1 + \rho\int_{-1}^{1} [tx_1 + t^3x_2]_+ \,\mathrm{d}t.$$

It suffices to consider the area $\{(x_1,x_2)|x_1 < 0\}$*. For* $x_1 < 0, 0 < -\frac{x_1}{x_2} \le 1$,

$$F^1_{int}(x_1,x_2) = x_1 + \rho\Big(\int_{-\sqrt{-\frac{x_1}{x_2}}}^{0} + \int_{\sqrt{-\frac{x_1}{x_2}}}^{1}\Big) tx_1 + t^3x_2 \,\mathrm{d}t = x_1\Big[1 - \frac{\rho}{2}\Big(-1 - \frac{x_2}{2x_1} - \frac{x_1}{x_2}\Big)\Big].$$

Since $\frac{1}{2s} + s \geq \sqrt{2}$, for $s \in (0,1]$, $F^1_{int}(x_1, x_2) \geq 0$ for ρ large enough. For $x_1 < 0, 1 < -\frac{x_1}{x_2}$, or $x_1 < 0, x_2 < 0$,

$$F^1_{int}(x_1, x_2) = x_1 + \rho \int_{-1}^{0} tx_1 + t^3 x_2 \,\mathrm{d}t = x_1[1 - \rho(\frac{1}{2} - \frac{1}{4}(-\frac{x_2}{x_1}))].$$

In both above cases, for some $\rho > 0$, $F^1_{int} \geq 0$ holds. Thus F^1_{int} is locally exact at $\bar{x}$.

Let $C(T)$ denote the space of all continuous functions u on T equipped with the max-norm $\|u\|_\infty = \max_{t \in T} |u(t)|$.

It is well known that the exact penalty function has close relationships with the concept of calmness. Corresponding to the two penalty functions F^p_{max} and F^p_{int}, we introduce the following two kinds of calmness, respectively.

For $u \in C(T)$, let $M(u) = \{x \in R^n \mid g(x,t) \leq u(t), t \in T\}$.

Definition 2.4.1. *Let $0 < p \leq 1$ and $\bar{x}$ be a local optimal solution of SIP.*

(i) *We say that SIP is pth-order max-type calm at $\bar{x}$, if there exists a positive scalar ρ such that, for any $u_k \in C(T)$ with $0 \neq u_k \to 0$ under the max-norm and any $x_k \in M(u_k)$ with $x_k \to \bar{x}$, the following relation holds:*

$$\frac{f(x_k) - f(\bar{x})}{\|u_k\|^p_\infty} + \rho \geq 0, \quad \textit{for all } k.$$

(ii) *We say that SIP is pth-order integral-type calm at $\bar{x}$, if there exists a positive scalar ρ such that, for any $u_k \in C(T)$ with $0 \neq |u_k|^p_{int} \to 0$ and any $x_k \in M(u_k)$ with $x_k \to \bar{x}$, the following relation holds:*

$$\frac{f(x_k) - f(\bar{x})}{|u_k|^p_{int}} + \rho \geq 0, \quad \textit{for all } k, \tag{2.45}$$

where $|u|^p_{int} = \int_T |u(t)|^p \,\mathrm{d}\mu(t)$.

We assert that local exact penalization and calmness are equivalent under a mild assumption.

Theorem 2.4.2. *Let $0 < p \leq 1$ and $\bar{x}$ be a local optimal solution of SIP. The following statements are equivalent:*

(a) *SIP is pth-order max-type calm at $\bar{x}$;*

(b) *F^p_{max} is a local exact penalty function at $\bar{x}$.*

Moreover, if supp(μ) $= T$, then the following statements are equivalent:

(c) *SIP is pth-order integral-type calm at $\bar{x}$;*

(d) *F^p_{int} is a local exact penalty function at $\bar{x}$.*

Proof. $(a) \Rightarrow (b)$. Loss of local exactness at $\bar{x}$ is to say that for any $\rho^k \to \infty$, there is x_k converging to $\bar{x}$ such that

$$f(x_k) + \rho^k \max_{t \in T} g^p_+(x_k, t) < f(\bar{x}) + \rho^k \max_{t \in T} g^p_+(\bar{x}, t) = f(\bar{x}).$$

This inequality implies the infeasibility of x_k, $\max_{t \in T} g^p_+(x_k, t) > 0$ and $\max_{t \in T} g^p_+(x_k, t) \to 0$ as $k \to \infty$. By choosing $u_k(t) = g_+(x_k, t)$, it is easy to check the failure of pth-order max-type calmness at $\bar{x}$.

$(b) \Rightarrow (a)$. Suppose that SIP is not pth-order max-type calm at $\bar{x}$. Then there exist sequences of $\{\rho^k\} \subset \mathbb{R}$, $\{u_k\} \subset C(T)$ and $\{x_k\} \subset \mathbb{R}^n$ with $\rho^k \to \infty$, $0 \neq u_k \to 0$, $x_k \in M(u_k)$ and $x_k \to \bar{x}$, such that

$$\frac{f(x_k) - f(\bar{x})}{\|u_k\|^p_\infty} + \rho^k < 0, \quad \text{for all } k.$$

That is

$$f(x_k) + \rho^k \|u_k\|^p_\infty < f(\bar{x}), \ \text{ for all } k.$$

Then

$$F^p_{max}(x^k) = f(x_k) + \bar{\rho} \max_{t \in T} g^p_+(x_k, t) \leq f(x_k) + \bar{\rho} \|u_k\|^p_\infty < f(\bar{x}) = F^p_{max}(\bar{x}), \ \text{ for all } k,$$

where $\bar{\rho}$ is the constant given in the definition of the local exactness of F^p_{max}. This contradicts the local exactness of F^p_{max} at $\bar{x}$. Thus SIP is pth-order max-type calm at $\bar{x}$.

$(c) \Rightarrow (d)$. Suppose to the contrary that there exist sequences $\{\rho_k\}$ and $\{x_k\}$ satisfying $\rho_k \to \infty$ and $x_k \to \bar{x}$ as $k \to \infty$ such that

$$f(x_k) + \rho_k \int_T \max\{g(x_k, t), 0\}^p \, \mathrm{d}\mu(t) < f(\bar{x}), \quad \text{for sufficiently large } k. \tag{2.46}$$

Let $u_k(t) = \max\{g(x_k, t), 0\}$. So, $u_k \in C(T)$, since g is continuous. Thus the above inequality yields

$$f(x_k) + \rho_k |u_k|^p_{int} = f(x_k) + \rho_k \int_T u_k(t)^p \, \mathrm{d}\mu(t) < f(\bar{x}). \tag{2.47}$$

It follows from (2.46) that, for sufficiently large k, x_k is infeasible, and hence $|u_k|_{int}^p \neq 0$, due to the fact $\text{supp}(\mu) = T$. Noting that when $\rho_k \to \infty$, $x_k \to \bar{x}$, and (2.47), we get $|u_k|_{int}^p \to 0$. Therefore,

$$\frac{f(x_k) - f(\bar{x})}{|u_k|_{int}^p} + \rho_k < 0, \quad \text{for sufficiently large } k,$$

contradicting the pth-order integral-type calmness of SIP at $\bar{x}$.

$(d) \Rightarrow (c)$. Suppose that SIP is not pth-order integral-type calm at $\bar{x}$. Then there exist sequences $\rho_k \to +\infty$, $u_k \in C(T)$ with $0 \neq |u_k|_{int}^p \to 0$, and $\{x_k\}$ with $x_k \to \bar{x}$ satisfying $g(x_k, t) \leq u_k(t)$ for all $t \in T$ such that

$$\frac{f(x_k) - f(\bar{x})}{|u_k|_{int}^p} + \rho_k < 0, \quad \text{for sufficiently large } k$$

or equivalently,

$$f(x_k) - f(\bar{x}) + \rho_k \int_T |u_k(t)|^p \,\mathrm{d}\mu(t) < 0.$$

Combining this with the fact that $g(x_k, t) \leq u_k(t)$ for $t \in T$ yields $\max\{g(x_k, t), 0\} \leq |u_k(t)|$ we obtain

$$f(x_k) - f(\bar{x}) + \rho_k \int_T |g(x_k, t)|^p \,\mathrm{d}\mu(t) < 0, \quad \text{for sufficiently large } k.$$

This contradicts the local exactness of F_{int}^p, since x_k converges to $\bar{x}$. □

Next, we discuss that how different perturbations will affect the calmness of the SIP problem. Let $u \in C(T)$ and $\tilde{u} \in \mathbb{R}$. Consider the following perturbation problems of SIP

$$(P_u) \qquad \min f(x) \quad \text{s.t.} \quad g(x, t) \leq u(t), \quad t \in T,$$

and

$$(P_{\tilde{u}}) \qquad \min\ f(x) \quad \text{s.t.} \quad g(x, t) \leq \tilde{u}, \quad t \in T.$$

The equivalence between the max-type calmness of problems P_u and $P_{\tilde{u}}$ can be established via the problem

$$(P_{\|u\|_\infty}) \qquad \min\ f(x) \quad \text{s.t.} \quad g(x, t) \leq \|u\|_\infty, \quad t \in T.$$

That calmness of P_u implies that of $P_{\tilde{u}}$ is obvious since the $\tilde{u}$ as a constant function belongs to $C(T)$. For the reverse implication, with the formulation of $P_{\|u\|_\infty}$, it follows from the definition of calmness. Besides, it is easy to see that the max-type calmness of $P_{\tilde{u}}$ reduces to the calmness of the following nonlinear program with one inequality constraint:

$$\min\ f(x) \quad \text{s.t.} \quad \phi(x) \leq \tilde{u}, \quad t \in T,$$

where $\phi(x) := \max_{t\in T} g(x,t)$.

The following example shows that only considering the right-hand-side constant perturbation is not enough to ensure the integral-type calmness.

Example 2.4.4. *Consider the semi-infinite programming*

$$\min -x \quad s.t. \quad t^2 x \leq t^3, t \in [0,1].$$

Then the feasible set $M = (-\infty, 0]$ and the optimal solution $\bar{x} = 0$. It is easy to compute that for $x > 0$ near 0, we have

$$F_{max}^{\frac{1}{3}} = -x + \rho \max[t^2 x - t^3]_+^{\frac{1}{3}} = -x + \frac{4\rho}{27} x$$

and

$$\bar{F}_{int}^{\frac{1}{3}} = -x + \rho\left(\int_0^1 [t^2 x - t^3]_+ \mathrm{d}t\right)^{\frac{1}{3}} = -x + \frac{\rho}{12} x^{\frac{4}{3}}.$$

That is $F_{max}^{\frac{1}{3}}$ is locally exact but $\bar{F}_{int}^{\frac{1}{3}}$ is not exact. Note that exactness of $F_{int}^{\frac{1}{3}}$ implies that of $\bar{F}_{int}^{\frac{1}{3}}$. So $F_{int}^{\frac{1}{3}}$ is not exact.

From the equivalence between exactness of penalty functions and calmness and the discussion above, we have that the problem is calm under the constant perturbation

$$\min -x \quad s.t. \quad t^2 x \leq t^3 + \alpha, t \in [0,1],$$

where α is a constant. However, it is not calm under the functional perturbation.

Hence, the functional perturbation and constant perturbation are essentially different for establishing the exactness of integral-type penalty function.

2.4.2 Optimality Conditions

In this section, we shall employ the local exactness of F_{max}^p, F_{int}^p and $\bar{F}_{int}^p (0 < p \leq 1)$ to develop first-order optimality conditions of SIP. Two cases are treated separately according to the different value p. We consider first the case when $p = 1$.

Theorem 2.4.3. *If F^1_{max} is locally exact at a local solution $\bar{x}$, then $\bar{x}$ is a KKT point.*

Proof. If F^1_{max} is exact at $\bar{x}$, then there exist $\delta > 0$ and $\bar{\rho} > 0$ such that

$$F^1_{max}(x) = f(x) + \rho \max_{t \in T} g_+(x,t) \geq f(\bar{x}), \quad \text{for all } x \in B(\bar{x}, \delta), \rho \geq \bar{\rho},$$

which is equivalent to

$$F^1_{max}(x) = f(x) + \rho \max_{t \in T^*} g(x,t) \geq f(\bar{x}), \quad \text{for all } x \in B(\bar{x}, \delta), \rho \geq \bar{\rho}, \tag{2.48}$$

where t^* is any given point not belonging to T, $T^* := T \cup \{t^*\}$ and $g(x, t^*) \equiv 0$. By Fermat's rule,

$$0 \in \nabla f(\bar{x}) + \rho \operatorname{conv}\{\nabla_x g(\bar{x}, t) \mid t \in T^*(\bar{x})\}, \quad \text{for any } \rho \geq \bar{\rho},$$

with $T^*(\bar{x}) = \{t \in T^* \mid g(\bar{x}, t) = \max_{t \in T^*} g(\bar{x}, t)\}$. By Carathéodory's theorem, there are numbers $\lambda_i \geq 0, \bar{t}_i \in T^*(\bar{x}), i = 1, \cdots, n+1$ such that $\sum_{i=1}^{n+1} \lambda_i \leq 1$ and

$$\nabla f(\bar{x}) + \sum_{i=1}^{n+1} \rho \lambda_i \nabla g(\bar{x}, \bar{t}_i) = 0.$$

This means that $\bar{x}$ is a KKT point. □

Below we give an example for which F^1_{max} is exact but F^1_{int} is not.

Example 2.4.5. *Consider the SIP problem with $f(x) = -x_1 - x_2, g(x,t) = tx_1 + t^2 x_2, T = [0,1]$. Then the feasible set $M = \{x \in \mathbb{R}^2 \mid x_1 + x_2 \leq 0, x_1 \leq 0\}$ and thus $x^* = (0,0)$ is the optimal solution. We also have that*

$$\max_{t \in [0,1]} g_+(x,t) = \begin{cases} x_1 + x_2, & \textit{if } x_1 + x_2 \geq 0, x_1 + 2x_2 \geq 0, \\ 0, & \textit{if } x_1 + x_2 \leq 0, x_1 \leq 0, \\ -\frac{x_1^2}{4x_2}, & \textit{otherwise}. \end{cases}$$

Noting that $-x_1 - x_2 - \frac{x_1^2}{4x_2} = -\frac{(2x_2 + x_1)^2}{4x_2}$, it is easy to see that for $\rho \geq 1$, F^1_{max} is exact at $\bar{x}$. $\bar{x}$ is a KKT point as well: by taking $\lambda = 1, t = 1$, $(-1,-1) + 1 \cdot (1,1) = 0$. However, for $x \in \{x | x_1 \leq 0, x_1 + x_2 \geq 0\}$, $F^1_{int}(x) = -x_1 - x_2 + \rho[\frac{1}{2}x_1 + \frac{1}{3}x_2 - \frac{1}{6}\frac{x_1^3}{x_2^2}] = \frac{\rho}{6}[x_1 - \frac{x_1^3}{x_2^2}] + (\frac{\rho}{3} - 1)[x_1 + x_2] = \frac{(x_1 + x_2)}{6}[(2\rho - 6) + \rho \frac{x_1}{x_2} - \rho \frac{x_1^2}{x_2^2}]$. So, by taking $x = (-s + s^2, s), s \downarrow 0$, we know that for ρ large enough $F^1_{int}(x) < 0$ and thus F^1_{int} is not exact at $\bar{x}$.

Compared with $p = 1$, the case of $p \in (0, 1)$ needs more effort. First, we briefly describe some notation that will be used in the sequel. Given a continuous function $h : \mathbb{R}^n \to \mathbb{R}$, upper Dini-directional derivative of h at a point x in the direction $d \in \mathbb{R}^n$ is defined by

$$D_+h(x; d) = \limsup_{\lambda \downarrow 0} \frac{h(x + \lambda d) - h(x)}{\lambda}.$$

The generalized lower and upper second-order directional derivatives of a $C^{1,1}$ function h at x in the direction $d \in \mathbb{R}^n$ are defined by

$$h_{\circ\circ}(x; d) = \liminf_{y \to x, \lambda \downarrow 0} \frac{\langle \nabla h(y + \lambda d), d\rangle - \langle \nabla h(y), d\rangle}{\lambda},$$
$$h^{\circ\circ}(x; d) = \limsup_{y \to x, \lambda \downarrow 0} \frac{\langle \nabla h(y + \lambda d), d\rangle - \langle \nabla h(y), d\rangle}{\lambda}.$$

Let

$$\begin{aligned} L_M(x) &= \{d \in R^n \mid \langle \nabla_x g(x, t), d\rangle \le 0, t \in T(x)\}, \\ T^{=}(x, d) &= \{t \in T(x) \mid \langle \nabla_x g(x, t), d\rangle = 0\}, \\ T^{<}(x, d) &= \{t \in T(x) \mid \langle \nabla_x g(x, t), d\rangle < 0\}. \end{aligned}$$

Before stating the main results, we present two lemmas for reference, one of which is compiled from [144] for the estimates of the directional derivative $D_+g_+^p(x, t; \cdot)$ at $x \in \mathbb{R}^n$.

Lemma 2.4.1 (Fatou's Lemma, [85, Corollary 5.7])**.** *Let $(X, \mathfrak{M}, \mu)$ be a measured space and $f_k\colon X \to [0, +\infty)$ be integrable for each k. Assume that $\liminf_{k\to\infty} \|f_k\|_1$ exists. Then $\int_X \liminf_{k\to\infty} f_k(t)\, d\mu(t) \le \liminf_{k\to\infty} \int_X f_k(t)\, d\mu(t)$.*

Lemma 2.4.2 ([144])**.** *Let $\bar{h}(x) = (\max\{h(x), 0\})^p$ with $0 < p < 1$ and h be continuously differentiable at x.*

(a) *If $h(x) < 0$, then $D_+\bar{h}(x; d) = 0$;*
(b) *If $h(x) = 0$ and $\langle \nabla h(x), d\rangle < 0$, then $D_+\bar{h}(x; d) = 0$;*
(c) *If $p = 0.5$, $h(x) = 0$ and $\langle \nabla h(x), d\rangle = 0$, then $D_+\bar{h}(x; d) \le \sqrt{\max\{\frac{1}{2}h^{\circ\circ}(x; d), 0\}}$;*
(d) *If $0.5 < p < 1$, $h(x) = 0$, $\langle \nabla h(x), d\rangle = 0$ and $h^{\circ\circ}(x; d)$ is finite, then $D_+\bar{h}(x; d) = 0$;*
(e) *If $0 < p < 0.5$, $h(x) = 0$, $\langle \nabla h(x), d\rangle = 0$ and $h^{\circ\circ}(x; d) < 0$, then $D_+\bar{h}(x; d) = 0$.*

Now we establish a necessary optimality condition for SIP by virtue of the exact penalty function F^p_{int}.

Theorem 2.4.4. *Let $p \in (0,1)$ and F^p_{int} be locally exact at $\bar{x}$. Under any one of the three assumptions below,*
(i) $p = 0.5$ *and* $g^{\circ\circ}(\bar{x}, t; d) \leq 0$ *for all* $t \in T^{=}(\bar{x}, d)$ *and* $d \in L_M(\bar{x})$,
(ii) $0.5 < p < 1$ *and* $g(\cdot, t)$ *is* $C^{1,1}$, *for all* $t \in T^{=}(\bar{x}, d)$, *and*
(iii) $0 < p < 0.5$ *and* $g^{\circ\circ}(\bar{x}, t; d) < 0$, *for all* $t \in T^{=}(\bar{x}, d)$ *and* $0 \neq d \in L_M(\bar{x})$, *we have*

$$\langle \nabla f(\bar{x}), d \rangle \geq 0, \text{ for all } d \in L_M(\bar{x}). \tag{2.49}$$

Proof. For $0 < p < 1$, and $0 \neq d \in L_M(\bar{x})$, we have

$$\begin{aligned} 0 &\leq D_+ F^p_{int}(\bar{x}; d) \\ &= \langle \nabla f(\bar{x}), d \rangle + \rho \limsup_{\lambda \downarrow 0} \int_T \frac{g^p_+(\bar{x} + \lambda d, t)}{\lambda} \, \mathrm{d}\mu(t) \\ &\leq \langle \nabla f(\bar{x}), d \rangle + \rho \limsup_{\lambda \downarrow 0} \int_{T \backslash T(\bar{x})} \frac{g^p_+(\bar{x} + \lambda d, t)}{\lambda} \, \mathrm{d}\mu(t) \\ &\quad + \rho \limsup_{\lambda \downarrow 0} \int_{T^{<}(\bar{x}, d)} \frac{g^p_+(\bar{x} + \lambda d, t)}{\lambda} \, \mathrm{d}\mu(t) + \rho \limsup_{\lambda \downarrow 0} \int_{T^{=}(\bar{x}, d)} \frac{g^p_+(\bar{x} + \lambda d, t)}{\lambda} \, \mathrm{d}\mu(t). \end{aligned} \tag{2.50}$$

Firstly, we claim that for any $t \in T$, $\frac{g^p_+(\bar{x} + \lambda d, t)}{\lambda} \to 0$ as $\lambda \to 0$. Note that $T = (T \backslash T(\bar{x})) \cup T^{<}(\bar{x}, d) \cup T^{=}(\bar{x}, d)$. Since, for $t \in T \backslash T(\bar{x})$, $g(\bar{x}, t) < 0$, thus by (a) of Lemma 2.4.2,

$$D_+ g^p_+(\bar{x}, t; d) = 0, \text{ for all } d \in R^n.$$

Similarly, for $t \in T^{<}(\bar{x}, d)$, $g(\bar{x}, t) = 0$ and $\langle \nabla_x g(\bar{x}, t), d \rangle < 0$, by (b) of Lemma 2.4.2, we still have

$$D_+ g^p_+(\bar{x}, t; d) = 0.$$

In the remaining, we consider the case $t \in T^{=}(\bar{x}, d)$, that is $g(\bar{x}, t) = \langle \nabla_x g(\bar{x}, t), d \rangle = 0$. We consider three subcases:
Subcase (i) $p = 0.5$. We have, by (c) of Lemma 2.4.2 and the assumption,

$$0 \leq D_+ g^p_+(\bar{x}, t; d) \leq \sqrt{\max\{\frac{1}{2} g^{\circ\circ}(\bar{x}, t; d), 0\}} = 0.$$

Subcase (ii) $0.5 < p < 1$. We have, by (d) of Lemma 2.4.2 and the assumption $g(\cdot, t)$ is $C^{1,1}$, $D_+ g^p_+(\bar{x}, t; d) = 0$.

Subcase (iii) $0 < p < 0.5$. We have, by (e) of Lemma 2.4.2 and the assumption $g^{\circ\circ}(\bar{x}, t; d) < 0, D_+ g_+^p(\bar{x}, t; d) = 0$.
In all, we have $D_+ g_+^p(\bar{x}, t; d) = 0$. Noting the nonnegativity of $\frac{g_+^p(\bar{x}+\lambda d, t)}{\lambda}$, thus we obtain the claim.

Secondly, we claim that for any constant $\epsilon > 0$ there exists $\lambda^* > 0$ such that $\frac{g_+^p(\bar{x}+\lambda d, t)}{\lambda} \leq \epsilon$ for all $(t, \lambda) \in T \times (0, \lambda^*)$. Indeed, by the continuity of $\frac{g_+^p(\bar{x}+\lambda d, t)}{\lambda}$ with respect to (t, λ), for any $t \in T$, there exists neighborhood V_t of t and $\lambda_t > 0$ such that $\frac{g_+^p(\bar{x}+\lambda d, t')}{\lambda} \leq \epsilon$, for all $(t', \lambda) \in V_t \times (0, \lambda_t)$. By the compactness of T, there are $t^1, \cdots, t^k$ belonging to T such that $T \subset \bigcup_k V_{t^k}$. By letting $\lambda^* = \min_{1 \leq i \leq k} \lambda_{t^i}$, we obtain the claim.

Thirdly, applying Fatou's Lemma to the sequence of functions $\{\epsilon - \frac{g_+^p(\bar{x}+\lambda d, t)}{\lambda} : \lambda \downarrow 0\}$, the last three terms in (2.50) vanish. Thus we have proved $\langle \nabla f(\bar{x}), d \rangle \geq 0$ for all $d \in L_M(\bar{x})$. $\square$

Remark 2.4.1. (i) *By Farkas Lemma [53], (2.49) is equivalent to*

$$-\nabla f(\bar{x}) \in \operatorname{cl}\operatorname{cone}\{\nabla_x g(\bar{x}, t), t \in T(\bar{x})\}.$$

Then (2.49) is weaker than the KKT-type optimality condition which is defined as

$$-\nabla f(\bar{x}) \in \operatorname{cone}\{\nabla_x g(\bar{x}, t), t \in T(\bar{x})\}.$$

(2.49) also implies that

$$\max\{\langle \nabla f(\bar{x}), d \rangle, \langle \nabla_x g(\bar{x}, t), d \rangle, t \in T(\bar{x})\} \geq 0, \textit{ for all } d \in \mathbb{R}^n,$$

which is equivalent to the FJ-tpye optimality condition $0 \in \operatorname{conv}\{\nabla f(\bar{x}), \nabla_x g(\bar{x}, t), t \in T(\bar{x})\}$ *due to the compactness of the set* $\{\nabla_x g(\bar{x}, t), t \in T(\bar{x})\}$. *However, it is easy to see that the reverse does not always hold. For example, consider the problem* $\min x_2 \text{ s.t. } t x_1 + x_2^2 \leq 0, t \in [-1, 1]$. *Indeed, only the FJ-type optimality condition holds at the unique feasible point* $\bar{x} = (0, 0)$, *where* $\nabla f(\bar{x}) = (0, 1)^T, \nabla_x g(\bar{x}, t) = (t, 0)^T, t \in [-1, 1]$. *So, the optimality condition (2.49) is one between the FJ-type and KKT-type optimalities.*

(ii) *In general, without any constraint qualifications, the following optimality condition of SIP holds at* $\bar{x}$

$$\langle \nabla f(\bar{x}), d \rangle \geq 0, \textit{ for all } d \in l_M(\bar{x}),$$

where $l_M(\bar{x}) := \{d \in R^n | \langle \nabla_x g(\bar{x}, t), d\rangle < 0, t \in T(\bar{x})\}$, which is equivalent to the FJ-type optimality condition. Imposing a constraint qualification which ensures that any $d \in L_M(\bar{x})$ is also a feasible direction, then we can also have the optimality condition (2.49). Here $d \in \mathbb{R}^n$ is said to be a feasible direction if there is a smooth feasible arc emanating from $\bar{x}$ with tangent d. Furthermore, with an additional constraint qualification, the KKT-type optimality condition can be obtained. For example, the (extended) Mangasarian-Fromovitz constraint qualification, i.e. $l_M(\bar{x}) \neq \emptyset$, serves both the roles. More details are referred to Krabs [82, 83] and Hettich and Kortanek [53].

(iii) *The extended Abadie CQ [88] is said to hold at $\bar{x}$ if $T_M(\bar{x}) = (\mathrm{cone} \cup_{t \in T(\bar{x})} \nabla_x g(\bar{x}, t))^\circ$, where $T_M(x)$ is the tangent cone of the feasible set M at $x \in M$ and K° denotes the negative polar cone of K. Since $\langle \nabla f(\bar{x}), d\rangle \geq 0$ for all $d \in T_M(\bar{x})$, see [128], then, under the extended Abadie CQ, the optimality condition (2.49) still holds. As a comparison, the Abadie extended CQ is not satisfied for Example 2.4.8, but assumptions in Theorem 2.4.4 are satisfied. The following example adapted from [144] shows that extended Abadie CQ is satisfied but assumptions of Theorem 2.4.4 are not.*

Example 2.4.6. *Consider the SIP problem*

$$\min x \quad s.t. \quad t(x^4 - x^2 - x) + x^8 \leq 0, t \in [1, 2].$$

It is easy to see that for some sufficiently small $\delta > 0$, $M \cap [-\delta, \delta] = [0, \delta]$. Thus $\bar{x} = 0$ is locally optimal. It is easy to check that ACQ holds at $\bar{x}$, and assumption (i) of Theorem 2.4.4 is not satisfied. Also $F_{int}^{\frac{1}{2}}$ is locally exact at $\bar{x}$. Note that $F_{int}^{\frac{1}{2}}(x) = x + \rho \int_1^2 (t(x^4 - x^2 - x) + x^8)_+^{\frac{1}{2}} \,\mathrm{d}t \geq x + \rho \int_1^2 (t(x^4 - x^2 - x))^{\frac{1}{2}} \,\mathrm{d}t = x + \frac{2\rho}{3}(2^{\frac{3}{2}} - 1)(x^4 - x^2 - x)^{\frac{1}{2}} \geq 0$ for all $x < 0$ small enough since the dominate term is $\sqrt{-x}$.

It should be noted that when $0 < p < 1$, we use integral-type exact penalty functions, instead of max-type exact penalty functions, to deal with necessary optimality conditions of SIP, while when $p = 1$ we use the max-type penalty function. The reason why we do so can be shown by the following example (letting $q = 1$ in Example 2.4.2.)

Example 2.4.7. *Consider the following linear SIP*

$$\min\ x \quad \text{s.t.} \quad 2tx - t^2 \leq 0, \quad t \in [-1, 1].$$

The optimal solution is $\bar{x} = 0$. It is easy to see that $F_{max}^{\frac{1}{2}}$ is exact at $\bar{x}$. We now show that $F_{int}^{\frac{1}{2}}$ is not exact. Indeed, for $x \in (-\frac{1}{2}, 0)$ sufficiently small, we have

$$F_{int}^{\frac{1}{2}}(x) = x + \rho \int_{-1}^{1} (2tx - t^2)_+^{\frac{1}{2}} \,\mathrm{d}t = x + \rho \int_{2x}^{0} (2tx - t^2)^{\frac{1}{2}} \,\mathrm{d}t = x + \frac{\rho\pi}{2} x^2 < 0.$$

Since the constraint function is linear, then the second-order condition $g^{\circ\circ}(\bar{x}, t; d) = 0$ is true for all t and d. But

$$\langle \nabla f(\bar{x}), d \rangle < 0,$$

whenever $d \in \mathbb{R}_-$. So, for this example, we cannot develop the optimality conditions by only assuming the exactness of $F_{max}^{\frac{1}{2}}$ and the second-order conditions presented in Theorem 2.4.4.

As we have mentioned before, sometimes we must resort to the lower-order penalty function, since the l_1 penalty function may fail to be exact. Therefore, we give an example to illustrate that Theorem 2.4.4 is applicable, but Theorem 2.4.3 fails. Note that, for this example, the extended Abadie CQ does not hold.

Example 2.4.8. *Consider the following SIP problem*

$$\min x^3 \quad \text{s.t.} \quad tx^6 + x^{12} \le 0, \quad t \in T = [0, 1].$$

The optimal solution is $\bar{x} = 0$. It is clear that F_{max}^{1} is not exact at $\bar{x} = 0$. The extended Abadie CQ is invalid at $\bar{x}$, since $\nabla_x g(\bar{x}, t) = 0$ for all $t \in T$ and

$$\{0\} = T_M(\bar{x}) \ne L_M(\bar{x}) = \mathbb{R}.$$

Theorem 2.4.4 is applicable for the case of $p = \frac{1}{2}$. In fact, by a simple calculation, we have for $x \ge 0$,

$$F_{int}^{\frac{1}{2}}(x) \ge x^3 + \rho \int_0^1 \max\{tx^6, 0\}^{\frac{1}{2}} \,\mathrm{d}t \ge 0;$$

and for $x < 0$,

$$F_{int}^{\frac{1}{2}}(x) \ge x^3 + \rho \int_0^1 \max\{tx^6, 0\}^{\frac{1}{2}} \,\mathrm{d}t = (1 - \rho \int_0^1 t^{\frac{1}{2}} \,\mathrm{d}t) x^3 = (1 - \frac{2}{3}\rho) x^3 \ge 0,$$

whenever $\rho \ge \frac{3}{2}$. So, $F_{int}^{\frac{1}{2}}$ is a local exact penalty function at $\bar{x}$. To be summarized, all assumptions given in Theorem 2.4.4 are satisfied.

Next we employ the local exactness of $\bar{F}_{int}(0 < p < 1)$ to develop the optimality condtition (2.49) of SIP. We first give a proposition that is needed in the proof of the necessary optimality conditions.

Proposition 2.4.1. *Let $g\colon \mathbb{R}_+ \to \mathbb{R}_+$ be a nonnegative function and $f\colon \mathbb{R}_+ \to \mathbb{R}$ be continuous and strictly increasing. Then*

$$\limsup_{\lambda\to\lambda_0} f(g(\lambda)) \le f(\limsup_{\lambda\to\lambda_0} g(\lambda)). \tag{2.51}$$

Proof. Let $\{\lambda^k\}$ be a sequence such that $\lim_{k\to\infty} \lambda_k = \lambda_0$ and

$$\limsup_{\lambda\to\lambda_0} f(g(\lambda)) = \lim_{k\to\infty} f(g(\lambda^k)).$$

Then $\limsup_{k\to\infty} g(\lambda^k) \le \limsup_{\lambda\to\lambda_0} g(\lambda)$. As the limit $\lim_{k\to\infty} f(g(\lambda^k))$ exists and f is strictly increasing, $\lim_{k\to\infty} g(\lambda^k)$ exists. If $\lim_{k\to\infty} g(\lambda^k) = \infty$, then $\limsup_{\lambda\to\lambda_0} g(\lambda) = \infty$. Thus (2.51) holds. Assume now that $\lim_{k\to\infty} g(\lambda^k) < \infty$. By the continuity and monotonicity of f,

$$\lim_{k\to\infty} f(g(\lambda^k)) = f(\lim_{k\to\infty} g(\lambda^k)) \le f(\limsup_{\lambda\to\lambda_0} g(\lambda)),$$

and the assertion (2.51) also holds. □

Theorem 2.4.5. *Let $p \in (0,1)$ and $\bar{F}^p_{int}$ be locally exact at $\bar{x}$. Under any one of the three assumptions below,*
(i) $p = 0.5$ *and* $g^{\circ\circ}(\bar{x}, t; d) \le 0$ *for all* $t \in T^=(\bar{x}, d)$ *and* $d \in L_M(\bar{x})$*,*
(ii) $0.5 < p < 1$ *and* $g(\cdot, t)$ *is* $C^{1,1}$*, for all* $t \in T^=(\bar{x}, d)$*, and*
(iii) $0 < p < 0.5$ *and* $g^{\circ\circ}(\bar{x}, t; d) < 0$*, for all* $t \in T^=(\bar{x}, d)$ *and* $d \in L_M(\bar{x})$*, we have*

$$\langle \nabla f(\bar{x}), d\rangle \ge 0 \text{ for all } d \in L_M(\bar{x}). \tag{2.52}$$

Proof. Given $d \in L_M(x^*)$, we have that

$$\begin{aligned} 0 \le D_+\bar{F}^p_{int}(x^*, d) =& \langle \nabla f(x^*), d\rangle + \rho \limsup_{\lambda\downarrow 0} \frac{(\int_T g_+(x^* + \lambda d, t)\, \mathrm{d}\mu(t))^p}{\lambda} \\ \le& \langle \nabla f(x^*), d\rangle + \rho \left(\limsup_{\lambda\downarrow 0} \frac{\int_T g_+(x^* + \lambda d, t)\, \mathrm{d}\mu(t)}{\lambda^{1/p}}\right)^p, \end{aligned} \tag{2.53}$$

where the last inequality follows from Proposition 2.4.1. Note that

$$\limsup_{\lambda\downarrow 0} \frac{\int_T g_+(x^* + \lambda d, t)\, \mathrm{d}\mu(t)}{\lambda^{1/p}} = \limsup_{\lambda\downarrow 0} \int_T \left(\frac{g^p_+(x^* + \lambda d, t)}{\lambda}\right)^{1/p} \mathrm{d}\mu(t).$$

Following exactly the proof in Theorem 2.4.4, we have any $t \in T$,

$$\frac{g_+^p(x^* + \lambda d, t)}{\lambda} \to 0 \quad \text{as } \lambda \to 0.$$

and that for any constant $\epsilon > 0$ there exists $\lambda^* > 0$ such that $\frac{g_+^p(x^*+\lambda d,t)}{\lambda} \leq \epsilon$ for any $(t, \lambda) \in T \times (0, \lambda^*)$. By applying Fatou's Lemma to the sequence $\{\epsilon^{1/p} - \left(\frac{g_+^p(x^*+\lambda d,t)}{\lambda}\right)^{1/p}\}$, we have

$$\begin{aligned}
&\liminf_{\lambda \downarrow 0} \int_T \left(\epsilon^{1/p} - \left(\frac{g_+^p(x^* + \lambda d, t)}{\lambda}\right)^{1/p}\right) \mathrm{d}\mu(t) \\
&\quad \geq \mu(T)\epsilon^{1/p} - \int_T \limsup_{\lambda \downarrow 0} \left(\frac{g_+^p(x^* + \lambda d, t)}{\lambda}\right)^{1/p} \mathrm{d}\mu(t) \\
&\quad \geq \mu(T)\epsilon^{1/p} - \int_T \left(\limsup_{\lambda \downarrow 0} \frac{g_+^p(x^* + \lambda d, t)}{\lambda}\right)^{1/p} \mathrm{d}\mu(t) \\
&\quad = \mu(T)\epsilon^{1/p} - \int_T \left(D_+ g_+^p(x^*, t; d)\right)^{1/p} \mathrm{d}\mu(t) \\
&\quad = \mu(T)\epsilon^{1/p},
\end{aligned}$$

where the second inequality again follows from Proposition 2.4.1. Thus

$$\limsup_{\lambda \downarrow 0} \int_T \left(\frac{g_+^p(x^* + \lambda d, t)}{\lambda}\right)^{1/p} \mathrm{d}\mu(t) = 0,$$

and so the second term of the last term in (2.53) vanishes. This completes the proof. □

2.5 A Numerical Scheme Based on Lower Order Penalty Functions

In this section, we propose a conceptual algorithm based on our lower order penalty functions F^p. Here we only use the integral-type penalty functions.

In general, the algorithm with penalty functions proceeds as follows, see for example Fiacco and McCormick [31].

Let $\epsilon > 0$ and the sequence $\{\rho_k\}$ be strictly monotonically increasing to $+\infty$.

Step 1. For $\rho_1 > 0$, find an unconstrained local minimum of $F^p(x, \rho_1)$. Denote it by $x(\rho_1)$.

Step 2. Given $x(\rho_k)$, find an unconstrained local minimum $x(\rho_{k+1})$ of $F^p(x, \rho_{k+1})$.

Step 3. Stop if $\|x_{\rho_{k+1}} - x_{\rho_k}\| < \epsilon$, otherwise return to Step 2.

Note that in Step 1 and 2 of the algorithm we need to solve an unconstrained problem

$$\min F^p(x, \rho_k) = f(x) + \rho_k \int_T g_+^p(x,t) \mathrm{d}\mu(t) \quad s.t. \quad x \in \mathbb{R}^n. \tag{2.54}$$

We can choose any unconstrained method to derive its local optimal solutions. As there is an integral involved, we may proceed with techniques from numerical integration. If T is an interval, say $[a, b]$, the most widely used integration method is Simpson's rule, which is a discretization method by approximating the integral with finite sums corresponding to some partition of the interval of integration $[a, b]$. If T is a subset of a multidimensional space, the curse of dimensionality occurs when one try to phrase the multiple integral as repeated one-dimensional integrals by appealing to Fubini's theorem. Monte Carlo and sparse grids are two known methods to overcome this curse. More details are referred to Stoer and Burlirsch [137].

For simplicity's sake, consider $T = [a, b]$. For any integral $\int_a^b h(t)dt$, the integrand h is replaced by an interpolation polynomial $P(t)$. Specifically, consider a uniform partition of $[a, b]$ given by

$$t_i = a + id, \qquad i = 0, \cdots, m, d = (b-a)/m.$$

The interpolation polynomial P_m is of degree m or less with

$$P(t_i) = h_i := h(t_i), i = 0, \cdots, m.$$

Then Newton-Cotes formulas

$$\int_a^b P_m(t)dt = d \sum_{i=0}^m h_i \alpha_i$$

with α_i being weights solely depending on m, provide approximate values for $\int_a^b h(t)dt$. When $m = 2$, it is Simpson's rule.

With above described integration technique and letting $h_m(x)$ be the approximate integral of interpolation polynomial of $g_+^p(x, t)$ obtained from Newton-Cotes formula, problem (2.54) can be approximated by

$$\min f(x) + \rho_k h_m(x) \quad s.t. \quad x \in \mathbb{R}^n.$$

Let x_m^k be a local solution of the above problem. Then x_m^k may provide approximate solution of x_{ρ_k}.

Chapter 3

Generalized Semi-Infinite Programming

3.1 Introduction

Problem description. The generalized semi-infinite programming problem (GSIP) is of the following form

$$\begin{aligned} &\min\ f(x) \\ &\text{s.t.}\ \ g(x, y) \leq 0, y \in Y(x), \end{aligned} \tag{3.1}$$

where $f : \mathbb{R}^n \to \mathbb{R}, g : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$, and $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is a set-valued mapping. GSIP features that its index set is variable dependent and in general is of infinite elements. When Y is not x-dependent, it reduces to the standard semi-infinite programming problem. Usually, we assume that Y is defined by a system of inequalities

$$Y(x) := \{y \in Y_0 \mid v(x, y) \leq 0\}, \tag{3.2}$$

where the function $v\colon \mathbb{R}^n \to \mathbb{R}^l$ and $Y_0 \subset \mathbb{R}^m$.

For simplicity, we will not consider the cases when there are multiple semi-infinite constraints and when there are equality constraints present in the definitions of (3.1) and (3.2). In our considerations, the set Y_0 is taken as either $\mathbb{R}^m$ or a compact subset of $\mathbb{R}^m$, which will be specified in later sections.

The lower level problem. The lower level problem associated with GSIP is

$$Q(x) \qquad \max g(x,y) \quad \text{s.t.} \quad y \in Y(x). \tag{3.3}$$

This is a parameterized optimization problem (x as a parameter). The estimates of the value function of $Q(x)$ and keeping track of the solution set $Y_0(x)$ of $Q(x)$ are crucial for the study of GSIP. Since the value function of $Q(x)$ is usually nonsmooth even though all defining functions are smooth, nonsmooth analysis techniques are involved.

Denote by M and $\text{cl}M$ the feasible set and the closure of the feasible set of GSIP, respectively. Let $Y_0(x) = \{y \in Y(x) \mid g(x,y) = 0\}$ the active index set at x, and $\phi(x)$ be the optimal value function of $Q(x)$:

$$\phi(x) = \begin{cases} \sup_{y \in Y(x)} g(x,y), & \text{if } Y(x) \neq \emptyset, \\ -\infty, & \text{otherwise.} \end{cases} \tag{3.4}$$

Note that when $Y(x) = \emptyset$ the point x is always feasible. GSIP problem (3.1) is equivalent to the following optimization problem with one constraint:

$$\min f(x) \quad \text{s.t.} \quad \phi(x) \leq 0. \tag{3.5}$$

It is shown that the feasible set M possesses topological structures not known to standard semi-infinite problems or finite problems. It turns out that a GSIP problem is a much harder problem than an SIP problem. Early studies [78, 55] neglecting these special structures managed to investigate the GSIP problem by transforming it into finite nonlinear programming under regularity conditions. To give an impression of the structure of GSIP, we list below some basic and useful properties related to Y and M:

(i) If Y is locally bounded and osc at x, then ϕ is usc at x;

(ii) If Y is isc at x, then ϕ is lsc at x;

(iii) M is closed if Y is isc on $\mathbb{R}^n$;

(iv) M can be reformulated as

$$M = [\text{pr}_x(\text{gph}Y \cap \mathcal{G}^C)]^C, \tag{3.6}$$

where pr_x denotes the projection onto the x-space, $\text{gph}Y$ is the graph of the mapping Y, $\mathcal{G} := \{(x,y) \mid g(x,y) \leq 0\}$, and A^C is the set complement of A in $\mathbf{R}^N$;

(v) An upper estimate of closure of M can be given as

$$\mathrm{cl}M \subset \{x \mid \sigma(x,y) \leq 0, y \in \mathbb{R}^m\}, \tag{3.7}$$

where

$$\sigma(x,y) = \min\{g(x,y), -v_1(x,y), \cdots, -v_l(x,y)\}; \tag{3.8}$$

(vi) Generically (for the function space with respect to (g,v)),

$$\mathrm{cl}M = \{x \mid \sigma(x,y) \leq 0, y \in \mathbb{R}^m\}. \tag{3.9}$$

Upper semi-continuity of ϕ is easy to obtain provided that the defining functions are continuous. Lower semi-continuity of ϕ, which attributes the closeness of M, is more demanding but also desirable. From the formula of M in (iv), the special structures of M such as the noncloseness and disjunctiveness, can be anticipated since it is obtained by some basic set operations of half closed/open planes if by taking all defining functions being linear. The symmetric representation formula of M in (vi) prompts a lot of recent research on GSIP [46, 47, 43, 69].

Since the feasible set M is not always closed, it is reasonable to consider the corresponding optimization problem on cl$M$, the closure of M, that is

$$\min f(x) \qquad \text{s.t.} \qquad x \in \mathrm{cl}M. \tag{3.10}$$

GSIP and problem (3.10) are equivalent in the sense that if f is continuous then they share the same optimal value.

Assume that GSIP is bounded below. Let $\bar{x} \in \mathrm{cl}M$ solve GSIP. Several cases occur:

1. $\bar{x}$ is an infeasible solution and $\phi(\bar{x}) > 0$;

2. $\bar{x}$ is a feasible solution with $Y(\bar{x}) = \emptyset$ and $\phi(\bar{x}) = -\infty$;

3. $\bar{x}$ is a feasible solution with $Y(\bar{x}) \neq \emptyset$ and $\phi(\bar{x}) < 0$;

4. $\bar{x}$ is a feasible solution with $Y(\bar{x}) \neq \emptyset$ and $\phi(\bar{x}) = 0$.

In the second case, usually $\bar{x}$ is an interior point of the domain of feasible set of GSIP problem provided that the data involved are continuous. In the third case, $\bar{x}$ is also an interior point if ϕ is upper semi-continuous, which is guaranteed in general. Thus, we in most times focus on the last situation.

Although we concentrate on necessary optimality conditions, there are some literature on sufficient optimality conditions such as [55, 121, 129].

Outline of this chapter is as follows. Section 3.2 provides a short introduction of calmness condition for GSIP problems. Section 3.3 proposes the approximate problems for GSIP via penalty functions of the lower level problems. Section 3.4 investigates optimality conditions for nonsmooth GSIP problems. Section 3.5 derives optimality conditions for GSIP via lower order penalty functions developed in Section 2.4. Section 3.6 derives the strong duality theory of GSIP with convex lower level problems via generalized augmented Lagrangians.

3.2 Calmness

As mentioned, GSIP can be equivalently reformulated as the following optimization problem with one nonsmooth constraint

$$\min f(x) \quad \text{s.t.} \quad \phi(x) \le 0, \tag{3.11}$$

where ϕ is the optimal value function of the lower level problem $Q(x)$. Let $\tilde{\phi}(x) = \max\{\phi(x), 0\} = \phi_+(x)$. Then GSIP can be rewritten as

$$\min f(x) \quad \text{s.t.} \quad \tilde{\phi}(x) \le (\text{or} =) 0. \tag{3.12}$$

It is easy to see that the following result holds.

Proposition 3.2.1. *If $\tilde{\phi}(x) = 0$ and ϕ is upper semicontinuous at x, then $\tilde{\phi}$ is continuous at x.*

Definition 3.2.1 (calmness)**.** *Let $p(u)$ be the value function of the following perturbed problem*

$$P(u) \qquad \min f(x) \quad \text{s.t.} \quad \phi(x) \le u. \tag{3.13}$$

GSIP is said to be calm at an optimal solution $\bar{x}$ if there are constants $\bar{\alpha} \ge 0$ and $\epsilon > 0$ such that for every $(x, u) \in \mathbb{R}^n \times \mathbb{R}$ with $\|x - \bar{x}\| \le \epsilon$ and $\phi(x) \le u$, we have

$$f(x) + \bar{\alpha}\|u\| \ge f(\bar{x}).$$

Remark 3.2.1. (i) *The definition of calmness can be independent of the optimal solution $\bar{x}$. We may say the family of perturbed problems $P(u)$ is calm at $u = 0$ if we have*

$$\liminf_{u \to 0} \frac{p(u) - p(0)}{\|u\|} > -\infty.$$

(ii) *The definition of calmness of GSIP at $\bar{x}$ in Clarke [22, Def. 6.4.1] is: there are constants $\bar{\alpha} \geq 0$ and $\epsilon > 0$ such that for every $(x, u) \in \mathbb{R}^n \times \mathbb{R}$ with $\|x - \bar{x}\| \leq \epsilon, |u| \leq \epsilon$ and $\phi(x) \leq u$, we have*

$$f(x) + \bar{\alpha}\|u\| \geq f(\bar{x}).$$

Due to the continuity property of $\tilde{\phi}$ at $\bar{x} = 0$ from Proposition 3.2.1, we have that the two versions of definitions of calmness are equivalent, as can be seen from [12].

A basic and important result about calmness is its equivalence to the exact penalization.

Theorem 3.2.1 (Burke [12, Theorem 1.1])**.** *Let $\bar{x}$ satisfy $\phi(\bar{x}) \leq 0$. Then GSIP is calm at $\bar{x}$ with the constants $\bar{\alpha}, \epsilon$ as in the Definition 3.2.1 if and only if $\bar{x}$ is a local minimum of radius ϵ for*

$$P_\alpha(x) := f(x) + \alpha\phi^+(x), \tag{3.14}$$

for all $\alpha \geq \bar{\alpha}$.

The function ϕ as the optimal value function of the lower level problem is in general hard to obtain and thus difficult to deal with.

Next we use the penalty function of the lower level problem to replace ϕ and obtain a relaxation of GSIP problem. Let $\bar{g}(x, y, c)$ be a penalty function associated the lower level problem $Q(x)$ and

$$\psi(x, c) := \sup_{y \in Y_0} \bar{g}(x, y, c). \tag{3.15}$$

For example, $\bar{g}(x, y, c) := g(x, y) - c\alpha(x, y)$ for $c > 0$ and some nonnegative function α which satisfies that $\alpha(x, y) = 0$ if and only if $v(x, y) \leq 0$. We may take $\alpha(x, y) = \sum_{i=}^{l}([v_i(x, y)]_+)^2$ or $\sum_{i=1}^{l}[v_i(x, y)]_+$. With the assumptions of the compactness of Y_0

and continuity of the data, the known result about penalization is that for x with $Y(x) \neq \emptyset$,

$$\phi(x) = \inf_{c} \psi(x, c)$$

where $\phi(x) = \sup_{y \in Y(x)} g(x, y)$, is the optimal value of the lower level problem $Q(x)$.

Since $\psi(x, c) \geq \phi(x)$ for any $c > 0$, we have

$$\{x \mid \exists c > 0 \text{ s.t. } \psi(x, c) \leq 0\} \subset M. \tag{3.16}$$

Consider the following perturbed problem

$$\min f(x) \quad \text{s.t.} \quad \psi(x, c) \leq u. \tag{3.17}$$

The *calmness* of above problem (3.17) at $\bar{x}$ is defined as

> there exists $\epsilon > 0, \alpha > 0$ such that for any (x, c) satisfying $\psi(x, c) \leq u, \|x - \bar{x}\| \leq \epsilon, |u| \leq \epsilon$, we have that
>
> $$f(x) - f(\bar{x}) \geq -\alpha |u|. \tag{3.18}$$

This calmness leads to the following *exactness*:

> for some $\epsilon' > 0, \alpha' > 0$, we have for each $x \in B(x_0, \epsilon')$, there is $c_x > 0$ such that for all $c \geq c_x$,
>
> $$f(x) + \alpha' \psi^+(x, c) \geq f(x_0). \tag{3.19}$$

The feasible set of the problem (3.17) is a subset of the feasible set of GSIP and thus the two problems are not equivalent. However, the calmness conditions for these two problems are the same.

Proposition 3.2.2. (i) *The calmness of problem* (3.17) *is equivalent to the exactness condition* (3.19)*;*

(ii) *The calmness of problem* (3.17) *at $\bar{x}$ is equivalent to the calmness of GSIP.*

Proof. (i). Exactness of (3.17) implies its calmness: proof is direct.
The reverse direction: loss of exactness is to say that for all $k \to \infty$, there is $\rho^k \to \infty, (x^k, c^k) \in \mathbb{R}^n \times \mathbb{R}_+, x^k \to x_0$ such that

$$f(x^k) + \rho^k \psi^+(x^k, c^k) < f(x_0),$$

this implies that $\rho^k \psi^+(x^k, c^k) \to 0$. By letting $u^k = \psi^+(x^k, c^k)$, it easily leads to a contradiction regarding the calmness assumption.

(ii). From inequality (3.19), the monotonicity of ψ with respect to c and the zero gap property, we have that (by letting $c \to \infty$)

$$f(x) + M\phi^+(x) \geq f(x_0).$$

The equivalence follows easily. $\square$

3.3 Lower Level Penalty Transform

Due to the complexity of the structure of GSIP, simple and tractable approximations for GSIP and its feasible set are important aspects of the research of GSIP. Stein [127] used the following function

$$\sigma(x, y) = \min\{g(x, y), -v_1(x, y), \cdots, -v_l(x, y)\} \tag{3.20}$$

to provide an upper estimate of the closure of the feasible set of GSIP and thus developed approximates for the tangent cone of the feasible set and first order necessary optimality conditions for GSIP with degenerate index sets. Recent researches based on the function σ gave sufficient conditions that characterize the closure of the feasible set, see [46, 47, 43, 69]. Still [136] described several approaches to transform GSIP but with 'good' lower level value functions into simpler problems, such as dual-, penalty-, discretization-, reduction-, and Karush-Kuhn-Tucker (KKT)-methods.

In this section, we consider inner and outer approximations of the feasible set of GSIP problem via the penalty functions of the lower level problem, consider corresponding approximating problems of GSIP, and derive some properties between the approximate solutions and the solutions of GSIP, and between the optimal vaules of the approximating problems and GSIP. Under some appropriate assumptions, we try to approximate and/or characterize the feasible set of GSIP (or the closure of the feasible set) by use of some simple functions – here we choose the penalty functions. Our analysis features that we make no assumptions that the feasible set is closed or of some regularity properties of the lower level problems. It turns out that the approximate problems can effectively approximate the optimization problems under its closure instead of the original problem and an asymptotic problem provides a lower estimate for GSIP.

As in Section 3.2, given any $x \in \mathbb{R}^n$, associate the lower level problem $Q(x)$ with the penalty problem with parameter $c > 0$:

$$\psi(x, c) := \max_{y \in Y_0} \bar{g}(x, y, c), \tag{3.21}$$

where $\bar{g}(x, y, c) := g(x, y) - c\alpha(x, y)$ is a penalty function for the lower level problem $Q(x)$. We may take $\alpha(x, y) = \sum_i ([v_i(x, y)]_+)^2$ for example. Due to the compactness of Y_0 and continuity of the data, the solution set $\tilde{Y}_x$ of $Q(x)$ is nonempty and for any $c > 0$ the solutions of its corresponding penalty problem exist and their cluster points belong to $\tilde{Y}_x$.

Lemma 3.3.1. (i) *For any x satisfying $Y(x) \neq \emptyset$ we have*

$$\phi(x) = \inf_c \psi(x, c), \tag{3.22}$$

where $\phi(x)$ is the optimal value function of $Q(x)$.

(ii) *For x satisfying $Y(x) = \emptyset$, we still have*

$$\phi(x) = -\infty = \inf_c \psi(x, c). \tag{3.23}$$

Some properties of ϕ and ψ are listed as follows.

(i) ϕ is usc and ψ is continuous and non-increasing in c.

(ii) For $y \in Y(x)$, $g(x, y) = \bar{g}(x, y, c)$ and thus $\psi(x, c) \geq \phi(x)$.

(iii) $M = \{x \mid \phi(x) \leq 0\}$.

(iv) $\{x \mid \phi(x) < 0\} = \{x \mid \exists c > 0 \text{ s.t. } \psi(x, c) < 0\} \subset \mathrm{int} M$.

(v) $\{x \mid \phi(x) \leq 0\} \supseteq \{x \mid \exists c > 0 \text{ s.t. } \psi(x, c) = 0\}$.

The fact that ϕ is only usc can be problematic. It leads to the nonclosedness of M. In general, the lower semi-continuity is a standard assumption for minimization problems and is not as easy to guarantee as the upper semi-continuity. Besides, the strong duality result for the lower level problem $Q(x)$ is equivalent to the lower semi-continuity of ϕ at x. This perspective hints the difficulty in dealing with GSIP. Below we provide some common results that ensure the lower semi-continuity of ϕ.

Proposition 3.3.1 ([59, 34])**.** *The following assertions hold:*

(i) *If the set-valued mapping Y is isc at x, then ϕ is lsc at x.*

(ii) *If $Y(\bar{x}) \neq \emptyset$, and MFCQ holds at some $\bar{y} \in \arg\max_{y \in Y(\bar{x})} g(\bar{x}, y)$, then ϕ is continuous at $\bar{x}$.*

(iii) *Under the assumption that*

$$\{y \in Y_0 \mid v(x, y) \leq 0\} \subset \mathrm{cl}\{y \in Y_0 \mid v(x, y) < 0\},$$

Y is isc at x. Thus, we have that ϕ is continuous at x.

(iv) *if Y_0 is convex, if each v_i is continuous and is convex w.r.t. y for every fixed x, and if there is a y such that $v(x, y) < 0$, then Y is isc at x.*

Remark 3.3.1. *For a usc function ϕ, neither*

$$\mathrm{cl}\{x \mid \phi(x) < 0\} \subset \{x \mid \phi(x) \leq 0\}$$

nor

$$\mathrm{cl}\{x \mid \phi(x) < 0\} \supset \{x \mid \phi(x) \leq 0\}$$

holds. For example, $\phi(x_1, x_2) = \begin{cases} -1, & \text{if } x_2 < 0, \\ 0, & \text{if } x_1 > 0, x_2 \geq 0, \\ 1, & \text{otherwise.} \end{cases}$

To some extent, the set

$$\{x \in \mathbb{R}^n \mid \exists c > 0 \text{ s.t. } \psi(x, c) < 0\}$$

provides an inner approximation of the feasible set M of GSIP.

Proposition 3.3.2. *If* $\mathrm{cl}\,\mathrm{int}M = \mathrm{cl}M$ *and* $\mathrm{int}M = \{x \mid \phi(x) < 0\}$, *then* $\mathrm{cl}M = \mathrm{cl}\left(\Pr\{(x, c) : \psi(x, c) \leq 0\}\right)$.

Proof. Since $\{x \mid \exists c > 0 \text{ s.t. } \psi(x,c) \le 0\} \subset M$, then $\mathrm{cl}\,\mathrm{Pr}\{(x,c) \mid \psi(x,c) \le 0\} \subset \mathrm{cl}M$. On the other hand, for $x \in \mathrm{cl}M$, there exists $x^k \in \mathrm{int}M$ such that $x^k \to x$. Then $\phi(x^k) < 0$ by assumption and thus there exists $c^k > 0$ such that $\psi(x^k, c^k) < 0$ due to the zero gap property of penalization. This leads to the other inclusion $\mathrm{cl}M \subset \mathrm{cl}\,\mathrm{Pr}\{(x,c) \mid \psi(x,c) \le 0\}$. □

Based on the previous discussions, it prompts us to consider the following two problems:

$$\begin{array}{llll} (P) & \min f(x) & \text{s.t.} & x \in \mathrm{cl}\,M, \\ (P_c) & \min f(x) & \text{s.t.} & \psi(x,c) \le 0. \end{array}$$

We intend to use problem (P_c) to approximate problem (P). Denote by M_c the feasible set of (P_c) and by $\mathrm{val}\,(P)$ the optimal value of problem (P). Then it is obvious that $M_{c_1} \subset M_{c_2} \subset M$ for all $0 < c_1 \le c_2$. However the relation $\mathrm{cl\,M} = \mathrm{cl} \cup_{\mathrm{c}>0} \mathrm{M_c} = \lim_{\mathrm{c}\to\infty} \mathrm{M_c}$ may not always hold.

Proposition 3.3.3. *Let* $\mathrm{clM} = \mathrm{cl\,int\,M}$, $\mathrm{int}\{x \mid \phi(x) = 0\} = \emptyset$ *or* $\{x \mid \phi(x) = 0\} \subset \mathrm{cl}\{x \mid \phi(x) < 0\}$. *Then the following assertions hold:*

(i) $\mathrm{val\,(P_c) \to val\,(P), as\, c \to \infty}$;

(ii) $\limsup_{c\to\infty}(\epsilon\text{-}\arg\min(P_c)) \subset \epsilon\text{-}\arg\min(P), \text{for } \epsilon > 0$;

(iii) $\limsup_{c\to\infty}(\epsilon_c\text{-}\arg\min(P_c)) \subset \arg\min(P), \text{for } \epsilon_{\mathrm{c}} \downarrow 0$;

(iv) $\limsup_{c\to\infty}(\arg\min(P_c)) \subset \arg\min(P)$.

Proof. (i) If $\mathrm{int}\{x \mid \phi(x) = 0\} = \emptyset$, then $\mathrm{int}M \subset \{x \mid \phi(x) < 0\}$ and thus $\mathrm{int}M = \{x \mid \phi(x) < 0\}$. Then assumptions of Proposition 3.3.2 are satisfied. Then $\mathrm{cl\,M} = \{\mathrm{x} \mid \exists \mathrm{c} > 0 \text{ s.t. } \psi(\mathrm{x},\mathrm{c}) \le 0\}$. Under the assumption $\{x \mid \phi(x) = 0\} \subset \mathrm{cl}\{x \mid \phi(x) < 0\}$, using basic properties of ϕ and ψ, we still have $\mathrm{cl\,M} = \{\mathrm{x} \mid \exists \mathrm{c} > 0 \text{ s.t. } \psi(\mathrm{x},\mathrm{c}) \le 0\}$. Thus $M_c \to \mathrm{cl}M$. And since $M_c \subset M$, the result (i) follows.

(ii) Let x be a cluster point of ϵ-$\arg\min(P_c)$, i.e., for some $x_{c_k} \in \epsilon$-$\arg\min(P_{c_k})$, $x_{c_k} \to x$. Then $x \in \mathrm{cl\,M}$. Assume the contrary, that $f(x) > \mathrm{val(P)} + \epsilon$. Then there is a

$\delta > 0$ such that for all $x' \in B(x,\delta) \cap \mathrm{cl}\,\mathrm{M}$, $f(x') > \mathrm{val(P)} + \epsilon + \Delta$ with $\Delta := \frac{f(x)-(\mathrm{val(P)}+\epsilon)}{2}$. For c_k large enough, x_{c_k} belongs to $B(x,\delta) \cap \mathrm{cl}\,\mathrm{M}$, and thus

$$f(x_{c_k}) > \mathrm{val(P)} + \epsilon + \Delta.$$

On the other hand, from (i),

$$\mathrm{val\,(P)} + \Delta \geq \mathrm{val\,(P_{c_k})}$$

for c_k large enough. Then we get

$$f(x_{c_k}) > \mathrm{val(P_{c_k})} + \epsilon,$$

which contradicts that x_{c_k} is ϵ-optimal for (P_{c_k}).

(iii) is a consequence of (ii). Since

$$\limsup_{c\to\infty}(\epsilon_c\text{-}\arg\min(P_c)) \subset \limsup_{c\to\infty}(\epsilon\text{-}\arg\min(P_c)),$$

then for any $\epsilon > 0$, $\limsup_{c\to\infty}(\epsilon_c\text{-}\arg\min(P_c)) \subset \limsup_{c\to\infty}(\epsilon\text{-}\arg\min(P))$. Letting $\epsilon \to 0$, (iii) holds. Similarly, since $\limsup_{c\to\infty}(\arg\min(P_c)) \subset \limsup_{c\to\infty}(\epsilon_c\text{-}\arg\min(P_c))$, then (iv) also holds. $\square$

On the other hand, for any $\epsilon > 0$, the set

$$\{x \in \mathbb{R}^n \mid \exists c > 0 \text{ s.t. } \psi(x,c) \leq \epsilon\}$$

provides an outer approximation of the feasible set M of GSIP. Under some stability assumption of the sets $\{\psi(x,c) \leq \epsilon\}$ w.r.t. ϵ (or the projection of them), we obtain the equivalence between GSIP and SIP.

Consider the following SIP problem

$$\min_{(x,c)\in\mathbb{R}^n\times\mathbb{R}^+} f(x) \quad \text{s.t.} \quad \bar{g}(x,y,c) \leq 0, y \in Y_0 \tag{3.24}$$

which can also be written as

$$\min f(x) \quad \text{s.t.} \quad \psi(x,c) \leq 0. \tag{3.25}$$

The corresponding perturbed problem:

$$\mathrm{SIP}(\epsilon) \qquad \min f(x) \quad \text{s.t.} \quad \psi(x,c) \leq \epsilon. \tag{3.26}$$

Another description of SIP(ϵ), the asymptotic case of SIP is the problem

$$(\mathrm{SIP}_a) \qquad \min\{\limsup f(x^k) : \limsup \psi(x^k, c^k) \leq 0\}. \tag{3.27}$$

In (SIP$_a$), we may restrict our consideration to the sequences $\{(x^k, c^k)\}$ with $\{c^k\}$ being non-decreasing. Otherwise, just replace $\{(x^k, c^k)\}$ by $\{(x^k, \max_{1\leq i\leq k} c^i)\}$.

Let $p(\epsilon) = \operatorname{val}\mathrm{SIP}(\epsilon) = \inf_{(x,c)}\{f(x) : \psi(x,c) \leq \epsilon\}$. The following inclusion is obvious: for any $\epsilon > 0$,

$$\{x \mid \psi(x,c) \leq 0 \text{ for some } c > 0\} \subset M \subset \{x \mid \psi(x,c) \leq \epsilon \text{ for some } c > 0\}. \tag{3.28}$$

Proposition 3.3.4. *The following properties are obvious:*

(i) $\operatorname{val}(\mathrm{SIP}) \geq \operatorname{val}(\mathrm{GSIP}),\ \operatorname{val}(\mathrm{SIP}) \geq \operatorname{val}\mathrm{SIP}(\epsilon)$*;*

(ii) $M = \cap_{\epsilon>0}\overline{M}_\epsilon$ *where* $\overline{M}_\epsilon = \Pr\{(x,c) \mid \psi(x,c) \leq \epsilon\}$*;*

(iii) $M \subseteq \lim_{\epsilon\to 0}\overline{M}_\epsilon$ *where* $\overline{M}_\epsilon = \operatorname{pr}\{(x,c) \mid \psi(x,c) \leq \epsilon\}$*.*

Proposition 3.3.5. $\operatorname{val}(\mathrm{SIP}_a) = \liminf_{\epsilon\to 0} p(\epsilon)$*.*

Proof. Let $\alpha = \operatorname{val}(\mathrm{SIP}_a)$ and $\beta = \liminf_{\epsilon\to 0} p(\epsilon)$. Show by contradiction that $\alpha < \beta$ and $\alpha > \beta$.

First, let $\alpha < \beta$. Then there is a $\delta > 0, \epsilon_0 > 0$ such that

$$\operatorname{val}(\mathrm{SIP}_a) + \delta < p(\epsilon), \forall \epsilon \in (0, \epsilon_0].$$

Let $\{(x_k, c_k)\}_k$ be the optimal solution of (SIP$_a$). Then there is $\epsilon_k \to 0$ such that

$$\psi(x_k, c_k) \leq \epsilon_k, f(x_k) \to \operatorname{val}(\mathrm{SIP}_a).$$

On the other hand, x_k is a feasible solution of SIP(ϵ) and thus

$$f(x_k) \geq p(\epsilon_k) > \delta + \alpha.$$

Letting $k \to \infty$, a contradiction is obtained and thus $\alpha < \beta$ does not hold. Similarly, we obtain that $\alpha > \beta$ also fails.

We also have $\beta = \liminf_{\epsilon\downarrow 0} p(\epsilon)$. □

In all, we have

$$p(0) = \mathrm{val}\,(\mathrm{SIP}) \geq \mathrm{val}\,(\mathrm{GSIP}) \geq \mathrm{val}\,(\mathrm{SIP}_a) = \liminf_{\epsilon \downarrow 0} p(\epsilon) \tag{3.29}$$

since val (SIP) $\geq$ val (GSIP) $\geq p(\epsilon)$ for any $\epsilon > 0$.

Theorem 3.3.1. (i) *Problem* (SIP_a) *provides a lower estimate for* GSIP.

(ii) *If p is lsc at $\epsilon = 0$, then* SIP *and* GSIP *are equivalent. Especially, if* (SIP_ϵ) *is calm at $\epsilon = 0$, then p is lsc at $\epsilon = 0$.*

Proof. (i) follows directly from relation (3.29).
(ii) If p is lsc at $\epsilon = 0$, then all quantities in (3.29) are equal. Thus the equivalence of SIP and GSIP follows. The calmness means that for some $M > 0$, the relation

$$p(\epsilon) - p(0) \geq -M|\epsilon|,$$

for all ϵ near 0, then p is lsc at $\epsilon = 0$. □

Note that there always exists a sequence $\{(x^k, c^k)\}$ such that

$$\limsup_{k\to\infty} \psi(x^k, c^k) \leq 0, \text{ and } f(x^k) \to \mathrm{val}\,\mathrm{SIP}_a. \tag{3.30}$$

Furthermore, if $\bar{x}$ is a limit point of a subsequence of $\{x_k\}$ and $\bar{x}$ is also feasible for GSIP, then $\bar{x}$ solves GSIP.

3.4 Nonsmooth Generalized Semi-Infinite Programming

In this section we mainly focus on GSIP problem with defining functions being Lipschitz continuous:

$$\min f(x) \quad \text{s.t.} \quad g(x,y) \leq 0, y \in Y(x), \tag{3.31}$$

where $Y(x) := \{y \in \mathbb{R}^m \mid v(x,y) \leq 0\}$, and derive its necessary optimality conditions via the generalized differentiation. We will consider the following two cases with different approaches:

1. developing optimality via the min-max formulation with Lipschitz lower level optimal value function;

2. developing optimality via bilevel formulation under the assumption of partial calmness(not necessarily with Lipschitz lower level value function).

With the nonlinear programming formulation

$$\min f(x) \quad \text{s.t.} \quad \phi(x) \leq 0,$$

we relate GSIP to the min-max problem

$$\min_x \max\{f(x) - f(\bar{x}), \phi(x)\}.$$

On the other hand, we can also relate GSIP to the following bilevel problem

$$\begin{aligned} \min_{(x,y)} f(x) \quad \text{s.t.} \quad & g(x,y) \leq 0, \\ & y \in \arg\max\{g(x,y) : y \in Y(x)\}. \end{aligned} \tag{3.32}$$

The problem (3.32) is a special bilevel optimization problem in that its upper level constraint is the same as the object of its lower level problem. However, there is a *slight difference* between GSIP problem (3.31) and problem (3.32). The feasible set of problem (3.32) is a subset of problem (3.31) in that the feasible set of problem (3.31) is the combination of the feasible set of problem (3.32) and the complement of dom$Y$. For more comparisons between GSIP problems and bilevel problems, see Stein and Still [130].

The bilevel problem (3.32) is equivalent to the following problem via the value function $\phi(x)$:

$$\min_{(x,y)} f(x) \quad \text{s.t.} \quad y \in Y(x), g(x,y) \leq 0, \phi(x) - g(x,y) \leq 0.$$

It can be rewritten as a nonlinear programming problem

$$\min_{(x,y)} f(x) \quad \text{s.t.} \quad (x,y) \in \Omega, G(x,y) \leq 0, \tag{3.33}$$

where $\Omega = \{(x,y) \mid g(x,y) \leq 0, v(x,y) \leq 0\}$ and $G(x,y) = \phi(x) - g(x,y)$. Dempe and Zemkoho [27] introduced the following WMFCQ

$$\partial G(\bar{x},\bar{y}) \cap (-\text{bd}N_\Omega(\bar{x},\bar{y})) = \emptyset,$$

to derived the necessary optimality conditions for the bilevel problem. They also took an alternate approach to use the concept of partial calmness, introduced by Ye and Zhu [146] which is weaker than WMFCQ. The partial calmness corresponding to the above bilevel problem is of the following form

$$f(x) - f(\bar{x}) + k|u| \geq 0, \forall (u, x, y) \in V_{(0,\bar{x},\bar{y})} : (x, y) \in \Omega, G(x, y) \leq u.$$

Finally, let's remark that a bilevel problem can be formulated as a GSIP problem. Consider a bilevel problem of the form

$$\begin{aligned} &\min_{x,y} \{F(x, y) : y \in S(x)\}, \\ &\qquad \text{where } S(x) = \{y \mid y \text{ solves } \min f(x, y) \text{ s.t. } g(x, y) \leq 0\}. \end{aligned} \tag{3.34}$$

The lower level optimal value function reformulation of the bilevel programming problem (3.34) is

$$\begin{aligned} \min_{x,y} F(x, y) \quad \text{s.t.} \quad & f(x, y) - v(x) \leq 0, \\ & g(x, y) \leq 0, \end{aligned}$$

where $v(x)$ is the value function of the lower level problem. Let

$$Y(x, y) = \{z \mid g(x, z) \leq 0\} \quad \text{(independent of } y).$$

Then the bilevel problem (3.34) is reformulated as the following GSIP problem

$$\begin{aligned} \min_{x,y} F(x, y) \quad \text{s.t.} \quad & f(x, y) - f(x, z) \leq 0, z \in Y(x, y), \\ & g(x, y) \leq 0. \end{aligned} \tag{3.35}$$

The equivalence between the bilevel problem (3.34) and GSIP problem (3.35) is easy to obtain. Note that $y^* \in S(x)$ is to say that $g(x, y^*) \leq 0$, and $f(x, y^*) \leq f(x, z)$ for all z satisfying $g(x, z) \leq 0$.

It is different from the case when transforming a GSIP problem into a bilevel problem that the set $Y(x, y) \equiv Y(x) = \emptyset$ makes no differences for the two problems. That is, x is not feasible for both problems. For the GSIP formulation, if $Y(x, y) = \emptyset$, then by convention that $\sup\{f(x, y) - f(x, z) \mid z \in Y(x, y)\} = -\infty$. On the other hand, $Y(x, y) = \emptyset$ means that there is no y such that $g(x, y) \leq 0$. Thus, x is not feasible for the GSIP problem (3.35).

More details of the value function reformulation for bilevel problem can be found in Chen and Florian [19] and Ye and Zhu [146].

Given a function $f\colon \mathbb{R}^n \to \overline{\mathbb{R}}$ and a point $\bar{x}$ with $f(\bar{x})$ finite. Recall that the regular subdifferential of f at $\bar{x}$ is defined by

$$\hat{\partial} f(\bar{x}) := \{u \in \mathbb{R}^n \mid \liminf_{x \to \bar{x}, x \neq \bar{x}} (f(x) - f(\bar{x}) - \langle u, x - \bar{x} \rangle)/\|x - \bar{x}\| \geq 0\}. \tag{3.36}$$

The general (basic, limiting) and singular subdifferential of f at $\bar{x}$ are defined respectively by

$$\partial f(\bar{x}) := \limsup_{x \xrightarrow{f} \bar{x}} \hat{\partial} f(x) \qquad \text{and} \qquad \partial^\infty f(\bar{x}) := \limsup_{x \xrightarrow{f} \bar{x}, \lambda \downarrow 0} \lambda \hat{\partial} f(x).$$

The upper regular subdifferential of f at $\bar{x}$ is defined by $\hat{\partial}^+ f(\bar{x}) := -\hat{\partial}(-f)(\bar{x})$, and the upper subdifferential of f at $\bar{x}$ $\partial^+ f(\bar{x}) := \limsup_{x \xrightarrow{f} \bar{x}} \hat{\partial}^+ f(x)$.

The following definitions are required for further development.

Definition 3.4.1 ([93, Definition 1.63])**.** *Let $S\colon X \rightrightarrows Y$ with $\bar{x} \in dom\, S$.*

(i) *Given $\bar{y} \in S(\bar{x})$, we say that the mapping S is inner semi-continuous at $(\bar{x}, \bar{y})$ if for every sequence $x_k \to \bar{x}$ there is a sequence $y_k \in S(x_k)$ converging to $\bar{y}$ as $k \to \infty$.*

(ii) *S is inner semi-compact at $\bar{x}$ if for every sequence $x_k \to \bar{x}$ there is a sequence $y_k \in S(x_k)$ that contains a convergent subsequence as $k \to \infty$.*

(iii) *S is μ-inner semicontinuous at $(\bar{x}, \bar{y})$ (μ-inner semicompact at $\bar{x}$) if in above two cases, $x_k \to \bar{x}$ is replaced by $x_k \to \bar{x}$ with $\mu(x_k) \to \mu(\bar{x})$.*

Here the concept of μ-inner semicontinuity/semicompactness is important for our considerations. It is typical that the value function ϕ of the lower level problem $Q(x)$ of GSIP is not continuous, even taking value $-\infty$.

Theorem 3.4.1 (subdifferentiation of maximum functions [93, Theorem 3.46])**.** *Consider the maximum function of the form*

$$(\max \phi_i)(x) = \max\{\phi_i(x) \mid i = 1, \cdots, l\}. \tag{3.37}$$

Let ϕ_i be lsc around $\bar{x}$ for $i \in I(\bar{x})$ and be usc at $\bar{x}$ for $i \notin I(\bar{x})$. Assume the qualification holds:

$$[\sum_{i \in I(\bar{x})} x_i^* = 0, x_i^* \in \partial^\infty \phi_i(\bar{x})] \Rightarrow x_i^* = 0, i \in I(\bar{x}). \tag{3.38}$$

Then

$$\partial(\max \phi_i)(\bar{x}) \subset \cup\{\sum_{i \in I(\bar{x})} \lambda_i \partial\phi_i(\bar{x}) \mid (\lambda_1, \cdots, \lambda_l) \in \Lambda(\bar{x})\}, \tag{3.39}$$

where $\Lambda(\bar{x}) = \{(\lambda_1, \cdots, \lambda_l) \mid \lambda_i \geq 0, \sum_{i=1}^l \lambda_i = 1, \lambda_i(\phi_i(\bar{x}) - (\max \phi_i)(\bar{x})) = 0\}$.

Note that the qualification (3.38) always holds if all related functions are locally Lipschitz.

The following two results are about continuity properties and estimates of subdifferentials of marginal functions which are crucial to our analysis for GSIP problems.

Proposition 3.4.1 (limiting subgradients of marginal functions [97])**.** *Consider the parametric optimization problem*

$$\mu(x) := \inf\{\varphi(x, y) \mid y \in G(x)\}, \tag{3.40}$$

and let $M(x) := \{y \in G(x) \mid \mu(x) = \varphi(x, y)\}$, $G(x) := \{y \in \mathbb{R}^m \mid \varphi_i(x, y) \leq 0, i = 1, \cdots, l\}$. *For simplicity, we don't consider the case with equality constraints involved.*

(i) *Assume that M is μ-inner semicontinuous at $(\bar{x}, \bar{y}) \in \mathrm{gph}M$, that φ and all φ_i are Lipschitz continuous around $(\bar{x}, \bar{y})$, and that the following qualification condition is satisfied:*

$$\text{only the vector } (\lambda_1, \cdots, \lambda_l) = 0 \in \mathbb{R}^l \text{ satisfies the relation } 0 \in \sum_{i=1}^l \lambda_i \partial\varphi_i(\bar{x}, \bar{y})$$
$$\text{for some } (\lambda_1, \cdots, \lambda_l) \in \mathbb{R}^l_+ \text{ with } \lambda_i \varphi_i(\bar{x}, \bar{y}) = 0, i = 1, \cdots, l. \tag{3.41}$$

One has the inclusions

$$\partial\mu(\bar{x}) \subset \{u^* \in X^* \mid (u^*, 0) \in \partial\varphi(\bar{x}, \bar{y}) + \sum_{i=1}^l \lambda_i \partial\varphi_i(\bar{x}, \bar{y})$$
$$\text{for some } (\lambda_1, \cdots, \lambda_l) \in \mathbb{R}^l_+ \text{ with } \lambda_i \varphi_i(\bar{x}, \bar{y}) = 0, i = 1, \cdots, l\}.$$

(ii) *Assume that M is μ-inner semicompact at $\bar{x}$, and all φ, and φ_i are Lipschitz continuous around $(\bar{x},\bar{y})$ for all $\bar{y} \in M(\bar{x})$ and the qualification* (3.41) *holds for all $(\bar{x},\bar{y}), \bar{y} \in M(\bar{x})$. Then*

$$\partial\mu(\bar{x}) \subset \cup_{\bar{y}\in M(\bar{x})}\{u^* \in X^* | (u^*, 0) \in \partial\varphi(\bar{x},\bar{y}) + \sum_{i=1}^{l} \lambda_i \partial\varphi_i(\bar{x},\bar{y}),$$
$$\textit{for some } (\lambda_1, \cdots, \lambda_l) \in \mathbb{R}^m_+ \textit{ with } \lambda_i\varphi_i(\bar{x},\bar{y}) = 0, i = 1, \cdots, l\}. \quad (3.42)$$

Proposition 3.4.2 (Lipschitz continuity of marginal functions [95, Theorem 5.2])**.** *Continue to consider the parametric problem* (3.40) *in Propostion 3.4.1. Then the following assertions hold:*

(i) *Assume that M is μ-inner semicontinuous at $(\bar{x},\bar{y}) \in$ gph$M$ and φ is locally Lipschitz around this point. Then μ is Lipschitz around $\bar{x}$ provided that it is lsc around $\bar{x}$ and G is Lipschitz-like around $(\bar{x},\bar{y})$.*

(ii) *Assume that M is μ-inner compact at $\bar{x}$ and φ is locally Lipschitz around $(\bar{x},\bar{y})$ for all $\bar{y} \in M(\bar{x})$. Then μ is Lipschitz around $\bar{x}$ provided that it is lsc around $\bar{x}$ and G is Lipschitz-like around $(\bar{x},\bar{y})$ for all $\bar{y} \in M(\bar{x})$.*

Now we are prepared to develop the optimality conditions for GSIP problem (3.31). Given a local solution $\bar{x}$ of problem (3.31), associate it with the following min-max problem

$$\min_x \max\{f(x) - f(\bar{x}), \phi(x)\}, \quad (3.43)$$

where $\phi(x) := \sup_{y\in Y(x)} g(x,y)$. Let $Y_0(x) := \{y \in Y(x) : \phi(x) = g(x,y)\}$. Denote by

$$F(x) = \max\{f(x) - f(\bar{x}), \phi(x)\}.$$

If $\bar{x}$ solves GSIP (3.31), then $\bar{x}$ also solves problem

$$\min\{F(x) : x \in \mathbb{R}^n\}$$

and thus by generalized Fermat's rule, cf. [116, Theorem 10.1], we have

$$0 \in \partial F(\bar{x}). \quad (3.44)$$

So, calculus for the maximum function and the estimate of subdifferentials are essential to proceed. From equation (3.44) and Theorem 3.4.1, there exists $\mu \in [0,1]$ such that (if ϕ is Lipschitz)

$$0 \in \mu\partial f(\bar{x}) + (1-\mu)\partial\phi(\bar{x}) \subset \mu\partial f(\bar{x}) + (1-\mu)\mathrm{co}\,\partial\phi(\bar{x}). \tag{3.45}$$

Note that for a Lipschitz function ϕ,

$$\mathrm{co}\,\partial\phi(\bar{x}) = -\mathrm{co}\,\partial(-\phi)(\bar{x}). \tag{3.46}$$

Theorem 3.4.2 (optimality for GSIP with Lipschitz lower level otpimal value function)**.** *Consider the GSIP problem* (3.31)*, and let $\bar{x} \in M$ be its locally optimal solution. Assume that all functions f, g and v are Lipschitz continuous, Y_0 is ϕ-inner semicompact at $\bar{x}$ and Y is Lipschitz-like at $(\bar{x}, \bar{y})$ for all $\bar{y} \in Y_0(\bar{x})$. Then there is $\bar{y}^j \in Y_0(\bar{x})$ and $\bar{\lambda}_0 \geq 0, \bar{\lambda}_j \geq 0, \bar{\alpha}_i^j \geq 0, i = 1, \cdots, l, j = 1, \cdots, k$ such that $\sum_{j=1}^k \bar{\lambda}_j = 1$ and*

$$0 \in (\bar{\lambda}_0\partial f(\bar{x}), 0) + \sum_{j=1}^{k} \bar{\lambda}_j[-\partial(-g)(\bar{x}, \bar{y}^j) - \sum_{i=1}^{l} \bar{\alpha}_i^j \partial v_i(\bar{x}, \bar{y}^j)]. \tag{3.47}$$

If in addition g and all components of v are regular at all $(\bar{x}, \bar{y}^j)$, then the optimality is of the form

$$0 \in \bar{\lambda}_0\partial f(\bar{x}) + \sum_{j=1}^{k} \bar{\lambda}_j[-\partial_x(-g)(\bar{x}, \bar{y}^j) - \sum_{i=1}^{l} \bar{\alpha}_i^j \partial_x v_i(\bar{x}, \bar{y}^j)], \tag{3.48}$$

$$0 \in \sum_{j=1}^{k} \bar{\lambda}_j[-\partial_y(-g)(\bar{x}, \bar{y}^j) - \sum_{i=1}^{l} \bar{\alpha}_i^j \partial_y v_i(\bar{x}, \bar{y}^j)]. \tag{3.49}$$

Note that $-\partial(-g) = \partial^+ g$.

Proof. Under regularity and Lipschitz continuity, since the following calculus rule for basic subgradients holds, see [116, Corollary 10.11]:

$$\partial_x f(x, y) = \{u \mid \exists\, v \text{ s.t. } (u, v) \in \partial f(x, y)\},$$

equations (3.48) and (3.49) follow directly from (3.47). Note that the function $-\phi$ is in position of μ in Proposition 3.4.1. Under our assumptions, by Proposition 3.4.2, $-\phi$ is Lipschitz continuous and the estimate of $\partial(-\phi)(\bar{x})$ is

$$\partial(-\phi)(\bar{x}) \subset \cup_{\bar{y}\in Y_0(\bar{x})}\{u^* \in X^* \mid (u^*, 0) \in \partial(-g)(\bar{x}, \bar{y}) + \sum_{i=1}^{l} \alpha_i \partial v_i(\bar{x}, \bar{y}),$$
$$\text{for some } (\alpha_1, \cdots, \alpha_l) \in \mathbb{R}_+^l \text{ with } \alpha_i v_i(\bar{x}, \bar{y}) = 0, i = 1, \cdots, l\}. \tag{3.50}$$

If $\bar{x} \in M$ solves GSIP, then it also solves $\min_x F(x)$. By equations (3.44)– (3.46), there exists $\mu \in [0, 1]$ such that

$$0 \in \mu \partial f(\bar{x}) - (1-\mu) \text{co}\, \partial(-\phi)(\bar{x}). \tag{3.51}$$

Combining (3.50) and (3.51), there is $\bar{y}^j \in Y_0(\bar{x})$ and $\lambda_j \geq 0, \alpha_i^j \geq 0, i = 1, \cdots, l, j = 1, \cdots, k$ such that $\sum_{j=1}^k \lambda_j = 1$ and

$$0 \in (\mu \partial f(\bar{x}), 0) + (1-\mu) \sum_{j=1}^k \lambda_j [-\partial(-g)(\bar{x}, \bar{y}^j) - \sum_{i=1}^l \alpha_i^j \partial v_i(\bar{x}, \bar{y}^j)]. \tag{3.52}$$

Letting $\bar{\lambda}_0 = \mu, \bar{\lambda}_j = (1-\mu)\lambda_j, \bar{\alpha}_i^j = \alpha_i^j, i = 1, \cdots, l, j = 1, \cdots, k$, we obtain the desired result. $\square$

Corollary 3.4.1. *In addition to the assumptions in Theorem 3.4.2, assume that f, g and v are continuously differentiable. Then the optimality condition at the optimal point $\bar{x}$ is that there exist $\bar{y}^j \in Y_0(\bar{x})$ and $\bar{\lambda}_0 \geq 0, \bar{\lambda}_j \geq 0, \bar{\alpha}_i^j \geq 0, i = 1, \cdots, l, j = 1, \cdots, k$ such that $\sum_{j=1}^k \bar{\lambda}_j = 1$ and*

$$0 = \bar{\lambda}_0 \nabla_x f(\bar{x}) + \sum_{j=1}^k \bar{\lambda}_j [\nabla_x g(\bar{x}, \bar{y}^j) - \sum_{i=1}^l \bar{\alpha}_i^j \nabla_x v_i(\bar{x}, \bar{y}^j)], \tag{3.53}$$

$$0 = \sum_{j=1}^k \bar{\lambda}_j [\nabla_y g(\bar{x}, \bar{y}^j) - \sum_{i=1}^l \bar{\alpha}_i^j \nabla_y v_i(\bar{x}, \bar{y}^j)]. \tag{3.54}$$

Next we consider the case where the lower level value function ϕ fails to be Lipschitz and give estimates of the subdifferential of ϕ and thus further derive the optimality conditions for GSIP. However, it requires to use the Clarke convexified subdifferential.

There are two different approaches to define the Clarke's normal cone. On one hand, it can be defined by the polar cone of the Clarke's tangent cone

$$\overline{N}_A(x) := \widehat{T}_A(x)^\circ,$$

where $\widehat{T}_A(x) = \liminf_{y \xrightarrow{A} x, t \searrow 0} \frac{A-y}{t}$ or defined via the generalized directional derivative of the (Lipschitzian) distant function $\text{dist}(\cdot, A)$, see Clarke [21]. On the other hand, it can be defined by the closed convex hull of the (general) normal cone

$$\overline{N}_A(x) := \text{cl co}\, N_A(x).$$

For this definition and the equivalence of the two definitions, see for example, Rockafellar and Wets [116].

The Clarke subgradients and Clarke horizon subgradients of f at x are defined by

$$\begin{aligned} \overline{\partial} f(x) &:= \{v \mid (v,-1) \in \overline{N}_{\operatorname{epi} f}(x, f(x))\}, \\ \text{and} \qquad \overline{\partial}^{\infty} f(x) &:= \{v \mid (v,0) \in \overline{N}_{\operatorname{epi} f}(x, f(x))\}. \end{aligned} \tag{3.55}$$

The relationship between the Clarke subsubdifferentials and basic subdifferentials is also referred to Mordukhovich [93, Theorem 3.57].

Proposition 3.4.3. *Let f be proper and lsc around $\bar{x} \in \operatorname{dom} f$. Then*

$$\overline{\partial} f(\bar{x}) = \operatorname{cl}\operatorname{co}[\partial f(\bar{x}) + \partial^{\infty} f(\bar{x})].$$

If, in particular, f is Lipschitz continuous at $\bar{x}$, then

$$\overline{\partial} f(\bar{x}) = \operatorname{cl}\operatorname{co}\partial f(\bar{x}).$$

The general/basic normal cone N_A enjoys the robustness property

$$N_A(\bar{x}) = \limsup_{x \to \bar{x}} N_A(\bar{x})$$

provided that the setting is finite dimension [93, page 11]. However, this is not true for the convexified cone $\overline{N}_A$, see for example Rockafellar [115]:

$$A := \{x \in \mathbb{R}^3 | x_3 = x_1 x_2 \, or \, x_3 = -x_1 x_2\} \text{ and } \bar{x} = (0,0,0).$$

The normal cone $\overline{N}_A(\bar{x})$ is just the x_3-axis, but $\overline{N}_A(x)$ is the x_2x_3-plane for all $x = (x_1, 0, 0)$. The following proposition is from Rockafellar [115].

Proposition 3.4.4. *If A is convex, or if $\overline{N}_A(\bar{x})$ is pointed, then the multifunction $\overline{N}_A$ is closed at $\bar{x}$, that is, for all $x^k \to \bar{x}, y^k \xrightarrow{\overline{N}_A(x^k)} \bar{y}$, one has $\bar{y} \in \overline{N}_A(\bar{x})$.*

Proposition 3.4.5. *The Clarke normal cone has the robustness property*

$$\overline{N}_A(x) = \limsup_{y \to x} \overline{N}_A(y)$$

provided that $N_A(x)$ is pointed.

Proof. It suffices to prove that $\limsup_{y\to x} \operatorname{co} N_A(y) \subset \operatorname{cl}\operatorname{co} \limsup_{y\to x} N_A(y) = \overline{N}_A(x)$. Let $v \in \limsup_{y\to x} \operatorname{co} N_A(y)$. Then there are $y_k \in A, v_{ik} \in N_A(y_k), i = 1, \cdots, n+1$ such that

$$\sum_{i=1}^{n+1} v_{ik} \to v \, as \, k \to \infty,$$

since the sets $N_A(y_k)$ are cones. Let $\lambda_k = \sum_{i=1}^{n+1} \|v_{ik}\|$. Then $\{\lambda_k\}$ is bounded, that is also to say $\{v_{ik}\}$ are bounded for all i. Otherwise, $\sum_{i=1}^{n+1} \frac{v_{ik}}{\lambda_k} \to 0$. That is $v_1 + \cdots + v_{n+1} = 0$, where v_i is the limit of $\{\frac{v_{ik}}{\lambda_k}\}_k$ for each i. Note that $v_i \in N_A(x)$ since $N_A(x) = \limsup_{y\to x} N_A(y)$. Thus $v_1 = \cdots = v_{n+1} = 0$ by the pointedness of $N_A(x)$. On the other hand, $\sum_i \|v_i\| = 1$. This is a contradiction. Thus the sequence $\{v_{ik}\}$ is bounded. By taking subsequences, we may assume that $v_{ik} \to v_i$. Then $v_i \in N_A(x)$ and $v = v_1 + \cdots + v_{n+1}$. This completes the proof. □

Proposition 3.4.6. *Consider the parametric optimization problem same as* (3.40)*:*

$$\mu(x) := \inf\{\varphi(x, y) \mid y \in G(x)\}$$

with corresponding solution mapping $M \colon X \rightrightarrows Y$. Let $\bar{x} \in \operatorname{dom} M$. Assume that the following conditions hold:

(i) *φ is lower semi-continuous at $\bar{x}$;*

(ii) *M is μ-inner semi-compact at $\bar{x}$ and $M(\bar{x})$ is nonempty and compact;*

(iii) *If $(u_i, w_i) \in \bar{\partial}^\infty \varphi(\bar{x}, \bar{y}_i), (v_i, -w_i) \in \overline{N}_{\operatorname{gph}G}(\bar{x}, \bar{y}_i), \bar{y}_i \in M(\bar{x}), i \le n+1$, and $\sum_i u_i + v_i = 0$, then $u_i = v_i = 0, w_i = 0$;*

(iv) *The cones $N_{\operatorname{gph}G}(\bar{x}, \bar{y})$ and $N_{\operatorname{epi}\varphi}(\bar{x}, \bar{y}, \varphi(\bar{x}, \bar{y}))$ for all $\bar{y} \in M(\bar{x})$ are pointed.*

Then we have the inclusion

$$\bar{\partial}\mu(\bar{x}) \subset \operatorname{co}(\bigcup_{\bar{y}\in M(\bar{x})} \{u \mid (u, 0) \in \bar{\partial}\varphi(\bar{x}, \bar{y}) + \overline{N}_{\operatorname{gph}G}(\bar{x}, \bar{y})\}).$$

Proof. (Sketch. The definition of $\bar{\partial} = \operatorname{cl}\operatorname{co}\{\partial + \partial^\infty\}$. The proof is divided into two parts. First the set on the right hand of the required inclusion, denoted by Λ, is closed. The second step is to justify that $\partial + \partial^\infty \subset \Lambda$.)
Let $\sum_{i=1}^{n+1} \lambda_i^k u_i^k \to u, \lambda_i^k \ge 0, \sum_i \lambda_i^k = 1$, and

$$(u_i^k, 0) = (x_{1i}^k, y_{1i}^k) + (x_{2i}^k, y_{2i}^k), i = 1,\, 2,\, \ldots,\, n+1, \tag{3.56}$$

where $(x^k_{1i}, y^k_{1i}) \in \bar{\partial}\varphi(\bar{x}, \bar{y}^k_i)$, $(x^k_{2i}, y^k_{2i}) \in \overline{N}_{\text{gph}G}(\bar{x}, \bar{y}^k_i), \bar{y}^k_i \in M(\bar{x}), i = 1, \cdots, n+1$. We have to show that $u \in \Lambda$. We may assume that $\bar{y}^k_i \to \bar{y}_i, i = 1, \cdots, n+1$, as $M(\bar{x})$ is compact. We show first that the sequence $\{z^k := (z^k_i)_i = (x^k_{1i}, y^k_{1i}, x^k_{2i}, y^k_{2i})\}_k$ is bounded. Suppose on the contrary that $\|z^k\| \to \infty$. Then for each i,

$$\frac{1}{\|z^k\|}(u^k_i, 0) = \frac{1}{\|z^k\|}(x^k_{1i}, y^k_{1i}) + (u^k_{2i}, v^k_{2i}), \text{ with } (u^k_{2i}, v^k_{2i}) = \frac{1}{\|z^k\|}(x^k_{2i}, y^k_{2i}). \tag{3.57}$$

Multiplying (3.57) by λ^k_i and taking summation over i, we have

$$\frac{1}{\|z^k\|}(\sum_i \lambda^k_i x^k_{1i}, \sum_i \lambda^k_i y^k_{1i}) + (\sum_i \lambda^k_i u^k_{2i}, \sum_i \lambda^k_i v^k_{2i}) \to (0, 0). \tag{3.58}$$

Note that for each i, by definition of $\bar{\partial}$,

$$(\frac{1}{\|z^k\|}(x^k_{1i}, y^k_{1i}), \frac{-1}{\|z^k\|}) \in \overline{N}_{\text{epi}\varphi}(\bar{x}, \bar{y}^k_i, \varphi(\bar{x}, \bar{y}^k_i)). \tag{3.59}$$

Let $(u^k_{1i}, v^k_{1i}) := \frac{1}{\|z^k\|}(x^k_{1i}, y^k_{1i}) \to (u_{1i}, v_{1i})$. Then by assumption (iv) and Proposition 3.4.5 one has $(u_{1i}, v_{1i}, 0) \in \overline{N}_{\text{epi}\varphi}(\bar{x}, \bar{y}_i, \varphi(\bar{x}, \bar{y}_i))$ and thus $(u_{1i}, v_{1i}) \in \bar{\partial}^\infty\varphi(\bar{x}, \bar{y}_i)$. Then

$$\frac{1}{\|z^k\|}(\sum_i \lambda^k_i x^k_{1i}, \sum_i \lambda^k_i y^k_{1i}) \to \sum \lambda_i(u_{1i}, v_{1i}), \tag{3.60}$$

where $\lambda_i := \lim_k \lambda^k_i, \sum \lambda_i = 1$. Let $(u^k_{2i}, v^k_{2i}) \to (u_{2i}, v_{2i}) \in \overline{N}_{\text{gph}G}(\bar{x}, \bar{y}_i)$. Then $v_{1i} + v_{2i} = 0$ from (3.56). Combining (3.58) and (3.60), we have

$$\sum \lambda_i((u_{1i}, v_{1i}) + (u_{2i}, v_{2i})) = 0. \tag{3.61}$$

Based on the assumption (iii), we have $(u_{1i}, v_{1i}) = (0, 0), (u_{2i}, v_{2i}) = (0, 0)$. This contradicts the fact that $(u_{1i}, v_{1i}, u_{2i}, v_{2i})$ is of norm 1 and thus $\{z^k\}$ is bounded. Then (x^k_{1i}, y^k_{1i}), (x^k_{2i}, y^k_{2i}) have convergent subsequences, say $(x^k_{1i}, y^k_{1i}) \to (x_{1i}, y_{1i}) \in \bar{\partial}\varphi(\bar{x}, \bar{y}_i), (x^k_{2i}, y^k_{2i}) \to (x_{2i}, y_{2i}) \in \overline{N}_{\text{gph}G}(\bar{x}, \bar{y}_i)$. Thus $u = \sum \lambda_i u_i$ with $u_i = x_{1i} + x_{2i}$. That is to say Λ is closed.

Next, we justify that $\partial + \partial^\infty \subset \Lambda$. Assume that $u_1 \in \partial\mu(\bar{x}), u_2 \in \partial^\infty\mu(\bar{x})$. Under the semi-compactness assumption, invoking [93, Theorem 1.108], one gets that

$$\partial\mu(\bar{x}) \subset \{u \mid (u, 0) \in \cup_{\bar{y} \in M(\bar{x})} \partial(\varphi(\bar{x}, \bar{y}) + \delta((\bar{x}, \bar{y}), \text{gph}G))\}, \tag{3.62}$$

$$\partial^\infty\mu(\bar{x}) \subset \{u \mid (u, 0) \in \cup_{\bar{y} \in M(\bar{x})} \partial^\infty(\varphi(\bar{x}, \bar{y}) + \delta((\bar{x}, \bar{y}), \text{gph}G))\}. \tag{3.63}$$

Employing the sum rule from [93, Theorem 3.36] to the two above leads to

$$\begin{aligned}(u_1 + u_2, 0) &\in [\partial\varphi(\bar{x}, \bar{y}) + N_{\text{gph}G}(\bar{x}, \bar{y})] + [\partial^\infty\varphi(\bar{x}, \bar{y}) + N_{\text{gph}G}(\bar{x}, \bar{y})] \\ &\subset \text{cl}\,\text{co}\,[\partial\varphi(\bar{x}, \bar{y}) + \partial^\infty\varphi(\bar{x}, \bar{y})] + \overline{N}_{\text{gph}G}(\bar{x}, \bar{y}) \\ &= \bar{\partial}\varphi(\bar{x}, \bar{y}) + \overline{N}_{\text{gph}G}(\bar{x}, \bar{y}),\end{aligned} \tag{3.64}$$

which completes the proof. □

Theorem 3.4.3 (convexified normal cone to inequality system [96])**.** *Consider G defined by inequality system $G(x) = \{y \mid \phi(x,y) \le 0\}$. Let ϕ be Lipschitz and the qualification (non-smooth MFCQ) at $(\bar{x},\bar{y})$ hold*

$$[\sum_{i\in I(\bar{x},\bar{y})} \lambda_i w_i = 0 \text{ with } w_i \in \partial\phi_i(\bar{x},\bar{y}), \lambda_i \ge 0] \Rightarrow \lambda_i = 0 \text{ for } i \in I(\bar{x},\bar{y}). \tag{3.65}$$

Then

$$\overline{N}_{\mathrm{gph}G}(\bar{x},\bar{y}) \subset \{\sum_{i\in I(\bar{x},\bar{y})} \lambda_i \bar{\partial}\phi_i(\bar{x},\bar{y}) \mid \lambda_i \ge 0\}.$$

As mentioned, GSIP can be relaxed into the following bilevel programming problem

$$\min_{(x,y)} f(x) \text{ s.t. } g(x,y) \le 0, v(x,y) \le 0, \phi(x) - g(x,y) \le 0. \tag{3.66}$$

The feasible set of above problem is a subset of the feasible set M of (3.31). Thus, if $\bar{x}$ solves GSIP and $\phi(\bar{x}) = 0$, and $Y(\bar{x}) \ne \emptyset$, then $\bar{x}$ also solves problem (3.32). The perturbed version of the above bilevel problem is

$$\min_{(x,y)} f(x) \text{ s.t. } g(x,y) \le 0, v(x,y) \le 0, \phi(x) - g(x,y) \le u. \tag{3.67}$$

Problem (3.32) is said to be *partially calm* at $(\bar{x},\bar{y})$ if

$$\begin{gathered} \text{there is } \kappa > 0 \text{ and a neighborhood } V \text{ of } (\bar{x},\bar{y},0) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R} \text{ such that} \\ \text{for all } (x,y,u) \in V \text{ feasible for (3.67) we have} \\ f(x) - f(\bar{x}) \ge -\kappa|u|. \end{gathered} \tag{3.68}$$

Under the partial calmness condition, problem (3.32) can be transformed into the problem below, for some constant $\kappa > 0$,

$$\min_{(x,y)} f(x) + \kappa[\phi(x) - g(x,y)] \text{ s.t. } g(x,y) \le 0, v(x,y) \le 0. \tag{3.69}$$

Theorem 3.4.4 (necessary conditions of optimality of GSIP)**.** *Let $\bar{x}$ be an optimal solution of GSIP with $\phi(\bar{x}) = 0$, and $Y(\bar{x}) \ne \emptyset$. Let the data functions f, g and v be Lipschitz and the partial calmness condition* (3.68) *hold at $(\bar{x},\bar{y})$ for some $\bar{y} \in Y_0(\bar{x})$. Assume that the following conditions hold:*

(i) *Qualificaiton* (3.65) *holds for $\Omega := \{(x,y) \mid v(x,y) \le 0, g(x,y) \le 0\}$ at $(\bar{x},\bar{y})$;*

(ii) *Y_0 is ϕ-inner semi-compact at $\bar{x}$ and $Y_0(\bar{x}) \ne \emptyset$;*

(iii) *If $(u_i, w_i) \in \bar{\partial}^\infty g(\bar{x},\bar{y}_i), (v_i, -w_i) \in \overline{N}_{\mathrm{gph}Y}(\bar{x},\bar{y}_i), \bar{y}_i \in Y_0(\bar{x}), i \le n+1$, and $\sum_i u_i + v_i = 0$, then $u_i = v_i = 0$ and $w_i = 0$;*

(iv) *The cones* $N_{\text{gph}Y}(\bar{x}, y)$ *and* $N_{\text{epi}\, g}(\bar{x}, y, g(\bar{x}, y))$ *are pointed for all* $y \in Y_0(\bar{x})$.

Then there is $\kappa > 0$, $\lambda_i \geq 0, \bar{y}_i \in Y_0(\bar{x}), i = 1, \cdots, r$ *such that* $\sum_{i=1}^r \lambda_i = 1$, *and*

$$0 \in (\bar{\partial} f(\bar{x}), 0) + \kappa \sum_{i=1}^r \lambda_i [\bar{\partial} g(\bar{x}, \bar{y}_i) - \bar{\partial} g(\bar{x}, \bar{y})] + \sum_{i=1}^r \overline{N}_{\text{gph}Y}(\bar{x}, \bar{y}_i) + \overline{N}_\Omega(\bar{x}, \bar{y}). \quad (3.70)$$

Proof. Let $\Omega = \{(x, y) \mid v(x, y) \leq 0, g(x, y) \leq 0\}$. Under our assumptions, GSIP can be relaxed into problem (3.32) and $(\bar{x}, y)$ also solves (3.32) for all $y \in Y_0(\bar{x})$. Due to the partial calmness of (3.32), we also have $(\bar{x}, \bar{y})$ solves (3.69). Under the qualification assumption, the necessary optimality condition for problem (3.69) is, see for example [26, Theorem 5.1] or [96, Theorem 6.2],

$$0 \in (\partial f(\bar{x}), 0) + \kappa(\partial \phi(\bar{x}), 0) + \kappa \partial(-g)(\bar{x}, \bar{y}) + N_\Omega(\bar{x}, \bar{y}). \quad (3.71)$$

If $v \in \partial^+(-\phi)(x)$, then $-v \in \bar{\partial}\phi(x)$. Indeed, by definition,

$$\begin{aligned} \partial^+(-\phi)(x) &= \{v | (-v, 1) \in N_{\text{hypo}(-\phi)}(x, -\phi(x))\} \\ &= \{v | (-v, -1) \in N_{\text{epi}\phi}(x, \phi(x))\} \\ &\subset \{v | (-v, -1) \in \overline{N}_{\text{epi}\phi}(x, \phi(x))\}. \end{aligned}$$

Thus, $\partial\phi(x) = -\partial^+(-\phi)(x) \subset \bar{\partial}\phi(x) = -\bar{\partial}(-\phi)(x)$ and from (3.71),

$$0 \in (\bar{\partial} f(\bar{x}), 0) - \kappa(\bar{\partial}(-\phi)(\bar{x}), 0) + \kappa \bar{\partial}(-g)(\bar{x}, \bar{y}) + \overline{N}_\Omega(\bar{x}, \bar{y}). \quad (3.72)$$

Applying Proposition 3.4.6 to $-\phi$, there are $\lambda_i \geq 0, \bar{y}_i \in Y_0(\bar{x}), i = 1, \cdots, r$ such that $\sum_{i=1}^r \lambda_i = 1$ and

$$(\bar{\partial}(-\phi)(\bar{x}), 0) \subset \sum_{i=1}^r \lambda_i [\bar{\partial}(-g)(\bar{x}, \bar{y}_i) + \overline{N}_{\text{gph}Y}(\bar{x}, \bar{y}_i)]. \quad (3.73)$$

So, noting that $\bar{\partial}(-g) = -\bar{\partial} g$,

$$\begin{aligned} 0 &\in (\bar{\partial} f(\bar{x}), 0) + \kappa \sum_{i=1}^r \lambda_i \bar{\partial} g(\bar{x}, \bar{y}_i) - \kappa \bar{\partial} g(\bar{x}, \bar{y}) + \kappa \sum_{i=1}^r \lambda_i \overline{N}_{\text{gph}Y}(\bar{x}, \bar{y}_i) + \overline{N}_\Omega(\bar{x}, \bar{y}) \\ &\subset (\bar{\partial} f(\bar{x}), 0) + \kappa \sum_{i=1}^r \lambda_i [\bar{\partial} g(\bar{x}, \bar{y}_i) - \bar{\partial} g(\bar{x}, \bar{y})] + \sum_{i=1}^r \overline{N}_{\text{gph}Y}(\bar{x}, \bar{y}_i) + \overline{N}_\Omega(\bar{x}, \bar{y}). \end{aligned}$$

This completes the proof. □

3.5 Lower Order Penalization

In this section we consider the following GSIP problem

$$\min f(x) \quad \text{s.t.} \quad g(x,y) \leq 0, y \in Y_0 \cap Y(x), \tag{GSIP}$$

where Y_0 is a nonempty closed subset of $\mathbb{R}^m$, $Y(x) = \{y \in \mathbb{R}^m \mid v(x,y) \leq 0\}$ and the functions $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ and $v\colon \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^l$ are twice continuously differentiable in $x \in \mathbb{R}^n$. Let val(GSIP) be the optimal value of the problem (GSIP) and M_{GSIP} be the feasible set of (GSIP). For any $x \in \mathbb{R}^n$, the lower level problem associated with (GSIP) is

$$Q(x) \qquad \max_{y \in Y_0} g(x,y) \quad \text{s.t.} \quad v(x,y) \leq 0.$$

Let $\phi(x)$ be the optimal value of the problem $Q(x)$. It is clear that

$$x \in M_{GSIP} \quad \text{iff} \quad \phi(x) \leq 0.$$

As in Polak and Royset [110], we will associate (GSIP) with an SIP problem via the augmented Lagrangian of the lower level problem. Let $\bar{f}(x,\mu,c) = f(x)$ and the augmented Lagrangian of the lower level problem

$$\bar{g}(x,y,\mu,c) = g(x,y) - \frac{1}{2c}\sum_{i=1}^{l}\{([cv_i(x,y)+\mu_i]_+)^2 - \mu_i^2\}, (\mu,c) \in \mathbb{R}^l \times \mathbb{R}_{++}.$$

Consider the following SIP problem

$$\min_{(x,\mu,c)\in\mathbb{R}^n\times\mathbb{R}^l\times\mathbb{R}_{++}} \bar{f}(x,\mu,c) \quad \text{s.t.} \quad \bar{g}(x,y,\mu,c) \leq 0, y \in Y_0. \tag{SIP$_\mathrm{g}$}$$

Let val(SIP$_\mathrm{g}$) be the optimal value of (SIP$_\mathrm{g}$) and M_{SIP_g} be the feasible set of (SIP$_\mathrm{g}$).

For $0 < p \leq 1$, let the integral-type double penalty function of (GSIP) be defined by

$$G^p_{int}(x,\mu,c) := \bar{f}(x,\mu,c) + \rho\int_{Y_0} \bar{g}^p_+(x,y,\mu,c)\,\mathrm{d}y.$$

Since g and v are twice continuously differentiable in x, then $\bar{g}$ is $C^{1,1}$ in (x,μ,c), see Hiriart-Urruty et al. [57]. This property allows us to apply Theorem 2.4.4 to (SIP$_\mathrm{g}$). As for (SIP), combining with the exact penalization assumption of $G^p_{int}(x,\mu,c)$, we develop the first-order optimality conditions for (GSIP).

Before proceeding, we will investigate the relations between optimal solutions of (GSIP) and ($\mathrm{SIP_g}$). Let $\bar{H}(x,\mu,c) = \sup_{y\in Y_0} \bar{g}(x,y,\mu,c)$. Under the convention that $\sup\emptyset = -\infty$, $\phi(x) = -\infty$ if $Y_0 \cap Y(x) = \emptyset$. It is easy to see that

$$\bar{H}(x,\mu,c) \geq \phi(x), \forall\, (x,\mu,c).$$

Thus we obtain the relation between the optimal values of ($\mathrm{SIP_g}$) and (GSIP) as follows

$$\mathrm{val(SIPg)} \geq \mathrm{val(GSIP)}.$$

Let $\nu(x,u) = \max_{y\in Y_0} \hat{g}(x,y,u)$, where

$$\hat{g}(x,y,u) = \begin{cases} g(x,y), & \text{if } v(x,y) \leq u, \\ -\infty, & \text{otherwise.} \end{cases}$$

Then $\nu(x,0) = \phi(x)$ is the optimal value of the lower level problem $Q(x)$.

Next we recall some concepts from Rockafellar [114]. Problem $Q(x)$ is said to satisfy the *quadratic growth condition* if there is a $c \geq 0$ such that $\bar{g}(x,y,0,c)$ is bounded above as a function of $y \in Y_0$. Problem $Q(x)$ is said to be *stable of degree* q (a nonnegative integer) if there is a neighborhood U of the origin in R^l and a C^q function $\pi_x : U \to R$ such that

$$\nu(x,u) \leq \pi_x(u), \forall u \in U, \text{ and } \nu(x,0) = \pi_x(0).$$

Theorem 3.5.1 (Rockafellar [114])**.** *Under the quadratic growth condition,*

$$\phi(x) = \min_{(\mu,c)} \bar{H}(x,\mu,c)$$

*iff the problem $Q(x)$ is stable of degree 2. Note that here the symbol '*min*' represents that the minimum is attained.*

Therefore, we have the following equivalence proposition.

Proposition 3.5.1. *Assume that $Y_0 \cap Y(x) \neq \emptyset$ for all $x \in \mathbb{R}^n$. If for all $x \in \mathbb{R}^n$, $Q(x)$ satisfies the quadratic growth condition and is stable of degree 2, then problems GSIP and SIP$_g$ are of the same optimal value, i.e.,* $\mathrm{val}(GSIP) = \mathrm{val}(SIPg)$, *and furthermore,*

(1) *if $\bar{x}$ is a local optimal solution of* (GSIP) *in a neighborhood $B(\bar{x},\delta)$ with respect to the feasible set of GSIP, then there exists $\bar{\mu} \in \mathbb{R}^l$ and $\bar{c} \in R$ such that $(\bar{x},\bar{\mu},\bar{c})$ is a local optimal solution of* ($\mathrm{SIP_g}$) *in the neighborhood $B(\bar{x},\delta) \times \mathbb{R}^{l+1}$ with respect to the feasible set of* ($\mathrm{SIP_g}$)*;*

(2) *if* $(\bar{x}, \bar{\mu}, \bar{c})$ *is a local optimal solution of* ($\mathrm{SIP_g}$) *in a neighborhood* $B(\bar{x}, \delta) \times \mathbb{R}^{l+1}$ *w.r.t the feasible set of* ($\mathrm{SIP_g}$), *then* $\bar{x}$ *is a local optimal solution of* (GSIP) *in the neighborhood* $B(\bar{x}, \delta)$ *w.r.t. the feasible set of* (GSIP).

Proof. (1) Given a local solution $\bar{x}$ of (GSIP), we need to show that for any $(x, \mu, c) \in B(\bar{x}, \delta) \times \mathbb{R}^{l+1}$ feasible for ($\mathrm{SIP_g}$), we have $\bar{f}(x, \mu, c) \geq \bar{f}(\bar{x}, \bar{\mu}, \bar{c})$ for some $(\bar{\mu}, \bar{c})$. That is, for $(x, \mu, c) \in B(\bar{x}, \delta) \times \mathbb{R}^{l+1}$ satisfying $\bar{H}(x, \mu, c) \leq 0$, we have $f(x) \geq f(\bar{x})$. By Theorem 3.5.1, there exists (μ_x, c_x) such that $\bar{H}(x, \mu, c) \geq \bar{H}(x, \mu_x, c_x) = \phi(x)$. Especially, for $\bar{x}$, there is $(\bar{\mu}, \bar{c})$ satisfying the corresponding relations. Thus $\phi(x) \leq 0$. That is x is feasible for (GSIP). Then by the optimality of $\bar{x}$ for (GSIP), $f(x) \geq f(\bar{x})$ and this completes the proof.
(2) The proof can proceed similarly. For any $x \in B(\bar{x}, \delta)$ satisfying $\phi(x) \leq 0$, there is (μ_x, c_x) such that $\bar{H}(x, \mu_x, c_x) = \phi(x) \leq 0$, that is (x, μ_x, c_x) is feasible for ($\mathrm{SIP_g}$). Then $f(x) = \bar{f}(x, \mu_x, c_x) \geq \bar{f}(\bar{x}, \bar{\mu}, \bar{c}) = f(\bar{x})$. □

Note that it is enough to assume that the conditions, $Y_0 \cap Y(x) \neq \emptyset$, the quadratic growth condition and the stablility of degree 2, hold for all x near a given point $\bar{x}$, in the proposition above, to guarantee the local equivalence between (GSIP) and ($\mathrm{SIP_g}$). Under the assumptions of Proposition 3.5.1, we transform (GSIP) into an equivalent SIP problem.

Assume that Y_0 is compact. Let $Y_0^* = \{y \in Y_0 : \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}) = 0\}$ and

$$\begin{aligned} D(\bar{x}, \bar{\mu}, \bar{c}) := \{ & d = (d_1, d_2, d_3) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R} : \\ & \langle \nabla_x \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}), d_1 \rangle + \langle \nabla_\mu \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}), d_2 \rangle + \nabla_c \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}) d_3 \leq 0, y \in Y_0^* \}, \end{aligned}$$

where

$$\begin{aligned} \nabla_x \bar{g}(x, y, \mu, c) &= \nabla_x g(x, y) - \nabla_x^T v(x, y)[cv(x, y) + \mu]_+, \\ \nabla_\mu \bar{g}(x, y, \mu, c) &= -\frac{1}{c}([cv(x, y) + \mu]_+ - \mu), \\ \nabla_c \bar{g}(x, y, \mu, c) &= \frac{1}{2c^2} \sum_{i=1}^{l} \{([cv_i(x, y) + \mu_i]_+)^2 - \mu_i^2 - 2c[cv_i(x, y) + \mu_i]_+ v_i(x, y)\}. \end{aligned}$$

Proposition 3.5.2. *Let* $\bar{x}$ *be a local optimal solution of* (GSIP) *and the assumptions of Proposition 3.5.1 hold. Let* $(\bar{\mu}, \bar{c})$ *be the corresponding multiplier and penalty parameter* *as in (1) of Proposition 3.5.1. Then the cone $D(\bar{x}, \bar{\mu}, \bar{c})$ is of the following form*

$$\{d = (d_1, d_2, d_3) \in \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R} : \langle \nabla_x g(\bar{x}, y) - \nabla_x^T v(\bar{x}, y)\bar{\mu}, d_1\rangle \leq 0, y \in Y_0^*\}. \quad (3.74)$$

Proof. Under the assumption of stability of degree 2, given any $\tilde{y} \in Y_0^*$, $\tilde{y}$ solves the problem $\max_{y\in Y_0} \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c})$ for which the following holds

$$\min_{(\mu,c)} \max_{y\in Y_0} \bar{g}(\bar{x}, y, \mu, c) = \max_{y\in Y_0} \min_{(\mu,c)} \bar{g}(\bar{x}, y, \mu, c)) = \max_{y\in Y_0\cap Y(\bar{x})} g(\bar{x}, y).$$

Or equivalently, that $(\tilde{y}, \bar{\mu}, \bar{c}), \tilde{y} \in Y_0^*$ is the saddle point of the augmented Lagrangian $\bar{g}(\bar{x}, \cdot, \cdot, \cdot)$. That is also to say that each $\tilde{y} \in Y_0^*$ solves the lower level problem

$$\max\{g(\bar{x}, y) : y \in Y_0, v(\bar{x}, y) \leq 0\}$$

and the augmenting multiplier $\bar{\mu}$ is also the Lagrange multiplier of the lower level problem. By the first order necessary optimality conditions, for $y \in Y_0^*$,

$$0 \in \nabla_y[g(\bar{x}, y) - \bar{\mu}^T v(\bar{x}, y)] + N_{Y_0}(y), \langle \bar{\mu}, v(\bar{x}, y)\rangle = 0, v(\bar{x}, y) \leq 0, \bar{\mu} \geq 0.$$

Define the index set

$$I(y) = \{i \in \{1, \cdots, l\} : \bar{c}v_i(\bar{x}, y) + \bar{\mu}_i > 0\}. \quad (3.75)$$

It is easy to see that $I(y) = \{i \in \{1, \cdots, l\} : \bar{\mu}_i > 0\}$ and

$$[\bar{c}v_i(\bar{x}, y) + \bar{\mu}_i]_+ = \bar{\mu}_i, i = 1, \cdots, l, \quad (3.76)$$

and thus $\nabla_\mu \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}) = 0, \nabla_c \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}) = 0, \forall y \in Y_0^*$. □

Theorem 3.5.2. *Let the assumptions of Proposition 3.5.1 hold. Let $\bar{x}$ be a local optimal solution of* (GSIP) *and*

$$G^1_{max}(x, \mu, c) := \bar{f}(x, \mu, c) + \max_{y\in Y_0} \bar{g}_+(x, y, \mu, c)$$

be locally exact at $(\bar{x}, \bar{\mu}, \bar{c})$ where the pair $(\bar{\mu}, \bar{c})$ is obtained from Proposition 3.5.1. Then the following KKT-type optimality condition holds: there exist $\bar{\lambda}_j \geq 0$ not all zero, $\bar{y}^j \in Y_0^*$, $1 \leq j \leq n$ such that

$$\nabla_x f(\bar{x}) + \sum_{j=1}^{n} \bar{\lambda}_j[\nabla_x g(\bar{x}, \bar{y}^j) - \nabla_x^T v(\bar{x}, \bar{y}^j)\bar{\mu}] = 0. \quad (3.77)$$

Proof. By Theorem 2.4.3, there exists $\lambda_j \geq 0$ and $y^j \in Y_0^*$, $1 \leq j \leq n+l+1$ such that $\sum \lambda_j = 1$ and

$$\nabla_x \bar{f}(\bar{x}, \bar{\mu}, \bar{c}) + \sum_j \lambda_j \nabla_x \bar{g}(\bar{x}, y^j, \bar{\mu}, \bar{c}) = 0.$$

That is, by noting (3.76), $\nabla_x f(\bar{x}) + \sum_j \lambda_j [\nabla_x g(\bar{x}, y^j) - \nabla_x^T v(\bar{x}, y^j)\bar{\mu}] = 0$. As $\nabla_x f(\bar{x})$ is a n-dimensional vector, by Carathéodory's theorem, there exist $\bar{\lambda}_j \geq 0$ not all zero, $\bar{y}^j \in \{y^1, \cdots, y^{n+l+1}\}$, $1 \leq j \leq n$ such that (3.77) holds. □

Theorem 3.5.3. *Let the assumptions of Proposition 3.5.1 hold. Let $\bar{x}$ be a local optimal solution of* (GSIP) *and G_{int}^p be locally exact at the point $(\bar{x}, \bar{\mu}, \bar{c})$ where the pair $(\bar{\mu}, \bar{c})$ is obtained from Proposition 3.5.1. Then, under one of the following assumptions,*
(i) $0.5 < p \leq 1$,
(ii) $p = 0.5$ *and* $\bar{g}^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) \leq 0$ *for all* $d \in D(\bar{x}, \bar{\mu}, \bar{c})$ *and* $y \in Y_0^*$ *with* $\langle \nabla \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}), d \rangle = 0$,
(iii) $0 < p < 0.5$ *and* $\bar{g}^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) < 0$ *for all* $0 \neq d \in D(\bar{x}, \bar{\mu}, \bar{c})$ *and* $y \in Y_0^*$ *with* $\langle \nabla \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}), d \rangle = 0$, *we have*

$$\langle \nabla \bar{f}(\bar{x}, \bar{\mu}, \bar{c}), d \rangle \geq 0, \forall d \in D(\bar{x}, \bar{\mu}, \bar{c}). \tag{3.78}$$

Proof. For $p = 1$. As the exactness of integral-type penalty function implies that of the max-type penalty function, Theorem 2.4.3 of Chapter 2 implies that $\nabla \bar{f}$ at $(\bar{x}, \bar{\mu}, \bar{c})$ is a positive linear combination of $\{\nabla \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}) : y \in Y_0^*\}$ and thus (3.78) holds. The rest cases of the proof follow directly from Theorem 2.4.4 of Chapter 2. □

Corollary 3.5.1. *Let the assumptions of Theorem 3.5.3 hold. Assume that the convex cone generated by $\{\nabla \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}) : y \in Y_0^*\}$ is closed, so is the case when Y_0^* is of finite elements. Then the KKT-type optimiality condition holds, that is, there exist $\lambda_j \geq 0, \alpha_j \geq 0$ and $y_j \in Y_0^*, j = 1, \cdots, k$, such that $\nabla f(\bar{x}) + \sum_{j=1}^k \lambda_j \nabla_x L(\bar{x}, y_j, \alpha_j) = 0$.*

Proof. By Farkas Lemma, see, e.g., Hettich and Kortanek [53], and the closedness of the convex cone of $\{\nabla \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}) : y \in Y_0^*\}$, there are $y_j \in Y_0^*, \lambda_j \geq 0, j = 1, \cdots, k$ such that

$$\nabla_x f(\bar{x}) + \sum_{j=1}^k \lambda_j \nabla_x \bar{g}(\bar{x}, y_j, \bar{\mu}, \bar{c}) = 0,$$

$$\sum_{j=1}^k \lambda_j \nabla_\mu \bar{g}(\bar{x}, y_j, \bar{\mu}, \bar{c}) = 0, \sum_{j=1}^k \lambda_j \nabla_c \bar{g}(\bar{x}, y_j, \bar{\mu}, \bar{c}) = 0.$$

Let $\alpha_{ij} = [\bar{c}v_i(\bar{x}, y_j) + \bar{\mu}_i]_+, \alpha_j = (\alpha_{1j}, \cdots, \alpha_{lj})^T$, and $L(x, y, \alpha) = g(x, y) - \alpha^T v(x, y)$. So, the system is

$$\nabla f(\bar{x}) + \sum_{j=1}^{k} \lambda_j \nabla_x L(\bar{x}, y_j, \alpha_j) = 0,$$

$$\sum_{j=1}^{k} \lambda_j \alpha_j = \bar{\mu} \sum_{j=1}^{k} \lambda_j,$$

$$\textstyle\sum_{j=1}^{k} \lambda_j [g(\bar{x}, y_j) - \alpha_j^T v(\bar{x}, y_j)] = 0. \quad \square$$

Remark 3.5.1. *The $y_j, j = 1 \cdots, k$ are also the solutions of the problems*

$$\max_{y \in Y_0} \bar{g}(\bar{x}, y, \bar{\mu}, \bar{c}).$$

Then, by the first order optimality conditions, we have $\nabla_y \bar{g}(\bar{x}, y_j, \bar{\mu}, \bar{c}) \in N_{Y_0}(y_j)$, i.e.,

$$\nabla_y L(\bar{x}, y_j, \alpha_j) \in N_{Y_0}(y_j), j = 1, \cdots, k,$$

where N denotes the normal cone. If Y_0 is of the form

$$\{y \in R^m \mid h(y) \leq 0\},$$

with $h \colon \mathbb{R}^m \to \mathbb{R}^r$ differentiable and some regularity condition holds, then one can further have that for each j, there is a $\beta_j \in \mathbb{R}^r$ such that

$$\nabla_y L(\bar{x}, y_j, \alpha_j) - \nabla_y^T h(y_j) \beta_j = 0,$$
$$\beta_j \geq 0, \beta_j^T h(y_j) = 0.$$

Remark 3.5.2. *If $(\bar{x}, \bar{\mu}, \bar{c})$ solves G^p_{int}, then so does $(\bar{x}, \bar{\mu}, c)$ for any $c \geq \bar{c}$. Since for any $(x, y, \mu), c_1 \leq c_2$, $\bar{g}(x, y, \mu, c_1) \geq \bar{g}(x, y, \mu, c_2)$, it follows that $G^p_{\text{int}}(x, \mu, c_1) \geq G^p_{\text{int}}(x, \mu, c_2)$.*

With Proposition 3.5.2, we further have the following result.

Corollary 3.5.2. *Under the assumptions of Theorem 3.5.3, the optimality condition takes the following form*

$$\langle \nabla f(\bar{x}), d \rangle \geq 0, \ \ \forall d \in \mathbb{R}^n : \langle \nabla_x g(\bar{x}, y) - \nabla_x^T v(\bar{x}, y) \bar{\mu}, d \rangle \leq 0, y \in Y_0^*.$$

In Theorem 3.5.3, we need to estimate the generalized upper second-order directional derivative $\bar{g}^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d)$ for $d \in D(\bar{x}, \bar{\mu}, \bar{c})$. To do so, it is enough to calculate the generalized second-order directional derivatives of the two parts,

$$g(x, y) + \sum_{i=1}^{m} \frac{\mu_i^2}{2c} \text{ and } -\sum_{i=1}^{m} \frac{([cv_i(x, y) + \mu_i]_+)^2}{2c}.$$

The first one is C^2 and its generalized second-order directional derivative is easy to compute. For the second part, denote by $h_i(x, y, \mu, c) = \frac{([cv_i(x,y)+\mu_i]_+)^2}{2c}$ for $i = 1, \cdots, m$.

Proposition 3.5.3. *Let $d \in D(\bar{x}, \bar{\mu}, \bar{c})$ and $y \in Y_0^*$. Then*

$$\bar{g}^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) = d_1^T[\nabla_x^2 g(\bar{x}, y) - \sum_{i=1}^{l} \bar{\mu}_i \nabla_x^2 v_i(\bar{x}, y)]d_1 - \sum_{i\in I(y)} (\sqrt{\bar{c}} d_1^T \nabla_x v_i(\bar{x}, y) + \frac{d_{2i}}{\sqrt{\bar{c}}})^2 + \sum_{i=1}^{l} \frac{d_{2i}^2}{\bar{c}}. \tag{3.79}$$

Proof. Let $d \in D(\bar{x}, \bar{\mu}, \bar{c})$. We claim that

$$(-\sum_{i=1}^{m} h_i)^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) = -\sum_{i\in I(y)} (h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) = -\sum_{i\in I(y)} d^T \nabla^2 h_i(\bar{x}, y, \bar{\mu}, \bar{c})d \tag{3.80}$$

where the second equality holds since h_i is C^2 for $i \in I(y)$. On one hand,

$$\begin{aligned}(-\sum_{i=1}^{m} h_i)^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) &= -(\sum_{i=1}^{m} h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) \\ &\leq -\sum_{i\in I(y)} (h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) - (\sum_{i\notin I(y)} h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) \\ &= -\sum_{i\in I(y)} (h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d).\end{aligned}$$

On the other hand,

$$(\sum_{i=1}^{m} h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) = \sum_{i:\bar{c}v_i(\bar{x},y)+\bar{\mu}_i\neq 0} (h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) + (\sum_{i:\bar{c}v_i(\bar{x},y)+\bar{\mu}_i=0} h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d)$$

since the h_i's are C^2 for $i : \bar{c}v_i(\bar{x}, y) + \bar{\mu}_i \neq 0$. Next, we claim that the second term of the right hand side of the above equation vanishes. By definition of generalized second-order directional derivative, for any sequences $(x^\nu, \mu^\nu, c^\nu) \to \bar{\mu}, \lambda^\nu \downarrow 0$ as $\nu \to \infty$, the last term is no larger than

$$\liminf_{\nu\to\infty} \frac{\langle \nabla(\sum_{i:\bar{c}v_i(\bar{x},y)+\bar{\mu}_i=0} h_i)((x^\nu, y, \mu^\nu, c^\nu) + \lambda^\nu d) - \nabla(\sum_{i:\bar{c}v_i(\bar{x},y)+\bar{\mu}_i=0} h_i)(x^\nu, y, \mu^\nu, c^\nu), d\rangle}{\lambda^\nu},$$

which is equal to zero by choosing any sequences $(x^\nu, \mu^\nu, c^\nu) \to (\bar{x}, \bar{\mu}, \bar{c}), \lambda^\nu \downarrow 0$ such that for all i with $\bar{c}v_i(\bar{x}, y) + \bar{\mu}_i = 0$,

$$\frac{c^\nu v_i(x^\nu, y) + \mu_i^\nu}{2c^\nu} < 0 \ \text{ and } \ \frac{(c^\nu + \lambda^\nu d_3)v_i(x^\nu + \lambda^\nu d_1, y) + (\mu_i^\nu + \lambda^\nu d_{2i})}{2(c^\nu + \lambda^\nu d_3)} < 0,$$

which implies that for all ν, $h_i((x^\nu, y, \mu^\nu, c^\nu) + \lambda^\nu d) = 0$ and $h_i(x^\nu, y, \mu^\nu, c^\nu) = 0$. The sequences can be chosen as follows. Let $(x^\nu, \mu^\nu, c^\nu) = (\bar{x}, \mu^\nu, \bar{c})$ with $\mu_i^\nu \uparrow \hat{\mu}_i$ for all i with $\bar{c}v_i(\bar{x}, y) + \bar{\mu}_i = 0$ and $\mu_i^\nu = \bar{\mu}_i$ for the other i's. Then the first strict inequality satisfies. This also means that the continuous function $F_i(x, \mu, c) := \frac{cv_i(x,y)+\mu_i}{2c}$ is strictly less than 0 at $(\bar{x}, \mu^\nu, \bar{c})$. Thus, for given $d = (d_1, d_2, d_3) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}$, there is $\lambda^\nu \downarrow 0$ such that $F_i(\bar{x} + \lambda^\nu d_1, \mu^\nu + \lambda^\nu d_2, c^\nu + \lambda^\nu d_3) < 0$. This is just the second strict inequality. Combining these two parts, formula (3.80) follows.
Note that

$$\alpha_{\circ\circ}(x; d) = \begin{cases} 0, & \text{if } \ \beta(\text{x}) \le 0, \\ 2[\beta(x)d^T\nabla^2\beta(x)d + \langle \nabla\beta(x), d\rangle^2], & \text{if } \ \beta(\text{x}) > 0. \end{cases} \tag{3.81}$$

for a function $\alpha(x) := ([\beta(x)]_+)^2$ with β being C^2. Therefore, we have

$$(h_i)_{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) = \bar{\mu}_i d_1^T \nabla_x^2 v_i(\bar{x}, y)d_1 + \bar{c}\langle \nabla_x v_i(\bar{x}, y), d_1\rangle^2 + \frac{d_{2i}^2}{\bar{c}} + \frac{\bar{\mu}_i^2 d_3^2}{\bar{c}^3}$$
$$+ 2d_{2i}\langle \nabla_x v_i(\bar{x}, y), d_1\rangle - 2\frac{\bar{\mu}_i d_{2i} d_3}{\bar{c}^2}, i \in I(y).$$

The conclusion follows easily. □

Proposition 3.5.4. *Assume that the functions $g(\cdot, y)$ and $-v(\cdot, y)$ are concave for each y. If $I(y) = \{1, \cdots, l\}$ and $\langle \nabla_x v_i(x, y), d_1\rangle = 0$ for $d \in D(\bar{x}, \bar{\mu}, \bar{c}), y \in Y_0^*$ and $i \in I(y)$, then assumption (ii) of Theorem 3.5.3 holds. Moreover, if at least one of $g(\cdot, y), -v_i(\cdot, y), i = 1, \cdots, l$ is stongly concave, then assumption (iii) of Theorem 3.5.3 holds.*

Proof. Under our assumptions and the formula in Proposition 3.5.3, we have

$$\bar{g}^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) = d_1^T[\nabla_x^2 g(\bar{x}, y) - \sum_{i=1}^{l} \bar{\mu}_i \nabla_x^2 v_i(\bar{x}, y)]d_1.$$

As the concavity is equivalent to the negative semi-definiteness of its Hessian, inequality $\bar{g}^{\circ\circ}(\bar{x}, y, \bar{\mu}, \bar{c}; d) \le 0$ holds. Also, note that the strong concavity of a function H is equivalent to say that for some $\sigma > 0$, $H(\cdot) + \sigma\|\cdot\|^2$ is concave, see for example [116], and thus our assertion follows. □

Finally, we present below an example which verifies our theroem.

Example 3.5.1. *Consider the following GSIP problem*

$$\min_{x\in\mathbb{R}} x^3 \ \text{s.t.} \ -y \le 0, y \in Y_0 \cap Y(x),$$

with $Y_0 = [-1,1], Y(x) = \{y \in \mathbb{R} : x^3 - y \le 0\}$. *Then* $M = \{x : x \ge 0\}$ *and* $Y_0 \cap Y(x) = \emptyset$, if $x > 1$. *Define* $\bar{g}(x,y,\mu,c) = -y - \frac{1}{2c}([c(x^3-y)+\mu]_+^2 - \mu^2)$ *and the double penalty function*

$$G_{\text{int}}^p(x,\mu,c) = x^3 + \rho\int_{-1}^{1}[\bar{g}(x,y,\mu,c)]_+^p \,\mathrm{d}y.$$

The solution is unique and the KKT multiplier for the lower level problem is unique, equal to $\{1\}$.

The value function of the perturbed problem of the lower level problem $Q(x)$ *is* $\nu(x,u) = u - x^3$ *for all points near* $(x,u) = (0,0)$, *and is twice continuously differentiable. Thus the lower level problem is stable of degree* 2 *(in a neighborhood of the origin). Therefore, the original GSIP can be equivalently transformed into an SIP in a neighborhood of the origin.*

We will verify that penalty function G_{int}^1 *is not locally exact but* $G_{\text{int}}^{\frac{1}{2}}$ *is locally exact at* $(\bar{x},\bar{\mu},\bar{c}) = (0,1,\bar{c})$ *for some* $\bar{c} > 0$. *For this we only need to consider points near* $(0,1,\bar{c})$. *It is also enough to consider the case when* $x \le 0$.

Let $x \le 0$. *The effective integral interval of* y *consists of the two parts:* $A := \{y : x \le 0, c(x^3-y)+\mu \le 0, \bar{g}(x,y,\mu,c) \ge 0\}$ *and* $B := \{y : x \le 0, c(x^3-y)+\mu \ge 0, \bar{g}(x,y,\mu,c) \ge 0\}$. *Then* $A = \{y : x \le 0, x^3+\mu/c \le y \le \mu^2/(2c)\}$ *and* $B = \{y : x \le 0, y \le x^3+\mu/c, y \in [x^3-(1-\mu+\sqrt{(1-\mu)^2-2cx^3})/c, x^3-(1-\mu-\sqrt{(1-\mu)^2-2cx^3})/c]\}$. *By choosing* $\bar{c}$ *large enough, and since* (x,μ,c) *is in a neighborhood of* $(0,1,\bar{c})$, *then* $A = \emptyset$ *and* $B = \{y : x \le 0, y \in [x^3-(1-\mu+\sqrt{(1-\mu)^2-2cx^3})/c, x^3-(1-\mu-\sqrt{(1-\mu)^2-2cx^3})/c]\}$. *So,*

$$\begin{aligned} G_{\text{int}}^p(x,\mu,c) &= x^3 + \rho\int_{x^3-(1-\mu+\sqrt{(1-\mu)^2-2cx^3})/c}^{x^3-(1-\mu-\sqrt{(1-\mu)^2-2cx^3})/c}[-\frac{c}{2}y^2 - (1-\mu-cx^3)y - \frac{c}{2}x^6 - x^3\mu]^{\frac{1}{p}}\,\mathrm{d}y \\ &= x^3 + \rho(\frac{c}{2})^{\frac{1}{p}}\int_{-\sqrt{(1-\mu)^2-2cx^3}/c}^{\sqrt{(1-\mu)^2-2cx^3}/c}(\frac{(1-\mu)^2-2cx^3}{c^2} - y^2)^{\frac{1}{p}}\,\mathrm{d}y. \end{aligned}$$

It is easy to calculate that

$$G^1_{\text{int}}(x,1,\bar c) = x^3 + \frac{\rho\bar c}{2}\int_{-\sqrt{-2\bar c x^3}/\bar c}^{\sqrt{-2\bar c x^3}/\bar c} -\frac{2x^3}{\bar c} - y^2\,\mathrm{d}y = x^3 + \frac{4\rho}{3}\cdot(\frac{2}{\bar c})^{\frac12}\cdot(-x)^{\frac92} = x^3 + \rho\cdot o(|x^3|)$$

$$\begin{aligned} G^{\frac12}_{\text{int}}(x,\mu,c) &= x^3 + \rho\cdot\sqrt{\frac{c}{2}}\cdot\frac{(1-\mu)^2 - x^3}{c^2}\cdot\int_{-1}^{1}\sqrt{1-y^2}\,\mathrm{d}y \\ &= x^3 + \rho\cdot\sqrt{\frac{c}{2}}\cdot\frac{(1-\mu)^2 - x^3}{c^2}\cdot\frac{\pi}{2} \\ &\ge x^3 + \rho\cdot\sqrt{\frac{c}{2}}\cdot\frac{1}{c^2}\cdot\frac{\pi}{2}\cdot(-x^3). \end{aligned}$$

Then $G^{\frac12}_{\text{int}}(x,\mu,c)\ge 0$ *near* $(0,1,\bar c)$*. Thus* G^1_{int} *is not exact and* $G^{\frac12}_{\text{int}}$ *is exact.*

We also have $D(\bar x,\bar\mu,\bar c) = \mathbb{R}^3$ *and* $\bar g^{\circ\circ}(\bar x, y,\bar\mu,\bar c; d)\equiv 0$*. Hence, all the conditions in the Theorem 3.5.3 (ii) are satisfied and thus the optimality conditions for GSIP hold.*

3.6 GSIP with Convex Lower Level Problem

In this section we consider a special class of GSIP problems, the one with convex lower level problems. In particular, we consider the following GSIP problem

$$\min f(x) \quad \text{s.t.} \quad g(x,y)\le 0, y\in Y(x), \tag{3.82}$$

where

$$Y(x) := \{y\in\mathbb{R}^m \mid v_i(x,y)\le 0, i=1,\cdots,l\},$$

and the real valued functions $-g(x,\cdot)$ and $v_i(x,\cdot), i=1,\cdots,l$ are convex.
Still, denote by $Q(x)$ the lower level problem

$$\max g(x,y) \quad \text{s.t.} \quad y\in Y(x). \tag{3.83}$$

Definition 3.6.1 (generalized augmenting function [60])**.** *A function* $\sigma\colon\mathbb{R}^m\to\mathbb{R}_+\cup\{\infty\}$ *is said to be a generalized augmenting function if it is proper, lower semicontinuous, level-bounded on* $\mathbb{R}^m$*,* $\arg\min_y\sigma(y)=\{0\}$*, and* $\sigma(0)=0$*.*

Definition 3.6.2 (NCP function)**.** *A function* $\phi\colon\mathbb{R}^2\to\mathbb{R}$ *is called an NCP function if it satisfies that*

$$\phi(a,b)=0 \quad \textit{if and only if} \quad a\ge 0, b\ge 0, ab=0.$$

The existence of a C^∞ NCP function is clear from a theorem by Whitney [10]. But smooth NCP functions are usually degenerate at the origin. Some common NCP functions include the natural residual(NR) or min-function [79]

$$\phi^{NR}(a,b) = \frac{1}{2}(a + b - \sqrt{(a-b)^2}) = \min\{a,b\},$$

the Fischer-Burmeister (FB) function [32]

$$\phi^{FB}(a,b) = a + b - \sqrt{a^2+b^2},$$

and also the penalized Fischer-Burmeister NCP-function [18], for $\lambda \in [0,1]$,

$$\phi_\lambda(a,b) = \lambda\phi^{FB}(a,b) + (1-\lambda)a_+b_+.$$

Smoothing functions of NR and FB NCP functions are respectively

$$\phi(a,b,\tau) = \frac{1}{2}(a + b - \sqrt{(a-b)^2 + 4\tau^2}),$$

$$\phi(a,b,\tau) = a + b - \sqrt{a^2+b^2+2\tau^2}.$$

If ϕ is either an NR or FB NCP function, then

$$\phi(a,b,\tau) = 0 \quad \text{if and only if} \quad a \geq 0, b \geq 0, ab = \tau^2.$$

For a zero (a,b) of $\phi(\cdot,\cdot,\tau)$ the gradient $\nabla\phi(\cdot,\cdot,\tau)$ w.r.t. (a,b) does not explicitly depend on τ and is given by $(a+b)^{-1}(b,a)$.

Next we describe two general assumptions for the convex lower level problem $Q(x)$.

Assumption 3.6.1. *The functions $-g(x,\cdot)$ and $v_i(x,\cdot), i = 1,\cdots,l$, are convex.*

Assumption 3.6.2 (Slater CQ)**.** *Given $x \in \mathbb{R}^n$, there is a $y \in \mathbb{R}^m$ such that $v_i(x,y) < 0$, for all $i = 1,\cdots,l$.*

Proposition 3.6.1 ([33])**.** *Let assumptions 3.6.1 and 3.6.2 hold for a given $x \in \mathbb{R}^n$. Then $y \in Y(x)$ is a global solution of the lower level problem $Q(x)$ if and only if there are $\gamma_i, i = 1,\cdots,l$, such that*

$$\nabla_y g(x,y) - \sum_{i=1}^{l} \gamma_i \nabla_y v_i(x,y) = 0, \gamma_i v_i(x,y) = 0, \gamma_i \geq 0, v_i(x,y) \leq 0, i = 1,\cdots,l. \tag{3.84}$$

It follows that the feasible set of the problem (3.82) is closed, see, e.g., (iv) of Proposition 3.3.1. Thus under assumptions 3.6.1 and 3.6.2 for each $x \in \mathbb{R}^n$, the GSIP problem (3.82) can be equivalently reformulated as

$$\begin{aligned} \min_{x,y,\gamma} \quad & f(x) \\ s.t. \quad & g(x,y) \le 0, \\ & \nabla_y g(x,y) - \sum_{i=1}^{l} \gamma_i \nabla_y v_i(x,y) = 0, \\ & \gamma_i v_i(x,y) = 0, \gamma_i \ge 0, v_i(x,y) \le 0, i = 1, \cdots, l. \end{aligned} \tag{3.85}$$

The above problem (3.85) is usually referred to as the mathematical program with complementarity (or equilibrium) constraints. The NCP function is usually introduced to reformulate the complementarity constraints. Let

$$\Phi(\gamma, -v(x,y), u) := (\phi(\gamma_1, -v_1(x,y), u_1), \cdots, \phi(\gamma_l, -v_l(x,y), u_l)^T.$$

Then the above mathematical program with complementarity constraints is reformulated as perturbed problem GSIP(u)

$$\begin{aligned} \min \quad & f(x) \\ s.t. \quad & g(x,y) \le 0, \\ & \nabla_y g(x,y) - \sum_{i} \gamma_i \nabla_y v_i(x,y) = 0, \\ & \Phi(\gamma, -v(x,y), u) = 0. \end{aligned} \tag{GSIP(u)}$$

Define $G(x,y,\gamma,u)\colon \mathbb{R}^{n+m+2l} \to \mathbb{R}^{1+m+l}$ by

$$G(x,y,\gamma,u) := \begin{pmatrix} g(x,y) \\ \nabla_y g(x,y) - \sum_i \gamma_i \nabla_y v_i(x,y) \\ \Phi(\gamma, -v(x,y), u) \end{pmatrix}. \tag{3.86}$$

Then GSIP(u) can be written as the following parameterized unconstrained optimization problem

$$\inf_{(x,y,\gamma)} F(x,y,\gamma,u), \tag{3.87}$$

where

$$F(x,y,\gamma,u) := f(x) + \delta_{\mathbb{R}_- \times \mathbf{0}_{\{m+l\}}}(G(x,y,\gamma,u)), \tag{3.88}$$

and let $p(u)$ be its optimal value function. Thus $p(0)$ is the optimal value of problem (3.82). The *generalized augmented Lagrangian* $l : \mathbb{R}^n \times \mathbb{R}^l \times \mathbb{R}^l \times (0, \infty) \to \bar{\mathbb{R}}$ as

$$l(x, y, \gamma, \omega, r) := \inf\{F(x, y, \gamma, u) + r\sigma(u) - \langle \omega, u \rangle : u \in \mathbb{R}^l\}, \tag{3.89}$$

where σ is a generalized augmenting function, and the *generalized augmented Lagrangian dual function* is defined as

$$\psi(\omega, r) := \inf_{(x,y,\gamma)} l(x, y, \gamma, \omega, r), \qquad \omega \in \mathbb{R}^l, y > 0. \tag{3.90}$$

Then both $l(x, y, \gamma, \omega, r)$ and $\psi(\omega, r)$ are concave and upper semi-continuous in $(\omega, r) \in \mathbb{R}^l \times \mathbb{R}_+$ and nondecreasing in r since they are pointwise infima of a collection of affine functions of (ω, r) which are nondecreasing in r. The *generalized augmented Lagrangian dual problem* is defined as

$$P(\omega, r) \quad \inf\{F(x, y, \gamma, u) + r\sigma(u) - \langle \omega, u \rangle : (x, y, \gamma, u) \in \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^l \times \mathbb{R}^l\}. \tag{3.91}$$

Let S and $V(\omega, r)$ be the solution sets of the problems (3.82) and $P(\omega, r)$, respectively. Note that $p(0)$ and $\psi(\omega, r)$ are the optimal values of the problems (3.82) and $P(\omega, r)$, respectively.

Besides, we have

$$\sup_{(\omega,r)} l(x, y, \gamma, \omega, r) = \begin{cases} f(x), & \text{if } (x, y, \gamma) \text{ is feasible for problem (3.85)}, \\ +\infty, & \text{otherwise.} \end{cases}$$

That is to say, the optimal value $p(0)$ of the GSIP problem (3.82) satisfies

$$p(0) = \inf_{(x,y,\gamma)} \sup_{(\omega,r)} l(x, y, \gamma, \omega, r). \tag{3.92}$$

The weak duality holds:

$$\psi(\omega, r) \leq p(0), \forall (\omega, r) \in \mathbb{R}^l \times (0, \infty), \tag{3.93}$$

since for any $(x, y, \gamma), l(x, y, \gamma, \omega, r) \leq F(x, y, \gamma, 0) = f(x)$.

Theorem 3.6.1. *Let the following assumptions hold*

(1) *The feasible set of the GSIP problem* (3.82) *is nonempty;*

(2) *For any* $\alpha \in \mathbb{R}$, *the set* $\{(x,y) : F(x,y,\gamma,u) \leq \alpha$ *for some* $\gamma \in \mathbb{R}^l\}$ *is bounded locally uniform in* u*;*

(3) *The vectors* $\{\nabla_y v_i(x,y) : v_i(x,y) = 0\}$ *are linearly independent for any* $y \in Y(x)$, *any* $x \in R^n$*;*

(4)

$$\exists(\bar{\omega},\bar{r}) : \inf_{(x,y,\gamma)} l(x,y,\gamma,\bar{\omega},\bar{r}) > -\infty. \tag{3.94}$$

Then

(i) S *is nonempty and compact;*

(ii) *for any* $r \geq \bar{r}+1$, $V(\bar{\omega},r)$ *is nonempty and compact, where* $(\bar{\omega},\bar{r})$ *is a pair meeting*

$$F(x,y,\gamma,u) + \bar{r}\sigma(u) - \langle\bar{\omega},u\rangle \geq m_0, \forall(x,y,\gamma,u), \tag{3.95}$$

for some $m_0 \in \mathbb{R}$*;*

(iii) *for each selection* $(x(r),y(r),\gamma(r),u(r)) \in V(\bar{\omega},r)$, *with* $r \geq \bar{r}+1$, *the optimal path* $(x(r),y(r),\gamma(r),u(r))$ *is bounded and its limit takes the form* $(x^*,y^*,\gamma^*,0)$, *where* $x^* \in S$*;*

(iv) $p(0) = \lim_{r\to\infty}\psi(\bar{\omega},r)$*;*

(v) *zero duality gap holds:*

$$p(0) = \sup_{(\omega,r)} \psi(\omega,r).$$

Proof. (i) By definition of F, we have F is lower semi-continuous. Since $F(x,y,\gamma,u)$ is proper and level-bounded in (x,y) locally uniformly in u, $F(\cdot,\cdot,\cdot,0)$ is also proper and level bounded in (x,y,γ). Then the solution set of the problem (3.85) is nonempty compact. Therefore, the set S as the projection of the solution set of problem (3.85) is also nonempty and compact.

(ii) Let $\bar{x}$ be such that $f(\bar{x})$ is finite. Let

$$U(r) := \{(x,y,\gamma,u) : F(x,y,\gamma,u) + r\sigma(u) - \langle\bar{\omega},u\rangle \leq f(\bar{x})\}.$$

It is obvious $U(r) \neq \emptyset$ and closed. Then $U(\bar{r}+1)$ is compact. If not, then assume that there exists a sequence $\{(x^k,y^k,\gamma^k,u^k)\} \subseteq U(\bar{r}+1)$ such that $\|(x^k,y^k,\gamma^k,u^k)\| \to \infty$.

Since $(x^k, y^k, \gamma^k, u^k) \subseteq U(\bar{r}+1)$, we have

$$F(x^k, y^k, \gamma^k, u^k) + \bar{r}\sigma(u^k) - \langle \bar{\omega}, u^k \rangle + \sigma(u^k) \leq f(\bar{x}). \tag{3.96}$$

Combining with (3.95), we get

$$\sigma(u^k) \leq f(\bar{x}) - m_0.$$

By the level boundedness of σ, we have that $\{u^k\}$ is bounded. Assume, or just by taking a subsequence if necessary, $u^k \to \bar{u}$. From (3.96), we have

$$F(x^k, y^k, \gamma^k, u^k) \leq f(\bar{x}) + \langle \bar{\omega}, u^k \rangle.$$

As F is level-bounded in (x, y) locally uniform in u, then $\{(x^k, y^k)\}$ is bounded. We may assume that $\{(x^k, y^k)\} \to (\bar{x}, \bar{y})$. $F(x^k, y^k, \gamma^k, u^k) < \infty$ is equivalent to

$$g(x^k, y^k) \leq 0, \tag{3.97}$$

$$\nabla_y g(x^k, y^k) - \sum_i \gamma_i^k \nabla_y v_i(x^k, y^k) = 0, \tag{3.98}$$

$$v_i(x^k, y^k) \leq 0, \gamma_i^k \geq 0, \tag{3.99}$$

$$v_i(x^k, y^k)\gamma_i^k = -(u_i^k)^2, i = 1, \cdots, l. \tag{3.100}$$

Then we have $g(\bar{x}, \bar{y}) \leq 0, v(\bar{x}, \bar{y}) \leq 0$.
Since (LICQ) holds for any point $y \in Y(\bar{x})$, then we have $\{\gamma^k\}$ is also bounded. Otherwise, let $\gamma^k \to \infty$ and $\frac{\gamma^k}{|\gamma^k|} \to \gamma, |\gamma| = 1$. Thus, from $v_i(x^k, y^k)\frac{\gamma_i^k}{|\gamma^k|} = -(u_i^k)^2 \frac{1}{|\gamma^k|}$, by letting $k \to \infty$, we have $\gamma_i = 0$ for inactive index i with $v_i(\bar{x}, \bar{y}) \neq 0$. For $\nabla_y g(x^k, y^k) - \sum_i \gamma_i^k \nabla_y v_i(x^k, y^k) = 0$, dividing by $|\gamma^k|$, and letting $k \to \infty$, we get $\sum_{i:v_i(\bar{x},\bar{y})=0} \gamma_i \nabla_y v_i(\bar{x}, \bar{y}) = 0$, contradicting the (LICQ) assumption. Then $\{(x^k, y^k, \gamma^k, u^k)\}$ is bounded and therefore $U(\bar{r}+1)$ is compact. It follows easily that, for any $r \geq \bar{r}+1$, $V(\bar{\omega}, r)$ is nonempty and compact.
Proofs of (iii)-(v) follow similarly as in Huang and Yang [60].
(iii) Given optimal path $(x(r), y(r), \gamma(r), u(r)), r \to \infty$, due to the compactness property proved in (ii), we may assume that a sequence $(x(r^k), y(r^k), \gamma(r^k), u(r^k))$, as $r^k \to \infty$, converges to the point $(x^*, y^*, \gamma^*, u^*)$. As was done in [60], we can check that $u^* = 0$, x^* is feasible for problem (3.82) and $x^* \in S$. □

Remark 3.6.1. *The perturbed feasible system we considered in Theorem 3.6.1 is*

$$\begin{aligned}
&g(x,y) \leq 0,\\
&\nabla_y g(x,y) - \sum_{i=1}^{l} \gamma_i \nabla_y v_i(x,y) = 0,\\
&-\gamma_i v_i(x,y) = u_i^2, \gamma_i \geq 0, v_i(x,y) \leq 0, i = 1, \cdots, l.
\end{aligned}$$

The more general perturbed system is

$$\begin{aligned}
&g(x,y) \leq \alpha,\\
&\nabla_y g(x,y) - \sum_{i=1}^{l} \gamma_i \nabla_y v_i(x,y) = \beta,\\
&-\gamma_i v_i(x,y) = u_i^2, \gamma_i \geq 0, v_i(x,y) \leq 0, i = 1, \cdots, l.
\end{aligned}$$

Replacing the parameter u by (α, β, u), the assertions of Theorem 3.6.1 still hold. Note that if we also perturb the constraints $v_i(x,y) \leq 0, i = 1, \cdots, l$, the same proofs won't hold and thus we are not sure that same results hold or not.

Algorithmically, Stein and Winterfeld [134] considered the following system of inner approximation of the feasible set M

$$\begin{aligned}
&g(x,y) + lu^2 \leq 0,\\
&\nabla_y g(x,y) - \sum_{i=1}^{l} \gamma_i \nabla_y v_i(x,y) = 0,\\
&-\gamma_i v_i(x,y) = u^2, \gamma_i \geq 0, v_i(x,y) \leq 0, i = 1, \cdots, l,
\end{aligned}$$

where l is the number of lower level constraints and u is a one-dimensional parameter.

Remark 3.6.2. *If $\lim_{\|(x,y)\| \to \infty} \max\{f(x), g(x,y), v_1(x,y), \cdots, v_l(x,y)\} = \infty$, then the boundedness condition (2) of Theorem 3.6.1 holds.*

Remark 3.6.3. (i) *If $\inf_{x \in \mathbb{R}^n} f(x) > -\infty$, then the condition (3.94) is satisfied with $(\bar{\omega}, \bar{r}) = (0, r)$, for any $r \in (0, \infty)$.*

(ii) *Let $\sigma(u) = \|u\|$.*

(a) *If for some $L > 0$, $p(u) \geq p(0) - L\|u\|$ for all u, then (3.94) holds. Indeed, under such conditions,*

$$\inf_{(x,y,\gamma)} l(x,y,\gamma,\omega,r) = \inf_u \{p(u) + r\sigma(u) - \langle \omega, u \rangle\} \geq \inf_u \{p(0) + (r - L - \|\omega\|)\|u\|\}.$$

So choosing r large enough for given ω such that $r - L - \|\omega\| > 0$ will fulfil the condition (3.94).
Especially, if $f(x) \geq p(0) - L\|u\|$ for all (x, y, γ, u) feasible for $GSIP_u$, then condition (3.94) holds.

(b) *If there exist $\alpha \geq 0, \beta \in \mathbb{R}$, $L > 0$ and a neighborhood V_0 of $u = 0$ such that $p(u) \geq p(0) - L\|u\|$ for all $u \in V_0$, and $p(u) \geq -\alpha\|u\| + \beta$ for all u, then (3.94) holds. It suffices to show that these conditions are equivalent to the one in (a).*

(c) *The following condition is also sufficient for (3.94): for all $\epsilon > 0$, there exists $(x(\epsilon), y(\epsilon), \gamma(\epsilon))$ such that*

$$F(x, y, \gamma, u) - F(x(\epsilon), y(\epsilon), \gamma(\epsilon), 0) + \epsilon \geq -L\|u\|, \forall(x, y, \gamma), \forall u.$$

(iii) *Conditions in (a)–(c) of (ii) are also necessary.*

(iv) *Likewise, for $\sigma(u) = \|u\|^2 = \sum u_i^2$ being the quadratic augmenting function, we can obtain some necessary and sufficient conditions for (3.94).*
(1) If for some u^* and some neighborhood V of $u = 0$, $p(u) - p(0) \geq \langle u^*, u\rangle - c\|u\|^2$ holds for all $u \in V$ and for some $\alpha > 0$ and $\beta \in \mathbb{R}$, $p(u) \geq -\alpha\|u\|^2 + \beta$ holds for all u, then (3.94) holds.
(2) If there exists u^* such that $p(u) - p(0) \geq \langle u^*, u\rangle - c\|u\|^2$ for all u, then (3.94) holds.

Just consider the following two conditions of any given function ϕ:

(a) *The proximal subgradients $\partial_p\phi(0) \neq \emptyset$, i.e., there exist u^* and some neighborhood V of $u = 0$ such that $\phi(u) - \phi(0) \geq \langle u^*, u\rangle - c\|u\|^2$ for all $u \in V$, and for some $\alpha > 0$ and $\beta \in \mathbb{R}$, $\phi(u) \geq -\alpha\|u\|^2 + \beta$ for all u.*

(b) *there exists u^* such that $\phi(u) - \phi(0) \geq \langle u^*, u\rangle - c\|u\|^2$ for all u.*

Then (a) $\Leftrightarrow$ (b):
Let (a) hold. Take any $u^* \in \partial_p\phi(0)$ and $V(0)$ specified in (a). If (b) fails, then there is $\{u_i\} \notin V(0)$ (which implies that $u_i \neq 0$) such that

$$\phi(u_i) - \phi(0) \leq \langle u^*, u_i\rangle - i\|u_i\|^2, \forall i.$$

Then by assumption of (a),

$$-\alpha\|u_i\|^2 + \beta \leq \langle u^*, u_i\rangle - i\|u_i\|^2, \forall i.$$

That is $(i-\alpha)\|u_i\|^2+\beta \le \langle u^*, u_i\rangle, \forall i$, *a contradiction. Conversely, if (b) holds, then it is clear that* $\partial_p\phi(0) \neq \emptyset$. *Pick* $\alpha > c$ *so large that* $-\|u^*\|\|u\| \ge -\alpha\|u\|^2, \forall u : \|u\| > 1$. *So taking* $\beta = -\|u^*\|$, *(a) holds.*

Exact penalization

Definition 3.6.3. *A vector* $\bar{\omega}$ *is said to support an exact penalty representation for GSIP problem* (3.82) *if, for all* $r > 0$ *sufficiently large, the problem is equivalent to minimizing* $l(x, y, \gamma, \bar{\omega}, r)$ *with respect to* (x, y, γ) *in the sense that*

$$\inf_{x\in M} f(x) = \inf_{(x,y,\gamma)} l(x, y, \gamma, \bar{\omega}, r), \mathrm{argmin}_x f(x) = \mathrm{pr}(\mathrm{argmin}_{(x,y,\gamma)} l(x, y, \gamma, \bar{\omega}, r)),$$

where M *denotes the feasible set of problem* (3.82) *and* pr *is the projection of* (x, y, γ) *onto the first argument.*

Theorem 3.6.2. (i) *If* $\bar{\omega}$ *supports an exact penalty representation for GSIP problem* (3.82)*, then there exists* $\hat{r} > 0$ *and a neighborhood* W *of* $0 \in \mathbb{R}^l$ *such that*

$$p(u) \ge p(0) + \langle \bar{\omega}, u\rangle - \hat{r}\sigma(u), \qquad \text{for all } u \in W. \tag{3.101}$$

(ii) *The converse is true if, in addition,*

(a) *p(0) is finite;*

(b) *there is an* $\bar{r}' > 0$ *such that*

$$\inf\{F(x, y, \gamma, u) + \bar{r}'\sigma(u) - \langle \bar{\omega}, u\rangle : (x, y, \gamma, u)\} > -\infty;$$

(c) *there is* $\delta > 0$ *and* $N > 0$ *such that* $\sigma(u) \ge \delta\|u\|$ *when* $\|u\| \ge N$.

Remark 3.6.4. *Assumption (c) is easy to be satisfied: replacing any given augmenting function* σ *by* $\sigma + \|\cdot\|$ *is enough.*

Proof. Note that $(\bar{\omega}, \bar{r}) \in \mathrm{argmax}\, \psi(\omega, r)$ is equivalent to the condition

$$p(u) \ge p(0) + \langle \bar{\omega}, u\rangle - \bar{r}\sigma(u), \forall u \in \mathbb{R}^l, \tag{3.102}$$

holding. The proof of (i) is easy. The exact penalty representation implies that $\psi(\bar{\omega}, r) = p(0)$, for any $r \ge \bar{r}$, where $\bar{r}$ assumes as a penalty threshold. Since $\psi(\omega, r)$ is usc, $\psi(\bar{\omega}, \bar{r}) \ge p(0)$. Combining with Theorem 3.6.1, we have $(\bar{\omega}, \bar{r})$ maximizes $\psi(\omega, r)$. For the proof of (ii), when $\bar{r}$ is replaced by any $r > \bar{r}$ in condition (3.102),

$$\mathrm{argmin}_u\{p(u) - \langle \bar{\omega}, u\rangle + r\sigma(u)\} = \{0\}.$$

As was done in [116], for such r, by defining function

$$g(x, y, \gamma, u) := F(x, y, \gamma, u) + r\sigma(u) - \langle \bar{\omega}, u \rangle$$

and its associated functions $h(u) := \inf_{(x,y,\gamma)} g(x, y, \gamma, u)$ and $k(x, y, \gamma) := \inf_u g(x, y, \gamma, u)$, and under assumption (3.102), we can obtain that $\bar{\omega}$ supports an exact penalty representation.

Then, to complete the proof, it is enough to show that condition (3.101) implies the stronger condition (3.102). This follows just the same as [116]. $\square$

As in Huang and Yang [60], for the special case that $\bar{\omega} = 0$ supports an exact penalty representation, the condition (c) of theorem 3.6.2 can be dismissed.

Theorem 3.6.3. (i) *If $\bar{\omega} = 0$ supports an exact penalty representation for GSIP problem* (3.82)*, then there exists $\hat{r} > 0$ and a neighborhood W of $0 \in \mathbb{R}^l$ such that*

$$p(u) \geq p(0) - \hat{r}\sigma(u), \qquad \text{for all } u \in W. \tag{3.103}$$

(ii) *The converse is true if, in addition,*

(a) *p(0) is finite;*

(b) *there is an $\bar{r}' > 0$ such that*

$$\inf\{F(x, y, \gamma, u) + \bar{r}'\sigma(u) : (x, y, \gamma, u)\} > -\infty.$$

Proof. (i) follows directly from Theorem 3.6.2. (ii) First, argue that there is some $\bar{r} > 0$, for all $r \geq \bar{r}$, $\inf_{x\in M} f(x) = \inf_{(x,y,\gamma)} l(x, y, \gamma, 0, r)$. Assume the contrary, for some $r_k \to \infty$, $\inf_{x\in M} f(x) > \inf_{(x,y,\gamma)} l(x, y, \gamma, 0, r_k)$. Then there is $(x^k, y^k, \gamma^k, u^k)$ such that

$$p(0) > F(x^k, y^k, \gamma^k, u^k) + r_k\sigma(u^k) \geq m_0 + (r_k - \bar{r}')\sigma(u^k), \tag{3.104}$$

where m_0 is the lower bound of $\inf\{F(x, y, \gamma, u) + \bar{r}'\sigma(u) : (x, y, \gamma, u)\}$. By the level-boundedness of σ, $\{u^k\}$ is bounded and $\sigma(u^k) < \frac{p(0)}{(r_k - \bar{r}')}$. Say $\bar{u}$ is a cluster point of $\{u^k\}$, then $\bar{u} = 0$. From (3.104), we also have the inequality $p(0) > p(u^k) + r_k\sigma(u^k)$ which contradicts (3.103). Thus $\inf_{x\in M} f(x) = \inf_{(x,y,\gamma)} l(x, y, \gamma, 0, r)$ holds.

For any $x^* \in \operatorname{argmin}_{x\in M} f(x)$, there is (y^*, γ^*) such that (x^*, y^*, γ^*) solves problem (3.85). From (3.103), (x^*, y^*, γ^*) also solve that $\inf l(x, y, \gamma, 0, r)$ for all $r \geq \hat{r}$. That is $x^* \in$

$\mathrm{pr}(\mathrm{argmin}_{(x,y,\gamma)} l(x, y, \gamma, 0, r))$ for all $r \geq \hat{r}$.
Next we show that there is $r^* > \hat{r} + 1$ such that

$$\mathrm{pr}(\mathrm{argmin}_{(x,y,\gamma)} l(x, y, \gamma, 0, r)) \subset \underset{x \in M}{\mathrm{argmin}}\, f(x).$$

Suppose to the contrary that there is $\hat{r} + 1 < r_k \to \infty$ and

$$(x^k, y^k, \gamma^k) \in \mathrm{argmin}_{(x,y,\gamma)} l(x, y, \gamma, 0, r_k)$$

such that $x^k \notin \mathrm{argmin}_{x \in M} f(x)$. The rest of the proof follows exactly as in [60]. $\square$

Next we introduce the concept of the strong stationary point of an MPCC problem, see [123].

Definition 3.6.4. *A vector (x, y, γ) feasible for problem (3.85) is called a strongly stationary point of the problem (3.85) if there is an element $(\alpha, \beta, \eta, \zeta) \in R \times R^m \times R^l \times R^l$ such that*

$$\begin{aligned}
\nabla_x f(x) + \alpha \nabla_x g(x, y) + \nabla_x \nabla_y^T L(x, y, \gamma)\beta - \sum_{1 \leq i \leq l} \eta_i \nabla_x v_i(x, y) &= 0, \\
\alpha \nabla_y g(x, y) + \nabla_y^2 L(x, y, \gamma)\beta - \sum_{1 \leq i \leq l} \eta_i \nabla_y v_i(x, y) &= 0, \\
-\nabla_y^T v(x, y)\beta + \sum_{1 \leq i \leq l} \zeta_i \mathbf{e}_i &= 0, \\
\alpha \geq 0, \alpha g(x, y) &= 0, \\
(\forall i : v(x, y) < 0) \quad \eta_i &= 0, \\
(\forall i : \gamma_i > 0) \quad \zeta_i &= 0, \\
(\forall i : \gamma_i = v_i(x, y) = 0) \quad \eta_i &\geq 0, \zeta_i \geq 0,
\end{aligned}$$

where $L(x, y, \gamma) = g(x, y) - \sum_{i=1}^l \gamma_i v_i(x, y)$ and $\mathbf{e}_i \in \mathbb{R}^l$ is the unit vector with i-th component 1.

Remark 3.6.5. *These above first order necessary optimality conditions involve the second order terms, differing all previously obtained results. As Henrion and Surowiec [49], when studying bilevel problems and mathematical programs with equilibrium constraints, points out, the first order optimality conditions involving the second terms may be more informative of the existence of the multipliers β.*

Consider a special case: Given a strongly stationary point $(\bar{x}, \bar{y}, \bar{\gamma})$ associated with $(\bar{\alpha}, \bar{\beta}, \bar{\eta}, \bar{\zeta})$, let

$$\{i \in \{1, \cdots, l\} : v_i(\bar{x}, \bar{y}) = 0\} = \emptyset.$$

Then from system (3.84)*, we know that* $\bar{\gamma} = 0$ *and from the definition of stationarity,* $\bar{\eta} = 0$*. Also we have that* $\bar{y}$ *is an interior point of* $Y(\bar{x})$*. Since that* $\bar{y}$ *is the optimal solution of the lower level problem* $Q(\bar{x})$*, then* $\nabla_y g(\bar{x}, \bar{y}) = 0$*. Therefore the strongly stationary system is reduced to*

$$\begin{aligned}
\nabla_x f(\bar{x}) + \bar{\alpha}\nabla_x g(\bar{x}, \bar{y}) + \nabla_x \nabla_y^T g(\bar{x}, \bar{y})\bar{\beta} = 0,\\
\bar{\alpha}\nabla_y g(\bar{x}, \bar{y}) + \nabla_y^2 g(\bar{x}, \bar{y})\bar{\beta} = 0,\\
\bar{\alpha} \geq 0, \bar{\alpha} g(\bar{x}, \bar{y}) = 0.
\end{aligned}$$

It is known that $(\bar{x}, \bar{y}, \bar{\gamma})$ is a strongly stationary point if and only if $(\bar{x}, \bar{y}, \bar{\gamma})$ is the KKT point of the following relaxed nonlinear programming

$$\begin{aligned}
\min_{x,y,\gamma} \quad & f(x)\\
s.t. \quad & g(x, y) \leq 0,\\
& \nabla_y g(x, y) - \sum_{i=1}^{l} \gamma_i \nabla_y v_i(x, y) = 0,\\
& \gamma_i = \begin{cases} = 0, & \text{if } \bar{\gamma}_i = 0, v_i(\bar{x}, \bar{y}) < 0,\\ \geq 0, & \text{otherwise,} \end{cases}\\
& v_i(x, y) = \begin{cases} = 0, & \text{if } v_i(\bar{x}, \bar{y}) = 0, \bar{\gamma}_i > 0,\\ \leq 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{3.105}$$

The tightened problem associated with (3.85) is

$$\begin{aligned}
\min_{x,y,\gamma} \quad & f(x)\\
s.t. \quad & g(x, y) \leq 0,\\
& \nabla_y g(x, y) - \sum_{i=1}^{l} \gamma_i \nabla_y v_i(x, y) = 0,\\
& \gamma_i = \begin{cases} = 0, & \text{if } \bar{\gamma}_i = 0,\\ \geq 0, & \text{otherwise,} \end{cases}\\
& v_i(x, y) = \begin{cases} = 0, & \text{if } v_i(\bar{x}, \bar{y}) = 0,\\ \leq 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{3.106}$$

The feasible set of the tightened problem (3.106) is a subset of that of problem (3.85), and the feasible set of (3.85) is that of the relaxed problem (3.105), see, e.g., [123].

Define $G(x, y, \gamma, u)\colon \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^{1+m+3l} \to \mathbb{R}^{1+m+3l}$ by

$$G(x, y, \gamma, u) := \begin{pmatrix} g(x, y) - a \\ -\lambda - c \\ v(x, y) - d \\ \nabla_y g(x, y) - \sum_i \gamma_i \nabla_y v_i(x, y) - b \\ -\lambda v(x, y) - e \end{pmatrix},$$

where $a \in \mathbb{R}$, $b \in \mathbb{R}^m, c, d, e \in \mathbb{R}^l$ and $u^T = (a, b^T, c^T, d^T, e^T)$, and

$$F(x, y, \gamma, u) := f(x) + \delta_{\mathbb{R}_-^{1+2l} \times \mathbf{0}_{\{m+l\}}}(G(x, y, \gamma, u)).$$

When taking $\sigma(u) = (\sum |u_i|)^p$ with $p > 0$,

$$\begin{aligned} l(x, y, \gamma, 0, r) = f(x) + r[g(x, y)_+ + \sum_i ((-\lambda_i)_+ + (v_i(x, y))_+ + |\lambda_i v_i(x, y)|) \\ + \|\nabla_y g(x, y) - \sum_i \gamma_i \nabla_y v_i(x, y)\|_1]^p, \end{aligned}$$

if (x, y, γ) is feasible for problem (3.85), otherwise, is equal to $+\infty$.

Proposition 3.6.2. *Let $\bar{\omega} = 0$ support an exact penalty representation for GSIP problem* (3.82)*, $p \in (\frac{1}{2}, 1]$ and $\sigma(u) = (\sum |u_i|)^p$. Assume that g and $\nabla_y v$ are $C^{1,1}$. Then at any solution $\bar{x}$ for* (3.82)*, there is $(\bar{y}, \bar{\gamma})$ such that $(\bar{x}, \bar{y}, \bar{\gamma})$ is a strongly stationary point.*

Proof. Since $\bar{x}$ is a solution of (3.82), then there is $(\bar{y}, \bar{\gamma})$ such that $(\bar{x}, \bar{y}, \bar{\gamma})$ is a solution of problem (3.85). From the definition of exact penalization, $(\bar{x}, \bar{y}, \bar{\gamma})$ minimizes the function $l(x, y, \gamma, 0, r)$ for r large enough, that is l is exact at $(\bar{x}, \bar{y}, \bar{\gamma})$ for r large enough. For any $p \in (\frac{1}{2}, 1]$, since the smoothness properties hold, p belongs to the indication set in Proposition 3.5 of [92]. Then the conclusion follows from Theorem 3.4 of [92]. □

Let $z := (x, y, \gamma) \in \mathbb{R}^{n+m+l}, d \in \mathbb{R}^{n+m+l}$ and

$$\begin{aligned} I_1 =& \{i : \gamma_i = 0, v_i(x, y) < 0\}, \\ I_2 =& \{i : \gamma_i > 0, v_i(x, y) = 0\}, \\ I_3 =& \{i : \gamma_i = v_i(x, y) = 0\}. \end{aligned}$$

Definition 3.6.5. *The second order necessary condition is satisfied at z if it is a strongly stationary point and*

$$d^T \left[\begin{pmatrix} \nabla_{xx} f(x) & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} + \alpha \begin{pmatrix} \nabla_{xx} g(x,y) & \nabla_{xy} g(x,y) & 0 \\ \nabla_{xy} g(x,y) & \nabla_{yy} g(x,y) & 0 \\ 0 & 0 & 0 \end{pmatrix} + \nabla_{zz}(\nabla_y^T L(z)\beta) \right.$$
$$\left. + \sum_{i \in I_2 \cup I_3} \eta_i \begin{pmatrix} \nabla_{xx} v_i(x,y) & \nabla_{xy} v_i(x,y) & 0 \\ \nabla_{xy} v_i(x,y) & \nabla_{yy} v_i(x,y) & 0 \\ 0 & 0 & 0 \end{pmatrix} \right] d \geq 0 \quad (3.107)$$

for all $d \in R^{n+m+l}$ being the critical direction, that is a direction $d \in \mathbb{R}^{n+m+l}$ satisfying $\langle \nabla f(x), d \rangle = 0$ and

$$\langle \nabla g(x,y), d \rangle \leq 0, \text{ if } g(x,y) = 0,$$
$$\langle \nabla(\nabla_y L(x,y,\gamma), d \rangle = 0,$$
$$\min\{\langle -\nabla v_i(x,y), d \rangle, i \in I_2 \cup I_3; \langle \nabla \gamma_j, d \rangle, j \in I_1 \cup I_3\} = 0.$$

The *strong Mangasarian Fromovitz constraint qualificaiton (SMFCQ)* holds at feasible point $z = (x, y, \gamma)$ if the SMFCQ holds at z for the tightened problem (3.106), that is there exists $(\alpha, \beta, \eta, \zeta)$ such that the vectors $\{\nabla(\nabla_y L(x,y,\gamma); \nabla g(x,y)$, if $\alpha > 0$; $\nabla \gamma_i, i : \gamma_i = 0; \nabla v_i(x,y), i : v_i(x,y) = 0\}$ are linearly independent and there exists $d \in \mathbb{R}^{n+m+l}$ orthogonal to these vectors such that $\nabla g(x,y)^T d < 0$, if $\alpha = g(x,y) = 0$.

Theorem 3.6.4 ([123])**.** (i) *At a local optimal solution z, if the SMFCQ is satisfied at z, then there exists a unique multiplier $(\alpha, \beta, \eta, \zeta)$ satisfying the strongly stationary conditions in Definition 3.6.4 such that the second order necessary condition* (3.107) *holds at z.*

(ii) *Let z be a strongly stationary point. If for every critical direction d there exists a multiplier $(\alpha, \beta, \eta, \zeta)$ satisfying the system of the strongly stationary point in Definition 3.6.4 and satisfying* (3.107) *with strict inequality, then z is a strict optimal solution.*

Chapter 4

Conclusions

This thesis contains various developments of the first order necessary optimality conditions of standard and generalized semi-infinite programming problems. One distinction of the thesis is the fully exploitation of roles of penalization techniques in terms of the fundamental theory of semi-infinite programming problems.

Firstly, by introducing two types of lower order penalty functions for standard semi-infinite programming problems, we establish relationship between the exactness of penalty functions and corresponding calmness conditions. Under exact penalization and some second order constraints assumptions, we derive necessary optimality conditions which are slightly weaker than the known KKT optimality conditions. The same technique is applied to generalized semi-infinite programming problems which are first transformed into equivalent semi-infinite programming problems via augmented Lagrangians of the lower level problems.

Secondly, we consider the derivation of optimality conditions of non-smooth generalized semi-infinite programming problems via the lower level value function reformulation. This approach heavily depend on the estimates of lower level value functions serving as the marginal functions of parametric optimization problems, especially the estimates of generalized subdifferentials. Under either min-max scheme or partial exactness of penalty functions, we derive optimality conditions for generalized semi-infinite programming problems involving the basic/limiting subdifferentials or Clarke generalized subdifferentials.

Thirdly, we consider a approximation scheme for generalized semi-infinite programming problems via the penalty functions of the lower level problems. It is known that the feasible set of generalized semi-infinite programming problems is in general with complicated structures and the value function of the lower level problem is typically improper. The penalty functions of simple constructions seem to provide good candidates for approximations of the irregular feasible set. Without any regularity assumptions on the feasible set, we can effectively approximate the generalized semi-infinite programming problem restricted to the closure of its feasible set.

Finally, we consider generalized semi-infinite programming problems with convex lower level problems which nevertheless entail no convexity of the feasible sets. Under mild assumptions, the generalized semi-infinite programming problems are transformed into mathematical programs with complementarity constraints via the optimality systems of lower level problems. We thus establish the strong duality theory and exact penalization representation via the generalized augmented Lagrangians for generalized semi-infinite programming problems. Combining lower order penalization, we also derive the optimality conditions involving the second order derivatives of the defining functions.

Bibliography

[1] J. Abadie, *On the Kuhn-Tucker theorem*, in Nonlinear Programming (NATO Summer School, Menton, 1964), North-Holland, Amsterdam, 1967, pp. 19–36.

[2] E. J. Anderson and P. Nash, *Linear Programming in Infinite-Dimensional Spaces: Theory and Applications*, Wiley-Interscience Series in Discrete Mathematics and Optimization, John Wiley & Sons Ltd., Chichester, 1987.

[3] A. Auslender, M. A. Goberna, and M. A. López, *Penalty and smoothing methods for convex semi-infinite programming*, Math. Oper. Res., 34 (2009), pp. 303–319.

[4] A. Ben-Tal, E. E. Rosinger, and A. Ben-Israel, *A Helly-type theorem and semi-infinite programming*, in Constructive approaches to mathematical models (Proc. Conf. in honor of R. J. Duffin, Pittsburgh, Pa., 1978), New York, 1979, Academic Press, pp. 127–135.

[5] A. Ben-Tal, M. Teboulle, and J. Zowe, *Second order necessary optimality conditions for semi-infinite programming problems*, in Semi-infinite programming (Proc. Workshop, Bad Honnef, 1978), vol. 15 of Lecture Notes in Control and Information Sci., Springer, Berlin, 1979, pp. 17–30.

[6] D. P. Bertsekas, *Constrained Optimization and Lagrange Multiplier Methods*, Computer Science and Applied Mathematics, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1982.

[7] D. P. Bertsekas and A. E. Ozdaglar, *Pseudonormality and a Lagrange multiplier theory for constrained optimization*, J. Optim. Theory Appl., 114 (2002), pp. 287–343.

[8] J. F. Bonnans and A. Shapiro, *Perturbation Analysis of Optimization Problems*, Springer Series in Operations Research, Springer-Verlag, New York, 2000.

[9] J. M. Borwein, *Direct theorems in semi-infinite convex programming*, Math. Program., 21 (1981), pp. 301–318.

[10] T. Bröcker, *Differentiable Germs and Catastrophes, Translated from the German, last chapter and bibliography by L. Lander*, vol. 17 of London Mathematical Society Lecture Note Series, Cambridge University Press, Cambridge, 1975.

[11] B. Brosowski, *Parametric Semi-Infinite Optimization*, Verlag Peter Lang, Frankfurt, 1982.

[12] J. V. Burke, *Calmness and exact penalization*, SIAM J. Control Optim., 29 (1991), pp. 493–497.

[13] J. V. Burke, *An exact penalization viewpoint of constrained optimization*, SIAM J. Control Optim., 29 (1991), pp. 968–998.

[14] M. J. Cánovas, A. L. Dontchev, M. A. López, and J. Parra, *Metric regularity of semi-infinite constraint systems*, Math. Program., 104 (2005), pp. 329–346.

[15] ——, *Isolated calmness of solution mappings in convex semi-infinite optimization*, J. Math. Anal. Appl., 350 (2009), pp. 829–837.

[16] M. J. Cánovas, D. Klatte, M. A. López, and J. Parra, *Metric regularity in convex semi-infinite optimization under canonical perturbations*, SIAM J. Optim., 18 (2007), pp. 717–732 (electronic).

[17] A. Charnes, W. W. Cooper, and K. Kortanek, *Duality, Haar programs, and finite sequence spaces*, Proc. Nat. Acad. Sci. U.S.A., 48 (1962), pp. 783–786.

[18] B. Chen, X. Chen, and C. Kanzow, *A penalized Fischer-Burmeister NCP-function*, Math. Program., 88 (2000), pp. 211–216.

[19] Y. Chen and M. Florian, *The nonlinear bilevel programming problem: formulations, regularity and optimality conditions*, Optimization, 32 (1995), pp. 193–209.

[20] T. D. Chuong, N. Q. Huy, and J.-C. Yao, *Subdifferentials of marginal functions in semi-infinite programming*, SIAM J. Optim., 20 (2009), pp. 1462–1477.

[21] F. H. Clarke, *Generalized gradients and applications*, Trans. Amer. Math. Soc., 205 (1975), pp. 247–262.

[22] ——, *Optimization and Nonsmooth Analysis*, Canadian Mathematical Society Series of Monographs and Advanced Texts, John Wiley & Sons Inc., New York, 1983.

[23] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, and P. R. Wolenski, *Nonsmooth Analysis and Control Theory*, vol. 178 of Graduate Texts in Mathematics, Springer-Verlag, New York, 1998.

[24] A. R. Conn and N. I. M. Gould, *An exact penalty function for semi-infinite programming*, Math. Program., 37 (1987), pp. 19–40.

[25] I. D. Coope and C. J. Price, *Exact penalty function methods for nonlinear semi-infinite programming*, in Semi-infinite programming, R. Reemtsen and J.-J. Rückmann, eds., vol. 25 of Nonconvex Optim. Appl., Kluwer Acad. Publ., Boston, MA, 1998, pp. 137–157.

[26] S. Dempe, J. Dutta, and B. S. Mordukhovich, *New necessary optimality conditions in optimistic bilevel programming*, Optimization, 56 (2007), pp. 577–604.

[27] S. Dempe and A. B. Zemkoho, *The generalized Mangasarian-Fromowitz constraint qualification and optimality conditions for bilevel programs*, J. Optim. Theory Appl., 148 (2011), pp. 46–68.

[28] N. Dinh, B. Mordukhovich, and T. T. A. Nghia, *Subdifferentials of value functions and optimality conditions for DC and bilevel infinite and semi-infinite programs*, Math. Program., 123 (2010), pp. 101–138.

[29] I. I. Eremin, *The penalty method in convex programming*, Cybernetics, 3 (1967), pp. 53–56.

[30] A. V. Fiacco and K. O. Kortanek, eds., *Semi-infinite programming and applications*, vol. 215 of Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, 1983.

[31] A. V. Fiacco and G. P. McCormick, *Nonlinear programming: Sequential unconstrained minimization techniques*, John Wiley and Sons, Inc., New York-London-Sydney, 1968.

[32] A. Fischer, *A special Newton-type optimization method*, Optimization, 24 (1992), pp. 269–284.

[33] R. Fletcher, *Practical Methods of Optimization*, A Wiley-Interscience Publication, John Wiley & Sons Ltd., Chichester, second ed., 1987.

[34] J. Gauvin and F. Dubeau, *Differential properties of the marginal function in mathematical programming*, Math. Programming Stud., (1982), pp. 101–119.

[35] K. Glashoff and S.-Å. Gustafson, *Linear Optimization and Approximation*, vol. 45 of Applied Mathematical Sciences, Springer-Verlag, New York, 1983.

[36] M. A. Goberna and M. A. López, *Linear Semi-Infinite Optimization*, vol. 2 of Wiley Series in Mathematical Methods in Practice, John Wiley & Sons Ltd., Chichester, 1998.

[37] M. Á. Goberna and M. A. López, eds., *Semi-Infinite Programming: Recent Advances*, vol. 57 of Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Dordrecht, 2001.

[38] M. A. Goberna, M. A. López, and J. Pastor, *Farkas-Minkowski systems in semi-infinite programming*, Appl. Math. Optim., 7 (1981), pp. 295–308.

[39] T. Graettinger, *Maneuverability constraints for design and control of robotic systems: A semi-infinite programming approach*, master's thesis, Carnegie-Mellon University, 1986.

[40] T. Graettinger and B. Krogh, *The acceleration radius: a global performance measure for robotic manipulators*, Robotics and Automation, IEEE Journal of, 4 (1988), pp. 60–69.

[41] J. Guddat, F. Guerra Vazquez, and H. T. Jongen, *Parametric Optimization: Singularities, Pathfollowing and Jumps*, B. G. Teubner, Stuttgart, 1990.

[42] J. Guddat, H. T. Jongen, and J. Rückmann, *On stability and stationary points in nonlinear optimization*, J. Austral. Math. Soc. Ser. B, 28 (1986), pp. 36–56.

[43] F. Guerra-Vázquez, H. T. Jongen, and V. Shikhman, *General semi-infinite programming: symmetric Mangasarian-Fromovitz constraint qualification and the closure of the feasible set*, SIAM J. Optim., 20 (2010), pp. 2487–2503.

[44] F. Guerra Vázquez and J.-J. Rückmann, *Extensions of the Kuhn-Tucker constraint qualification to generalized semi-infinite programming*, SIAM J. Optim., 15 (2005), pp. 926–937 (electronic).

[45] F. Guerra Vázquez, J.-J. Rückmann, O. Stein, and G. Still, *Generalized semi-infinite programming: a tutorial*, J. Comput. Appl. Math., 217 (2008), pp. 394–419.

[46] H. Günzel, H. T. Jongen, and O. Stein, *On the closure of the feasible set in generalized semi-infinite programming*, CEJOR Cent. Eur. J. Oper. Res., 15 (2007), pp. 271–280.

[47] H. Günzel, H. T. Jongen, and O. Stein, *Generalized semi-infinite programming: the symmetric reduction ansatz*, Optim. Lett., 2 (2008), pp. 415–424.

[48] S. P. Han and O. L. Mangasarian, *Exact penalty functions in nonlinear programming*, Math. Program., 17 (1979), pp. 251–269.

[49] R. Henrion and T. Surowiec, *On calmness conditions in convex bilevel programming*, Appl. Anal., 90 (2011), pp. 951–970.

[50] M. R. Hestenes, *Multiplier and gradient methods*, J. Optim. Theory Appl., 4 (1969), pp. 303–320.

[51] R. Hettich, ed., *Semi-Infinite Programming*, vol. 15 of Lecture Notes in Control and Information Sciences, Berlin, 1979, Springer-Verlag.

[52] R. Hettich and H. T. Jongen, *Semi-infinite programming: conditions of optimality and applications*, in Optimization techniques (Proc. 8th IFIP Conf.,

Würzburg, 1977), Part 2, vol. 7 of Lecture Notes in Control and Information Sci., Springer, Berlin, 1978, pp. 1–11.

[53] R. Hettich and K. O. Kortanek, *Semi-infinite programming: theory, methods, and applications*, SIAM Rev., 35 (1993), pp. 380–429.

[54] R. Hettich and G. Still, *Semi-infinite programming models in robotics*, in Parametric optimization and related topics, II (Eisenach, 1989), vol. 62 of Math. Res., Akademie-Verlag, Berlin, 1991, pp. 112–118.

[55] R. Hettich and G. Still, *Second order optimality conditions for generalized semi-infinite programming problems*, Optimization, 34 (1995), pp. 195–211.

[56] R. Hettich and P. Zencke, *Numerische Methoden der Approximation und semi-infiniten Optimierung*, B. G. Teubner, Stuttgart, 1982.

[57] J.-B. Hiriart-Urruty, J.-J. Strodiot, and V. H. Nguyen, *Generalized Hessian matrix and second-order optimality conditions for problems with $C^{1,1}$ data*, Appl. Math. Optim., 11 (1984), pp. 43–56.

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

[59] W. W. Hogan, *Point-to-set maps in mathematical programming*, SIAM Rev., 15 (1973), pp. 591–603.

[60] X. X. Huang and X. Q. Yang, *A unified augmented Lagrangian approach to duality and exact penalization*, Math. Oper. Res., 28 (2003), pp. 533–552.

[61] N. Q. Huy, N. D. Giang, and J.-C. Yao, *Subdifferential of optimal value functions in nonlinear infinite programming*, Appl. Math. Optim., 65 (2012), pp. 91–109.

[62] N. Q. Huy and D. S. Kim, *Stability and augmented Lagrangian duality in nonconvex semi-infinite programming*, Nonlinear Anal., 75 (2012), pp. 163–176.

[63] N. Q. Huy and J.-C. Yao, *Semi-infinite optimization under convex function perturbations: Lipschitz stability*, J. Optim. Theory Appl., 148 (2011), pp. 237–256.

[64] A. D. Ioffe, *Necessary and sufficient conditions for a local minimum. I. A reduction theorem and first order conditions*, SIAM J. Control Optim., 17 (1979), pp. 245–250.

[65] ——, *Necessary and sufficient conditions for a local minimum. II. Conditions of Levitin-Miljutin-Osmolovskiĭ type*, SIAM J. Control Optim., 17 (1979), pp. 251–265.

[66] ——, *Necessary and sufficient conditions for a local minimum. III. Second order conditions and augmented duality*, SIAM J. Control Optim., 17 (1979), pp. 266–288.

[67] ——, *Second order conditions in nonlinear nonsmooth problems of semi-infinite programming*, in Semi-infinite programming and applications (Austin, Tex., 1981), vol. 215 of Lecture Notes in Econom. and Math. Systems, Springer, Berlin, 1983, pp. 262–280.

[68] H. T. Jongen, J.-J. Rückmann, and O. Stein, *Generalized semi-infinite optimization: a first order optimality condition and examples*, Math. Program., 83 (1998), pp. 145–158.

[69] H. T. Jongen and V. Shikhman, *Generalized semi-infinite programming: the nonsmooth symmetric reduction ansatz*, SIAM J. Optim., 21 (2011), pp. 193–211.

[70] H. T. Jongen, F. Twilt, and G.-W. Weber, *Semi-infinite optimization: structure and stability of the feasible set*, J. Optim. Theory Appl., 72 (1992), pp. 529–552.

[71] H. T. Jongen, W. Wetterling, and G. Zwier, *On sufficient conditions for local optimality in semi-infinite programming*, Optimization, 18 (1987), pp. 165–178.

[72] C. Kanzow and A. Schwartz, *Mathematical programs with equilibrium constraints: enhanced Fritz-John conditions, new constraint qualifications, and improved exact penalty results*, SIAM J. Optim., 20 (2010), pp. 2730–2753.

[73] A. Kaplan and R. Tichatschke, *On a class of terminal variational problems*, in Parametric optimization and related topics, IV (Enschede, 1995), vol. 9 of Approx. Optim., Lang, Frankfurt am Main, 1997, pp. 185–199.

[74] ——, *On the numerical treatment of a class of semi-infinite terminal problems*, Optimization, 41 (1997), pp. 1–36.

[75] H. Kawasaki, *An envelope-like effect of infinitely many inequality constraints on second-order necessary conditions for minimization problems*, Math. Program., 41 (1988), pp. 73–96.

[76] ——, *The upper and lower second order directional derivatives of a sup-type function*, Math. Program., 41 (1988), pp. 327–339.

[77] ——, *Second order necessary optimality conditions for minimizing a sup-type function*, Math. Program., 49 (1990), pp. 213–229.

[78] D. Klatte, *Stability of stationary solutions in semi-infinite optimization via the reduction approach*, in Advances in optimization (Lambrecht, 1991), vol. 382 of Lecture Notes in Econom. and Math. Systems, Springer, Berlin, 1992, pp. 155–170.

[79] D. Klatte and B. Kummer, *Nonsmooth Equations in Optimization. Regularity, Calculus, Methods and Applications*, vol. 60 of Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Dordrecht, 2002.

[80] V. L. Klee, Jr., *The critical set of a convex body*, Amer. J. Math., 75 (1953), pp. 178–188.

[81] M. Kojima, *Strongly stable stationary solutions in nonlinear programs*, in Analysis and computation of fixed points (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1979), vol. 43 of Publ. Math. Res. Center Univ. Wisconsin, Academic Press, New York, 1980, pp. 93–138.

[82] W. Krabs, *Nonlinear optimization and approximation*, J. Approximation Theory, 9 (1973), pp. 316–326.

[83] ——, *Optimization and Approximation*, Translated from the German by Philip M. Anselone and Ronald B. Guenther, A Wiley-Interscience Publication, John Wiley & Sons Ltd., Chichester, 1979.

[84] ——, *On time-minimal heating or cooling of a ball*, in Numerical methods of approximation theory, Vol. 8 (Oberwolfach, 1986), vol. 81 of Internat. Schriftenreihe Numer. Math., Birkhäuser, Basel, 1987, pp. 121–131.

[85] S. Lang, *Real and Functional Analysis*, vol. 142 of Graduate Texts in Mathematics, Springer-Verlag, New York, third ed., 1993.

[86] E. Levitin and R. Tichatschke, *A branch-and-bound approach for solving a class of generalized semi-infinite programming problems*, J. Global Optim., 13 (1998), pp. 299–315.

[87] E. S. Levitin, *On the reduction of nonconvex problems of generalized semi-infinite mathematical programming to convex problems of semi-infinite programming*, Automation and Remote Control, 59 (1998), pp. 22–27.

[88] W. Li, C. Nahak, and I. Singer, *Constraint qualifications for semi-infinite systems of convex inequalities*, SIAM J. Optim., 11 (2000), pp. 31–52.

[89] M. López and G. Still, *Semi-infinite programming*, European J. Oper. Res., 180 (2007), pp. 491–518.

[90] M. A. López and E. Vercher, *Optimality conditions for nondifferentiable convex semi-infinite programming*, Math. Program., 27 (1983), pp. 307–319.

[91] Z.-Q. Luo, J.-S. Pang, D. Ralph, and S.-Q. Wu, *Exact penalization and stationarity conditions of mathematical programs with equilibrium constraints*, Math. Program., 75 (1996), pp. 19–76.

[92] K. W. Meng and X. Q. Yang, *Optimality conditions via exact penalty functions*, SIAM J. Optim., 20 (2010), pp. 3208–3231.

[93] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation. I: Basic Theory*, vol. 330 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2006.

[94] ——, *Variational Analysis and Generalized Differentiation. II: Applications*, vol. 331 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 2006.

[95] B. S. Mordukhovich and N. M. Nam, *Variational stability and marginal functions via generalized differentiation*, Math. Oper. Res., 30 (2005), pp. 800–816.

[96] B. S. Mordukhovich, N. M. Nam, and H. M. Phan, *Variational analysis of marginal functions with applications to bilevel programming*, J. Optim. Theory Appl., 152 (2012), pp. 557–586.

[97] B. S. Mordukhovich, N. M. Nam, and N. D. Yen, *Subgradients of marginal functions in parametric mathematical programming*, Math. Program., 116 (2009), pp. 369–396.

[98] B. S. Mordukhovich and T. T. A. Nghia, *DC optimization approach to metric regularity of convex multifunctions with applications to infinite systems*, J. Optim. Theory Appl., 155 (2012), pp. 762–784.

[99] ——, *Subdifferentials of nonconvex supremum functions and their applications to semi-infinite and infinite programs with Lipschitzian data*, SIAM J. Optim., 23 (2013), pp. 406–431.

[100] B. S. Mordukhovich and H. M. Phan, *Tangential extremal principles for finite and infinite systems of sets, I: basic theory*, Math. Program., 136 (2012), pp. 3–30.

[101] ——, *Tangential extremal principles for finite and infinite systems of sets II: applications to semi-infinite and multiobjective optimization*, Math. Program., 136 (2012), pp. 31–63.

[102] B. S. Mordukhovich and R. T. Rockafellar, *Second-order subdifferential calculus with applications to tilt stability in optimization*, SIAM J. Optim., 22 (2012), pp. 953–986.

[103] L. Neralić and O. Stein, *On regular and parametric data envelopment analysis*, Math. Methods Oper. Res., 60 (2004), pp. 15–28.

[104] T. Pietrzykowski, *An exact potential method for constrained maxima*, SIAM J. Numer. Anal., 6 (1969), pp. 299–304.

[105] ——, *The potential method for conditional maxima in the locally compact metric spaces*, Numer. Math., 14 (1969/1970), pp. 325–329.

[106] ——, *A generalization of the potential method for conditional maxima on the Banach, reflexive spaces*, Numer. Math., 18 (1971/72), pp. 367–372.

[107] E. Polak, *On the mathematical foundations of nondifferentiable optimization in engineering design*, SIAM Rev., 29 (1987), pp. 21–89.

[108] E. Polak, *Optimization: Algorithms and Consistent Approximations*, vol. 124 of Applied Mathematical Sciences, Springer-Verlag, New York, 1997.

[109] E. Polak, J. E. Higgins, and D. Q. Mayne, *A barrier function method for minimax problems*, Math. Program., 54 (1992), pp. 155–176.

[110] E. Polak and J. O. Royset, *On the use of augmented Lagrangians in the solution of generalized semi-infinite min-max problems*, Comput. Optim. Appl., 31 (2005), pp. 173–192.

[111] B. N. Pshenichnyi, *Necessary conditions for an extremum*, vol. 4 of Pure and Applied Mathematics. Translated from the Russian by Karol Makowski. Translation edited by Lucien W. Neustadt, Marcel Dekker Inc., New York, 1971.

[112] L. Qi, A. Shapiro, and C. Ling, *Differentiability and semismoothness properties of integral functions and their applications*, Math. Program., 102 (2005), pp. 223–248.

[113] R. Reemtsen and J.-J. Rückmann, eds., *Semi-Infinite Programming*, vol. 25 of Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Boston, MA, 1998.

[114] R. T. Rockafellar, *Augmented Lagrange multiplier functions and duality in nonconvex programming*, SIAM J. Control, 12 (1974), pp. 268–285.

[115] ——, *The Theory of Subgradients and Its Applications to Problems of Optimization. Convex and Nonconvex Functions*, vol. 1 of R & E, Heldermann Verlag, Berlin, 1981.

[116] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*, vol. 317 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], Springer-Verlag, Berlin, 1998.

[117] W. W. Rogosinski, *Moments of non-negative mass*, Proc. Roy. Soc. London Ser. A, 245 (1958), pp. 1–27.

[118] A. Rubinov and X. Yang, *Lagrange-type functions in constrained non-convex optimization*, vol. 85 of Applied Optimization, Kluwer Academic Publishers, Boston, MA, 2003.

[119] J.-J. Rückmann and J. A. Gómez, *On generalized semi-infinite programming*, Top, 14 (2006), pp. 1–59.

[120] J. J. Rückmann and A. Shapiro, *First-order optimality conditions in generalized semi-infinite programming*, J. Optim. Theory Appl., 101 (1999), pp. 677–691.

[121] J.-J. Rückmann and A. Shapiro, *Second-order optimality conditions in generalized semi-infinite programming*, Set-Valued Anal., 9 (2001), pp. 169–186.

[122] ——, *Augmented Lagrangians in semi-infinite programming*, Math. Program., 116 (2009), pp. 499–512.

[123] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity*, Math. Oper. Res., 25 (2000), pp. 1–22.

[124] A. Shapiro, *Second-order derivatives of extremal-value functions and optimality conditions for semi-infinite programs*, Math. Oper. Res., 10 (1985), pp. 207–219.

[125] ——, *First and second order optimality conditions and perturbation analysis of semi-infinite programming problems*, in Semi-infinite programming, vol. 25 of Nonconvex Optim. Appl., Kluwer Acad. Publ., Boston, MA, 1998, pp. 103–133.

[126] ——, *Semi-infinite programming, duality, discretization and optimality conditions*, Optimization, 58 (2009), pp. 133–161.

[127] O. Stein, *First-order optimality conditions for degenerate index sets in generalized semi-infinite optimization*, Math. Oper. Res., 26 (2001), pp. 565–582.

[128] ——, *Bi-Level Strategies in Semi-Infinite Programming*, vol. 71 of Nonconvex Optimization and its Applications, Kluwer Academic Publishers, Boston, MA, 2003.

[129] O. Stein and G. Still, *On optimality conditions for generalized semi-infinite programming problems*, J. Optim. Theory Appl., 104 (2000), pp. 443–458.

[130] O. Stein and G. Still, *On generalized semi-infinite optimization and bilevel optimization*, European J. Oper. Res., 142 (2002), pp. 444–462.

[131] ——, *Solving semi-infinite optimization problems with interior point techniques*, SIAM J. Control Optim., 42 (2003), pp. 769–788.

[132] O. Stein and A. Tezel, *The semismooth approach for semi-infinite programming under the reduction ansatz*, J. Global Optim., 41 (2008), pp. 245–266.

[133] ——, *The semismooth approach for semi-infinite programming without strict complementarity*, SIAM J. Optim., 20 (2009), pp. 1052–1072.

[134] O. Stein and A. Winterfeld, *Feasible method for generalized semi-infinite programming*, J. Optim. Theory Appl., 146 (2010), pp. 419–443.

[135] G. Still, *Generalized semi-infinite programming: theory and methods*, European J. Oper. Res., 119 (1999), pp. 301–313.

[136] G. Still, *Solving generalized semi-infinite programs by reduction to simpler problems*, Optimization, 53 (2004), pp. 19–38.

[137] J. Stoer and R. Bulirsch, *Introduction to numerical analysis*, vol. 12 of Texts in Applied Mathematics, Springer-Verlag, New York, third ed., 2002. Translated from the German by R. Bartels, W. Gautschi and C. Witzgall.

[138] J. Warga, *A necessary and sufficient condition for a constrained minimum*, SIAM J. Optim., 2 (1992), pp. 665–667.

[139] G.-W. Weber, *Generalized semi-infinite optimization: on some foundations*, Vychisl. Tekhnol., 4 (1999), pp. 41–61.

[140] W. Wetterling, *Definitheitsbedingungen für relative Extrema bei Optimierungs- und Approximationsaufgaben*, Numer. Math., 15 (1970), pp. 122–136.

[141] X. Q. Yang, *Second-order conditions in $C^{1,1}$ optimization with applications*, Numer. Funct. Anal. Optim., 14 (1993), pp. 621–632.

[142] ——, *Generalized second-order derivatives and optimality conditions*, Nonlinear Anal., 23 (1994), pp. 767–784.

[143] ——, *On second-order directional derivatives*, Nonlinear Anal., 26 (1996), pp. 55–66.

[144] X. Q. Yang and Z. Q. Meng, *Lagrange multipliers and calmness conditions of order p*, Math. Oper. Res., 32 (2007), pp. 95–101.

[145] J. J. Ye and S. Y. Wu, *First order optimality conditions for generalized semi-infinite programming problems*, J. Optim. Theory Appl., 137 (2008), pp. 419–434.

[146] J. J. Ye and D. L. Zhu, *Optimality conditions for bilevel programming problems*, Optimization, 33 (1995), pp. 9–27.

[147] W. I. Zangwill, *Non-linear programming via penalty functions*, Management Sci., 13 (1967), pp. 344–358.

[148] X. Y. Zheng and X. Q. Yang, *Lagrange multipliers in nonsmooth semi-infinite optimization problems*, Math. Oper. Res., 32 (2007), pp. 168–181.